

# Lawrence Berkeley National Laboratory

## Recent Work

### Title

A New Precision Measurement of the Small-scale Line-of-sight Power Spectrum of the Ly $\alpha$  Forest

### Permalink

<https://escholarship.org/uc/item/3gs1h3nf>

### Journal

Astrophysical Journal, 852(1)

### ISSN

0004-637X

### Authors

Walther, M  
Hennawi, JF  
Hiss, H  
[et al.](#)

### Publication Date

2018

### DOI

10.3847/1538-4357/aa9c81

Peer reviewed

# A NEW PRECISION MEASUREMENT OF THE SMALL-SCALE LINE-OF-SIGHT POWER SPECTRUM OF THE Ly $\alpha$ FOREST

MICHAEL WALTHER<sup>1,2,3</sup>, JOSEPH F. HENNAWI<sup>1,2</sup>, HECTOR HISS<sup>2,3</sup>, JOSE OÑORBE<sup>2</sup>, KHEE-GAN LEE<sup>4</sup>, ALBERTO RORAI<sup>5</sup>, JOHN O'MEARA<sup>6</sup>

*Draft version October 2, 2017*

## Abstract

We present a new measurement of the Ly $\alpha$  forest power spectrum at  $1.8 < z < 3.4$  using 74 Keck/HIRES and VLT/UVES high-resolution, high-S/N quasar spectra. We developed a custom pipeline to measure the power spectrum and its uncertainty, which fully accounts for finite resolution and noise, and corrects for the bias induced by masking missing data, DLAs, and metal absorption lines. Our measurement results in unprecedented precision on the small-scale modes  $k > 0.02 \text{ s km}^{-1}$ , inaccessible to previous SDSS/BOSS analyses. It is well known that these high- $k$  modes are highly sensitive to the thermal state of the intergalactic medium, however contamination by narrow metal lines is a significant concern. We quantify the effect of metals on the small-scale power, and find a modest effect on modes with  $k < 0.1 \text{ s km}^{-1}$ . As a result, by masking metals and restricting to  $k < 0.1 \text{ s km}^{-1}$  their impact is completely mitigated. We present an end-to-end Bayesian forward modeling framework whereby mock spectra with the same noise, resolution, and masking as our data are generated from Ly $\alpha$  forest simulations. These mocks are used to build a custom emulator, enabling us to interpolate between a sparse grid of models and perform MCMC fits. Our results agree well with BOSS on scales  $k < 0.02 \text{ s km}^{-1}$  where the measurements overlap. The combination of BOSS' percent level low- $k$  precision with our 5 – 15% high- $k$  measurements, results in a powerful new dataset for precisely constraining the thermal history of the intergalactic medium, cosmological parameters, and the nature of dark matter. The power spectra and their covariance matrices are provided as electronic tables.

*Keywords:* intergalactic medium, cosmology: observations, reionization, quasars: absorption lines

## 1. INTRODUCTION

The Lyman Alpha (Ly $\alpha$ ) forest (Gunn & Peterson 1965; Lynds 1971) is the premier probe of diffuse baryons in the intergalactic medium (IGM) at high redshifts. Its fluctuations can be accurately described in the current  $\Lambda$ CDM framework and while on large scales it is mostly sensitive to cosmology and structure formation, on small scales it probes the thermal state of the IGM due to two effects: Doppler broadening due to thermal motions and pressure smoothing (sometimes called "Jeans" broadening), which affects the underlying baryon distribution and depends on the integrated thermal history of the IGM (Gnedin & Hui 1998; Kulkarni et al. 2015; Oñorbe et al. 2017).

The major process heating the IGM is considered to be reionization. In this evolutionary phase neutral hydrogen and helium are ionized thus finally allowing photons to freely travel through the Universe. Excess energy from ionizing photons is distributed into the intergalactic gas leading to strong thermal evolution during reionization.

Studying the IGM during reionization provides important insight into the physical state of baryons at that time and thus allows to constrain the environment in which galaxies and quasars (QSOs) form.

Hydrogen reionization (as well as He I reionization which has a similar ionization threshold) was mostly complete by redshift  $z = 6$  (Fan et al. 2006; Becker et al. 2015) with a likely midpoint around 8.5 (Planck Collaboration et al. 2016a). Although it is widely accepted that ionizing photons from stars in the first galaxies are responsible for reionizing H I (e.g. Bouwens et al. 2015; Oesch et al. 2014), a competing scenario whereby reionization was driven by faint AGN has also recently been debated (Madau & Haardt 2015, but see D'Aloisio et al. 2017) as larger samples of quasar candidates were obtained (but not always spectroscopically confirmed) at possibly high redshifts (Giallongo et al. 2015 vs. Weigel et al. 2015).

For He II, however, reionization is delayed until  $z \sim 4$  because of its much higher ionization threshold of 54.4 eV. As stars typically not producing enough photons above this energy (Stanway et al. 2016; Topping & Michael Shull 2015) quasars are assumed to be the major source of He II reionization, delaying the full ionization of helium until quasars become sufficiently abundant (Furlanetto & Oh 2008; McQuinn et al. 2009). While observations of the He II Ly $\alpha$  forest show that He II has to be ionized at  $z = 2.7$  (Worseck et al. 2011) with a large fraction already ionized at  $z = 3.4$  (Worseck et al. 2016) there are only indirect constraints on the start of this process by analyses of quasar proximity zones (Bolton

mwalther@physics.ucsb.edu

<sup>1</sup> Physics Department, Brodia Hall, University of California Santa Barbara, Santa Barbara, CA 93106-9530, USA

<sup>2</sup> Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany

<sup>3</sup> International Max Planck Research School for Astronomy & Cosmic Physics at the University of Heidelberg

<sup>4</sup> Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley, CA 94720, USA

<sup>5</sup> Cambridge Institute of Astronomy, Madingley Rd, Cambridge CB3 0HA, UK

<sup>6</sup> Saint Michael's College, One Winooski Park, Colchester, VT 05439, USA

et al. 2010, 2012). Due to the observational challenges of working in the ultraviolet (UV), as well as the universe being mostly opaque to these photons at  $z > 4$ , direct observations of He II at earlier times will have to wait for the next generation of space telescopes. However, the signatures of reionization imprinted on the thermal state of the IGM are still observable.

In the standard picture of thermal evolution cold IGM gas (few K) is strongly heated during H I and He I reionization (by few times 10 000 K), subsequently cools and then experiences additional heating during He II reionization (McQuinn et al. 2009; Compostella et al. 2013; Puchwein et al. 2015; Upton Sanderbeck et al. 2016; McQuinn & Upton Sanderbeck 2016; Oñorbe et al. 2017). The combined effects of photoionization heating, Compton cooling, and adiabatic cooling due to the expansion of the universe lead to a net cooling of intergalactic gas between and after the reionization phases which has so far not been conclusively observed. The combination of those effects also results in a tight power law temperature-density relation for most of the IGM gas (Hui & Gnedin 1997; Puchwein et al. 2015; McQuinn & Upton Sanderbeck 2016) long after reionization events:

$$T(\Delta) = T_0 \Delta^{\gamma-1} \quad (1)$$

for overdensity  $\Delta = \rho/\bar{\rho}$ , temperature at mean density  $T_0$  and an expected slope  $\gamma \approx 1.6$ . During reionization events this slope is expected to be shallower due to the additional photoionization heating which may also result in a large scatter and a more complicated density dependence (Compostella et al. 2013; McQuinn & Upton Sanderbeck 2016). This would be mostly caused by reionization not occurring uniformly, but being intrinsically patchy and therefore leading to significant fluctuations in the ultraviolet background (UVB) (Davies & Furlanetto 2016; Suarez & Pontzen 2017) as well as temperatures (D’Aloisio et al. 2015) during reionization events. Hydrodynamical simulations show that within several hundred Myr the gas relaxes to the tight power law relation of eqn. (1). Unfortunately there is so far no consensus about the thermal evolution of the IGM during and after He II reionization, although many measurements have been performed.

Several statistical properties of the Ly $\alpha$  forest can be used to characterize the thermal state of the IGM typically by constraining the smoothness of the forest. Measurements have been performed using the flux probability density function (PDF) (Bolton et al. 2008; Viel et al. 2009; Lee et al. 2015; Rorai et al. 2017a), wavelet decompositions of the forest (Theuns et al. 2002; Lidz et al. 2010; Garzilli et al. 2012), the curvature of the smoothed transmission (Becker et al. 2011; Boera et al. 2014), the quasar pair phase angle distribution (Rorai et al. 2013, 2017b), and decomposition of the forest into individual absorption lines (Haehnelt & Steinmetz 1998; Schaye et al. 2000; Bryan & Machacek 2000; Ricotti et al. 2000; McDonald et al. 2001; Rudie et al. 2012; Bolton et al. 2014). There is a new measurement of thermal parameters using the last approach (Hiss et al. 2017) on the same dataset we’ll use in this work that also highlights some systematics of this technique (e.g. due to blending of absorption lines).

In the past the different types of thermal state mea-

surements led to very different results with temperatures at mean density varying by a factor of  $\sim 2$  and  $\gamma$  ranging from  $\gamma < 1$  (i.e. an inverted  $T - \rho$  relation) in several PDF analyses (Bolton et al. 2008; Viel et al. 2009) that might need new physics to be explained (Puchwein et al. 2012, but also see Rorai et al. 2017a for an alternative explanation based on different sensitivities of the techniques) to the expected asymptotic value. On the other hand, the curvature technique provides the so far strongest temperature constraint at some characteristic overdensities  $\Delta_*$ , but is insensitive to  $\gamma$  (but see Boera et al. 2016, for an updated approach) and one therefore needs to rely on external measurements of  $\gamma$  to assess  $T_0$ .

To overcome those issues we measure the small-scale (high wavenumber  $k$ ) cutoff of the Ly $\alpha$  forest flux power spectrum to determine the thermal state of the IGM. On smaller scales than this cutoff there is no structure left in the Ly $\alpha$  forest due to a degenerate combination of spectral resolution as well as thermal and cosmological properties of the IGM. But compared to the other methods this is not only a smoothness measurement (as there are constraints on large scales as well) which allows for breaking of degeneracies. While the thermal history of the IGM has the strongest effect on the cutoff scale the small scale power is also sensitive to the nature of dark matter (Viel et al. 2013; Iršič et al. 2017b) and the power spectrum in general can be used to deduce constraints for a variety of cosmological parameters, e.g. the mass of neutrinos (Palanque-Delabrouille et al. 2015; Yeche et al. 2017; Baur et al. 2017).

The existing measurements of the Ly $\alpha$  forest power spectrum can be divided into two groups. Either they were obtained from small datasets of high-resolution spectra (McDonald et al. 2000; Croft et al. 2002; Kim et al. 2004; Viel et al. 2008, 2013) and therefore have sparse redshift sampling as well as lacking precision (especially for large scale modes). Or they have been determined using large datasets from the Sloan Digital Sky Survey (SDSS) (McDonald et al. 2006) or Baryon Oscillation Spectroscopic Survey (BOSS) surveys (Palanque-Delabrouille et al. 2013, hereafter PD+13), but lack the spectroscopic resolution needed to measure the small-scale cutoff of the IGM flux power spectrum. While the latter lead to high precision ( $\sim 2\%$ ) measurements they only weakly constrain the thermal state of the IGM on their own due to the limitation in spectral resolution. On the other hand previous high-resolution measurements have been used to put constraints on the thermal state of the IGM (Zaldarriaga et al. 2001), but typically lack the precision to constrain cosmological parameters. There have also been recent measurements using medium resolution X-SHOOTER data (Iršič et al. 2017c) at  $3 < z < 4.2$  that in principle probe the IGM small scale power. Lastly, there have been HST/Cosmic Origins Spectrograph (COS) observations to probe the low-redshift IGM (Gaikwad et al. 2016) which is not accessible from the ground due to the atmospheric UV cutoff.

In this work we perform a new power spectrum analysis on a large sample of archival high-resolution spectra and combine this with the existing low- and medium-resolution measurements to enable an accurate measurement of thermal evolution in the IGM in a follow-up study (Walther et al. 2017b). This paper is organized

**Table 1**  
UVES spectra from Dall’Aglio et al. (2008)  
used for our analysis

Object	$z_{\text{QSO}}$	median S/N per 6 km s <sup>-1</sup>
HE1341–1020	2.137	58.1
Q0122–380	2.192	56.4
PKS1448–232	2.222	57.4
PKS0237–23	2.224	102.4
HE0001–2340	2.278	65.9
Q0109–3518	2.406	70.0
HE1122–1648	2.407	171.6
HE2217–2818	2.414	93.6
Q0329–385	2.437	58.4
HE1158–1843	2.459	66.7
Q2206–1958	2.567	74.5
Q1232+0815	2.575	45.8
HE1347–2457	2.615	62.0
HS1140+2711	2.628	88.9
Q0453–423	2.663	77.6
PKS0329–255	2.705	48.0
Q1151+068	2.758	49.1
Q0002–422	2.768	75.0
HE0151–4326	2.787	98.1
Q0913+0715	2.788	54.4
Q1409+095	2.843	24.7
HE2347–4342	2.886	152.3
Q1223+178	2.955	33.4
Q0216+08	2.996	36.8
HE2243–6031	3.011	118.8
CTQ247	3.026	69.1
HE0940–1050	3.089	69.6
Q0420–388	3.120	116.2
CTQ460	3.141	40.9
Q2139–4434	3.208	31.2
Q0347–3819	3.229	83.9
PKS2126–158	3.285	63.6
Q1209+0919	3.291	30.2
Q0055–269	3.665	75.7
Q1249–0159	3.668	69.7
Q1621–0042	3.708	77.7
Q1317–0507	3.719	42.0
PKS2000–330	3.786	150.8

as follows. The dataset of high-resolution spectra is described in § 2. There we also discuss our additional treatment of the data to remove metal line contaminants. In § 3 we present our approach to measure the power spectrum and describe our forward modeling procedure required to accurately interpret our measurement. In § 4 we present our new power spectrum measurement, quantify the impact of metal line contamination, and compare our results to previous work. Our conclusions and outlook are given in § 5.

## 2. HIGH-RESOLUTION QUASAR DATASET

In this section we will describe the dataset we used for our measurement. First we explain how our quasar sample was constructed. Then we describe which parts of the selected data were used and how we masked regions of the data to remove contaminants like e.g. metals or damped Ly $\alpha$  absorption systems (DLAs) Finally we will explain how we regulate the mean flux of our spectra.

### 2.1. Dataset

Our measurement of the power spectrum was performed using 38 high-resolution quasar spectra (see Table 1) from Dall’Aglio et al. (2008) observed with the Ultraviolet and Visual Echelle Spectrograph (UVES,

**Table 2**  
HIRES spectra from KODIAQ (O’Meara et al.  
2015) used for our analysis

Object	$z_{\text{QSO}}$	median S/N per 6 km s <sup>-1</sup>
J122824+312837	2.200	87.3
J110610+640009	2.203	58.5
J162645+642655	2.320	103.7
J141906+592312	2.321	36.7
J005814+011530 <sup>c</sup>	2.495	36.2
J162548+264658 <sup>c</sup>	2.518	43.9
J121117+042222	2.526	33.6
J101723–204658	2.545	70.3
J234628+124859	2.573	75.1
J101155+294141 <sup>a</sup>	2.620	129.9
J082107+310751	2.625	64.0
J121930+494052	2.633	90.3
J143500+535953	2.635	65.0
J144453+291905	2.669	133.7
J081240+320808	2.712	48.8
J014516–094517A	2.730	76.8
J170100+641209 <sup>c</sup>	2.735	81.8
J155152+191104	2.830	30.2
J012156+144820	2.870	54.5
Q0805+046 <sup>b</sup>	2.877	26.8
J143316+313126	2.940	53.8
J134544+262506	2.941	34.7
J073621+651313	3.038	25.7
J194455+770552 <sup>a</sup>	3.051	30.4
J120917+113830	3.105	31.4
J114308+345222 <sup>a</sup>	3.146	31.9
J102009+104002	3.168	35.9
J1201+0116 <sup>b</sup>	3.233	30.1
J080117+521034 <sup>a</sup>	3.236	43.2
J095852+120245 <sup>a</sup>	3.298	44.8
J025905+001126	3.365	26.3
Q2355+0108 <sup>b</sup>	3.400	58.3
J173352+540030	3.425	57.3
J144516+095836 <sup>a</sup>	3.530	24.6
J142438+225600 <sup>a</sup>	3.630	29.3
J193957–100241	3.787	65.5

<sup>a</sup> objects are part of DR2, but a pre-DR2 reduction has been used

<sup>b</sup> objects are not part of DR1 or DR2, but reduced in the same way

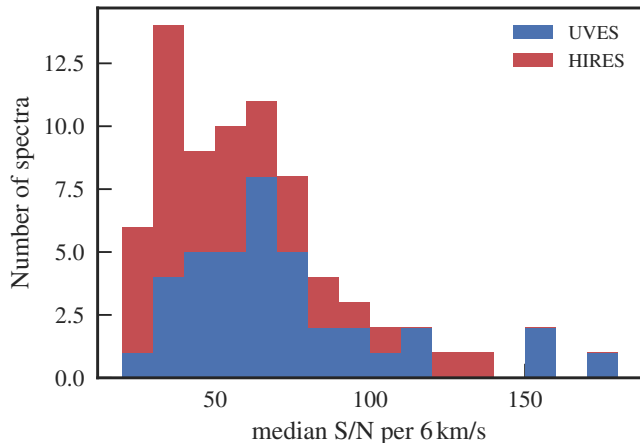
<sup>c</sup> objects are part of DR1, but a pre-DR1 reduction has been used

Dekker et al. 2000) at the Very Large Telescope (VLT), and 36 spectra (see Table 2) from the Keck Observatory Database of Ionized Absorption toward Quasars (KODIAQ) project (Lehner et al. 2014) observed with the High Resolution Echelle Spectrometer (HIRES, Vogt et al. 1994) at Keck. For the latter we used the highest S/N part of DR1 (O’Meara et al. 2015) and additional data beyond DR1 (mostly early reductions of objects in DR2 O’Meara et al. 2017) reduced in the same way.

