Origin of Metals around Galaxies. I. Catalogs of Metal-line Absorption Doublets from High-resolution Quasar Spectra

Lluís Mas-Ribas[1] , Signe Riemer-Sørensen[1] , Joseph F. Hennawi[2,3] , Jordi Miralda-Escudé[4,5], John M. O'Meara[6] ,
Ignasi Pérez-Ràfols[7], Michael T. Murphy[8], and John K. Webb[9]

[1] Institute of Theoretical Astrophysics, University of Oslo, Postboks 1029, NO-0315 Oslo, Norway; l.m.ribas@astro.uio.no
[2] Department of Physics, University of California, Santa Barbara, CA 93106, USA
[3] Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany
[4] Institut de Ciències del Cosmos, Universitat de Barcelona (UB-IEEC), Barcelona E-08028, Catalonia, Spain
[5] Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain
[6] Saint Michael's College, One Winooski Park, Colchester, VT 05439, USA
[7] Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, F-13388, Marseille, France
[8] Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia
[9] School of Physics, University of New South Wales, Sydney NSW 2052, Australia
Received 2017 October 30; revised 2018 May 22; accepted 2018 May 23; published 2018 July 20

Abstract

We present the first paper of the series Origin of Metals around Galaxies, which aims to explore the origin of the metals observed in the circumgalactic and intergalactic media. In this work we extract and build catalogs of metal absorbers that will be used in future analyses, and make our results publicly available to the community. We design a fully automatic algorithm to search for absorption metal-line doublets of the species C IV, N V, Si IV, and Mg II in high-resolution ($R \gtrsim 30,000$) quasar spectra without human intervention, and apply it to the high-resolution and signal-to-noise ratio spectra of 690 quasars, observed with the UVES and HIRES instruments. We obtain 5656 C IV doublets, 7919 doublets of Mg II, 2258 of Si IV, and 239 of N V, constituting the largest high-resolution metal-doublet samples to date, and estimate the dependence of their completeness and purity on various doublet parameters such as equivalent width and redshift, using real and artificial quasar spectra. The catalogs include doublets with rest-frame line-equivalent widths down to a few mÅ, all detected at a significance above 3σ, and covering the redshifts between $1 < z \lesssim 5$, properties that make them useful for a wide range of chemical evolution studies.

Key words: catalogs – intergalactic medium – large-scale structure of universe – line: identification – quasars: absorption lines

1. Introduction

Metals are precise probes for revealing the physical properties of the medium that they inhabit, ranging from small regions within galaxies out to the largest scales of the intergalactic medium (IGM), and from the present-day back to the redshifts of cosmic reionization (e.g., Tytler et al. 1995; Songaila & Cowie 1996; Pettini et al. 2001).

Despite their importance, our knowledge of the origin of metals is restricted to those detected in massive galaxies and their immediate surroundings, where local star formation and feedback processes create and distribute them, respectively. It remains unclear whether the main polluting mechanisms of the IGM are powerful feedback processes from distant massive and bright galaxies, or the in situ enrichment by (currently) undetectable small-galaxy populations embedded in such a medium (Pettini 2004; Bertone et al. 2005; Bouché et al. 2006, 2007; Pratt et al. 2017). Adelberger et al. (2003) proposed that galactic superwinds from massive galaxies were the mechanism responsible for the abundance of C IV absorption systems detected around $z = 3$ Lyman break galaxies, but Porciani & Madau (2005) argued that early metal pollution driven by dwarf galaxies at $6 < z < 12$ was a more likely scenario. The latter origin was also supported by the observations of Simcoe (2011), suggesting that half of the intergalactic metals observed at $z \sim 2.4$ were produced at higher redshifts, between $z \sim 4.3$ and $z \sim 2.4$, consistent with the enrichment by small galaxies (with dark matter halos of

masses $M_h \lesssim 10^{11} M_\odot$) found in the simulations by Wiersma et al. (2010). More recent simulations by Oppenheimer et al. (2012) also favored an IGM enriched by similar low-mass galaxies through supernova feedback, arguing that metals produced in massive systems might not be able to escape the galactic environment (see also Hayward & Hopkins 2017).

A powerful method to probe the origin of metals around galaxies and in the IGM is the study of their clustering signature (e.g., Adelberger et al. 2005; Scannapieco et al. 2006; Wild et al. 2008; Martin et al. 2010; Pérez-Ràfols et al. 2015). In detail, the clustering of cosmic structure follows that of the underlying dark-matter density field as $\xi(r) \simeq b^2 \xi_{DM}(r)$, where $\xi(r)$ and $\xi_{DM}(r)$ denote the distance-dependent clustering (i.e., the two-point correlation function) of structure and dark matter, respectively, and the term b is the so-called *bias factor* (Sheth & Tormen 1999; Tinker et al. 2010). This bias depends on the mass of the dark matter halo hosting the structure of interest (Zehavi et al. 2005), the redshift, and the spatial scale, although in the linear regime (large scales) the scale dependence is small (Peebles 1980; Peacock & Dodds 1996). One expects the metals and the galaxy population that produces them to inhabit the same dark matter halos and show the same clustering, and therefore, to also have the same bias factor, although the evolution of a population of halos of fixed mass modifies their bias factors if the metals are observed a long time after they were produced. In conclusion, the measurement of the bias factor of the intergalactic metals can be used as an indicator of their progenitor galaxies.

Shedding light on the sources responsible for enriching the IGM through the measurement of the metal bias factor is the main goal of our project Origin of Metals around Galaxies (OMG). We aim to obtain the metal bias by spatially cross-correlating the intergalactic metals found in a specifically built sample of quasar spectra, with the Lyman-alpha (Lyα) forest from the quasar spectra in the DR12 catalog (DR12Q, Pâris et al. 2017) of the BOSS/SDSS-III Collaboration (Eisenstein et al. 2011; Dawson et al. 2013; Alam et al. 2015). In principle, we could calculate the auto-correlation function of the metals alone, but the number of detectable metal absorbers is small (a few thousands) and this would result in weak constraints. Since the number of useful pixels in the BOSS quasar spectra denoting the Lyα forest is large (∼27 millions; Busca et al. 2013), the cross-correlation between the two tracers is a better choice. Overall, this cross-correlation will equal $\xi_{\text{metal-Ly}\alpha} \simeq b_{\text{metal}} b_{\text{Ly}\alpha} \xi_{\text{DM}}$, which is the product of the clustering of the underlying dark matter density field, the bias factor of the Lyα forest $b_{\text{Ly}\alpha}$, and the bias factor of the population of metal absorbers. The first term can be calculated from the theory of structure formation within the Lambda Cold Dark Matter (ΛCDM) cosmological model (e.g., Smith et al. 2003), and the Lyα forest bias has been constrained in recent works by Blomqvist et al. (2015), Delubac et al. (2015), and Bautista et al. (2017), so the bias factor of the metal population can be determined from the clustering.

For the calculation of the IGM metal bias, we will use weak C IV λλ1548,1550 metal absorption doublets detected in high-resolution quasar spectra, motivated by the following factors. (i) We need the largest possible number of metal systems spatially overlapping with the Lyα forest data in order to reduce the statistical uncertainties of the measurements and obtain tight constraints on the bias factor. C IV is the optimal ion for this purpose due to the large cosmic abundance of carbon, which typically results in the detection of several C IV doublets in one quasar spectrum, the exact number depending on the signal-to-noise ratio (S/N) and quasar redshift. Furthermore, the doublet nature of the absorption systems facilitates their identification, and the value of the C IV rest-frame wavelength enables coverage of a broad redshift range when the doublets are identified in ultraviolet spectra. (ii) The requirement of weak absorption systems arises from the fact that the strength of the absorption lines is broadly correlated with the metal content (e.g., Mas-Ribas et al. 2017), and in turn the mass of the host medium. We expect strong absorbers to trace massive galaxies instead of the IGM. In Mas-Ribas et al. (2017), we found a mean rest-frame equivalent width of $\langle W_r \rangle \sim 0.43$ Å for the C IV λ1548 line of the carbon doublet by stacking the spectra of ∼27,000 damped Lyman-alpha systems (DLAs; see Wolfe et al. 2005, for a review), a type of object associated with galaxies with dark matter halo masses in the range $8.7 \lesssim \log(M_h/M_\odot) \lesssim 11.8$ (Pérez-Ràfols et al. 2018; see also Font-Ribera et al. 2012). Furthermore, Vikas et al. (2013) argued that C IV systems with $W_r > 0.28$ Å and up to $W_r \sim 5$ Å inhabit dark matter halos with masses of $\log(M_h/M_\odot) \gtrsim 11.3$–13.4, and Gontcho et al. (2017) recently found indications suggesting that weaker absorbers may represent smaller halos when comparing their findings with those by Vikas et al.[10] In view of these results,

it seems plausible to consider C IV doublets with $W_r \gtrsim 0.3$ Å to be tracers of massive galaxies and not desirable for our purposes, although the exact threshold should be revisited when doing the cross-correlation. (iii) The highest possible purity (>90%) is needed for obtaining a reliable bias factor value because each false-positive detection contributes to the bias calculation with a null value that, overall, results in an underestimation of the true result. Summarizing, we need to search for a large number of C IV doublets with small equivalent widths, while ensuring the minimum number of false-positive detections, which is only possible with spectra with very high resolution and S/N. This metal search is the purpose of this first paper of the OMG series.

