

# The XQR-30 metal absorber catalogue: 778 absorption systems spanning $2 \lesssim z \lesssim 6.5$

Rebecca L. Davies <sup>1,2</sup>★ E. Ryan-Weber <sup>1,2</sup> V. D’Odorico <sup>3,4,5</sup> S. E. I. Bosman <sup>1,6</sup> R. A. Meyer <sup>1,6</sup> G. D. Becker <sup>1,7</sup> G. Cupani <sup>3</sup> M. Bischetti <sup>3,8</sup> A. M. Sebastian <sup>1,2</sup> A.-C. Eilers <sup>1,9</sup> E. P. Farina <sup>1,10</sup> F. Wang <sup>11</sup> J. Yang <sup>11</sup> and Y. Zhu <sup>1</sup>

<sup>1</sup>Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia

<sup>2</sup>ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia

<sup>3</sup>INAF – Osservatorio Astronomico di Trieste, Via Tiepolo 11, I-34143 Trieste, Italy

<sup>4</sup>IFPU – Institute for Fundamental Physics of the Universe, via Beirut 2, I-34151 Trieste, Italy

<sup>5</sup>Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy

<sup>6</sup>Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

<sup>7</sup>Department of Physics & Astronomy, University of California, Riverside, CA 92521, USA

<sup>8</sup>Dipartimento di Fisica, Sezione di Astronomia, Università di Trieste, via Tiepolo 11, I-34143 Trieste, Italy

<sup>9</sup>MIT Kavli Institute for Astrophysics and Space Research, 77 Massachusetts Ave., Cambridge, MA 02139, USA

<sup>10</sup>Gemini Observatory, NSF’s NOIRLab, 670 N A’ohoku Place, Hilo, Hawai’i 96720, USA

<sup>11</sup>Steward Observatory, University of Arizona, 933 N Cherry Ave., Tucson, AZ 85721, USA

Accepted 2022 December 8. Received 2022 November 22; in original form 2022 August 12

## ABSTRACT

Intervening metal absorption lines in the spectra of  $z \gtrsim 6$  quasars are fundamental probes of the ionization state and chemical composition of circumgalactic and intergalactic gas near the end of the reionization epoch. Large absorber samples are required to robustly measure typical absorber properties and to refine models of the synthesis, transport, and ionization of metals in the early Universe. The Ultimate XSHOOTER legacy survey of quasars at  $z \sim 5.8$ –6.6 (XQR-30) has obtained high signal-to-noise spectra of 30 luminous quasars, nearly quadrupling the existing sample of 12 high quality  $z \sim 6$  quasar spectra. We use this unprecedented sample to construct a catalogue of 778 systems showing absorption in one or more of Mg II (360 systems), Fe II (184), C II (46), C IV (479), Si IV (127), and N V (13) which span  $2 \lesssim z \lesssim 6.5$ . This catalogue significantly expands on existing samples of  $z \gtrsim 5$  absorbers, especially for C IV and Si IV which are important probes of the ionizing photon background at high redshift. The sample is 50 per cent (90 per cent) complete for rest-frame equivalent widths  $W \gtrsim 0.03 \text{ \AA}$  ( $0.09 \text{ \AA}$ ). We publicly release the absorber catalogue along with completeness statistics and a PYTHON script to compute the absorption search path for different ions and redshift ranges. This data set is a key legacy resource for studies of enriched gas from the era of galaxy assembly to cosmic noon, and paves the way for even higher redshift studies with JWST and 30 m-class telescopes.

**Key words:** intergalactic medium – quasars: absorption lines – early Universe.

## 1 INTRODUCTION

High redshift quasars are extremely valuable tools for studying the properties of the Universe during the later stages of the epoch of reionization ( $z \sim 6$ ). Intervening metal absorption lines in quasar spectra encode the redshift, chemical content, and ionization state of gas in and around galaxies along the line of sight. The rest-frame equivalent widths of absorption lines do not decrease as the quasar light moves through space, providing a means to robustly detect and study faint galaxies in the distant Universe that are below the flux limit of current state-of-the-art emission line surveys. The chemical abundance ratios reflect the properties of the stars that synthesized the

metals, offering key insights into the nature of early stars and galaxies (e.g. Cooke et al. 2011; Becker et al. 2012; Bañados et al. 2019; Welsh et al. 2022). The ionization state of the absorbing gas is sensitive to the shape and strength of the ionizing radiation field, providing independent constraints on the timing and sources of reionization (e.g. Finlator et al. 2016; García et al. 2017; Doughty et al. 2018; Götberg et al. 2020). The overall incidence of metal absorbers traces the average metal content of galaxies and the circumgalactic medium (CGM), and can be used to test models of the production and transport of metals in the early Universe (see Becker, Bolton & Lidz 2015a and Tumlinson, Peeples & Werk 2017 for reviews).

Over the last decade, the increasing availability of  $z \gtrsim 6$  quasar spectra has propelled our understanding of the Universe during its infancy. Measurements of Ly $\alpha$  forest dark gaps (e.g. Songaila & Cowie 2002; Furlanetto, Hernquist & Zaldarriaga 2004; Gallerani et al. 2008; Gnedin, Becker & Fan 2017; Nasir & D’Aloisio 2020; Zhu et al. 2021; Zhu et al. 2022) and transmission statistics (e.g.

\* E-mail: [davies@rebeccalouise.net](mailto:davies@rebeccalouise.net)

† NASA Hubble Fellow.

‡ Strittmatter Fellow.

Fan et al. 2006; Becker et al. 2015b; Eilers, Davies & Hennawi 2018; Yang et al. 2020; Bosman et al. 2022) as well as Ly $\alpha$  damping wing absorption (e.g. Davies et al. 2018; Greig et al. 2022) suggest that signatures of reionization persist below  $z \sim 6$ ; in agreement with constraints from independent state-of-the-art probes (e.g. the luminosity function of Ly $\alpha$  emitters; Konno et al. 2018, the fraction of Lyman Break galaxies showing Ly $\alpha$  emission; Mason et al. 2018, and the optical depth to reionization; Planck Collaboration 2020).

Precision measurements of absorber properties at  $z \sim 6$  offer another independent avenue to probe the reionization history of the Universe. If the CGM is primarily ionized by the UV background (rather than by the radiation field of the host galaxy), then we would expect to see a transition from primarily neutral or low-ionization metal lines while reionization is ongoing to more highly ionized lines once reionization is complete (e.g. Oppenheimer, Davé & Finlator 2009; Finlator et al. 2015, 2016; Doughty et al. 2018). There is growing observational evidence for such a shift of metals towards lower ionization environments at  $z \gtrsim 5$ . The incidence of high-ionization absorbers such as C IV  $\lambda\lambda 1548, 1550\text{\AA}$  and Si IV  $\lambda\lambda 1393, 1402\text{\AA}$  declines smoothly over  $1.5 \lesssim z \lesssim 5$  before dropping rapidly at  $z \gtrsim 5$  (e.g. Sargent, Boksenberg & Steidel 1988; Songaila 2001; Boksenberg, Sargent & Rauch 2003; Simcoe 2006; Ryan-Weber, Pettini & Madau 2006; Ryan-Weber et al. 2009; Becker, Rauch & Sargent 2009; Simcoe 2011; D’Odorico et al. 2013; Díaz et al. 2016; Cooper et al. 2019; Meyer et al. 2019; D’Odorico et al. 2022; Davies et al. 2023), whereas the frequency of weak O I  $\lambda 1302\text{\AA}$  absorbers (tracing neutral gas) increases over this same range (e.g. Becker et al. 2006, 2019) and the incidence of Mg II  $\lambda\lambda 2796, 2803\text{\AA}$  absorbers remains approximately constant (e.g. Matejek & Simcoe 2012; Bosman et al. 2017; Chen et al. 2017; Codoreanu et al. 2017; Zou et al. 2021). There is also evidence for an increase in the C II  $\lambda 1304\text{\AA}/\text{C IV}$  ratio at  $z \gtrsim 5.7$  (e.g. Cooper et al. 2019; Simcoe et al. 2020). All of these results suggest that metals at  $z \gtrsim 5$  typically inhabit lower ionization environments than those at lower redshifts.

Metal absorbers also offer a unique probe of stellar populations at high redshift. The chemical abundance ratios of neutral absorbers reflect the yields of stellar and supernova nucleosynthesis, which are strongly dependent on both the masses and metallicities of the progenitor stars (e.g. Cooke et al. 2011). Above some redshift, we expect to observe a shift in the abundance patterns of neutral absorbers to reflect the chemical signatures of metal-free Population III stars (e.g. Simcoe, Sargent & Rauch 2004; Greif et al. 2007; Welsh et al. 2022). Several metal-poor absorption systems have been identified at  $z > 5.5$ , but their abundance ratios appear to be consistent with those measured at lower redshifts (e.g. Becker et al. 2012; D’Odorico et al. 2018; Bañados et al. 2019; Simcoe et al. 2020), suggesting that it may be necessary to push observations to even lower column densities and/or higher redshifts in order to uncover the very metal poor systems harbouring the chemical signatures of the first stars in the Universe.

Finally, higher ionization species such as C IV and Si IV probe enriched gas in galaxy halos and are therefore a unique tracer of outflows in the early Universe. The observed incidence of strong C IV absorbers at  $z \gtrsim 5$  is significantly larger than predicted by cosmological hydrodynamical simulations, which suggests that either the metal outflow rate in the feedback models is too low or the assumed ionizing photon background does not reflect the true conditions in the early Universe (e.g. Keating et al. 2016; Rahmati et al. 2016; Finlator et al. 2020; D’Odorico et al. 2022). Suresh et al. (2015) find that if the feedback is strengthened to reproduce the observed incidence of C IV absorbers, the galaxy star formation rates are reduced to levels that are inconsistent with measured values. This

discrepancy has been further strengthened by the recent observational claims of very massive galaxies at  $z \gtrsim 10$  revealed by JWST (e.g. Adams et al. 2023; Atek et al. 2023; Furtak et al. 2022; Harikane et al. 2022; Labbe et al. 2022; Naidu et al. 2022; Yan et al. 2022). This highlights the power of C IV statistics to provide independent constraints on feedback models.

The Ultimate XSHOOTER legacy survey of quasars at  $z \sim 5.8$ –6.6 (XQR-30; PI: D’Odorico; D’Odorico et al., in preparation) has obtained 30 high signal-to-noise (S/N;  $\geq 10$  per  $10\text{ km s}^{-1}$  spectral pixel at  $1285\text{\AA}$  rest-frame) quasar spectra with the goal of characterizing the properties of the Universe during the second half of reionization. This European Southern Observatory (ESO) large program quadruples the existing sample of 12 archival spectra in the same redshift range observed at comparable S/N and spectral resolution ( $\sim 30\text{ km s}^{-1}$ ). The combined sample of 42 quasars (hereafter referred to as the ‘enlarged XQR-30’ or E-XQR-30 sample) provides the largest homogeneous collection of high quality quasar spectra in the early Universe.

In this paper we present a catalogue of metal absorption systems in the E-XQR-30 spectra. The catalogue was constructed using a combination of automated line finding, custom filtering algorithms and visual inspection. The paper is structured as follows. The quasar sample and data processing are described in Section 2 and the steps taken to construct the absorber catalogue are outlined in Section 3. In Section 4 we present the physical properties of the absorber sample and explore how the sample size and absorption search path compare to previous studies in the literature. The completeness and false-positive rate in the catalogue are characterized in Section 5.2. The released products are described in Section 6 and the work is summarized in Section 7.

Throughout this work we adopt the Planck 2018  $\Lambda\text{CDM}$  cosmology with  $H_0 = 67.7\text{ km s}^{-1}\text{ Mpc}^{-1}$  and  $\Omega_m = 0.31$  (Planck Collaboration 2020).

## 2 OBSERVATIONS AND DATA PROCESSING

### 2.1 Sample

The majority of the spectra used in this work were obtained as part of XQR-30, an ESO Large Program (PI: D’Odorico) which acquired deep, medium resolution ( $R \simeq 10\,000$ ; full width at half-maximum, FWHM  $\simeq 30\text{ km s}^{-1}$ ) observations of 30 bright ( $J_{AB} \leq 19.8$ ) quasars at  $5.8 \lesssim z \lesssim 6.6$  using XSHOOTER (Vernet et al. 2011). The other 12 spectra were sourced from archival XSHOOTER observations of quasars in the same redshift and magnitude range that have similar spectral resolution and S/N to the XQR-30 spectra. The full sample is described in detail in D’Odorico et al. (in preparation). All 42 quasars were targeted based on their redshift and magnitude, with no prior knowledge of intervening absorber properties.

### 2.2 Data reduction and processing

#### 2.2.1 Extracting 1D spectra

The data reduction process is described in Becker et al. (2019). Briefly, for each exposure, a composite dark frame was subtracted before performing sky subtraction on the un-rectified frame (Kelson 2003). An initial 1D spectrum was extracted using the optimal weighting method of Horne (1986) and flux calibration was performed using measurements of a standard star. Telluric corrections were computed using the Cerro Paranal Advanced Sky Model (Noll et al. 2012; Jones et al. 2013) and applied back to the 2D spectra. A

final 1D spectrum was extracted for each of the XSHOOTER spectroscopic arms (VIS and NIR) using all exposures simultaneously to maximize the bad pixel rejection efficiency. The extracted spectra have a velocity sampling of  $10 \text{ km s}^{-1}$  per pixel which corresponds to approximately three spectral pixels per resolution element (see Section 2.2.2). The archival observations were re-reduced using the same method for consistency.

For each quasar, the VIS and NIR spectra were combined into a single spectrum by scaling the NIR spectrum to match the median value of the VIS arm over 990–1015 nm, truncating both spectra at 1015 nm, and stitching them together. We note that the absorption line properties are measured from continuum-normalized spectra (as described in the following section) and therefore the absolute flux scaling of the combined spectrum has no impact on our measurements. In some cases we found small ( $5$ – $20 \text{ km s}^{-1}$ ) offsets between the redshift centroids of transitions of the same ion and absorption system falling in the VIS arm and the NIR arm (e.g. Si II  $\lambda 1260$  versus Si II  $\lambda 1526$ ). These offsets are believed to originate from small spatial misalignments between the slits. When such velocity offsets were identified, the NIR spectra (for which the uncertainties on the wavelength solutions are slightly larger) were shifted by  $\pm 5$ – $20 \text{ km s}^{-1}$  to align the absorption profiles across both arms. A more detailed description of the velocity shifting and its impact on the relative centroids of the line profiles is given in Appendix A.

### 2.2.2 Spectral resolution measurements

When fitting the absorption lines it became apparent that some spectra have significantly higher spectral resolution than the nominal value (which depends on the slit width). This occurs when the average seeing is significantly smaller than the slit width (0.9 arcsec in the VIS arm and 0.6 arcsec in the NIR arm for the XQR-30 spectra). We estimated the true velocity resolutions of the final extracted VIS and NIR spectra for each quasar using seeing data for the individual exposures as described in D’Odorico et al. (in preparation). Briefly, we used high S/N exposures to measure the relationship between the average telluric line width (a single value per exposure and spectral arm) and the seeing for each slit width and spectral arm, used these relationships to estimate the width of the line spread function for each exposure, created a composite profile weighted by the inverse variance of the individual exposures, and measured the width of the composite profile to obtain the final spectral resolution.

The adopted spectral resolutions are listed in Table B1. The resolutions for the VIS spectra range from  $R = 9500$ – $13\,700$  (median 11 400, corresponding to  $\text{FWHM} = 26 \text{ km s}^{-1}$ ) and are higher than the nominal resolution of  $R = 8800$  ( $\text{FWHM} = 34 \text{ km s}^{-1}$ ) for the 0.9 arcsec slit. The NIR resolutions range from  $R = 7600$ – $11\,000$  (median 9750, corresponding to  $\text{FWHM} = 31 \text{ km s}^{-1}$ ) and are comparable to or higher than the nominal resolution of  $R = 7770$  ( $\text{FWHM} = 39 \text{ km s}^{-1}$ ) for the 0.6 arcsec slit. The high spectral resolution of our observations allows us to resolve the metal absorption lines more effectively than some of the largest recent surveys of metal absorbers at  $z \gtrsim 5$  that have velocity resolutions of  $\text{FWHM} \sim 50 \text{ km s}^{-1}$  (e.g. Chen et al. 2017; Cooper et al. 2019; see Section 4.2).

### 2.2.3 Quasar redshifts

We adopted the best quasar redshifts that were available when the construction of the absorption line catalogue commenced. The redshifts were taken from emission line measurements where available (primarily [C II]  $\lambda 158 \mu\text{m}$  as well as CO, Mg II, and Ly $\alpha$  + N V), or from the apparent start of the Ly $\alpha$  forest (see Table B1). The adopted

redshifts are consistent with the final emission redshifts published in D’Odorico et al. (in preparation) to within  $\Delta v \lesssim 300 \text{ km s}^{-1}$ . The redshift values could not be updated if better measurements became available because we only search for absorbers redward of the quasar Ly $\alpha$  emission line and up to a maximum redshift of  $5000 \text{ km s}^{-1}$  above of the quasar redshift (see Section 3.2). However, these small redshift differences do not have any significant impact on the final absorber catalogue.

## 3 ABSORPTION LINE CATALOGUE

The absorption line catalogue was constructed by performing an automated search for candidate systems in each spectrum, filtering the list of candidates using a combination of custom algorithms and visual inspection, and fitting Voigt profiles to obtain the column density ( $\log N \equiv \log N/\text{cm}^{-2}$ ) and Doppler ( $b$ ) parameter of individual absorption components. Many of the steps use routines from ASTROCOOK,<sup>1</sup> a PYTHON library and application for identifying and fitting absorption systems in quasar spectra (Cupani et al. 2020). The following sections provide detailed explanations of all steps.

Throughout the analysis we use only the wavelength region redward of the Ly $\alpha$  emission line of each quasar. Neutral hydrogen in the  $z \sim 6$  intergalactic medium results in saturation of the Ly $\alpha$  forest, making it impossible to identify individual absorption systems at shorter wavelengths.

### 3.1 Continuum normalization

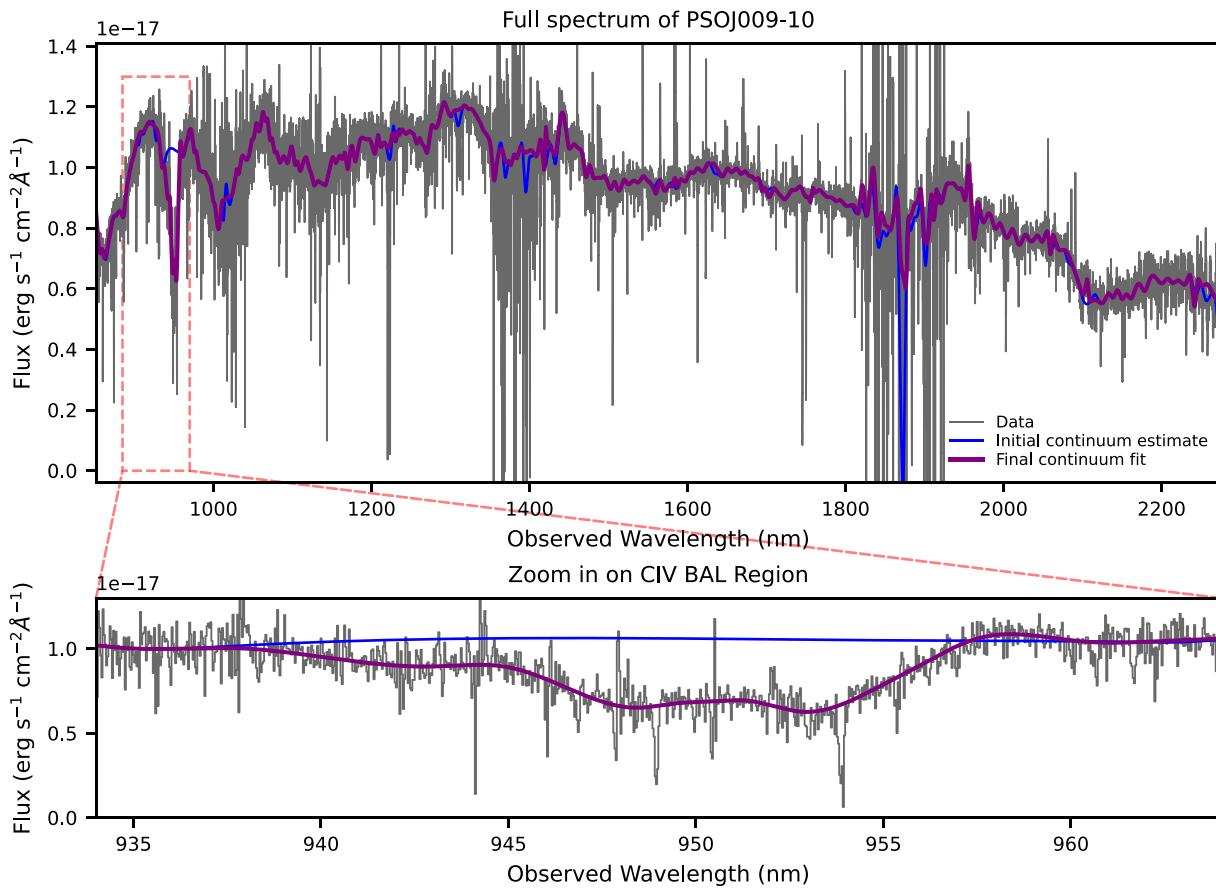
A first estimate of the continuum shape is obtained using the `lines_find` and `nodes_cont` recipes in ASTROCOOK. The `lines_find` routine creates a list of potential lines by convolving the spectrum with Gaussian profiles of different widths (standard deviation  $0 < \sigma \leq 100 \text{ km s}^{-1}$ ) and identifying troughs with a depth  $> 5 \times$  the median error. The `nodes_cont` recipe masks the identified lines, calculates the median wavelength and flux in windows of fixed velocity width, and inserts nodes at these locations. Finally, univariate spline interpolation is applied to the nodes to estimate the shape of the continuum at the wavelength sampling of the original spectra. We adopt a window width of  $770 \text{ km s}^{-1}$  which gives good initial results for our data.