Most of the UVES spectra have full coverage between the atmospheric cutoff at  $\lambda_{\text{obs}} \sim 3100 \text{ \AA}$  and  $\lambda_{\text{obs}} \sim 1 \mu\text{m}$ . This allows us to use a large range of spectrum redward of the Ly $\alpha$  forest to search for metal lines as well as enabling us to search for Lyman Limit Systems (LLSs) using higher order Lyman series transitions, in many cases even exploiting coverage of the Lyman-limit.

For the KODIAQ data the typical red spectral coverage ends at  $\lambda_{\text{obs}} \sim 6000 \text{ \AA}$ , while the blue spectral cutoff is comparable to UVES  $\lambda_{\text{obs}} \sim 3100 \text{ \AA}$ . For a few cases in both datasets, however, even the Ly $\alpha$  forest was not fully covered.

The objects used in our analysis were chosen to have



**Figure 1.** Histogram of the median S/N per  $6 \text{ km s}^{-1}$  in the Ly $\alpha$  forest region used divided by dataset. Note that we only used spectra with S/N > 20 in the dataset

a median S/N > 20 per  $6 \text{ km s}^{-1}$  interval inside the Ly $\alpha$  forest region covered by the spectra. We also chose to omit spectra with known broad absorption lines (BALs). Finally we omitted sightlines with  $3.0 < z < 3.5$  that were color selected to avoid potential biases due to their increased abundance of LLSs (Worseck & Prochaska 2011).

The distribution of S/N for the dataset used in our analysis is shown in Figure 1. Many objects have a much larger S/N than our cut (up to about a median S/N of 150 per  $6 \text{ km s}^{-1}$  inside the Ly $\alpha$  forest in a sightline). This very high S/N enables us to perform the measurement without strong systematics due to the limited accuracy of our noise model affecting our measurement at the scales of interest.

The nominal FWHM spectral resolution of the data varies between  $3.1 \text{ km s}^{-1}$  and  $6.3 \text{ km s}^{-1}$  with a typical value around  $6 \text{ km s}^{-1}$ . Therefore all Ly $\alpha$  absorption lines are resolved and we expect thermal broadening and broadening due to pressure smoothing to be the effects determining the smoothness of lines, and not smoothing due to finite spectroscopic resolution (see e.g. Bolton et al. 2014, for a measurement of line width for thermally broadened lines).

The data was already reduced and continuum fit, and we briefly summarize the details here. The UVES data were fit with both a global power law in non-absorbed regions and local cubic splines fitted automatically with spline point separations depending on continuum slope. Systematic biases in this technique were estimated and corrected using a Monte Carlo analysis on mock data by Dall’Aglio et al. (2008). The High Resolution Echelle Spectrometer (HIRES) continua were hand-fitted by John O’Meara one echelle order at a time by placing Legendre polynomial anchor points at non-absorbed positions (for estimates of continuum fitting errors, see Kirkman et al. 2005; Faucher-Giguère et al. 2008b). Afterwards a 4th to 12th order polynomial was fit through the anchors. Further details about the reduction and continuum fitting techniques can be found in the respective data papers (Dall’Aglio et al. 2008; O’Meara et al. 2015).

We will use this high-resolution dataset to measure the small scale power spectrum of the Ly $\alpha$  forest in red-

shift bins of size  $\Delta z = 0.2$  with central redshifts between  $\bar{z} = 1.8$  and  $\bar{z} = 3.4$ . Each of the bins contains at least eight quasar spectra, with the majority (all but the low and high redshift edges of our sample) containing more than 14 quasar spectra. The datasets also contain eight spectra that cover higher redshifts which we did not analyze due to the small amount of data available (two spectra cover  $z \gtrsim 4.0$ , the rest only cover the forest for  $z < 3.6$  just half a bin further than our analysis).

In this work and our companion paper (Walther et al. 2017b) we compare our power spectrum measurements to measurements from lower spectral resolution data from BOSS (PD+13) and from the XQ-100 survey Iršič et al. (2017c) based on X-SHOOTER data, and also conduct joint model fits. To facilitate this comparison we use the same binning in redshift from  $\bar{z} = 2.2$  to  $\bar{z} = 3.4$ .

## 2.2. Spectral Masking Procedure

To prepare the data for the power spectrum computation we restrict our attention to the restframe wavelength range of  $1050 \text{ \AA} < \lambda_r < 1180 \text{ \AA}$ . This was done to exclude the Ly $\alpha$  proximity zone, accounting for possibly large redshift errors, as well as to exclude the Ly $\beta$  and O VI  $\lambda 1035$  emission lines and possible blueshifted absorption from these as well as increased continuum fitting errors close to emission lines. This is the same range used in PD+13 and is considered a conservative choice for the Ly $\alpha$  forest region.

We masked parts of the spectrum to reduce contaminations due to low-quality data, high-column density absorbers such as DLAs, and metal lines. If a pixel is already masked during the reduction (due to e.g. cosmic rays or gaps in spectral coverage) it stays masked. According to McDonald et al. (2005) excluding DLAs and LLSs from power spectrum calculation only changes the power by < 2% on the scales of interest for our analysis. We nevertheless excluded absorbers with clearly visible damping wings, i.e. DLAs and super-LLSs, from our spectra to make sure those do not influence the result. For this we masked the core and wings of these strong absorbers by eye until the wings are below the noise level of the spectra. In most cases this leads to the exclusion of big continuous spectral regions at the boundary of a redshift bin or the whole data of this object falling inside a redshift bin. Where a DLA mask removed only the bin center we used the longer of the two remaining spectral regions at the bin boundaries to compute the power. Therefore spectra with DLAs are only shorter, but with no additional gaps in the data. Finally we removed spectra that were shorter than 10% of a redshift bin corresponding to a minimum length of  $\Delta z \sim 0.02$  (or  $3500 \text{ km s}^{-1}$ ) to avoid noisy contributions from short spectra that consist of only few absorption lines.

Metal lines in the Ly $\alpha$  forest are expected to increase the power primarily on small scales ( $k \gtrsim 0.1 \text{ s km}^{-1}$ , see McDonald et al. 2000; Lidz et al. 2010) due to their narrower widths compared to Ly $\alpha$ , but due to correlations induced by the rest-frame velocity separations of different transitions (e.g. between Si III and H I-Ly $\alpha$  or between the C IV doublet lines) there is contamination on larger scales as well (Croft et al. 1999; McDonald et al. 2006).

To reduce the impact of metal absorption inside our sightlines we masked metal lines in the forest region. We first identified metal absorption lines in the Ly $\alpha$  forest by

**Table 3**  
List of metal transitions that were masked

Ion	$\lambda_{\text{rest}}/\text{\AA}$	Metal transition	$\lambda_{\text{rest}}/\text{\AA}$
O VI <sup>a</sup>	1031.926	Si IV <sup>a</sup>	1402.770
C II	1036.337	Si II	1526.707
O VI	1037.617	C IV <sup>a</sup>	1548.195
N II	1083.990	C IV <sup>a</sup>	1550.770
Fe III	1122.526	Fe II	1608.451
Fe II	1144.938	Al II	1670.787
Si II	1190.416	Al III	1854.716
Si II	1193.290	Al III	1862.790
N I	1200.710	Fe II	2344.214
Si III <sup>a</sup>	1206.500	Fe II	2374.461
N V	1238.821	Fe II	2382.765
N V	1242.804	Fe II	2586.650
Si II <sup>a</sup>	1260.422	Fe II	2600.173
O I	1302.169	Mg II	2796.352
Si II	1304.370	Mg II	2803.531
C II	1334.532	Mg I	2852.964
C II*	1335.708	Ca I	3934.777
Si IV <sup>a</sup>	1393.755	Ca I	3969.591

<sup>a</sup> strongest transitions, therefore used for all the masking techniques

having two of the authors (H. Hiss and M. Walther) visually inspect the spectra and mask metal contamination in several ways. First, we look for DLAs inside the forest and mask all strong metal absorption lines corresponding to the DLA redshifts. For this we used all metal transitions in [Table 3](#) and masked a region of  $60 \text{ km s}^{-1}$  in each direction around each identified absorber. Then we search for absorption from common doublet transitions (Si IV, C IV, Mg II, Al III, Fe II) redwards of the forest where the spectrum is mostly clean and mask all associated metal lines analogous to our procedure for DLAs (using the same metal catalog and mask width), until there are no doublet features left redward of the forest. We also masked out a region of  $200 \text{ km s}^{-1}$  in each direction around redshift zero to get rid of metal contamination from the Milky Way (in this case the only relevant transition is Ca II for  $2.33 < z_{\text{QSO}} < 2.78$ ).

Additionally, we used an automated partial LLS (pLLS) finder written by John O’Meara to identify strong absorption systems. This finder works by searching for pixels with zero flux (within some threshold) at the corresponding positions of Ly $\alpha$ ,  $\beta$ ,  $\gamma$  and higher Lyman transitions if available and grouping them into systems of LLS candidates. The candidates were then visually inspected by one of the authors (John O’Meara) and compared to theoretical line profiles of absorbers with  $\log(N_{\text{H I}}) = 15, 16, 17$  in Ly $\alpha$  to Ly $\gamma$ , and systems which appeared consistent were positively identified as pLLSs. For these systems, associated metal absorbers from a reduced line list (see [Table 3](#) absorbers marked with a) were masked with the same velocity window size as above. Note that the hydrogen absorption arising from the pLLSs identified in this way was not masked regardless of their  $N_{\text{H I}}$ .

Lastly we perform a line-fitting analysis (see [Hiss et al. 2017](#)) using a semi-automatic wrapper around VPFIT ([Carswell & Webb 2014](#)) on the same set of Ly $\alpha$  spectra. The result of this is a distribution of line widths  $b$  and column density  $N_{\text{H I}}$  for all fitted lines assuming that all absorption is due to hydrogen. For H I gas at a given column density lines are broadened both thermally (due

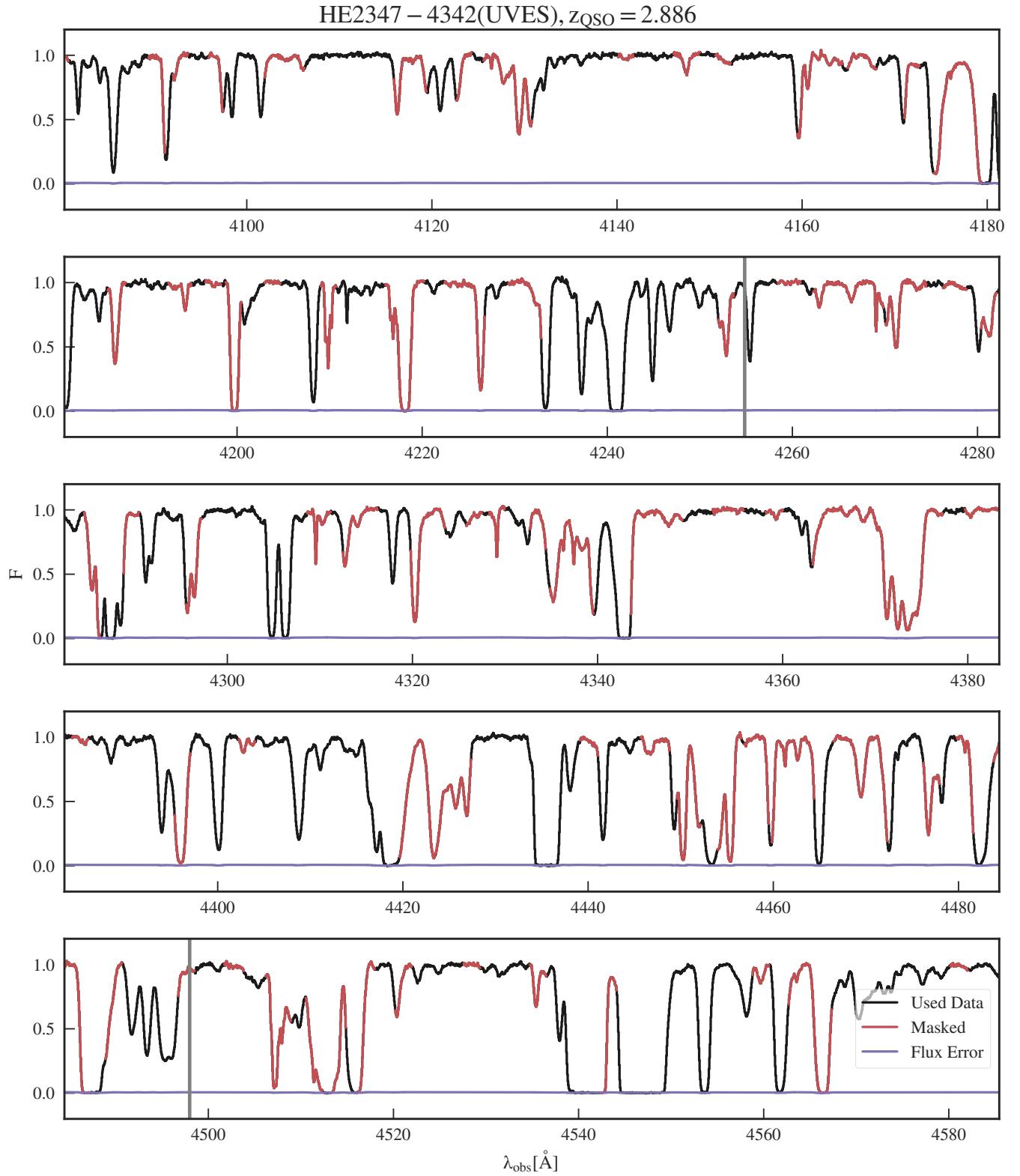
to finite pressure broadening and instantaneous temperature Doppler broadening) and hydrodynamically due to local gas motions. There is a minimum broadening  $b_{\text{cut}}(N_{\text{H I}})$  populated by absorbers with zero line-of-sight peculiar velocity which are purely thermally broadened (see e.g. [Schaye et al. 2000](#); [Hiss et al. 2017](#), for more details). Therefore  $b(N_{\text{H I}})$  should cut off for  $b < b_{\text{cut}}(N_{\text{H I}})$  and all remaining lines narrower than this cutoff can be attributed to fitting artifacts at the edges of strong absorbers, noise fluctuations, or narrow metal absorption lines. This cutoff is fit using an iterative procedure similar to the one used in [Rudie et al. \(2012\)](#). We identified all the lines in previously unmasked spectral regions fulfilling  $b < 11 \text{ km s}^{-1} (N_{\text{H I}}/10^{12.95} \text{ cm}^{-2})^{0.15}$  as these are narrower than the thermal cutoff (see [Hiss et al. 2017](#), for details). For each of these lines (that are in the Ly $\alpha$  forest region) we checked if they could be identified using a second metal transition clearly lining up at the same redshift. In practice, these identifications were most easily made when both metal transitions form a doublet. Given a positive identification, masking was then performed as for metal absorbers that we identified redward of the Ly $\alpha$  forest, i.e. determining the redshift of the absorber and using the full list of metal lines in [Table 3](#).

Neither of these techniques produces a fully metal free Ly $\alpha$  forest, and we briefly elaborate on these limitations. The search for metals redwards of the forest only finds transitions if both doublet counterparts are visible at redder wavelength than the Ly $\alpha$  emission line. If the doublet counterpart of a line falls into a spectral region that was not covered or is blended with a different line (from either a different absorber redshift or e.g. a telluric absorption feature) the contaminating system will often go undetected. For example, for a C IV $\lambda$ 1548/1550 system at  $z > 2.5$  that might contaminate the forest, the Mg II  $\lambda$ 2796/2803 doublet would land at  $\lambda_{\text{obs}} > 1 \mu\text{m}$  and hence outside of the spectral coverage of our spectra, so only if the absorber shows Fe II or Al III doublets would this C IV absorber be masked. On the other hand especially for the higher redshift bins a large fraction of the spectral range redwards of the forest is contaminated by telluric absorption making doublet identification very challenging. The largest problem for this method is therefore limited usable spectral coverage in the red.

