The cosmological bias factor of damped Lyman alpha systems: dependence on metal line strength

Ignasi Pérez-Ràfols,[1]★ Jordi Miralda-Escudé,[2,3] Andreu Arinyo-i-Prats [iD],[3] Andreu Font-Ribera[4] and Lluís Mas-Ribas[5]

[1]*Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France*
[2]*Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain*
[3]*Institut de Ciències del Cosmos, Universitat de Barcelona/IEEC, Barcelona, E-08028 Catalonia, Spain*
[4]*Department of Physics and Astronomy, University College London, Gower Street, London, UK*
[5]*Institute of Theoretical Astrophysics, University of Oslo, PO Box 1029 Blindern, NO-0315 Oslo, Norway*

ABSTRACT

We measure the cosmological bias factor of damped Ly α absorbers (DLAs) from their cross-correlation with the Ly α forest absorption, as a function of the DLA metal strength, defined from an average of equivalent widths of the strongest detectable low-ionization metal lines. A clear increase of the bias factor with metal strength is detected, as expected from a relation of metallicity and velocity dispersion with host halo mass. The relation is stronger after the metal strength is corrected for the H I column density, to make it more related to metallicity instead of metal column density. After correcting for the effects of measurement errors of the metal strength parameter, we find that the bias factor of DLAs with the weakest metal lines is close to unity, consistent with an origin in dwarf galaxies with host halo masses $\sim 10^{10}\,\mathrm{M}_\odot$, whereas the most metal rich DLAs have a bias factor as large as $b_{\mathrm{DLA}} \sim 3$, indicative of massive galaxies or galaxy groups in host haloes with masses $\sim 10^{12}\,\mathrm{M}_\odot$. Our result confirms the physical origin of the relation of bias factors measured from cross-correlation studies to the host haloes of the absorbers.

Key words: intergalactic medium – cosmological parameters – large-scale structure of Universe – cosmology: observations.

1 INTRODUCTION

The Ly α forest is a fluctuating absorption seen bluewards of the Ly α rest-frame wavelength of the source caused by intergalactic hydrogen. Generally, regions of higher density give rise to absorption features of higher hydrogen column density. When the column density is as high as $N_{\mathrm{H\,I}} \geq 2 \times 10^{20}\,\mathrm{cm}^{-2}$ the hydrogen becomes self-shielded against the external cosmic ionizing background, and the gas is mostly in atomic form. These systems are called Damped Ly α Absorbers (DLAs) (Wolfe et al. 1986), and their damped profiles are measurable even in low-resolution and low-signal-to-noise spectra, providing a robust method to measure their column densities. The contribution of these systems to the cosmic density of atomic gas is $\Omega_{\mathrm{DLA}} \simeq 10^{-3}$ at redshifts $2 < z < 3$ (Péroux et al. 2003; Prochaska, Herbert-Fort & Wolfe 2005; Noterdaeme et al. 2009, 2012; Prochaska & Wolfe 2009; Zafar et al. 2013; Crighton et al. 2015; Padmanabhan, Choudhury & Refregier 2016), corresponding to ~ 2 per cent of all the baryons in the Universe.

Absorbers of this high column density naturally arise in haloes where gas cools and forms dense clouds, and are therefore crucial to understand the galaxy formation process from gas that is accreted into haloes and expelled in galactic winds.

Even though the contribution of the DLAs to the matter density is well understood, the sizes of galaxies and the masses of haloes hosting the absorbing gas cannot be inferred from the properties of the absorption lines, and it was widely believed until recently that the majority of DLAs were hosted in dwarf galaxies and low-mass haloes. One of the avenues to determine the characteristic masses of DLA host haloes is to use the cross-correlation of DLAs with other tracers of large-scale structure to measure their cosmological bias factor. The standard cold dark matter (ΛCDM) model of structure formation predicts the bias as a function of halo mass: the larger the halo mass, the larger the bias factor. This has been exploited by means of the cross-correlation with the Ly α forest transmission fluctuations by Font-Ribera et al. (2012) and Pérez-Ràfols et al. (2018). The most recent value of the bias factor obtained from the analysis of the final twelfth Data Release (DR12) of the Baryon Oscillations Spectroscopic Survey (BOSS) survey from Sloan Digital Sky Survey III (SDSS-III) is $b_{\mathrm{DLA}} = 1.97 \pm 0.08$, implying

★ E-mail: ignasi.perez@lam.fr

substantially higher halo masses than previously believed. Models where DLAs are present in a broad range of halo masses, including dwarf galaxies but also massive galaxies and galaxy groups, are consistent with this value of the bias factor if the average halo cross-section to produce a DLA increases with the halo mass M_h at least as steeply as $\Sigma_{DLA} \propto M_h$. The results also show there is no dependence of the DLA bias on hydrogen column density, and no evolution with redshift to within measurement errors.

There are other observable properties of DLAs, however, that we expect the bias factor to depend on. The velocity dispersion of matter in dynamical equilibrium in the halo should increase with the host halo mass as $\sigma \propto M_h^{1/3}$ for a fixed halo collapse epoch. The mass–metallicity relation observed between the stellar mass and the metal abundance of stars, already in place at redshifts $z \sim 3$ (Maiolino et al. 2008), suggests that a corresponding relation probably holds between the metallicity in the gas phase and the host halo mass. The strength of metal absorption lines should increase with metallicity, and also with the velocity dispersion of the absorbing gas for saturated lines. Therefore, the bias factor should increase with the strength of metal lines.

Studies of these metal lines have shown that DLA metallicities are typically low and distributed over a broad range, and that the mean metallicity decreases with redshift (Kulkarni & Fall 2002; Vladilo 2002; Prochaska et al. 2003; Kulkarni et al. 2005; Rafelski et al. 2012, 2014; Jorgenson, Murphy & Thompson 2013; Neeleman et al. 2013). The gas velocity dispersion can be measured when several absorption components are seen (Prochaska & Wolfe 1997; Prochaska et al. 2008), although their relation with a halo velocity dispersion is uncertain and may be affected by disc dynamics. A relation of metallicity and velocity dispersion has been found from these observations (Neeleman et al. 2013). These detailed studies, however, can only be done with high-resolution and signal-to-noise spectra, which are not possible to obtain for the large samples of DLAs that are required to detect the large-scale cross-correlation with the Lyα forest.