The initial continuum estimates are manually refined by adding and removing nodes as necessary using the ASTROCOOK GUI. Manual refinement is most commonly required in regions where:

- (i) Strong skyline or telluric residuals cause the measured flux to be highly variable, resulting in large oscillations in the estimated continuum level.
- (ii) Clustered narrow absorption lines fill a significant fraction of the  $770 \text{ km s}^{-1}$  window, leaving relatively few unmasked channels to estimate the median flux level.
- (iii) The spectrum has a significant slope which is not well captured at the adopted  $770 \text{ km s}^{-1}$  sampling.
- (iv) There is a prominent broad absorption line (BAL) feature.

A significant fraction of the E-XQR-30 spectra show BAL features (Bischetti et al. 2022; Bischetti et al., in preparation). When processing the spectra of BAL quasars, we attempt to adjust the continuum fit to follow the BAL troughs because this increases the probability of recovering underlying narrow absorption lines. Fig. 1 shows the initial continuum estimate and final continuum fit for the BAL quasar PSOJ009-10. The top panel shows the fits over

<sup>1</sup><https://github.com/DAS-OATs/astrocook>



**Figure 1.** Continuum fit to the spectrum of the BAL quasar PSOJ009-10. The top panel shows the spectrum over the entire range redward of the Ly $\alpha$  emission line, and the bottom panel shows a zoom-in around the C IV BAL feature. In both panels the observed spectrum is shown in grey, the initial continuum estimate returned by ASTROCOOK is shown in blue, and the purple curve shows the final continuum fit after manual adjustments which were primarily required around the BAL feature and in regions with prominent skyline residuals.

the full spectral range redward of Ly $\alpha$  and the bottom panel shows a zoom-in around the C IV BAL feature, where the continuum fit has been manually adjusted to follow the BAL trough. Absorbers found within BAL regions are flagged in the final catalogue (see Section 6.2 and Column 14 in Table B3). The final continuum fits are made publicly available with this paper (see Section 6).

### 3.2 Identification of candidate metal absorption systems

After dividing out the continuum, we search for candidate absorption systems in six of the most commonly observed ions: Mg II, Fe II, C II (combined with O I and Si II  $\lambda 1260\text{\AA}$  when accessible), C IV, Si IV, and N V (specific transitions listed in Table 1). Of these, Fe II, C II, Si II, and O I trace low-ionization gas, C IV, Si IV, and N V trace high-ionization gas, and Mg II traces a range of ionization states. Searching for a combination of ions spanning a wide range of ionization energies enables many interesting science cases. For example, the C IV/Si IV ratio is sensitive to the shape of the ionizing photon background (e.g. Songaila, Hu & Cowie 1995; Songaila 1998; Bolton & Viel 2011; Finlator et al. 2016; Doughty et al. 2018; Graziani, Maselli & Maio 2019), and the C IV/C II ratio directly probes the distribution of carbon between ionization states (see e.g. Prochaska et al. 2015; Cooper et al. 2019; Berg et al. 2021).

We construct a list of candidate absorption systems in each ion by searching for coincident absorption in pairs of transitions. For the

ions observed as doublets (Mg II, C IV, Si IV, and N V), we search for matching absorption in the two transitions of the doublet. For Fe II, we search for matching absorption in any pair of the four strongest transitions ( $\lambda 2344, 2382, 2586$ , and  $2600\text{\AA}$ ). This increases the probability of detection compared to searching for only the two strongest lines because one or more transitions may fall in a region contaminated by skyline or telluric residuals. There is only one strong C II transition in the probed spectral region, so it is necessary to search for associated low-ionization transitions of other ions to confirm the redshift. We search for C II  $\lambda 1334$  paired with O I  $\lambda 1302$ , Si II  $\lambda 1260$ , Si II  $\lambda 1526$ , or Al II  $\lambda 1670$ . If either of Si II  $\lambda 1260$  or O I  $\lambda 1302$  fall within the probed wavelength range but are not detected, the system is rejected because C II  $\lambda 1334$ , O I  $\lambda 1302$ , and Si II  $\lambda 1260$  are expected to have comparable equivalent widths (e.g. Becker et al. 2019). Al II  $\lambda 1670$  and Si II  $\lambda 1526$  are not required but enable the detection of lower redshift systems for which Si II  $\lambda 1260$  and/or O I  $\lambda 1302$  fall outside of the probed wavelength range. The considered transitions and redshift search interval for each ion are listed in Table 1. At this stage we do not search for systems that show O I or Si II absorption without associated C II absorption. We later search for additional transitions associated with all candidate absorption systems (see Section 3.4), meaning that O I and Si II lines could be detected as associated transitions of Mg II or Fe II systems. However, we note that all O I and Si II  $\lambda 1260$  systems in our final catalogue have associated C II absorption, and we recover all O I systems reported by previous surveys (see Section 5.4).

**Table 1.** Transitions and redshift ranges used when searching for candidate absorption systems, unmasked total absorption path lengths (see Section 4.3) and system detection statistics.

Ion	Transitions	z range	$\Delta X$ (Intervening)		$\Delta X$ (all)		Systems
			XQR-30	E-XQR-30	XQR-30	E-XQR-30	
Mg II	Mg II $\lambda 2796$ and $\lambda 2803$	1.944–6.760 (6.381)	400.7	553.3	445.0	612.4	360
Fe II	Pairs of Fe II $\lambda 2344$ , $\lambda 2382$ , $\lambda 2586$ , $\lambda 2600$	2.183–6.759 (6.380)	422.8	594.3	473.3	665.1	184
C IV	C IV $\lambda 1548$ and $\lambda 1550$	4.317–6.760 (6.339)	145.6	211.8	193.8	280.3	479
Si IV	Si IV $\lambda 1393$ and $\lambda 1402$	4.907–6.715 (6.339)	83.6	120.6	126.0	182.5	127
C II	C II $\lambda 1334$ paired with O I $\lambda 1302$ , Al II $\lambda 1670$ , Si II $\lambda 1260$ , $\lambda 1526$	5.169–6.760 (6.381)	55.4	77.8	105.9	148.6	46
N V	N V $\lambda 1238$ and $\lambda 1242$	5.645–6.760 (N/A)	0.0	0.0	33.1	47.6	13

Candidate absorption systems are identified by searching for pairs of transitions with  $\geq 98$  per cent probability of absorption at the same redshift, as described in Section 3.2. Candidate C II systems are rejected if either O I  $\lambda 1302$  or Si II  $\lambda 1260$  fall within the probed wavelength range but are not detected. The bracketed values in Column 3 indicate the maximum search redshift for intervening absorbers. We note that the maximum search redshifts differ between ions because the search paths exclude regions that are masked due to skyline contamination or poorly corrected telluric absorption, as well as regions showing broad absorption line (BAL) features (see Section 4.3).  $\Delta X$ (intervening) gives the path length for intervening absorbers alone (separated by at least  $10\,000\,\text{km s}^{-1}$  from the quasar redshift), whereas  $\Delta X$ (all) also includes the search path for proximate absorbers. We note that Fe II is accessible over a smaller  $z$  range than Mg II but has a larger  $\Delta X$  because Fe II systems can be detected in any two of the four listed transitions, meaning that masking removes less of the search path.

For each line pair, the lower end of the redshift search interval is fixed by the wavelength of the quasar Ly $\alpha$  line and the maximum search redshift is set to  $5000\,\text{km s}^{-1}$  above the quasar redshift.<sup>2</sup> N V is produced by photons with energies  $\geq 77.5\,\text{eV}$  and is therefore expected to be found in systems illuminated by the radiation field of a quasar (e.g. Perrotta et al. 2016). Because of the small wavelength separation between N V  $\lambda 1238$  and Ly $\alpha$   $\lambda 1216$ , the search interval for N V is restricted to ‘proximate’ systems with velocity separations between  $-5460\,\text{km s}^{-1}$  and  $+5000\,\text{km s}^{-1}$  from the quasar redshift (see Section 3.9).

The search for candidate systems is performed using the ASTROCOOK recipe `sysys_new_from_like`. The redshift search interval is divided into bins of a specified width (we adopted  $\Delta z = 10^{-4}$  corresponding to  $\Delta v \simeq 4\text{--}10\,\text{km s}^{-1}$ ). The quasar spectrum is then resampled onto this redshift grid for both transitions of the relevant line pair. For each redshift bin, the absorption S/N (defined as  $(1\text{--flux})/\text{error}$ ) is calculated for each transition and the Gaussian Error Function is used to determine the probability that a feature with this S/N does not trace noise in the spectrum, evaluated as  $P_i = \text{ERF}([\text{S}/\text{N}]/\sqrt{2})$ . The absorption probabilities for the two transitions are combined using an empirical expression<sup>3</sup> to estimate the probability that both transitions show absorption at that redshift. Finally, a peak-finding algorithm is used to identify peaks in redshift space that exceed a given threshold probability and are separated by a specified minimum number of redshift bins. We tested a range of parameters and adopted a threshold probability of  $P > 0.98$  and a minimum peak separation of four bins because these values were found to maximize the recovery rate of visually identifiable absorbers whilst minimizing the number of obviously spurious candidates (e.g. those associated with skyline residuals).

When performing this search, we manually mask wavelength regions with particularly bad skyline or telluric residuals (primarily in the gaps between the  $J$ ,  $H$ , and  $K$  NIR bands at  $\lambda \simeq 1345\text{--}1450\,\text{nm}$  and  $1800\text{--}1940\,\text{nm}$ ). Including these regions would result in a large number of spurious candidate systems. However, individual skylines in other spectral regions are left unmasked. Some spectra show strong absorption features close in wavelength to the quasar Ly $\alpha$  emission line, making it difficult to determine the strength of the overlying continuum. In such cases we mask the affected regions which can extend up to  $\sim 10\text{\AA}$  (observed) redward of the Ly $\alpha$  emission. The masked regions are listed in Table B2.

### 3.3 Automatic filtering of candidate list

#### 3.3.1 Overview

The initial line search outputs six files, each of which records the properties of candidate absorption systems detected in one of the six primary ions. Each entry of each file contains the redshift of one candidate system and the pair of transitions in which the absorption was detected (e.g. 6.1459, [O I  $\lambda 1302$ , C II  $\lambda 1334$ ]). These individual file entries are hereafter referred to as ‘absorption pairs’.

More than 50 per cent of the initial candidates are likely to be false-positives (see Section 5.3). We therefore use a custom filtering algorithm to decrease the amount of contamination in the sample prior to further processing. The filtering algorithm specifically targets spurious absorption systems with two main origins:

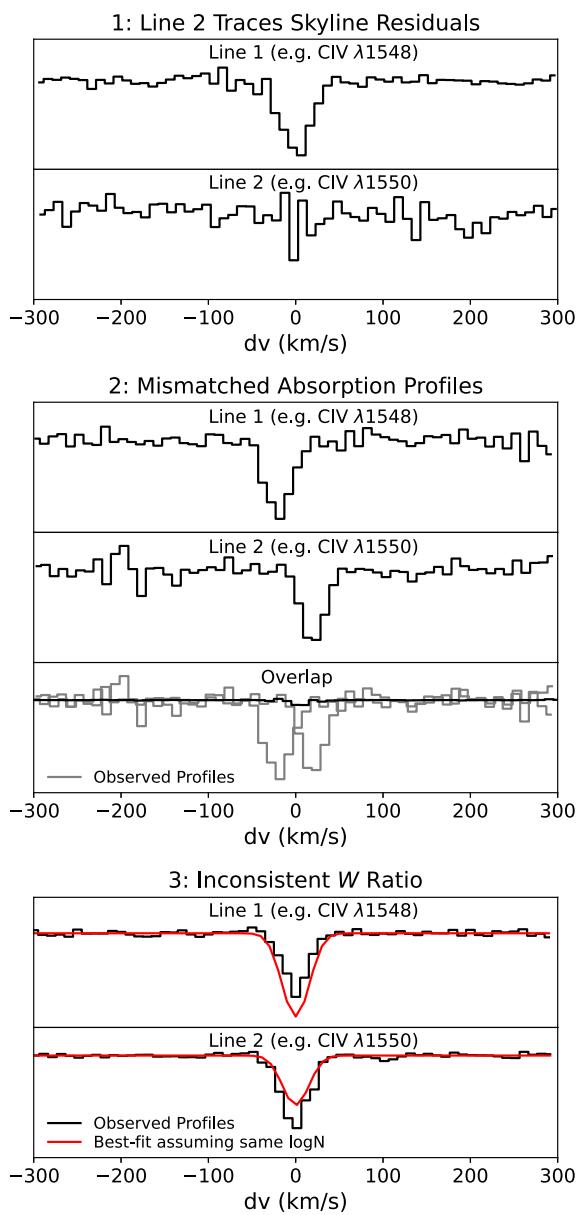
(i) Negative skyline residuals falsely identified as absorption lines (see example 1 in Fig. 2). Candidate systems are classified as spurious if one or both absorption profiles span less than three  $10\,\text{km s}^{-1}$  spectral channels (approximately one resolution element).

(ii) Chance alignment of unassociated absorption lines. This occurs when two absorption lines fall at approximately the velocity separation expected for a given line pair, but actually trace absorption in other ions at different redshifts. Such systems are identified by examining whether the absorption profiles of the two transitions are aligned in velocity space (example 2 in Fig. 2) and, for transitions of the same ion, whether their equivalent width ratio is consistent with expectations from quantum mechanics (example 3 in Fig. 2).

The filtering algorithm is applied to each absorption pair individually. We note that the rejection of any given pair does not

<sup>2</sup>Our final catalogue includes seven absorption systems that lie above the redshift of the relevant quasar. In five of these cases the quasars have redshifts measured from Ly $\alpha$  emission/absorption, which are less accurate than [C II] or CO redshifts.

<sup>3</sup> $P = 1 - (1 - P_1 \times P_2)^2$ . The absorption probabilities for the individual transitions are multiplied, the product is subtracted from 1 which maps peaks to valleys, the expression is squared to enhance extremes and again subtracted from 1 to map the valleys back to peaks. We note that in future versions of ASTROCOOK this expression will be replaced with  $\text{ERF}^{-1}(P_1 \times P_2)$ .



**Figure 2.** Illustration of the criteria used to filter the lists of candidate absorption pairs. The absorption profile of each transition is resampled onto a  $10 \text{ km s}^{-1}$  grid centered at the redshift of the candidate. *Top:* Line 2 is a skyline residual. A candidate is rejected if either absorption line does not span at least three  $10 \text{ km s}^{-1}$  velocity channels (about one resolution element; Section 3.3.2). *Middle and bottom:* In each candidate, both transitions are real absorption lines but they are associated with unrelated absorption systems. In the middle panel, the absorption profiles are misaligned in velocity space. We reject candidates that do not have significant absorption overlapping between the two transitions (Section 3.3.3). The candidate in the bottom panel is rejected because the observed equivalent widths (black) are inconsistent with Voigt profile fits assuming a single column density (red; Section 3.3.4).

preclude the acceptance of other absorption pairs at the same redshift. As described in Section 3.2, the absorption pairs mostly trace two transitions of the same ion, except in the case of  $\text{CII} \lambda 1334$  where the accompanying line is one of  $\text{OI} \lambda 1302$ ,  $\text{AlII} \lambda 1670$ ,  $\text{SiII} \lambda 1260$ , or  $\text{SiII} \lambda 1526$ .

The automatic filtering reduces the number of candidate absorption systems by  $\sim 35\text{--}50$  per cent depending on the ion. It is not possible to

determine what percentage of the rejected systems are real. Based on the mock absorber catalogue used for the completeness calculations (see Section 5.1), we estimate that this filtering only reduces the sample completeness by a few per cent while lowering the false-positive rate by  $\sim 20$  per cent, from an initial value of  $\sim 70$  per cent to a post-filtering value of  $\sim 50$  per cent (see Section 5.3).

### 3.3.2 Spurious candidates associated with skyline residuals

A significant fraction of the spurious candidates arise due to unmasked negative skyline residuals. The top panel of Fig. 2 shows an example of such a candidate. Each row shows the absorption profile for one transition of the absorption pair (labelled Line 1 and Line 2), resampled onto a velocity grid centered at the expected absorption wavelength. In this candidate, Line 1 traces real absorption but Line 2 is clearly a skyline residual. We minimize contamination from skyline residuals by rejecting candidates for which the putative absorption does not span at least three  $10 \text{ km s}^{-1}$  velocity channels (approximately one resolution element) in both transitions. We emphasize that this filtering does not preclude Line 1 from being accepted as part of a different absorption pair.

### 3.3.3 Velocity alignment of absorption profiles

The majority of the remaining spurious candidates arise when two real absorption lines are observed at a velocity separation similar to that expected for the relevant absorption pair, but they actually trace other transitions at different redshifts. For example,  $\text{OI} \lambda 1302$  and  $\text{SiII} \lambda 1304$  are separated by  $507 \text{ km s}^{-1}$  and may therefore masquerade as  $\text{CIV} \lambda\lambda 1548, 1550$  which has a separation of  $499 \text{ km s}^{-1}$ . Such cases are identified in two ways. First, we examine whether the absorption profiles of the two transitions are aligned in velocity space by computing the overlap signal,

$$\text{Overlap}(v) = 1 - (1 - \text{Line 1}(v)) \times (1 - \text{Line 2}(v)). \quad (1)$$

The middle panel of Fig. 2 shows the absorption profiles of a candidate absorption pair (top and middle rows) and the overlap signal between them (bottom row). The two absorption profiles are offset in velocity space, and as a result there is no significant overlap signal.

We use ASTROCOOK to fit a Voigt profile to the overlap signal for each candidate absorption pair (see Section 3.6). An absorption pair is only rejected if the best-fitting column density is equal to the lower bound adopted in the fitting ( $\log(N/\text{cm}^{-2}) = 10$ ), because this indicates that the measured  $\log N$  is an upper limit and therefore that there is no significant overlap between the absorption profiles. We emphasize that this criterion does not require that the two absorption profiles have exactly the same velocity structure, it only requires that there is some overlapping absorption. This is to prevent the spurious rejection of systems that have multiple kinematic components and/or are blended with unassociated absorption lines at different redshifts. The absorption pairs trace either two transitions of the same ion which are expected to have the same kinematics, or in the case of pairs detected using  $\text{CII}$ , two low-ionization transitions that are expected to have very similar kinematics (see Section 3.6).

