Damped Ly α absorbers and atomic hydrogen in galaxies: the view of the GAEA model

Serafina Di Gioia [1,2,3,4]★ Stefano Cristiani,[2,3,4] Gabriella De Lucia [2] and Lizhi Xie [5]

[1]*Dipartimento di Fisica dell' Università di Trieste, Sezione di Astronomia, via Tiepolo 11, I-34131 Trieste, Italy*
[2]*INAF–Osservatorio Astronomico di Trieste, Via G.B. Tiepolo, 11, I-34143 Trieste, Italy*
[3]*IFPU–Institute for Fundamental Physics of the Universe, via Beirut 2, I-34151 Trieste, Italy*
[4]*INFN–National Institute for Nuclear Physics, via Valerio 2, I-34127 Trieste*
[5]*Tianjin Normal University, Binshuixidao 393, Tianjin, China*

ABSTRACT

Using the GAEA semi-analytic model, we analyse the connection between Damped Ly α systems (DLAs) and H I in galaxies. Our state-of-the-art semi-analytic model is tuned to reproduce the local galaxy H I mass function, and that also reproduces other important galaxy properties, including the galaxy mass–gas metallicity relation. To produce catalogues of simulated DLAs we throw 10^5 random lines of sight in a composite simulated volume: dark matter haloes with $\log(\frac{M_{200}}{M_\odot}) \geq 11.5$ are extracted from the Millennium Simulation, while for $9.2 \leq \log(\frac{M_{200}}{M_\odot}) < 11.5$ we use the Millennium II, and for $8 \leq \log(\frac{M_{200}}{M_\odot}) < 9.2$ a halo occupation distribution model. At $2 < z < 3$, where observational data are more accurate, our fiducial model predicts the correct shape of the column density distribution function, but its normalization falls short of the observations, with the discrepancy increasing at higher redshift. The agreement with observations is significantly improved increasing both the H I masses and the disc radii of model galaxies by a factor of 2, as implemented 'a posteriori' in our $2M-2R$ model. In the redshift range of interest, haloes with $M_{200} \geq 10^{11} M_\odot$ give the major contribution to Ω_{DLA}, and the typical DLA host halo mass is $\sim 10^{11} M_\odot$. The simulated DLA metallicity distribution is in relatively good agreement with observations, but our model predicts an excess of DLAs at low metallicities. Our results suggest possible improvements for the adopted modelling of the filtering mass and metal ejection in low-mass haloes.

Key words: methods: numerical – intergalactic medium – galaxies: evolution – quasars: absorption lines.

1 INTRODUCTION

In the modern cosmological framework, large-scale structure develops hierarchically, due to the growth of gravitational instabilities in a fluid dominated by dark matter and dark energy (e.g. Peebles & McCrea 1981; Peebles 1984; Springel & Hernquist 2003). Galaxies form and evolve in this cosmological scenario, and it is nowadays accepted that a crucial element to understand the physical processes driving galaxy evolution is cold gas. In particular, the cold gas distribution in galaxies at different cosmic epochs should be quantified, understanding how galaxies accrete and lose their gas as a function of cosmic time and environment. These questions have been subject of intense research activities in the past decades (Silk & Mamon 2012; Conselice et al. 2013; Fraternali 2014; Spring & Michałowski 2017; Sorini et al. 2018; Whitney et al. 2019).

With present and upcoming facilities (e.g. MUSE, ALMA, and ELT), allowing us to trace the gaseous components of galaxies out to their outskirts, this is an ideal time to study the cycle of gas (and metals) in and around galaxies. Hydrogen is the most abundant element in the galactic cold phase, and can be detected in emission (21 cm line – mostly in the local Universe) or in absorption (Ly α line, in optical for $z \geq 1.65$ and in UV for lower redshifts). Due to the sensitivity of current instrumentation, the detection in emission (21 cm line) is strongly biased towards the brightest galaxies/highest column densities, and is limited to relatively low redshift (up to $z = 0.06$). In the last decade, the H I content of galaxies has been characterized for a large sample of local galaxies thanks to surveys like HIPASS, ALFALFA, GASS (Meyer et al. 2004; Giovanelli et al. 2005; Catinella et al. 2010, 2013, 2018). These surveys have also allowed studies of the correlation between H I and galaxy stellar mass or other galaxy properties (e.g. star formation rate and environment).

