



**HAL**  
open science

# High-redshift Damped Ly $\alpha$ Absorbing Galaxy Model Reproducing the N H – Z Distribution

Jens-Kristian Krogager, Palle Møller, Lise B Christensen, Pasquier  
Noterdaeme, Johan P U Fynbo, Wolfram Freudling

► **To cite this version:**

Jens-Kristian Krogager, Palle Møller, Lise B Christensen, Pasquier Noterdaeme, Johan P U Fynbo, et al.. High-redshift Damped Ly $\alpha$  Absorbing Galaxy Model Reproducing the N H – Z Distribution. Monthly Notices of the Royal Astronomical Society, 2020, 495 (3), pp.3014-3021. 10.1093/mnras/staa1414 . hal-02997808

**HAL Id: hal-02997808**

**<https://hal.sorbonne-universite.fr/hal-02997808v1>**

Submitted on 10 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# High-redshift Damped Ly $\alpha$ Absorbing Galaxy Model Reproducing the $N_{\text{H I}} - Z$ Distribution

Jens-Kristian Krogager<sup>1</sup>, Palle Møller<sup>2</sup>, Lise B. Christensen<sup>3</sup>, Pasquier Noterdaeme<sup>1</sup>, Johan P. U. Fynbo<sup>3</sup>, Wolfram Freudling<sup>2</sup>

<sup>1</sup>*Institut d'Astrophysique de Paris, CNRS-SU, UMR7095, 98bis bd Arago, FR-75014 Paris, France; E-mail: krogager@iap.fr*

<sup>2</sup>*European Southern Observatory, Karl-Schwarzschildstrasse 2, D-85748 Garching bei München, Germany*

<sup>3</sup>*DARK, Niels Bohr Institute, University of Copenhagen, Lyngbyvej 2, DK-2100 Copenhagen Ø, Denmark*

19 May 2020

## ABSTRACT

We investigate how damped Lyman- $\alpha$  absorbers (DLAs) at  $z \sim 2 - 3$ , detected in large optical spectroscopic surveys of quasars, trace the population of star-forming galaxies. Building on previous results, we construct a model based on observed and physically motivated scaling relations in order to reproduce the bivariate distributions of metallicity,  $Z$ , and H I column density,  $N_{\text{H I}}$ . Furthermore, the observed impact parameters for galaxies associated to DLAs are in agreement with the model predictions. The model strongly favours a metallicity gradient, which scales with the luminosity of the host galaxy, with a value of  $\gamma^* = -0.019 \pm 0.008 \text{ dex kpc}^{-1}$  for  $L^*$  galaxies that gets steeper for fainter galaxies. We find that DLAs trace galaxies over a wide range of galaxy luminosities, however, the bulk of the DLA cross-section arises in galaxies with  $L \sim 0.1 L^*$  at  $z \sim 2.5$  consistent with numerical simulations.

**Key words:** galaxies: high-redshift — galaxies: statistics — quasars: absorption lines

## 1 INTRODUCTION

The properties of neutral gas are crucial for a complete understanding of galaxy evolution. Locally, the neutral gas can be studied directly through the H I 21-cm transition (e.g., Zwaan et al. 2005; Walter et al. 2008). At high redshift the neutral gas phase is best studied through Ly $\alpha$  absorption towards background quasars. The so-called damped Ly $\alpha$  absorbers (DLAs) with  $N_{\text{H I}} > 2 \times 10^{20} \text{ cm}^{-2}$  are of particular interest for studies of galactic environments (Wolfe et al. 2005). Due to their high column density of neutral hydrogen, DLAs arise in predominantly neutral gas (Viegas 1995). As a result, ionization corrections for metallicity measurements are negligible.

The distribution function of  $N_{\text{H I}}$  for DLAs,  $f(N_{\text{H I}})$ , has been studied in detail (Prochaska & Wolfe 2009; Noterdaeme et al. 2012; Bird et al. 2017) and provides a measurement of the cosmic mass density of neutral hydrogen in DLAs,  $\Omega_{\text{DLA}}$ . It is found that  $\Omega_{\text{DLA}}$  makes up  $\sim 80\%$  of the total mass density of H I at  $z > 2.2$  (Noterdaeme et al. 2009).

The plethora of low-ionization metal absorption lines make it possible to obtain accurate measurements of the metallicity,  $Z$ , in the neutral gas (Rafelski et al. 2012; Jorgenson et al. 2013; De Cia et al. 2018). It is found that the average  $Z$  increases towards lower redshifts in agreement with expectations from the build up of metals via star formation (Dvorkin et al. 2015).

In order to properly interpret the properties of DLAs it is important to know which galaxies give rise to DLAs over cosmic time. One way to understand the galaxy population associated to DLAs is through numerical simulations. Large cosmological simulations are able to match the observables from DLAs at  $z \sim 3$  fairly well

(e.g. Pontzen et al. 2008; Altay et al. 2013; Bird et al. 2013, 2014, 2015; Rahmati & Schaye 2014). However, the redshift evolution and the details of the bivariate distribution of  $N_{\text{H I}}$  and  $Z$  are still not well-understood (Hassan et al. 2020) and the properties of DLAs in simulations depend strongly on the rather ad-hoc feedback mechanisms from supernovae and quasars assumed in the simulations.

Another way to study DLA galaxy properties is by cross-correlation analyses with Ly $\alpha$  forest absorbers in order to measure the DLA bias,  $b_{\text{DLA}}$  (Font-Ribera et al. 2012; Pérez-Ràfols et al. 2018a). Based on a large statistical analysis, Pérez-Ràfols et al. (2018a) find an average bias of  $b_{\text{DLA}} = 2.0 \pm 0.1$ , which translates to a rather large halo mass ( $\sim 10^{11} M_{\odot}$ ) if all DLAs reside in halos of the same mass; However, the inferred DLA halo mass depends on the distribution function for DLA cross-section as a function of halo mass. A power-law scaling between DLA cross-section and halo mass with index larger than unity implies that DLAs instead reside in a large range of halo masses, where the minimum mass depends critically on the assumed power-law index (Pérez-Ràfols et al. 2018a). Moreover, the inferred  $b_{\text{DLA}}$  depends on metal line strength (Pérez-Ràfols et al. 2018b) indicating that more metal-enriched DLAs reside in more massive halos, consistent with the mass-metallicity relation inferred for DLAs (Møller et al. 2013; Christensen et al. 2014).

The most direct method to examine the galaxies associated to DLAs is through direct detections (see compilation by Møller & Christensen 2020). Yet, detections of high- $z$  DLA-galaxies have been scarce due to their intrinsically faint nature (Fynbo et al. 1999; Haehnelt et al. 2000; Schaye 2001; Krogager et al. 2017). Recently,

however, powerful integral field spectrographs enable more effective follow-up of DLA-galaxies by probing large areas around the background quasar (Fumagalli et al. 2017).

