A metal-line strength indicator for damped Lyman alpha (DLA) systems at low signal-to-noise

Andreu Arinyo-i-Prats1, Lluís Mas-Ribas1,2, Jordi Miralda-Escudé1,3, Ignasi Pérez-Ràfols4 and Pasquier Noterdaeme5

1Institut de Ciències del Cosmos, Universitat de Barcelona/IEEC, Barcelona E-08028, Catalonia, Spain
2Institute of Theoretical Astrophysics, University of Oslo, Postboks 1029, 0315 Oslo, Norway
3Institució Catalana de Recerca i Estudis Avançats, Barcelona E-08010, Catalonia, Spain
4Aix Marseille Univ, CNRS, LAM, Laboratoire d’Astrophysique de Marseille, F-13013 Marseille, France
5Institut d’Astrophysique de Paris, UPMC & CNRS, F-75014 Paris, France

Accepted 2018 August 29. Received 2018 August 29; in original form 2018 January 9

ABSTRACT
The bias factor of damped Ly α (DLAs) systems, derived from the cross-correlation with the Ly α forest in absorption spectra of the Baryon Oscillation Spectroscopic Survey of SDSS-III, has been used to infer the characteristic mass of DLA host haloes. So far, no dependence of this bias factor with hydrogen column density $N_{\text{HI}}$ or other parameters has been detected. With the aim of measuring the dependence of the bias factor on the strength of metal lines, we define the metal strength parameter $S$, based on optimally combining equivalent widths of 17 metal lines to obtain the highest possible signal-to-noise ratio of $S$ for individual candidate DLAs (defined as having $N_{\text{HI}} \geq 10^{20}$ cm$^{-2}$). We present the distribution of metal strength for these DLAs and the dependence of its mean value on $N_{\text{HI}}$ and redshift. We search for systematic effects and variations in the catalogue purity by examining the dependence of the $S$ distribution on the spectral signal-to-noise and the estimated error on $S$. A catalogue of DLAs with measured equivalent widths for the selected 17 metal lines and the value of $S$ are made publicly available, which are used in a separate paper to measure the dependence of the DLA bias factor on the $S$ parameter. The relation of the metal strength on the gas metal abundances and velocity dispersion is complicated by the saturation of metal lines, and remains to be determined in future work.

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

1 INTRODUCTION
Observations of damped Lyman alpha (DLAs) systems in quasar absorption spectra are crucial for understanding galaxy formation at different epochs (e.g. Wolfe et al. 1986; Wolfe, Gawiser & Prochaska 2005). Most of the atomic hydrogen in the Universe is present in DLAs, or absorption systems with $N_{\text{HI}} > 2 \times 10^{20}$ cm$^{-2}$, which are self-shielded against the external ionizing background. This atomic hydrogen accounts for ~2 to 3 per cent of all the baryonic matter in the Universe, a fraction that remains roughly constant during the epoch of the maximum rate of star formation at $2 < z < 4$ (Noterdaeme et al. 2012) and is comparable to the fraction of baryons in stars at that epoch (Shapley 2011).

The bias factor of a population of objects measures the amplitude of the large-scale fluctuations in the density of these objects, compared to the underlying mass fluctuations. The bias factor is related to the mass of the haloes hosting the galaxies that give rise to the DLA absorber: the larger the bias, the more massive the host halo. A moderately high mean large-scale bias factor for DLAs $b_{\text{DLA}} \simeq 2$ has been measured at $z \simeq 2.3$ by Font-Ribera et al. (2012) and Pérez-Rafols et al. (2018), using the cross-correlation of DLAs with the Ly α forest absorption on comoving scales larger than $10 \text{h}^{-1}$ Mpc.

The measured mean bias factor implies that many DLAs are hosted by haloes of massive galaxies or galaxy groups, and that

1Following previous works, we call DLAs all absorbers with $N_{\text{HI}} \geq 10^{20}$ cm$^{-2}$, as determined following Noterdaeme et al. (2009). These are actually candidate DLAs because of the imperfect catalogue purity, and are also often designated sub-DLAs in the literature when $N_{\text{HI}} < 2 \times 10^{20}$ cm$^{-2}$, but we call them DLAs generally because the column density threshold normally used is only a convention.
an extended distribution of atomic gas is generally present in these haloes out to radii larger than the size of visible galaxies to account for the large cross-sections required to explain the observed rate of incidence. Although chemical abundances and kinematic studies have associated a portion of DLAs with dwarf galaxies (Berg et al. 2015; Cooke, Pettini & Jorgenson 2015; Skúladóttir et al. 2018), the measured mean bias factor is not consistent with nearly all DLAs being located in dwarf galaxies. This contradiction can be explained if strong galaxy winds expel the gas from low-mass haloes to reduce the contribution of dwarf galaxies to DLAs, and spread the gas out to large radius in massive haloes to increase their DLA cross-section (e.g. Barnes & Haehnelt 2014; Bird et al. 2015).

The Ly α absorption profile used to detect a DLA yields information only about the hydrogen column density, $N_{\text{HI}}$. The DLA bias factor has not been found to depend on $N_{\text{HI}}$ in any discernible way (Pérez-Rafols et al. 2018b). However, other properties of a DLA can be derived from metal lines. The abundances are roughly similar to those of halo stars and external globular clusters in the Milky Way, although differences among these populations can be present when examined in detail (Rafelski et al. 2012; Berg et al. 2015; Skúladóttir et al. 2018). The metal-line absorption profiles reveal a rich and diverse velocity structure characterized by multiple components, with a total velocity width ranging from 10 to 200 km s$^{-1}$ (Prochaska & Wolfe 1997, 1998; Wolfe 1998). The derived metallicities are distributed over a broad range ($10^{-3} \leq Z/Z_{\odot} \leq 10^{-1}$), and their average declines slowly with redshift (e.g. Prochaska & Wolfe 2002; Rafelski et al. 2012, 2014).

Most studies of the DLA metal lines are based on high-resolution, high signal-to-noise quasar spectra, which are necessary to detect the small equivalent width lines and to reveal the complexity of the velocity profiles. However, the large samples of DLAs required to measure cross-correlations that are used to infer the DLA linear bias factor can only be obtained at present with low resolution and low signal-to-noise spectra similar to those of the Baryon Oscillation Spectroscopic Survey (BOSS) of SDSS-III (Eisenstein et al. 2011; Dawson et al. 2013; Smeet al. 2013; Alam et al. 2015), which allow only measurements of the equivalent widths of the strongest metal lines with large error bars.

In this paper, we aim to define a parameter characterizing the strength of the metal lines of a DLA, which can be measured with the highest possible signal-to-noise ratio by combining all the metal absorption lines that are usually observable for DLAs. If this metal strength can be measured with sufficient accuracy for individual DLAs, even in low signal-to-noise spectra of a large spectroscopic survey similar to BOSS, then the mean bias factor and other average properties of DLAs can be measured as a function of this metal strength. So far, no dependence of the DLA bias factor on any property has been measured, but we expect the DLA host halo mass to increase with metallicity and velocity dispersion of the absorption lines, because of the well-known mass–metallicity relation for galaxies. In fact, a relation of metallicity and velocity dispersion is also present in DLAs (Neeleman et al. 2013), so we may reasonably expect an increase of the bias factor with metal strength to be detected for DLAs. As we shall explain below, caution is required in interpreting any such relation because the metal strength parameter we will define depends only on the mean metal-line equivalent widths that are measurable in low-resolution absorption spectra, which are affected both by metal abundances and the gas velocity dispersion, and have a broad range of saturation levels. Nevertheless, the mean absorption spectrum can be measured for groups of DLAs with different metal strength using the technique discussed in Mas-Ribas et al. (2017), which contains information that can help disentangle variations of the gas velocity dispersion and metal abundances with metal strength.