Studies based on absorption lines are not affected by the same observational limits of emission-line studies: the Ly α line results from a transition between the $2^2 P$ state and the $1^2 S$ (ground) state of the hydrogen atom ($\lambda = 1215$ Å), and it is possible to observe it from the ground at $z \gtrsim 1.6$. The systems characterized by the strongest absorption lines are the Damped Ly α systems (DLAs), defined as hydrogen absorbers with column density $N_{H I} > 10^{20.3}$ atoms cm^{-2}. These strong absorbers are typically associated with low-ionization metal line complexes (Prochaska, Castro & Djorgovski 2003; Noterdaeme et al. 2012; Rafelski et al. 2012), suggesting that they are part of a gaseous medium affected by chemical enrichment, like the ISM and the CGM in galaxies.

Large spectroscopic surveys, such as the Sloan Digital Sky Survey (SDSS; Schneider et al. 2010) and BOSS (Eisenstein et al. 2011),

have greatly improved the statistics for samples of high-redshift absorbers ($1.5 < z < 4.5$), tightening the constraints on the shape of the column density distribution function (CDDF), the comoving line density of DLAs, and the evolution of the neutral gas density (e.g. Storrie-Lombardi & Wolfe 2000; Péroux et al. 2003; Noterdaeme et al. 2012; Crighton et al. 2015). These studies have demonstrated that DLAs contain ~ 80 per cent of the neutral gas available for star formation (Storrie-Lombardi & Wolfe 2000; Péroux et al. 2003; Prochaska, Herbert-Fort & Wolfe 2005; Prochaska & Wolfe 2009; Noterdaeme et al. 2012; Zafar et al. 2013), so DLAs studies provide us with an estimate of the gas available for star formation from $z = 5$ to now.

Rafelski et al. (2012) and Neeleman et al. (2013) have estimated the metallicities for a sample of DLAs in the redshift interval ($2 < z < 4$), and investigated their mean metallicity evolution. Recently, these measurements have been updated by De Cia et al. (2018) who developed a procedure to estimate DLA metallicities corrected for dust depletion.

In the last decades, numerous follow-up observations have been carried out to identify the counter-parts of DLAs, mainly at low redshift (e.g. Chen & Lanzetta 2003; Rao et al. 2011a; Rahmani et al. 2016). Different techniques have been used: narrow-band imaging of the fields around the background quasar (Møller & Warren 1998; Kulkarni et al. 2007; Fumagalli et al. 2010; Rahmani et al. 2016), long-slit spectroscopy to search for emission lines from the galaxy associated with the DLA system (e.g. Møller et al. 2002; Fynbo et al. 2010, 2011; Noterdaeme et al. 2012; Srianand et al. 2016; Krogager et al. 2017), integral field spectroscopy (Péroux et al. 2011; Wang, Kanekar & Prochaska 2015), and sub-millimetre observations with ALMA (Neeleman et al. 2019). The detection rate in blindly selected samples remains very low (Fumagalli et al. 2015), but increases when strong cuts on the DLA metallicity are applied (Krogager et al. 2017). These results suggest that DLAs are likely associated with low-luminosity galaxies, most of which are below current observational capabilities (Krogager et al. 2017).

The occurrence of strong H I absorbers detected at high impact parameters ($b > 30$ kpc) from their likely host galaxies (Christensen et al. 2019; Møller & Christensen 2019; Péroux et al. 2019) provides insight into their origin and clustering properties. While in early DLA studies it was commonly believed that they originate from the absorption of gas settled in the discs of massive galaxies (Prochaska & Wolfe 1997), there is now ample observational evidence that small and intermediate mass galaxies provide a non-negligible contribution to DLA statistics (Krogager et al. 2017), in accordance with the predictions of the theoretical study by Rahmati & Schaye (2014).