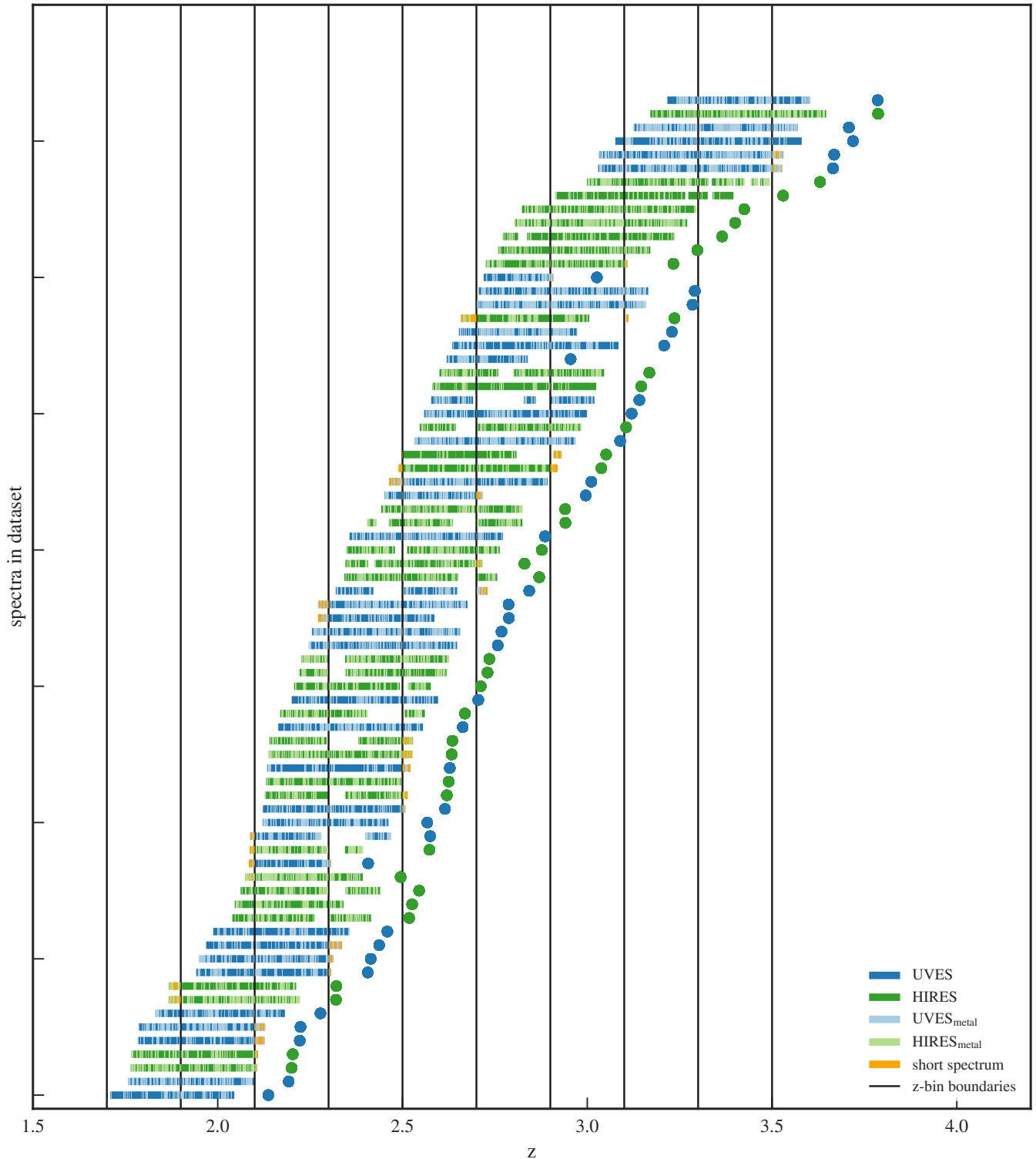
For the automated pLLS finder at least Ly $\alpha$  to Ly $\gamma$  need to be detectable to identify a pLLS. This leads to a minimal redshift of  $z \gtrsim 2.1$  for absorbers that can be identified as they need to be at observed wavelength higher than the atmospheric cutoff  $\lambda_{\text{obs}} \approx 3000 \text{ \AA}$ . For reduced spectral coverage in the blue this minimal redshift is correspondingly higher.

The last of our metal masking procedures only recognizes metal absorption lines significantly narrower than the cutoff in the  $b(N_{\text{H I}})$  distribution and requires a second metal transition for identification. Therefore singlet lines are not masked by this technique unless another transition from a different metal species clearly lines up with them. Also metal absorbers were not removed if all components are broadened above the  $b(N_{\text{H I}})$  threshold adopted, and are therefore not recognized as metal candidates.

In [Figure 2](#) we show an example to illustrate how much of a typical spectrum is masked. The black line shows

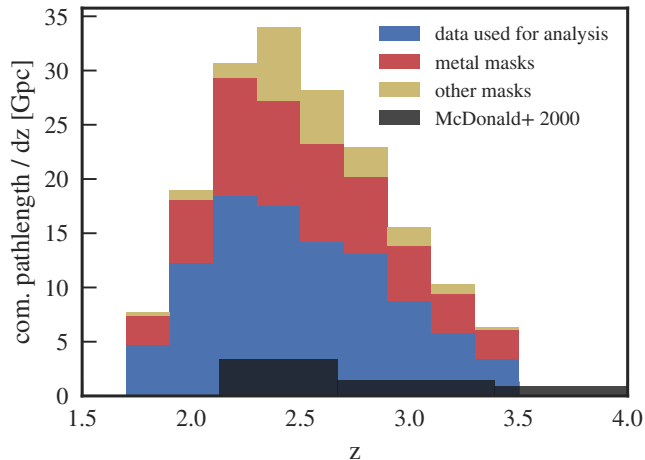


**Figure 2.** The Ly $\alpha$  forest region for one of our spectra (HE2347-4342) with regions masked due to possible metal contamination in red. The purple line shows the error level of the spectrum. Gray vertical lines show the boundaries of our redshift bins. Note that due to our approach not only metals are masked, but also coincidental pieces of the Ly $\alpha$  forest.



**Figure 3.** Redshift coverage of the dataset used colored by spectrograph. Blue (green) lines show the spectral coverage of each used spectrum in the Ly $\alpha$  forest for the UVES (HIRES) subset. Circles mark the corresponding quasar redshifts. Most of the long gaps in this figure are due to missing data between e.g. non-overlapping echelle orders or masked out DLAs while the light colored parts show masks due to possible metal contamination (see § 2.2). Orange lines show regions that were ultimately rejected because of limited forest coverage. Vertical lines mark boundaries of the redshift bins used in the analysis.





**Figure 4.** The pathlength of used data in our analysis compared to the pathlength of spectra masked due to several reasons (metals: red, other: yellow). About 30%-40% of each redshift bin were masked due to possible metal contamination. Far less due to other things, e.g. DLAs, bad reductions or only short available pathlength. As a comparison the amount of data in [McDonald et al. \(2000\)](#) is shown in black for their redshift binning. Normalization is such that the area (and not the height) of each bar corresponds to the total path inside it.

the non-masked part of the spectrum while the red line shows parts that are masked due to possible metal contamination. One can see that many narrow lines in the spectrum are masked, but also that many regions of the forest coincidentally overlap with positions where metal lines associated with our identified absorbers could lie, but that don't actually contain any visible metal absorption.

In [Figure 3](#) we present an overview of the dataset. For each quasar spectrum the emission redshift is shown as well as the full coverage of the spectrum after masking pre-existing gaps in the data and DLAs (which therefore appear as white gaps). The remaining spectrum is divided into the data used (dark colors), masked data due to possible metal contamination (bright colors), and spectra not used due to failing our requirement of spectral extent (yellow). We also show the usable Ly $\alpha$  forest pathlength after applying the full masking procedure compared to the available pathlength when not masking metals in [Figure 4](#). Up to  $\sim 40\%$  of the forest gets removed by applying our metal masking procedure strongly reducing our dataset size and therefore the precision of our measurement results. Although not perfect, our procedure significantly reduces the metal line contamination in our spectra, which decreases the amount of small-scale power compared with an unmasked dataset as we will see in [§ 4.1](#). However, this masking also changes the power spectrum due to the application of a complex window function in configuration space. To correct for this effect we forward model the masking (see [§ 3.4](#)).

### 2.3. Mean Flux Regulation and Continuum Uncertainties

In principle the estimation of quasar continua in the data is subject to errors as well. We perform our power spectrum measurement on the flux contrast

$$\delta_F = \frac{F - \bar{F}}{\bar{F}}. \quad (2)$$

, with  $\bar{F}$  being the mean transmission of the Ly $\alpha$  forest. Because of this, any global misplacement of the continuum will be divided out as long as the mean transmission measurement we divide by is measured on the same spectrum. In addition, incorrect placement of the continuum could lead to gradients or wiggles in the data that could source additional large scale power (on scales  $k < 0.001 \text{ s km}^{-1}$ ). However, this will not strongly impact the small scale power measurement we want to perform in this work<sup>7</sup> ([Lee et al. 2012](#)).

Nevertheless, we perform a mean flux regulation on the dataset using the technique of [Lee et al. \(2012\)](#) which enables us to easily divide out the mean transmission as well as possible gradients in the continuum fits. For this the transmission of the Ly $\alpha$  forest region (excluding masked pixels as well as possible proximity regions as discussed earlier) of each quasar sightline is first fit by a linear relation  $f(z)$  times the mean transmission function  $\bar{F} = \exp(-\tau_{\text{eff}})$  with

$$\tau_{\text{eff}} = C + \tau_0 \left( \frac{1+z}{1+z_0} \right)^\beta \quad (3)$$

following the functional form and parameters from [Becker et al. \(2013\)](#). Afterwards  $f(z)$  is divided out so that the mean flux evolution of each spectrum follows the same relationship.

The flux contrast  $\delta_F$  is now easily obtained by using [eqn. \(2\)](#) on mean flux regulated spectra using [eqn. \(3\)](#) for the mean flux evolution instead of dividing each spectrum by a mean transmission estimate for this spectrum. This allows us to divide out the mean flux at each pixel analytically based on the fit of the mean flux evolution across each spectrum. While this in principle leads to a reduced large-scale power ([Lee et al. 2012](#)), we are not measuring the affected scales in this work.

The resulting spectra are finally divided into our redshift bins of  $\Delta z_{\text{chunk}} = 0.2$  (with the first bin starting at  $z = 1.7$ ) to increase redshift resolution to a level where we could closely monitor thermal evolution. We will henceforth call these pieces of spectra "spectral chunks".

## 3. POWER SPECTRUM MEASUREMENT AND FORWARD MODELING PROCEDURE

In this section we describe our procedure for measuring the flux power spectrum. First we explain how we measure the raw power spectrum. Next we discuss the impact of our masking (especially the masking of metal contaminants) on this measurement. After this we discuss the models we use in most of this work, how we generate mock spectra from them, and approximate the data covariance matrix by combining information from both data and models. Finally, we discuss how we create a fast emulator of our model power spectra, and use this to fit models to our data, allowing us to correct for the masking.

### 3.1. Measuring the Power Spectrum

<sup>7</sup> We performed tests on model simulations at  $z = 3$  including quasar continua. Comparing the power spectrum using the true continuum vs. a hand-fitted continuum showed that only the very largest scales (smallest  $k$ ) were affected by the procedure.

We calculate the power spectrum from the flux contrast  $\delta_F$  defined in [eqn. \(2\)](#). As our spectra are not periodic and are not regularly sampled because of masking, we Fourier transform  $\delta_F$  using a Lomb-Scargle periodogram ([Lomb 1976](#); [Scargle 1982](#)), allowing us to compute the raw power  $P_{\text{raw}}(k)$  of each spectrum for a linearly spaced set of modes from the fundamental mode given by the length of the spectral chunk to the Nyquist limit. We subtract the noise power  $P_{\text{noise}}(k)$  from this raw power, and divide the difference by the window function  $W_R$  resulting from finite spectroscopic resolution  $R$ , following the FFT method described in [PD+13](#):

$$P_{\text{data}}(k) = \left\langle \frac{P_{\text{raw}}(k) - P_{\text{noise}}(k)}{W_R^2(k, R, \Delta v)} \right\rangle, \quad (4)$$

with

$$W_R(k, R, \Delta v) = \exp\left(-\frac{1}{2}(kR)^2\right) \frac{\sin(k\Delta v/2)}{(k\Delta v/2)}, \quad (5)$$

and  $\Delta v$  refers to the pixel scale of the spectra in velocity units.

The noise power  $P_{\text{noise}}$  is measured by creating 100 realizations of Gaussian random noise generated from the  $1\sigma$  error vector of each quasar spectrum. The resolution assumed for the window function correction is the nominal slit-resolution of the spectrograph which is different from the actual spectral resolution of the data, which also depends on the seeing of the observations. For the typical resolution in our dataset (so a resolving power of 50,000 or equivalently  $R = 2.55 \text{ km s}^{-1}$ ) we get  $W_R^2(0.1 \text{ s km}^{-1}) \sim 0.94$ . Given this small correction, even  $\sim 20\%$  error in our knowledge of the resolution (due to the unknown seeing), would only lead to a small  $\lesssim 4\%$  correction of the power at  $k < 0.1 \text{ s km}^{-1}$  (for further discussion of this point see [§ 4.2](#) and [Appendix A](#)). Therefore, the resolution uncertainty does not significantly affect our measurement of the  $k$ -modes we consider.

<sup>8</sup> We chose to use the same logarithmically spaced  $k$ -bins as in [McDonald et al. \(2000\)](#) for our analysis. Note that this is different from the linear spacing adopted by ([PD+13](#)) and ([Iršič et al. 2017b](#)).

We adopt the same Fourier normalization convention as the BOSS measurement ([PD+13](#)), such that the variance in flux contrast is  $\sigma_{\delta_F}^2 = \sigma_F^2/\bar{F}^2 = \int_{-\infty}^{\infty} P(k)dk/(2\pi)$ . Note that this differs from the conventions used by some older high-resolution measurements (see [Appendix C](#)) leading to additional normalization factors needed when comparing to those results.

### 3.2. The Window Function Resulting from Masking

As described in [§ 2.2](#) we masked out parts of the data in real space due to metal contamination. This is a different approach than the one used by [PD+13](#) who estimated the

<sup>8</sup> The averaging  $\langle \dots \rangle$  is performed over the individual periodograms of all spectral chunks inside a redshift bin, and also the average over all modes  $k$  inside logarithmic bins in  $k$ , where equal weights are given to each individual mode from any spectrum. As the fundamental mode of a shorter spectrum is at larger  $k$ , there are less modes available in a given band power for shorter spectra and they are therefore effectively downweighted by performing the average over modes.

metal power from transitions with  $\lambda > \lambda_{\text{Ly}\alpha, \text{rest}}$  by measuring the power redwards of the forest in lower redshift quasar spectra. The power measured in this way was then subtracted from the measurement, but this method can never account for transitions with  $\lambda < \lambda_{\text{Ly}\alpha, \text{rest}}$  which always fall into the forest (e.g. the Si III  $\lambda 1206$  line that leads to the Si III correlation feature in those measurements, see [McDonald et al. 2006](#)).

Masking spectra in configuration space with a window function  $W_m$  leads to the measured  $P_{\text{masked}}$  for each spectrum being effectively a convolution of the true power  $P_{\text{true}}$  with the square of the Fourier transform  $\tilde{W}_m$  of this window function:

$$P_{\text{masked}} = P_{\text{true}} * \tilde{W}_m^2. \quad (6)$$

Determining the true power thus requires deconvolving the window function or adopting a different technique to measure the power taking into account the windowing (like the minimum variance estimator used in [McDonald et al. 2006](#)). In this work we opt to use a simpler approach by generating forward models of the data with and without the windowing applied to be able to determine the effect this window function has on our data. In the end, we correct our data for those effects by dividing the measurement by a correction function based on these models. In the remainder of this section we describe how our forward models are generated, how we estimate the data covariance matrix based on those models, and how we fit the data using those models.

### 3.3. Simulations and Mock Spectra

We use simulations of the Ly $\alpha$  forest for two reasons. First, we need to simulate the effect of noise, resolution and the window function due to masking on the power spectrum. Second, we want to connect the information encoded in the power spectrum to a thermal history of the IGM (parametrized by the temperature at mean density  $T_0$ , the slope of the temperature-density relation  $\gamma$  and the pressure smoothing scale  $\lambda_P$ ). For this purpose, one generally needs to run hydrodynamical simulations with different thermal histories. However, these are computationally expensive and at least for the first point we only need a model that is flexible enough to provide a good fit to the observed power spectra, but it need not necessarily provide the correct thermal parameters. Because dark matter (DM) only simulations of the Ly $\alpha$  forest are more flexible and computationally inexpensive to generate, for all the forward modeling in this paper, we use approximate DM only simulations (a single box with different thermal parameters generated in post-processing) with a semi-numerical approach to paint on the thermal state of the IGM ([Croft et al. 1998](#); [Meiksin & White 2001](#); [Hui & Gnedin 1997](#); [Gnedin & Hui 1998](#); [Gnedin et al. 2003](#); [Rorai et al. 2013](#)) This fast simulation scheme allows us to generate a grid of  $\sim 500$  combinations of thermal parameters in a reasonable amount of time.

These DM-based simulations are however, not sufficiently accurate to infer the thermal state of the IGM, as they produce significant biases in thermal parameters when fitted to mock data based on hydrodynamical simulations (see e.g. [Sorini et al. 2016](#), for a detailed comparison between hydrodynamical and dark matter only

simulations). They do, however, provide a good fit to the data (compare lines and same color errorbars in [Figure 6](#) which we’ll discuss in more detail later) and we therefore use them to correct for the window function in our measurement as well as testing our analysis procedure and our fast power spectrum emulator (interpolation scheme, see [§ 3.6](#)). For inference of IGM thermal parameters a grid of hydrodynamical simulations will be used in a companion paper ([Walther et al. 2017b](#)).

For the DM only simulations we use an updated version of the TreePM code described in [White \(2002\)](#) to simulate the evolution of dark matter particles from initial conditions at  $z = 150$  up to  $z = 1.8$ . We use a simulation with  $L_{\text{box}} = 30 h^{-1} \text{Mpc}$  and  $2048^3$  particles with a Plummer equivalent smoothing of  $1.2 h^{-1} \text{kpc}$  based on a [Planck Collaboration et al. \(2014\)](#) cosmology with  $\Omega_{\text{m}} = 0.30851, \Omega_{\text{b}} h^2 = 0.022161, h = 0.6777, n_{\text{s}} = 0.9611$  and  $\sigma_8 = 0.8288$ . We do not include uncertainties in cosmological parameters in our models as CMB measurement errors on these parameters ([Hinshaw et al. 2013; Planck Collaboration et al. 2016b](#)) are much smaller than current constraints on thermal parameters.