This paper focuses on the definition of the metal strength parameter and the study of its distribution in the DLA sample of the BOSS survey. A second paper (Pérez-Rafols et al. 2018a) will measure the mean bias factor as a function of this metal strength parameter. In a third future paper, we plan to analyse other mean DLA properties using the stacked absorption spectrum of DLAs with different measured metal strength. This first paper is organized as follows: Section 2 specifies the DLA catalogue and the quasar spectra from BOSS that we use. In Section 3 we describe in detail our method to measure line equivalent widths of a selected set of 17 low-ionization metal lines. In Section 4 we define the metal strength, and a second quantity that is $N_{\text{HI}}$-corrected for the effect of the hydrogen column density on the mean strength of metal lines. Results are presented in Section 5, where we discuss the distribution of this metal strength parameter and several systematic effects due to impurities in the DLA catalogue, and we make publicly available a catalogue with our measurements of equivalent widths and the metal strength parameter for each DLA. Finally, we present the conclusions in Section 6.

2 DATA SAMPLE

We use the quasar spectra in the complete SDSS-III BOSS Data Release 12 (DR12), from the Quasar Catalogue DR12Q (Páris et al. 2017). A detailed description of the SDSS telescope and the BOSS instrument obtaining the spectra is found in Gunn et al. (1998, 2006), Dawson et al. (2013), Smeet al. (2013), and the method to select quasar targets is described in Ross et al. (2012).

We use the DR12 extension of the candidate DLA catalogue of Noterdaeme et al. (2012b), containing a total of 34 050 DLAs with a column density $\log(N_{\text{HI}}) \geq 20$. The method to detect these DLAs, described in Noterdaeme et al. (2009), is an automatic profile recognition procedure using Spearman correlation analysis with a Voigt profile. Only the Ly α absorption line is considered to decide if an absorption feature is included as a DLA in the catalogue. This ensures that there is no selection bias in favour of DLAs with strong metal lines. The presence of metal lines is, however, used to refine the accuracy of the measured redshift of the detected DLAs. This catalogue will be referred to as DR12-DLA from now on and we designate all the candidate DLAs in our sample as DLAs for simplicity.

The DLA sample we use in this work is very similar to the sample of Mas-Ribas et al. (2017), who used an earlier version of the DR12-DLA catalogue with minor differences in the selected DLAs; their fig. 1 showed the distribution in redshift and $N_{\text{HI}}$ in their sample. Most DLAs are at redshifts $1.9 < z < 3.5$, with a small fraction extending to higher redshifts. DLAs are usually defined as having $\log(N_{\text{HI}}) \geq 20.3$ (roughly the column density at which the gas becomes mostly neutral due to self-shielding, although this depends on the gas density; see Wolfe et al. 1986), nevertheless we use all the systems going down to column densities $N_{\text{HI}} \geq 10^{20}$ cm$^{-2}$ to increase our sample size and decrease the statistical error. We note that no evidence of a change of the bias factor with column density is found in Pérez-Rafols et al. (2018b).

In our sample a fraction of the candidates are expected to be false DLAs arising from a combination of noise, regions of strong Ly α forest absorption, and blended Lyman limit systems that are confused with a DLA in low signal-to-noise spectra. The purity of the catalogue, or fraction of DLAs that are real, is expected to decline as the signal-to-noise ratio and the column density decrease.
The catalogue gives, for each detected DLA, the quasar and DLA redshift, the continuum-to-noise ratio (hereafter, CNR) of the spectrum in the Ly α forest region (as defined in Noterdaeme et al. 2012b), and the DLA column density. For each DLA in the catalogue, we use the corresponding quasar spectrum from BOSS to re-measure the associated metal-line equivalent widths. We use the co-added spectra, which have wavelength bins of width \( \Delta \log_\lambda(\lambda) = 10^{-1} \), corresponding to a velocity width of 69.05 km s\(^{-1}\).

3 MEASURING METAL-LINE EQUIVALENT WIDTHS

In this section we describe the method used to compute the equivalent widths \( W \) of a selection of 17 metal lines of DLAs. A weighted average of these equivalent widths will be used as a definition of the metal strength of each DLA. We do not aim to detect individual metal lines, but to measure an equivalent width at the expected central wavelength, given the known redshift of the DLA. Equivalent widths are measured and used even if their value is consistent with zero or negative, due to noise. Briefly, our method uses a measurement window around the central wavelength of the absorption line over which the equivalent width is integrated, and two windows around it to determine a continuum from a linear regression of the measured flux. In Section 3.1 we explain the selection of the 17 lines. The windows are described in Section 3.2. In Section 3.3 we specify which of the 17 lines are used for each individual DLA depending on how the windows are placed in the quasar spectrum, and Section 3.4 gives the details of how the equivalent widths and their errors are calculated.

The DR12-DLA catalogue of Noterdaeme et al. (2009) includes measurements of \( W \) for 10 metal lines. However, their method provides a biased estimate of these equivalent widths because they are only measured when their detection is considered significant. This means that a metal line may be included in the catalogue when the value of \( W \) has been increased by noise, or may be dropped when it has been reduced, systematically affecting the average \( W \). Negative equivalent widths caused by noise also need to be included in the catalogue to avoid bias. Our method will select the lines to be measured depending only on the location of the windows to be used for the continuum determination and the equivalent width integration, but not on the value of \( W \) that is derived.

3.1 List of selected metal lines

The 17 metal lines we use to evaluate our metal strength for each DLA are selected to be transitions of low-ionization species that lie in the DLA rest-frame wavelength interval between 1260 and 3000 Å, with a mean equivalent width measured from the stacked spectrum in Mas-Ribas et al. (2017) of \( W > 0.05 \) Å. The wavelength lower limit is set to avoid the Ly α forest and the Ly α quasar emission line, and the upper limit is determined by the maximum wavelength reached by the BOSS spectrograph (Smee et al. 2013) at the lowest redshifts at which DLAs are found, combined with the absence of strong absorption lines beyond our longest wavelength line, Mg I at 2852 Å.

The low-ionization species are either neutral atoms or ions that are once or twice ionized, but we exclude higher ionization lines like C IV and Si IV. We use this combination of low-ionization transitions to obtain a metal strength that reflects a property of the low-ionization gas with the highest possible signal-to-noise, by combining \( W \) measurements of all the available lines. We exclude high-ionization lines because these are known to arise in very low-density gas and reflect a more extended gas distribution with different kinematics (e.g. Ellison et al. 2000; Fox et al. 2007; Tumlinson, Peeples & Werk 2017).

The 17 selected line transitions are listed in Table 1, where we give the equivalent width computed from the stacked DR12 spectrum designated as total in Mas-Ribas et al. (2017), compared to the mean equivalent width \( W \) measured with our method, as described below. The first three lines in Table 1 are blends of lines that are too close to be measured separately, and we list the total equivalent widths for the blend. Their wavelengths are that of the Si II transition in the first blend (which is much stronger than Fe II), the average of the O I and Si II transitions in the second blend (which are comparably strong), and the C II transition in the third blend (which is much stronger than the transition of the metastable state C II*). The rest of the quantities will be discussed in Section 4.