Font-Ribera et al. (2012) carried out a cross-correlation analysis of DLAs (selected from the BOSS survey) with the Ly α forest and obtained constraints on the DLA cross-section as a function of halo mass. The bias they find implies a typical DLA host halo mass of $\sim 10^{12}$ M$_\odot$ at $z = 2$. In 2018, Pérez-Ràfols et al. (2018a) updated the results by Font-Ribera et al. (2012) finding a typical DLA halo mass of $\sim 4 \times 10^{11}$ M$_\odot$. In the meantime, Arinyo-i-Prats et al. (2018) developed a new method to classify the metal strength of DLAs and studying the dependence of the bias on the metallicity of the absorbers. Pérez-Ràfols et al. (2018b) showed that the linear bias associated with DLAs decreases as their metallicity decreases.

In the last 20 years, a number of theoretical studies have used hydrodynamical simulations to investigate the nature of strong H I absorbers and DLAs in particular (e.g. Gardner et al. 1997, 2001; Haehnelt, Steinmetz & Rauch 1998; Nagamine, Springel & Hernquist 2004; Pontzen et al. 2008; Razoumov 2009; Tescari et al. 2009; Fumagalli et al. 2011; Cen 2012; van de Voort et al. 2012;

Altay et al. 2013; Rhodin et al. 2019; Hassan et al. 2020). The resolution of the simulations adopted has increased over time, but the approach typically needs to resort to different layers of sub-grid prescriptions to model the high H I column densities of DLAs. Some studies overcome the absence of a full cosmological distribution of absorbers by combining results from small-scale simulations with analytic parametrizations of the halo mass function (HMF) to predict statistical properties of the DLA population (e.g. Gardner et al. 1997, 2001), or to study the nature of the host galaxies (e.g. Pontzen et al. 2008). This approach can lead to biased results, requiring some strong assumptions about the environments that can give rise to DLA absorbers. In addition, it does not account for the potentially large scatter in the distribution of absorbers for haloes of similar properties.

Studies based on hydrodynamical simulations have pointed out an important contribution to the DLA population, typically increasing with increasing redshift, from gas that is not associated with the ISM of galaxies. There is no consensus on the quantitative estimate of such a contribution that ranges, depending on the study, between ~ 20 per cent (Villaescusa-Navarro et al. 2018) to more than ~ 50 per cent (Fumagalli et al. 2011; van de Voort et al. 2012). Most numerical and observational studies indicate a major contribution to the DLA population at $2 \leq z \leq 3$ from haloes with virial masses of 10^{10}–10^{12} M$_\odot$ (Cooke et al. 2006; Pontzen et al. 2008; Barnes & Haehnelt 2009; Font-Ribera et al. 2012).

In this study, we focus on an alternative theoretical approach provided by semi-analytic models of galaxy formation. While these are unable to resolve the internal structure of galaxies and do not model the hydrodynamical processes self-consistently, they can easily access to much larger cosmological volumes than hydrodynamical simulations. In addition, a fast exploration of the parameter space and an efficient investigation of the influence of different specific assumptions are possible, thanks to the very limited computational costs. We take advantage of the state-of-the-art semi-analytic model GAlaxy Evolution and Assembly [GAEA; De Lucia & Blaizot (2007), De Lucia et al. (2014), Hirschmann, De Lucia & Fontanot (2016)], coupled to large cosmological *N*-body simulations, and analyse the properties of host DLA galaxies, as well as their connection with dark matter haloes. GAEA accounts for an explicit partition of the cold gas between atomic and molecular hydrogen, but assumes that all cold gas is associated with galaxy discs. Our approach therefore ignores the contribution to DLAs from filamentary structures or gas outflows, and tests to what extent current estimates of DLA statistics can be explained by the gas in galaxy discs.