The model is generated for snapshots with  $z$  between 1.8 and 3.4 with a separation of 0.2, the same as our power spectrum measurements, and provides a dark matter density and velocity field. However, the relevant quantities for absorption in the IGM are baryonic density and temperature fields. The results of previous hydrodynamic simulations suggested a computation of the relevant baryonic fields using scaling relations on the dark matter quantities ([Hui & Gnedin 1997; Gnedin & Hui 1998; Gnedin et al. 2003](#)). This basically consists of smoothing the DM density field with a Gaussian kernel to mimic pressure support as well as rescaling the densities to get the right  $\Omega_{\text{b}}$ . Temperatures are then introduced by applying the power law temperature-density relation from [eqn. \(1\)](#) to the density field (see [Rorai et al. 2013](#), section 2.2 for the exact procedure).

Given that the UVB is not known perfectly, we created a sequence of models with different mean transmissions  $\bar{F}$  spanning an  $\sim 5\sigma$  range around the current observational constraints by [Becker et al. \(2013\)](#), [Faucher-Giguère et al. \(2008b\)](#) and [Kirkman et al. \(2005\)](#). This was done by rescaling the optical depths of the full set of skewers to match the desired  $\bar{F}$ . In total we have a parameter grid of  $\sim 500$  different thermal parameter combinations ( $T_0, \lambda_{P_2}, \gamma$ ) with each of those evaluated for 5 different values of  $\bar{F}$ .

### 3.4. Forward Modeling Approach

As the observed spectra are much longer than the simulation box we first divided each spectral chunk of  $\Delta z = 0.2$  into regions smaller than our box and assigned a random simulated model skewer to each of the pieces. The skewers were then assumed to fall on the respective position of the spectrum and truncated to have the same length as the spectral chunk. A model of a single real spectrum therefore consists of 4–8 simulation skewers. We generated  $\sim 15,000$  skewers of the same length as the simulation box for each parameter combination, therefore we have the equivalent of  $\gtrsim 1900$  mock spectra (or equivalently  $\gtrsim 100$  times our whole dataset) available at each set of parameters. This step is required to enable

us to add the wavelength dependent noise and masks to the models.

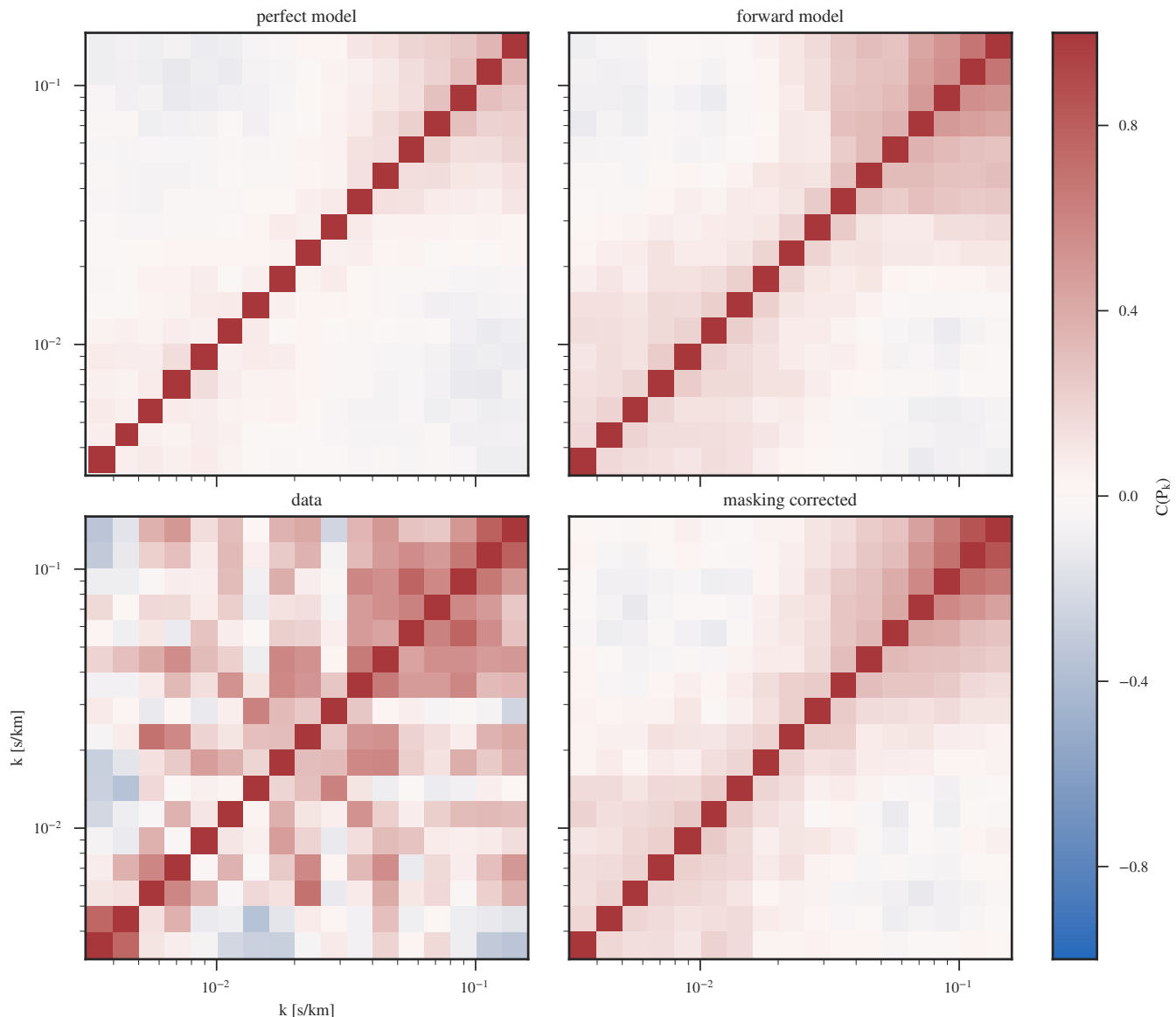
While the overall mean flux of the box is a free parameter of our models, we renormalize the flux in each pixel of the skewer again to account for the slight redshift evolution of the mean flux along the skewer with respect to the mean flux of the simulation snapshot. To do this the fluxes are converted back to optical depth by  $\tau = -\log(F)$ . We then use the best fit relation to the  $\tau$  evolution by [Becker et al. \(2013, see eqn. \(3\)\)](#) to compute the fractional change in  $\tau_{\text{eff}}$  between box redshift and the individual pixel redshifts and rescaled  $\tau$  at each pixel with the corresponding value. After this we convert back to fluxes. The same mean flux evolution function is then later taken out in the power spectrum computation when we, analogous to our procedure on the data, divide by the mean flux in order to compute the flux contrast. We then convolved these spectra with the respective instrument resolution, and interpolate them onto the same wavelength grid as the observed data. The result of these steps are mock Ly $\alpha$  forest spectra with the same noise properties, spectral coverage, and masking of our data. We henceforth call these the ‘forward models’.

We also compute the power spectrum of the same number of model skewers without any noise, masking, or degradation of resolution which we compare to the power spectrum of our mocks to validate our power spectrum pipeline, and to determine the window function correction. We will henceforth refer to these as ‘perfect models’.

For fitting the BOSS data, there is no need to create full forward models, because all the imperfections our forward model approach treats are already accounted for by the BOSS power spectrum pipeline, and therefore we simply compare the BOSS data to the perfect models.

### 3.5. Covariance Matrix Estimation

In addition to measurements of power spectrum, our statistical analysis requires a covariance matrix. The full covariance matrix consists of  $n_{\text{bins}} \cdot (n_{\text{bins}} + 1)/2$  independent entries where  $n_{\text{bins}} \simeq 15$  is the number of band powers of the wavenumber  $k$ . For a dataset like ours that consists of only  $n_{\text{QSO}} \sim 10$  quasars, this estimate will be extremely noisy. [McDonald et al. \(2000\)](#) did tests on bootstrapped covariance matrices based on spectra that were subdivided into 5 spectral chunks (which then were treated as independent data in the same redshift bin) and concluded that there was still significant statistical uncertainties on the estimates of individual covariance matrix entries. Also [Iršič et al. \(2017c\)](#) tested the bootstrap covariance estimation technique on models with different amounts of skewers. When using only 100 model skewers, which is comparable to the size of their dataset, noise in their correlation matrix is still clearly visible. They also provide a correlation matrix of their measurement that looks similar to this model estimate. As our typical redshift bin has less data available than theirs and as our data is additionally masked for metal absorption, we do not believe that we can estimate reliable covariance matrices from our data (see [Figure 5](#) and discussion below). To circumvent this problem we use a hybrid approach (in a similar way as [Lidz et al. 2006](#)) and only measure the diagonal values  $\sigma_{\text{data}}^2$  of the covariance from the data itself while computing the off-diagonal correla-



**Figure 5.** The correlation matrices of the non-window corrected, metal-masked Ly $\alpha$ -forest flux power at  $z = 2.8$  as measured from our DM perfect model with parameters close to the best fit (upper left, see § 3.7 for how the fit was done), the forward model (upper right) at the same parameters, the data (lower left), and the final masking corrected measurement based on the forward model (lower right, see § 3.8 for an explanation about how this was obtained). We can see that the data correlation matrix is far noisier than the others due to the limited data sample size and therefore not usable for model fitting. For the forward model correlations we observe that  $k$ -bins that are close together are mildly correlated ( $\approx 20\%$ ) and bins on very different scales are mildly anticorrelated. For  $k \gtrsim 0.03$  correlations get far stronger due to the power spectrum cutoff as well as the stronger influence of contaminants, e.g. metals, noise and finite spectroscopic resolution. For the perfect model correlations are far weaker except for neighboring  $k$ -bins or bins in the cutoff region. As the exact position of this increase in correlation depends on the power spectrum cutoff position (and therefore e.g. on thermal parameters) we interpolate between correlation matrices when doing model fitting. The final masking corrected correlation matrix looks very similar to the forward model case except for additional correlations in the smallest scale bins due to the masking correction.

tions  $R_{ij}$  from simulations (see § 3.3) for which we can obtain sufficient statistics:

$$R_{ij} = \frac{\langle (P_{m,\text{single}}(k_i) - P_m(k_i))(P_{m,\text{single}}(k_j) - P_m(k_j)) \rangle}{\sigma_m(k_i)\sigma_m(k_j)} \quad (7)$$

where  $P_{m,\text{single}}$  is the power estimated from a single model skewer,  $P_m$  is its mean over all model skewers,  $\sigma_m$  its standard deviation and the average  $\langle \dots \rangle$  is performed over all model skewers. The covariance matrix is

then computed as:

$$C_{ij} = \sigma_{\text{data}}(k_i)\sigma_{\text{data}}(k_j)R_{ij}. \quad (8)$$

To estimate the data uncertainties  $\sigma_{\text{data}}$  we use a bootstrapping technique resampling the data. For this we draw 1000 random sets of quasars with replacement from those contributing to any given redshift bin, and for each compute the power spectrum using eqn. (4). For the correlation matrix  $R$ , we use the same mock spectra generated with our forward modeling procedure (the ‘forward models’ in § 3.4) to determine the correlation matrix according to eqn. (7). No bootstrapping was performed,

since for the mocks we have access to different forward modeling realizations of the same dataset. Note that we have a correlation matrix for each model, because the shape of the power spectrum impacts the correlations leading to stronger correlations for modes on scales smaller than the power spectrum cutoff.

In [Figure 5](#) we show the correlation matrix for one of our redshift bins determined from the model with parameters closest to our best-fit at this redshift. The typical correlation between neighboring bins is  $\sim 15\%$  and decreases strongly for bins further apart. For comparison, we also compute the correlation matrix from our dataset at this redshift as well. This correlation matrix is noisy, and is not used anywhere else in our analysis but is shown here for the sake of illustration. However, one sees that it qualitatively agrees well with our simulated correlation matrix, validating our approach. We also show the correlations of a perfect model and comparing to the forward model we can see that the additional masking added significant correlations on the 20% level between non-neighboring  $k$ -bins.

We also tested our bootstrap estimation of the diagonal elements of the covariance using random simulated datasets with a size comparable to our measurement dataset that were drawn from our full set of models without replacement. We found that the variance of the power spectrum determined from the ensemble of simulated datasets was in good agreement with the bootstrap estimate obtained from a single mock dataset. Therefore we are confident that our bootstrap estimates of the diagonal elements of the covariance are converged and reflect the actual uncertainties of the power spectrum.

### 3.6. Fast Emulation of Model Power Spectra

To be able to fit the power spectrum measurement we need to create a model for the power in the full region of interest for thermal parameters  $(T_0, \gamma, \lambda_P)$  and the mean transmission of the forest  $\bar{F}$ . As simulations are relatively expensive there is no way to run a full simulation for every combination of thermal parameters at which the power spectrum needs to be evaluated during the fitting process. Therefore we generate a grid of simulations with different thermal parameters (using our semi-numerical DM only simulations) and adopt a fast emulation procedure to compute the power spectrum for parameters between the grid points (in a similar way to what is used for the cosmic calibration framework by [Heitmann et al. 2006, 2009, 2013; Habib et al. 2007](#)). Following the procedure described in [§ 3.3](#), we generated Ly $\alpha$  forest skewers for a grid of thermal parameters  $(T_0, \gamma, \lambda_P)$  and mean flux values  $\bar{F}$ . We computed the power spectrum from perfect skewers generated from these models, as well as for mock data run through the forward modeling pipeline described in [§ 3.4](#).

For both sets of power spectra (perfect and forward-modeled) we perform a principal component analysis (PCA) of the full set of model power spectra. As result the power at any point  $\theta = \{T_0, \gamma, \lambda_P, \bar{F}\}$  in our model grid can be written:

$$P(k, \theta) = \sum_i \omega_i(\theta) \Phi_i(k), \quad (9)$$

where both the eigenvectors  $\Phi_i(k)$  and the PCA weights

$\omega_i(\theta)$  are a result of the PCA decomposition. Note that the PCA weights depend on the gridpoints in our model parameter space, and given a suitable interpolation scheme, they can be used to generate model predictions at any point in this space. Along these lines, and following [Habib et al. \(2007\)](#), the PCA weights  $\omega_i$  are then used as an input to train a Gaussian process interpolation scheme. From this Gaussian process we can later evaluate new weights for parameter combinations that lie between or original gridpoints, allowing us to evaluate the power at any location. For more details on this approach see a companion paper to this analysis ([Walther et al. 2017a](#)) or e.g. [Rorai et al. \(2013\)](#) and [Habib et al. \(2007\)](#). The Gaussian process interpolation uses a squared exponential kernel with smoothing lengths chosen to be larger than the separation between model grid points of the parameter space. To speed up computations only the first 9 principal components are actually used for the analysis. We found that the additional errors due to discarding the higher order components is less than 1%.

In general the same approach can be used to emulate models from hydrodynamical simulations, but this means that a separate hydrodynamical simulation must be run for each parameter combination in the grid which is far more costly than the approach we chose for this work. Nevertheless, we will use an emulator based on full hydrodynamic simulations to infer thermal parameters from our power spectrum measurements in a companion paper ([Walther et al. 2017b](#)).

### 3.7. Parameter Exploration

To explore the parameter space and fit the measured data power spectra from both BOSS and high-resolution datasets we use a Bayesian Markov Chain Monte Carlo (MCMC) approach with a Gaussian multivariate likelihood:

$$\begin{aligned} \mathcal{L} &\equiv P(\text{data}|\text{model}) & (10) \\ &\propto \prod_{\text{datasets}} \frac{1}{\sqrt{\det(C)}} \exp\left(-\frac{\Delta^T C^{-1} \Delta}{2}\right) \\ \Delta &= \mathbf{P}_{\text{data}} - \mathbf{P}_{\text{model}}, \end{aligned}$$

where  $\mathbf{P}_{\text{data}}$  and  $C$  are the power spectrum and covariance matrix obtained for any dataset, and the product is over all datasets considered (i.e. BOSS and our high-resolution measurement, but including more than two datasets would just add factors to this product).

For the high-resolution dataset  $C^{-1}$  is estimated using our hybrid approach (see [§ 3.5](#)) for each model on the parameter grid and therefore is a model dependent quantity. Inside the likelihood computation we use nearest neighbor interpolation in model parameter space to obtain an inverse covariance estimate for each combination of model parameters (which is then just the matrix at the closest model grid point). For the covariance matrix of the BOSS dataset we use the tabulated values from [PD+13](#) which are directly measured from the data, and are therefore model independent. The model power  $\mathbf{P}_{\text{model}}$  is determined for any parameter combination using our emulation procedure. Note that we actually have two distinct emulators, one for the high-resolution data (using the full forward-modeling procedure), and one for

the BOSS data (using our perfect simulation skewers).

It is well known that correlated absorption of hydrogen Ly $\alpha$  and Si III at 1206.5 Å leads to a bump in the Ly $\alpha$  forest flux correlation function  $\xi(\Delta v)$  at  $\Delta v = 2271 \text{ km s}^{-1}$  (McDonald et al. 2006; PD+13)<sup>9</sup>. This imprints wiggles on the power spectrum with separation  $\Delta k = 2\pi/\Delta v = 0.0028 \text{ s km}^{-1}$ . Following McDonald et al. (2006), we model this contamination with a multiplicative correction to the power

$$P(k) = (1 + a_{\text{Si III}}^2 + 2a \cos(\Delta v k))P_{\text{HI}}(k), \quad (11)$$

with  $a_{\text{Si III}}$  being a free nuisance parameter for the strength of the correlation. In previous works this was typically expressed as  $a_{\text{Si III}} = f_{\text{Si III}}/(1 - \bar{F})$  with  $f_{\text{Si III}}$  being a redshift independent quantity that was fit using the entire dataset. We adopt this same parameterization but opt to fit for a unique value of  $f_{\text{Si III}}$  at each redshift. Therefore we have five free parameters  $T_0, \gamma, \lambda_J, \bar{F}, f_{\text{Si III}}$  to fit to our power spectrum measurements in each redshift bin.