The only DLA sample that is large enough at present to allow for an accurate determination of the DLA-Lyα forest cross-correlation is the one obtained from the BOSS survey (Noterdaeme et al. 2009, 2012). In the BOSS spectra, metal lines are practically unresolved (the characteristic width of the broadest DLA systems is comparable to the BOSS spectrograph resolution), and the signal to noise is most often below $\sim$3. These spectra allow only a detection of metal lines and rough measurements of their equivalent widths, although many average DLA properties can be derived by stacking the absorption spectra of many DLAs (Mas-Ribas et al. 2017).

In this paper, we will use an estimate of the metal strength of individual DLAs obtained by averaging the equivalent widths of several metal lines, which we have previously defined and studied in Arinyo-i-Prats et al. (2018), to classify the DLAs into several groups of different metal strength. We will then measure the bias factor for these groups to look for a dependence on the metal strength. It is not possible in general to tell the dependence of the metal strength (or equivalent widths of the strongest observable lines) on the metal abundance, and the velocity dispersion of the absorbing gas in each individual DLA without obtaining spectra of much higher quality. In a future paper, we plan to investigate the average relation of this metal strength parameter to metallicity and velocity dispersion from the study of stacked DLA absorption spectra, following the technique of Mas-Ribas et al. (2017). For now, we simply note that metallicity and velocity dispersion should be correlated with each other and should both increase with host halo mass.

We start by describing the data sets used to derive the DLA bias in Section 2. The estimator for the cross-correlation as well as the model used to derive the DLA bias is described in Section 3. The results are presented and discussed in Section 4, and we summarize our conclusions in Section 5. Throughout this paper, we use a flat ΛCDM cosmology, with $\Omega_m = 0.3156$, $\Omega_b = 0.0492$, $h = 0.6727$, $n_s = 0.9645$, and $\sigma_8 = 0.831$, as reported by Planck Collaboration XIII (2016).

2 DATA SAMPLE

As in Pérez-Ràfols et al. (2018), we use the DLA catalogue obtained with the technique of Noterdaeme et al. (2009) from the final DR12 of BOSS (Dawson et al. 2013), from SDSS-III (Gunn et al. 1998; York et al. 2000; Gunn et al. 2006; Eisenstein et al. 2011; Bolton et al. 2012; Smee et al. 2013). These DLAs were searched in the final DR12 quasar catalogue, Pâris et al. (2017) that used the quasar target selection of BOSS, as summarized in Ross et al. (2012).

The Lyα forest absorption, used as tracer of the underlying density field around DLAs, is measured from the same set of 157 922 quasar spectra as in Pérez-Ràfols et al. (2018), which have $\sim$27 million pixels in the Lyα forest. Because the wavelength resolution of the BOSS spectra is better than required to measure the cross-correlation function at the scales we are interested in, we actually use *analysis pixels* in all of our calculations, described in Busca et al. (2013), which are the average of every three pixels of the actual co-added spectra and have a width $\Delta_v = (\Delta\lambda/\lambda)c \simeq 207 \, \text{km s}^{-1}$.

Our DLA sample is defined starting from sample $C1$ of Pérez-Ràfols et al. (2018), which includes 23 342 DLA candidates with column density $N_{HI} \geq 10^{20} \, \text{cm}^{-2}$. Note that we use a column density threshold lower than the standard DLA threshold of $N_{HI} \geq 2 \times 10^{20} \, \text{cm}^{-2}$. The reason is that the larger number of DLAs we obtain implies a smaller statistical error, and that no evidence of a change of the bias factor with column density was found in Pérez-Ràfols et al. (2018). This threshold is in any case a conventional definition since the transition to a mostly atomic and self-shielded medium with increasing N_{HI} is a gradual one. This $C1$ sample is drawn from an early version of the DR12 extension of the DLA catalogue from Noterdaeme et al. (2012), after applying the first three cuts described in Pérez-Ràfols et al. (2018), which are as follows:

(i) The DLA redshift is in the range $2.0 \leq z_{DLA} < 3.5$.

(ii) We require an average (CNR) in the Lyα forest region CNR $\geq$ 3.

(iii) We exclude DLAs found in quasar spectra with positive Balnicity Index, as listed in the DR12Q catalogue of Pâris et al. (2014).

The additional cuts 4–6 described in Pérez-Ràfols et al. (2018) are not applied here, because they were found not to significantly affect the measured bias factor.

The goal of this paper is to measure the dependence of the DLA bias factor on the metal strength parameter S, as defined in Arinyo-i-Prats et al. (2018). Briefly, the S parameter is an average of the equivalent widths of metal lines associated with a DLA, optimally weighted to obtain the best possible signal to noise, and normalized so that $S = 1$ represents an average of the DLA metal strength. The metal strength S is therefore a quantity that depends both on the column density of the measured metal species and the velocity dispersion of the DLAs, because of line saturation effects on the measured equivalent widths. The precise relation of S to the metal column density and velocity dispersion is uncertain, and can be constrained from the stacked spectra of DLAs in different intervals

Table 1. Properties of DLA samples *SA*, *S*1, *S*2, and *S*3. The fourth column gives the average value and dispersion of *S* in each sample, computed by weighting the samples with weights equal to ϵ_S^{-1}. Bias values are given at the reference redshift $z_{\rm ref} = 2.3$. Results for sample *C*1 of Pérez-Ràfols et al. (2018) are given for reference. The χ^2 parameter of the fit of the cross-correlation in each sample is in the last column.