3.2 Measurement and continuum windows

We use a fixed measurement window to integrate the equivalent width for all the lines, with a width that is normally set to 15 pixels in the co-added spectra, except for the O I–Si II blend for which we use a window width of 27 pixels. The measurement window is centred on the pixel that includes the central wavelength of the line listed in Table 1 multiplied by \( 1 + z_d \), where \( z_d \) is the DLA redshift, and includes seven pixels on each side (13 for O I–Si II). The width of 15 pixels, corresponding to a velocity width 1036 km s\(^{-1}\), is generally wide enough to include most of the absorption components in DLAs, as can be seen from the metal-line profiles in the stacked absorption spectrum of DLAs (see their fig. 10 Mas-Ribas et al. 2017), which have a dispersion of \( \sim 100 \) to 150 km s\(^{-1}\) (only moderately wider than the point spread function of the BOSS spectrograph wavelength resolution). However, the equivalent widths we measure are underestimated if there are large DLA redshift errors in the DLA catalogue we use, which shift the metal lines partly outside the measurement window. In addition, the O I–Si II blended line requires the wider window mentioned above because the two lines are resolved and are spread over a wider interval than the other blends. We will show below the impact on the mean equivalent widths of decreasing the measurement window width to 10 pixels or increasing it to 20 or 25 pixels (Fig. 2 and Section 5). For these cases, we also change the O I–Si II blend to 25, 29, and 31 pixels, respectively.

The windows for determining the continuum are both set to a width of 21 pixels, on the left and the right of the measurement window. We generally leave 1 pixel that is not used between the end of the measurement window and the start of each continuum window, with the exception of some lines for which the continuum window placed in this way would include another DLA metal line that would systematically lower the estimated continuum. These exceptions are listed in Table 2, where we give the space left between the centre of the line and the start of the two continuum windows on each side for these set of lines (for all other lines, the normal space left is 8.5 pixels, corresponding to half of the 15 pixel width of the measurement window plus one). The size of the two continuum windows remains fixed at 21 pixels for all these lines.

3.3 Excluded metal lines for each DLA

DLA metal lines often fall in regions of the spectrum where the equivalent width measurement is subject to large systematic errors, due to the presence of quasar emission lines or of sky lines from
Table 1. Metal lines and their principal characteristics used for this work. First column: Name of the metal line. Second column: Wavelength in Angstroms (from Table 2 of Prochaska et al. 2001). The first three lines are blends of two lines and we give the wavelength of the strongest line, except for O I–Si II (which are comparably strong) where we give the average of the two wavelengths. In the third blend, C II is a metastable state of C II. Third column: Equivalent width measured from the stacked spectrum (Mas-Ribas et al. 2017). Fourth column: Mean equivalent width $W$ calculated with our method as described in Section 3 (for spectra with continuum-to-noise ratio CNR > 2). Fifth column: Mean contribution $C_k$ of each line to the total metal strength in the DLAs where $W$ of the line is measured (see Section 4 for details). Sixth column: Fraction $x_k$ of the 34 050 DLAs in which each line is actually measured. Seventh column: Number of DLAs in which each line has been measured.

| Name                  | $\lambda$ [Å] | Stacked $W$ [Å] | $\bar{W}$ [Å] | $C_k$  | $x_k$  | $N_k$ |
|-----------------------|---------------|----------------|--------------|-------|-------|------|
| Si II–Fe II 1260      | 1260.42       | 0.623 ± 0.008  | 0.693 ± 0.010 | 0.58  | 0.18  | 6057 |
| O I–Si II 1303        | 1303.20       | 0.793 ± 0.012  | 0.818 ± 0.012 | 0.46  | 0.23  | 7807 |
| C II–C II 1334        | 1334.53       | 0.630 ± 0.007  | 0.596 ± 0.006 | 0.40  | 0.47  | 16 020 |
| Si II 1526            | 1526.71       | 0.443 ± 0.004  | 0.396 ± 0.005 | 0.28  | 0.71  | 21 440 |
| Fe II 1608            | 1608.45       | 0.228 ± 0.004  | 0.218 ± 0.004 | 0.11  | 0.71  | 24 013 |
| Al II 1670            | 1670.79       | 0.452 ± 0.005  | 0.41 ± 0.005  | 0.27  | 0.70  | 23 940 |
| Si II 1808            | 1808.01       | 0.059 ± 0.008  | 0.0575 ± 0.005 | 0.015 | 0.63  | 21 608 |
| Al II 1854            | 1854.72       | 0.117 ± 0.006  | 0.118 ± 0.006 | 0.031 | 0.60  | 20 317 |
| Al II 1862            | 1862.79       | 0.067 ± 0.006  | 0.095 ± 0.006 | 0.020 | 0.59  | 20 131 |
| Fe II 2344            | 2344.21       | 0.520 ± 0.014  | 0.404 ± 0.014 | 0.103 | 0.17  | 5732 |
| Fe II 2374            | 2374.46       | 0.282 ± 0.014  | 0.278 ± 0.018 | 0.034 | 0.15  | 5024 |
| Fe II 2382            | 2382.76       | 0.67 ± 0.03    | 0.623 ± 0.019 | 0.15  | 0.14  | 4907 |
| Fe II 2586            | 2586.65       | 0.46 ± 0.02    | 0.47 ± 0.03   | 0.06  | 0.11  | 3597 |
| Fe II 2600            | 2600.17       | 0.72 ± 0.02    | 0.61 ± 0.03   | 0.10  | 0.10  | 3458 |
| Mg II 2796            | 2796.35       | 1.15 ± 0.03    | 1.13 ± 0.05   | 0.18  | 0.047 | 1598 |
| Mg II 2803            | 2803.53       | 1.07 ± 0.03    | 1.02 ± 0.05   | 0.15  | 0.048 | 1629 |
| Mg II 2852            | 2852.06       | 0.23 ± 0.03    | 0.31 ± 0.04   | 0.021 | 0.040 | 1370 |

Table 2. Range around each metal absorption line at which the two continuum windows start, given in pixel numbers ($p$) and velocity interval ($s$). The two continuum windows of any of the two lines in each of the four pairs start a number $p$ of pixels to the right and to the left of the line centre. When measuring one line in any of these four pairs, the other line is avoided by leaving it in between the integrating and the continuum windows.

| Name                  | $\lambda$ [Å] | s [km s$^{-1}$] | p |
|-----------------------|---------------|----------------|---|
| Al III–Al III         | 1854.72–1862.79 | 2071.5         | 30 |
| Fe II–Fe II           | 2374.46–2382.76 | 1726.3         | 25 |
| Fe II–Fe II           | 2586.65–2600.17 | 2071.5         | 30 |
| Mg II–Mg II           | 2796.35–2803.53 | 1553.6         | 22.5 |

Table 3. Values for the linear regression fitted to the dependence of $W$ on log($N(D)$) in equation (6). The first two columns use values of $W$ from the stacking technique shown in Mas-Ribas et al. (2017), and the latter two columns use $W$ measured with our method, which are used for the analysis in this paper.