We used the publicly available MCMC package `emcee` (Foreman-Mackey et al. 2013) based on an affine invariant ensemble sampler (Goodman & Weare 2010) to sample the posterior distribution:

$$P(\text{model}|\text{data}) = \frac{P(\text{data}|\text{model})P(\text{model})}{P(\text{data})} \quad (12)$$

with  $P(\text{data}|\text{model})$  being the combined likelihood of both datasets for a given set of model parameters, and  $P(\text{model})$  being the prior on that combination of parameters. We assume that the priors on parameters are independent, therefore  $P(\text{model})$  is just the product of the individual priors for each parameter. For the thermal parameters we adopt flat priors on  $\gamma$  in the range  $0.5 < \gamma < 2.1$ ,  $\log(\lambda_J)$  in the range  $22 \text{ ckpc} < \lambda_J < 150 \text{ ckpc}$  and  $\log(T_0)$  in the range  $3000 \text{ K} < T_0 < 20000 \text{ K}$ . For the mean flux we adopt Gaussian priors based on the most recent measurements of Becker et al. (2013), Faucher-Giguère et al. (2008b) and Kirkman et al. (2005) (depending on redshift) for  $\bar{F}$ . For the Si III correlations we used a Gaussian prior defined by the best-fit value in  $f_{\text{Si III}}$  and  $1\sigma$  region of Palanque-Desabrouille et al. (2013). All priors for all redshifts are listed in Table 4. Note that as long as a good fit to the data can be obtained (i.e. the model is sufficiently flexible) the exact values of thermal parameters (and therefore also the exact priors chosen) don't matter in the context of this work as the fits are only used to correct out the window function as we discuss in the next subsection. In a companion paper doing thermal parameter analysis based on hydrodynamical simulations we will loosen the priors on both  $\bar{F}$  and  $f_{\text{Si III}}$  to not bias our results. Fits were always performed for individual redshifts and no correlation between thermal properties at different redshifts was assumed.

### 3.8. The Raw Power Spectrum and Window Function Correction

The impact of masking on the power depends on the underlying shape of the power spectrum. We use the fits to the underlying power based on our DM models

<sup>9</sup> Note that this correlation should generally be weaker in our high-resolution dataset as we masked some of the Si III absorption

to determine the impact of the window function on the shape of the power spectrum. We emphasize here that we use this window-function corrected power mainly for visualization purposes only as the covariance in the measurement is subtly changed by the correction. While we do provide an approximate covariance matrix in Appendix D, ideally a full forward model of the power spectrum should be used when doing parameter estimations.

Our measurement, emulation, and window function procedure are illustrated in Figure 6. On the left side we show a comparison between our raw data power spectrum (i.e. before window function correction) and the BOSS measurement as points. After applying our fitting procedure we obtain an MCMC chain from which we can draw parameter combinations  $\Theta = \{T_0, \lambda_P, \gamma, \bar{F}, f_{\text{Si III}}\}$  that are compatible with the data. Feeding random draws from this chain into our emulator routines for both the forward model and the perfect model produces the black and purple bands. We can see that these provide good fits to our dataset and the BOSS dataset, respectively. For a single draw  $\Theta$  from the posterior, we can then measure a window function correction  $f_{\text{window}}$  using both emulators as:

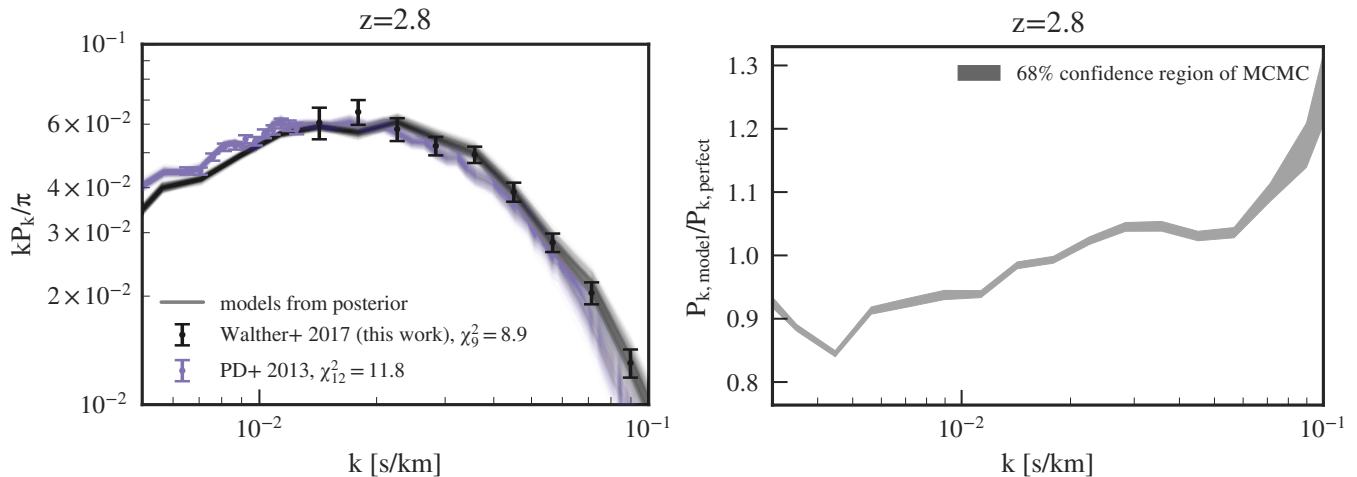
$$f_{\text{window}}(k) = \frac{P_{\text{forward model}}(k, \Theta)}{P_{\text{perfect model}}(k, \Theta)}. \quad (13)$$

The gray bands on the right side of Figure 6 show the resulting 16% and 84% quantiles of this quantity. The dominant effect of windowing our data is a strong increase in power on the small scale ( $k \gtrsim 0.07 \text{ s km}^{-1}$ ) end of the measurement.

To extract the underlying (i.e. window function corrected) power spectrum from our raw measurement we proceed as follows. We generate 1000 random draws from a multivariate normal distribution with a mean given by the raw  $P(k)$  measurement (which includes the effect of the window function) and a covariance matrix defined in eqn. (7) (based on both the raw measurement and the best-fit model). We then obtain the window function correction  $f_{\text{window}}(k)$  for 1000 draws from the posterior. By multiplying each  $P(k)$  draw with each  $f_{\text{window}}(k)$  draw we obtain samples of the distribution describing the window corrected measurement. We take the mean of these samples to represent our window corrected power, and the covariance of these samples gives the respective window corrected covariance. These results are tabulated in Appendix D. We also show the resulting correlation matrix applying this procedure at  $z = 2.8$  in Figure 5. Compared to the uncorrected forward model we can see additional correlation at the smallest scales ( $k \gtrsim 0.07 \text{ s km}^{-1}$ ) where the influence of the window correction is strongest.

## 4. A NEW POWER SPECTRUM MEASUREMENT

In this section we present our final window-function corrected power spectrum measurements over the redshift range  $1.8 \leq z \leq 3.4$ . First, we discuss the impact of metal line contamination on the power spectrum measurement. Then we compare our power spectrum measurement to the lower- $k$  measurements from BOSS as well as the XQ-100 dataset, which represent the state-of-the-art from low- and medium-resolution data. Finally, we compare our results to previous high-resolution measurements.



**Figure 6.** Left: The measured power spectrum at redshift  $z = 2.8$  compared to the best-fit. The black points show the raw power of the high-resolution data and are fitted with mock observations including the same noise and masking as in the data. The purple points are showing the PD+13 measurement including their metal, noise and resolution corrections. These are then fitted with models generated from noiseless, non-masked skewers. Lines show models randomly drawn from the posterior MCMC-chain. Right: The window function estimate from a comparison between the power from noiseless, mask-free models and the full forward-modeled mock observations. The band shows the 68 % contour for models drawn from the posterior MCMC-chain. The median is applied as a correction factor to the raw data. The width of the band is propagated into the data errorbars.

#### 4.1. The Final Power Spectrum Measurement and the Effect of Metals on the Data

In previous work the effect of metals on the small scale power spectrum has been largely ignored. While McDonald et al. (2000) computed the power spectrum on a dataset with and without metals masked and also considered the effect of masking on the power spectrum, they did not combine both results to see the net-effect of metal contamination. However, they did note that especially for  $k > 0.1 \text{ s km}^{-1}$  the effect of metals as well as noise becomes too strong for their measurement to be usable. Motivated by this conclusion, later measurements by Croft et al. (2002), Kim et al. (2004) and Viel et al. (2008, 2013) also ignored these small scale (larger  $k$ ) modes. For lower resolution measurements using SDSS (McDonald et al. 2006) or BOSS (PD+13) spectra, these modes are well above the resolution limit of the data, and this has not been an issue. Instead, for those measurements the power in metals is estimated from lower redshift data (using a spectral range redwards of the Ly $\alpha$  forest). As this procedure cannot treat metal lines that are always inside the forest large scale correlations between Si III and Ly $\alpha$  are the dominant effect of metal contamination in this case.

In Figure 7 we show a comparison between our power spectrum measurement applying the metal masking procedure described in § 2.2 and performing the analysis without masking metals. Both measurements have been noise subtracted following the discussion in § 3 and corrected for their respective window functions according to § 3.8. We also show the BOSS measurement as a comparison, for  $z > 2.2$  where those measurements exist.

It is clear that particularly at small scales ( $k \gtrsim 0.1 \text{ s km}^{-1}$ ) metal lines significantly contribute to the measured power spectrum. This increase in small scale power in general leads to an underestimation of the small-scale thermal cutoff, and naively fitting this metal-contaminated power would result in lower overall IGM temperatures (i.e. more small-scale structure and hence

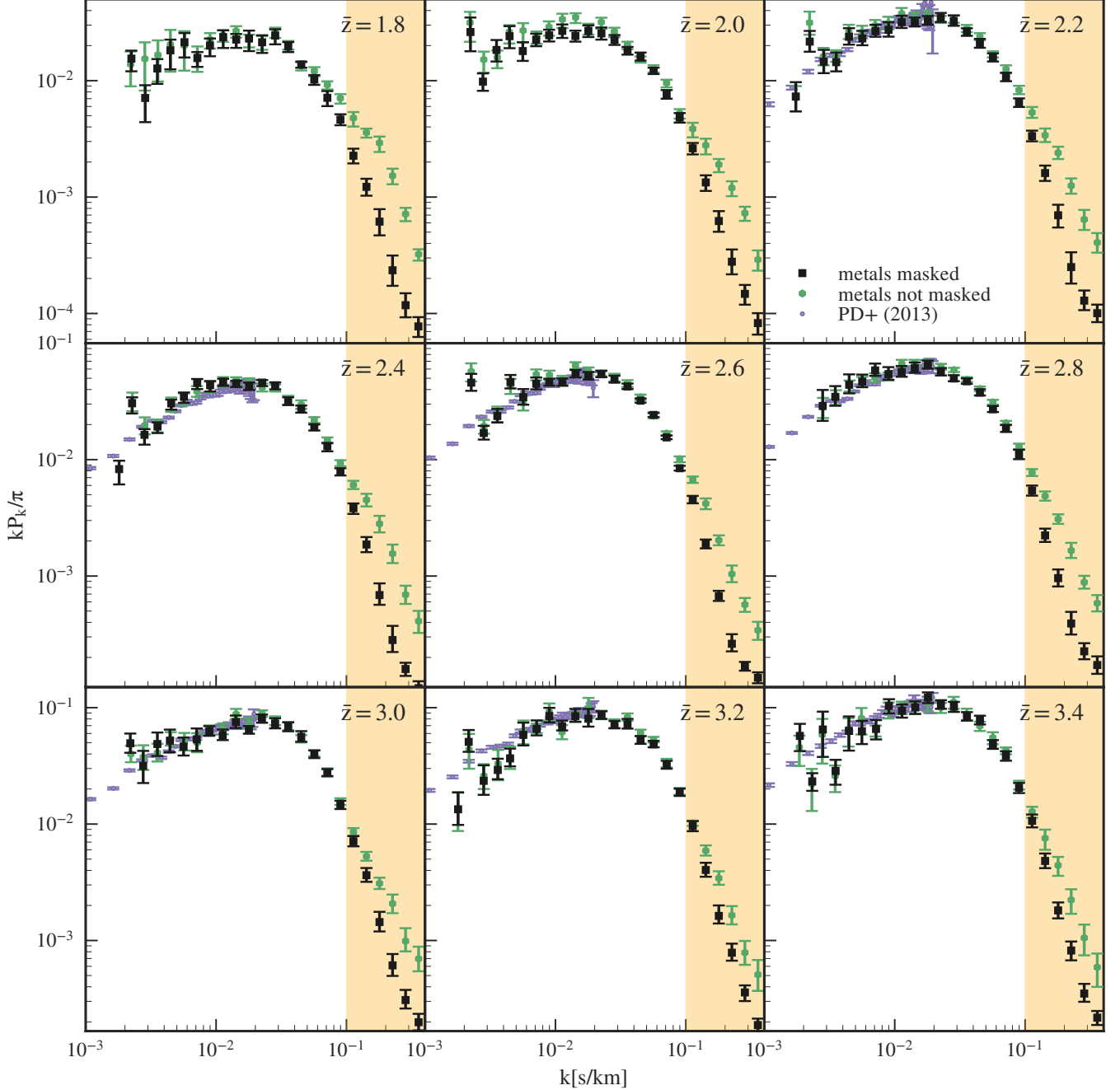
less thermal broadening and/or pressure smoothing). For all further analysis we therefore use the metal-masked power (black dots in the figure), and our model fitting is conservatively restricted to modes with  $k < 0.1 \text{ s km}^{-1}$ , where the impact of metal line contamination is relatively weak, and hence our metal-masked power is relatively insensitive to the fact that we may not have masked all of the metals (see discussion in § 2.2).

Figure 7 also indicates that the impact of metal-line contamination is not restricted to small-scales. Multiple ionic metal-line transitions are typically associated with a given absorber redshift, and because the velocity separations are thousands of  $\text{km s}^{-1}$ , these large-scale velocity correlations can impact the power spectrum on large-scales (low- $k$ ) as well. Masking metals in the data at e.g.  $z=2.6$  decreases this effect lowering the power spectrum and therefore increases the agreement with the PD+13 measurement.

To summarize, metal-line contamination of the power spectrum does not significantly impact our results, given that we mask the metals and conservatively restrict our power spectrum fits to low  $k < 0.1 \text{ s km}^{-1}$  where the impact of residual (i.e. metals we missed in our masking) metal-line contamination should be negligible. In principle, a more careful treatment of metal-lines (i.e. via improved masking, or forward modeling the metals, or subtracting the red-side metal-line power) could allow one to access even smaller scales (higher  $k$ ), although this would also require a very careful treatment of the noise and instrumental resolution whose effect also increases for smaller scales (larger  $k$ ).

#### 4.2. Comparison to Previous Low and Medium Resolution Measurements

In Figure 8 we show our new metal and window corrected measurement of the high-resolution power spectrum compared to the BOSS measurement from PD+13. Note that different power spectrum bins are correlated and the errorbars only reflect the diagonal elements of the covariance matrix. Where both measurements over-



**Figure 7.** Our power spectrum measurement of spectra with masked metals (black squares) and with metals left inside the forest (green hexagons) as well as the BOSS measurement (purple errorbars). Both high-resolution measurements are corrected for their respective window function. In most redshift bins contamination due to metal absorption clearly leads to an increased small scale (large  $k$ ) power which would bias a potential temperature measurement towards lower temperatures. The orange region shows the region excluded in our further analysis as this effect gets far larger than our statistical uncertainties. On larger scales power is removed as well when masking the metal lines leading to an overall better agreement with the BOSS measurement.