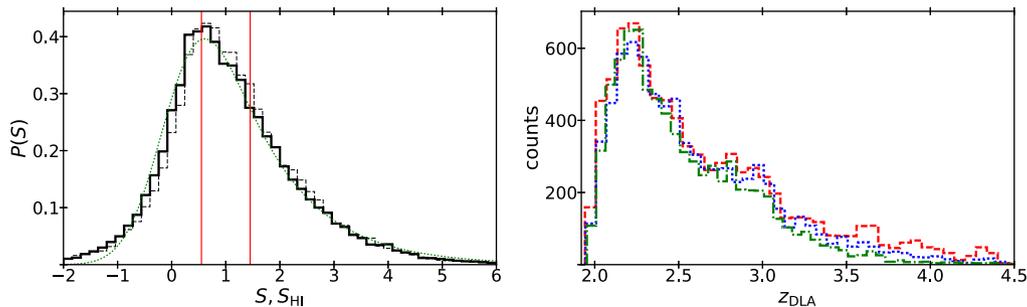| Name	Range	DLA number	$\overline{S}$	$b_{\rm DLA}$	χ^2 (d.o.f.)
SA	$-2 < S$	18 221	1.13 ± 1.21	2.00 ± 0.09	3001.23 (2896-1)
*S*1	$-2 < S < 0.55$	6039	0.01 ± 0.48	1.67 ± 0.16	2859.52 (2896-1)
*S*2	$0.55 \leq S < 1.45$	5936	0.97 ± 0.25	2.03 ± 0.16	2941.40 (2896-1)
*S*3	$1.45 \leq S$	6246	2.44 ± 0.91	2.27 ± 0.15	2957.17 (2896-1)
*C*1	–	23 342	–	1.97 ± 0.08	3061.04 (2896-1)

Figure 1. Left: Distribution of the metal strength *S* for the DLAs in sample SA (thick black line), and of the H I-corrected metal strength $S_{\rm H\,I}$ (thin dashed line). Vertical red lines indicate the bins used to construct the sub-samples *S*1–*S*3. Green dotted line shows an exponential profile with $\lambda_S = 1.2$ convolved with a Gaussian with error $\epsilon_S = 0.6$ (see Section 4.3), which adequately describes the measured distribution. Right: Distribution of $z_{\rm DLA}$ for the DLAs in samples *S*1 (red dashed line), *S*2 (blue dotted line), and *S*3 (green dashed–dotted line).

of *S*, which we plan to study in another paper. We note that the *S* parameter depends on the DLA sample that is used, because it is normalized to have a mean value of 1 in the sample, and is therefore a measure of the metal strength compared to the mean of other DLAs. However, our DLA sample is a representative set of absorbers intercepted on random lines of sight, so the quantity *S* is an actual estimate of the metal line strength compared to the average DLA in the Universe.

The true value of *S* should be positive for every DLA, but the frequently large spectral noise may occasionally render it negative. Cases of DLAs with negative *S* are also included to avoid biasing our sample. There are a small number of DLAs for which it is not possible to measure the equivalent width of any line for various reasons (see Arinyo-i-Prats et al. 2018; for more details). Removing these objects reduces our sample to 23 312 DLAs.

To measure the dependence of the bias with the metal strength, we separate our DLA sample into three bins in *S*, defined as $S < 0.55$, $0.55 \leq S < 1.45$, and $1.45 \leq S$, chosen to have similar numbers of DLAs in each bin. We label the sub-samples of DLAs in each bin as *S*1, *S*2, and *S*3, as shown in Table 1. We estimate for each DLA the error of the value of *S*, ϵ_S, measured from the equivalent widths of several metal lines, as described in detail in Arinyo-i-Prats et al. (2018). To avoid having too many objects that are placed in the wrong bin because of the measurement error ϵ_S, we require this error to be smaller than a threshold value, which we set for our standard case to be $\epsilon_S < 0.5$, although we shall examine the dependence of our results on this value. In addition we remove any DLAs with a value $S < -2$, because we have found that these systems suffer from systematic errors with a highly non-Gaussian distribution tail. After removing all the DLAs with an error $\epsilon_S > 0.5$ or $S < -2$, our total sample is reduced to the *SA* sample, with 18 221 DLAs, of which roughly 6000 are in each of the three sub-samples *S*1, *S*2, and *S*3 (see Table 1). Note that the values of *S* are defined to have a mean value of 1 for our whole C1 sample, and that removing

some objects affects this mean. We do not renormalize the values of *S*, as seen in table 1 (e.g. sample *SA* has a mean *S* slightly larger than 1).

The left-hand panel of Fig. 1 shows the distribution of the *SA* DLA sample in metal strength. The bins separating the sub-samples *S*1 to *S*3 are indicated as red solid lines. The right-hand panel shows the redshift distribution of sub-samples *S*1, *S*2, and *S*3. The redshift distributions are similar, although there is an excess of low *S* systems at high redshift. This excess can be not only due to the known increase of metallicity with cosmic time, but also by an increased contamination of false DLAs which have weak or absent metal lines. In any case, this small difference in the redshift distribution should not significantly affect differences in the bias factor of the three sub-samples, because no redshift evolution of the DLA bias was detected in Pérez-Ràfols et al. (2018).

The green dotted line in the left-hand panel models the measured distribution of *S* as an exponential distribution $P(S) = e^{-S/\lambda_S}/\lambda_S$, with $\lambda_S = 1.2$, convolved with a Gaussian with error $\epsilon_S = 0.6$, which we find fits the measured distribution adequately. The required error ϵ_S is larger than the statistical errors calculated from the spectra when measuring the equivalent widths involved in the calculation of *S*, which indicates the likely presence of unaccounted systematic errors. We will come back to this question in Section 4.3. The fact that the true distribution of *S* can be modelled as an exponential is not surprising, because the equivalent width distribution of metal lines are usually reasonably fitted by this form.

3 CROSS-CORRELATION: MEASUREMENT AND MODEL

In this section, we summarize our method to compute the cross-correlation of DLAs and the Lyα forest transmission fluctuation $\delta_i = 1 - F_i/\bar{F}$ (where F_i is the transmission fraction at any pixel i of a quasar spectrum and $\bar{F}$ is its mean value), the covariance matrix

of this cross-correlation, and the model used to infer the DLA bias from a fit to the cross-correlation.