### 3.3.4 Equivalent width ratio test

For candidate absorption pairs where Line 1 and Line 2 are transitions of the same ion (i.e. pairs detected in all ions *except*  $\text{CII}$ ), we also check whether the relative equivalent widths ( $W$ ) of the two absorption profiles are consistent with expectations from quantum

mechanics. We fit Voigt profiles to the two transitions assuming that they have the same  $z$ ,  $b$  parameter, and  $\log N$ , measure the equivalent widths of these best-fitting profiles, and use these to calculate the ‘expected’ ratio and difference between the equivalent widths of the stronger and weaker transition ( $W_{\text{strong}}/W_{\text{weak}}$  and  $W_{\text{strong}} - W_{\text{weak}}$ ). By computing the expected parameters in this way we naturally account for changes in the relative equivalent width as the absorption optical depth increases. We then repeat the Voigt profiles fits, tying the  $z$  and  $b$  parameter as before but this time allowing the two transitions to have different  $\log N$  values. This second set of Voigt profiles is then used to calculate the ‘measured’  $W$  ratio and difference. We reject pairs for which the measured  $W$  ratio and/or difference are inconsistent with the expected values at the  $> 1\sigma$  level (see Section 3.8 for a description of the  $W$  error calculations). The bottom panel of Fig. 2 shows an example of a candidate absorption pair for which the equivalent widths of the observed profiles (black) are inconsistent with expectations for absorption at a single column density (red).

If enforced uniformly, this criterion could incorrectly reject systems for which one or both of the transitions is blended with unassociated strong absorption at a different redshift. When multiple transitions fall within the full width at ten percent maximum (FWTM) of a given absorption profile, we conservatively treat the expected equivalent width calculated for that profile as a lower limit. This also applies to absorption profiles with multiple kinematic components. A multicomponent absorption profile may be identified as a single candidate system (e.g. when there is one dominant component accompanied by other weaker components, such as in rows 2 and 3 of Fig. 3) or may be split into multiple candidates (e.g. when the absorption profile shows multiple clear peaks as seen in Fig. B1). In the first scenario, the measured  $W$  ratio reflects the relative equivalent widths of the dominant component. In the second scenario, the  $W$  ratio test is applied to each component separately. As before, if any component is within the FWTM of another component, the expected  $W$  is treated as an upper limit.

The equivalent width check cannot be applied to C II pairs because we probe only one C II transition. To minimize the fraction of spurious C II systems we instead require candidates to be detected in at least three low-ionization transitions (see Section 3.4).

### 3.3.5 Removal of duplicates

Noise in the spectrum can cause the peak-finding algorithm to identify a single absorption component multiple times. To minimize the number of duplicates in our candidate list we remove candidate systems of the same ion with redshift separation  $|\Delta z| < 5 \times 10^{-4}$  (corresponding to velocity separations  $\lesssim 20\text{--}50 \text{ km s}^{-1}$ ; approximately one resolution element).

## 3.4 Identification of candidate associated transitions

The next step is to search for additional transitions<sup>4</sup> associated with each candidate absorption system. We do this using the ASTROCOOK recipe `syscts_complete_from_like` which employs the same algorithm used to identify candidate absorption systems

<sup>4</sup>We consider all of the transitions listed in Table 1 as well as Fe II  $\lambda 1608$ , Fe II  $\lambda 2374$ , C II\*  $\lambda 1335$ , Si II  $\lambda 1304$ , Si II  $\lambda 1808$ , Si II\*  $\lambda 1264$ , Mg I  $\lambda 2852$ , Al III  $\lambda 1854$ , Al III  $\lambda 1862$ , Mn II  $\lambda 2576$ , Mn II  $\lambda 2594$ , Mn II  $\lambda 2606$ , S II  $\lambda 1250$ , S II  $\lambda 1253$ , S II  $\lambda 1259$ , Co II  $\lambda 1466$ , Co II  $\lambda 1574$ , Co II  $\lambda 1941$ , Co II  $\lambda 2012$ , Cr II  $\lambda 2056$ , Cr II  $\lambda 2062$ , Cr II  $\lambda 2066$ , Ca II  $\lambda 3934$ , Ca II  $\lambda 3969$ , Na I  $\lambda 5891$ , and Na I  $\lambda 5897$ .

in Section 3.2, but this time calculates the probability that an additional transition is present at the redshift of a given system. When performing this search, the wavelength masks used in Section 3.2 are removed to enable the identification of associated transitions lying in noisy spectral regions. Searching for transitions at the redshifts of previously identified systems significantly increases the probability that the recovered absorption is real compared to a blind search. We only accept associated transitions for which the absorption significance exceeds  $2\sigma$ .

At the conclusion of this step, we reject any C II systems that are not detected in at least two other strong low-ionization transitions<sup>5</sup> (this removes  $\sim 20$  per cent of the candidates). We do this because the equivalent width ratio test described in Section 3.3 cannot be applied to C II, so the probability of spurious candidates resulting from the chance alignment of unassociated transitions in velocity space is elevated compared to the other ions. Requiring the detection of three transitions (including C II  $\lambda 1334$ ) significantly reduces the number of spurious candidate C II systems.

## 3.5 Visual inspection

The remaining candidate systems were visually inspected by five members of the XQR-30 collaboration independently. For each ion in each absorption system, the checkers were asked to indicate whether they thought any real absorption was present. This step is necessary, even for systems with candidate absorption detected in several transitions, because the associated transitions found in the previous step could trace skyline residuals or unrelated absorption lines at other redshifts as discussed in Section 3.3. Any absorption classified as spurious by at least three of the five checkers was removed. This visual inspection further reduces the number of candidate absorption systems by  $\sim 50\text{--}75$  per cent depending on the ion.

## 3.6 Voigt profile fitting

Finally, we use the ASTROCOOK GUI to fit Voigt profiles and obtain the column density and  $b$  parameter for each component of each absorption system. At this stage there are a total of 964 candidate absorbers of which 505 are detected in Mg II, 262 in Fe II, 263 in C II, 727 in C IV, 182 in Si IV, and 31 in N V. The candidate list indicates which transitions show absorption at which redshifts, but it was often necessary to add multiple Voigt profile components per candidate in order to reproduce the observed velocity structure of the absorption profiles.

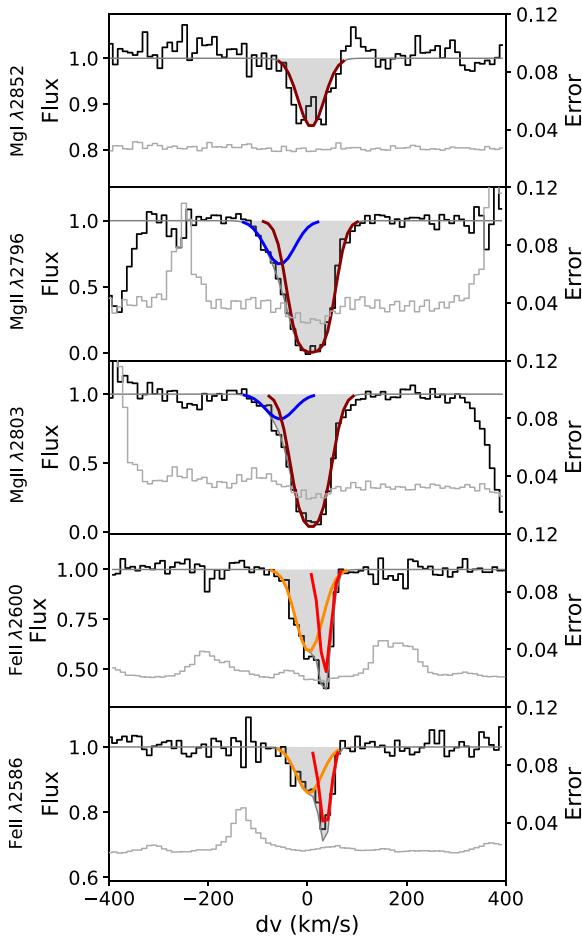
The Voigt parameters are allowed to vary in the ranges  $b = 5\text{--}200 \text{ km s}^{-1}$  and  $N = 10^{10}\text{--}10^{18} \text{ cm}^{-2}$ . If the best-fitting  $b$  parameter of any component is below the minimum measurable value,<sup>6</sup> we fix it to that minimum value. The reported uncertainties on  $\log N$  and  $b$  are the  $1\sigma$  statistical errors returned by the LMFIT least squares minimization routine used within ASTROCOOK.

Where possible, the kinematics of all low-ionization (neutral or singly ionized) transitions and the kinematics of all high-ionization (doubly ionized or higher) transitions are tied within each component.<sup>7</sup> Many systems show both low-ionization and high-ionization

<sup>5</sup>O I  $\lambda 1302$ , Si II  $\lambda 1260$ , Si II  $\lambda 1526$ , Al II  $\lambda 1670$ , Fe II  $\lambda 2382$ , Fe II  $\lambda 2600$ , Mg II  $\lambda 2796$ , and Mg II  $\lambda 2803$

<sup>6</sup>The minimum measurable value is  $\sim 1/3$  of the  $b$  parameter corresponding to the resolution element,  $b_{\text{min}} \simeq \frac{1}{3} \frac{c}{R \times 2\sqrt{\ln 2}}$

<sup>7</sup>The version of ASTROCOOK used in this paper (v1.0.0) ties the  $b$  parameter by fitting a single value for all transitions, which is equivalent to assuming that the turbulent component is dominant.



**Figure 3.** Four component fit to a low-ionization absorption system at  $z = 2.72$ . Each colour indicates a different kinematic component. The Mg II profiles (rows 2 and 3) show a blueshifted wing that is not detected in the other ions. The Fe II profiles (rows 4 and 5) are well fit by two components (orange and red). However, the corresponding central absorption trough in Mg II (rows 2 and 3) can only be fit with a single component (brown) due to saturation. The Mg I absorption profile (row 1) is fit using the same central component (brown) because the low S/N prevents a robust multicomponent decomposition.

transitions, and the transitions in each ionization group are fit separately. However, for  $\sim 15$  per cent of absorption systems it is not possible to fit all ions from one or both ionization categories using the same kinematic structure.<sup>8</sup> Some rare systems show evidence of intrinsic differences between the kinematics of different low-ionization ions, but in most cases the main limitations are saturation and S/N, as illustrated in Fig. 3. In this low-ionization absorption system at  $z = 2.72$ , the Mg II absorption profile (rows 2 and 3) shows a blue-shifted wing which is not detected in either Fe II or Mg I. ASTROCOOK does not provide upper limits on column densities of undetected components, so if the blue wing is fit to the Fe II and Mg I profiles, the best-fitting  $\log N$  tends towards the minimum allowed value in both cases. Therefore, we remove these components.

In this system, the absorption at  $\Delta v \simeq 0$  is also fit with a differing number of components depending on the ion. The Fe II absorption profile is well fit by two components (shown as orange and red

<sup>8</sup> Of these 15 per cent,  $\sim 80$  per cent are low-ionization systems and  $\sim 20$  per cent are high-ionization.

curves). However, the Mg II absorption profiles are saturated, and attempting to fit them using three components (the blue wing plus the two Fe II components) results in large errors on the column densities of the individual components, indicating that the multicomponent decomposition is poorly constrained. Therefore, the Mg II profile is fit using two components: the blue wing, and a component encompassing the whole central absorption trough (brown curve). The Mg I absorption is relatively weak and the low S/N prevents a robust multicomponent decomposition, so it is fit using the same single central component.

We emphasize that the kinematics are always tied when like-ionization transitions are fit using the same kinematic structure (same number of components at the same velocities). Differing kinematics are only used in cases where it is not possible to obtain well constrained fits for all low-ionization or high-ionization ions with a single set of kinematic parameters.

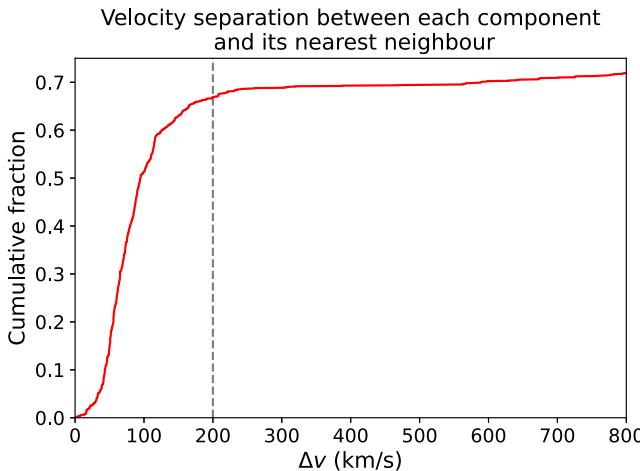
We find six systems for which different transitions of the same ion cannot be fit using the same  $\log N$ , perhaps due to partial coverage or errors in continuum normalization. These systems were not rejected by the equivalent width filtering criterion either because the discrepancy was smaller than the associated uncertainty or because there were multiple candidate absorption systems contributing to the absorption profile(s) (see Section 3.3). For these systems, we fit each transition independently and the  $\log N$  and  $d\log N$  given in the catalogue represent the average and half-range of the individual  $\log N$  values, respectively. This results in larger errors which more accurately reflect the uncertainty compared to the errors returned by ASTROCOOK. The  $\log N$  values for these systems are flagged (with ‘M’ for multiple in Column 10; see Table B3) to indicate that they are a combination of multiple measurements. The reported equivalent widths are measured from the best-fitting Voigt profile for each individual line.

After completing the fitting we perform an automated search for unidentified absorption features in the quasar spectra. Specifically, we searched for regions where the measured flux was significantly below the best continuum + line model obtained from the manual Voigt profile fitting. In some cases, unidentified absorption was attributed to less common transitions that were not part of the initial linelist but were clearly detected in a few systems (e.g. Zn II and Ti II). In other cases, the absorption could be matched to candidate systems that were rejected at the visual inspection phase due to low S/N and/or skyline contamination. In such cases the rejected systems were added back as the best explanation for the observed absorption. The final catalogue includes flags to indicate whether each system was automatically or manually identified, and whether or not it passed the visual inspection phase (see Section 6.2 and Column 14 of Table B3).

The numbers of systems that were manually added and removed during the fitting stage were sufficiently small that the total number of candidate systems was approximately unchanged.

### 3.7 Grouping into systems

We group the components into systems following a method similar to that described in D’Odorico et al. (2022). For each quasar we iterate through the list of systems and combine components that are separated by less than  $200 \text{ km s}^{-1}$  into a single system characterized by the total equivalent width and equivalent-width-weighted average redshift of the constituent components. This process is repeated until no components are separated by less than  $200 \text{ km s}^{-1}$ . The grouping is applied to all transitions of all ions (both low and high ionization)



**Figure 4.** Cumulative distribution of the velocity separation between each absorption component and the closest neighbouring component of the same transition. The cumulative fraction increases rapidly up to a velocity separation of  $200 \text{ km s}^{-1}$  and then plateaus, indicating that  $200 \text{ km s}^{-1}$  is a physically motivated threshold to adopt when grouping components into systems.

simultaneously, meaning that a single system can contain both low and high ionization transitions.

The  $200 \text{ km s}^{-1}$  threshold was chosen by examining the distribution of velocity separations between each absorption component and the nearest neighbouring component of the same transition, shown in Fig. 4. The cumulative fraction increases rapidly up to a velocity separation of  $200 \text{ km s}^{-1}$  and then plateaus, indicating that the vast majority of physically associated components have velocity separations less than  $200 \text{ km s}^{-1}$ . We note that the catalogue contains all the information required to generate the line profiles and implement a different grouping method if needed.

### 3.8 Calculating equivalent width and $\Delta v_{90}$

We measure the rest-frame equivalent width ( $W$ ) of each transition in each system using the Voigt profile fits. This avoids errors associated with skyline spikes and blending with absorption systems at other redshifts. The error on the equivalent width is computed by summing the flux error in quadrature over the region enclosing 90 per cent of the absorption optical depth.

$\Delta v_{90}$  is defined as the velocity interval enclosing 90 per cent of the absorption optical depth (Prochaska & Wolfe 1997). We measure  $\Delta v_{90}$  using the apparent optical depth ( $\tau_a$ ) profile of the strongest transition of each ion in each system. We construct the cumulative distribution function (CDF) for  $\tau_a$  as a function of velocity, and calculate the difference between the velocities at which the CDF equals 0.95 and 0.05.

### 3.9 Classifying absorbers as intervening or proximate

Absorbers are typically classified as either intervening or proximate. Proximate (also known as ‘associated’ or ‘intrinsic’) absorbers are traditionally defined to lie within  $\sim 5000 \text{ km s}^{-1}$  of the quasar redshift. Due to their close proximity to the quasar, such absorbers may have different ionization states or metal abundance patterns to the underlying absorber population (e.g. Foltz et al. 1986; Berg et al. 2016; Perrotta et al. 2016). For this reason, it is important to distinguish between intervening and proximate absorbers when investigating e.g. the evolution of absorber properties over cosmic time.

Perrotta et al. (2016) found an  $8\sigma$  excess of C IV absorbers (compared to the average incidence) at velocity separations extending down to  $-10000 \text{ km s}^{-1}$  from the quasar redshift (see also Wild et al. 2008). The excess at the largest absolute velocity offsets ( $-10000$  to  $-5000 \text{ km s}^{-1}$ ) is dominated by weak C IV systems (with equivalent width  $W_{1548} < 0.2 \text{ \AA}$ ), suggesting that the velocity threshold for proximate absorbers should be extended when weak systems are included. Our absorber sample is 90 per cent complete at  $W > 0.1 \text{ \AA}$  (see Section 5.2) and most of the C IV absorbers in our sample are in the  $W_{1548} < 0.2 \text{ \AA}$  regime (see Section 4.2). Therefore, we classify all systems that have velocity offsets between  $-10000 \text{ km s}^{-1}$  and  $+5000 \text{ km s}^{-1}$  from the quasar redshift as proximate absorbers. However, we note that the data published with this paper enable users to calculate the velocity offset of each absorber and apply a different velocity threshold if desired.

## 4 SAMPLE PROPERTIES

### 4.1 Overview

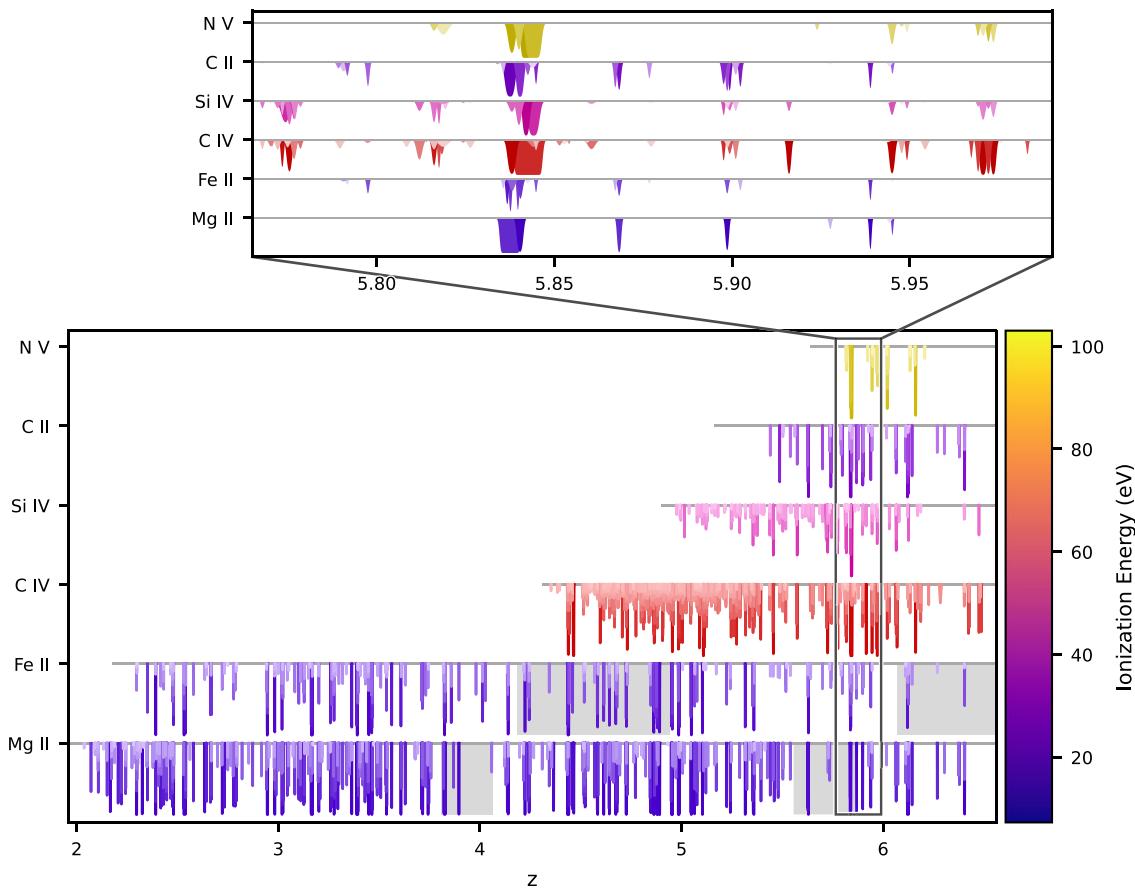
The final catalogue contains a total of 778 systems including 479 C IV absorbers, 360 Mg II absorbers, and 46 C II absorbers (see Table 1). This sample represents a significant increase in the number of  $z > 5$  absorbers observed at high spectral resolution and S/N compared to existing samples, as discussed in Section 4.3. Fig. 5 illustrates the significant statistical power of the full data set. The bottom panel displays all absorption systems detected in each of the primary ions (Mg II, Fe II, C IV, Si IV, C II, and N V) in all E-XQR-30 spectra combined. The horizontal lines show the redshift range over which each ion can be detected, and the grey bands indicate regions where one or more transitions of the relevant ion fall in noisy spectral regions (see Section 5.2). Each system is plotted using a hue that represents the energy required to ionize the relevant ion (according to the colour scale on the right). The number of systems is so large that many of them fall on top of one another in this collapsed redshift space. To ensure that all systems are visible, stronger systems are plotted in darker shades in the background with weaker systems in lighter shades.