Since we only observe the brightest DLA-galaxies, it is necessary to extrapolate from the individual associations to obtain the global properties of DLAs. Fynbo et al. (2008) have carried out a successful modelling approach to reproduce the  $Z$  and impact parameter distribution of DLAs at  $z \approx 3$ . Padmanabhan & Refregier (2017) have studied the  $H\text{I}$  distribution of DLAs using an analytical formalism to link halo properties to  $H\text{I}$  mass and cross-section. Their model reproduces well the redshift evolution of  $dn_{\text{DLA}}/dz$  but overproduces the number of high  $N_{\text{HI}}$  systems.

While the models by Fynbo et al. (2008) and Padmanabhan & Refregier (2017) are successful at predicting the distributions of  $Z$  and  $N_{\text{HI}}$  independently, so far there have been no attempts to model both of these key properties simultaneously. Here we therefore extend the original model by Fynbo et al. (2008) to include a statistical prescription for  $N_{\text{HI}}$  in order to describe the bivariate  $N_{\text{HI}}-Z$  distribution. We furthermore include the effects of a dust bias in optical quasar selection affecting the observed DLA properties (Pei et al. 1991; Murphy & Bernet 2016; Krogager et al. 2019). In this paper, we perform a Bayesian analysis to constrain the model parameters including priors on parameters that have already been constrained independently.

The paper is organized as follows: In Sect. 2, we describe the compilation of data used to constrain the model; in Sect. 3, we present the details of our model and the parameter estimation; We discuss the results and implications of our work in Sect. 4; and lastly, we summarize our findings in Sect. 5.

Throughout this paper, we assume a flat  $\Lambda$ CDM cosmology with  $H_0 = 68 \text{ kms}^{-1} \text{ Mpc}^{-1}$ ,  $\Omega_\Lambda = 0.69$  and  $\Omega_M = 0.31$  (Planck Collaboration et al. 2016).

## 2 LITERATURE DATA

In order to constrain our model, we use the observed  $N_{\text{HI}}$  distribution function derived by Noterdaeme et al. (2012) based on a statistical sample of  $\sim 3500$  DLAs at redshifts  $2 < z < 3$  detected in  $\sim 40000$  quasar spectra. However, only a small subset of these have robust measurements of metallicity,  $Z$ . We compile a sample of  $Z$  measurements from the literature combining the three largest samples available by Quiret et al. (2016), Jorgenson et al. (2013) and Rafelski et al. (2012). We remove measurements based on limits and those based only on iron as this element tends to deplete heavily onto dust grains. We furthermore only consider DLAs in the redshift range  $2 < z < 3$ . This sample of 178 DLAs will hereafter be referred to as the ‘full DLA sample’. The selection effects are discussed in more detail in Sect. 4. All values of metallicities are in units of Solar metallicity,  $Z_\odot$ , unless stated otherwise. We have corrected all the measurements described above to the same Solar reference values using measurements by Asplund et al. (2009) and the recommendations by De Cia et al. (2016) as to whether photospheric or meteoritic values are used.

In order to compare impact parameter predictions from our model, we use the sample of high-redshift ( $z_{\text{abs}} \gtrsim 2$ ) DLAs with confirmed emission counterparts compiled by Møller & Christensen (2020) and Krogager et al. (2017). We restrict the sample to the subset with  $\log(N_{\text{HI}} / \text{cm}^{-2}) > 20.3$ . To this sample, we add three counterparts reported by Ranjan et al. (2020) at  $z \sim 2.3$  together with one DLA counterpart by Srianand et al. (2016) at  $z = 3.247$  and one by Fumagalli et al. (2017) at  $z = 3.25$ . This sample will hereafter be referred to as the ‘DLA galaxy sample’.

## 3 MODELLING HIGH-REDSHIFT DLAS

The model described here is based on the work by Fynbo et al. (2008). We here offer a short summary of the model framework and refer the reader to the original work for further details. In this work, we only consider redshifts between  $2 < z < 3$ . The model uses a selection probability of DLAs given by  $P_{\text{DLA}} \propto \sigma_{\text{DLA}} \phi(L)$ , where  $\sigma_{\text{DLA}}$  is the effective cross-section of DLAs and  $\phi$  is the UV luminosity function. For the luminosity function, a Schechter function of the form  $\phi(L) = \phi_0 (L/L^*)^\alpha \exp(-L/L^*)$  is used.

We assume  $\sigma_{\text{DLA}}$ , given by  $\pi R_{\text{DLA}}^2$ , to scale with luminosity, through a Holmberg relation  $R_{\text{DLA}} = R_{\text{DLA}}^* (L/L^*)^t$ , and vanish below a limiting luminosity  $L_{\text{min}}$ . In what follows, all quantities marked by \* are referring to the given quantity of an  $L^*$  galaxy, e.g., the radial extent of DLA cross-section for an  $L^*$  galaxy is denoted  $R_{\text{DLA}}^*$ . The absolute value of  $R_{\text{DLA}}^*$  is obtained by requiring that the incidence rate,  $dn_{\text{DLA}}/dz$ , matches the observed value at  $z \approx 2.5$ . We calculate  $dn/dz$  as:

$$\frac{dn}{dz} = \pi R_{\text{DLA}}^{*2} \phi_0 c (1+z)^2 H^{-1}(z) \int_{L_{\text{min}}}^{\infty} L^{\alpha+2t} e^{-L/L^*} dL, \quad (1)$$

where  $L$  is in units of  $L^*$  and the Hubble parameter is given as:

$$H(z) = H_0 \sqrt{\Omega_M (1+z)^3 + \Omega_\Lambda}. \quad (2)$$

We here use the observed value of  $dn_{\text{DLA}}/dz = 0.21 \pm 0.04$  at  $z = 2.5$  (Zafar et al. 2013).

Galaxies are then sampled from the luminosity function weighted by  $\sigma_{\text{DLA}}$ , and an impact parameter,  $b$ , is drawn randomly with a probability  $P(b) \propto b$  for  $b \leq R_{\text{DLA}}(L)$ , that is, the probability is weighted by area.

A central metallicity,  $Z_0$  is assigned to each galaxy assuming a metallicity–luminosity ( $Z-L$ ) relation:  $\log(Z_0) = \log Z_0^* + \beta \times M_{\text{UV}}$ . A radial metallicity gradient is then assumed in order to obtain a value of the metallicity,  $Z_{\text{abs}}$ , at the impact parameter where the absorption system would be observed. This gradient is taken to be luminosity dependent with a variable power-law index:  $\gamma = \gamma^* L^{q_z}$ . While the original work by Fynbo et al. (2008) assumed a fixed value of  $q_z = -1$  following Boissier & Prantzos (2001), we keep this index as a free parameter in order to quantify whether a universal gradient (i.e.,  $q_z = 0$ ) is preferred over a luminosity dependent gradient.

### 3.1 Including $H\text{I}$ and dust

The neutral gas in galaxies is roughly expected to follow an exponential distribution with radius:  $N_{\text{HI}}(r) = N_{\text{HI},0} \exp(-r/r_{\text{HI}})$  (Walter et al. 2008). The scale length,  $r_{\text{HI}}$ , is calculated by demanding that  $N_{\text{HI}}(r = R_{\text{DLA}}) = 2 \times 10^{20} \text{ cm}^{-2}$ . For this reason,  $r_{\text{HI}}$  is not a free parameter in this model. The central  $N_{\text{HI}}$  value is however kept as a free variable with an adopted fiducial value of  $N_{\text{HI},0} = 10^{22} \text{ cm}^{-2}$ .