3.1 Measurement of the cross-correlation

We use the same estimator for the cross-correlation of DLAs and Lyα transmission fluctuation, ξ, as in Font-Ribera et al. (2012) and Pérez-Ràfols et al. (2018), where the method is described in more detail. After dividing the plane of the parallel and perpendicular components of the separation vector from a DLA to a Lyα forest pixel, $(r_{\parallel}, r_{\perp})$, into bins that we designate by A, the cross-correlation at the bin A is

$$\xi^{A} = \frac{\sum_{i \in A} w_i \delta_i}{\sum_{i \in A} w_i}, \tag{1}$$

where the sum is over all DLAs and over all pixels i located within a bin A of the separation $\mathbf{r}$ from a DLA (note that a given Lyα forest pixel may appear several times in this sum, as many as DLAs there are within the separation A from the pixel). Here, the fluctuations δ_i have been corrected for continuum fitting in an operation that we call 'projection', and the weights w_i are defined to optimize the accuracy of the measurement of ξ^{A} (see Pérez-Ràfols et al. 2018 and references therein for a detailed description).

The covariance of the cross-correlation at two bins A and B is equal to

$$C^{AB} \equiv \left\langle \xi^{A} \xi^{B} \right\rangle - \left\langle \xi^{A} \right\rangle \left\langle \xi^{B} \right\rangle$$
$$= \frac{1}{N^{AB}} \sum_{i \in A} \sum_{j \in B} w_i w_j \zeta_{ij}, \tag{2}$$

where ζ_{ij} is the Lyα forest autocorrelation at two pixels i and j, and each of the two sums is again understood to be over all Lyα forest pixels and all the DLAs at separations within the bins A or B. To compute C^{AB}, we include only pairs of Lyα pixels that are on the same spectrum, i.e. separated only by a parallel component. We neglect the contribution to the covariance matrix of pixels from different forests. The normalization factor is

$$N^{AB} = \sum_{i \in A} \sum_{j \in B} w_i w_j. \tag{3}$$

3.2 Model of the cross-correlation

We model the cross-correlation starting from the cross-power spectrum, and then Fourier Transform it to obtain the model cross-correlation. We compute the poser spectrum assuming the linear theory of redshift space distortions of Kaiser (1987) as in Font-Ribera et al. (2012) and Pérez-Ràfols et al. (2018)

$$P_{\mathrm{DLA,Ly\alpha}}(\mathbf{k}, z) = b_{\mathrm{DLA}} \left(1 + \beta_{\mathrm{DLA}} \mu_k^2\right) b_{\mathrm{Ly\alpha}} \left(1 + \beta_{\mathrm{Ly\alpha}} \mu_k^2\right)$$
$$\times P_L(\mathbf{k}, z) G(\mathbf{k}) S(k_{\parallel}), \tag{4}$$

where b_{DLA} and $b_{\mathrm{Ly\alpha}}$ are the bias factors of DLAs and the Lyα forest, respectively, β_{DLA} and $\beta_{\mathrm{Ly\alpha}}$ their redshift space distortion parameter, $\mu_k = k_{\parallel}/k$ is the cosine of the angle of the Fourier mode vector with respect to the line of sight, and $P_L(\mathbf{k}, z)$ is the linear matter power spectrum. The functions G and S are added to account for the smoothing that occurs due to the spectrograph resolution, the binning of the Lyα forest spectra, and the binning of the $(r_{\parallel}, r_{\perp})$ plane to measure the cross-correlation. Details of the assumed form for G and S are found in Section 4 of Pérez-Ràfols et al. (2018). We use comoving bin sizes $\Delta_{\parallel} = \Delta_{\perp} = 2\,h^{-1}$ Mpc to compute the cross-correlation function, out to maximum separation components $r_{\parallel} < 80\,h^{-1}$ Mpc and $r_{\perp} < 80\,h^{-1}$ Mpc. In addition, we apply to the

model the same projection operation discussed above that is applied to the data to correct for continuum fitting effects, as explained in appendix B.3 of Pérez-Ràfols et al. (2018).

We use the publicly available code BAOFIT to fit the model, which computes the linear power spectrum P_L using CAMB (Lewis & Challinor 2011; Kirkby et al. 2013). The model is evaluated at the mean values of $r_{\parallel}$ and $r_{\perp}$ of each bin, at the mean redshift of our sample.

As in Pérez-Ràfols et al. (2018), we fix $b_{\mathrm{Ly\alpha}}(1 + \beta_{\mathrm{Ly\alpha}}) = -0.325$ at the reference redshift $z_{\mathrm{ref}} = 2.3$, and $\beta_{\mathrm{Ly\alpha}} = 1.663$, as found from a detailed measurement of the Lyα forest transmission autocorrelation in Bautista et al. (2017). We assume the amplitude of the cross-correlation evolves proportionally to $(1 + z)^{0.9}$, corresponding to $b_{\mathrm{Ly\alpha}} \propto (1 + z)^{2.9}$ as found previously in McDonald et al. (2006), Palanque-Delabrouille et al. (2013), and a constant DLA bias, consistently with the results of Pérez-Ràfols et al. (2018). We also assume a non-evolving $\beta_{\mathrm{Ly\alpha}}$, and we fix $\beta_{\mathrm{DLA}} b_{\mathrm{DLA}} = f(\Omega) = 0.9689$.

In summary, we fix all the parameters in equation (4) except for the DLA bias factor. Our reported errors on b_{DLA} include only measurement errors of the DLA-Lyα cross-correlation, and not the errors on the Lyα bias parameters or any other ingredients in our modelling. Our derived values of b_{DLA} decrease with the amplitude of P_L and with the value of $b_{\mathrm{Ly\alpha}}$ assumed in our model. However, in this paper we are interested in the variation of b_{DLA} with the metal strength S, which is not affected by these parameters (except for the small effect caused by the variation of β_{DLA} with b_{DLA}, which means that b_{DLA} is not exactly inversely proportional to the assumed amplitude of P_L and $b_{\mathrm{Ly\alpha}}$).

4 RESULTS AND DISCUSSION

We now present the results of our fits to the measured cross-correlations for each sample. We exclude from the fit bins at very small separations, with $r = \sqrt{r_{\parallel}^2 + r_{\perp}^2} \leq 5\,h^{-1}$ Mpc. The measured DLA bias factors are listed in Table 1 and shown in the left-hand panel of Fig. 2. We also include the measurement for sample $C1$ of Pérez-Ràfols et al. (2018) in Table 1 as a reference value.