| Name                  | $a_s$ stacking | $b_s$ stacking | $a_k$ | $b_k$ |
|-----------------------|----------------|----------------|-------|-------|
| Si II–Fe II 1260      | 0.316          | 0.439          | 0.267 | 0.568 |
| O I–Si II 1303        | 0.591          | 0.482          | 0.651 | 0.489 |
| C II 1334             | 0.311          | 0.443          | 0.391 | 0.403 |
| Si II 1526            | 0.327          | 0.270          | 0.324 | 0.231 |
| Fe II 1608            | 0.263          | 0.089          | 0.287 | 0.0744 |
| Al II 1670            | 0.338          | 0.273          | 0.368 | 0.233 |
| Si II 1808            | 0.101          | 0.0041         | 0.110 | 0.00105 |
| Al II 1854            | 0.091          | 0.068          | 0.147 | 0.0478 |
| Al II 1862            | 0.052          | 0.038          | 0.0399 | 0.0759 |
| Fe II 2344            | 0.471          | 0.263          | 0.568 | 0.123 |
| Fe II 2374            | 0.374          | 0.085          | 0.337 | 0.104 |
| Fe II 2383            | 0.559          | 0.370          | 0.513 | 0.360 |
| Fe II 2586            | 0.531          | 0.179          | 0.574 | 0.122 |
| Fe II 2600            | 0.491          | 0.455          | 0.528 | 0.304 |
| Mg II 2796            | 0.595          | 0.841          | 1.008 | 0.583 |
| Mg II 2803            | 0.650          | 0.731          | 0.923 | 0.576 |
| Mg II 2853            | 0.217          | 0.092          | 0.208 | 0.135 |

We do not eliminate proximate DLAs, with a redshift close to that of the quasar. These systems may be affected by the quasar sky. These systems may be affected by the quasar sky. We discard the metal line whenever the central wavelength of any sky line is within the measurement window or in one of the two adjacent pixels, to avoid the increased systematic error on the equivalent width caused by skylines. Following previous analyses of BOSS quasar spectra, we use the set of 872 skylines.²

²https://github.com/igmhub/picca/blob/master/etc/dr14-line-sky-mask.txt
3.4 Calculation of the metal-line equivalent width and error

The quasar continuum around a metal absorption is estimated by doing a linear regression of the flux values \( f_i \) in every pixel \( i \) in the two continuum windows to the left and right of the measurement window, which contain a total of 42 pixels. First a linear regression\(^3\) is computed applying the inverse variance \( w_i = 1/\sigma_i^2 \) of the pixels \( i \) as weights, where \( \sigma_i \) is the noise. This yields a preliminary determination of a continuum \( C_{\lambda_k} \), from which we compute the preliminary transmission fraction \( F_{\lambda_k} = f_i/C_{\lambda_k} \). Next, we eliminate outliers in the continuum windows, which may result from highly noisy flux measurements in skylines. Note that DLA metal lines are discarded only when skylines are present in the measurement window, but not in the continuum window. We eliminate any pixels for which \( |F_{\lambda_k} - 1| > 4 \) as outliers. We repeat the determination of the continuum and we check again for outliers, until none are left. If this process leaves less than five pixels in any of the two continuum windows, the line is discarded and not measured for the DLA being analysed.

Then we repeat the determination of the continuum with non-weighted linear regression with the pixels that are left after elimination of outliers. The reason for not using weights in our final determination of the continuum is that the weights that minimize the impact of more noisy pixels also introduce a bias that systematically lowers the continuum, because pixels with less flux are assigned lower noise owing to the photon noise contribution calculated by the BOSS pipeline. This systematic effect causes equivalent widths to be underestimated due to the lowering of the continuum, by an amount that increases with spectral noise and introduces an unwanted artificial correlation of our metal strength parameter and its error, so we decided not to use weights to obtain the final continuum. With the final continuum estimation \( C_m \), we compute the transmission fraction \( F_i = f_i/C_m \). On the other hand, not using the weights for our final continuum implies that outliers caused by skylines or other large errors can strongly distort the determination of the continuum and they need to be eliminated.

We checked that different ways of computing the linear regression, weighted or unweighted, and varying the outlier condition to \( |F_{\lambda_k} - 1| > (2, 3, 5) \), have only percent effects on the mean equivalent widths.

Fig. 1 illustrates the procedure in an example of an Al\(\text{II} \) 1670 absorption line. The red vertical bars show the limits of the measurement window, and the blue bars are the outer limits of the continuum windows. Green points show the values of the flux in the upper panel. The unweighted linear regression is the magenta line. The lower panel shows the result of dividing the flux values by this linear regression, giving the transmitted fraction \( F_i \). The error on this transmitted fraction, \( \sigma_{F_i} \), is computed by dividing \( \sigma_i \) by the same linear regression.

Finally, we obtain the rest-frame equivalent width of a line \( k \) simply by summing the absorbed fraction over all the \( N_m \) pixels of the measurement window,

\[
W_k = \sum_{i=1}^{N_m} (1 - F_i) \lambda_k p_0, \tag{1}
\]

where \( \lambda_k \) is the central rest-frame wavelength of the metal line being measured, and \( p_0 = 10^{-4} \times \log (10) = 2.303 \times 10^{-4} \) is the width of the BOSS pixel in log \( \lambda \). Note that some pixels in the measurement window may have \( F_i > 1 \), and in fact, some equivalent widths we compute are negative because of the noise. Still, as explained before, they need to be included to have unbiased mean properties. The error on this equivalent width is computed as

\[
\epsilon_k = \sqrt{\sum_{i=1}^{N_m} \sigma_{F_i}^2 / \lambda_k}, \tag{2}
\]

where \( \sigma_{F_i}^2 = \sigma_i^2 / C_i^2 \), of the pixels in the two continuum windows that are used to determine the continuum linear regression. We use the noise in the continuum windows instead of the measurement window to estimate the error \( \epsilon_k \) because of the correlation of \( \sigma_{F_i} \) and \( F_i \) discussed earlier, which can induce an artificial correlation of the estimated value of \( W \) and its error. We note that the error \( \epsilon_k \) accounts only for pixel noise, assuming that it is uncorrelated in all the pixels, but excludes any systematic errors due to the continuum fitting.

---

\(^3\)http://www.statsmodels.org/stable/generated/statsmodels.regression.line

ar_model.WLS.html
We have tested our method by inserting mock lines in empty parts of a spectrum and successfully recovering their equivalent widths within the estimated uncertainty. We also checked that mean equivalent widths estimated with our method in spectral intervals where no absorption lines are inserted are consistent with zero.

As a final test, in Fig. 2 we compare the mean equivalent width we derive for each of the 17 lines in Table 1 with the value obtained from the stacked absorption spectrum presented in Mas-Ribas et al. (2017), and we also check the sensitivity of this mean equivalent width to the size of our measurement window by varying it from our standard width of 15 pixels to 10 and to 20 pixels. The mean values generally agree. The largest discrepancy occurs for the Fe II 2344 line, for which our method yields an equivalent width ∼20 per cent smaller than in Mas-Ribas et al. (2017). Our derived $W$ generally increases with the measurement window width, as expected because of the effect of redshift errors. The overall agreement is relatively good for a 15 pixel width, which we choose as the optimal and standard one from this plot.

3.5 Comparison with high-resolution spectra

Our method to measure metal-line equivalent widths is intended to be a fast and simple automatic procedure, and is often affected by systematic errors due to superposition with metal lines from other absorbers and deviations of the continuum from our linear interpolation model. While these systematic errors can strongly perturb equivalent widths of individual systems, their effects are less severe on the sample as a whole, where they simply add to the noise and do not cause large variations of the statistical distribution of the metal strengths we derive, as we shall see below.