Motivated by observations of local  $H\text{I}$  discs by the THINGS survey (Walter et al. 2008), we include a stochastic term in the radial  $N_{\text{HI}}$  distribution to account for local fluctuations in column density as well as inclination effects which are not explicitly modelled. The fluctuations are implemented as a log-normal scatter around the smooth average radial profile:

$$\log N_{\text{HI}}(r) = \log N_{\text{HI},0} - \frac{\log(e)}{r_{\text{HI}}} r + \mathcal{N}(0, \sigma_{\text{HI}}),$$

where the log-normal scatter  $\sigma_{\text{HI}}$  is a free variable in our model with a fiducial value of 0.3 dex.

Since we now have a prescription for both  $N_{\text{HI}}$  and  $Z$ , we can calculate the expected amount of optical dust extinction along

the absorption sightline,  $A(V)$ . This value is obtained following Zafar & Møller (2019) assuming a constant dust-to-metals ratio,  $\log \kappa_Z = -21.4$ .

Lastly, we include a dust bias to account for the fact that quasars behind dusty DLAs are systematically under-represented due to the complex colour and magnitude selection criteria. We calculate the selection probability as  $P_{\text{QSO}} = \text{sech}(x^2)$ , where  $x = A(V)/A(V)_{\text{crit}}$ . This functional form fits very well the calculated selection probability by Krogager et al. (2019) for a value of  $A(V)_{\text{crit}} = 0.25$  mag. The model distributions are filtered according to  $P_{\text{QSO}}$  to produce a mock observable model distribution.

In total, the original model contains 9 free parameters:  $\{\phi_0, M_{\text{UV}}^*, \alpha, L_{\text{min}}, t, \beta, \gamma^*, Z^*, q_Z\}$ , and with the above modifications, we have effectively added the following 3 parameters:  $\{N_{\text{HI},0}, \sigma_{\text{HI}}, A(V)_{\text{crit}}\}$ .

### 3.2 Constraining model parameters

Due to the significant degeneracies in the parameters we constrain the model parameters using a Bayesian approach. This also allows us to include priors since we have independent constraints on many parameters. For this purpose, we use the Python package EMCEE (Foreman-Mackey et al. 2013).

We use the observed distributions of  $N_{\text{HI}}$  and  $Z$  (see Sect. 2) to statistically constrain the model parameters. For a given set of parameters, we draw a large sample of 50,000 DLAs and then calculate the model  $f(N_{\text{HI}})$  in the same bins as the data, using only DLAs that pass the mock quasar selection as implemented here using  $P_{\text{QSO}}$  (see Sect. 3.1). The distribution function is normalized by requiring that the integral  $\int_{N_{\text{DLA}}}^{\infty} f(N_{\text{HI}}) dN$  matches the observed value of  $\int_{N_{\text{DLA}}}^{\infty} f(N_{\text{HI}}) dN$ , where  $N_{\text{DLA}} = 2 \times 10^{20} \text{ cm}^{-2}$ . This normalization ensures a correct absolute scaling of  $f(N_{\text{HI}})$  in order to match the observed  $dn_{\text{DLA}}/dz$ . We then calculate the likelihood assuming Gaussian statistics given the uncertainties quoted by Noterdaeme et al. (2012).

Similarly, we obtain a model distribution for  $Z$  which we compare to the observed distribution using a Kolmogorov–Smirnov (KS) test. Although the  $p$ -value from a KS test is not an exact estimator of the formal likelihood, the two are sufficiently correlated (Krueger & Heck 2017) allowing us to use  $P_{\text{KS}}$  as an estimate of the likelihood. In our case we find that the  $p$ -value provides tighter constraints than other likelihood estimators (such as kernel density estimators).

Along with the constraints from the  $N_{\text{HI}}$  and  $Z$  distributions, we have the following independent constraints:

(i) The average reddening for DLAs, which pass the optical quasar selection criteria,  $\langle E(B - V) \rangle_{\text{obs}}$ , must not exceed 21 mmag at the  $3\text{-}\sigma$  level (Murphy & Liske 2004). We use this conservative upper limit since there is significant disagreement among various measurements (see Murphy & Bernet 2016);

(ii) The average metallicity of extremely strong DLAs (ESDLAs,  $\log(N_{\text{HI}}) > 21.7$ , Noterdaeme et al. 2014) is observed to be  $\langle Z \rangle = -1.30 \pm 0.05$  (Ranjan et al. 2020).

The joint likelihood is then taken as the product of the independent likelihoods taking into account the higher number of degrees of freedom for the  $N_{\text{HI}}$  data.

#### 3.2.1 Priors

All priors used in our statistical analysis are summarized in Table 1. For parameters where we have no prior knowledge we use flat priors

over a reasonable range of parameter space. We have verified that the choice of prior ranges do not affect the results and all values are constrained well within the chosen ranges. For the parameters with more restrictive priors, the details of the priors are given below.

We constrain the shape of the luminosity function following observations from Malkan et al. (2017). The parameters  $M_{\text{UV}}^* = -20.9$  and  $\phi_0 = 1.7 \times 10^{-3} \text{ Mpc}^{-3}$  are kept fixed as they agree very well from one survey to another (Malkan et al. 2017, table 3). On the other hand, we choose to keep  $\alpha$  as a free parameter since this value shows large dispersion among various surveys. The average value and the standard deviation are used as a prior on  $\alpha$ . In this work, we adopt a fiducial value of  $L_{\text{min}} = 10^{-4} L^*$ . This value is consistent with numerical simulations (Bird et al. 2013) and semi-analytical modelling (Dvorkin et al. 2015).

The slope of the  $Z - L$  relation,  $\beta$ , is constrained from observations of galaxies at  $z \sim 0.5 - 1$  (Kobulnicky & Kewley 2004; Hidalgo 2017). We obtain an average value of  $\langle \beta \rangle = 0.21^1$  (weighted by individual uncertainties) with a standard deviation of 0.05. This average value is taken as our prior and is in agreement with results from previous modelling (Krogager et al. 2017). Although the redshift range studied by Hidalgo (2017) is lower than what we try to model here, there is evidence that the slope of the related mass–metallicity relation does not evolve significantly with redshift (Maiolino et al. 2008). It is therefore reasonable to assume that the slope of the metallicity–luminosity relation would also remain constant with redshift.

The normalization of the  $Z - L$  relation,  $Z^*$ , is constrained from observations of the mass–metallicity relation at  $z \sim 2.2$  (Maiolino et al. 2008). We find that galaxies around  $M^*$  have roughly Solar metallicity. In order to take into account the observed scatter as well as systematics, we use a weak prior on  $Z^*$ :  $\log Z^* = 0.0 \pm 0.2$ .