The top panel is a zoom-in on the region around  $5.77 \lesssim z \lesssim 6$  and illustrates the kinematics and ionization structure of a few example systems. In this panel, every second component is plotted using a slightly darker shade to better highlight the locations of individual components. The absorption systems show a wide variety of properties, including narrow single component low-ionization systems seen only in Mg II, Fe II, and C II, multicomponent high-ionization systems seen in C IV, Si IV, and N V, and multiphase systems detected in both low and high-ionization transitions. The rich multiphase nature of this data set will provide crucial insights into the ionization state of the CGM during the later stages of reionization. We observe absorption lines with a variety of strengths, from strongly saturated systems spanning a wide velocity range to very weak systems just above the detection limit. The large equivalent width range will enable us to robustly measure the properties of systems down to low metallicities and/or column densities.

### 4.2 Absorber statistics

#### 4.2.1 Primary ions

The physical properties of the Mg II, C IV, Si IV, and C II absorbers are summarized in Fig. 6. The redshift, equivalent width,  $\Delta v_{90}$  and  $\log N$  distributions show the properties of absorption systems, whilst



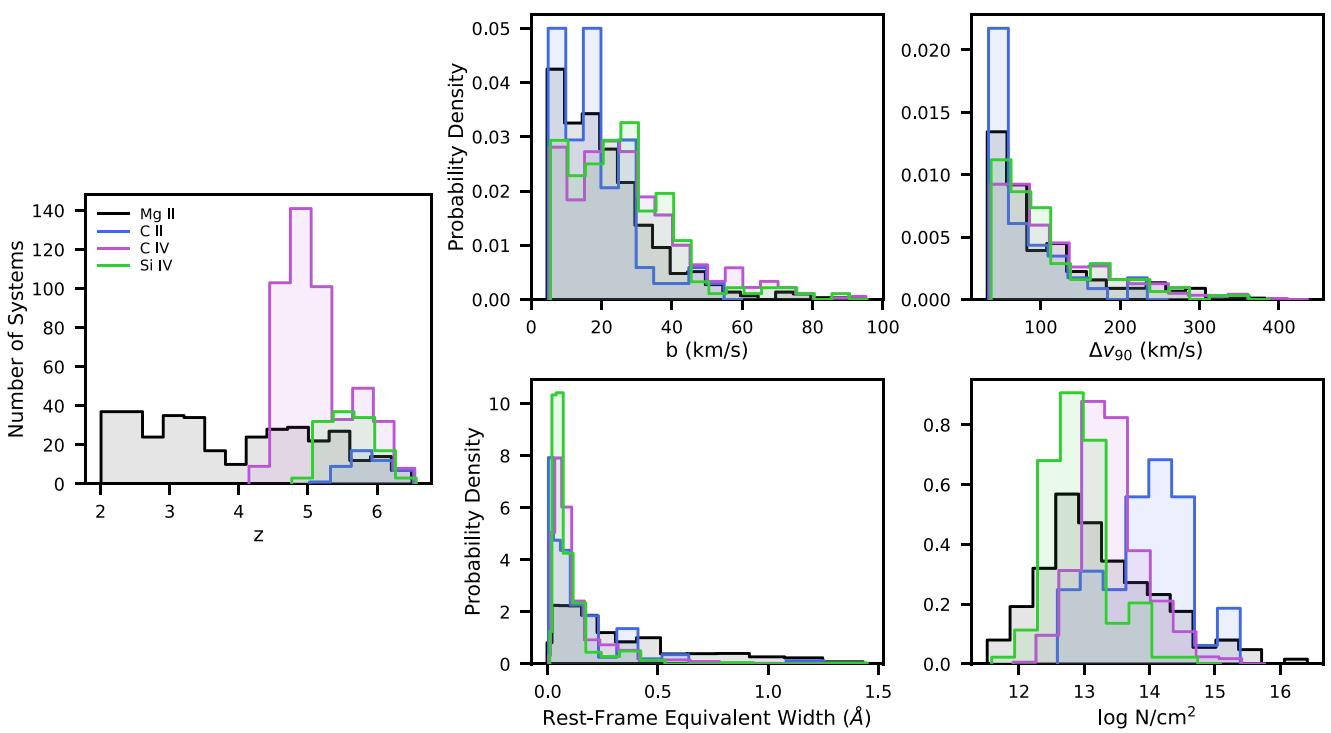
**Figure 5.** Visualisation of the absorber catalogue presented in this paper. In the bottom panel, each row shows all of the absorption systems detected in one of the six primary ions across all of the E-XQR-30 spectra combined. For each system, the hue represents the energy required to ionize the relevant ion (according to the colour scale on the right). To increase the contrast between different systems, stronger systems are plotted in darker shades in the background and weaker systems are plotted in lighter shades in the foreground. The horizontal lines show the redshift range over which each ion can be detected. The grey shaded bands indicate areas where one or more transitions of the relevant ion fall in noisy spectral regions (see Section 5.2). The top panel shows a zoom-in on the redshift range  $5.77 \lesssim z \lesssim 6$ , and illustrates the range of different kinematics and ionization structures found within the absorber sample. In this panel, every second component is plotted using a slightly darker shade for clarity. An animated version of this plot can be found in the GitHub repository.

the  $b$  parameter distribution necessarily reflects the properties of individual Voigt profile components.

The left-most panel shows the distributions of absorber redshifts. Mg II absorbers (black) are detected relatively uniformly across the probed redshift range, with a couple of dips corresponding to wavelength regions with elevated contamination from skyline and/or telluric residuals (see Section 5.2). This data set therefore allows us to self-consistently probe the evolution of Mg II absorption over 2.5 Gyr of early cosmic history. Si IV (green) and C II (blue) absorbers do not show any strong variations in incidence as a function of redshift, which may be due to the small redshift intervals over which these ions are accessible. The incidence of C IV absorbers (purple) peaks strongly at  $z \sim 4.9$  before declining rapidly towards higher redshifts. This is caused by a combination of the difference in resolution and S/N between the XSHOOTER VIS and NIR spectra, and an intrinsic drop in the occurrence of C IV absorbers at  $z \gtrsim 5$  (e.g. Songaila 2001; Simcoe 2006; Ryan-Weber et al. 2006, 2009; D’Odorico et al. 2013; Díaz et al. 2016; Codoreanu et al. 2018; Meyer et al. 2019; D’Odorico et al. 2022; see Davies et al. 2023 for a detailed analysis of C IV absorber statistics using the E-XQR-30 sample). The decline in C IV absorption at lower redshifts is the result of C IV shifting into the inaccessible Ly $\alpha$  forest region.

The  $b$  parameter distributions for all four ions (upper middle panel) show broad peaks at  $10-20 \text{ km s}^{-1}$  with prominent tails towards larger values. The fraction of systems in the tail (with  $b \gtrsim 30 \text{ km s}^{-1}$ ) is higher for C IV and Si IV (32–35 per cent) than Mg II and C II (12–21 per cent), consistent with the expectation that higher ionization absorbers should arise from warmer and more turbulent material on average (see also Wolfe et al. 1993; Rauch et al. 1996; Lehner et al. 2014; Muzahid 2014; Fox et al. 2015; Cooper et al. 2019; Pradeep et al. 2020). We do not observe strong peaks in the lowest  $b$  parameter bin because there were relatively few cases where the  $b$  parameter could not be reliably measured and had to be fixed to the minimum measurable value. In other words, the majority of components are at least partially resolved in our data.

The  $\Delta v_{90}$  distributions (upper right-hand panel) peak at  $35-60 \text{ km s}^{-1}$ , indicating that the most common type of absorption system is a single Voigt component with a  $b$  parameter of  $5-20 \text{ km s}^{-1}$ . Systems with multiple kinematic components are also prevalent as indicated by the long tails which extend up to velocity widths of  $400 \text{ km s}^{-1}$ . The  $\Delta v_{90}$  distributions are similar for all ions, with a slight preference towards lower velocity widths in C II (and to a lesser extent Mg II) absorbers which may reflect their typically lower  $b$  parameters.



**Figure 6.** Distributions of redshift ( $z$ ; left),  $b$  parameter (upper middle),  $\Delta v_{90}$  (upper right), rest-frame equivalent width ( $W$ ; lower middle), and column density ( $\log N$ ; lower right) for Mg II (black), C II (blue), C IV (purple), and Si IV (green) absorbers, including both intervening and proximate systems. The equivalent width and  $\Delta v_{90}$  are measured for the strongest line of each doublet. The  $z$ ,  $\Delta v_{90}$ ,  $W$ , and  $\log N$  distributions reflect the properties of absorption systems (described in Section 3.7) whilst the  $b$  parameter distribution is necessarily based on measurements for individual Voigt profile components. The same bin boundaries are used for all ions but the distributions are artificially offset from one another for clarity.

The rest-frame equivalent width distributions (lower middle) similarly peak at  $W \simeq 0.1 \text{ \AA}$  with long tails towards larger values. The Si IV distribution is most strongly concentrated in the weak absorber ( $W < 0.2 \text{ \AA}$ ) regime. On the other hand, Mg II shows the flattest equivalent width distribution with a prominent tail extending to the ultra-strong regime ( $W > 1 \text{ \AA}$ ). Strong Mg II absorbers trace the evolution of the cosmic star formation rate density (e.g. Steidel & Sargent 1992; Matejek & Simcoe 2012; Zhu & Ménard 2013; Chen et al. 2017; Mathes, Churchill & Murphy 2017; Zou et al. 2021) and are often associated with star-forming galaxies (e.g. Zibetti et al. 2007; Noterdaeme, Srianand & Mohan 2010; Ménard et al. 2011). They have been shown to trace extended galaxy discs (e.g. Zabl et al. 2019), outflows (e.g. Nestor et al. 2011; Nielsen, Churchill & Kacprzak 2013; Rubin et al. 2014; Lan & Mo 2018; Schroetter et al. 2019), and galaxy group environments (e.g. Rubin et al. 2010; Gauthier 2013; Nielsen et al. 2018; Dutta et al. 2020; Lee, Hwang & Song 2021; Guha et al. 2022). We convert  $W$  to  $N(\text{H I})$  using the empirical scaling relation published in Lan & Fukugita (2017) and find that 56/360 of the Mg II absorbers could potentially be associated with Damped Lyman Alpha absorption. However, the scaling relation is constrained using measurements over  $0 \lesssim z \lesssim 4.5$  and therefore the extrapolation to higher redshifts is uncertain.

The high equivalent width Mg II absorbers do not dominate the high end of the  $\log N$  distribution (lower right) due to the large oscillator strength of Mg II. Instead, the majority of systems with  $\log N > 14$  are C II absorbers. The C IV systems generally have column densities intermediate between those of C II and Mg II, and the Si IV absorbers have the lowest median column density.

Fig. 7 shows plots of  $W$  as a function of  $z$  for all four ions. Black markers show measurements for individual systems, and red points

and error bars show the median and  $1\sigma$  spread in bins of  $z$ . The dark and light grey coloured regions indicate areas where the sample is  $< 50$  per cent and 50–90 per cent complete (see Section 5.2 for details). There is a large scatter in the rest-frame equivalent width at fixed redshift for all ions, and there is no evidence for any significant variation in the typical  $W$  probed at different redshifts.

#### 4.2.2 Associated transitions

We detect a significant number of additional transitions associated with the primary absorption systems. In descending order, we find 80 systems with Al II absorption, 77 with Si II, 46 with Mg I, 29 with O I (all with associated C II), 16 with Al III, 8 with Mn II, 6 with C II\*, 4 with Ca II, 4 with S II, 3 with Cr II, 2 with Si II\*, 2 with Ti II, and 2 with Zn II.

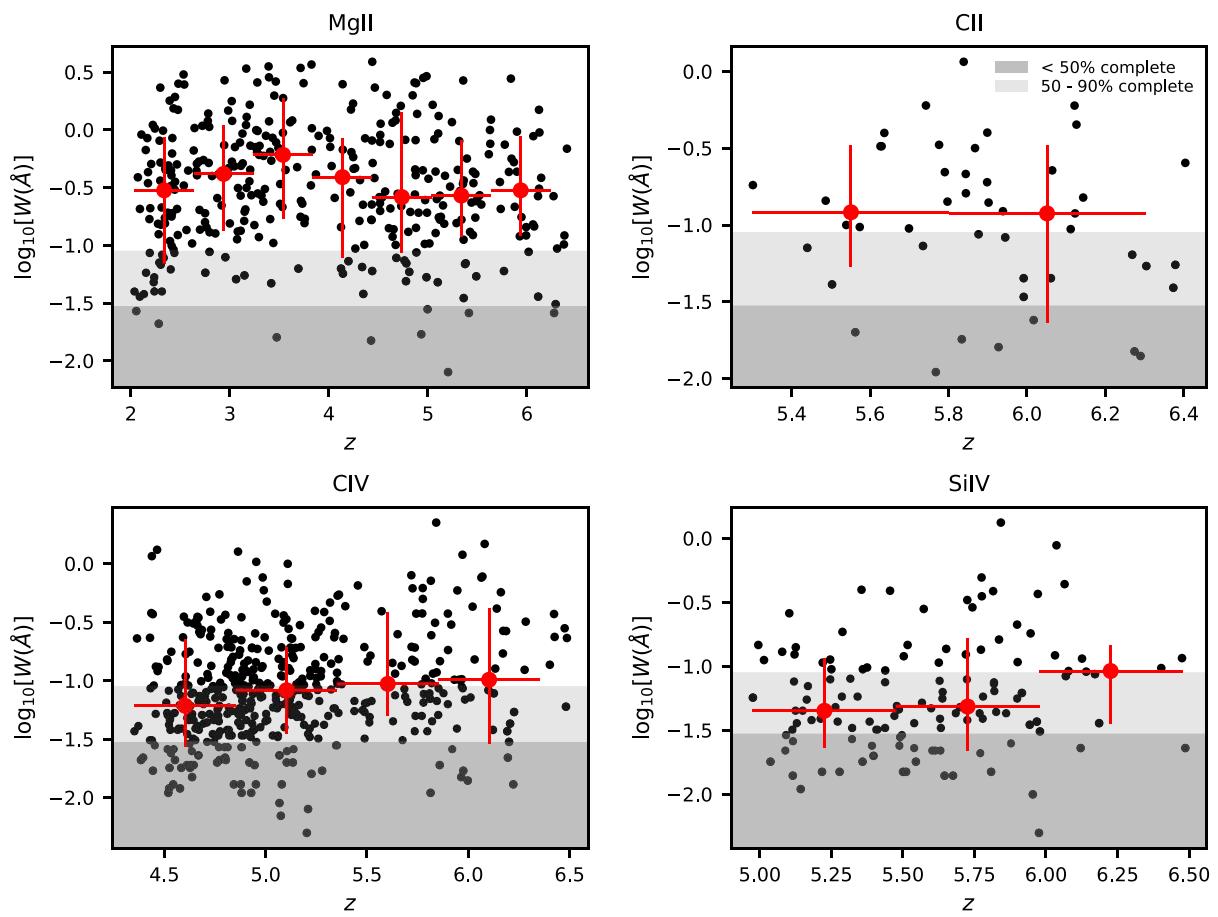
#### 4.3 Total absorption path

A key quantity for absorption line surveys is the absorption path  $\Delta X$ , which gives the redshift search interval in a comoving coordinate system. The absorption distance between  $z = 0$  and  $z = z$  is given by

$$X(z) = \frac{2}{3\Omega_m} (\Omega_m (1+z)^3 + \Omega_\Lambda)^{1/2} \quad (2)$$

(e.g. Bahcall & Peebles 1969). Throughout this paper, all mentions of absorption path refer to the total  $\Delta X$ .

The total absorption path for a given ion in a spectrum with no masked regions can be computed by simply taking the difference between the absorption distances at the maximum and minimum



**Figure 7.** Rest-frame equivalent width as a function of redshift for Mg II (top left), C II (top right), C IV (bottom left), and Si IV (bottom right) absorbers. Black markers show measurements for individual systems. Red markers and error bars indicate the median and  $1\sigma$  spread of values in redshift bins. The dark grey and light grey shaded areas indicate the regions where the sample is < 50 per cent and 50–90 per cent complete, respectively.

probed redshifts. Table 1 lists the total absorption path for each of the six primary ions, for the full E-XQR-30 sample as well as the XQR-30 quasars alone. We also list the path lengths for intervening absorbers which are relevant for studies of absorber statistics. In all cases, *regions that were masked due to bad skyline or telluric contamination were subtracted from the absorption path* (see Section 3.2). Regions dominated by BAL features are also excluded when measuring  $\Delta X$  because the completeness of narrow absorbers in these regions is poorly defined, making them difficult to incorporate in statistical measurements. The BAL regions were identified by searching for contiguous troughs with normalized flux less than 0.9 over a velocity interval of at least  $2000 \text{ km s}^{-1}$  (Bischetti et al. 2022; Bischetti et al., in preparation) and are listed in Table B4.

The sample presented in this paper represents a significant increase in the probed absorption path for C IV and Si IV absorbers at  $z \gtrsim 5$ , which are key probes of the ionizing photon background near the end of the epoch of reionization (e.g. Songaila et al. 1995; Songaila 1998; Boksenberg et al. 2003; Bolton & Viel 2011; Finlator et al. 2016; Doughty et al. 2018; Graziani et al. 2019). Prior to this work, the most extensive search for Si IV at  $z \gtrsim 4.9$  covered an absorption search path of  $\Delta X = 32.7$  (D’Odorico et al. 2022), which is a factor of  $\sim 4$  smaller than the intervening path length covered by the current sample. For C IV, our search path for intervening absorbers is a factor of  $\sim 2.6$  ( $\sim 1.4$ ) larger than any previous survey at  $z \gtrsim 5.2$  ( $4.3 \lesssim z \lesssim 5.2$ ) (see Davies et al. 2023).

The current sample also significantly increases the number of high-redshift, low-ionization absorbers measured at high S/N and spectral resolution. The largest survey to date of Mg II absorption at high redshift contains 38 quasars at  $z \gtrsim 5.8$  observed at a spectral resolution of  $\sim 50 \text{ km s}^{-1}$  (Chen et al. 2017). Similarly, the largest survey of C II absorption at high redshift comprises 47 quasars at  $z \gtrsim 5.7$  observed at  $\sim 50 \text{ km s}^{-1}$  resolution with a S/N of  $\sim 5$ –10 (Cooper et al. 2019). The largest sample of O I and C II absorbers observed at  $\sim 30 \text{ km s}^{-1}$  resolution contains 39 quasars at  $z \gtrsim 5.8$  (Becker et al. 2019). 21/39 of these quasars are part of our sample, and adding the rest of the quasars in our sample increases the absorption path for  $z \sim 6$  O I and C II absorbers by  $\sim 50$  per cent.