There is a clear increase of the bias factor with S. A simple linear fit, shown by the red solid line, yields $b_{\mathrm{DLA}}(S) = (0.25 \pm 0.06)S + (1.71 \pm 0.09)$, with a slope that is greater than zero at 4σ. In the rest of this section we discuss how this relation changes when we use the metal strength corrected for its dependence on $N_{\mathrm{H\,I}}$ (Section 4.1), the possible impact of impurities on these results (Section 4.2), and the correction we compute for the bias–metal strength relation that is caused by the measurement errors in S (Section 4.3). Finally, we discuss how this result relates to the mass–metallicity relation of galaxies in Section 4.4.

4.1 H I-corrected metal strength

The metal strength is expected to increase with the column density of each metal species and with the velocity dispersion of the absorbing gas. At the same time, the theory of halo formation in theories with hierarchical clustering tells us that the bias factor should increase with the host halo mass. The host halo mass is directly related to the velocity dispersion, and we also expect a mass-metallicity relation to exist, as observed in galaxies, implying that more massive haloes should have a higher metal abundance in the gas phase. However, at fixed metallicity, the column density of metal species should increase with $N_{\mathrm{H\,I}}$, and should in fact be directly proportional to $N_{\mathrm{H\,I}}$ if the average ionization conditions in DLAs do not change with
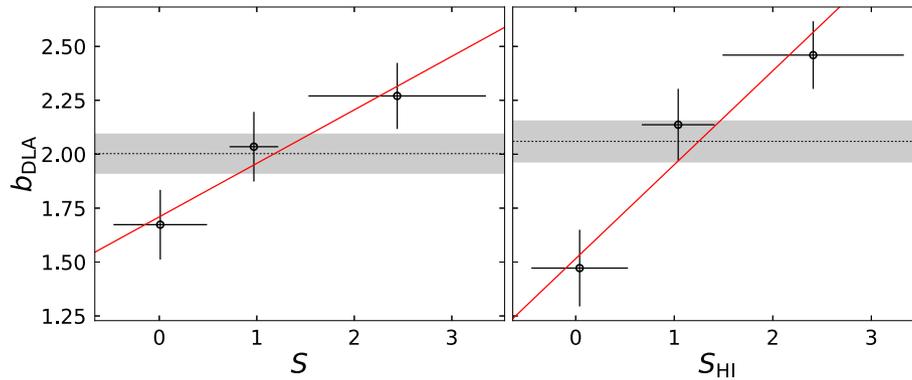
Figure 2. *Left*: Bias of the DLAs for the three bins in S of samples $S1$, $S2$, and $S3$ (see Table 1), with errors shown as vertical bars. The value for sample SA is shown by the black dashed line, with its error indicated by the shaded area. Red solid line shows a linear fit to the points. *Right*: Same as in left-hand panel for the corrected metal strength S_{H_I}.

Table 2. Properties of the DLA sub-samples SA_{H_I}, $S1_{H_I}$, $S2_{H_I}$, and $S3_{H_I}$, obtained by dividing the DLAs into the same bins in S_{H_I} as the ones used for S shown in Table 1. The result for the bias factor is given in the fourth column, and the χ^2 of the fit in the fifth column.

Name	Number of DLAs	$\overline{S}_{H_I}$	b_{DLA}	χ^2 (d.o.f.)
SA_{H_I}	16 666	1.19 ± 1.20	2.06 ± 0.10	3041.56 (2896-1)
$S1_{H_I}$	4951	-0.01 ± 0.51	1.47 ± 0.18	2894.13 (2896-1)
$S2_{H_I}$	5630	1.01 ± 0.36	2.14 ± 0.17	2947.77 (2896-1)
$S3_{H_I}$	5953	2.41 ± 0.92	2.46 ± 0.16	3038.79 (2896-1)

N_{H_I}. Therefore, if these arguments are correct, a corrected metal strength S that makes it independent of N_{H_I} should increase the variation of the bias factor with this parameter.

We use the corrected metal strength S_{H_I} as defined by Arinyo-i-Prats et al. (2018), by empirically determining a linear fit to the dependence of the average value of S on N_{H_I} and subtracting it from S. We repeat the measurement made above for S_{H_I}, dividing the DLAs into the same three bins used for S. This is justified because the distributions of S_{H_I} and S are very similar (see Fig. 1). The properties of the new samples for the corrected metal strength are specified in Table 2. The total number of DLAs is lower when we use S_{H_I} because of the cuts $S_{H_I} > -2$ and $\epsilon_{S,H_I} < 0.5$, which eliminate a larger number of DLAs than for the case when we use S. The dependence of the bias on S_{H_I} is shown in the right-hand panel of Fig. 2 and in Table 2.

The result shows that indeed, the bias factor varies more strongly with the corrected metal strength. The same linear fit that was obtained previously for S now gives the result $b_{DLA}(S_{H_I}) = (0.44 \pm 0.13) S_{H_I} + (1.52 \pm 0.19)$, with a substantially steeper slope. This is consistent with our interpretation that we have detected a physical dependence of the bias factor with metallicity and velocity dispersion, and that metallicity is better characterized by the corrected metal strength S_{H_I} than by S. Note that, as found by Pérez-Ràfols et al. (2018), there is no dependence of the bias factor on N_{H_I} within measurement errors, which might have affected the relation of bias and S_{H_I} if it were present.

4.2 Dependence on the cuts in ϵ_S and continuum-to-noise ratio

The dependence of the bias factor on S and S_{H_I} we have detected might be contaminated by the lack of purity of the DLA catalogue we use, i.e. the fact that our DLAs are only candidates and some

of them may be false DLAs arising from spectral noise or from other absorption systems that are confused with DLAs. For e.g., if a fraction of DLAs in our catalogue were simply arising from noise, their bias factor and their metal strength would both be zero, creating an artificial correlation of the bias and the metal strength.

To test for the possible presence of this contaminating effects, we check for variations of our results with the cut in the error ϵ_S of the metal-strength parameter, and with the CNR in the Ly α forest region of the spectrum that the DLA is detected in.