We use a fiducial value of  $A(V)_{\text{crit}} = 0.25$  mag, derived using the calculation of selection probability as a function of  $A(V)$  from Krogager et al. (2019). The value of 0.25 mag corresponds to a limiting  $A(V)$  of 0.57 mag for a selection probability of  $P_{\text{QSO}} = 0.01$ , i.e., DLAs with  $A(V)$  larger than 0.57 mag have a selection probability less than 1% in SDSS-II (up until DR7). Since the ‘full DLA sample’ is observed with larger telescopes than the SDSS this  $A(V)$  limit may differ with respect to the fiducial value. We therefore keep this value as a free parameter and assign a rather arbitrary logarithmic uncertainty for the prior of 0.3 dex. The results do not depend strongly on the chosen width of the prior distribution.

Using the priors mentioned above, we obtain an initial estimate of the parameters using EMCEE to explore parameter space with 100 walkers for 600 steps. The posterior probability distribution shows a strong one-to-one anti-correlation between the parameters  $t$  and  $q_Z$  with a Spearman correlation coefficient of  $-0.92$ . A similar anti-correlation is implemented in the original model by Fynbo et al. (2008) following Boissier & Prantzos (2001), who use  $q_Z = -t$ . Based on the observed anti-correlation we adopt the constraint:  $q_Z = -t$ . Hence,  $q_Z$  is no longer considered a free parameter.

Although the average metallicity gradient of DLA hosts has been inferred by Christensen et al. (2014), we do not include this constraint as a prior on  $\gamma^*$ . Since these authors have assumed a constant gradient with no luminosity dependence and no separation between low- and high redshift DLA galaxies, including their obtained metallicity gradient as a prior in this work could possibly bias our results.

<sup>1</sup> Note the change of sign in our definition with respect to Hidalgo (2017).

**Table 1.** Summary of free model parameters

Parameter	Prior <sup>a</sup>	Best-fit <sup>b</sup>	Ref.
$\alpha$	$-1.70 \pm 0.20$	$-1.70^{+0.20}_{-0.20}$	(1)
$t = -q_z$	$[-5, +5]$	$0.51^{+0.27}_{-0.17}$	(2)
$\beta$	$0.21 \pm 0.05$	$0.20^{+0.04}_{-0.03}$	(2)
$\gamma^* / \text{dex kpc}^{-1}$	$[-1, +1]$	$-0.019^{+0.008}_{-0.008}$	(3)
$\log Z_0^* / Z_\odot$	$0.0 \pm 0.2$	$-0.02^{+0.18}_{-0.18}$	(3)
$\log(N_{\text{H I}, 0} / \text{cm}^{-2})$	$[20, 24]$	$20.91^{+0.29}_{-0.27}$	(3)
$\sigma_{\text{H I}}$	$[0, 2]$	$0.54^{+0.06}_{-0.09}$	(3)
$\log A(V)_{\text{crit}}$	$-0.6 \pm 0.3$	$-0.55^{+0.22}_{-0.20}$	(4)
Fixed parameters:		Value	Ref.
$\phi_0 / \text{Mpc}^{-3}$		$1.7 \times 10^{-3}$	(5)
$M_{\text{UV}}^*$		$-20.9$	(5)
$L_{\text{min}} / L^*$		$10^{-4}$	(5)
$\log \kappa_z$		$-21.4$	(6)

<sup>a</sup> Gaussian priors are given as  $\mu \pm \sigma$ ; Flat priors are given as  $[\text{min}, \text{max}]$ .

<sup>b</sup> Best-fit values are stated as the median value with 16-th and 84-th percentiles as confidence intervals.

**References for priors:** (1) Malkan et al. (2017); (2) Hidalgo (2017); (3) Maiolino et al. (2008); (4) Krogager et al. (2019); (5) Malkan et al. (2017); (6) Zafar & Möller (2019).

### 3.3 Results

We obtain the best-fit solution using EMCEE with 100 walkers for 800 iterations of which we discard the first 200 iterations for which the ensemble has not converged. The values of the optimized model parameters are given in Table 1. These are stated as the median value of the posterior probability distribution together with the 16th and 84th percentiles as  $1-\sigma$  confidence intervals.

The results of the best-fit model are shown in Fig. 1. We find a very good agreement between the model and the data. For comparison, we also show the impact parameter distribution as a function of  $N_{\text{H I}}$  and  $Z$  in Fig. 2. Since the ‘DLA galaxy’ sample is not complete and suffers from strong and inhomogeneous selection effects (mainly high-metallicity galaxies have been targeted and identified), we do not include the impact parameters in the formal modelling. Nonetheless, it is interesting to compare the model predictions to the observations. We find that the observations indeed overlap with the model predictions which lends qualitative support to the best-fit model, in particular since these observations have not been used to constrain the model parameters.

## 4 DISCUSSION

### 4.1 Radial distribution of $N_{\text{H I}}$

The best-fit value of  $N_{\text{H I}, 0} = 8_{-4}^{+8} \times 10^{20} \text{ cm}^{-2}$  is consistent with local  $\text{H I}$  observations from the THINGS survey who report typical values of  $\sim 10 M_\odot \text{ pc}^{-2}$  corresponding to  $\sim 10^{21} \text{ cm}^{-2}$  (Walter et al. 2008). The inferred amount of scatter in  $\log(N_{\text{H I}})$  is high,  $\sigma_{\text{H I}} = 0.54^{+0.07}_{-0.09}$  dex, compared to the fairly smooth radial profiles presented by Walter et al. (2008). This is expected since the

individual radial profiles presented by Walter et al. (2008) have been azimuthally averaged. Moreover, the  $\text{H I}$ -emission studies provide beam-averaged measurements (typically 100 – 500 pc for the THINGS galaxies) which smoothes out small-scale structure. The very small scales probed by quasar sightlines ( $\lesssim 1$  pc; e.g. Balashev et al. 2011) may therefore show much larger local variations. A large degree of randomness in  $N_{\text{H I}}$  is also expected since the neutral medium is highly turbulent (Elmegreen & Scalo 2004) and our model samples the whole galaxy population, not just a single galaxy.