## 5 COMPLETENESS AND FALSE-POSITIVE RATE

In order to recover the intrinsic statistical properties of absorber populations it is necessary to quantify the completeness (the fraction of real absorption systems that are recovered) and false-positive rate (the fraction of recovered systems that are spurious) in the sample. These quantities depend on both the data quality and analysis method and are impacted to different degrees by the automated system finding and filtering, visual inspection, and Voigt profile fitting. We produce a set of mock spectra for which the absorber properties are known *a priori* (described in Section 5.1), and process these

spectra using the same steps applied to the observed data (outlined in Section 3) to characterize the completeness as a function of equivalent width, redshift, column density, and  $b$  parameter for different ions (Section 5.2), as well as the overall false-positive rate in the sample (Section 5.3).

## 5.1 Mock spectra

Each mock spectrum is constructed by taking the continuum fit for one of the observed spectra, perturbing it by the measured error spectrum, and then inserting Mg II, Fe II, C II, C IV, Si IV, and N V absorption systems with a range of redshifts, column densities, and  $b$  parameters. For C II systems we also insert O I, Si II, Al II, Mg II, and Fe II lines, with column density ratios determined by fitting scaling relations to existing measurements of low-ionization  $z > 5$  absorbers from Cooper et al. (2019).

This method does not account for systematic errors originating from skyline or telluric subtraction. However, the wavelength regions with the worst systematics in the observed spectra are masked when searching for absorption systems (see Section 3.2) and are not included in the survey absorption path (see Section 4.3), so we do not characterize the completeness in these regions. This approach also assumes that the error on the continuum fit is negligible, which is a reasonable assumption for non-BAL quasars but breaks down in wavelength regions with prominent BAL features. Therefore, the completeness statistics presented in this section only apply to absorbers falling in wavelength regions that are not masked and are not coincident with BAL features.

The number of systems inserted for each ion is set so that the automatic line finder identifies approximately the same number of candidate absorption systems in the mock spectra as in the observed spectra. This is important because the number of absorption lines per unit wavelength directly influences the probability of blending and the chance alignment of unassociated transitions. The inserted systems are uniformly distributed in redshift, from the minimum redshift where each ion is detectable to 5000 km s $^{-1}$  above the quasar redshift.

The  $b$  parameter of each inserted system is randomly selected from three possible values that approximately represent the 16th, 50th, and 84th percentile of the values measured from the Voigt profile fitting. For the high-ionization transitions these values are 13 km s $^{-1}$ , 24 km s $^{-1}$ , and 47 km s $^{-1}$ , while for the low-ionization transitions they are 7 km s $^{-1}$ , 18 km s $^{-1}$ , and 32 km s $^{-1}$ . The log  $N$  values are distributed uniformly over intervals chosen to span the full completeness range from 0 per cent to 100 per cent for each ion.

To ensure robust statistics we produce 20 mock spectra per quasar which all have the same noise statistics but differing populations of inserted absorbers (resulting in a total of 840 mock spectra). All 840 mock spectra were run through the automatic line finding and filtering procedures described in Sections 3.2 and 3.3 to produce preliminary catalogues and calculate the completeness and false-positive rate as a function of absorber properties. Visual inspection and Voigt profile fitting were performed on 1/20 spectra for each quasar to characterize the impact of these steps on the final completeness and false-positive rate (see Section 5.3).

We consider an inserted system to be successfully recovered if the mock catalogue contains a system of the same ion within 50 km s $^{-1}$  of the inserted redshift. The 50 km s $^{-1}$  window was chosen by manually matching inserted and recovered systems for a single mock spectrum and calculating the velocity offsets between the inserted and recovered redshifts. The 50 km s $^{-1}$  window is large enough to correctly match the majority of recovered systems to

the corresponding inserted systems, but sufficiently small to avoid spurious matches. The largest velocity offsets occur when multiple strongly blended absorption components are recovered as a single system, with a redshift that may be significantly different from the central redshifts of the individual absorption components.

## 5.2 Completeness statistics

The sample completeness for a given ion varies most strongly with equivalent width and column density but is also dependent on the  $b$  parameter (because a narrower line will be detected at higher peak S/N than a broader line at fixed log  $N$ ) and the redshift (because the noise level is strongly wavelength-dependent due to the presence of significant skyline contamination in some wavelength regions). In the following sections we explore how the completeness depends on each of these parameters.

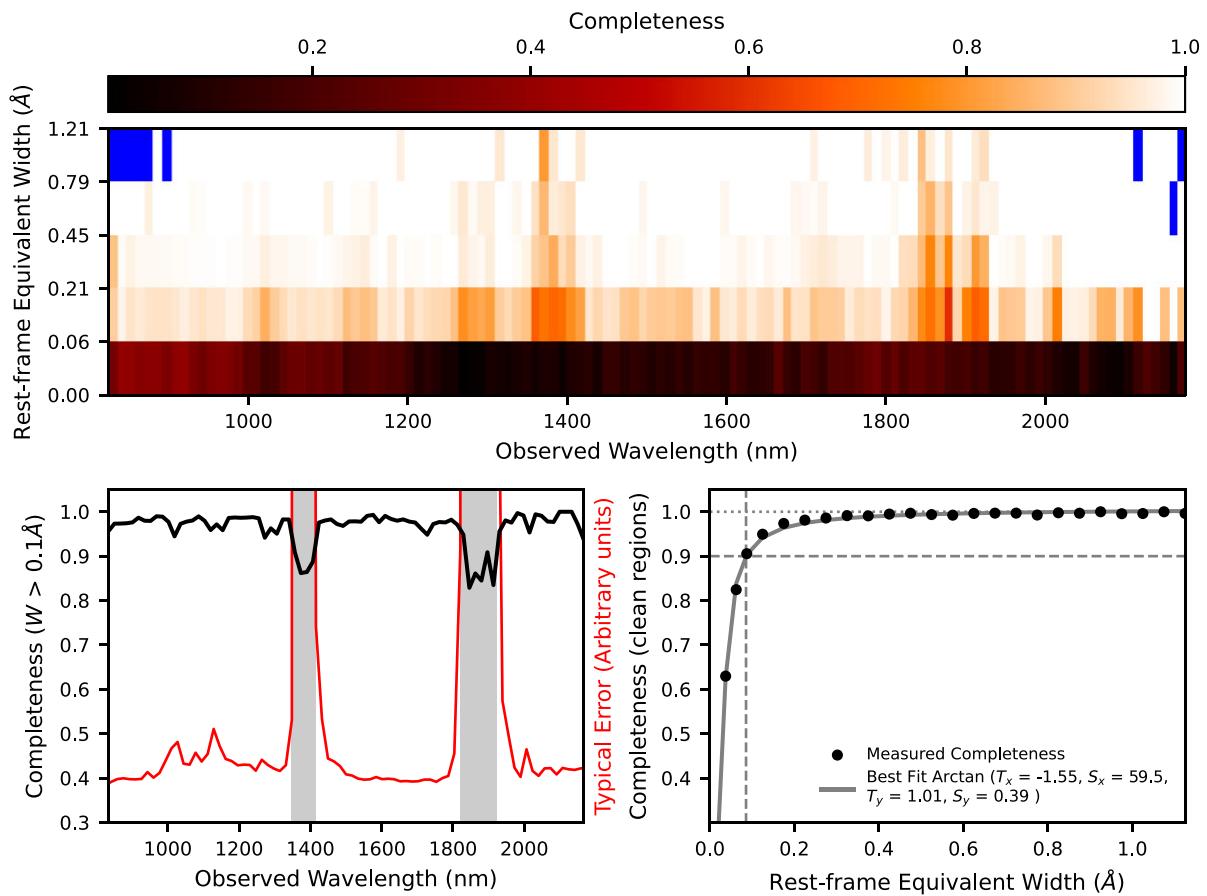
### 5.2.1 Completeness as a function of equivalent width and redshift

Fig. 8 shows how the completeness varies as a function of rest-frame equivalent width and observed wavelength for all ions combined. The colour scale reflects the completeness in each region of 2D parameter space, with blue patches indicating regions that do not contain sufficient data to make meaningful measurements. The completeness is generally very high ( $\gtrsim 90$  per cent) for systems with  $W \gtrsim 0.2$  Å, with the exception of two wavelength regions around  $\sim 1400$  nm and  $\sim 1900$  nm which correspond to regions of elevated sky contamination between the  $J$ ,  $H$  and  $K$  observing bands. We reiterate that the completeness in these regions is measured using only about half of the E-XQR-30 quasars because the remainder show strong skyline/telluric residuals which were masked (see Table B2). *If we had not masked the most contaminated wavelength ranges, the completeness in the regions between the observing bands would be significantly lower (even for the relatively strong absorbers considered here) while the total absorption search path would be correspondingly larger.* The completeness also decreases at lower equivalent widths.

To more accurately measure the extents of the wavelength regions that are most strongly impacted by low completeness, we compute the completeness as a function of wavelength using all absorbers with  $W > 0.1$  Å, as shown in the bottom left-hand panel of Fig. 8. We consider absorbers in this equivalent width range because their overall completeness is high, meaning that any dips in completeness due to elevated errors can be robustly quantified. We identified regions of low completeness by selecting wavelength channels where the average completeness is less than 90 per cent, and then expanding these regions to include any neighbouring channels with completeness below 95 per cent. The low-completeness regions, shown as grey bands, cover 1348–1416 nm and 1821–1922 nm.

The presence of low-completeness wavelength bands translates directly to reductions in the completeness of individual ions over the corresponding redshift intervals. The contaminated bands lie sufficiently far into the near-infrared that none of C II, C IV, or Si IV are shifted to those wavelengths within the redshift range probed by our sample. However, the low-completeness bands translate to dips in the Mg II completeness of our data set over  $3.81 \leq z \leq 4.05$  and  $5.50 \leq z \leq 5.86$ . When quantifying the completeness statistics for Mg II in the following section we exclude these contaminated regions of redshift space.

Finally, we examine the completeness as a function of equivalent width (excluding the contaminated wavelength bands), as shown in the bottom right-hand panel of Fig. 8. The completeness rises



**Figure 8.** Top: Completeness as a function of rest-frame equivalent width ( $W$ ) and observed wavelength for all ions combined. We emphasize that these statistics are computed excluding the masked spectral regions listed in Table B2. The spectral range is divided into 120 bins each with a width of  $\sim 11$  Å. The equivalent width bins are defined such that the square roots of the bin boundaries are equally spaced. Bins that have fewer than three systems are shown in blue. The completeness is a strong function of equivalent width and also dips at around 1400 nm and 1900 nm in the gaps between the  $J$ ,  $H$ , and  $K$  observing bands where sky noise is elevated. Bottom left: We measure the extents of the contaminated wavelength bands by plotting the completeness (black) and typical flux error (red) as a function of wavelength for systems with  $W > 0.1$  Å. These systems are chosen because their overall completeness is high, meaning that any dips in completeness can be robustly identified. The low completeness regions are shown as grey bands and cover 1348–1416 nm and 1821–1922 nm. Bottom right: Completeness as a function of  $W$  for clean wavelength regions (excluding the grey bands in the bottom left-hand panel). The dependence is fit using an arctan function (equation 3) with best-fitting parameters given in the legend. The completeness exceeds 90 per cent (50 per cent) for  $W \gtrsim 0.09$  Å (0.03 Å).

steeply at  $W \lesssim 0.1$  Å before plateauing at  $W \gtrsim 0.2$  Å. The variation in completeness as a function of wavelength is well fit by an arctan function, parametrized as follows:

$$\text{Completeness}(W) = S_y (\arctan(S_x W + T_x) + T_y) \quad (3)$$

where  $T_{x,y}$  and  $S_{x,y}$  represent the translation and scaling of each axis, respectively. The best fit is shown in grey and the parameters are listed in the figure legend. Based on this fit, the completeness exceeds 90 per cent (50 per cent) for  $W \gtrsim 0.09$  Å (0.03 Å).

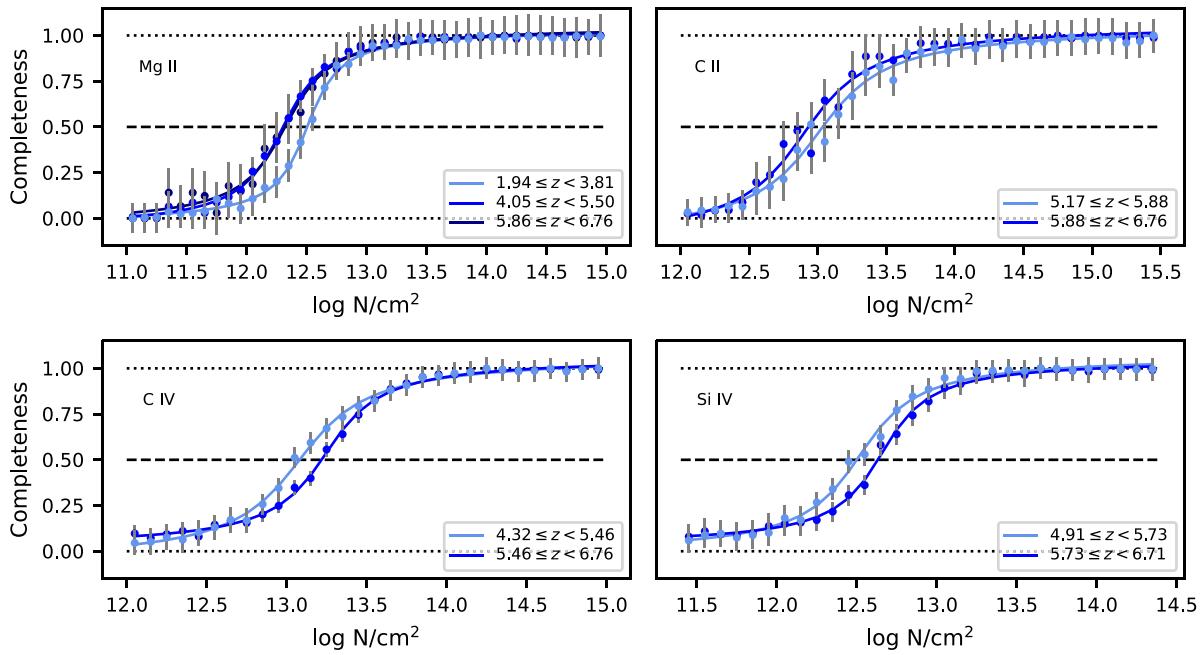
### 5.2.2 Completeness as a function of column density and $b$ parameter

Next, we compute the completeness as a function of column density for Mg II, C II, C IV, and Si IV. The average error on the observed flux varies as a function of wavelength, so we calculate the completeness in redshift bins. For C II, C IV, and Si IV we use two redshift bins divided at the path-length-weighted mean redshift of the survey. For Mg II, we exclude the redshift ranges  $3.81 \leq z \leq 4.05$  and  $5.50 \leq z \leq 5.86$  where sky contamination was found to be most significant.

The results are shown in Fig. 9. In all cases, the completeness asymptotes towards 0 at low  $\log N$  and 1 at high  $\log N$ , with an almost linear increase in the transition region. Therefore, we again characterize the completeness variations by fitting arctan functions which are plotted in the same colour as the completeness measurements for each redshift bin. The best-fitting parameters for the arctan functions are listed in Table 2.

We find that the sub-sample of Mg II absorbers at  $z \geq 4.05$  is slightly more complete at intermediate  $\log N$  than the subsample at  $z \leq 3.81$ . This may be because the lower redshift systems fall at shorter observed wavelengths where they are more likely to be blended with other absorption lines, making them more difficult to recover. For C II we do not find any significant difference in completeness between the two redshift bins. For C IV and Si IV the higher redshift bins are less complete at intermediate  $\log N$ , likely due to the increase in the average flux error towards longer wavelengths.

The inserted systems for each ion have one of three  $b$  parameter values, so we can also fit the completeness as a function of  $\log N$  for each  $b$  parameter as shown in Fig. 10. As before, the best-fitting parameters are listed in Table 2. For all ions we find that completeness at intermediate  $\log N$  decreases with increasing  $b$  parameter, which

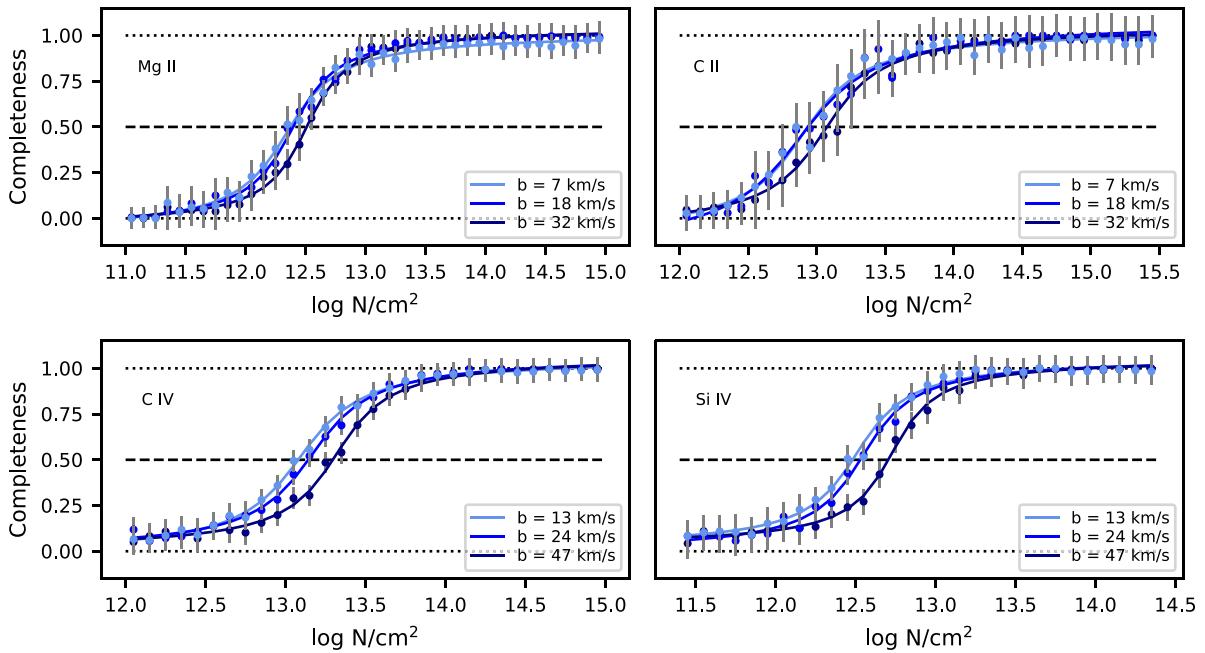


**Figure 9.** Completeness as a function of column density in different redshift bins for Mg II, C II, C IV, and Si IV absorbers. For Mg II, the completeness is computed in three redshift bins and we do not report values for the redshift ranges impacted by strong sky contamination ( $3.81 \leq z \leq 4.05$  and  $5.50 \leq z \leq 5.86$ ). For C II, C IV and Si IV the completeness is computed in two redshift bins divided at the path-length-weighted mean redshift of the survey. We quantify the variation in completeness as a function of  $\log N$  by fitting arctan functions which are over-plotted in the same colour as the completeness measurements for each redshift bin. The parameters of the arctan curves are listed in Table 2.