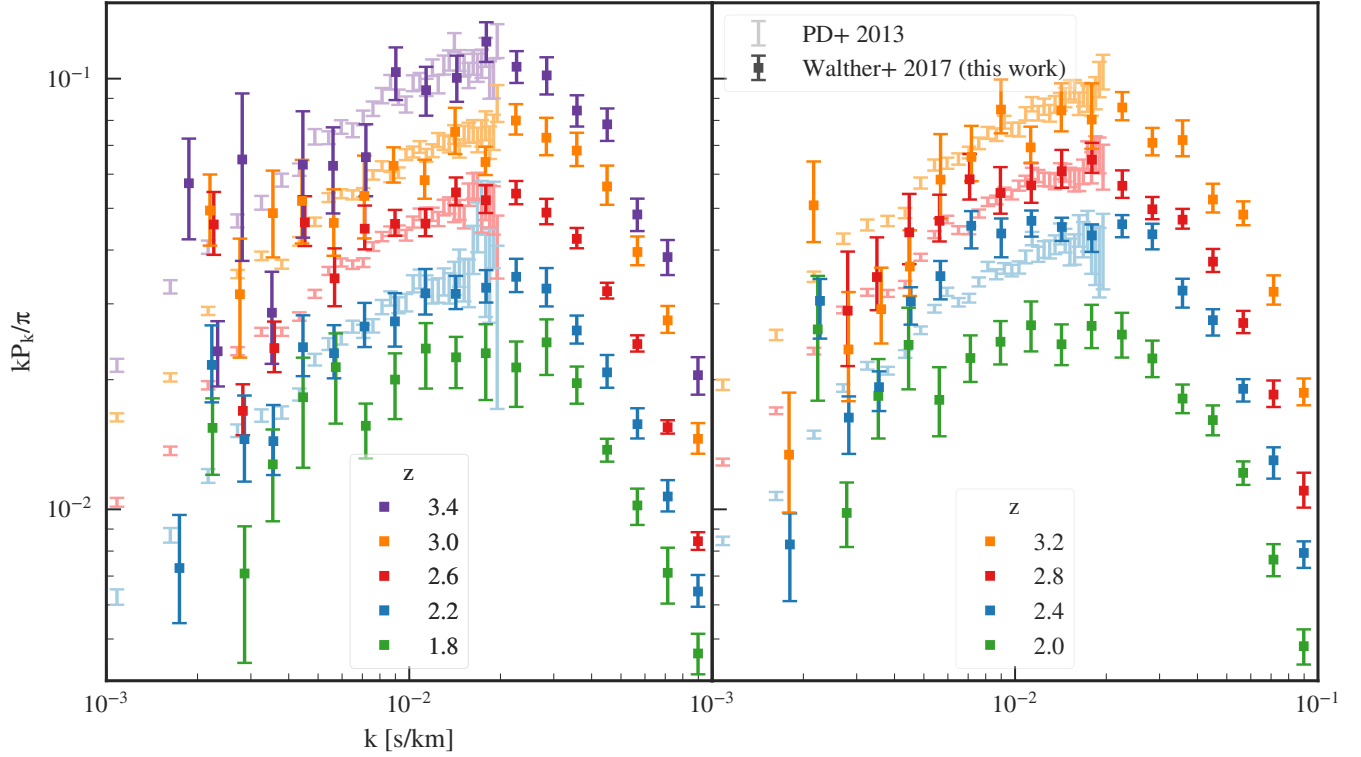
lap ( $k \lesssim 0.02 \text{ s km}^{-1}$ ,  $2.2 < z < 3.4$ ) we find good agreement with the BOSS power spectrum (which has a typical accuracy of  $\sim 2\%$ ) within our errors at modes  $k > 0.01 \text{ s km}^{-1}$  and the agreement is generally good for larger scale (lower- $k$ ) modes as well. The discrepancies at low- $k$  are most likely due to the fact that our measurements are rather noisy for small wavenumbers<sup>10</sup>.

<sup>10</sup> This is a result of two factors: 1) the density of modes for a one-dimensional power spectrum measurement is uniform, and so

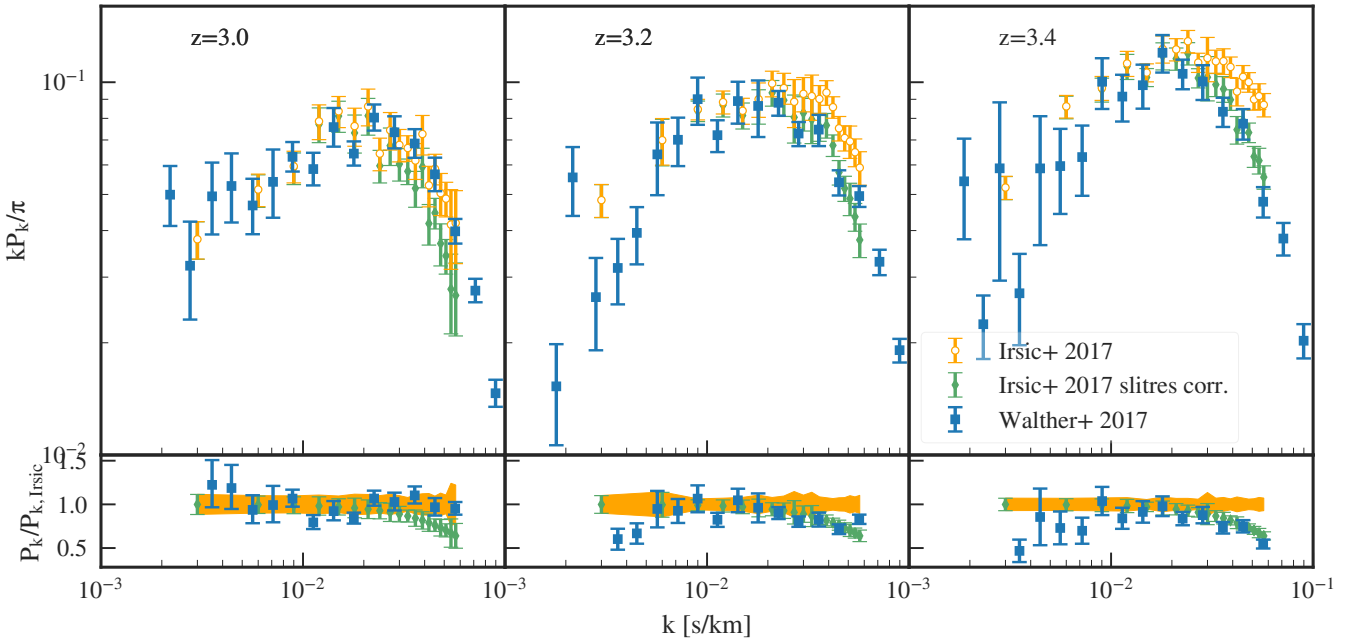
We also compare to the recent result of Iršič et al. (2017c) based on the XQ-100 dataset (López et al. 2016). Specifically, Iršič et al. (2017c) measured the power at  $3.0 \leq z \leq 4.2$ , and their redshift bins at  $z = 3.0, 3.2$ ,

for linearly spaced bins there are equal numbers of modes at each  $k$ . However, our  $k$ -bins are logarithmically spaced, so in general our error bars are larger in a relative sense at lower- $k$ . 2) For some of our spectra either limited spectral coverage, or our masking procedure tend to reduce the amount of path length available for measuring the largest scale modes, making them more noisy.





**Figure 8.** Our new measurement of the power spectrum (dark squares) for  $1.8 \leq \bar{z} \leq 3.4$  (different colors), metal lines were masked our analysis and the power introduced by masking was removed using forward modeling of our measurement (see § 3.4). Also shown are the measurements of PD+13 (BOSS, bright errorbars on the left). We can see that at most overlapping redshifts there is good agreement except for a mild disagreement with BOSS on large scales (small  $k$ ) for  $z \sim 2.4$ .



**Figure 9.** Our new measurement of the power spectrum (blue squares) for  $3.0 \leq \bar{z} \leq 3.4$  as in Figure 8 compared to the measurement of Irsić et al. 2017c (XQ-100, orange open circles). To address the disagreement at small scales (high  $k$ ) for  $z \geq 3.2$  between our high-resolution and the XQ-100 data, we also show the XQ-100 data assuming a different resolution correction (see main text for details) by using a distribution of seeings and assuming an underestimation of the X-SHOOTER slit-resolution (green diamonds). The bottom panel shows the same comparison, but normalized by the (untreated) Irsic+2017 measurement. The quoted Irsic measurement errors are shown as an orange band.

and 3.4 match those used in our analysis. Those are compared in [Figure 9](#).

While the agreement between our measurements at  $z = 3$  is good, at  $z = 3.2$  and  $z = 3.4$  [Iršič et al. \(2017c\)](#) measures  $\sim 10\%$  to  $50\%$  higher power than we do at small scales  $k \gtrsim 0.02 \text{ s km}^{-1}$ , which is clearly statistically significant given the error bars for the respective datasets. Note however that this disagreement is restricted to high- $k$ , and we agree well with [Iršič et al. \(2017c\)](#) at intermediate  $k$  in all redshift bins.

While it is difficult to be certain about the source of this discrepancy, given the different methodology used for measuring the power, the  $k$  dependence of this disagreement provides a very important clue. Note that X-SHOOTER data is significantly lower resolution than the dataset presented here and the resolution corrections become significant at higher- $k$ .

Possible uncertainties of spectroscopic resolution can come from several sources: 1) as X-SHOOTER is a slit-spectrograph its resolution is seeing dependent, so the seeing itself needs to be accurately determined to get an accurate resolution, 2) seeing changes between different observations and correcting assuming a constant resolution across the dataset might bias the measurement (see [Appendix A](#) for more details on those points), 3) the UVB slit resolution quoted for X-SHOOTER might be significantly underestimated (see [Appendix B](#)). Because of those problems, a correction based on the nominal slit resolution can overproduce small scale power in the data analysis. However, [Iršič et al. \(2017c\)](#) already used a higher resolution than provided in the XQ-100 data release paper when performing his corrections<sup>11</sup>. Using the [Iršič et al. \(2017c\)](#) value for the resolution leads to agreement between both XQ-100 power spectrum analyses by [Iršič et al. \(2017c\)](#) and the independent determination of the power spectrum from the same dataset by [Yeche et al. \(2017\)](#). [Yeche et al. \(2017\)](#) determined the spectral resolution by assuming the XQ-100 resolution, and estimating the seeing on object by object basis analyzing the width of the Ly $\alpha$  forest in the spacial direction of the 2d-spectra.

The influence of such resolution errors on the resolution correction factor  $W_R^2$  (see [eqn. \(5\)](#)) can be found by simple error propagation:

$$\begin{aligned} \Delta \ln W_R^2 &= -k^2 \Delta(R^2) \\ &\simeq -2k^2 R^2 \Delta \ln(R) \end{aligned} \quad (14)$$

and propagates to an error on the estimated power spectrum  $P$  as:

$$\Delta \ln P = -\Delta \ln W_R^2. \quad (15)$$

Assuming the nominal resolving power of 4350 (according to [López et al. 2016](#), listing a slightly higher value than the ESO Instrument description) using the X-SHOOTER  $1''$  slit on the UVB arm as a worst case scenario a spectrum with  $10\%$  higher resolution than assumed when performing the correction will lead to a  $\sim 45\%$  ( $28\%$  assum-

ing a resolving power of 5350) overestimate of the power in the resolution corrected measurement. As our own power spectrum measurement is based on  $\sim 10$  times higher resolution data a comparable error in the knowledge of the resolution will have a  $\sim 100$  times smaller effect (so  $\sim 0.4\%$  at  $k = 0.05 \text{ s km}^{-1}$  and  $\sim 1.5\%$  at  $k = 0.1 \text{ s km}^{-1}$ ). Therefore lack of knowledge of the precise resolution of the spectrograph a significant concern for the X-SHOOTER measurement, but can be safely ignored for our study.

To determine the influence of possible errors in the resolution estimates due to points 1 and 2 we divided out the resolution correction from the [Iršič et al. \(2017c\)](#) power spectrum measurement and corrected using different assumptions. First, we used the nominal slit resolution  $R = 4350$  from the X-SHOOTER spectrograph, generated Gaussian distributed seeing values (with a mean of  $0.75''$  and FWHM of  $0.2''$  similar to the distribution shown in [Yeche et al. 2017](#)) for each we computed the resolution correction  $W_R^2$  according to [eqn. \(5\)](#) and obtained the mean correction which we then used as the new resolution correction. This gives basically identical results (that we don't show in our comparison figure for clarity) to the original [Iršič et al. \(2017c\)](#) measurement showing that the seeing estimate used for their measurement is in agreement with the distribution determined by [Yeche et al. \(2017\)](#) and cannot be the reason for the disagreement with our measurement. In addition we also generate a measurement assuming a higher slit resolution of  $R = 5350$  (due to point 3 and in agreement with measurements based on calibration spectra, see [Appendix B](#)) and otherwise performing the same analysis. This comparison is shown as red diamonds in [Figure 9](#) and we can see that the agreement between our high-resolution measurement and XQ-100 at  $z = 3.2$  and  $z = 3.4$  is good in this case. However, the agreement in the  $z = 3.0$  bin without assuming a different X-SHOOTER resolution correction is unclear, but might be due to possible variations in the resolution between Echelle orders. Note that for the other spectral arms (that cover the Ly $\alpha$  forest for  $z > 3.6$ ) this is a less severe problem as data from those is intrinsically higher resolution. Also note that for those arms one can in principle obtain the resolution of the science observation from the width of telluric absorption lines (see [Bosman et al. 2017](#)), but those are rare in the UVB arm. This can also explain the agreement between XQ-100 measurements and older high-resolution data by [McDonald et al. \(2000\)](#) at  $z \sim 3.8$  (but as the redshift bins of both measurements are significantly different, this comparison is tricky) and [Viel et al. \(2013\)](#) at  $z = 4.2$ .

Because of the severe impact the resolution correction can have on the XQ-100 power spectrum measurement we caution against using the smallest scales ( $k > 0.02 \text{ s km}^{-1}$ ) of this measurement (at least for the lowest redshift bins) for parameter studies. This is especially true for measurements that rely on the accurate determination of the power spectrum cutoff, like e.g. determining the thermal state of the IGM, or the nature of dark matter (e.g. [Iršič et al. 2017b,a](#); [Baur et al. 2017](#)). Although to be fair, for the latter most of the sensitivity comes from the higher redshift ( $z \gtrsim 4$ ) bins where the resolution of X-SHOOTER is higher and addition-

<sup>11</sup> They used a FWHM resolution of  $50 \text{ km s}^{-1}$  instead of the nominal  $c/R = 69 \text{ km s}^{-1}$  where  $R$  is the quoted resolving power of the X-SHOOTER spectrograph. Communication with the main author revealed that this value was obtained by visual inspection of the raw science data before the official values were available in a similar way as for the [Yeche et al. \(2017\)](#) analysis.

ally high-resolution ( $R \simeq 50,000$ ) data from Viel et al. (2013) are used.

In Figure 4 we can see, that our metal removal and masking correction techniques do not change the data hugely (the difference is covered by our error bars) at the redshifts and scales where we disagree with Iršič et al. (2017c). Additionally, the changes due to metal masking do not exhibit the same shape as the discrepancy. Finally, we also checked the raw power spectra not corrected for any masking and could not get a small-scale power as high as in the Iršič et al. (2017c) result. We are therefore confident of metal masking and window correction not being the reason for the discrepancies.

Our measurements probe the small-scale cutoff in the power ( $0.02 \lesssim k \lesssim 0.1$ ) in all redshift bins with a typical precision of 5 – 15%. The position of this cutoff is still at far larger scales than the expected cutoff due to the spectroscopic resolution of our data and the observed cutoff in the power results from thermal and/or pressure broadening of the absorption lines (or e.g. warm/fuzzy dark matter). In the next subsection we also compare to previous high-resolution measurements to make sure our measurement agrees with existing data on small scales as well.

#### 4.3. Comparison to Previous High-Resolution Measurements

Previous power spectrum measurements based on high-resolution data were obtained by different groups using redshift bins of different size and location. They also differ in the Fourier normalization convention used (leading to factors of 2 between some measurements), the field of which the power is computed (flux contrast or transmission leading to additional factors of  $\bar{F}^2$  in the power) as well as whether metals were masked and noise was subtracted. We therefore show comparisons at the quoted redshifts for the old high-resolution measurements and linearly interpolate our results to those redshifts and renormalize the different measurements to the power spectrum convention that we use (see Appendix C).

For authors that chose to study the  $F$  field instead of  $\delta_F$ , there is ambiguity of the mean flux. While the mean flux of the IGM is clearly the same, the mean flux of the data is sensitive to where the continuum is placed. It is well known that hand continuum fits to high-resolution spectra are biased low (Faucher-Giguère et al. 2008a), and this systematic effect is a bigger issue at higher redshift. If one measures the power spectrum of  $F$ , and the continuum fits are biased low, then the power will be biased high. McDonald et al. (2000) provide measurements of their mean flux, and we can therefore easily correct this effect, whereas Croft et al. (2002) does not. Thus a direct comparison to Croft’s measurements is not straightforward, but we do our best by simply assuming their continua are unbiased and multiplying in the mean flux of the IGM measured by (Becker et al. 2013) at the respective redshifts of their measurements. The differences between power spectrum conventions clearly limit the precision at which our measurements will agree with previous work.