### 4.2 DLA impact parameters

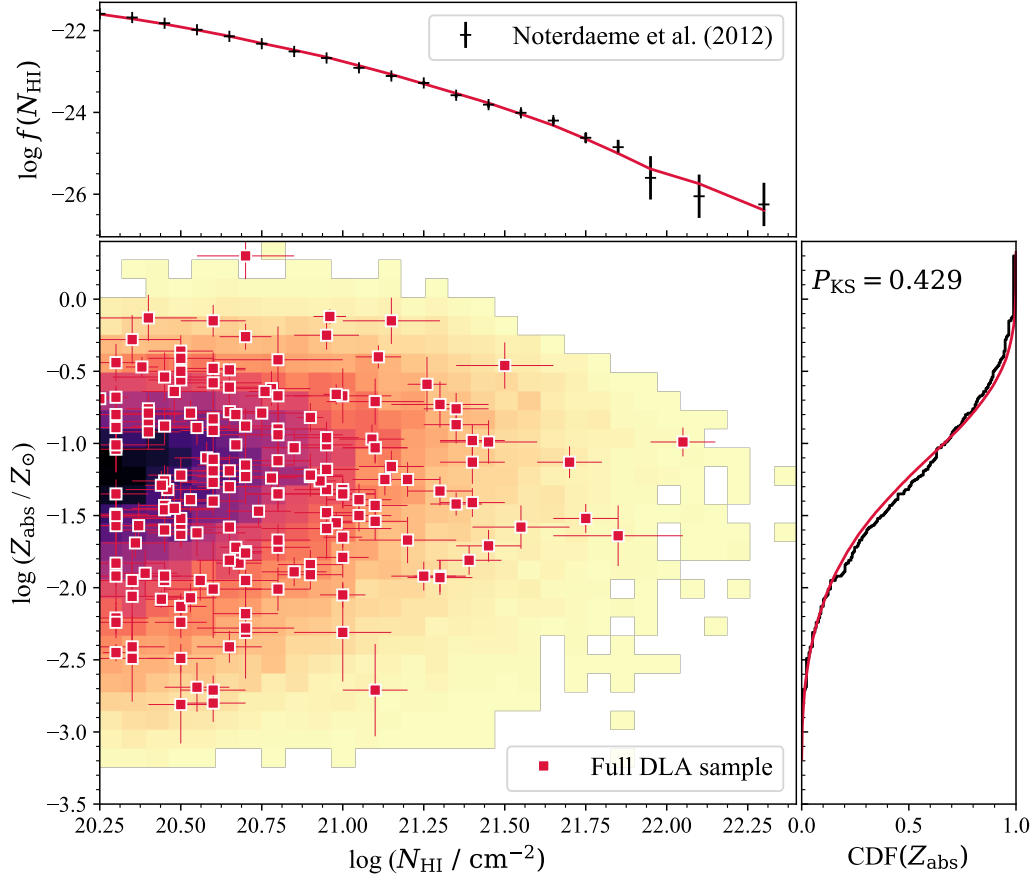
The distribution of  $N_{\text{H I}}$  and impact parameter is in qualitative agreement with the simulation by Rahmati & Schaye (2014). In order to make a fair comparison, we restrict our model to only consider DLA hosts with similar star formation rates (SFRs) as Rahmati & Schaye (SFR  $> 0.004$ ). For this purpose, we calculate SFR based on the UV luminosity included in our model following Kennicutt (1998), i.e.,  $\text{SFR} \propto L_{\text{UV}}$ . We find that the median impact parameter increases when looking at DLA hosts with larger SFRs (see top panel of Fig. 3). This is similar to the simulations by Rahmati & Schaye (2014), yet the median impact parameters from our best-fit model are  $\sim 2$  times larger. The simulations by Rahmati & Schaye (2014) do not provide explicit information on the absorption metallicity and it is therefore not certain whether they match the bivariate  $N_{\text{H I}}-Z$  distribution.

Our results are in better agreement with the recent simulation by Rhodin et al. (2019), who find larger impact parameters for high-redshift absorbers. The observed anti-correlation between  $\log(N_{\text{H I}})$  and impact parameter is however not recovered at  $z > 1$  in the simulation by Rhodin et al. (2019). Their simulation only addresses one Milky Way type progenitor, which is more massive than the average DLA host in our work, and it is therefore difficult to perform a one-to-one comparison between their simulation and our work.

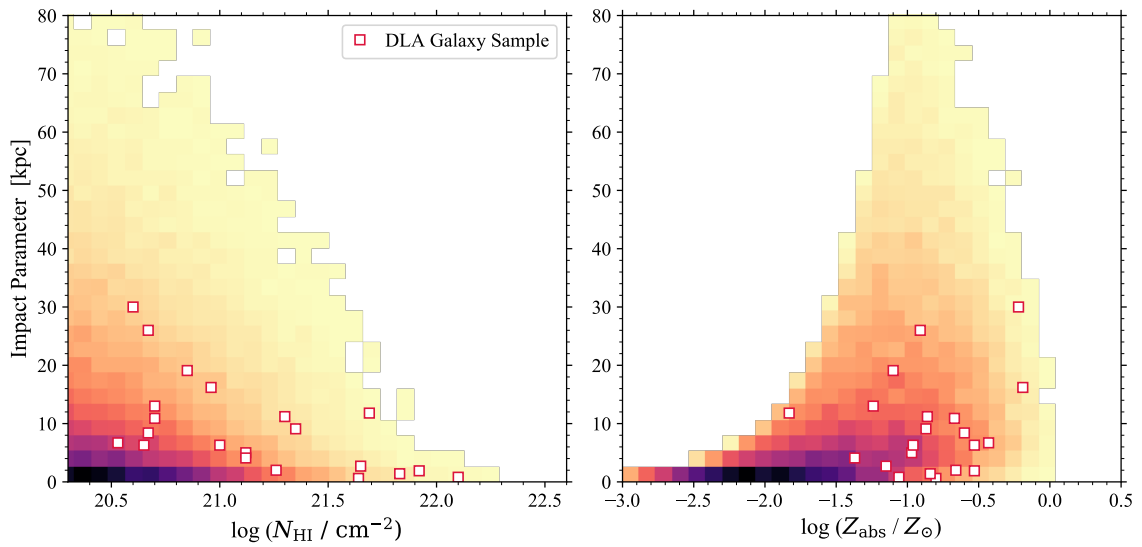
We have also investigated how the metallicity–impact parameter distribution varies when considering only DLA hosts with SFR larger than  $0.2 M_\odot \text{ yr}^{-1}$ . This limit corresponds broadly to the SFR limit obtained in the ‘DLA galaxy sample’. In the lower panel of Fig. 3, we show the distribution of impact parameters as a function of absorber metallicity for DLA hosts with SFR larger than  $0.2 M_\odot \text{ yr}^{-1}$ . By restricting the model distribution to this SFR limit we obtain a good match to the observations. However, it is not possible to quantify the agreement in more detail given the inhomogeneous sample selection of the DLA galaxy sample, combined with the fact that the detections are based on different emission lines (e.g.,  $\text{Ly}\alpha$ ,  $\text{H}\alpha$ ,  $[\text{O III}]$ ) yielding inhomogeneous detection limits.