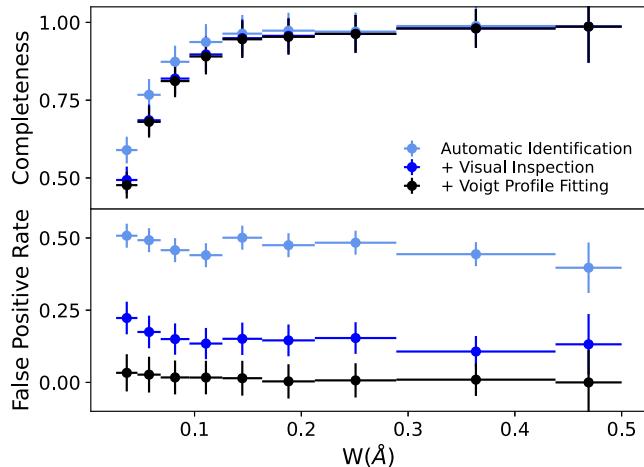
**Table 2.** Parameters of the arctan functions (equation 3) fit to the completeness as a function of  $\log N$  for Mg II, C II, C IV, and Si IV.

subset	$T_x$	$S_x$	$T_y$	$S_y$	50 per cent completeness limit $\left(\log_{10} \left( \frac{N}{\text{cm}^{-2}} \right) \right)$
Mg II					
$1.94 \leq z < 3.81$	-62.8	5.02	1.48	0.34	12.50
$4.05 \leq z < 5.50$	-45.4	3.70	1.38	0.36	12.30
$5.86 \leq z < 6.76$	-43.7	3.55	1.44	0.35	12.32
$b = 7 \text{ km s}^{-1}$	-35.7	2.91	1.43	0.35	12.27
$b = 18 \text{ km s}^{-1}$	-77.4	6.34	1.54	0.33	12.21
$b = 32 \text{ km s}^{-1}$	-58.4	4.69	1.54	0.34	12.45
C II					
$5.17 \leq z < 5.88$	-35.5	2.73	1.26	0.37	13.04
$5.88 \leq z < 6.76$	-35.5	2.75	1.19	0.39	12.93
$b = 7 \text{ km s}^{-1}$	-33.5	2.62	1.06	0.40	12.84
$b = 18 \text{ km s}^{-1}$	-34.7	2.69	1.12	0.40	12.90
$b = 32 \text{ km s}^{-1}$	-51.5	3.95	1.45	0.34	13.05
C IV					
$4.32 \leq z < 5.46$	-40.9	3.13	1.37	0.36	13.08
$5.46 \leq z < 6.76$	-52.0	3.93	1.61	0.33	13.22
$b = 13 \text{ km s}^{-1}$	-53.4	4.06	1.61	0.33	13.14
$b = 24 \text{ km s}^{-1}$	-52.6	3.98	1.58	0.34	13.17
$b = 47 \text{ km s}^{-1}$	-60.0	4.49	1.66	0.33	13.34
Si IV					
$4.91 \leq z < 5.73$	-46.4	3.70	1.50	0.35	12.50
$5.73 \leq z < 6.71$	-56.7	4.48	1.64	0.33	12.63
$b = 13 \text{ km s}^{-1}$	-67.8	5.39	1.72	0.32	12.56
$b = 24 \text{ km s}^{-1}$	-53.4	4.24	1.53	0.34	12.57
$b = 47 \text{ km s}^{-1}$	-65.3	5.10	1.72	0.32	12.77

We fit the completeness in bins of redshift and for each of the individual  $b$  parameter values that were selected from when inserting mock systems (see Section 5.1).



**Figure 10.** Same as Fig. 9 but separated by  $b$  parameter. For each ion, the three  $b$  parameter values are those that were selected from when inserting absorption systems into the mock spectra (see Section 5.1).



**Figure 11.** Completeness and false positive rate as a function of rest-frame equivalent width at different stages of the data processing: after automatic system identification and filtering (light blue), after visual inspection (blue), and after Voigt profile fitting (black). The visual inspection and fitting remove  $< 10$  per cent of real systems but  $\gtrsim 90$  per cent of spurious candidates, reducing the false positive rate from  $\sim 40$ – $50$  per cent to  $< 5$  per cent in the final catalogue.

is to be expected because the S/N per spectral channel is lower when the absorption is spread over a larger range of velocities.

We note that completeness statistics for ions not included in Table 2 can be computed using the equivalent width scaling given in the bottom right-hand panel of Fig. 8.

### 5.3 False-positive rate and impact of visual inspection

The completeness statistics presented in the previous section were computed using the automatically produced preliminary catalogues for the full set of 840 mock spectra. However, it is possible that the

visual inspection and Voigt profile fitting applied to the observed spectra could modify the completeness. To investigate this, we use the subset of 42 mock spectra that were fully processed to calculate the completeness as a function of equivalent width at three different stages of the data processing: after the automatic processing, after the visual inspection, and after the Voigt profile fitting. The results are shown in the top panel of Fig. 11. For equivalent widths below  $\sim 0.1$   $\text{\AA}$  the visual inspection reduces the completeness by  $\lesssim 10$  per cent, while for stronger systems the impact is negligible. The Voigt profile fitting does not significantly impact the completeness at any equivalent width.

In contrast, the bottom panel of Fig. 11 shows that the visual inspection and Voigt profile fitting play crucial roles in reducing the false positive rate in the final sample. The visual inspection reduces the false-positive rate from  $\sim 40$ – $50$  per cent to  $\sim 20$ – $30$  per cent at all equivalent widths whilst removing very few real systems, therefore filtering the catalogue very effectively. The remaining false positives originate from cases where a single absorption line has multiple associated candidates transitions. These are removed in the Voigt profile fitting phase, resulting in a final false positive rate of  $< 5$  per cent at all equivalent widths. Contamination from false positives can therefore be treated as negligible in the final catalogue.

### 5.4 Cross-check with previously published catalogues

Many of the absorption systems reported in this paper are published here for the first time. However, absorption line catalogues have previously been published for the archival spectra as well as for some of the XQR-30 quasars using lower quality spectra. We examine how the completeness of our sample compares to previous studies by cross-checking our catalogue with absorber lists previously published for the same quasars.

Becker et al. (2019) catalogued low-ionization absorbers in 21 of the quasars in our sample. They report 11 O I absorbers, all of which appear in our catalogue associated with automatically identified C II

systems. The improved spectra used in this analysis enable us to identify one additional O I absorber in the set of overlapping spectra.

Cooper et al. (2019) searched for C II absorbers in 11 of the quasars in our sample. They report robust detections of C II absorption in five systems, all of which are recovered by our automated line finding procedure and appear in the final catalogue. We also obtain  $> 3\sigma$  detections of C II absorption in four systems for which Cooper et al. (2019) reported upper limits, as well as 11 additional systems not reported in Cooper et al. (2019).

Chen et al. (2017) catalogued Mg II absorbers in 19 of the quasars in our sample. They find 70 Mg II absorbers, of which we recover 66. One missed system coincides with a region of elevated noise around the Mg II  $\lambda 2796$  line, while the other three systems were found to be better explained by transitions of other ions at different redshifts. Our comprehensive, simultaneous search for low and high-ionization systems allows us to more robustly identify the origin(s) of individual absorption lines. We identify 95 additional Mg II systems in the quasars surveyed by Chen et al. (2017).

Catalogues of C IV and Si IV absorption have been published for 12/42 of the E-XQR-30 quasars (D’Odorico et al. 2013, 2022; Codoreanu et al. 2018; Meyer et al. 2019). Where multiple catalogues have been published for a given spectrum we compare with the most recent one. The existing catalogues contain a combined total of 148 C IV absorbers and Si IV absorbers, of which we recover 129 (87 per cent). Of the unrecovered systems, 10/19 (53 per cent) are not found in the automatic line finding stage. 8/10 of the unidentified doublets have column densities at least 0.15 dex below the completeness limit of our data set, while the remaining 2/10 lie in regions with significant skyline residuals. Most of these systems were previously identified by eye. 4/19 (21 per cent) of the unrecovered systems were found to be better explained by transitions of other ions at different redshifts. Finally, 3/19 (16 per cent) of the unrecovered systems were rejected by the automatic filtering algorithm due to blending with other much stronger absorption systems, and 2/19 (11 per cent) were rejected by the checkers due to low S/N. Systems that were missed in the initial catalogue generation but were plausibly present at low S/N in our data were manually added, with catalogue flags used to indicate that they are not part of the automatically identified primary sample (see Section 6.2). Our catalogue adds approximately 50 C IV systems and 10 Si IV systems to those previously reported for the same quasars.

## 6 DESCRIPTION OF RELEASED PRODUCTS

This paper is accompanied by a GitHub repository (<https://github.com/XQR-30/Metal-catalogue>) which contains the metal absorber catalogues and all of the data required to use these catalogues for scientific studies. Specifically, the repository includes:

- (i) A metal absorber catalogue for each quasar, as well as a full catalogue for the entire sample (provided in CSV format).
- (ii) Plots of the Voigt profile fits for individual absorption systems (provided in PDF format).
- (iii) Machine-readable versions of Tables B1, B2 and B4 (provided in CSV or JSON format).
- (iv) A PYTHON tool to calculate the absorption path over a given redshift range for any transition or primary ion and any value of  $\Omega_m$ .
- (v) Quasar continuum fits (provided in PDF format).

A condensed excerpt of the absorber catalogue for J0408-5632 is shown in Table B3 to illustrate the contents of the data columns. The plots of the Voigt profile fits for the absorption systems listed in Table B3 are shown in Fig. B1.

## 6.1 Absorber properties

Columns 1–13 of Table B3 give the IDs and physical properties of the detected absorption systems. In the full catalogue, there is an additional entry at the start of each row (labelled Column 0) indicating which quasar the relevant absorption system is associated with. We illustrate the format of the table entries using System 9 in J0408-5632 as an example. Fig. B1 shows that this system is detected in Mg II, Fe II, Si II, Al II, and C IV. The low-ionization species all share the same central two-component kinematic structure, although only the blue-most component is fit to Fe II due to the low S/N (the detection is listed as an upper limit). Mg II also shows a weaker redshifted component not seen in the other ions. C IV is fit completely independently of the low-ionization species with four Voigt components.

The first row of Table B3 lists the overall properties of the Mg II absorption: the optical-depth-weighted mean absorption redshift, the rest-frame equivalent width of the strongest transition (in this case Mg II  $\lambda 2796$ ), the total  $\log N$ , calculated by summing the column densities of the individual Voigt components, and  $\Delta v_{90}$ . The  $b$  parameter is not listed because it can only be measured for individual components. The next row lists the equivalent width of Mg II  $\lambda 2803$ . If further Mg II transitions were detected, their equivalent widths would be listed in subsequent rows. Similar entries for Fe II, Si II, Al II, and C IV are found further down the table.

The next three rows in the table list the column density,  $\Delta v_{90}$ , and  $b$  parameter of the individual Voigt profile components which are labelled 9a to 9c. In these IDs, the ‘9’ denotes the system number and the letters are the component labels. These labels make it possible to link components with the same kinematics across different ions. If we examine the properties of component 9a for Mg II, Fe II, Si II, and Al II, we see they all have the same  $b$  parameter and redshift, indicating that their kinematics were tied during fitting. It was not always possible to fit all high-ionization and all low-ionization ions using the same kinematic structure (due to e.g. saturation, low S/N, or intrinsic differences between the velocity profiles; see Section 3.6). In these cases, the component labels (letters) would differ between ions. We note that C IV, which was fit independently from the low-ionization species, has a different set of components labelled 9d to 9g.

If the optical depth measured from the weakest transition of a given ion is  $\tau > 2$ , then we classify the line as saturated and flag the column density measurement with ‘S’ (saturated) in Column 10. If no transitions of a given ion are detected at  $\geq 3\sigma$ , then the reported equivalent width values are  $3\sigma$  upper limits. The  $W$  upper limits are converted to  $\log N$  values assuming that the absorption falls on the linear part of the curve of growth, and the minimum of the  $\log N$  values for the individual transitions is reported as the overall upper limit for that ion. In such cases, the  $W$  and  $\log N$  values are flagged with ‘U’ (upper limit) in Columns 7 and 10, respectively.

## 6.2 Sample flags

The remaining three columns of the table provide various flags to assist users in determining which subset of absorbers should be used for a given science case. The completeness and false positive rate statistics presented in Sections 5.2 and 5.3 only apply to systems that are recovered by the automatic line finder, pass the visual inspection, and do not lie in masked wavelength regions or BAL regions. Objects that do not pass the visual inspection have a much higher likelihood of being false positives (see Fig. 11). The completeness in the masked wavelength regions is low due to significant sky contamination (see Fig. 8). There is no straightforward way to quantify the impact

of BAL absorption on our ability to recover underlying narrow absorption. For these reasons, Column 14 gives four flags indicating, from left to right, whether each ion in each absorption system (1) was recovered by the automatic line finder, (2) passed the visual inspection (3) does not lie in a masked wavelength region, and (4) does not lie in a BAL region.

In addition, the completeness is only well defined for the ‘primary’ ions that we search for in the initial candidate finding stage: Mg II, Fe II, C II, C IV, Si IV, and N V, as well as O I and Si II  $\lambda 1260$  because they are required for the acceptance of C II systems when they lie within the probed wavelength region (see Section 3.2). Absorption in other ions can only be recovered at redshifts where absorption is already detected in one or more of the primary ions, making the completeness poorly defined.

With this in mind we define a ‘primary’ sample consisting only of primary ions in systems that meet all of the outlined criteria (all flags in Column 14 set to ‘Y’). The completeness and false positive rate for this sample are well understood, and only objects in this primary sample should be used for statistical studies requiring accurate completeness estimates. Column 15 indicates whether or not a given ion in a given system is part of the primary sample. The final column in the table indicates whether each absorption system is proximate (P) or intervening (I).

### 6.3 Example usage: redshift evolution of metal ions

The metal absorber catalogue presented in this work significantly expands on existing samples of metal absorbers at  $z \gtrsim 5$  (see Section 5.4) and is therefore well suited to study the evolution of metal absorber properties over  $5 \lesssim z \lesssim 6$ . The evolution of C IV absorber properties over  $4.3 \lesssim z \lesssim 6.3$  is presented in an accompanying paper by Davies et al. (2023), and the redshift evolution of low-ionization absorbers will be examined in a follow-up paper by Sebastian et al. (in preparation).

To illustrate how the catalogue can be used to quantify the evolution in metal absorbers properties, we outline the basic procedure to measure the cosmic mass density of intervening C IV absorbers ( $\Omega_{\text{CIV}}$ ) over an arbitrary redshift interval  $z_1$ – $z_2$ :

- (i) Extract  $\log N$  values (Column 8) for all rows that have:
  - (a) A valid Component ID in Column 2 (not ‘-’)
  - (b)  $z$  in Column 3 within the redshift interval of interest; i.e.  $z_1 \leq z \leq z_2$
  - (c) ‘C IV’ listed under Ion/Species in Column 4
  - (d) ‘Y’ listed under Primary Sample in Column 15
  - (e) ‘T’ listed under Proximate or Intervening in Column 16
- (ii) Remove components for which  $\log N$  is less than the 50 per cent completeness limit listed in Table 2.
- (iii) Group the remaining components into bins of  $\log N$  with a width of  $\text{dlog } N = 0.1$ .
- (iv) For each bin, calculate the total  $\log N$  of all components and correct for completeness using equation (3) and the parameters listed in Table 2. If the redshift interval of interest is large enough that it spans multiple redshift bins for the completeness correction, this step can be performed for each redshift bin individually.
- (v) Sum the total  $\log N$  from all redshift and  $\log N$  bins to obtain the overall total column density.

The final step is to determine the total absorption path over the interval  $z_1$ – $z_2$ . For each quasar, the unmasked absorption search path is given by  $\Delta X = X(z_{\text{upper}}) - X(z_{\text{lower}})$ , where  $X(z)$  is defined in equation (2).  $z_{\text{lower}}$  is the larger of  $z_1$  and the minimum redshift at which C IV  $\lambda 1548$  falls redward of the quasar Ly $\alpha$

line; i.e.  $z_{\text{lower}} = \max(z_1, \lambda_{\text{Ly}\alpha}(1 + z_{\text{em}})/\lambda_{\text{CIV} 1548} - 1)$ .  $z_{\text{upper}}$  is the smaller of  $z_2$  and  $10000 \text{ km s}^{-1}$  blueward of the quasar redshift (the maximum redshift for intervening absorbers); i.e.  $z_{\text{upper}} = \min(z_2, (1 + z_{\text{em}}) \times \exp\left(\frac{-10000 \text{ km s}^{-1}}{c}\right) - 1)$ .<sup>9</sup> Once the unmasked absorption path is calculated, any masked regions (listed in Table B2) and BAL regions (listed in Table B4) that fall within the relevant wavelength range must be subtracted to obtain the masked absorption path. Finally, the masked absorption paths of all quasars can be summed to obtain the total absorption path. The GitHub repository includes a PYTHON routine to calculate the absorption path over a given redshift range for any transition or primary ion and any value of  $\Omega_m$ .

## 7 SUMMARY

We have presented a catalogue of 778 metal absorption systems found in high quality XSHOOTER spectra of 42 bright quasars at  $5.8 \lesssim z \lesssim 6.6$ . Thirty of the quasars were observed as part of the VLT Large Program XQR-30 (PI: D’Odorico), and the remaining 12 have archival spectra with similar S/N ( $\gtrsim 10$  per  $10 \text{ km s}^{-1}$  spectral pixel at  $1285 \text{ \AA}$ ) and spectral resolution ( $\sim 30 \text{ km s}^{-1}$ ).

The absorber catalogue was constructed by performing an automated search for Mg II, Fe II, C II, C IV, Si IV, and N V absorption, automatically filtering the list of candidates using custom routines, searching for additional transitions at the redshifts of the candidate systems, visually inspecting the candidates to remove spurious systems, and fitting Voigt profiles to the absorption lines. Many of these steps make use of routines from the PYTHON library ASTROCOOK (Cupani et al. 2020).

We quantify the completeness and false-positive rate in the absorber sample by producing a set of 840 mock spectra for which the absorber properties are known a-priori, and processing them using the same steps applied to the observed data. The final false-positive rate is negligible (< 5 per cent) at all equivalent widths, and the sample is 50 per cent (90 per cent) complete for  $W(\text{rest}) \gtrsim 0.03 \text{ \AA}$  ( $0.09 \text{ \AA}$ ). We provide functional forms that can be used to calculate the completeness as a function of rest-frame equivalent width, and as a function of  $\log N$  for Mg II, C II, C IV and Si IV, in different redshift intervals and for a range of  $b$  parameters.

The final catalogue includes 479 C IV absorbers, 360 Mg II absorbers, and 46 C II absorbers. The sample covers a factor of  $\sim 3$ – $4$  larger absorption path than any previously published search for C IV and Si IV absorbers at  $z \gtrsim 5$ , and significantly expands existing samples of low-ionization absorbers observed at high S/N and spectral resolution. The number of high S/N spectra observed at comparable spectral resolution is unlikely to increase significantly in the near future because the space density of luminous quasars drops rapidly at  $z \gtrsim 6$  (e.g. Wang et al. 2019; Ren & Trenti 2021). Therefore, the catalogue presented here is a crucial legacy resource for studies of metal absorber properties during the later stages of reionization. The catalogue is also a key reference sample for evolutionary studies of enriched gas in galaxy halos. The redshifts of the absorption systems span  $2 \lesssim z \lesssim 6.5$ , enabling the self-consistent comparison of absorber properties all the way from the end of reionization to the peak epoch of star formation.

The redshift evolution of C IV absorber properties is presented in an accompanying paper by Davies et al. (2023), and the evolution of low-ionization absorber properties will be published in a follow-up

<sup>9</sup>This conversion between  $\text{dv}$  and  $\text{dz}$  was adopted for consistency with ASTROCOOK. It is derived as follows:  $\text{dv}/c \simeq \text{d}\lambda/\lambda = \text{dln}\lambda$ , which gives  $\text{dz} = (1 + z_{\text{em}})\text{d}\lambda \simeq (1 + z_{\text{em}})\exp(\text{dv}/c)$ .

paper by Sebastian et al. (in preparation). To maximize the scientific impact of this data set, we have publicly released the complete absorption line catalogue along with all of the information required to use the data for statistical analysis. This includes machine-readable versions of the published tables as well as a PYTHON tool to calculate the total absorption path for any ion/transition and redshift interval.

The unprecedent statistical power of this data set makes it possible to robustly measure the typical properties of metal absorbers near the end of reionization, opening up a myriad of possibilities for ground-breaking research. These data have the potential to facilitate valuable new insights on the formation redshifts of Population III stars, the strength and hardness of the UV background at  $z \sim 6$ , the properties of star-formation driven outflows in the early Universe, and the formation redshifts of the first galaxies. By propelling our understanding of absorber properties at the current redshift frontier, this data set will create a firm basis for future studies of the even more distant Universe with *JWST* and 30 m-class telescopes.