The specific questions that we will address in our study include the following:

(i) What is the typical virial mass of dark matter haloes that host DLAs?

(ii) To what extent can we reproduce the observed DLA statistical properties, by only considering the ISM associated with galaxies?

(iii) What drives the evolution of $\Omega_{\rm DLA}$ with z, and what is the contribution to this quantity of galaxies with different mass?

The paper is organized as follows. In Section 2, we briefly present the semi-analytical model and the *N*-body simulations used in our study. We then discuss some basic predictions of our model, and the method that we have used to quantify the contribution of dark matter haloes that are not resolved by our simulations. In Section 3, we describe the methodology adopted to create our simulated sample of DLAs, and discuss model predictions in Section 4. In Section 5, we discuss our results in the framework of recent studies, and highlight model improvements/developments
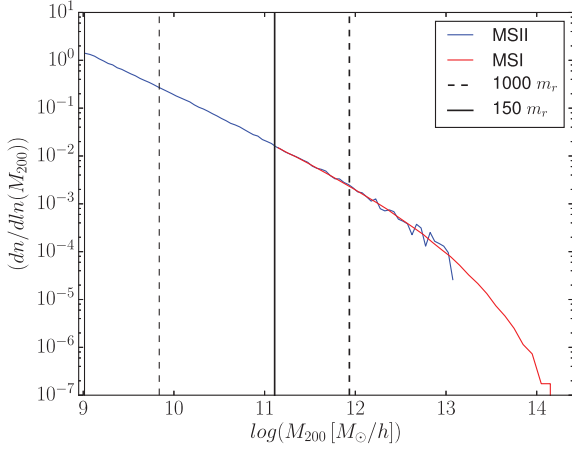
Figure 1. Comparison between the HMF, at $z = 2$, derived from the MSI and MSII (red and blue solid lines, respectively). The vertical solid and dashed lines correspond to 150 times and 1000 times, respectively, the DM particle mass for the two simulations. In our analysis, we will assume that haloes with more than 150 particles are well resolved in both simulations.

that could lead to a better agreement between model predictions and observational results. Finally, in Section 6, we give a summary of our results.

2 PROPERTIES OF THE SIMULATED GALAXIES

2.1 The *N*-body simulations

The adopted physical model for the evolution of galaxies and their baryonic components is coupled to the output of cosmological dark matter simulations, as detailed in De Lucia & Blaizot (2007). In this study, we use dark matter merger trees from two cosmological *N*-body simulations: the Millennium simulation (MSI; Springel et al. 2005) and the Millennium II simulation (MSII; Boylan-Kolchin et al. 2009).

Both the MSI and the MSII assume a *WMAP*1 cosmology, with $\Omega_m = 0.25$, $\Omega_b = 0.045$, $\Omega_\lambda = 0.75$, $h = 0.73$, and $\sigma_8 = 0.9$. Recent measurements from Planck (Planck Collaboration et al. 2016) and *WMAP*9 (Bennett et al. 2013) provide slightly different cosmological parameters and, in particular, a larger value for Ω_m and a lower one for σ_8. As shown in previous work (Wang 2008; Guo et al. 2013), however, these differences are expected to have little influence on model predictions, once model parameters are tuned to reproduce a given set of observables in the local Universe.

The particle mass is $m_{\mathrm{DM}} = 8.61 \times 10^8 \, \mathrm{M}_\odot \, h^{-1}$ for MSI and $m_{\mathrm{DM}} = 6.89 \times 10^6 \, \mathrm{M}_\odot \, h^{-1}$ for MSII, and the box size length $L_{\mathrm{box}} = 500 \, \mathrm{c \, Mpc} \, h^{-1}$ and $L = 100 \, \mathrm{c \, Mpc} \, h^{-1}$, respectively. In Fig. 1, we show the HMF predicted from the two simulations at $z = 2$, where the halo mass is defined as the mass contained in a sphere that encloses an overdensity corresponding to 200 times the critical density of the Universe (M_{200}). In the following, we will consider as resolved all haloes that contain at least 150 particles. This corresponds to $\sim 10^{11} \, \mathrm{M}_\odot \, h^{-1}$ for the MSI and $\sim 10^9 \mathrm{M}_\odot \, h^{-1}$ for the MSII. Below, we will combine the two simulations by selecting galaxies in haloes more massive than $10^{11.5} \, \mathrm{M}_\odot$ from the MSI, and those residing in less massive haloes from the MSII. To investigate equal physical volumes in the MSI and MSII, we will sub-divide the MSI box in 125 sub-boxes, with volume equal to that of the MSII.