The comparison between high-resolution measurements is shown in Figure 10. While overall agreement is good considering the different approaches, some com-

parisons show disagreement. The strongest mismatch is with (Croft et al. 2002) at  $z = 2.13$  on all scales. At similar redshifts we agree with both Kim et al. (2004) and PD+13 which hints toward an incorrect mean flux or a problem with the Croft et al. (2002) measurement at this redshift (which is not part of their fiducial sample). We can also see that the shape of the different measurements on scales smaller (larger  $k$ ) than the cutoff matches between the four high-resolution datasets for scales  $0.01 \lesssim k \lesssim 0.08$ . On smaller scales the difference in treatment of metals and S/N of the datasets as well as removal of noise power can probably explain the deviations between these measurements. We do note that our  $z = 3.0$  bin exhibits a cutoff at slightly smaller scales (larger  $k$ ) than the one of McDonald et al. (2000), but agree with Croft et al. (2002) at essentially the same redshift. This shows that there are clear limitations of comparisons to the previous measurements from high-resolution datasets due to the different conventions

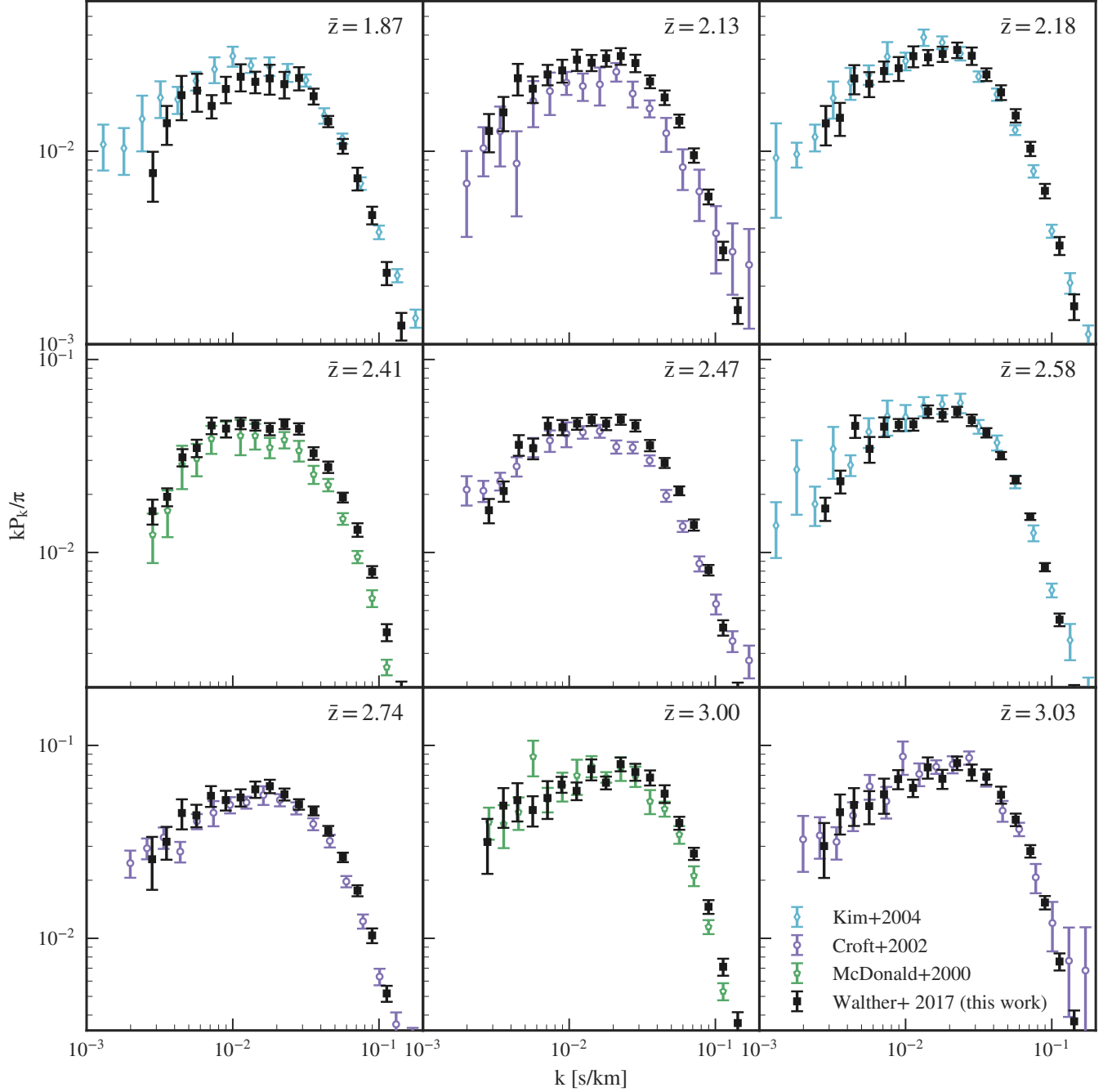
To summarize, we find reasonable agreement between our measurements and previous analyses. We will use our new result for parameter estimations in future works.

## 5. SUMMARY

In this work we presented a new measurement of the Ly $\alpha$  forest power spectrum at  $1.8 \leq z \leq 3.4$  from archival high-resolution spectra obtained with the UVES and HIRES spectrographs. The pathlength of  $\sim 20$  cGpc covered by this dataset (see Figure 4) is several times larger than the previous measurements McDonald et al. (2000), Croft et al. (2002), and Kim et al. (2004). This allows us to measure the small scale cutoff in the power spectrum and its redshift evolution with unprecedented precision.

We developed a custom pipeline to accurately measure the power spectrum and its uncertainty, which fully corrects finite resolution and noise. Some regions of quasar spectra must be masked because of missing data, DLAs, and metal absorption line contaminants, which we identify and mask using several methods. If left uncorrected, this masking alters the shape of the power spectrum, particularly on the small-scales (high- $k$ ) of interest for studying the thermal state of the IGM. To obtain unbiased estimates of the power spectrum and its associated noise, we adopt a forward modeling approach. We post-process a DM simulation (see § 3.3 and § 3.4) and generate a grid of different Ly $\alpha$  forest models with the same noise and resolution as our data. The same masks are applied to these mock spectra and we use a custom emulator (§ 3.6) to perform MCMC fits (§ 3.7) to our measurements. These models are sufficiently flexible that they provide a good fit to the data, although the resulting parameters are not physically meaningful. These model fits are then used to correct our power spectrum and its covariance for the impact of masking (§ 3.8). Our analysis shows that metal line contaminants significantly alter the shape of the raw power spectrum on small-scales  $k > 0.1 \text{ s km}^{-1}$ . Although our masking mitigates the effect of this contamination, we nevertheless restrict further analysis of the power spectrum to  $k < 0.1 \text{ s km}^{-1}$ . Our power spectrum measurements in 9 redshift bins covering  $1.8 < z < 3.4$  are presented in Appendix D.

We compared our new measurements to previous results from both low-/medium (§ 4.2) and high-resolution (§ 4.3) spectrographs. Our measurements agree well



**Figure 10.** Our new measurement of the power spectrum (black squares) compared to the existing measurements of [McDonald et al. \(2000\)](#), [Croft et al. \(2002\)](#) and [Kim et al. \(2004\)](#). Our measurement has been interpolated between the 2 neighboring redshift bins to the same mean redshift as the other datasets. The old measurements by [McDonald et al. \(2000\)](#) and [Croft et al. \(2002\)](#) have been rescaled by  $\bar{F}^2$  as they were obtained on the flux itself instead of the flux contrast  $\delta_F$  which is the cause for different overall normalizations in some bins which is the cause for different overall normalizations in some bins. The [Croft et al. \(2002\)](#) measurement has been rescaled by an additional factor of two to match the Power spectrum normalization convention we use in this paper. It is worth to remark that the older measurements were also performed in wider redshift bins ( $0.3 \lesssim \Delta z \lesssim 0.6$ ). Also [Kim et al. \(2004\)](#) performed their measurement on a subset of our dataset. Notice that the approaches to noise, metal and resolution correction vary between all 4 datasets as well.

with the BOSS power spectrum (PD+13) for the low wavenumbers  $k < 0.02 \text{ s km}^{-1}$  probed by that low-resolution dataset. Given the extremely high  $\sim 2\%$  precision of the BOSS study, we consider this an important validation of our approach. Our measurements significantly disagree with the recent study of (Iršič et al. 2017c) based on the medium-resolution XQ-100 dataset. This disagreement is restricted to redshift  $z = 3.2$  and  $z = 3.4$  and is present only for the higher  $k \gtrsim 0.02 \text{ s km}^{-1}$  modes. Given the direction of the discrepancy, and the fact that only the highest  $k$  modes are affected, we argue that the disagreement most likely results from over-correcting the effect of spectral resolution on the power spectrum, which can ultimately be attributed to improper characterization of the X-SHOOTER spectrograph's resolution. Comparing our results to previous high-resolution efforts (McDonald et al. 2000; Croft et al. 2002; Kim et al. 2004), we mostly find good agreement, although some combinations of dataset and redshift bin are discrepant at the 10 – 50% level. We do not believe these differences are a significant source of concern, as they likely arise from the challenges in comparing measurements covering significantly different redshift ranges, adopting different mean flux normalization conventions (Appendix C), and other systematics that may have plagued previous work.

Combining our precise small-scale (high- $k$ ) measurement with the larger scale power measured by PD+13 and Iršič et al. (2017c) results in a powerful new dataset for placing percent level constraints on the thermal evolution of the IGM, as well as cosmological parameters. In particular, the greatly improved high- $k$  precision enabled by our work will help break degeneracies between the two, and enable improved constraints on neutrino masses, as well alternatives to the CDM paradigm such as warm (Viel et al. 2013; Iršič et al. 2017b) or fuzzy dark matter (Hui et al. 2017; Iršič et al. 2017a). In a companion paper we will compare these measurements to a grid of hydrodynamical simulations to place precision constraints on the thermal evolution of the IGM, and search for the thermal signature of HeII reionization. As our work only covers the redshift range  $1.8 < z < 3.4$ , future efforts should focus on filling in the gaps at both lower  $z \lesssim 1.8$  at  $3.4 \lesssim z \lesssim 6$ , which would enable studies of thermal state of the IGM from just after hydrogen reionization until the present.

#### ACKNOWLEDGMENTS

We thank Martin White for providing the collisionless simulation used for our models.

We thank the members of the ENIGMA group<sup>12</sup> at the Max Planck Institute for Astronomy (MPIA) and University of California Santa Barbara (UCSB) for insightful suggestions and discussions.

Joseph F. Hennawi acknowledges generous support from the Alexander von Humboldt foundation in the context of the Sofja Kovalevskaja Award. The Humboldt foundation is funded by the German Federal Ministry for Education and Research.

Some 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 used resources of the National Energy Research Scientific Computing Center (NERSC), which is supported by the Office of Science of the U.S. Department of Energy under Contract no. DE-AC02-05CH11231.

We would like to thank ESO for making the ESO data archive publicly available.

#### REFERENCES

- Baur, J., Palanque-Delabrouille, N., Yèche, C., et al. 2017, arXiv preprint arXiv:1706.03118
- Becker, G. D., Bolton, J. S., Haehnelt, M. G., & Sargent, W. L. W. 2011, MNRAS, 410, 1096
- Becker, G. D., Bolton, J. S., Madau, P., et al. 2015, MNRAS, 447, 3402
- Becker, G. D., Hewett, P. C., Worseck, G., & Prochaska, J. X. 2013, MNRAS, 430, 2067
- Boera, E., Murphy, M. T., Becker, G. D., & Bolton, J. S. 2014, MNRAS, 441, 1916
- . 2016, MNRAS, 456, L79
- Bolton, J. S., Becker, G. D., Haehnelt, M. G., & Viel, M. 2014, MNRAS, 438, 2499
- Bolton, J. S., Becker, G. D., Raskutti, S., et al. 2012, MNRAS, 419, 2880
- Bolton, J. S., Becker, G. D., Wyithe, J. S. B., Haehnelt, M. G., & Sargent, W. L. W. 2010, MNRAS, 406, 612
- Bolton, J. S., Viel, M., Kim, T.-S., Haehnelt, M. G., & Carswell, R. F. 2008, MNRAS, 386, 1131
- Bosman, S. E., Becker, G. D., Haehnelt, M. G., et al. 2017, MNRAS
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, ApJ, 803, 34
- Bryan, G. L., & Machacek, M. E. 2000, ApJ, 534, 57
- Carswell, R. F., & Webb, J. K. 2014, Astrophysics Source Code Library, ascl:1408.015
- Compostella, M., Cantalupo, S., & Porciani, C. 2013, MNRAS, 435, 3169
- Croft, R. A., Weinberg, D. H., Pettini, M., Hernquist, L., & Katz, N. 1999, ApJ, 520, 1
- Croft, R. A. C., Weinberg, D. H., Bolte, M., et al. 2002, ApJ, 581, 20
- Croft, R. A. C., Weinberg, D. H., Katz, N., & Hernquist, L. 1998, ApJ, 495, 44
- Dall'Aglio, A., Wisotzki, L., & Worseck, G. 2008, A&A, 491, 465
- D'Aloisio, A., McQuinn, M., & Trac, H. 2015, ApJ, 813, L38
- D'Aloisio, A., Upton Sanderbeck, P. R., McQuinn, M., Trac, H., & Shapiro, P. R. 2017, MNRAS, 468, 4691
- Davies, F. B., & Furlanetto, S. R. 2016, MNRAS, 460, 1328
- Dekker, H., D'Odorico, S., Kaufer, A., Delabre, B., & Kotzlowski, H. 2000, in Astronomical Telescopes and Instrumentation, Proc. SPIE(International Society for Optics and Photonics), 534–545
- Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, ApJ, 132, 117
- Faucher-Giguère, C.-A., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008a, ApJ, 688, 85
- Faucher-Giguère, C.-A., Prochaska, J. X., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008b, ApJ, 681, 831
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
- Furlanetto, S. R., & Oh, S. P. 2008, ApJ, 681, 1
- Gaikwad, P., Khaire, V., Choudhury, T. R., & Raghunathan, S. 2016, MNRAS, stw3086
- Garzilli, A., Bolton, J. S., Kim, T.-S., Leach, S., & Viel, M. 2012, MNRAS, 424, 1723
- Giallongo, E., Grazian, A., Fiore, F., et al. 2015, A&A, 578, A83
- Gnedin, N., & Hui, L. 1998, MNRAS, 296, 44
- Gnedin, N. Y., Baker, E. J., Bethell, T. J., et al. 2003, ApJ, 583, 525
- Goodman, J., & Weare, J. 2010, CAMCoS, 5, 65
- Gunn, J. E., & Peterson, B. A. 1965, ApJ, 142, 1633
- Habib, S., Heitmann, K., Higdon, D., Nakhleh, C., & Williams, B. 2007, Phys. Rev. D, 76, arXiv:astro-ph/0702348
- Haehnelt, M. G., & Steinmetz, M. 1998, MNRAS, 298, L21
- Heitmann, K., Higdon, D., Nakhleh, C., & Habib, S. 2006, ApJL, 646, L1
- Heitmann, K., Higdon, D., White, M., et al. 2009, ApJ, 705, 156
- Heitmann, K., Lawrence, E., Kwan, J., Habib, S., & Higdon, D. 2013, ApJ, 780, 111
- Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19
- Hiss, H., Walther, M., & Hennawi, J. F. 2017, in prep.
- Hui, L., & Gnedin, N. Y. 1997, MNRAS, 292, 27

<sup>12</sup> <http://enigma.physics.ucsb.edu/>

- Hui, L., Ostriker, J. P., Tremaine, S., & Witten, E. 2017, *Physical Review D*, 95, arXiv:1610.08297
- Iršič, V., Viel, M., Haehnelt, M. G., Bolton, J. S., & Becker, G. D. 2017a, arXiv preprint arXiv:1703.04683
- Iršič, V., Viel, M., Haehnelt, M. G., et al. 2017b, *Phys. Rev. D*, 96, 023522
- Iršič, V., Viel, M., Berg, T. A., et al. 2017c, *MNRAS*, stw3372
- Kim, T.-S., Viel, M., Haehnelt, M. G., Carswell, R. F., & Cristiani, S. 2004, *MNRAS*, 347, 355
- Kirkman, D., Tytler, D., Suzuki, N., et al. 2005, *MNRAS*, 360, 1373
- Kulkarni, G., Hennawi, J. F., Oñorbe, J., Rorai, A., & Springel, V. 2015, *ApJ*, 812, 30
- Lee, K.-G., Suzuki, N., & Spergel, D. N. 2012, *AJ*, 143, 51
- Lee, K.-G., Hennawi, J. F., Spergel, D. N., et al. 2015, *ApJ*, 799, 196
- Lehner, N., O’Meara, J. M., Fox, A. J., et al. 2014, *ApJ*, 788, 119
- Lidz, A., Faucher-Giguère, C.-A., Dall’Aglio, A., et al. 2010, *ApJ*, 718, 199
- Lidz, A., Heitmann, K., Hui, L., et al. 2006, *ApJ*, 638, 27
- Lomb, N. R. 1976, *Ap&SS*, 39, 447
- López, S., D’Odorico, V., Ellison, S. L., et al. 2016, *A&A*, 594, A91
- Lynds, R. 1971, *ApJL*, 164, L73
- Madau, P., & Haardt, F. 2015, *ApJL*, 813, L8
- McDonald, P., Miralda-Escudé, J., Rauch, M., et al. 2001, *ApJ*, 562, 52
- , 2000, *ApJ*, 543, 1
- McDonald, P., Seljak, U., Cen, R., Bode, P., & Ostriker, J. P. 2005, *MNRAS*, 360, 1471
- McDonald, P., Seljak, U., Burles, S., et al. 2006, *ApJS*, 163, 80
- McQuinn, M., Lidz, A., Zaldarriaga, M., et al. 2009, *ApJ*, 694, 842
- McQuinn, M., & Upton Sanderbeck, P. R. 2016, *MNRAS*, 456, 47
- Meiksin, A., & White, M. 2001, *MNRAS*, 324, 141
- Modigliani, A., Bramich, D., Vernet, J., & Ballester, P. 2017, X-Shooter Pipeline User Manual Issue 12.15, ,
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2014, *ApJ*, 786, 108
- O’Meara, J. M., Lehner, N., Howk, J. C., et al. 2017, *ArXiv e-prints*
- , 2015, *AJ*, 150, 111
- Oñorbe, J., Hennawi, J. F., & Lukić, Z. 2017, *ApJ*, 837, 106
- Palanque-Delabrouille, N., Yèche, C., Borde, A., et al. 2013, *A&A*, 559, A85
- Palanque-Delabrouille, N., Yèche, C., Baur, J., et al. 2015, *J. Cosmology Astropart. Phys.*, 2015, 011
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, *A&A*, 571, A16
- Planck Collaboration, Adam, R., Aghanim, N., et al. 2016a, *A&A*, 596, A108
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016b, *A&A*, 594, A13
- Puchwein, E., Bolton, J. S., Haehnelt, M. G., et al. 2015, *MNRAS*, 450, 4081
- Puchwein, E., Pfrommer, C., Springel, V., Broderick, A. E., & Chang, P. 2012, *MNRAS*, 423, 149
- Ricotti, M., Gnedin, N. Y., & Shull, J. M. 2000, *ApJ*, 534, 41
- Rorai, A., Hennawi, J. F., & White, M. 2013, *ApJ*, 775, 81
- Rorai, A., Becker, G. D., Haehnelt, M. G., et al. 2017a, *MNRAS*, 466, 2690
- Rorai, A., Hennawi, J. F., Oñorbe, J., et al. 2017b, *Science*, 356, 418
- Rudie, G. C., Steidel, C. C., & Pettini, M. 2012, *ApJL*, 757, L30
- Scargle, J. D. 1982, *ApJ*, 263, 835
- Schaye, J., Theuns, T., Rauch, M., Efstathiou, G., & Sargent, W. L. W. 2000, *MNRAS*, 318, 817
- Sorini, D., Oñorbe, J., Lukić, Z., & Hennawi, J. F. 2016, *ApJ*, 827, 97
- Stanway, E. R., Eldridge, J. J., & Becker, G. D. 2016, *MNRAS*, 456, 485
- Suarez, T., & Pontzen, A. 2017, arXiv preprint arXiv:1706.02716
- Theuns, T., Zaroubi, S., Kim, T.-S., Tzanavaris, P., & Carswell, R. 2002, *MNRAS*, 332, 367
- Topping, M. W., & Michael Shull, J. 2015, *ApJ*, 800, 97
- Upton Sanderbeck, P. R., D’Aloisio, A., & McQuinn, M. J. 2016, *MNRAS*, 460, 1885
- Viel, M., Becker, G. D., Bolton, J. S., & Haehnelt, M. G. 2013, *Phys. Rev. D*, 88
- Viel, M., Becker, G. D., Bolton, J. S., et al. 2008, *Phys. Rev. Lett.*, 100
- Viel, M., Bolton, J. S., & Haehnelt, M. G. 2009, *MNRAS*, 399, L39
- Vogt, S. S., Allen, S. L., Bigelow, B. C., et al. 1994, in 1994 Symposium on Astronomical Telescopes & Instrumentation for the 21st Century, *Proc. SPIE(International Society for Optics and Photonics)*, 362–375
- Walther, M., Hennawi, J. F., & Onorbe, J. 2017a, in prep.
- Walther, M., Hennawi, J. F., Onorbe, J., & Lukic, Z. 2017b, in prep.
- Weigel, A. K., Schawinski, K., Treister, E., et al. 2015, *MNRAS*, 448, 3167
- White, M. 2002, *ApJS*, 143, 241
- Worseck, G., & Prochaska, J. X. 2011, *ApJ*, 728, 23
- Worseck, G., Prochaska, J. X., Hennawi, J. F., & McQuinn, M. 2016, *ApJ*, 825, 144
- Worseck, G., Prochaska, J. X., McQuinn, M., et al. 2011, *ApJL*, 733, L24
- Yeche, C., Palanque-Delabrouille, N., Baur, J., & BourBoux, H. 2017, arXiv preprint arXiv:1702.03314
- Zaldarriaga, M., Hui, L., & Tegmark, M. 2001, *ApJ*, 557, 519