1.1. Observational Requirements

The first aspect of our calculations is obtaining a large enough sample of high-resolution quasar spectra. In Pérez-Ràfols et al. (2018), we recently cross-correlated ∼14,000 DLAs (data set A) with the Lyman-alpha forest sample by Busca et al. (2013), which resulted in 1σ uncertainties of around 10% for a bias factor value of $b_{\text{DLA}} \sim 2$. Assuming a similar bias factor for C IV and DLAs, and that we can detect 10 C IV systems in each quasar spectrum on average, consistent with our results presented below, obtaining similar constraints would require a search of 1400 high-resolution and S/N quasar spectra. In practice, however, not all the detected C IV systems will be weak, which results in the need for even a larger number of spectra. As we will show below, our sample consists of 690 quasars with high-resolution ($R \gtrsim 30,000$) spectra and median S/N ratios above 10. Despite this number of spectra being lower than our rough estimate, it represents the largest compilation of quasar spectra of this resolution to date, followed by the partially overlapping sample of 602 quasars by Mathes et al. (2017), who investigated the properties of Mg II absorbers. The use of our quasar sample will constitute a strong improvement compared to previous works, which were only able to use a large number of spectra for the case of low-resolution (∼2000, typical of the BOSS quasar spectra) and low S/N, and therefore could only identify the strong absorbers that probe the vicinity of massive galaxies.

Since our methodology for searching quasar spectra can be applied to metal doublets of different species in a similar fashion, we perform calculations to obtain doublet catalogs of C IV, Si IV, N V, and Mg II, and make these catalogs publicly available at https://github.com/lluism/OMG, which will enable a large number of additional metal studies by the community. We emphasize that, given our goal, we will pay special attention to building catalogs with high purity, and will be less concerned about completeness.

This paper is structured as follows. In Section 2 we present the quasar data, and in Section 3 we detail the search code designed to find the doublets. The capabilities of the code are tested in Section 4, where we estimate the completeness and purity expected from the search. The resulting metal-doublet catalogs are presented in Section 5, before concluding in Section 6.

2. Quasar Spectra

Our data are composed of four high-resolution quasar spectra data sets: we use the continuum-normalized spectra of the 170 quasars in the first public data release of the Keck Observatory Database of Ionized Absorption toward Quasars (KODIAQ DR1; O'Meara et al. 2015), and those of the 130 additional

---

[10] The aforementioned findings are also supported by studies of Mg II absorbers at redshifts $z \lesssim 2$; Mathes et al. (2017) argued that weak and strong Mg II absorbers represent different objects given their different evolutions with redshift, and Kacprzak et al. (2011) suggested that weak Mg II absorbers trace cold filaments that are being accreted onto the galaxies.

Table 1
Quasar Samples

Sample	S/N[a]	$\bar{z}$	Num. of Quasars
UVES	13.8	2.3	414
HIRES	23.3	2.3	71
Kodiaq DR1	10.7	2.7	170
Kodiaq DR2	19.2	2.7	130

Note.
[a] Median S/N considering the entire wavelength range of the spectra.

quasars incorporated in the second public data release (KODIAQ DR2; O'Meara et al. 2017).[11] We also consider 414 and 71 quasars with continuum-normalized spectra from observations with the Ultraviolet and Visual Echelle Spectrograph (UVES; Dekker et al. 2000) at the Very Large Telescope (VLT), and the High Resolution Echelle Spectrometer (HIRES; Vogt et al. 1994) at Keck, respectively (Murphy et al. 2003; M. T. Murphy et al. 2018, in preparation; King et al. 2012). Table 1 indicates the number of quasars for each sample, their median quasar redshift, and the median signal-to-noise ratio of all the spectra considering their entire wavelength range. The upper and lower panels of Figure 1 display the quasar redshift and spectral signal-to-noise ratio distributions, respectively, for the four samples, and the dashed vertical lines denote the median of each distribution.

The Kodiaq and HIRES samples were both built from studies using the same instrument (i.e., HIRES spectrograph), but we name the samples in this way because they contain different spectra. Some quasars are observed and contained in more than one sample, but we consider them all for the search, because different observations cover different wavelength ranges, and some parts of the spectrum with no detection or high noise in one sample are clearly detectable in another one. In practice, there are 690 different quasars in the samples. We will look for repeated doublets after the search, and will keep those with the highest significance in the detection. In all cases, we only use the data outside the Lyα forest, i.e., redward of the Lyα emission line of the quasars.

2.1. Quasar Continuum

The continuum (unabsorbed) spectral energy distribution (SED) of each quasar is used to obtain its normalized spectrum as

$$f_\lambda = \frac{F_\lambda}{C_\lambda},$$
$$\sigma_\lambda = \frac{E_\lambda}{C_\lambda}, \qquad (1)$$

where F_λ, E_λ, and C_λ are the observed quasar flux, uncertainty, and continuum at a given wavelength (pixel), respectively, and f_λ and σ_λ are the normalized flux and uncertainty.

For the KODIAQ data, we use the continua provided along with the observed spectra and computed as described in O'Meara et al. (2015). Briefly, the spectra are continuum-fitted by hand, using Legendre polynomials of order generally between 4 and 12, depending on the specific characteristics of every spectrum. The continua for the UVES and HIRES spectra
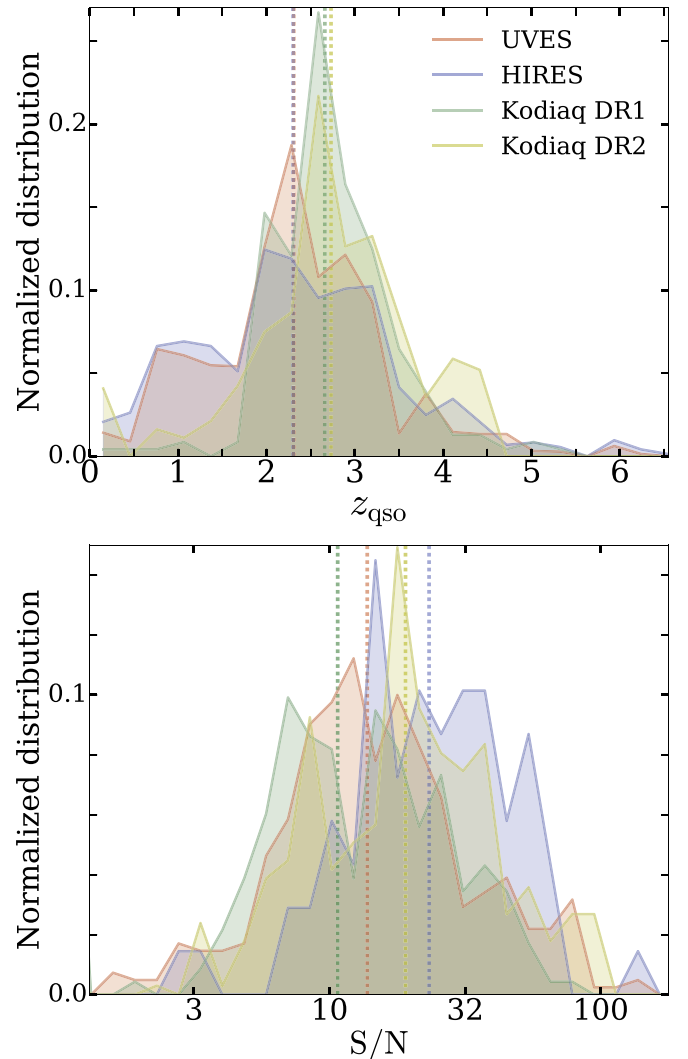
Figure 1. Upper panel: redshift distributions for the four quasar samples in our study. The vertical dashed lines denote the median of each sample. Lower panel: distributions for the signal-to-noise ratio (S/N) of each quasar sample. The vertical dashed lines represent the median of the samples, which was calculated considering the whole wavelength range covered by each spectrum.

were obtained by fitting spline curves to regions that are free from absorption features (as described in Bagdonaite et al. 2014 and Riemer-Sørensen et al. 2015). For the UVES spectra, a 6–8 order Chebyshev polynomial was generally used. Because our analysis only uses continua outside the Lya forest, where absorption lines are rare, we expect the impact of continuum uncertainty to be negligible, but we perform a more quantitative test in Section 4.3.