### 4.3 DLA cross-section and halo properties

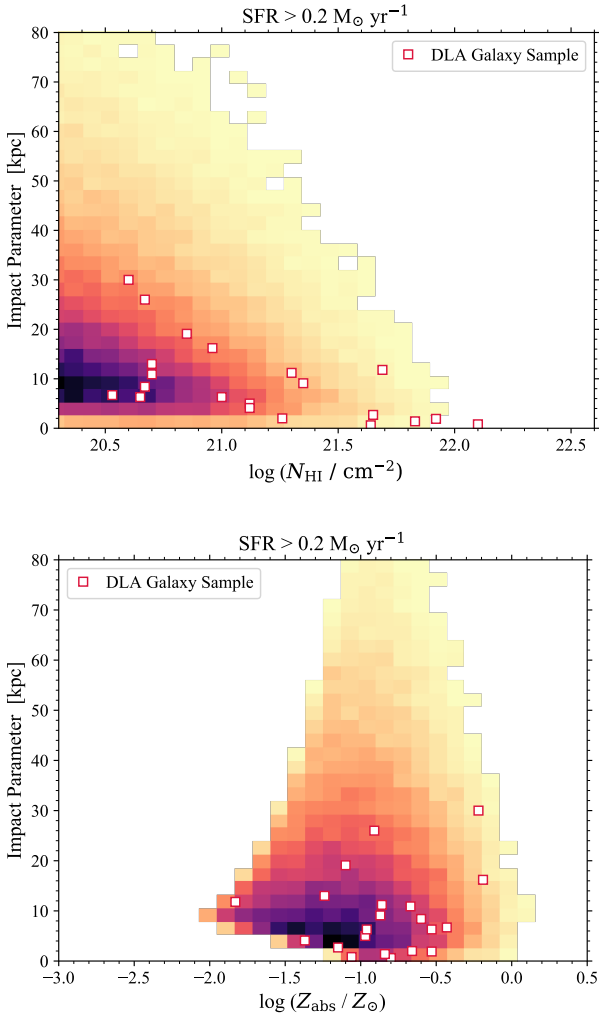
In the following, we shall compare the cross-section of DLAs,  $\sigma_{\text{DLA}}$ , as a function of their host luminosity to results from numerical simulations. Bird et al. (2014) find that  $\sigma_{\text{DLA}}$  as a function of halo mass is well-reproduced by a power-law with an index between 0.8 and 1. The upper range of their results is in good agreement with the luminosity scaling we infer for  $\sigma_{\text{DLA}}$  of 1.02 ( $\sigma_{\text{DLA}} \propto L^{2t}$ , and  $t = 0.51$ ) assuming a constant mass-to-light ratio. The normalization of the power-law relation inferred by Bird et al. (2014) is  $\sim 100 \text{ kpc}^2$  at  $M_{\text{h}} = 10^{10} M_\odot$ . Pontzen et al. (2008) find a steeper relation for  $\sigma_{\text{DLA}}$  as function of halo mass for low masses which flattens at masses above  $10^{10} M_\odot$  and the DLA cross-section at  $10^{10} M_\odot$  in



**Figure 1.** Model prediction for the bivariate  $N_{\text{HI}}$  and  $Z_{\text{abs}}$  distribution.  $Z_{\text{abs}}$  here refers to the metallicity at the given impact parameter in contrast to the central metallicity,  $Z_0$ , that would be probed by emission line measures. The colors of the model distribution indicates the number of model points in the given bin normalized to a linear scale from 0 to 1. The marginalized  $Z_{\text{abs}}$  distribution is shown in the right panel as the cumulative distribution function (CDF, in black) together with the best-fit model (in red).



**Figure 2.** Model prediction for impact parameter as a function of  $N_{\text{HI}}$  (left) and  $Z_{\text{abs}}$  (right) for all model galaxies ( $\text{SFR} > 0 M_{\odot} \text{yr}^{-1}$ ). The color scale of the model distributions follows that of Fig. 1. The sample of DLA galaxies is shown as open, square points. We note that the ‘DLA galaxy’ sample is neither complete nor representative and hence should only be compared qualitatively to the underlying model distribution.



**Figure 3.** Model prediction for impact parameter as a function of  $N_{\text{HI}}$  (top) and  $Z_{\text{abs}}$  (bottom) considering only DLA hosts with  $\text{SFR} > 0.2 M_{\odot} \text{ yr}^{-1}$ .

their simulation is  $\sim 50 \text{ kpc}^2$ . As our analysis is based on the DLA galaxy luminosity, we need to assume a halo-mass-to-light ratio in order to compare our model to the simulations. The best-fit model yields a characteristic scale of DLA cross-section for a  $L^*$  galaxy of  $R_{\text{DLA}}^* = 31 \text{ kpc}$ . In order to reproduce the simulations by Pontzen et al. (2008) and Bird et al. (2014), we therefore need to assume  $M_{\text{h}}/L_{\text{UV}} \sim 20 - 40 M_{\odot}/L_{\odot}$  for  $L^*$  galaxies. This  $M_{\text{h}}/L_{\text{UV}}$  ratio is consistent with what is found in the literature (Vale & Ostriker 2006; Mason et al. 2015).

We then compare the distribution of DLA galaxy luminosities from our model to the halo mass distribution from the simulations by Bird et al. (2014). For our best-fit model, we find that the bulk of the DLA cross-section is contributed by galaxies around  $0.1 L^*$ . Assuming the average inferred  $M_{\text{h}}/L_{\text{UV}} = 30$ , we obtain a bulk halo mass for DLAs of  $6 \times 10^{10} M_{\odot}$ . This is consistent with the peak of the halo-mass distribution presented by Bird et al. (2014, their fig. 2). The minimum halo mass contributing to the DLA cross-section in the simulations by Bird et al. (2014) is  $M_{\text{min}} \approx 10^9 M_{\odot}$  consistent with modelling studies of the halo properties of DLAs (e.g., Pontzen et al. 2008; Barnes & Haehnelt 2009, 2014; Padmanabhan & Refregier 2017). Using the  $M_{\text{h}}-L_{\text{UV}}$  relation by Mason et al. (2015), we find that a minimum halo mass of  $10^9 M_{\odot}$

corresponds to  $L_{\text{min}} = 4 \times 10^{-5} L^*$ . This is in good agreement with the fiducial value of  $L_{\text{min}} = 10^{-4} L^*$  assumed in this work, when taking the significant scatter of the halo mass relations into account.

The halo properties of DLA galaxies can furthermore be studied by analysing the cross-correlation of DLAs and  $\text{Ly}\alpha$  forest absorbers (Font-Ribera et al. 2012; Pérez-Ràfols et al. 2018a). Based on SDSS DR12, Pérez-Ràfols et al. (2018a) find that the observed bias is consistent with a minimum halo mass of  $M_{\text{min}} \sim 10^9 M_{\odot}$  if the DLA cross-section scales with halo mass as a power-law with index  $a \approx 1.05$ . This might be slightly at odds with the lower range of the power-law index inferred by Bird et al. (2014) although not ruled out. It is on the other hand consistent with the results of our model within the rather large uncertainty on  $t$ .

#### 4.4 Metallicity gradients

The best-fit value of the metallicity gradient for  $L^*$  galaxies,  $\gamma^* = -0.019 \pm 0.008 \text{ dex kpc}^{-1}$ , is in good agreement with the previous measurements of the average metallicity gradient for DLAs of  $\gamma = -0.022 \pm 0.004$  (Christensen et al. 2014). This agreement is consistent with the fact that most of the high-redshift galaxies analysed by Christensen et al. (2014) are fairly bright and have luminosities around  $L^*$ . Similar estimates are reported by Péroux et al. (2012), although the scatter among individual measurements is significant. When studying emission-selected galaxies at  $z \gtrsim 1$ , a large range in metallicity gradients has been observed in galaxy discs (e.g., Swinbank et al. 2012; Stott et al. 2014; Curti et al. 2020). Curti et al. (2020) report flat or negative gradients for the majority of galaxies, and only in a few cases do they observe inverted metallicity gradients (i.e., more metal-rich at larger radii). The authors find a slight trend of steeper, negative metallicity gradients for more massive galaxies, contrary to the relation found in our model, where the metallicity gradient flattens for more luminous (and thus massive) galaxies;  $\gamma \propto L^{-0.5}$ . This disagreement might however be a result of sample selection effects together with the very different ways by which metallicity gradients are measured in absorption and emission as well as the physical scales they probe.