## APPENDIX

### A. IMPACTS OF SEEING ON POWER SPECTRUM MEASUREMENTS

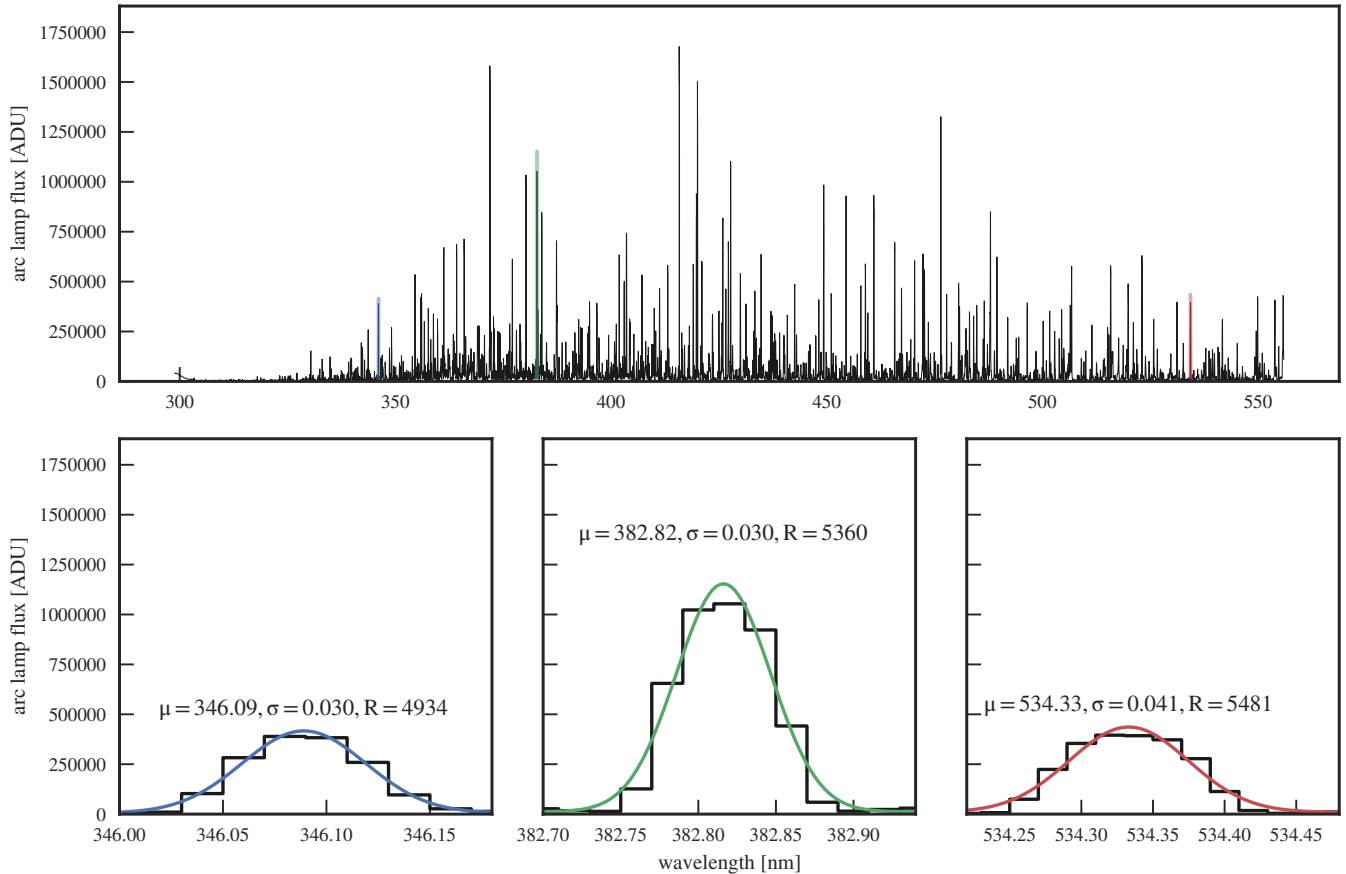
For slit spectrographs, it is very difficult to know the exact resolution because it depends on the seeing, and also the resolution can vary at the  $\sim 10\%$  across the echelle orders, which is typically never carefully quantified or taken into account (e.g. the X-SHOOTER pipeline user manual Modigliani et al. 2017, shows variations in the slit resolution with wavelength at about this level). Note that the change of resolution with seeing is not a problem for fiber spectrographs (such as e.g. BOSS) that allow measurements of the resolution of the science data on sky fibers. Assuming the same resolution for each object (whereas the objects actually have different resolutions) generally will also increase the weight of higher resolution data in the power spectrum averages. This is because higher resolution data has a smaller scale (higher  $k$ ) cutoff. Therefore, a higher power by a factor that is exponential in  $k$  as well as in  $R$  would be measured. When performing the mean over all objects the higher resolution objects (which

now have been overcorrected) will bias the power estimate as  $\langle \exp(k^2 R^2) \rangle > \exp(k^2 \langle R^2 \rangle)$  (if  $R$  is not constant throughout the dataset). As explained in Yeche et al. (2017) this problem can be weakened by measuring the seeing from the data (e.g. by measuring the width of the object in the spacial direction) and correcting each spectrum using its correct seeing.

A similar biasing due to mixing different resolutions can appear already during data reduction because co-adding overlapping echelle orders as well as different observations will give higher weight to data with better seeing as this typically has higher S/N as well. If resolution varies strongly between echelle orders or observations, the final co-added spectrum might be dominated by the best resolution obtained. Therefore, a knowledge of the spectrograph resolution for each individual object on the  $< 10\%$  level as well as an individual resolution correction for each object are needed to provide an accurate measurement.

### B. SLIT RESOLUTION OF THE X-SHOOTER SPECTROGRAPH

The quoted resolution in the X-SHOOTER manual was originally 5100 for the same slit/arm combination and was changed to 4260 during a recalibration run in



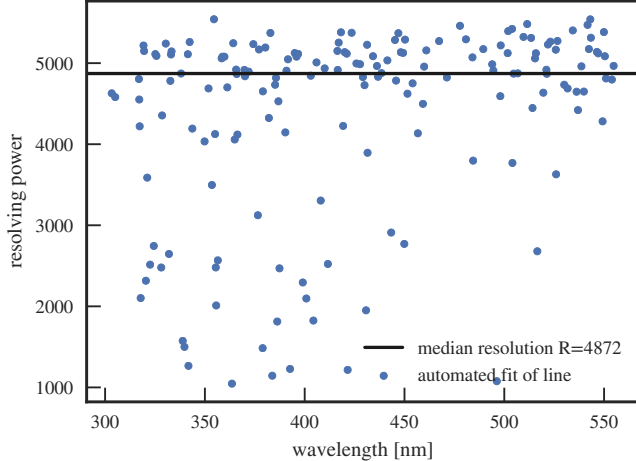
**Figure 11.** A reduced slit-arc spectrum of the UVB channel with the same slit width and binning as the XQ-100 survey and reduced with the `xsh_scired_slit_stare` recipe of the ESO X-Shooter Pipeline using the same calibrations as for science data reductions (except for the response curves and disabling sky-subtraction). The full arc spectrum is shown in the top panel, colored lines indicate the position of zoom-ins in the bottom panel. The bottom panels also show Gaussian fits (colored lines) to three of the arc lines with the best-fit parameters (mean  $\mu$ , standard deviation  $\sigma$ , both in nm) as well as the resulting resolving power  $R$  printed as text. We can clearly see that the resolution of each of those fits is exceeding the nominal value (taken from the XQ-100 data release paper) of  $R = 4350$  and is varying over the spectrum.

2011 (so before XQ-100 data was taken). The full reason for this change in value between calibrations is unclear to the authors. However, the X-SHOOTER Pipeline manual (Modigliani et al. 2017) shows values more consistent with the higher resolution. The manual also claims an underestimation of resolution by the pipeline for the 1x1 binning in the UVB arm. Additionally, the reduction QC plots in the ESO archive for 1x1 binning and 1x2 binning (which XQ-100 used) for all measurements between 2012 and 2014 shows  $\sim 20\%$  higher values for the pipeline resolution values with the 1x2 binning. We contacted ESO about this issue and found out that the difference in estimated resolution for different binnings is still an open problem. A determination of the instruments' slit resolution to the accuracy needed for small scale power analyses is therefore not available.

This motivated us to estimate the X-SHOOTER resolving power for the UVB arm and the configuration used in the XQ-100 dataset ( $0.1''$  slit width and a  $1 \times 2$  binning) from a slit-arc spectrum (taken at 2012-05-20T17:06:41.424) reduced in the same way as one would do for science data using the ESO Pipeline recipe `xsh_scired_slit_stare`. These frames are taken on a regular basis using the same slits used for science targets. A normal reduction process (at least with the ESO

pipeline) does not need those frames as the wavelength calibration is performed using pinhole arcs which cannot be used to determine the resolution of science data. In Figure 11 we show the reduced slit-arc spectrum as well as zoom-ins to three random, non-blended lines in different parts of the spectrum. We fitted the lines with Gaussian profiles which show that the resolution of X-SHOOTER in the UVB arm a) varies within the arm by at least  $\sim 10\%$  and b) can be  $\sim 25\%$  higher than the value quoted in the XQ-100 data release paper (López et al. 2016). We also performed a quick automated fit for all lines in the arc line list individually without checking for line blends and other kinds of contamination. These might make some fits broader than a single line leading to determining a lower resolution. The resulting resolution with respect to wavelength is shown in Figure 12. We do not observe a clear trend with wavelength and note that the bulk of the distribution agrees with  $R \sim 5000$ .

These tests are in basic agreement with the QC plots in the ESO archive as well as the pipeline manual. We therefore assume that the resolution values given in the automatic QC plots are right and use their approximate median of  $R = 5350$  when performing further tests of the XQ-100 results. If one only determines the seeing and estimates spectral resolution by combining the seeing



**Figure 12.** Full slit resolution with respect to wavelength for the X-SHOOTER spectrograph estimated from the spectrum in Figure 11. We show the resolution for a quick fit of individual lines in the arc spectrum based on the line list provided by ESO. Obviously unphysical or unconverged fits have been removed. As fits to blends have not been removed, spurious low-resolution fits are included inside the figure.

with the slit resolution quoted on the ESO website or X-SHOOTER manual one might therefore under-determine the true resolution of the spectra. In principle one should be able to obtain the slit resolution from sky-lines in the science observations as well (albeit with less available lines and worse signal to noise than for the arc spectra) as motions in the atmosphere are slower than the  $\sim 1 \text{ km s}^{-1}$  resolution accuracy we want to obtain.

### C. NORMALIZATION CONVENTIONS FOR THE POWER SPECTRUM

While Kim et al. (2004) as well as the SDSS/BOSS/XQ-100 measurements already used the same normalization as we do, Croft et al. (2002) and McDonald et al. (2000) measure the power of the transmission  $F$  instead of  $\delta_F$ . This leads to an additional factor  $F^2$  between their measurements and the more recent ones. The Croft et al. (2002) measurement also has a different normalization convention by a factor 2 compared to McDonald et al. (2000) which we corrected for. The former also does not provide a measurement of the mean transmission of their sample. Therefore we renormalized the McDonald et al. (2000) measurement using the provided mean transmission of their sample and rescaled the Croft et al. (2002) with the external mean transmission measurement by Becker et al. (2013).

### D. DATA PRODUCTS

Truncated data tables showing our high-resolution measurement at  $z = 2.8$  are shown in Table 5 (including our metal masking approach) and Table 6 (without metal masking). We also show the first and last column of the correlation matrix at  $z = 2.8$  in Table 7. Note

that the correlation matrix is based on the best-fitting DM only simulation. It is therefore a model dependent quantity although the agreement of the fit with the data is good. Thus, for fitting models to the measured power spectra you might want to estimate the correlation matrix from the actual model fitted in the analysis. The full

**Table 4**  
Priors used for each fitting parameter

Parameter	type of prior	lower limit	upper limit	$\mu$	$\sigma$
$\log(T_0/\text{K})$	flat	3.48	4.30		
$\gamma$	flat	0.5	2.1		
$\log(\lambda_J/\text{ckpc})$	flat	1.34	2.18		
$f_{\text{Si III}}$	Gaussian	-0.002	0.018	0.008	0.001
$\bar{F}(z = 1.8)$	Gaussian	0.871	0.931	0.901	0.006
$\bar{F}(z = 2.0)$	Gaussian	0.785	0.976	0.881	0.019
$\bar{F}(z = 2.2)$	Gaussian	0.818	0.921	0.869	0.010
$\bar{F}(z = 2.4)$	Gaussian	0.766	0.859	0.812	0.009
$\bar{F}(z = 2.6)$	Gaussian	0.723	0.813	0.768	0.009
$\bar{F}(z = 2.8)$	Gaussian	0.683	0.771	0.727	0.009
$\bar{F}(z = 3.0)$	Gaussian	0.640	0.724	0.682	0.008
$\bar{F}(z = 3.2)$	Gaussian	0.581	0.661	0.621	0.008
$\bar{F}(z = 3.4)$	Gaussian	0.528	0.602	0.565	0.007

**Note.** — priors in  $\bar{F}$  are based on Kirkman et al. (2005) for  $z = 1.8$ , Faucher-Giguère et al. (2008b) for  $z = 2$  and Becker et al. (2013) for the higher redshifts

**Table 5**  
Measured flux power spectrum at  $z = 2.8$  after masking metals and removing the window function due to masking. Note that the full table that also includes the other redshifts will be available in the electronic edition of the journal.

$\bar{z}$	$\bar{k}$ $\text{s km}^{-1}$	$kP_k\pi^{-1}$	$\sigma_{kP_k\pi^{-1}}$
-----------	---------------------------------	----------------	-------------------------

**Table 6**  
Measured flux power spectrum at  $z = 2.8$  without masking of metals and after removing the window function due to masking. Note that the full table that also includes the other redshifts will be available in the electronic edition of the journal.

$\bar{z}$	$\bar{k}$ $\text{s km}^{-1}$	$kP_k\pi^{-1}$	$\sigma_{kP_k\pi^{-1}}$
-----------	---------------------------------	----------------	-------------------------

tables including all redshifts and  $k$ -bins will be available in machine readable form.



**Table 7**

Correlation matrix at  $z = 2.8$  for the measurement in [Table 5](#). Note that the full table that also includes the other redshifts and matrix elements will be available in the electronic edition of the journal.

$\bar{z}$	$\bar{k}$ s km <sup>-1</sup>	$R_{1,j}$	...	$R_{n,j}$