3. Blind Doublet Search Code

We design a code to detect metal absorption doublets in quasar spectra, based on the public code by Cooksey et al. (2008, 2010),[12] which has proven to be efficient for this purpose (e.g., Cooksey et al. 2011, 2013; Seyffert et al. 2013). Contrary to the Cooksey code, our code is built to be fully automatic, without the need for human intervention, which allows us to search a large number of spectra in a short time.
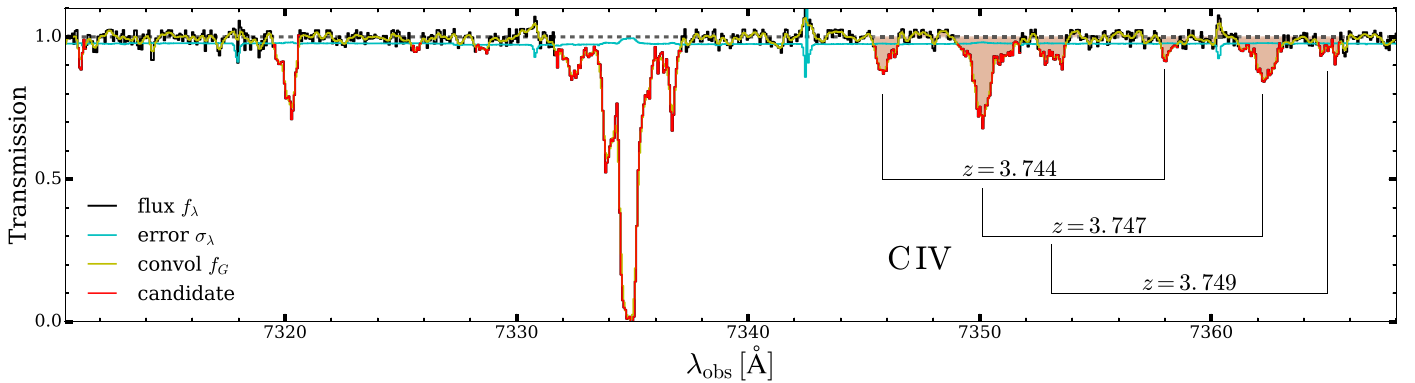
Figure 2. Wavelength range of a quasar spectrum in the UVES sample. The black line denotes the normalized spectrum, and the cyan line is the value $1-\sigma_\lambda$ at every pixel. The yellow line denotes the Gaussian-convolved flux, and the red lines are the candidate absorption features. The regions pertaining to a C IV doublet, between ~7345 Å and ~7366 Å, are represented as shaded red areas, and are listed as three doublets in our resulting catalog, although they may belong to the same absorption system. The horizontal dashed line marks the position of the unabsorbed continuum, i.e., a transmission of 100%.

In Section 3.1 below, we describe the code, focusing on the procedure for the search and identification of C IV $\lambda\lambda$1548,1550 doublets, although the code also searches for the additional doublets Si IV $\lambda\lambda$1393,1402, N V $\lambda\lambda$1238,1242, and Mg II $\lambda\lambda$2796,2803, simultaneously. In Section 3.2, the requirements that the doublet candidates have to accomplish in order to be considered real and included in the final catalogs are presented.

3.1. Automatic Metal-doublet Search

The code starts by correcting possible outliers in the spectrum: the normalized pixel fluxes are limited to the values $-\sigma_\lambda \leqslant f_\lambda \leqslant 1 + \sigma_\lambda$. Pixels with flux values outside this range are reset to the upper/lower limit. Additionally, we limit the flux between $-1 \leqslant f_\lambda \leqslant 2$ to account for unreasonable flux values usually associated with extremely large uncertainties.

After these corrections, we convolve the spectrum using a Gaussian kernel with an FWHM of three times the FWHM of the instrument, and group adjacent pixels whose Gaussian-convolved flux is $f_G \leqslant 1 - \sigma_\lambda$, also including the pixels outside this group that reside within one kernel sigma from both group sides. Only groups with a minimum number of pixels between three and five, depending on the sample, are considered. Each of these groups then represents an absorption candidate within a range defined by a minimum and maximum wavelength, λ_l and λ_h, respectively. Figure 2 displays an example of the search described above for an arbitrary wavelength interval and a spectrum from the UVES sample. The black line denotes the normalized flux and the cyan line denotes the value unity minus uncertainty at every pixel ($1-\sigma_\lambda$). The yellow line represents the Gaussian-convolved flux, and the red line denotes the candidate absorption features whose convolved flux resides below the $1-\sigma_\lambda$ line.

Under the assumption that these candidates are C IV λ1548 absorption lines (or the shortest wavelength line of the doublet for other species), we calculate their redshift and equivalent width as follows. The absorption redshift is calculated as (Cooksey 2010)

$$1 + z_{abs} = \frac{\sum_{\lambda_l}^{\lambda_h} \lambda_i \ln\left(\frac{1}{f_{\lambda i}}\right)}{\lambda_r \sum_{\lambda_l}^{\lambda_h} \ln\left(\frac{1}{f_{\lambda i}}\right)}, \quad (2)$$

where λ_r is the rest-frame wavelength of the metal line, and λ_i is the observed wavelength of the pixel i. The logarithmic term

denotes the pixel optical depth and acts as a weight, i.e., pixels with large optical depth will contribute more to the redshift calculation. For the calculation of the weights, we limit the minimum flux to $f_\lambda \geqslant (0.2\sigma_\lambda > 0.05)$ to avoid extremely large values of the optical depth where the line saturates (Cooksey et al. 2010).

The rest-frame equivalent width and its uncertainty from error propagation equate

$$W_r = \frac{1}{(1 + z_{abs})} \sum_{\lambda_l}^{\lambda_h} (1 - f_{\lambda i}) \delta\lambda_i,$$

$$\sigma_{W_r}^2 = \frac{1}{(1 + z_{abs})^2} \sum_{\lambda_l}^{\lambda_h} \sigma_{\lambda i}^2 \delta\lambda_i^2, \quad (3)$$

where $\delta\lambda_i$ is the wavelength interval between pixels.

We use the redshift measurements to calculate the expected position of the associated C IV λ1550 lines (or the higher-wavelength lines of the other doublets). In practice, the actual position of the associated lines fluctuates around the calculated value due to the effects of blends with other lines and noise, and the fact that we imposed a flux limit for the weights in the redshift calculation. A match is accepted when the velocity offset between the observed and calculated positions of the associated lines is $|\delta_v| \leqslant 7$ km s^{-1}. This value, as those of other parameters, is set to maximize the number of detections while avoiding false-positive detections (see Section 4).

Our code also computes the column density of each feature, as in Cooksey et al. (2008), using the apparent optical depth method by Savage & Sembach (1991). However, we do not present these measurements because this method is valid only when the absorption lines are unsaturated (Fox et al. 2005; although see also Cooksey et al. 2008), which occurs only for ~50%–60% of the doublets.

3.2. Doublet Acceptance Criteria

All the previous candidate doublets are finally considered real if they accomplish all the following requirements (with the corresponding values for each species):

1. The redshift of the metal absorption lines is below that of the quasar, $z_{abs} \leqslant z_{em}$, where z_{em} is the redshift of the quasar.