## ACKNOWLEDGEMENTS

We thank the referee for their valuable suggestions which improved the clarity of this paper. RLD acknowledges the support of a Gruber Foundation Fellowship research support grant. SEIB and RAM acknowledge funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement no. 740246 ‘Cosmic Gas’). GDB and YZ were supported by the National Science Foundation through grant AST-1751404. ACE and FW acknowledge support by NASA through the NASA Hubble Fellowship grants #HF2-51434 and HF2-51448.001-A, awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. EPF is supported by the international Gemini Observatory, a program of NSF’s NOIRLab, which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation, on behalf of the Gemini partnership of Argentina, Brazil, Canada, Chile, the Republic of Korea, and the United States of America. This research was supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. Based on observations collected at the European Organisation for Astronomical Research in the Southern Hemisphere under ESO Programme IDs 0100.A-0625, 0101.B-0272, 0102.A-0154, 0102.A-0478, 084.A-0360(A), 084.A-0390(A), 084.A-0550(A), 085.A-0299(A), 086.A-0162(A), 086.A-0574(A), 087.A-0607(A), 088.A-0897(A), 091.C-0934(B), 096.A-0095(A), 096.A-0418(A), 097.B-1070(A), 098.B-0537, 098.B-0537(A), 1103.A-0817, 294.A-5031(B), 60.A-9024(A). This research made use of NASA’s Astrophysics Data System as well as ASTROCOOK (Cupani et al. 2020), ASTROPY (Astropy Collaboration 2018), MATPLOTLIB (Hunter 2007), NUMPY (Harris et al. 2020), and SCIPY (Virtanen et al. 2020).

## DATA AVAILABILITY

The metal absorber catalogues, data tables, quasar continuum fits and accompanying PYTHON scripts are released with this paper and can be downloaded from this GitHub repository: <https://github.com/XQR-30/Metal-catalogue>. The reduced spectra will be released with the XQR-30 survey paper (D’Odorico et al., in preparation).

## REFERENCES

Adams N. J. et al., 2023, *MNRAS*, 518, 4755  
 Astropy Collaboration, 2018, *AJ*, 156, 123  
 Atek H. et al., 2023, *MNRAS*, 519, 1201  
 Bahcall J. N., Peebles P. J. E., 1969, *ApJ*, 156, L7  
 Bañados E., Decarli R., Walter F., Venemans B. P., Farina E. P., Fan X., 2015, *ApJ*, 805, L8  
 Bañados E. et al., 2016, *ApJS*, 227, 11  
 Bañados E. et al., 2019, *ApJ*, 885, 59  
 Becker G. D., Sargent W. L. W., Rauch M., Simcoe R. A., 2006, *ApJ*, 640, 69  
 Becker G. D., Rauch M., Sargent W. L. W., 2009, *ApJ*, 698, 1010  
 Becker G. D., Sargent W. L. W., Rauch M., Carswell R. F., 2012, *ApJ*, 744, 91  
 Becker G. D., Bolton J. S., Lidz A., 2015a, *PASA*, 32, e045  
 Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015b, *MNRAS*, 447, 3402  
 Becker G. D. et al., 2019, *ApJ*, 883, 163  
 Berg T. A. M. et al., 2016, *MNRAS*, 463, 3021  
 Berg T. A. M. et al., 2021, *MNRAS*, 502, 4009  
 Bischetti M. et al., 2022, *Nature*, 605, 244  
 Boksenberg A., Sargent W. L. W., Rauch M., 2003, preprint ([arXiv:astro-ph/0307557](https://arxiv.org/abs/astro-ph/0307557))  
 Bolton J. S., Viel M., 2011, *MNRAS*, 414, 241  
 Bosman S. E. I., Becker G. D., Haehnelt M. G., Hewett P. C., McMahon R. G., Mortlock D. J., Simpson C., Venemans B. P., 2017, *MNRAS*, 470, 1919  
 Bosman S. E. I., Fan X., Jiang L., Reed S., Matsuoka Y., Becker G., Haehnelt M., 2018, *MNRAS*, 479, 1055  
 Bosman S. E. I. et al., 2022, *MNRAS*, 514, 55  
 Carilli C. L. et al., 2010, *ApJ*, 714, 834  
 Carnall A. C. et al., 2015, *MNRAS*, 451, L16  
 Chen S.-F. S. et al., 2017, *ApJ*, 850, 188  
 Codoreanu A., Ryan-Weber E. V., Crighton N. H. M., Becker G., Pettini M., Madau P., Venemans B., 2017, *MNRAS*, 472, 1023  
 Codoreanu A., Ryan-Weber E. V., García L. Á., Crighton N. H. M., Becker G., Pettini M., Madau P., Venemans B., 2018, *MNRAS*, 481, 4940  
 Cooke R., Pettini M., Steidel C. C., Rudie G. C., Nissen P. E., 2011, *MNRAS*, 417, 1534  
 Cooper T. J., Simcoe R. A., Cooksey K. L., Bordoloi R., Miller D. R., Furesz G., Turner M. L., Bañados E., 2019, *ApJ*, 882, 77  
 Cupani G., D’Odorico V., Cristiani S., Russo S. A., Calderone G., Taffoni G., 2020, in Guzman J. C., Ibsen J., eds, Proc. SPIE Conf. Ser. Vol. 11452, Software and Cyberinfrastructure for Astronomy VI. SPIE, Bellingham, p. 114521U  
 D’Odorico V. et al., 2013, *MNRAS*, 435, 1198  
 D’Odorico V. et al., 2018, *ApJ*, 863, L29  
 D’Odorico V. et al., 2022, *MNRAS*, 512, 2389  
 Davies F. B. et al., 2018, *ApJ*, 864, 142  
 Davies R. L. et al., 2023, *MNRAS*, 521, 314  
 Decarli R. et al., 2018, *ApJ*, 854, 97  
 Díaz C. G., Ryan-Weber E. V., Codoreanu A., Pettini M., Madau P., 2016, Bol. Asociacion Argentina Astron. La Plata Argentina, 58, 54  
 Doughty C., Finlator K., Oppenheimer B. D., Davé R., Zackrisson E., 2018, *MNRAS*, 475, 4717  
 Dutta R. et al., 2020, *MNRAS*, 499, 5022  
 Eilers A.-C., Davies F. B., Hennawi J. F., 2018, *ApJ*, 864, 53  
 Eilers A.-C. et al., 2021, *ApJ*, 914, 74  
 Fan X. et al., 2006, *AJ*, 132, 117  
 Finlator K., Thompson R., Huang S., Davé R., Zackrisson E., Oppenheimer B. D., 2015, *MNRAS*, 447, 2526  
 Finlator K., Oppenheimer B. D., Davé R., Zackrisson E., Thompson R., Huang S., 2016, *MNRAS*, 459, 2299  
 Finlator K., Doughty C., Cai Z., Díaz G., 2020, *MNRAS*, 493, 3223  
 Foltz C. B., Weymann R. J., Peterson B. M., Sun L., Malkan M. A., Chaffee Frederic H. J., 1986, *ApJ*, 307, 504  
 Fox A. J. et al., 2015, *ApJ*, 799, L7

Furlanetto S. R., Hernquist L., Zaldarriaga M., 2004, *MNRAS*, 354, 695

Furtak L. J., Shuntov M., Atek H., Zitrin A., Richard J., Lehnert M. D., Chevallard J., 2022, preprint ([arXiv:2208.05473](https://arxiv.org/abs/2208.05473))

Gallerani S., Ferrara A., Fan X., Choudhury T. R., 2008, *MNRAS*, 386, 359

García L. A., Tescari E., Ryan-Weber E. V., Wyithe J. S. B., 2017, *MNRAS*, 470, 2494

Gauthier J.-R., 2013, *MNRAS*, 432, 1444

Gnedin N. Y., Becker G. D., Fan X., 2017, *ApJ*, 841, 26

Götberg Y., de Mink S. E., McQuinn M., Zapartas E., Groh J. H., Norman C., 2020, *A&A*, 634, A134

Graziani L., Maselli A., Maio U., 2019, *MNRAS*, 482, L112

Greif T. H., Johnson J. L., Bromm V., Klessen R. S., 2007, *ApJ*, 670, 1

Greig B., Mesinger A., Davies F. B., Wang F., Yang J., Hennawi J. F., 2022, *MNRAS*, 512, 5390

Guha L. K., Srianand R., Dutta R., Joshi R., Noterdaeme P., Petitjean P., 2022, *MNRAS*, 513, 3836

Harikane Y. et al., 2022, preprint ([arXiv:2208.01612](https://arxiv.org/abs/2208.01612))

Harris C. R. et al., 2020, *Nature*, 585, 357

Horne K., 1986, *PASP*, 98, 609

Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90

Jones T. A., Ellis R. S., Schenker M. A., Stark D. P., 2013, *ApJ*, 779, 52

Keating L. C., Puchwein E., Haehnelt M. G., Bird S., Bolton J. S., 2016, *MNRAS*, 461, 606

Kelson D. D., 2003, *PASP*, 115, 688

Konno A. et al., 2018, *PASJ*, 70, S16

Kurk J. D. et al., 2007, *ApJ*, 669, 32

Labbe I. et al., 2022, preprint ([arXiv:2207.12446](https://arxiv.org/abs/2207.12446))

Lan T.-W., Fukugita M., 2017, *ApJ*, 850, 156

Lan T.-W., Mo H., 2018, *ApJ*, 866, 36

Lee J. C., Hwang H. S., Song H., 2021, *MNRAS*, 503, 4309

Lehner N., O'Meara J. M., Fox A. J., Howk J. C., Prochaska J. X., Burns V., Armstrong A. A., 2014, *ApJ*, 788, 119

Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018, *ApJ*, 856, 2

Matejek M. S., Simcoe R. A., 2012, *ApJ*, 761, 112

Mathes N. L., Churchill C. W., Murphy M. T., 2017, preprint ([arXiv:1701.05624](https://arxiv.org/abs/1701.05624))

Mazzucchelli C. et al., 2017, *ApJ*, 849, 91

Ménard B., Wild V., Nestor D., Quider A., Zibetti S., Rao S., Turnshek D., 2011, *MNRAS*, 417, 801

Meyer R. A., Bosman S. E. I., Kakiuchi K., Ellis R. S., 2019, *MNRAS*, 483, 19

Muzahid S., 2014, *ApJ*, 784, 5

Naidu R. P. et al., 2022, *ApJ*, 940, L14

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

Nestor D. B., Johnson B. D., Wild V., Ménard B., Turnshek D. A., Rao S., Pettini M., 2011, *MNRAS*, 412, 1559

Nielsen N. M., Churchill C. W., Kacprzak G. G., 2013, *ApJ*, 776, 115

Nielsen N. M., Kacprzak G. G., Pointon S. K., Churchill C. W., Murphy M. T., 2018, *ApJ*, 869, 153

Noll S., Kausch W., Barden M., Jones A. M., Szyszka C., Kimeswenger S., Vinther J., 2012, *A&A*, 543, A92

Noterdaeme P., Srianand R., Mohan V., 2010, *MNRAS*, 403, 906

Oppenheimer B. D., Davé R., Finlator K., 2009, *MNRAS*, 396, 729

Perrotta S. et al., 2016, *MNRAS*, 462, 3285

Planck Collaboration, 2020, *A&A*, 641, A6

Pradeep J., Sankar S., Umarsree T. M., Narayanan A., Khaire V., Gebhardt M., Sameer Charlton J. C., 2020, *MNRAS*, 493, 250

Prochaska J. X., Wolfe A. M., 1997, *ApJ*, 487, 73

Prochaska J. X., O'Meara J. M., Fumagalli M., Bernstein R. A., Burles S. M., 2015, *ApJS*, 221, 2

Rahmati A., Schaye J., Crain R. A., Oppenheimer B. D., Schaller M., Theuns T., 2016, *MNRAS*, 459, 310

Rauch M., Sargent W. L. W., Womble D. S., Barlow T. A., 1996, *ApJ*, 467, L5

Reed S. L. et al., 2019, *MNRAS*, 487, 1874

Ren K., Trenti M., 2021, *ApJ*, 923, 110

Rubin K. H. R., Weiner B. J., Koo D. C., Martin C. L., Prochaska J. X., Coil A. L., Newman J. A., 2010, *ApJ*, 719, 1503

Rubin K. H. R., Prochaska J. X., Koo D. C., Phillips A. C., Martin C. L., Winstrom L. O., 2014, *ApJ*, 794, 156

Ryan-Weber E. V., Pettini M., Madau P., 2006, *MNRAS*, 371, L78

Ryan-Weber E. V., Pettini M., Madau P., Zych B. J., 2009, *MNRAS*, 395, 1476

Sargent W. L. W., Boksenberg A., Steidel C. C., 1988, *ApJS*, 68, 539

Schroetter I. et al., 2019, *MNRAS*, 490, 4368

Simcoe R. A., 2006, *ApJ*, 653, 977

Simcoe R. A., 2011, *ApJ*, 738, 159

Simcoe R. A., Sargent W. L. W., Rauch M., 2004, *ApJ*, 606, 92

Simcoe R. A., Onoue M., Eilers A.-C., Banados E., Cooper T. J., Furesz G., Hennawi J. F., Venemans B., 2020, preprint ([arXiv:2011.10582](https://arxiv.org/abs/2011.10582))

Songaila A., 1998, *AJ*, 115, 2184

Songaila A., 2001, *ApJ*, 561, L153

Songaila A., Cowie L. L., 2002, *AJ*, 123, 2183

Songaila A., Hu E. M., Cowie L. L., 1995, *Nature*, 375, 124

Steidel C. C., Sargent W. L. W., 1992, *ApJS*, 80, 1

Suresh J., Bird S., Vogelsberger M., Genel S., Torrey P., Sijacki D., Springel V., Hernquist L., 2015, *MNRAS*, 448, 895

Tumlinson J., Peebles M. S., Werk J. K., 2017, *ARA&A*, 55, 389

Venemans B. P. et al., 2020, *ApJ*, 904, 130

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

Virtanen P. et al., 2020, *Nat. Methods*, 17, 261

Wang R. et al., 2010, *ApJ*, 714, 699

Wang R. et al., 2013, *ApJ*, 773, 44

Wang J., Koribalski B. S., Serra P., van der Hulst T., Roychowdhury S., Kamphuis P., Chengalur J. N., 2016, *MNRAS*, 460, 2143

Wang F. et al., 2019, *ApJ*, 884, 30

Welsh L., Cooke R., Fumagalli M., Pettini M., 2022, *ApJ*, 929, 158

Wild V. et al., 2008, *MNRAS*, 388, 227

Wolfe A. M., Turnshek D. A., Lanzetta K. M., Lu L., 1993, *ApJ*, 404, 480

Yan H., Ma Z., Ling C., Cheng C., Huang J.-s., Zitrin A., 2022, preprint ([arXiv:2207.11558](https://arxiv.org/abs/2207.11558))

Yang J. et al., 2019, *ApJ*, 880, 153

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

Zabl J. et al., 2019, *MNRAS*, 485, 1961

Zhu G., Ménard B., 2013, *ApJ*, 770, 130

Zhu Y. et al., 2021, *ApJ*, 923, 223

Zhu Y. et al., 2022, *ApJ*, 932, 76

Zibetti S., Ménard B., Nestor D. B., Quider A. M., Rao S. M., Turnshek D. A., 2007, *ApJ*, 658, 161

Zou S. et al., 2021, *ApJ*, 906, 32

## SUPPORTING INFORMATION

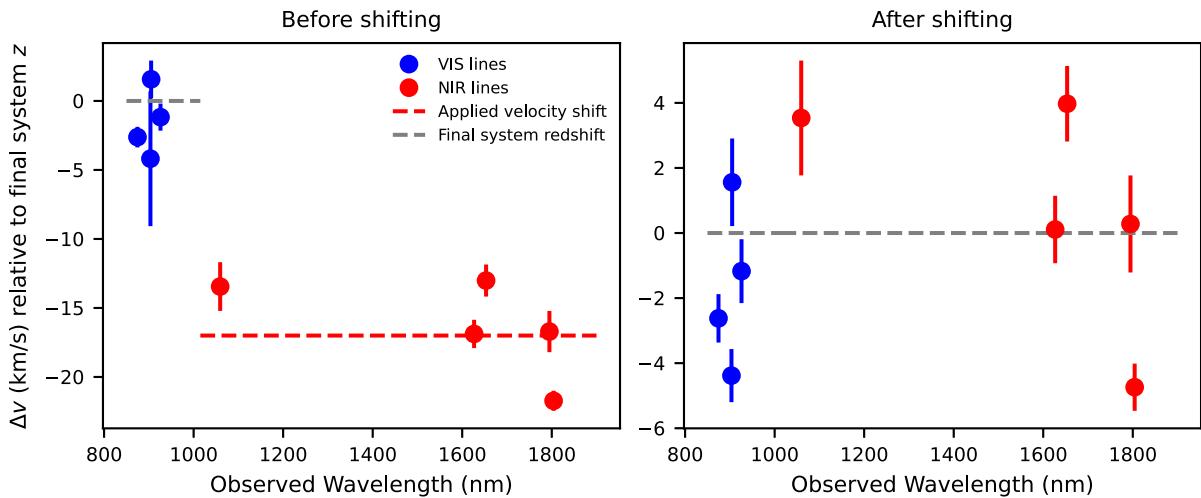
Supplementary data are available at [MNRAS](https://mnras.oxfordjournals.org) online.

**Table B3.** A condensed excerpt of the absorber catalogue for J0408-5632.

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

## APPENDIX A: VELOCITY SHIFTING

When processing the XSHOOTER spectra we found that in some cases, the redshift centroids of transitions of the same ion and absorption system falling in the VIS arm and the NIR arm (e.g.  $\text{Si II } \lambda 1260$  versus  $\text{Si II } \lambda 1526$ ) were slightly offset in velocity (5–20  $\text{km s}^{-1}$ ). This is demonstrated in the left-hand panel of Fig. A1 which compares the centroid velocities of transitions falling in the VIS (blue markers) and NIR (red markers) arms for a single-component low-ionization absorption system (the line profiles of

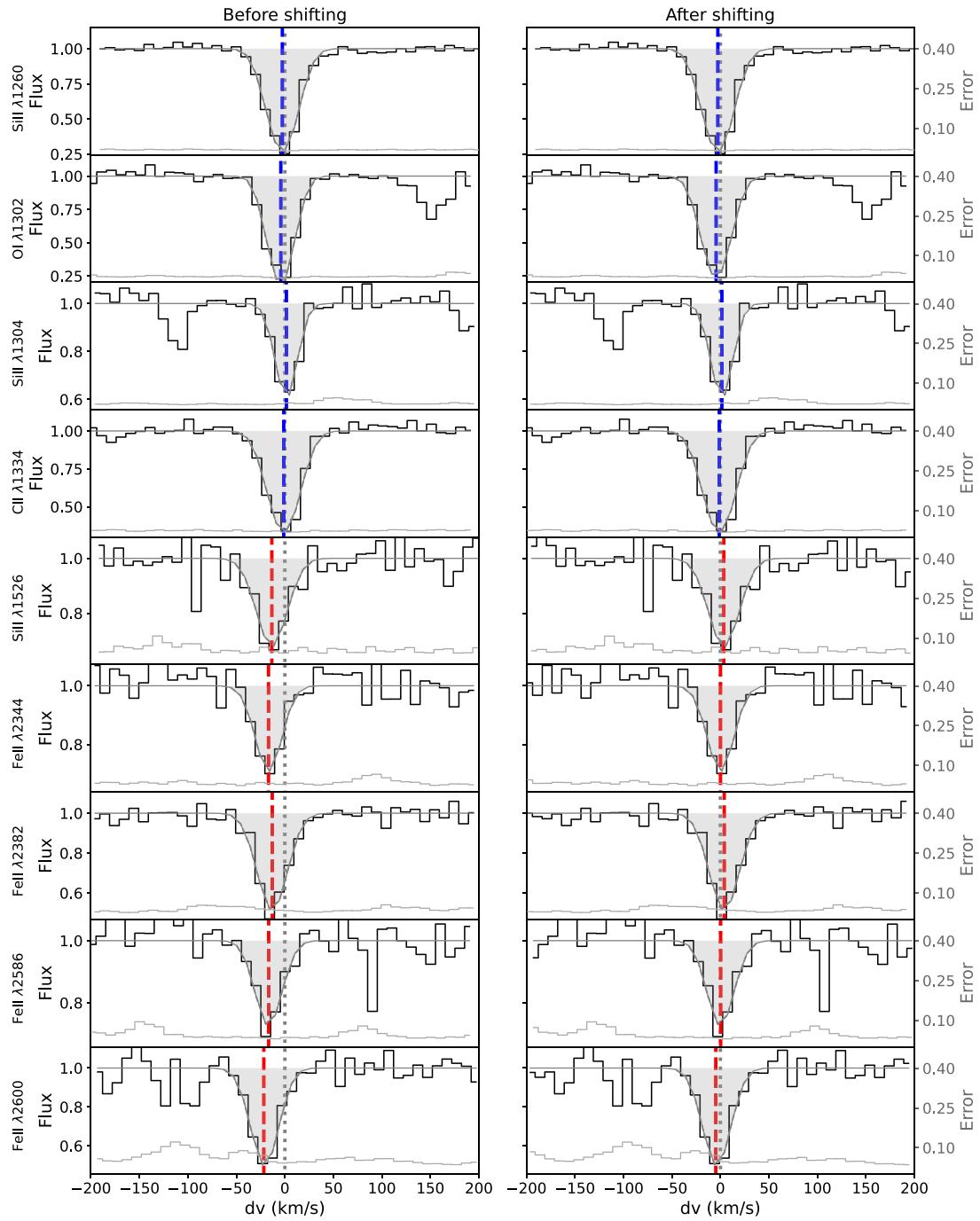


**Figure A1.** Left: Difference between the centroid velocities of VIS and NIR transitions (blue and red markers, respectively) as a function of observed wavelength for a single-component low-ionization absorption system at  $z = 5.939$  in the spectrum of SDSSJ2310+1855. The velocities are relative to the redshift of the absorption system in the final catalogue (grey dashed line). The NIR transitions are offset by a median velocity of  $17 \text{ km s}^{-1}$  (red dashed line). Right: Velocity offsets after applying the  $17 \text{ km s}^{-1}$  shift to the NIR spectrum. There is no evidence for any systematic variation in the measured centroid velocity as a function of wavelength. The final centroid velocities vary by  $\sim 10 \text{ km s}^{-1}$  which is the width of a spectral pixel.

the individual transitions are shown in the left-hand column of Fig. A2). The velocities are relative to the redshift of the absorption system in the final catalogue (grey dashed line). There is a clear offset between the velocities of transitions falling in the VIS and NIR arms, with a magnitude of  $\sim 17 \text{ km s}^{-1}$  (red dashed line).