We compare in this subsection the equivalent widths we measure in a few examples of the low-resolution BOSS spectra containing DLAs, with the values measured in spectra of the same quasars with much higher signal-to-noise and resolution, as a test of the most common errors that occur. For this we use the public data release of the Keck Observatory Database of Ionized Absorption toward Quasars (KODIAQ, O’Meara et al. 2015, 2017).4 We searched for SDSS spectral counterparts of the KODIAQ data base. This presents some difficulties because the KODIAQ spectra are usually shorter than SDSS spectra, with some parts of the spectrum missing. Moreover, some of the absorption lines are near emission lines of the quasar, which we exclude in our analysis of the SDSS spectra, leaving few lines that can be directly compared for each DLA.

We investigated eight spectra of DLAs that we found to be present in both catalogues (KODIAQ and DR12-DLA). We required the KODIAQ spectra to be longer than 5000 Å, the quasar to be at redshift lower than 3.1, and the DLA to have a column density $\log(N_{\text{HI}}) < 21.0$. Out of the eight spectra, six have between three and seven metal lines that can have their equivalent widths compared, with a total of 28 lines. Of these 28 lines, we find that 21 have equivalent widths determined from SDSS and KODIAQ that are compatible within a 3σ error. Notice that these SDSS quasars have unusually high CNR (16–78), which is not representative of SDSS average spectra because quasars in the KODIAQ sample are selected to be bright. Although the fraction of 25 per cent of lines showing errors above 3σ indicates a common presence of the systematics we have mentioned, the errors in these spectra are smaller than in the majority of quasars, which are fainter.

We find that two reasons generally account for the discrepancies in the seven lines with substantial disagreement. In four cases a metal line from another absorber is present in the measurement and/or continuum windows, affecting the inferred equivalent width. As an example, Fig. 3 presents an extreme case of contamination of an Si II 1526 line in the upper and middle panels. The upper panel shows the transmitted fraction in the KODIAQ spectrum (where the continuum was determined by the usual spline fitting method of

4https://koa.ipac.caltech.edu/workspace/TMP_AqLrga_13996/kodiaq13996.html
Metal-line strength indicator for DLA systems

Figure 3. Examples of metal lines with discrepant equivalent widths when measured in KODIAQ and SDSS spectra. The upper and middle panels show the contaminating effect of metal lines from unrelated absorbers, and the bottom panel illustrates the effect of continuum deviations from a linear fit. Upper panel: Normalized flux near an Si II 1526 line in a KODIAQ spectrum (KODIAQ J082619+314848 Wolfe-Dec-27-2006). Middle panel: SDSS spectrum of the same quasar spectrum and wavelength interval (SDSS 55589-4446-0120). Bottom panel: Spectrum around an Al ii 1670 line (SDSS 56426-6756-0814). Black lines are the flux in SDSS spectra, and the derived transmitted fraction in the KODIAQ spectrum. Grey is the noise level of the spectra. Green, red, and blue vertical lines mark the expected line centre, measurement window edges, and continuum windows edges, respectively. Magenta line in the SDSS spectra is the fitted linear continua.

4 DEFINITION OF THE METAL STRENGTH

We now define a metal strength for each DLA intended to provide a weighted average of the strength of the observable associated metal lines with an optimal signal-to-noise, which can be used in spectra with high noise where individual lines are generally barely detectable. If the metal lines that have been measured for a given DLA are labelled by the index \( k \), with equivalent widths \( W_k \) and error \( \epsilon_k \) from equations (1) and (2), the metal strength is defined as

\[
S = \frac{\sum_k \left( \frac{W_k}{\epsilon_k} \right)^2 \cdot \frac{1}{W_k}}{\sum_k \left( \frac{W_k}{\epsilon_k} \right)^2},
\]

where \( \bar{W} \) is the mean equivalent width of line \( k \) over all the DLAs in the sample for which it is measured. This sum excludes any lines that were discarded for a given DLA, so in general it has fewer than
17 lines. The metal strength is equal to unity if all the lines of a DLA have the mean value $\overline{W}_k$, and is in general a measure of the ratio of the equivalent widths to their mean value. Each line is weighted by $(\overline{W}_k/\epsilon_k)^2$, which is the expected squared signal-to-noise ratio of the measurement if the line has an equivalent width equal to the mean. The error of this metal strength is computed from the errors of each individual line as

$$
\epsilon_k = \left( \sum_k \left( \frac{W_k}{\epsilon_k} \right)^2 \right)^{-1/2}.
\quad (4)
$$

This metal strength parameter reflects how strong the metal lines of a DLA are, and is chosen as a quantity that can be measured with an optimal signal-to-noise ratio from spectra similar to those in the BOSS survey. The metal strength is expected to increase both with the metal column density and the velocity dispersion. A larger velocity dispersion implies a wider line, which is less saturated at a fixed metal column density, thereby yielding a larger equivalent width.

The metal strength depends on the mean values $\overline{W}_k$, which may depend on the DLA sample that is chosen. We fix these mean values throughout this work to the ones we obtain for our entire sample, listed in the fourth column of Table 1. These mean values should reflect an intrinsic property of DLAs with the frequency at which they are intercepted along random lines of sight, if the sample selection is done independently of the metal lines, as it is in our case.

To visualize the metal lines that are most important to determine $S$, we define the mean contribution $C_k$ of metal line $k$ as follows:

$$
C_k = \frac{1}{N_k} \sum_{i=0}^{N_k} \left( \frac{W_{k,i}}{\epsilon_{k,i}} \right)^2 / \sum_j \left( \frac{W_{j,i}}{\epsilon_{j,i}} \right)^2,
\quad (5)
$$

where the sum over $i$ includes all the $N_k$ DLAs in which the metal line $k$ has been measured, $\epsilon_{k,i}$ is the error of the metal line $j$ in the DLA $i$, and the sum in the denominator adds the weights of all metal lines $j$ that have been measured in the DLA $i$. This means that $C_k$ is the mean contribution to the metal strength $S$ from line $k$, restricted only to the fraction $x_k$ of DLAs in which this metal line has been measured. The average contribution to all the DLA sample is the product $C_k x_k$, which adds to unity when summed over the 17 metal lines. These quantities are listed in Table 1, and shown also in Fig. 4.

The lines that are important for most DLAs, with the highest value of $C_k x_k$, are the six lines with shortest wavelength (Si II 1526, C II 1334, Al II 1670, O I–Si II 1303, Si II–Fe II 1260, and Fe II 1608 ordered by their contribution $C_k x_k$). A few of the longer wavelength ones are important only in a small fraction of DLAs. The fraction $x_k$ is small for the shortest wavelength line, the Si II–Fe II blend, because any lines that are close to the quasar Ly $\alpha$ emission line or that fall in the Ly $\alpha$ forest region are excluded. Then, $x_k$ increases with increasing line wavelength until it starts decreasing rapidly above $\sim 2000 \AA$ due to the exclusion of metal lines when a sky line falls within the measurement window, and the large abundance of sky lines in the red part of the spectrum.

The distribution of the number of lines available for the determination of $S$ in equation 3 (which satisfy the criteria specified in Section 3.3) is shown in Fig. 5 as the grey solid line. The average number of contributing lines is between six and seven. The black dashed line is the distribution of the number of measured lines among only the eight lines with $C_k > 0.15$ in Table 1. In most DLAs, there are only two or three of these highly contributing lines that can be used. Fig. 5 also shows the fraction of DLAs for which no line is measured and no estimate of $S$ is therefore available. These DLAs are excluded and not used in the rest of this work.