2. Both lines in the doublet reside outside the Lyα forest region.

3. The equivalent width ratio between the doublet lines, R_W, is within the range $1 - \sigma_{R_W} \leqslant R_W \leqslant 2 + \sigma_{R_W}$, where σ_{R_W} is the uncertainty of the ratio. The equivalent width ratio and uncertainty from error propagation are calculated as (Cooksey et al. 2010)

$$R_W = \frac{W_{r,1548}}{W_{r,1550}},$$

$$\sigma_{R_W}^2 = R_W^2 \left[\left(\frac{\sigma_{W_{r,1548}}}{W_{r,1548}} \right)^2 + \left(\frac{\sigma_{W_{r,1550}}}{W_{r,1550}} \right)^2 \right]. \quad (4)$$

These limits denote the extreme cases of completely saturated, $R_W = 1$, and unsaturated lines, $R_W = 2$.

4. The significance of the detection for the strongest doublet line is $W/\sigma_W \geqslant 3$.

The red shaded regions of Figure 2 illustrate three separate C IV doublet features, a central component at $z \sim 3.747$, plus two more within a velocity offset ranging between 100–200 km s^{-1}. We include these features in the final catalog as three different doublets, thus allowing the users to either group and treat them as a single absorption system or individually. There is another C IV doublet at the positions ≈ 7320 Å and ≈ 7332 Å ($z \sim 3.728$) not identified by the code. The strong absorber centered at ≈ 7335 Å is blended enough with the rightmost wavelength line of the doublet that the code can consider them as a single feature. This results in the determination of a higher (incorrect) redshift for the redder doublet feature compared to that of the blue one, and, notably, fails the requirement of the equivalent width ratio (condition number 3 above).

4. Code and Data Validation

We present a series of tests to assess the capabilities of our search code and the validity of our results. The impact of false-positive detections on the purity is assessed with the real spectra in Section 4.1, and we make use of mock spectra to further test the completeness in Section 4.2, and the effects of the quasar continuum placement in Section 4.3.

4.1. Searching for False Positives in Real Spectra

We run our search code on the four quasar samples, setting the theoretical separation between the two lines of the doublets to a slightly larger value than the true one. This implies that all the doublets identified by the code in this case will be false detections, and will provide a robust estimation for the purity of the results.

We set the distances between the lines of the doublet to values a few Angstrom larger than the theoretical ones, the exact value depending on the species, and avoid coincidences of these new separations with the values for doublets of other species. Because the separation in wavelength between lines changes with redshift, it is possible that for a specific redshift the separation of one species matches the value of the separation of another species at a different redshift. However, this also happens in reality, and so we enable this possibility in the analysis. We have checked that changing the separation from a few to a couple of tens of Angstrom produces negligible differences in the test results.

After obtaining the catalogs of false positives, we compute the purity as $1 - \text{false}/\text{real}$, where false and real refer to the numbers of false detections and detections in the real search (with the true doublet line separation), respectively. Figure 3 shows the results for every species (from left to right), binned in terms of the equivalent width, redshift, and significance of the detection (from top to bottom panels), using the values of the lower-wavelength lines in the doublets. The colored solid lines denote each of the four samples, and the dashed black line represents the median of the four. The middle panels indicate that there is no correlation between the purity and the redshift of the absorbers, while the upper panels show an increase of purity with equivalent width. The equivalent widths where a purity above 90% is reached varies strongly with the species and the sample. The significance of the detection (bottom panels) appears as the most robust observable, since the trends for the different samples present a narrow scatter (dispersion) around the median value. For C IV and Mg II, a median purity of $>90\%$ is reached at $W/\sigma_W \sim 32$, and at $W/\sigma_W \sim 10$ is reduced to $\sim 60\%$. For Si IV, a $W/\sigma_W \gtrsim 100$ is required to reach a median purity above 90%, similar to the case of N V, although in the latter case the number of detections is too small to establish firm conclusions.

4.2. Searching Realistic Mock Spectra

We perform tests creating and analyzing mock spectra that precisely reproduce the spectra from our different real-quasar data sets.

4.2.1. Generation of Mock Spectra

To create mock spectra, we use the publicly available code QSOSim10 (Milakovic et al. 2017), designed to simulate quasar spectra from the 10th Data Release quasar catalog of the BOSS survey (DR10Q; Pâris et al. 2014). Briefly, the code computes a quasar continuum by taking the redshift, spectral slope, and magnitude parameters from the DR10 quasar catalog for every object, and adding up to 59 broad emission lines to it. The Lyα forest is then created using statistical models for the redshift and column density distributions of neutral hydrogen, which are derived from studies of high-resolution spectra. Individual metal absorption lines associated with the neutral hydrogen absorption systems are included by means of the photoionization code CLOUDY v13.03 (Ferland et al. 2013). In detail, for every hydrogen absorber at a given redshift and hydrogen column density above $N_{H I} \geqslant 10^{15}$ cm^{-2}, the code reads its average metallicity inferred from absorption spectroscopy studies and introduces these three parameters into CLOUDY. Considering the radiation background at the corresponding redshift, CLOUDY then yields the metal column densities of 30 atomic species and about 500 ionization states. Finally, QSOSim10 creates the absorption lines of these column densities with a line width sampled from a normal distribution centered at 6 km s^{-1} and standard deviation of 1.5 km s^{-1}. Blends within different lines are allowed, but the individual lines are created using Voigt profiles. The resulting mock spectrum is finally convolved by accounting for the BOSS spectrograph resolution, and the noise from the observations and instrumentation is included. We refer the reader to Milakovic et al. (2017) for more detailed descriptions of the code and comparisons with the real BOSS spectra.
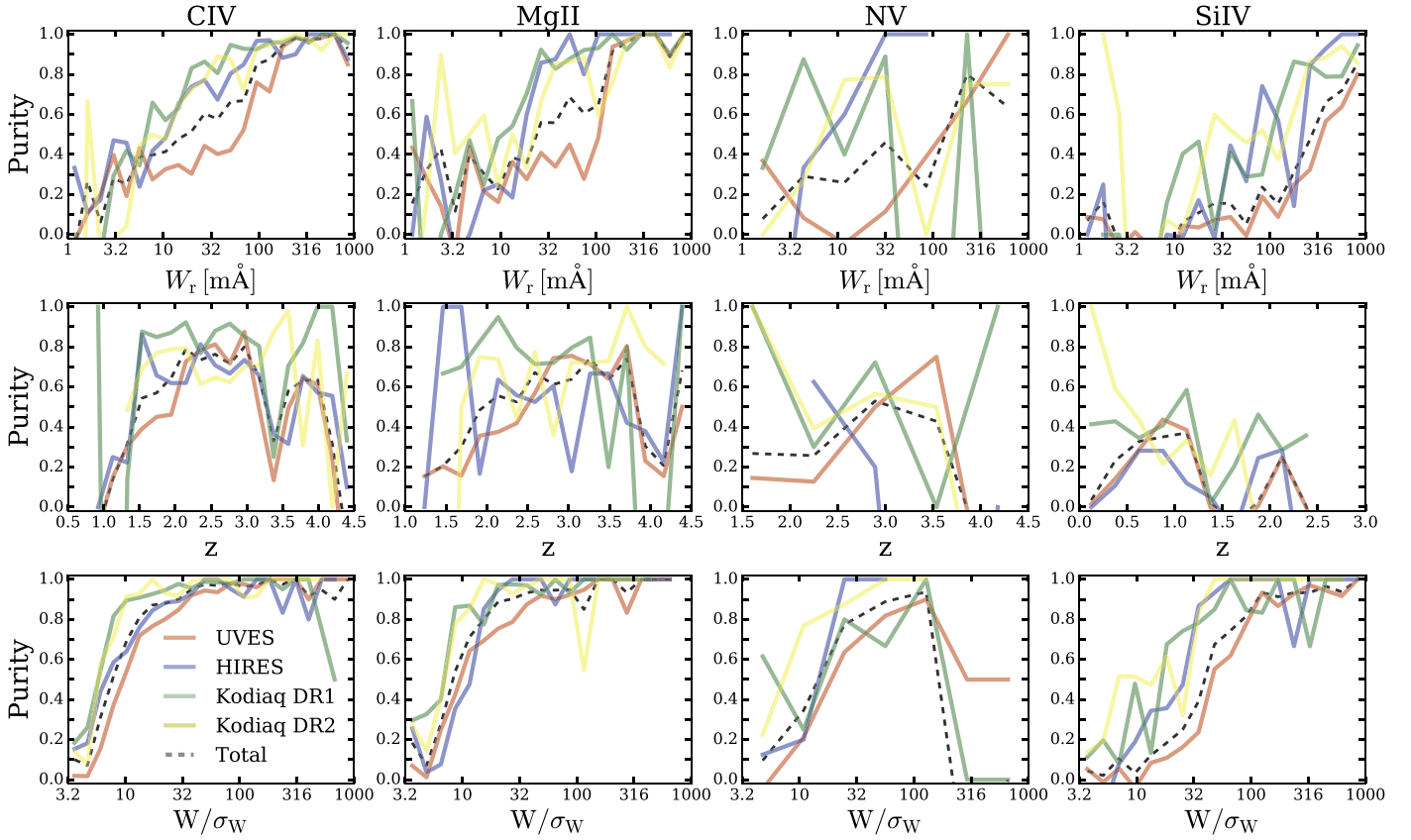
Figure 3. Purity results from the search of false doublets and a comparison with the total number of detections in the four quasar samples. The results are presented for the four species (from left to right), as are the equivalent width, redshift, and significance of the detection (from top to bottom), using the values of the lower-wavelength lines in the doublets. The colored solid lines denote each of the four samples, and the dashed black line represents the median of the four. The number of bins ranges between 15–25 for visualization purposes, and the scales are logarithmic for the cases of the equivalent widths (top panels) and significance (bottom panels), and linear for the redshifts (middle panels). The best indicator for the purity is the significance of the detection (bottom panels), showing purities above 90% at $W/\sigma_W \sim 32$ for C IV and Mg II.