We start modifying the cut in ϵ_S. As ϵ_S increases, the true sub-samples with different values of S are increasingly mixed in the sub-samples we construct from the measured S. We show in Fig. 3 the bias as a function of S and S_{H_I}, for thresholds $\epsilon_S < (0.3, 0.5, 0.7)$. There is indeed a small reduction of the change in bias with the metal strength as the maximum allowed ϵ is raised, that can be explained by the mixing of sub-samples. However, we do not see a reduction of b_{DLA} at the smallest S or S_{H_I} that might have been caused by a large fraction of fake DLAs among the sub-sample where metal lines are not significantly detected.

We now move our attention to the dependence of the bias-metal strength relation on the minimum value required for the continuum-to-noise ratio of the spectrum in the Ly α forest region where each DLA is detected. The purity of the catalogue should increase with CNR, so any effect of impurities should decrease rapidly with the threshold we impose on CNR. We show this in Fig. 4, where we see again an increased variation of DLA bias with metal strength as the minimum required CNR is increased.

Furthermore, in this case we see a larger variation of the bias for the smallest value of S: a higher threshold in CNR results in a lower bias factor for the DLAs with the smallest metal strength. The variation is highly significant: taking into account that the sub-samples with CNR > 4 contain 70 per cent of the same DLAs in the sub-samples with CNR > 2, the expected statistical fluctuation
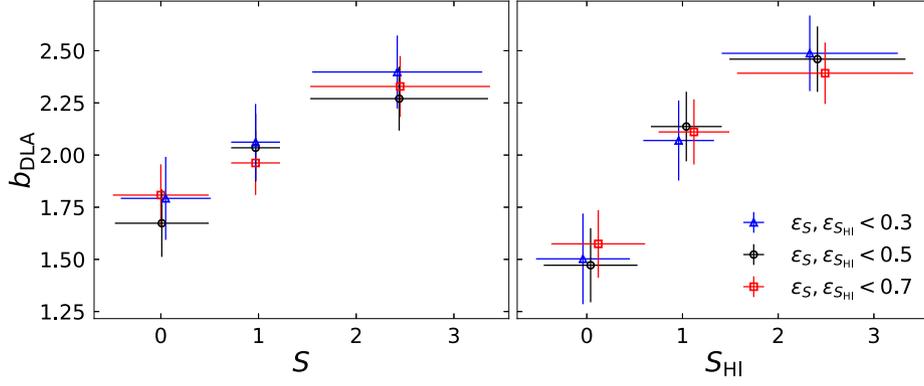
Figure 3. Bias of the DLAs against S obtained by fitting samples $S1$, $S2$, and $S3$ when the cut in ϵ is changed. Right-hand panel is for sub-samples in S_{H_1}.
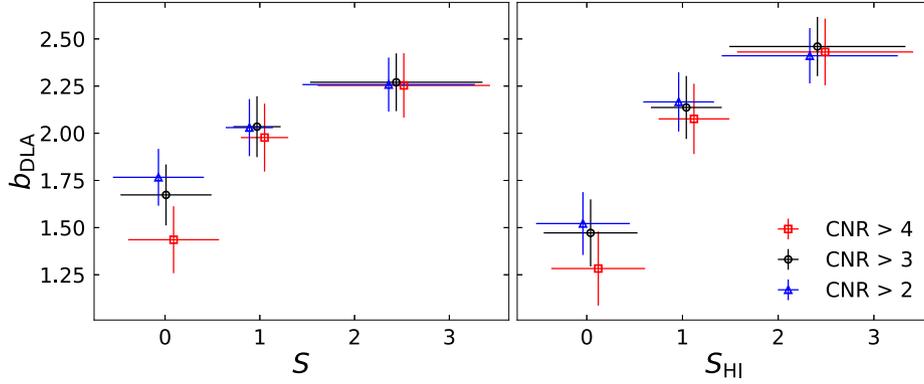

Figure 4. Bias of the DLAs against S (left-hand panel) and S_{H_1} (right-hand panel) for different cuts in CNR. Red squares (CNR > 4) and blue triangles (CNR > 2) have been horizontally shifted to avoid overlap.

in the difference of bias factors is 0.4σ, where σ stands for the error in the largest sample, with CNR > 2 (indicated by the blue error bar in the figure). The expected fluctuation comes from the fact that the sub-sample with CNR > 2 can be considered as the combination of two independent samples (one with CNR > 4, and another one with $2 < $ CNR ≤ 4) that are combined with weight $\sigma^2_{\text{CNR}>4}/\sigma^2_{\text{CNR}>2}$ to construct the final estimator. This implies that the typical fluctuation is $\delta\sigma_{\text{CNR}>2}$, where $\delta = 1 - \sigma^2_{\text{CNR}>4}/\sigma^2_{\text{CNR}>2}$. This variation is opposite to our expectation if impurities caused only by noise were present in the catalogue, in which case false DLAs would have metal strength and bias consistent with zero.

Our interpretation of this result is that the contaminants that are included in the catalogue as the CNR threshold is lowered are mostly regions of broad Ly α absorption in the forest, which correspond to absorbers of lower column density than DLAs but with sufficiently broad velocity dispersions to create a broad absorption line that is consistent with a DLA in the most noisy spectra. These absorbers generally have weaker associated metal lines because of their lower column densities and higher ionization level (most of the lines used to measure the metal strength are of low-ionization species; see Arinyo-i-Prats et al. 2018), and they are more highly biased than the DLAs with the weakest metal lines because they are associated with collapsed regions, or regions in the process of collapsing, with high velocity dispersion. This can explain why the presence of contaminants increases the bias factor, and actually decreases the variation we measure of the bias factor with metal strength. These type of absorbers and their average metal lines are likely similar to the systems studied by Pieri et al. (2014); a detailed study of the

bias factor of these strong Ly α absorbers is separately being done (Blomqvist et al. in preparation).

4.3 Correction for the effect of metal strength measurement errors

The errors in the measurement of S_{H_1} mix the samples $S1_{H_1}$, $S2_{H_1}$ and $S3_{H_1}$, as each DLA is classified to one of these samples based on the measured S_{H_1}. This mixing among the samples should flatten the dependence of the DLA bias on S_{H_1}. Hints of this effect are seen in Fig. 3. In this sub-section we compute a correction for this effect based on the distribution of S.