### 4.1 Correction for the dependence of $\overline{W}_k$ on $N_{HI}$

We expect the mean equivalent width of any metal line to increase with $N_{HI}$, because for a fixed metal abundance the metal column density should be proportional to the hydrogen column density, and metal-line equivalent widths should increase with metal column densities for fixed internal velocity distributions. This dependence was confirmed in Mas-Ribas et al. (2017) from stacked spectra of DLAs in different bins of $N_{HI}$ (see their Table 3). Our goal is to classify DLAs depending on a parameter that reflects only their metal content and velocity dispersion (which affects the degree of saturation of the line), but not on $N_{HI}$.

---

**Figure 4.** Dark grey histogram: Mean contribution $C_k$ of each line to the metal strength of DLAs in which the line is measured. Light grey histogram: Mean contribution to all the DLA sample, equal to $C_k x_k$, where $x_k$ is the fraction of DLAs in which the line is measured. The sum of all $C_k x_k$ adds to unity. Lines are ordered in the histogram by their value of $C_k x_k$, with the most contributing lines on the left.

**Figure 5.** Grey solid line: Distribution of the number of lines measured for a DLA among all 17 lines in Table 1. Black dashed line: Distribution of the number of lines used among the eight lines with a mean contribution $C_k > 0.15$ (as defined below in equation 5). When the number of lines is zero, $S$ cannot be measured and the DLA is excluded.
For this purpose, we define the $N_{\text{HI}}$-corrected metal strength, $S_{\text{HI}}$, in the same way as $S$ in equation (3), but replacing the constant values of $W_k$ by mean equivalent widths that depend linearly on $\log(N_{\text{HI}})$:

$$W_k(N_{\text{HI}}) = a_k(\log(N_{\text{HI}}) - 20) + b_k.$$ (6)

We fit the parameters $a_k$ and $b_k$ by dividing the DLA sample into the following five intervals of $\log(N_{\text{HI}})$: [20.0, 20.20), [20.2, 20.4), [20.4, 20.6), [20.6, 21.0), and [21.0, 22.2]. Then, for each metal line, we compute the mean value of $W_k$ and $\log(N_{\text{HI}})$ in the five intervals, and we make an unweighted linear regression of these five points.

The result is shown in Table 3, in columns 4 and 5. These linear regressions to the mean equivalent width as a function of $N_{\text{HI}}$ were also obtained in Mas-Ribas et al. (2017) from stacked spectra, and their results are shown in columns 2 and 3. The values are in general similar, and differences can be attributed to the different methods of determining the quasar continuum and weighting the DLAs.

The mean equivalent widths calculated from equation (6) for the column density of each DLA are then used to compute the $N_{\text{HI}}$-corrected metal strength $S_{\text{HI}}$ with the same equation (3) as for $S$.

The average of the corrected $S_{\text{HI}}$ is by construction independent of $N_{\text{HI}}$, and provides a way of measuring a dependence of any other property of DLAs with the metal strength without being affected by the impact of any dependence on $N_{\text{HI}}$. This will be useful to interpret the dependence of the DLA bias factor on the metal strength in Pérez-Rafols et al. (2018a), where it will be shown that the dependence on $S_{\text{HI}}$ is stronger than the dependence on $S$.

5 RESULTS

We now present the distribution of the metal strength parameter $S$ and its variation with column density and redshift. We also examine if the metal strength distribution has any dependence on the CNR in the Lyα forest, and the computed error $\epsilon_S$, as a test of any possible variation of the purity of the DLA catalogue that might affect derived statistical properties of DLAs as a function of $S$. As the main product of this work, we present the catalogue of metal-line equivalent widths and the metal strength parameter with its error, both $N_{\text{HI}}$-corrected and uncorrected for its $N_{\text{HI}}$ dependence, for all the DLAs in DR12-DLA.

5.1 The metal strength distribution

The distribution of the metal strength $S$ is shown in the top panel of Fig. 6, including only DLAs with an error $\epsilon_S < 0.5$. The excluded DLAs with larger error are systems in low signal-to-noise spectra, and systems where some of the best lines for measuring $S$ were excluded according to the criteria stated in Section 3.3. We exclude DLAs with large errors to better see their intrinsic properties when they are less affected by measurement errors. The distribution of $S$ we measure is the convolution of the true distribution with the error distribution. Systems with negative $S$ should be a consequence of the error, and this is consistent with the shape of the curves at $S < 0$, except at $S < -1$ where the extended tail indicates the presence of non-Gaussian errors that are likely due to systematics in the continuum fitting. The $S$ distribution becomes wider with the measurement window width owing to increased errors. The $N_{\text{HI}}$ correction practically does not affect this distribution.

The bottom panel shows the distribution of $\epsilon_S$. The error increases with the measurement window simply because of the increased number of pixels in equation (2). The majority of DLAs have errors $\epsilon_S < 0.5$, and including the ones with larger errors only distorts the distribution of $S$ in the top panel further from the true one. The subset of DLAs with $\epsilon_S < 0.5$ can be safely divided into smaller subsets corresponding to intervals in $S$ to measure other DLA properties (like the bias factor or the mean equivalent width of other metal lines) as a function of $S$, with a tolerable level of mixing of subsets due to observational error.

Results in the rest of the paper are shown only for DLAs with $\epsilon_S < 0.5$, unless otherwise specified.

5.2 Metal strength dependence on $N_{\text{HI}}$ and redshift

Fig. 7 shows the mean value of $S$ in several column density intervals as the green points. As expected, $S$ increases with $N_{\text{HI}}$ simply due to the increasing metal column densities. After applying the correction discussed in Section 4.1 (blue points), the mean value of $S_{\text{HI}}$ becomes indeed nearly constant. Note that the mean value of $S$ should be unity for the whole DLA sample; when we eliminate DLAs with $\epsilon_S > 0.5$, the mean value of $S$ increases slightly above unity because of a decrease of the mean $S$ with $\epsilon_S$ that will be discussed below.
A. Arinyo-i-Prats et al.

Figure 7. Correction to reduce the dependence of \( S \) with the column density. Green points show mean metal strength for different column density intervals (the intervals are indicated by the horizontal error bars). Blue points show the mean of the H I–N\(_{\text{HI}}\) metal strength correction \( S_{\text{HI}} \), as defined in Section 4.1.

Figure 8. Evolution of the mean metal strength \( S \) (green) and \( S_{\text{HI}} \) (blue) with redshift intervals (intervals are indicated by the horizontal error bars).

The evolution of the mean metal strength with redshift is shown in Fig. 8. The mean \( S \) is roughly constant at \( z < 3.4 \), and then drops rapidly at higher redshift. Some of this decline should be due to the well-known decrease of the average DLA metallicity with redshift (e.g. Kulkarni et al. 2005; Rafelski et al. 2012; Neeleman et al. 2013). A decreasing velocity dispersion with redshift may also contribute because the mean value of \( S \) increases with velocity dispersion at a fixed metal column density owing to the effects of line saturation. However, the dependence of the purity of the catalogue on redshift may be the main effect causing the rapid drop at \( z > 3.4 \). Only \( \sim 10 \) per cent of our DLA sample is at \( z > 3.5 \) (see Fig. 1 of Mas-Ribas et al. 2017), and the difficulty in detecting DLAs at high redshift because of the increased mean absorption of the Ly \( \alpha \) forest, which is likely decreasing the purity of the catalogue. More complete studies revising the DLA detection method will be needed to resolve this issue.