For our purposes, we modify `QSOSim10` to create three sets of 1000 mock quasar spectra each, with different spectral resolutions and S/Ns broadly covering the range of values in our real data samples. We consider spectral resolutions $R = 100,000$ and $R = 50,000$, which we name HIres and MDres, respectively, referring to high resolution and medium resolution. Since `QSOSim10` computes BOSS noise for each spectrum, we adopt these BOSS values and simply apply a constant reduction factor for our mock samples, i.e., factors 3, 20, and 50, named LOsnr, MDsnr, and HIsnr, respectively, accounting for low, medium, and high S/Ns. Table 2 lists the median S/N values, the resolution, and number of spectra for each mock sample. Figure 4 illustrates the S/N distributions as shaded areas, and the distributions of the four real-quasar samples are overplotted as solid lines. The mock HIres-MDsnr sample (blue shaded region) broadly matches the HIRES sample (blue line), and the shaded red and shaded green regions cover the higher and lower tails, respectively, of the distributions of the real samples, which will be useful to determine the dependencies on S/N.

4.2.2. Completeness and Purity in Mock Spectra Searches

We run our search code on the mock spectra to assess the purity and completeness of the results for all the species. We define the completeness as the fraction of mock absorbers, i.e., actual matches, that are detected by our code (considering only
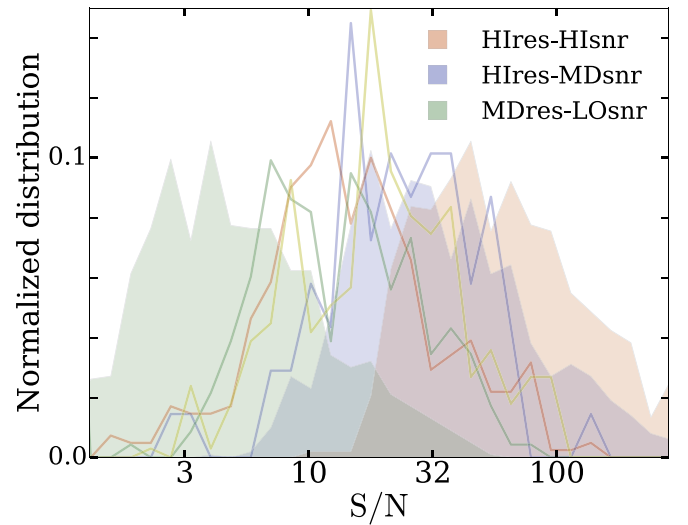

Figure 4. The S/N distributions for the three mock quasar samples are denoted as *shaded regions*, and the distributions of the real samples are overplotted as solid lines, for comparison. The blue region broadly matches the HIRES sample distribution, and the red and green areas cover the tails of the distributions of the real samples.

detections of at least 3σ significance), and for several equivalent width bins. The purity denotes the ratio between the number of the doublets detected by our code that are real
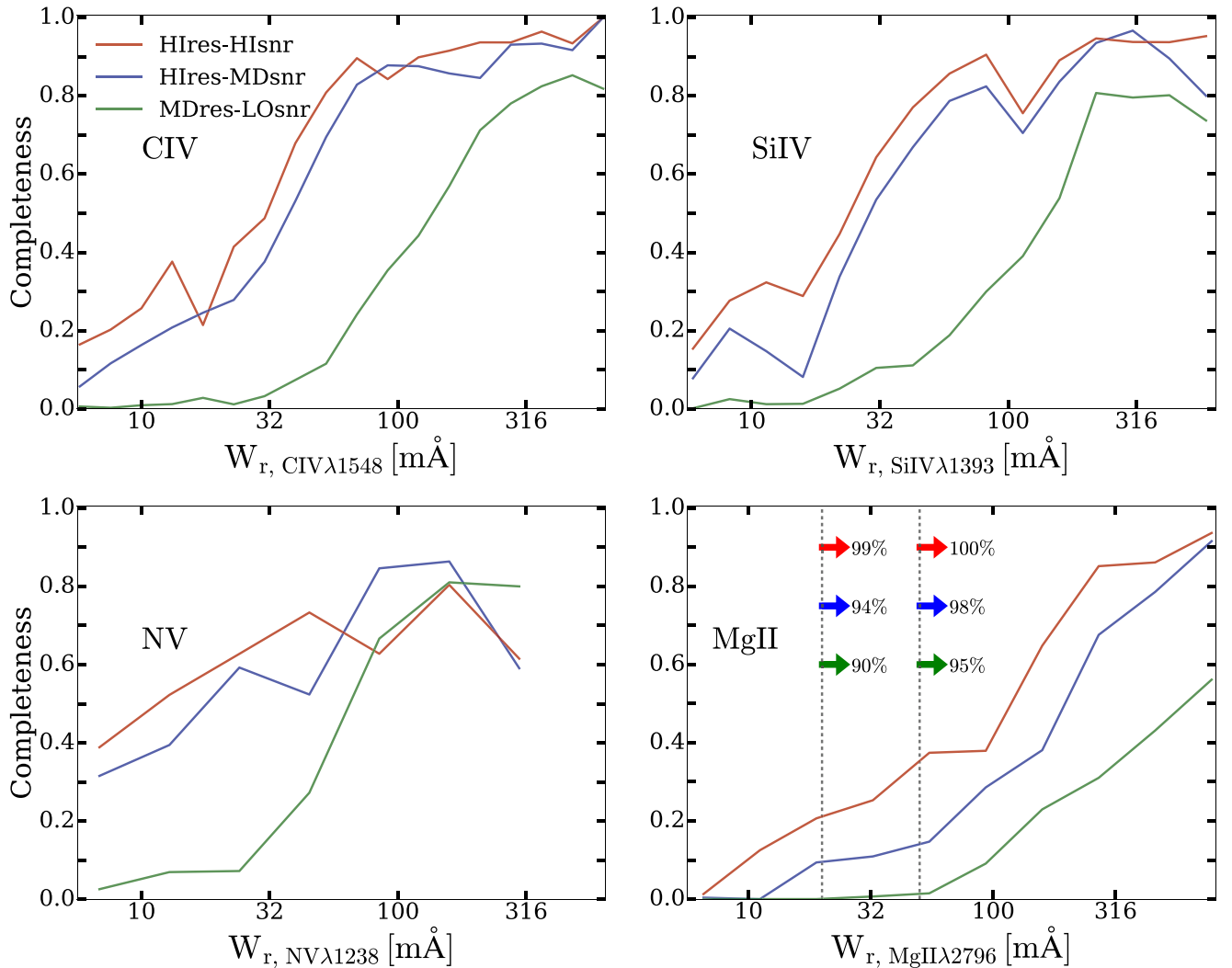
Figure 5. Completeness fractions with respect to rest-frame equivalent width, for the low-wavelength line of the C IV, Si IV, N V, and Mg II doublets, and for the three mock catalogs. The lower right panel also indicates the purity considering the absorbers with equivalent widths $W_r > 20$ mÅ and $W_r > 50$ mÅ, for the three mock samples of Mg II. The arrows have the same color-coding as the lines. The horizontal axes are logarithmically spaced to facilitate the visualization.