The intrinsic distribution of S can be modelled as an exponential,

$$P_{\text{in}}(S) = e^{-S/\lambda_S}/\lambda_S, \tag{5}$$

where λ_S is a characteristic value for S. To take into account the effect of errors, we convolve this intrinsic distribution with a Gaussian to obtain the observed distribution,

$$P(S) = \int_{-S}^{\infty} d\delta_S \, P_{\text{in}}(S + \delta_S) \frac{e^{-\delta_S^2/2\epsilon_S^2}}{\sqrt{2\pi\epsilon_S^2}}. \tag{6}$$

The distribution with parameters $\lambda_S = 1.2$ and $\epsilon_S = 0.6$ reasonably reproduces the observed distribution as shown in the left-hand panel of Fig. 1. This good match is not surprising, since we know that the equivalent width distribution of metal lines is reasonably fitted by an exponential function, and S was defined to make its average be close to unity. The typical error in this Gaussian is somewhat larger than the actual errors reported in the catalogue (see figure 5 of Arinyo-i-Prats et al. 2018). This indicates that the true errors may be larger
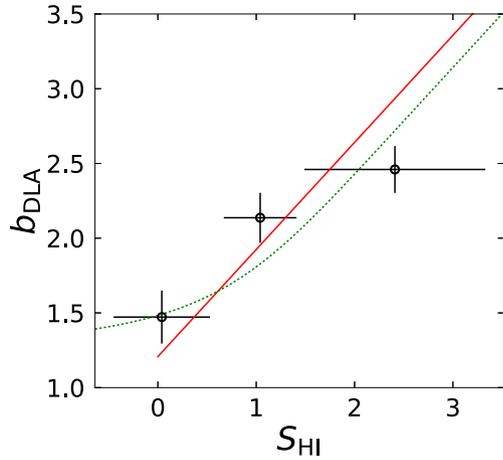
Figure 5. Bias of DLAs versus H I-corrected metal strength, $S_{H I}$. Green dotted line shows a fit to the points assuming an intrinsic linear relation for the bias, shown as the red solid line, and applying equation (7).

than the estimate calculated in Arinyo-i-Prats et al. (2018) from the statistical errors of the flux in spectral pixels, probably because of continuum fitting and other systematic errors. The correction presented here takes this increased errors into account by calibrating them from the observed distribution of S.

We now use this distribution to compute a correction for the measured DLA bias, given an intrinsic or true relation $b_{DLA, in}(S)$

$$b_{DLA}(S) = \frac{1}{\mathcal{N}} \int_{-S}^{\infty} d\delta_S \, b_{DLA,in} \, (S + \delta_S) \, e^{-(S+\delta_S)/\lambda_S} \frac{e^{-\delta_S^2/2\epsilon_S^2}}{\sqrt{2\pi \epsilon_S^2}}, \quad (7)$$

where

$$\mathcal{N} = \int_{-S}^{\infty} d\delta_S \, e^{-(S+\delta_S)/\lambda_S} \frac{e^{-\delta_S^2/2\epsilon_S^2}}{\sqrt{2\pi \epsilon_S^2}}. \quad (8)$$

We apply this correction to the measurement of the bias using $S_{H I}$. We assume that the intrinsic bias–metal strength relation is linear, and we fit the predicted observed relation to the three values of our chosen bins. The result is shown in Fig. 5. The best-fitting solution is $b_{DLA\,in}(S) = (0.72 \pm 0.30)S_{H I} + (1.21 \pm 0.24)$, and the corresponding observed relation is the green dotted line. The predicted observed relation has a positive second derivative, whereas the observed points show a negative second derivative, although this is not highly significant. In any case, there is no reason why the intrinsic relation should be linear, and the true relation can only be predicted from detailed cosmological simulations of the haloes giving rise to DLAs, but more data and an increased number of bins would be necessary to measure the shape of this relation more accurately. Nevertheless, the simple correction we have applied highlights two important results. First, the true relation of the DLA bias to the metal strength is substantially steeper than what we have directly measured because of the spreading effect of measurement errors. Secondly, the intrinsic estimated value of the DLA bias at zero metal strength is 1.21 ± 0.24, and is consistent with unity, which is the value expected for low-mass haloes hosting dwarf galaxies ($M \sim 10^9$–$10^{10} M_\odot$; see figure 8 of Pérez-Ràfols et al. 2018). This is consistent with the observed mass–metallicity relation, where more metal poor galaxies have lower stellar mass (Maiolino et al. 2008), and suggests that DLAs are hosted by haloes with a very broad range of masses, with metal-line strength strongly correlating with host halo mass.

4.4 Implications on the mass–metallicity relation

The relation of the DLA bias with the metal strength can arise from an increase of the DLAs metallicity with host halo mass, or also an increase of the velocity dispersion. The theory of ΛCDM for structure formation makes a clear prediction of how the velocity dispersion must increase with halo mass, although uncertainties can be introduced in the velocity dispersion of the gas depending on its density profile. The relation of metallicity to stellar mass has been empirically found in galaxies, and can be measured at present even at the relatively high redshifts of our DLA sample (e.g. Ledoux et al. 2006; Maiolino et al. 2008; Møller et al. 2013). We expect that future studies will be able to discern the contribution of the velocity dispersion and metallicity correlations with halo mass by comparing the mean metal strength of several lines in our different samples (see Mas-Ribas et al. 2017), or from individual measurements at high resolution and signal to noise of an adequate sample of DLAs to clarify how velocity dispersion and metallicity affect the metal strength parameter we have used (see e.g. Prochaska et al. 2008). This promises to allow precise measurements of a mass-metallicity relation for the gas phase, in terms of halo mass instead of stellar mass.