2.2 The semi-analytical model GAEA

The GAlaxy Evolution and Assembly (GAEA) semi-analytic model, at the basis of this work, is an evolution of the model originally described in De Lucia & Blaizot (2007), with significant updates that have been published in the last years (see, in particular, De Lucia et al. 2014; Hirschmann et al. 2016). In this study, we use the version of the model that includes an explicit treatment of the partition of cold gas in its atomic and molecular components (Xie et al. 2017). Specifically, we adopt the fiducial run presented in the work by Xie et al. (2017), based on the empirical prescriptions by Blitz & Rosolowsky (2006). We will refer to this as the BR run in the following.

The GAEA model describes the evolution of four different baryonic reservoirs associated with a dark matter halo: (i) a hot gas reservoir that can grow due to cosmological accretion and stellar feedback, and from which gas cools on to the gaseous discs of central galaxies; (ii) a cold gas component associated with model galaxies from which stars form, and whose mass is affected by gas recycling due to stellar evolution and by stellar feedback; (iii) a stellar component for each model galaxy; and (iv) an ejected component that stores the gas that has been removed from the inter-stellar medium (ISM) of galaxies (i.e. cannot participate to star formation), and that can be later re-accreted on to the hot component associated with the parent dark matter halo.

The BR prescription, described by Xie et al. (2017), allows a partition of the cold gas into atomic (H I) and molecular (H₂) hydrogen, and has been tuned to reproduce the observed H I mass function at $z = 0$. The ratio of molecular to atomic hydrogen, $R_{\mathrm{mol}} = \Sigma_{\mathrm{H2}}/\Sigma_{\mathrm{H1}}$, depends on four physical properties of model galaxies: the mass of the *cold gas* (M_{CG}, that in our model corresponds to gas with temperature below 10^4 K), the galaxy stellar mass ($M_\star$), the size of the gaseous disc ($R_{\mathrm{CG,d}}$), and the size of the stellar disc ($R_{\star,\mathrm{d}}$). Using the empirical relation by Blitz & Rosolowsky (2006), the molecular fraction can be expressed as

$$R_{\mathrm{mol}} = \left(\frac{P_{\mathrm{ext}}}{P_0} \right)^\alpha,$$

where P_0 is the external pressure of molecular clumps and its logarithmic value is assumed to be $\log(P_0/k_B[\mathrm{cm}^{-3}\mathrm{K}]) = 4.54$, $\alpha = 0.92$, $P_{\mathrm{ext}} = \frac{\pi}{2} G \Sigma_{\mathrm{CG}} \left[\Sigma_{\mathrm{CG}} + f_\sigma \Sigma_\star \right]$, $\Sigma_\star$ is the stellar surface density, and Σ_{CG} is the cold gas surface density. The latter is estimated in 21 logarithmic annuli (see the original paper by Xie et al. 2017 for details).

In our model, $R_{\mathrm{CG,d}}$ and $R_{\star,\mathrm{d}}$ are estimated from the specific angular momentum of the gaseous (J_{CG}) and stellar ($J_\star$) disc component, respectively, assuming both are well described by an exponential profile:

$$R_{\mathrm{CG,d}} = \frac{J_{\mathrm{CG}}/M_{\mathrm{CG}}}{2 V_{\mathrm{max}}}, \tag{1}$$

$$R_{\star,\mathrm{d}} = \frac{J_\star/M_\star}{2 V_{\mathrm{max}}}, \tag{2}$$

where V_{max} is the maximum circular velocity of the dark matter halo.