Table 2
Mock Spectra

Sample	R	S/N[a]	Num. of Spectra	Purity all (%)				Purity $W_r > 10$ mÅ (%)			
				C IV	Si IV	N V	Mg II	C IV	Si IV	N V	Mg II
mock HIres-HIsnr	100,000	109	1000	95	97	87	89	100	100	99	99
mock HIres-MDsnr	100,000	43	1000	97	98	89	81	100	99	100	93
mock MDres-LOsnr	50,000	7	1000	97	100	94	85	98	100	99	87

Note.
[a] Median S/N considering the entire wavelength range of the spectra.

and the total number of detections. In all cases, we only consider the doublets outside the Lyα forest. We have also used these mock spectra to explore and find the optimal search parameters.

Figure 5 displays the completeness of the search as a function of the equivalent width of the bluer member of each doublet species, and for the three mock samples. The two highest S/N samples (red and blue lines) show little difference, with the HIsnr sample (red lines) presenting a completeness ~5%–15% higher than the MDsnr sample (blue lines), owing to the better S/N of the spectra, which allows for the detection of weaker features. For the lowest-resolution and low-S/N sample (*green lines*), the completeness is, in general, significantly lower than that of the other two samples, especially at small equivalent width values. Given the distributions of S/N for the real spectra, we expect most of these to show completeness curves between the green and red lines in Figure 5.
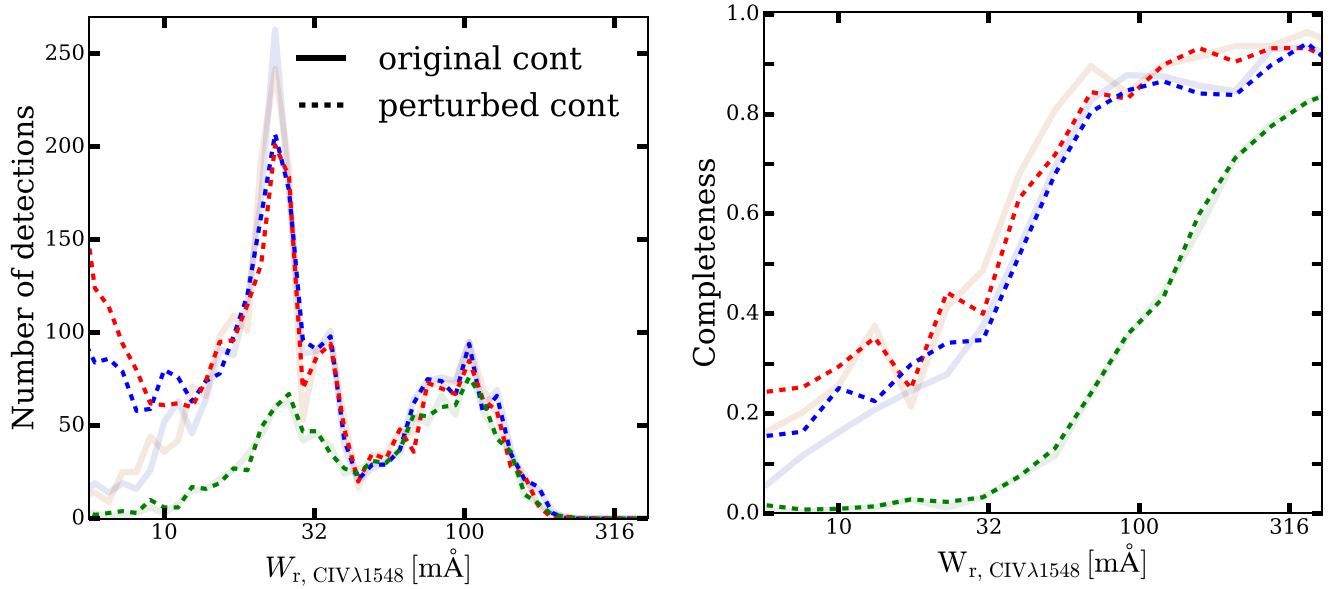
Figure 6. Left panel: comparison between the number detections for the continuum-perturbed (dashed lines) and unperturbed cases (solid lines). There is a slight increase in the number of detections at small equivalent widths ($W_r \lesssim 15$ mÅ) when the continuum variations are larger than the noise (blue and red lines). Right panel: comparison of the completeness between the three mock quasar samples without and with variations in the quasar continua. The completeness increases slightly for the two high-resolution samples (red and blue lines) at low equivalent widths, although this trend is contributed by false positives misidentified as real doublets. The low S/N sample (green lines) is not affected because of the small variations in the continuum. The horizontal axes are logarithmically spaced to facilitate the visualization.

The overall purity for each species and mock sample is indicated in Table 2, and is generally above 90% for the C IV, Si IV, and N V species, and $\gtrsim 80\%$ for Mg II. The last four columns show purity values considering only absorbers with equivalent width for the strongest line in the doublet $W_r > 10$ mÅ, reaching values within 99%–100% in all C IV, Si IV, and N V mock samples. For Mg II, the lower right panel of Figure 5 indicates additional purity values for absorbers with equivalent widths $W_r > 20$ mÅ and $W_r > 50$ mÅ, with the same color-coding as the lines in the plot. In general, purity values from the mock catalogs are higher than those computed from real spectra, likely because mocks have lower complexity, especially in the absorption line profiles and in the lack of effects, such as sky lines or outliers, that can contaminate pixels.

4.3. Continuum Placement Effects

Obtaining the quasar continuum is particularly difficult in the Lyα forest region of the spectra, where multiple hydrogen and metal absorption features overlap and reduce the average transmission. Faucher-Giguère et al. (2008) estimated the continuum uncertainty in this region to be a few percent in spectra with S/N ∼ 10, increasing toward lower S/N and higher redshifts. Outside the Lyα forest, however, most of the spectrum is free of absorption and the continuum can be precisely calculated, its uncertainty being a small fraction of the noise in those regions.

To test the effect of continuum placement on the doublet search, we introduce variations in the continuum for each of our normalized mock spectra as follows: we consider eight evenly spaced wavelength positions (nodes) in the normalized spectrum, the first and the last nodes corresponding to the first and last pixels, respectively. For every node, we then compute a small perturbation by randomly sampling a Gaussian distribution centered at 1, corresponding to the value of the

nominal unabsorbed normalized flux, and standard deviation $\sigma_C = 0.01$, corresponding to a 1% variation in the flux. Finally, the flux of the eight perturbed nodes is connected using linear interpolation, and the flux of each pixel in the spectrum is multiplied by the value of the interpolation at the corresponding pixel wavelength. Selecting $\sigma_C = 0.01$ for the region outside the forest is likely an overestimation of the uncertainty in the continuum, especially for the high S/N spectra, but we adopt this value as an upper limit and interpret our results as a worst case.

Figure 6 shows the results after searching the 1000 mock spectra of each sample with variations in the continua (dashed lines) and using the original unperturbed mocks (solid lines), for the case of C IV. The left panel shows the number of detections at every wavelength bin. The low S/N samples (green lines) show no differences between the perturbed and unperturbed cases because the noise dominates the signal and variations of the continuum are embedded within the noise. For the two high S/N cases, the overestimated continuum variations result in an increase of the number of low-equivalent-width doublets, $W_r \lesssim 15$ mÅ, detected in the search. The small variations in the completeness shown in the right panel indicate that most of the newly detected features are not real.

Even assuming these large continuum variations, the results are unaffected at equivalent widths above $W_r \sim 15$ mÅ, concluding that the continuum uncertainties are not a concern for our analysis above this threshold. We have not observed apparent effects on the redshift distributions due to changes in the continua in any case.

5. Metal-doublet Catalogs

We present here the doublet catalogs resulting from our search.