5 SUMMARY AND CONCLUSIONS

In this paper, we make use of the metal strengths obtained for a large number of DLAs in Arinyo-i-Prats et al. (2018) to measure the bias of DLAs as a function of metal strength. We divide the total DLA sample into three subsamples of different metal strength, and we measure the cross-correlation of each sub-sample with the Ly α forest. We fit a linear theory model to the cross-correlations to derive the bias factor of DLAs as a function of S, and as a function of $S_{H I}$, the metal strength corrected for the dependence on $N_{H I}$ (Arinyo-i-Prats et al. 2018). Our main results are summarized as follows:

(i) We find a clear dependence of the DLA bias on S. A linear fit yields $b_{DLA}(S) = (0.25 \pm 0.06)S + (1.71 \pm 0.09)$. For the first time, we find a dependence of the bias factor of DLAs with their metal content.

(ii) The dependence on $S_{H I}$ is even stronger, $b_{DLA}(S_{H I}) = (0.44 \pm 0.13)S_{H I} + (1.52 \pm 0.19)$, which confirms that the effect we detect is real and is related to metallicity rather than metal column density. These detections are statistically significant at more than 3σ.

(iii) We note that the presence of contaminants, which are likely strong Ly α absorbers that are confused with DLAs at low signal to noise, can significantly increase the value of the bias for the lowest bin in S or $S_{H I}$, further increasing the slope of the true linear relation.

(iv) The large errors in the measurement of the metal strength that are inevitable in our large sample of DLAs observed at low signal to noise causes a spreading over S that dilutes the dependence of bias on metal strength. Calculating a simple correction for this spreading effect, we fit a linear relation and obtain $b_{DLA\,in}(S_{H I}) = (0.73 \pm 0.31)S_{H I} + (1.21 \pm 0.24)$. This suggests that the lowest metallicity DLAs have a bias close to unity, which is characteristic of the lowest mass haloes hosting dwarf galaxies, and that the average DLA bias of $b_{DLA} = 1.98 \pm 0.08$ found in Pérez-Ràfols et al. (2018) is the result of averaging the low bias of the low-metallicity DLAs in dwarf galaxies with DLAs in very massive haloes, with high metallicity and velocity dispersion, and bias factor as high as 2.5–3. Physically motivated models for the true bias–metal strength relation from cosmological simulations will be highly valuable in the future.

(v) The bias–metal strength relation we have investigated for DLAs is most likely related to the observed stellar mass–metallicity relation in galaxies (Maiolino et al. 2008). We will investigate the relation between metal strength and metallicity in a future paper. More complete studies promise to measure a relation of the metallicity in the gas phase with halo mass, which can help establish the physical origin of this relation both in the star and gas contents of galaxies.

REFERENCES

Arinyo-i-Prats A., Mas-Ribas L., Miralda-Escude J., Perez-Rafols I., Noterdaeme P., 2018, preprint (arXiv:1801.02605)
Bautista J. E., et al., 2017, A&A, 603, A12
Bolton A. S. et al., 2012, AJ, 144, 144
Busca N. G. et al., 2013, A&A, 552, A96
Crighton N. H. M. et al., 2015, MNRAS, 452, 217
Dawson K. S. et al., 2013, AJ, 145, 10
Eisenstein D. J. et al., 2011, AJ, 142, 72
Font-Ribera A. et al., 2012, J. Cosmol. Astropart. Phys., 11, 59
Gunn J. E. et al., 1998, AJ, 116, 3040
Gunn J. E. et al., 2006, AJ, 131, 2332
Jorgenson R. A., Murphy M. T., Thompson R., 2013, MNRAS, 435, 482
Kaiser N., 1987, MNRAS, 227, 1
Kirkby D. et al., 2013, J. Cosmol. Astropart. Phys., 3, 24
Kulkarni V. P., Fall S. M., 2002, ApJ, 580, 732
Kulkarni V. P., Fall S. M., Lauroesch J. T., York D. G., Welty D. E., Khare P., Truran J. W., 2005, ApJ, 618, 68
Ledoux C., Petitjean P., Fynbo J. P. U., Møller P., Srianand R., 2006, A&A, 457, 71
Lewis A., Challinor A., 2011, CAMB: Code for Anisotropies in the Microwave Background. Astrophysics Source Code Library
Maiolino R. et al., 2008, A&A, 488, 463
Mas-Ribas L. et al., 2017, ApJ, 846, 4
McDonald P. et al., 2006, ApJS, 163, 80
Møller P., Fynbo J. P. U., Ledoux C., Nilsson K. K., 2013, MNRAS, 430, 2680
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., 2012, A&A, 547, L1
Padmanabhan H., Choudhury T. R., Refregier A., 2016, MNRAS, 458, 781
Palanque-Delabrouille N. et al., 2013, A&A, 559, A85
Pieri M. M. et al., 2014, MNRAS, 441, 1718
Planck Collaboration et al. 2016, A&A, 594, A13
Prochaska J. X., Chen H.-W., Wolfe A. M., Dessauges-Zavadsky M., Bloom J. S., 2008, ApJ, 672, 59
Prochaska J. X., Gawiser E., Wolfe A. M., Castro S., Djorgovski S. G., 2003, ApJ, 595, L9
Prochaska J. X., Herbert-Fort S., Wolfe A. M., 2005, ApJ, 635, 123
Prochaska J. X., Wolfe A. M., 1997, ApJ, 487, 73
Prochaska J. X., Wolfe A. M., 2009, ApJ, 696, 1543
Pâris I. et al., 2014, A&A, 563, A54
Pâris I. et al., 2017, A&A, 597, A79
Pérez-Ràfols I. et al., 2018, MNRAS, 473, 3019
Péroux C., McMahon R. G., Storrie-Lombardi L. J., Irwin M. J., 2003, MNRAS, 346, 1103
Rafelski M., Neeleman M., Fumagalli M., Wolfe A. M., Prochaska J. X., 2014, ApJ, 782, L29
Rafelski M., Wolfe A. M., Prochaska J. X., Neeleman M., Mendez A. J., 2012, ApJ, 755, 89
Ross N. P. S. et al., 2012, ApJS, 199, 3
Smee S. A. et al., 2013, AJ, 146, 32
Vladilo G., 2002, A&A, 391, 407
Wolfe A. M., Turnshek D. A., Smith H. E., Cohen R. D., 1986, ApJS, 61, 249
York D. G. et al., 2000, AJ, 120, 1579
Zafar T., Péroux C., Popping A., Milliard B., Deharveng J. M., Frank S., 2013, A&A, 556, A141

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