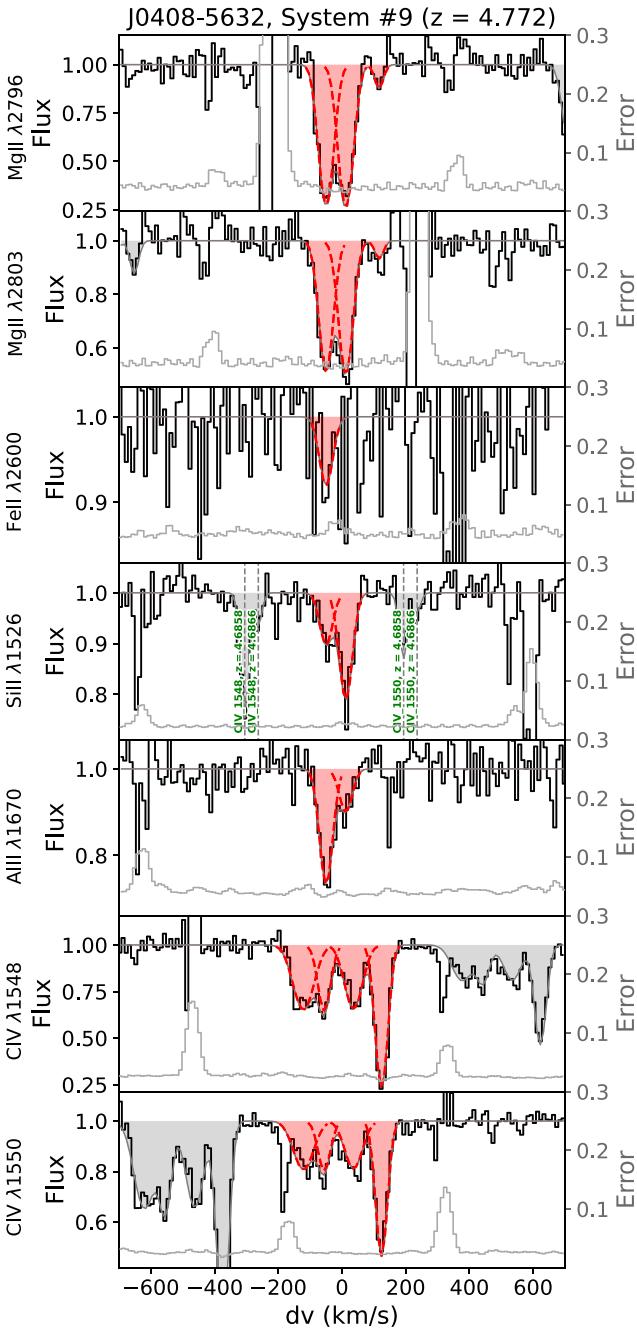
For each quasar where such a velocity offset was identified, a single velocity shift ( $\pm 5\text{--}20 \text{ km s}^{-1}$ ) was applied to the whole NIR spectrum (for which the uncertainties on the wavelength solutions are slightly larger) to maximize the inter-arm alignment for all absorption systems. The right-hand panel of Fig. A1 shows the relative centroid velocities of the transitions in the example system after shifting the

NIR spectrum. There is no evidence for any systematic variation in the measured centroid velocity as a function of wavelength (this is also apparent from the fits to the shifted line profiles in the right-hand column of Fig. A2). Higher order dispersion effects may be present but do not appear to leave any detectable signature at the spectral resolution and sampling of our observations. The final centroid velocities are spread over a  $\sim 10 \text{ km s}^{-1}$  range which is the width of a spectral pixel. This is significantly larger than the error on the absolute wavelength calibration. This remaining velocity spread may instead reflect uncertainties on the line profile shape originating from factors that are not accounted for in the error budget, such as continuum normalization.



**Figure A2.** Profiles of VIS and NIR transitions from a single-component low-ionization absorption system at  $z = 5.939$  in the spectrum of SDSSJ2310+1855. The centroid velocities of the VIS and NIR transitions are marked with blue and red dashed lines, respectively. The velocity zero-point (grey dotted line) represents the redshift of the absorption system in the final catalogue. The left-hand column shows measurements from the original spectra. The centroids of the NIR transitions are offset by  $\sim 17 \text{ km s}^{-1}$  relative to the VIS transitions (see Appendix A and Fig. A1 for further information). The right-hand column shows the profiles after shifting the NIR spectrum by  $17 \text{ km s}^{-1}$ . The absorption profiles are very well aligned across both arms of the spectrograph, indicating that a single absolute velocity shift captures the majority of the observed centroid variation.

## APPENDIX B: ACCOMPANYING DATA TABLES AND FIGURES



**Figure B1.** Voigt profile fits to the absorption system listed in Table B3. Each row shows data for a single transition. The black and grey curves show the continuum-normalized flux and error, respectively. The flux and error values are plotted over different y ranges which are shown on the left- and right-hand y-axes, respectively. The x-axis shows the velocity offset from the median redshift of the Voigt components. Red dashed lines indicate individual Voigt components and red filled regions show the total absorption from all components combined. Grey dashed lines and filled regions indicate absorption associated with systems at other redshifts, which are labelled with green text.

**Table B1.** Adopted emission redshifts and spectral resolutions for the XQR-30 and archival (E-XQR-30) quasars.

QSO ID	Sample	z	Ref	R(VIS)	R(NIR)
SDSSJ0927+2001	E-XQR-30	5.772	1,2 <sup>a</sup>	12 900	9400
PSOJ065+01	XQR-30	5.79	3	10 700	9700
PSOJ308-27	XQR-30	5.7985	4	11 800	10 600
SDSSJ0836+0054	E-XQR-30	5.81	5	13 100	10 200
PSOJ242-12	XQR-30	5.83	6	10 700	9700
PSOJ025-11	XQR-30	5.85	6	10 500	9700
PSOJ183-12	XQR-30	5.86	6	11 900	10 000
PSOJ023-02	XQR-30	5.90	6	10 900	9700
PSOJ108+08	XQR-30	5.9485	4	12 200	9800
PSOJ089-15	XQR-30	5.957	4	12 000	9700
VDESJ2250-5015	XQR-30	5.9767	4	10 800	9600
ULASJ0148+0600	E-XQR-30	5.98	7	13 300	10 000
PSOJ029-29	XQR-30	5.984	6	10 800	9900
PSOJ007+04	XQR-30	6.0008	8	11 500	9700
SDSSJ2310+1855	XQR-30	6.0031	9	12 700	9800
PSOJ009-10	XQR-30	6.0039	8	11 500	10 500
ATLASJ029-36	XQR-30	6.02	10	10 100	9200
SDSSJ0818+1722	E-XQR-30	6.02	11	11 000	7600
SDSSJ1306+0356	E-XQR-30	6.0337	8	12 000	9600
VDESJ0408-5632	XQR-30	6.0345	4	11 300	9700
PSOJ158-14	XQR-30	6.0679	12	10 800	9100
SDSSJ0842+1218	XQR-30	6.0763	8	11 400	10 000
PSOJ239-07	XQR-30	6.1106	12	11 300	11 000
CFHQSJ1509-1749	E-XQR-30	6.1225	8	11 800	8000
ULASJ1319+0950	E-XQR-30	6.133	9	13 700	9800
PSOJ217-16	XQR-30	6.1498	8	11 500	10 200
PSOJ217-07	XQR-30	6.1663	4	11 200	9900
PSOJ359-06	XQR-30	6.1722	12	11 300	10 000
PSOJ060+24	XQR-30	6.18	6	11 500	10 300
PSOJ065-26	XQR-30	6.1877	8	11 700	10 500
SDSSJ1030+0524	E-XQR-30	6.308	5	12 300	8400
SDSSJ0100+2802	E-XQR-30	6.3258	13	11 400	10 300
ATLASJ025-33	E-XQR-30	6.3379	8	11 200	9300
J2211-3206	XQR-30	6.3394	8	10 600	9100
J1535+1943	XQR-30	6.38	3	11 600	9700
J1212+0505	XQR-30	6.4386	14	11 900	10 200
(PSOJ183+05)					
J0439+1634	E-XQR-30	6.5188	15	9500	8200
VDESJ0224-4711	XQR-30	6.526	16	11 200	9400
PSOJ036+03	E-XQR-30	6.5412	17	10 700	9200
PSOJ231-20	XQR-30	6.5864	8	10 800	9800
PSOJ323+12	XQR-30	6.5881	18	10 900	9800
J0923+0402	XQR-30	6.6336	4	11 600	10 200

We adopted the best redshift measurements that were available when commencing construction of the absorption line catalogue. The redshifts are consistent with those published in D'Odorico et al. (in preparation) to within  $300 \text{ km s}^{-1}$  and these small differences do not have any significant impact on the final data set. The spectral resolution values are estimated using seeing measurements as described in Section 2.2.2. References: (1) Bosman et al. (2018), (2) Wang et al. (2010), (3) Zhu et al. (2021), (4) D'Odorico et al. (in preparation), (5) Kurk et al. (2007), (6) Bañados et al. (2016), (7) Becker et al. (2015b), (8) Decarli et al. (2018), (9) Wang et al. (2013), (10) Carnall et al. (2015), (11) Carilli et al. (2010), (12) Eilers et al. (2021), (13) Wang et al. (2016), (14) Venemans et al. (2020), (15) Yang et al. (2019), (16) Reed et al. (2019), (17) Bañados et al. (2015), (18) Mazzucchelli et al. (2017).

<sup>a</sup>This redshift is reported in Bosman et al. (2018) based on the CO(2-1) measurement from Wang et al. (2010).

**Table B2.** Wavelength regions masked during system finding.

QSO ID	Ly $\alpha$	Telluric(Y)	Masked regions (nm)		
			Telluric (J–H)	Telluric (H–K)	Telluric (K)
SDSSJ0927+2001			1345–1450	1800–1940	
PSOJ065+01	825.0–827.5				
SDSSJ0836+0054			1345–1450	1800–1940	
PSOJ025-11	825.0–836.0				1998–2030, 2100–2280
PSOJ183-12	830.0–838.8				
PSOJ023-02		1008–1030	1345–1450	1800–1940	1998–2030
PSOJ089-15	845–850			1800–1940	
J2250-5015	845.0–851.5		1345–1485	1800–1940	2100–2280
PSOJ007+04			1350–1375		1998–2030, 2100–2280
SDSSJ2310+1855			1345–1450	1800–1940	
ATLASJ029-36	853.0–854.5	1000–1025	1345–1450	1800–1975	2100–2280
SDSSJ0818+1722			1345–1450	1800–1940	
SDSSJ1306+0356			1345–1450	1800–1940	
J0408-5632			1350–1415	1800–1975	1998–2030, 2100–2280
PSOJ239-07	865–880				
CFHQSJ1509-1749				1800–1940	2100–2280
PSOJ359-06	870–872.6			1800–1940	
PSOJ060+24			1345–1450		
SDSSJ1030+0524			1345–1450	1800–1940	
SDSSJ0100+2802			1345–1450	1800–1940	2045–2080
J2211-3206			1345–1450	1800–1940	
J1535+1943	895.0–897.8		1345–1485	1800–1975	
J1212+0505			1350–1415	1820–1940	1998–2030
J0439+1634			1350–1420	1810–1940	1998–2030, 2045–2080
PSOJ036+03			1345–1450	1800–1940	
J0923+0402			1345–1485	1800–1975	

The wavelength regions listed in this table were manually masked during the initial system finding due to either strong absorption just redward of the quasar Ly $\alpha$  emission line preventing the continuum level from being accurately determined, or significant contamination by skylines or telluric absorption (see discussion in Section 3.2). No spectral regions were masked for the 16 quasars not included in this table.

**Table B3.** A condensed excerpt of the absorber catalogue for J0408-5632.

Syst	Comp	Ion/		Rest-frame	Column		Doppler <i>b</i>	Sample	Primary	(P)roximate		
ID	ID	Redshift	species	equivalent width $W(\text{\AA})$	Flag	$\log_{10} \left( \frac{N}{\text{cm}^{-2}} \right)$	parameter	flags	sample	or (I)ntervening		
(1)	(2)	(3)	(4)	(5-6)	(7)	(8-9)	(10)	(11)	(12-13)	(14)	(15)	(16)
9	–	4.77151	Mg II 2796	0.821 ± 0.013	–	13.53 ± 0.02	–	128	–	YYYY	Y	I
–	–	4.77151	Mg II 2803	0.530 ± 0.013	–	–	–	–	–	–	–	–
–	9a	4.77078	Mg II	–	–	13.21 ± 0.02	–	70	23.5 ± 2.0	YYYY	Y	I
–	9b	4.77198	Mg II	–	–	13.22 ± 0.02	–	71	23.7 ± 1.8	YYYY	Y	I
–	9c	4.77399	Mg II	–	–	12.03 ± 0.14	–	47	6.2 ± 0.0	YYYY	Y	I
9	–	4.77078	Fe II 2600	0.042	U	12.47	U	71	–	YYNY	N	I
–	9a	4.77078	Fe II	–	–	12.47	U	71	23.5 ± 2.0	YYNY	N	I
9	–	4.77159	Si II 1526	0.082 ± 0.005	–	13.51 ± 0.04	–	112	–	YYYY	N	I
–	9a	4.77078	Si II	–	–	13.00 ± 0.07	–	67	23.5 ± 2.0	YYYY	N	I
–	9b	4.77198	Si II	–	–	13.35 ± 0.04	–	67	23.7 ± 1.8	YYYY	N	I
9	–	4.77110	Al II 1670	0.105 ± 0.007	–	12.44 ± 0.04	–	110	–	YYYY	N	I
–	9a	4.77078	Al II	–	–	12.31 ± 0.04	–	67	23.5 ± 2.0	YYYY	N	I
–	9b	4.77198	Al II	–	–	11.84 ± 0.11	–	67	23.7 ± 1.8	YYYY	N	I
9	–	4.77197	C IV 1548	0.548 ± 0.007	–	14.29 ± 0.02	–	290	–	YYYY	Y	I
–	–	4.77197	C IV 1550	0.322 ± 0.008	–	–	–	–	–	–	–	–
–	9d	4.76942	C IV	–	–	13.63 ± 0.06	–	102	40.6 ± 7.4	YYYY	Y	I
–	9e	4.77067	C IV	–	–	13.43 ± 0.09	S	63	21.6 ± 5.2	YYYY	Y	I
–	9f	4.77244	C IV	–	–	13.56 ± 0.04	S	89	34.6 ± 4.3	YYYY	Y	I
–	9g	4.77413	C IV	–	–	13.95 ± 0.02	S	58	18.8 ± 1.2	YYYY	Y	I

Full catalogues for all quasars and a full catalogue for the entire sample are available in the online supplementary material and GitHub repository. Column numbers match those in the published catalogues. Absorption components are grouped into systems as described in Section 3.7. For each ion in each absorption system, we provide the total  $\log N$  and  $\Delta v_{90}$  as well as the total  $W$  of each transition. We also list the  $\log N$ ,  $\Delta v_{90}$  and  $b$  parameter of each individual Voigt profile component. (1) System ID. (2) Component ID, consisting of a system number and a component label (letter). Components with the same ID were fit using the same  $z$  and  $b$  parameter (Section 3.6). (3) Optical-depth-weighted mean redshift (systems) or Voigt profile redshift (components). We do not quote errors on the redshifts which are typically on the order of  $10^{-5}$ . (4) Ion or species. (5–7) Equivalent width measurement, error, and flag reported for each transition in each system. Measurements for ions detected at  $< 3\sigma$  are flagged as upper limits ('U'). This quantity is not reported for individual components. (8–10) Column density measurement, error, and flag, reported for both systems and individual components. Flags indicate upper limits (for ions detected at  $< 3\sigma$ ; 'U'), saturated profiles (optical depth  $\tau > 2$ ; 'S'), and ions for which  $\log N$  had to be fit for different transitions independently ('M'; see Section 3.6). (11)  $\Delta v_{90}$  measurement. (12–13)  $b$  parameter measurement and error, only reported for individual components. (14) Sample flags indicating whether (a) the system was found by the automatic line finder (Section 3.2), (b) the system was classified as real by at least 3/5 expert checkers (Section 3.5), (c) the relevant transitions lie in spectral regions free of significant sky contamination (Section 3.2), and (d) the relevant transitions lie in regions free of BAL features (Section 4.3). (15) Flag indicating whether the system is part of the primary sample (Section 6.2). (16) Flag indicating whether the absorption system is proximate (P) or intervening (I) (Section 3.9).

**Table B4.** Wavelength regions showing BAL features.

QSO ID	C IV	BAL regions (nm)	
		Si IV	N V
PSOJ065+01	918.6–995.0	865.1–896.7	–
SDSSJ0836+0054	956.6–965.6	–	–
PSOJ183-12	975.9–1010.1	–	–
PSOJ023-02	981.4–1042.2	–	–
PSOJ089-15	971.0–1072.2	883.8–967.5	848.2–858.5
J2250-5015	900.8–1001.2	854.5–880.8	–
ULASJ0148+0600	980.9–1023.0	–	–
SDSSJ2310+1855	986.3–1016.5	–	–
PSOJ009-10	937.5–956.7	846.1–863.2	–
J0408-5632	985.1–1002.1	–	–
SDSSJ0842+1218	995.1–1022.4	–	–
PSOJ239-07	1079.9–1099.4	983.6–992.0	865.9–880.3
CFHQSJ1509-1749	997.2–1017.7	–	–
PSOJ217-07	947.7–1029.0	885.1–900.3	–
J2211-3206	1043.7–1100.8	959.6–992.3	–
J0439+1634	1079.4–1120.5	991.2–1010.0	–
PSOJ231-20	1162.2–1171.5	–	931.2–938.6
J0923+0402	1065.3–1157.7	979.8–1044.6	–

BAL regions were identified by searching for contiguous troughs with normalized flux less than 0.9 over a velocity interval of at least  $2000 \text{ km s}^{-1}$  (Bischetti et al. 2022). The C IV and Si IV+N V BAL regions are computed directly from the velocity intervals and redshifts published in Bischetti et al. (2022) and Bischetti et al. (in preparation), respectively.