5.3 Dependence of the \( S \) distribution on \( \varepsilon_S \) and the purity of the DLA catalogue

Our main goal in defining and measuring the metal strength \( S \) for individual DLAs is to be able to measure average quantities of the DLAs as a function of \( S \). An important consideration to make is the purity of the catalogue. For example, false DLAs arising purely from spectral noise should have a mean value of \( S \) and bias factor equal to zero, and this can induce a spurious increase of the bias factor with \( S \). Other false DLAs may be dense Ly \( \alpha \) forest regions with a column density and mean \( S \) much lower than DLAs, and a different bias factor.

A first test we can do to investigate possible variations of the purity with \( S \) is to check if the mean value of \( S \) depends on the CNR ratio of the Ly \( \alpha \) forest region, where the DLA absorbers are searched. We expect the purity of the catalogue to increase with CNR. This is shown in the top panel of Fig. 9, both for the uncorrected and \( N_{\text{HI}}\)-corrected cases. The mean \( S \) is consistent with a constant value for CNR > 3, suggesting that as long as we restrict any analysis to this minimum CNR, there should not be strong effects of the catalogue impurity caused by spectral noise. The bottom panel shows that, as expected, the error \( \varepsilon_S \) decreases with CNR (since they both reflect the noise amplitude in the spectrum, although on different wavelength ranges). In the uncorrected and corrected cases, the sample is restricted to DLAs with \( \varepsilon_S < 0.5 \), and \( \varepsilon_{\text{HI}} < 0.5 \), respectively.

A second test we do is to examine if the \( S \) distribution varies with the error \( \varepsilon_S \). In principle, the true distribution of \( S \) should not depend on \( \varepsilon_S \), which is determined using only the flux measured in spectral pixels in the continuum windows and is independent of the
Metal-line strength indicator for DLA systems

Figure 10. Normalized distributions of corrected and uncorrected metal strength for different cuts in its error. Upper panel: Distributions of the DLA metal strength for three intervals of the error $\varepsilon_S$ for systems with $\varepsilon_S < 0.5$. Blue line lower error interval, green dashed line mid-error interval, red semi-dashed line upper error interval. Lower panel: Same as upper panel but for the $N_{HI}$-corrected metal strength $S_{HI}$.

However, in reality there is a fairly strong dependence as we show in Fig. 10, where we have divided DLAs into three intervals of $\varepsilon_S$ ([0.0, 0.16), [0.16, 0.33), and [0.33, 0.5]). If the true $S$ distribution stayed constant, we would expect only an increased width of the distribution as $\varepsilon_S$ increases because of the convolution with the error, but we also observe a decrease of the mean value of $S$ with increasing $\varepsilon_S$. The effect is nearly the same for the $N_{HI}$-corrected $S_{HI}$, as seen in the bottom panel.

This effect is more clearly seen by examining the mean value of $S$ as a function of $\varepsilon_S$, shown in Fig. 11. Our standard upper limit $\varepsilon_S < 0.5$ is shown as the dashed vertical bar, and we divide the range $0 < \varepsilon_S < 0.5$ into five intervals. Blue points in the top panel show that in our standard case, the mean value of $S$ declines from 1.2 to 0.8 as $\varepsilon_S$ increases up to its upper threshold of 0.5. Eliminating DLAs with CNR > 3 diminishes the dependence of $S$ on $\varepsilon_S$ only slightly, as shown by the red points in the top panel. This is consistent with the small decrease of $S$ for CNR < 3 we saw in Fig. 9, which cannot explain the much larger effect of $\varepsilon_S$ on the mean $S$. In other words, the degree of catalogue impurity caused by spectral noise is relatively small, and the large variation of $S$ with $\varepsilon_S$ must be caused by a different systematic variation of the selected absorption systems with $\varepsilon_S$.

The green points in the top panel in Fig. 11 show the visually identified DLAs. This visual identification (Pâris et al. 2012) has a larger impact reducing the dependence of $S$ on $\varepsilon_S$. The reason is probably that the visual identification was influenced by the presence of metal lines: if metal lines are visually identified, an absorber is more likely to be flagged as DLA.

The middle panel of Fig. 11 shows the change in the dependence of the mean $S$ with $\varepsilon_S$ when we restrict the sample to DLAs with $N_{HI} \geq 10^{20.3} \text{cm}^{-2}$ (green) and to DLAs with $z < 3.2$ (red). Lower panel: Standard sample (blue) and $N_{HI}$-corrected (green) metal strength. The shaded (grey) area covers the region where $\varepsilon_S \geq 0.5$, these systems are excluded from our analysis.
DLAs, and a more appreciable change when the low column density DLAs are eliminated. However, the strong decline of the mean $S$ with $\epsilon_S$ is present in all these subsets, with only a small reduction of this dependence. The bottom panel compares the $N_{\text{HI}}$-corrected $S_{\text{HI}}$ dependence on $\epsilon_{\text{SHI}}$ (green), with the previous standard uncorrected case. The correction tends to slightly decrease this dependence.

As we now show, the principal cause of the decline of the mean $S$ with $\epsilon_S$ is that the DLA redshifts are subject to large errors when no metal lines are identified. The Ly $\alpha$ profile of the DLA is often contaminated by neighbouring Ly $\alpha$ forest absorption, which may shift the best fit of the DLA redshift. Metal lines, which are narrow, usually pin down the redshift much more accurately. When there is a large redshift error, the metal lines may be partly shifted outside our measurement window, reducing the equivalent width that we compute. This is shown in Fig. 12, where the mean $S$ dependence on $\epsilon_S$ is plotted for four widths of the measurement window: 10, 15, 20, and 25 pixels. The drop in the mean $S$ with $\epsilon_S$ is gradually reduced with the measurement window width. For a 25 pixel width, this drop is reduced to only $\sim 10$ per cent for $\epsilon_S < 0.5$. The redshift error required to shift a metal line by half of the 25 pixel width is $\sim 850 \text{ km s}^{-1}$, so from the difference in the results for the 20 and 25 pixels, we conclude that redshift errors can reach this value for $\sim 10$ per cent of the DLAs. The remaining dependence can be explained by a small fraction of even larger redshift errors and the increase of the mean $S$ with CNR seen in Fig. 9. We have checked that this reduction of the $S - \epsilon_S$ dependence when the window is increased is not explained by other effects, such as pollution from lines unrelated to the DLA entering the measurement window.

![Figure 12. Mean metal strength intervals as a function of its error $\epsilon_S$, for four different values of the width of the measurement window: 10, 15, 20, and 25 pixels (blue, green, red, and turquoise, respectively). The shaded (grey) area covers the region where $\epsilon_S \geq 0.5$, this systems are excluded from our analysis.](https://academic.oup.com/mnras/article-abstract/481/3/3921/5088371)

### 5.4 Catalogue of metal lines and DLA metal strength

We have produced a new catalogue of the DLAs in DR12-DLA, including the metal-line equivalent widths computed with the method described in Section 3. An example of the entries in this catalogue is presented for three DLAs in Table 4: we list the quasar spectrum identification and quasar redshift, the CNR, the DLA redshift and...
column density, the derived value of $S$ and its error $\epsilon_S$, the same values corrected for $N_{\text{HI}}$ as explained in Section 4.1, and the equivalent width and errors of each of the 17 metal lines. When lines are not measured for the reasons described in Section 3.3, the entries for $W$ and its error are set to zero. We make the catalogue publicly available (with metal equivalent widths or not) which can be found on the GitHub repository https://github.com/andreuandreu/Catalogue_Metal_Strength_DLAs_SDSS-BOSS_DR12.