Fig. 2 shows the H I mass function predicted by GAEA at $z = 0$, and compares model predictions with observational results by Zwaan et al. (2005) and Martin et al. (2010). The model runs used in this paper are based on the Millennium I (red line) and Millennium II (blue line) simulations (see the next section) that resolve DM haloes down to $\sim 10^{11} \, \mathrm{M}_\odot$ and $\sim 10^9 \, \mathrm{M}_\odot$, respectively. We consider as completeness limit for the cold gas mass at $z = 0$ the values $M_{\mathrm{CG}} \sim 10^8 \, \mathrm{M}_\odot$ and $M_{\mathrm{CG}} \sim 10^7 \, \mathrm{M}_\odot$ for the MSI and MSII, respectively (for details, see Spinelli et al. 2019).
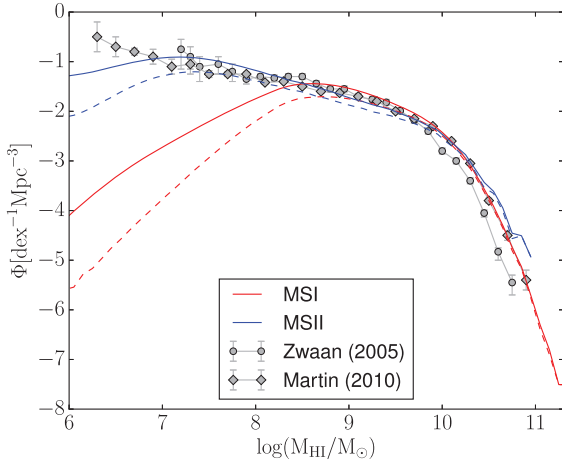
Figure 2. The H I mass function predicted for the MSI (red) and MSII (blue) simulations at $z = 0$. Solid and dashed lines are used for all model galaxies and for centrals only, respectively. Dark grey symbols with error bars show the observational measurements by Zwaan et al. (2005) and Martin et al. (2010). These are based on the blind H I surveys HIPASS (Meyer et al. 2004, limited to $z < 0.04$) and ALFALFA (Giovanelli et al. 2005, limited to $z < 0.06$), respectively. We apply the same stellar mass cuts adopted in Spinelli et al. (2019).

As mentioned above, the observed H I mass function in the local Universe has been used as the primary constraint for the BR model. Previous works have shown that the same model is able to reproduce a number of important additional observational constraints including scaling relations between the atomic/molecular mass and stellar mass, and the observed evolution of the mass–metallicity relation up to $z \sim 3$ (Hirschmann et al. 2016; Xie et al. 2017; Zoldan et al. 2017). This is relevant for our study that will include an analysis of the metallicities predicted for DLAs.

2.3 H I cosmic density

We have estimated the comoving density of the atomic hydrogen, and cold and ejected gas (Ω_{HI}, Ω_{CG}, and $\Omega_{ejected}$) in our simulated universe, summing the corresponding gaseous components of all model galaxies residing in haloes above our adopted resolution limits (for each observable component X, $\Omega_X(z) = \frac{\rho_X(z)}{\rho_c(0)}$). In particular, we have summed the comoving gas density measured in the MSII box, considering haloes in the mass range $10^{9.2} \leq M_{200} < 10^{11.5}\,M_\odot$ to the average comoving gas density measured from haloes with $M_{200} \geq 10^{11.5}\,M_\odot$ in the 125 sub-boxes of the MSI, each with a volume equal to the volume of the MSII box. Fig. 3 shows these model predictions together with observational measurements of Ω_{HI} from Crighton et al. (2015). In this figure, we have corrected for the critical density value corresponding to the cosmology adopted by Crighton et al. (2015).