As mentioned in Section 2, a fraction of the doublets in the samples are repeated because some spectra are included in
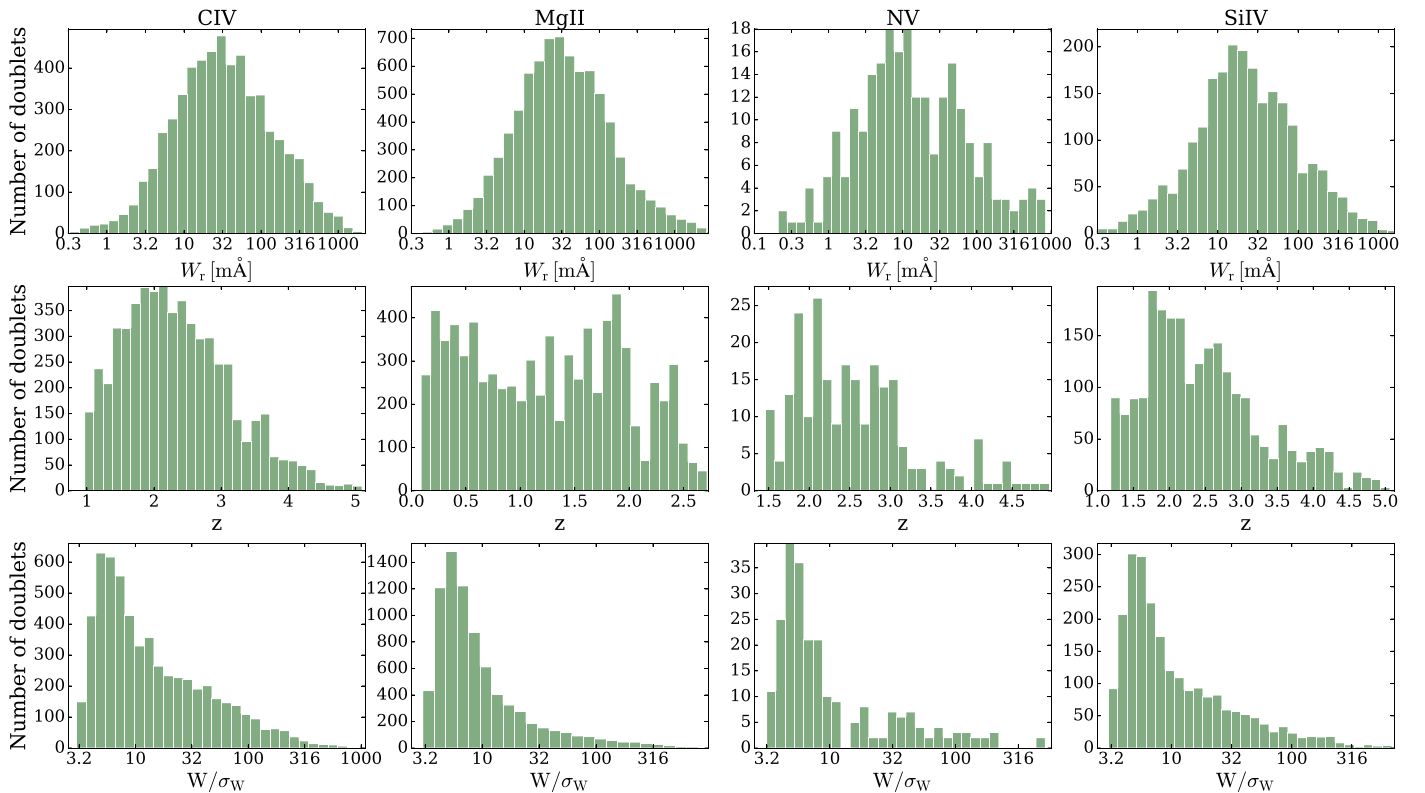
Figure 7. Rest equivalent width, W_r, redshift, z, and detection significance, W/σ_W, distributions for the higher-frequency line of the doublets in the total catalogs. The horizontal axes for equivalent width and detection significance are logarithmically spaced to facilitate the visualization.

various surveys, and for the case of the KODIAQ samples, the same quasar has several spectra from different observational campaigns. We consider two doublets to be the same when they accomplish all the following requirements: (i) the doublets appear in observations of the same quasar, (ii) the average redshifts of the doublets are offset by $\leqslant 50\,\mathrm{km\,s^{-1}}$, implying redshift variations of $\Delta z/(1+z) \sim 10^{-4}$, and (iii) the difference between the equivalent width of the respective absorption lines in the two doublets, for both lines of the doublet, is $\Delta W < 20\%$. When two doublets are considered to be the same feature, we keep the one with the highest total detection significance, expressed as $W_l/\sigma_{W_l} + W_{rg}/\sigma_{W_{rg}}$, where l and rg denote the lower- and higher-wavelength lines in the doublet, respectively. Variations of a factor of two in these two threshold values lead to variations of a few percent in the total number of doublets.

Figure 7 shows the distributions of rest-frame equivalent width, W_r, doublet redshift, z, and significance of the detection, W/σ_W, for the lower-wavelength lines in the doublet, and the four species in the final total catalog, including the spectra from all samples. The redshift of the doublets is computed as the weighted mean of the redshift of the two lines in the doublet, using the significance of the detections as the weighting parameter. Table 3 quotes the number of doublets for each species and for each catalog, and in the final catalog after accounting for the repeated doublets.

Table 4 illustrates the structure of the final C IV catalog available at https://github.com/lluism/OMG. The name, R.A., and decl. of the quasar are quoted in the first to the third columns, respectively, and the average redshift of the doublet is displayed in the fourth column. The fifth through tenth columns denote the rest-frame equivalent width and its

Table 3
Doublet Catalogs

Sample	Number of Doublets			
	C IV	Si IV	N V	Mg II
UVES	3291	1394	156	6482
HIRES	849	290	25	481
KODIAQ DR1	1190	430	34	550
KODIAQ DR2	1515	598	75	1004
Total	5656	2258	239	7919

uncertainty (in mÅ), and detection significance, for the C IV $\lambda1548$ and C IV $\lambda1550$ lines, respectively. We list the detection significances computed from the wavelength and uncertainty values in Angstrom, at the observer frame, using the redshift for each individual line, and before rounding. In some cases, these values differ from the simple division of the values quoted in the table. The last column denotes the quasar sample from where the doublet is obtained. The individual catalogs and distribution plots for each species and quasar sample are publicly available at the link indicated above.

6. Conclusions

In this first paper of the series OMG, we build and publicly release the large metal-line doublet catalogs that we will analyze in future papers.

We have designed a metal-doublet search algorithm, based on that by Cooksey et al. (2008), that searches for metal-line doublets

Table 4
Example C IV Doublet Catalog

Quasar	R.A.	Decl.	$\bar{z}$	$W_{r,1548}$	$\sigma_{W_{r,1548}}$	$W_{48}/\sigma_{W_{48}}$	$W_{r,1550}$	$\sigma_{W_{r,1550}}$	$W_{50}/\sigma_{W_{50}}$	Sample
J082540+354414	08:25:40.12	+35:44:14.20	3.073	7.81	1.21	6.22	5.31	1.11	4.95	KODIAQ
J000150-015940	00:01:50.00	−01:59:40.00	1.023	99.52	20.40	4.87	66.32	17.31	3.82	UVES
J045142-132033	04:51:42.60	−13:20:33.00	2.847	27.41	1.40	19.55	11.60	1.33	8.95	HIRES

Note. First column: quasar name. Second column: R.A. Third column: decl. Fourth column: average redshift of the doublet. Fifth column: rest-frame equivalent width of C IV λ1548 (in mÅ). Sixth column: rest-frame uncertainty for the equivalent width of C IV λ1548 (in mÅ). Seventh column: detection significance for the C IV λ1548 line. Eighth column: rest-frame equivalent width of C IV λ1550 (in mÅ). Ninth column: rest-frame uncertainty for the equivalent width of C IV λ1550 (in mÅ). Tenth column: detection significance for the C IV λ1550 line. Eleventh column: quasar sample from where the doublet is obtained.