6 CONCLUSIONS

We have presented a way to classify DLAs according to the strength of their metal lines. A total of 17 metal lines are used to obtain a metal strength parameter that is optimized to be measured with the best possible accuracy. This allows DLAs to be classified from spectra in surveys with large numbers of objects, where the spectral signal-to-noise is often poor. A catalogue of this parameter and several metal equivalent widths in the DR12-DLA catalogue previously constructed from the BOSS survey using the method of Noterdaeme et al. (2009) is made publicly available.

Our main motivation to present this catalogue is to allow for studies of the dependence of any property of DLAs as a function of the metal strength $S$, which after correcting for its dependence on $N_{\text{HI}}$ to obtain the new parameter $S_{\text{HI}}$, should be a function of the metal abundance and velocity dispersion of the absorbing gas. The first quantity we have measured as a function of $S$ is the bias factor, which is presented in Pérez-Rafols et al. (2018a), and is the first detection of a dependence of the bias factor on any DLA property.

In the future we plan to measure also the mean stacked absorption spectrum of the high-ionization DLA metal lines with the technique of Mas-Ribas et al. (2017) for different values of $S$ or $S_{\text{HI}}$.

Our catalogue presents individual equivalent width measurements, allowing the user to build other combinations of metal-line strengths. One of the applications should be to investigate correlations among the strength of equivalent widths of different metal lines. Although a fraction of these equivalent width measurements may be affected by contamination of absorbers not associated with the DLA, and despite the large noise for individual DLAs, the large size of the sample can allow for detailed statistical studies.

An important limitation in using this catalogue is its imperfect purity. Some of the DLAs may arise due to spectral noise, and some others may be absorbers of low column density with an absorption profile that arises from a velocity distribution of absorbing gas in the Ly $\alpha$ forest but is consistent with a DLA. This can introduce spurious variations of any property we wish to measure as a function of $S$, if the catalogue purity varies with $S$. The fact that the mean $S$ as a function of CNR is constant within 5 per cent at CNR > 3 suggests that spectral noise does not introduce substantial purity variations at CNR > 3. We measure a dependence of $S$ on $\epsilon_S$ which we find is mostly due to large redshift errors that are not corrected when no metal lines are significantly detected, which cause the metal lines to be partially shifted out of our measurement window. The impact this may have when measuring the dependence of any DLA property on the $S$ parameter needs to be borne in mind in future studies.

ACKNOWLEDGEMENTS

We would like to thank George Becker for invaluable guidance and help and Andreu Font-Ribera and Mat Pieri for useful discussions. This work has been supported in part by Spanish grants AYA-2012-33938 and AYA 2015-71091-P.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is http://www.sdss3.org/.

SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, University of Cambridge, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

IPR was supported by the A=MI`DEX project (ANR-11-IDEX-0001-02) funded by the ‘Investissements d’Avenir’ French Government program, managed by the French National Research Agency (ANR), and by ANR under contract ANR-14-ACHN-0021.

REFERENCES

Alam S. et al., 2015, ApJS, 219, 12
Barnes L. A., Haehnelt M. G., 2014, MNRAS, 440, 2313
Berg T. A. M., Ellison S. L., Prochaska J. X., Venn K. A., Dessauges-Zavadsky M., 2015, MNRAS, 452, 4326
Bird S., Haehnelt M., Neeleman M., Genel S., Vogelsberger M., Hernquist L., 2015, MNRAS, 447, 1834
Cooke R. J., Pettini M., Jorgenson R. A., 2015, ApJ, 800, 12
Dawson K. S. et al., 2013, AJ, 145, 10
Eisenstein D. J. et al., 2011, AJ, 142, 72
Ellison S. L., Songaila A., Schaye J., Pettini M., 2000, AJ, 120, 1175
Ellison S. L., Prochaska J. X., Hennawi J., Lopez S., Usher C., Wolfe A. M., Russell D. M., Beam C. R., 2010, MNRAS, 406, 1435
Font-Ribera A. et al., 2012, J. Cosmol. Astropart. Phys., 11, 59
Fox A. J., Ledoux C., Petitjean P., Srianand R., 2007, A&A, 473, 791
Gunn J. E. et al., 1998, AJ, 116, 3040
Gunn J. E. et al., 2006, AJ, 131, 2332
Kulkarni V. P., Fall S. M., Laureosch J. T., York D. G., Welty D. E., Khare P., Truran J. W., 2005, ApJ, 618, 68
Mas-Ribas L. et al., 2017, ApJ, 846, 4
Neeleman M., Wolfe A. M., Prochaska J. X., Rafelski M., 2013, ApJ, 769, 54
Noterdaeme P., Petitjean P., Ledoux C., Srianand R., 2009, A&A, 505, 1087
Noterdaeme P. et al., 2012a, A&A, 547, L1
Noterdaeme P. et al., 2012b, A&A, 547, L1
O’Meara J. M. et al., 2015, AJ, 150, 111
O’Meara J. M., Lehner N., Howk J. C., Prochaska J. X., Fox A. J., Peeples M. S., Tumlinson J., O’Shea B. W., 2017, AJ, 154, 114
Paris I. et al., 2012, A&A, 548, A66
Paris I. et al., 2017, A&A, 597, A79
Pérez-Rafols I., Miralda-Escudé J., Arinyo-i-Prats A., Font-Ribera A., Mas-Ribas L., 2018a, MNRAS, 480, 4702
Pérez-Rafols I. et al., 2018b, MNRAS, 473, 3019
Prochaska J. X., Wolfe A. M., 1997, ApJ, 487, 73
Prochaska J. X., Wolfe A. M., 1998, ApJ, 507, 113
Prochaska J. X., Wolfe A. M., 2002, ApJ, 566, 68
Prochaska J. X. et al., 2001, ApJS, 137, 21
Prochaska J. X., Hennawi J. F., Herbert-Fort S., 2008, ApJ, 675, 1002

MNRAS 481, 3921–3934 (2018)
Rafelski M., Wolfe A. M., Prochaska J. X., Neeleman M., Mendez A. J., 2012, ApJ, 755, 89
Rafelski M., Neeleman M., Fumagalli M., Wolfe A. M., Prochaska J. X., 2014, ApJ, 782, L29
Ross N. P. et al., 2012, ApJS, 199, 3
Shapley A. E., 2011, ARAA, 49, 525
Skúladóttir Á., Salvadori S., Pettini M., Tolstoy E., Hill V., 2018, A&A, 615, A137
Smee S. A. et al., 2013, AJ, 146, 32
Tumlinson J., Peeples M. S., Werk J. K., 2017, ARA&A, 55, 389
Wolfe A. M., 1998, in Zaritsky D., ed., ASP Conf. Ser. Vol. 136, Galactic Halos, Astron. Soc. Pac., San Francisco. p. 159
Wolfe A. M., Turnshek D. A., Smith H. E., Cohen R. D., 1986, ApJS, 61, 249
Wolfe A. M., Gawiser E., Prochaska J. X., 2005, ARA&A, 43, 861

SUPPORTING INFORMATION

Supplementary data are available at MNRAS online.

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

This paper has been typeset from a \TeX/\LaTeX file prepared by the author.