The study by Curti et al. (2020) has very few galaxies at  $z > 2$ , and all of these are highly star-forming ( $\text{SFR} \sim 50 M_{\odot} \text{ yr}^{-1}$ ) and rather massive ( $M_{\star} \sim 10^{10} M_{\odot}$ ). DLA hosts in our model probe faint galaxies with low star-formation activity corresponding to stellar masses in the range of  $\lesssim 10^9 M_{\odot}$ . It is thus plausible that the differences derived in the luminosity (or mass) dependence is due to the very different sample characteristics in terms of stellar mass. We furthermore note that while luminosity and mass are correlated, it is not a one-to-one correspondence due to variations in star-formation histories and dust attenuation.

Beyond sample selection effects, the emission samples further present a mix of various emission line diagnostics which have complicated systematic effects (e.g., Kewley & Ellison 2008). In studies combining emission and absorption, similar systematic effects come into play (see Péroux et al. 2012; Rahmani et al. 2016). Lastly, the observations of high-redshift galaxies in emission only probe the inner few kpc of the brightest galaxies. In contrast, our analysis takes into account the average metallicity gradient of the whole DLA galaxy population out to large distances. We thus conclude that the metallicity gradient for DLA galaxies included in our model is not directly comparable to the observed emission-line-derived metallicity gradients, and any differences might therefore be ascribed to differences in sample selection and methodology. Further investigation is needed to analyse these systematics in detail; However, this is beyond the scope of this work.

Both studies by Curti et al. (2020) and Stott et al. (2014) report a correlation between the metallicity gradient and the specific star formation rate, indicating that more vigorously star-forming galaxies have flatter gradients (or even inverted gradients). Isolating more regular galaxies on the so-called ‘star formation main sequence’, Stott et al. (2014) find an average metallicity gradient of  $\langle \gamma \rangle = -0.020 \pm 0.004$  in surprising agreement with our results. This might indicate that the bulk of DLAs does not probe strongly star-bursting galaxies, consistent with the low star formation activity observed in direct detections of high-redshift DLAs (Krogager et al. 2017; Rhodin et al. 2018).

#### 4.5 Completeness of observations

As eluded to above, one complication in our modelling is the complex selection effects of the observational data. The  $N_{\text{HI}}$  data by Noterdaeme et al. (2012) are fairly complete and homogeneous based on BOSS data release 9. These data are limited by a signal-to-noise criterion in order to have enough signal in the Ly $\alpha$ -forest to detect the DLAs. This translates to an effective magnitude limit in the  $g$  band. On the other hand, the  $Z$  measurements are very heterogeneous as they are obtained by different follow-up campaigns which are often preselected on different and poorly quantified criteria. However, in order to obtain high-resolution data with current 8–10 m class telescopes, the background quasars are required to be brighter than  $r \lesssim 20$ . Both of these effective magnitude limits on the  $N_{\text{HI}}$  and  $Z$  samples are well-reproduced by the selection probability as implemented in this work. Yet, we caution that the value of  $A(V)_{\text{crit}}$  might be different for the  $N_{\text{HI}}$  and  $Z$  samples. The only way to properly overcome these selection effects is by obtaining large and homogeneous samples of DLAs with measurements of both  $N_{\text{HI}}$  and  $Z$ .

## 5 SUMMARY

In this work, we have presented an extension of the model by Fynbo et al. (2008) with the aim of reproducing the joint distribution of  $N_{\text{HI}}$  and  $Z_{\text{abs}}$  for DLAs at redshifts  $z = 2 - 3$ . The model assumes that the galaxies giving rise to DLAs are drawn from the population of star-forming galaxies following the UV luminosity function. The effective DLA cross-section,  $\sigma_{\text{HI}}$ , around each galaxy is assumed to scale with the luminosity of the galaxy:  $\sigma_{\text{HI}} = \pi R_{\text{DLA}}^2$ , where  $R_{\text{DLA}}$  scales with luminosity as  $R_{\text{DLA}} \propto L^l$ . Furthermore, the galaxies are assumed to follow a metallicity–luminosity relation and exhibit an average radial metallicity gradient.

We have included a simple prescription for the radial column density profile of HI. We have found that a log-normal scatter around this average radial HI profile is needed in order to match the high  $N_{\text{HI}}$  tail of the distribution. We furthermore include a selection bias due to dust obscuration in optically selected quasar samples as quantified by Krogager et al. (2019). The model contains 8 free parameters (and 4 parameters which are kept fixed at their assumed fiducial values) summarized in Table 1. In order to constrain these parameters, we use an MCMC sampler to obtain the posterior probabilities.

The best-fit model provides a good fit to the data as seen in Fig. 1 and we find that the modelled distribution of impact parameters agrees well with observations even though these were not used to constrain the model. This agreement is highlighted when considering only model galaxies that would be bright enough in the UV to be detected with current facilities (see Fig. 3).

By converting the UV luminosities in our analysis to halo masses (following Mason et al. 2015), we compare the distribution of DLA galaxy luminosities in our model to the halo mass distribution seen in numerical simulations (Pontzen et al. 2008; Bird et al. 2014). We find that the model distribution is consistent with numerical simulations in terms of cross-section and its scaling with luminosity (halo mass) as well as the lower luminosity (mass) limit for DLA cross-section of  $\sim 10^{-4} L^*$  ( $\sim 10^9 M_{\odot}$ ).

Lastly, we find strong evidence for a negative radial metallicity gradient which scales inversely with luminosity, i.e., more luminous galaxies have flatter gradients. While this trend with luminosity is somewhat at odds with the mass dependence seen in observations (e.g., Curti et al. 2020), the average value of the gradient for  $L^*$  galaxies,  $\gamma^* = -0.019 \pm 0.008 \text{ dex kpc}^{-1}$ , is in agreement with other works (Christensen et al. 2014; Stott et al. 2014).

## ACKNOWLEDGEMENTS

We thank the anonymous referee for the very constructive report. We would like to thank Matt Lehnert for fruitful discussions. The research leading to these results has received funding from the French *Agence Nationale de la Recherche* under grant no ANR-17-CE31-0011-01 (project ‘‘HIH2’’ – PI Noterdaeme).