of C IV $\lambda\lambda$1548,1550, Mg II $\lambda\lambda$2796, 2803, Si IV $\lambda\lambda$1393,1402, and N V $\lambda\lambda$1238,1242, in quasar spectra in a fully automatic fashion, i.e., without human intervention. We have tested the purity of the results by looking for false doublets using line separations larger than the true ones in real spectra, and have created three large samples of mock quasar spectra with different resolution and S/N values to estimate the completeness. We have also assessed the impact of variations in the determination of the intrinsic quasar continua on the purity. Finally, we have applied our code to the rest-frame wavelength ranges outside the Lyα forest of the spectra of 690 quasars, compiled from the KODIAQ data sets and observations with the UVES and HIRES instruments. Our findings can be summarized as follows:

1. We build catalogs with 5656 C IV $\lambda\lambda$1548, 1550 doublets, 7919 doublets of Mg II $\lambda\lambda$2796,2803, 2258 of Si IV $\lambda\lambda$1393,1402, and 239 of N V $\lambda\lambda$1238,1242, available at https://github.com/lluism/OMG.
2. We present the purity of the individual samples with the equivalent width, redshift, and significance of the detection, the latter being the parameter that purity depends on the most.
3. We expect contamination to the purity arising from errors in the calculation of the continuum to be negligible outside the Lyα forest. In cases of large continuum errors, if any, this effect may introduce false-positive detections with small equivalent widths ($\lesssim$15 mÅ) in spectra with high signal-to-noise ratios (S/N $\gtrsim$ 10).

In the upcoming work, we will perform the cross-correlation of the weak C IV absorbers found here with the Lyα forest from BOSS. These calculations will enable us to obtain the bias factor of the intergalactic absorbers at different equivalent widths, and thus place constraints on their progenitor galaxy population.

ORCID iDs

Lluís Mas-Ribas ● https://orcid.org/0000-0003-4584-8841
Signe Riemer-Sørensen ● https://orcid.org/0000-0002-5308-7651
Joseph F. Hennawi ● https://orcid.org/0000-0002-7054-4332
John M. O'Meara ● https://orcid.org/0000-0002-7893-1054

References

Adelberger, K. L., Shapley, A. E., Steidel, C. C., et al. 2005, ApJ, 629, 636
Adelberger, K. L., Steidel, C. C., Shapley, A. E., & Pettini, M. 2003, ApJ, 584, 45
Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, ApJS, 219, 12
Bagdonaite, J., Ubachs, W., Murphy, M. T., & Whitmore, J. B. 2014, ApJ, 782, 10
Bautista, J. E., Busca, N. G., Guy, J., et al. 2017, A&A, 603, A12
Bertone, S., Stoehr, F., & White, S. D. M. 2005, MNRAS, 359, 1201
Blomqvist, M., Kirkby, D., Bautista, J. E., et al. 2015, JCAP, 11, 034
Bouché, N., Lehnert, M. D., Aguirre, A., Péroux, C., & Bergeron, J. 2007, MNRAS, 378, 525
Bouché, N., Lehnert, M. D., & Péroux, C. 2006, MNRAS, 367, L16
Busca, N. G., Delubac, T., Rich, J., et al. 2013, A&A, 552, A96
Cooksey, K. L., Kao, M. M., Simcoe, R. A., O'Meara, J. M., & Prochaska, J. X. 2013, ApJ, 763, 37
Cooksey, K. L., Prochaska, J. X., Chen, H.-W., Mulchaey, J. S., & Weiner, B. J. 2008, ApJ, 676, 262
Cooksey, K. L., Prochaska, J. X., Thom, C., & Chen, H.-W. 2011, ApJ, 729, 87
Cooksey, K. L., Thom, C., Prochaska, J. X., & Chen, H.-W. 2010, ApJ, 708, 868
Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10
Dekker, H., D'Odorico, S., Kaufer, A., Delabre, B., & Kotzlowski, H. 2000, Proc. SPIE, 4008, 534
Delubac, T., Bautista, J. E., Busca, N. G., et al. 2015, A&A, 574, A59
Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72
Faucher-Giguère, C.-A., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008, ApJ, 688, 85
Ferland, G. J., Porter, R. L., van Hoof, P. A. M., et al. 2013, RMXAA, 49, 137
Font-Ribera, A., Miralda-Escudé, J., Arnau, E., et al. 2012, JCAP, 11, 059
Fox, A. J., Savage, B. D., & Wakker, B. P. 2005, AJ, 130, 2418
Gontcho, S. G. A., Miralda-Escudé, J., Font-Ribera, A., et al. 2017, arXiv:1712.09886
Hayward, C. C., & Hopkins, P. F. 2017, MNRAS, 465, 1682
Kacprzak, G. G., Churchill, C. W., Barton, E. J., & Cooke, J. 2011, ApJ, 733, 105
King, J. A., Webb, J. K., Murphy, M. T., et al. 2012, MNRAS, 422, 3370
Martin, C. L., Scannapieco, E., Ellison, S. L., et al. 2010, ApJ, 721, 174
Mas-Ribas, L., Miralda-Escudé, J., Pérez-Ràfols, I., et al. 2017, ApJ, 846, 4
Mathes, N. L., Churchill, C. W., & Murphy, M. T. 2017, arXiv:1701.05624
Milakovic, D., Webb, J. K., & Dumont, V. 2017, QSOSIM, 10: Simulated Quasar Spectra Generator, Zenodo, doi:10.5281/zenodo.439394
Murphy, M. T., Webb, J. K., & Flambaum, V. V. 2003, MNRAS, 345, 609
O'Meara, J. M., Lehner, N., Howk, J. C., et al. 2015, AJ, 150, 111
O'Meara, J. M., Lehner, N., Howk, J. C., et al. 2017, arXiv:1707.07905

Oppenheimer, B. D., Davé, R., Katz, N., Kollmeier, J. A., & Weinberg, D. H. 2012, MNRAS, 420, 829

Pâris, I., Petitjean, P., Aubourg, É., et al. 2014, A&A, 563, A54

Pâris, I., Petitjean, P., Ross, N. P., et al. 2017, A&A, 597, A79

Peacock, J. A., & Dodds, S. J. 1996, MNRAS, 280, L19

Peebles, P. J. E. 1980, The Large-scale Structure of the Universe (Princeton, NJ: Princeton University Press)

Pérez-Ràfols, I., Font-Ribera, A., Miralda-Escudé, J., et al. 2018, MNRAS, 473, 3019

Pérez-Ràfols, I., Miralda-Escudé, J., Lundgren, B., et al. 2015, MNRAS, 447, 2784

Pettini, M., Ellison, S. L., Schaye, J., et al. 2001, ApSSS, 277, 555

Pettini, M. 2004, Cosmochemistry. The Melting pot of the Elements, ed. C. Esteban et al. (Cambridge: Cambridge University Press)

Porciani, C., & Madau, P. 2005, ApJL, 625, L43

Pratt, C. T., Stocke, J. T., Keeney, B. A., & Danforth, C. W. 2017, arXiv:1706.05103

Riemer-Sørensen, S., Webb, J. K., Crighton, N., et al. 2015, MNRAS, 447, 2925

Savage, B. D., & Sembach, K. R. 1991, ApJ, 379, 245

Scannapieco, E., Pichon, C., Aracil, B., et al. 2006, MNRAS, 365, 615

Seyffert, E. N., Cooksey, K. L., Simcoe, R. A., et al. 2013, ApJ, 779, 161

Sheth, R. K., & Tormen, G. 1999, MNRAS, 308, 119

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

Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, MNRAS, 341, 1311

Songaila, A., & Cowie, L. L. 1996, AJ, 112, 335

Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, ApJ, 724, 878

Tytler, D., Fan, X.-M., Burles, S., et al. 1995, in Proc. ESO Workshop, QSO Absorption Lines, ed. G. Meylan (Berlin: Springer), 289

Vikas, S., Wood-Vasey, W. M., Lundgren, B., et al. 2013, ApJ, 768, 38

Vogt, S. S., Allen, S. L., Bigelow, B. C., et al. 1994, Proc. SPIE, 2198, 362

Wiersma, R. P. C., Schaye, J., Dalla Vecchia, C., et al. 2010, MNRAS, 409, 132

Wild, V., Kauffmann, G., White, S., et al. 2008, MNRAS, 388, 227

Wolfe, A. M., Gawiser, E., & Prochaska, J. X. 2005, ARA&A, 43, 861

Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2005, ApJ, 630, 1