## REFERENCES

- Altay G., Theuns T., Schaye J., Booth C. M., Dalla Vecchia C., 2013, *MNRAS*, **436**, 2689
- Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, *ARA&A*, **47**, 481
- Balashev S. A., Petitjean P., Ivanchik A. V., Ledoux C., Srianand R., Noterdaeme P., Varshalovich D. A., 2011, *MNRAS*, **418**, 357
- Barnes L. A., Haehnelt M. G., 2009, *MNRAS*, **397**, 511
- Barnes L. A., Haehnelt M. G., 2014, *MNRAS*, **440**, 2313
- Bird S., Vogelsberger M., Sijacki D., Zaldarriaga M., Springel V., Hernquist L., 2013, *MNRAS*, **429**, 3341
- Bird S., Vogelsberger M., Haehnelt M., Sijacki D., Genel S., Torrey P., Springel V., Hernquist L., 2014, *MNRAS*, **445**, 2313
- Bird S., Haehnelt M., Neeleman M., Genel S., Vogelsberger M., Hernquist L., 2015, *MNRAS*, **447**, 1834
- Bird S., Garnett R., Ho S., 2017, *MNRAS*, **466**, 2111
- Boissier S., Prantzos N., 2001, *MNRAS*, **325**, 321
- Christensen L., Møller P., Fynbo J. P. U., Zafar T., 2014, *MNRAS*, **445**, 225
- Curti M., et al., 2020, *MNRAS*, **492**, 821
- De Cia A., Ledoux C., Mattsson L., Petitjean P., Srianand R., Gavignaud I., Jenkins E. B., 2016, *A&A*, **596**, A97
- De Cia A., Ledoux C., Petitjean P., Savaglio S., 2018, *A&A*, **611**, A76
- Dvorkin I., Silk J., Vangioni E., Petitjean P., Olive K. A., 2015, *MNRAS*, **452**, L36
- Elmegreen B. G., Scalo J., 2004, *ARA&A*, **42**, 211
- Font-Ribera A., et al., 2012, *J. Cosmology Astropart. Phys.*, **11**, 59
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, **125**, 306
- Fumagalli M., et al., 2017, *MNRAS*, **471**, 3686
- Fynbo J. U., Møller P., Warren S. J., 1999, *MNRAS*, **305**, 849
- Fynbo J. P. U., Prochaska J. X., Sommer-Larsen J., Dessauges-Zavadsky M., Møller P., 2008, *ApJ*, **683**, 321
- Haehnelt M. G., Steinmetz M., Rauch M., 2000, *ApJ*, **534**, 594
- Hassan S., Finlator K., Davé R., Churchill C. W., Prochaska J. X., 2020, *MNRAS*, **492**, 2835
- Hidalgo S. L., 2017, *A&A*, **606**, A115
- Jorgenson R. A., Murphy M. T., Thompson R., 2013, *MNRAS*, **435**, 482
- Kennicutt Jr. R. C., 1998, *ARA&A*, **36**, 189
- Kewley L. J., Ellison S. L., 2008, *ApJ*, **681**, 1183
- Kobulnicky H. A., Kewley L. J., 2004, *ApJ*, **617**, 240



- Krogager J.-K., Møller P., Fynbo J. P. U., Noterdaeme P., 2017, *MNRAS*, **469**, 2959
- Krogager J.-K., Fynbo J. P. U., Møller P., Noterdaeme P., Heintz K. E., Pettini M., 2019, *MNRAS*, **486**, 4377
- Krueger J. I., Heck P. R., 2017, *Frontiers in Psychology*, **8**, 908
- Maiolino R., et al., 2008, *A&A*, **488**, 463
- Malkan M. A., et al., 2017, *ApJ*, **850**, 5
- Mason C. A., Trenti M., Treu T., 2015, *ApJ*, **813**, 21
- Møller P., Christensen L., 2020, *MNRAS*, p. 120
- Møller P., Fynbo J. P. U., Ledoux C., Nilsson K. K., 2013, *MNRAS*, **430**, 2680
- Murphy M. T., Bernet M. L., 2016, *MNRAS*, **455**, 1043
- Murphy M. T., Liske J., 2004, *MNRAS*, **354**, L31
- Noterdaeme P., Petitjean P., Ledoux C., Srianand R., 2009, *A&A*, **505**, 1087
- Noterdaeme P., et al., 2012, *A&A*, **547**, L1
- Noterdaeme P., Petitjean P., Pâris I., Cai Z., Finley H., Ge J., Pieri M. M., York D. G., 2014, *A&A*, **566**, A24
- Padmanabhan H., Refregier A., 2017, *MNRAS*, **464**, 4008
- Pei Y. C., Fall S. M., Bechtold J., 1991, *ApJ*, **378**, 6
- Pérez-Ràfols I., et al., 2018a, *MNRAS*, **473**, 3019
- Pérez-Ràfols I., Miralda-Escudé J., Arinyo-i-Prats A., Font-Ribera A., Mas-Ribas L., 2018b, *MNRAS*, **480**, 4702
- Péroux C., Bouché N., Kulkarni V. P., York D. G., Vladilo G., 2012, *MNRAS*, **419**, 3060
- Planck Collaboration et al., 2016, *A&A*, **594**, A13
- Pontzen A., et al., 2008, *MNRAS*, **390**, 1349
- Prochaska J. X., Wolfe A. M., 2009, *ApJ*, **696**, 1543
- Quirot S., et al., 2016, *MNRAS*, **458**, 4074
- Rafelski M., Wolfe A. M., Prochaska J. X., Neeleman M., Mendez A. J., 2012, *ApJ*, **755**, 89
- Rahmani H., et al., 2016, *MNRAS*, **463**, 980
- Rahmati A., Schaye J., 2014, *MNRAS*, **438**, 529
- Ranjan A., Noterdaeme P., Krogager J. K., Petitjean P., Srianand R., Balashev S. A., Gupta N., Ledoux C., 2020, *A&A*, **633**, A125
- Rhodin N. H. P., Christensen L., Møller P., Zafar T., Fynbo J. P. U., 2018, *A&A*, **618**, A129
- Rhodin N. H. P., Agertz O., Christensen L., Renaud F., Fynbo J. P. U., 2019, *MNRAS*, **488**, 3634
- Schaye J., 2001, *ApJ*, **559**, L1
- Srianand R., Hussain T., Noterdaeme P., Petitjean P., Krühler T., Japelj J., Pâris I., Kashikawa N., 2016, *MNRAS*, **460**, 634
- Stott J. P., et al., 2014, *MNRAS*, **443**, 2695
- Swinbank A. M., Sobral D., Smail I., Geach J. E., Best P. N., McCarthy I. G., Crain R. A., Theuns T., 2012, *MNRAS*, **426**, 935
- Vale A., Ostriker J. P., 2006, *MNRAS*, **371**, 1173
- Viegas S. M., 1995, *MNRAS*, **276**, 268
- Walter F., Brinks E., de Blok W. J. G., Bigiel F., Kennicutt Robert C. J., Thornley M. D., Leroy A., 2008, *AJ*, **136**, 2563
- Wolfe A. M., Gawiser E., Prochaska J. X., 2005, *ARA&A*, **43**, 861
- Zafar T., Møller P., 2019, *MNRAS*, **482**, 2731
- Zafar T., Péroux C., Popping A., Milliard B., Deharveng J. M., Frank S., 2013, *A&A*, **556**, A141
- Zwaan M. A., van der Hulst J. M., Briggs F. H., Verheijen M. A. W., Ryan-Weber E. V., 2005, *MNRAS*, **364**, 1467