Our simulated estimate of Ω_{HI} is a factor of ~ 2.5 below the observational estimates based on DLA surveys up to $z \sim 2$, and further decreases at higher redshift. The low z behaviour of the predicted Ω_{HI} is not surprising, because the GAEA model is tuned to reproduce the H I mass function observed in the local Universe by Martin et al. (2010), whose integrated value is a factor of ~ 2 lower than the H I cosmic density estimate by Lah et al. (2007) at $z \sim 0.24$ (also based on emission lines measurements), and lower than estimates based on statistical analysis of DLAs at higher redshift (Rao, Turnshek & Nestor 2006; Noterdaeme et al. 2012).
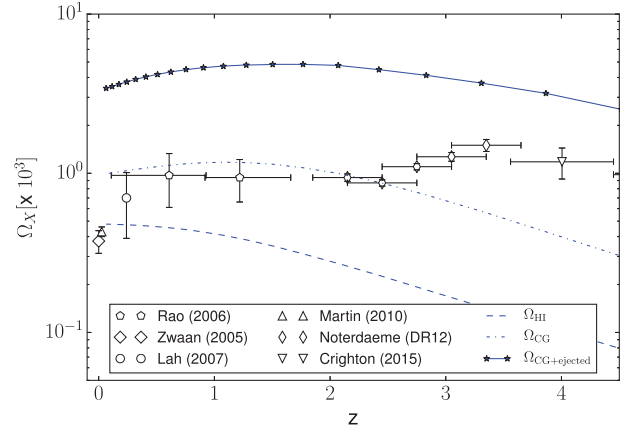
Figure 3. The comoving density evolution of the atomic hydrogen, cold and ejected gas (dashed, dot–dashed, and solid lines, respectively), obtained by summing the corresponding components of all model galaxies down to the resolution limits of the two simulations (see the text for details). Model predictions are compared with observational measurements of Ω_{HI} collected by Crighton et al. (2015).

The decrease of Ω_{HI} at high redshift ($z > 3$) is more difficult to explain. Spinelli et al. (2019) show that the largest contribution to Ω_{HI} in our model is given by haloes with mass $10^{10}\,M_\odot \leq M_{200} \leq 10^{12}\,M_\odot$, and that Ω_{HI} decreases with increasing redshift for more massive haloes while it flattens for less massive haloes. The decrease of Ω_{HI} at higher redshift is found also for independent semi-analytical models that consider a similar mass range of dark matter haloes contributing to the H I density (e.g. Lagos et al. 2011; Berry et al. 2014).

A possible solution to this problem is to increase the contribution of intermediate and low-mass haloes to Ω_{HI} at high redshift, that we may be underestimating because of the adopted physical prescriptions and resolution limits of our simulations. It is difficult to quantify precisely the impact of resolution on our results, as it can affect both the missing H I content of the unresolved isolated haloes and the H I content of the satellite galaxies hosted in the resolved haloes. Considering the resolution limit of the MSII, and the observed scaling relation between the H I to stellar mass ratio and galaxy stellar mass, we expect that the largest contribution should come from the H I content of unresolved haloes. In the next section, we explain how we compute an estimate of such a contribution.

2.4 Minimal HOD model

To quantify whether low-mass haloes (i.e. below $\sim 10^9\,M_\odot$) can significantly increase the H I density in our simulated universe, we populated the MSII box with haloes below its resolution using a simple halo occupation distribution (HOD model – see Berlind & Weinberg 2002 for an historical review).

The number of low-mass haloes to be added, and their mass distribution, have been derived integrating the HMF by Tinker et al. (2008) in the range $10^8\,M_\odot \leq M_{200} < 10^{9.2}\,M_\odot$, and using the cosmological parameters adopted for the Millennium simulations. We have checked that the shape and normalization of the Tinker HMF are consistent with those derived from the MSII and MSI. This can be appreciated in Fig. 4, where we compare the Tinker HMF (solid lines) with that measured from the MSII (dashed lines), at five different redshifts.
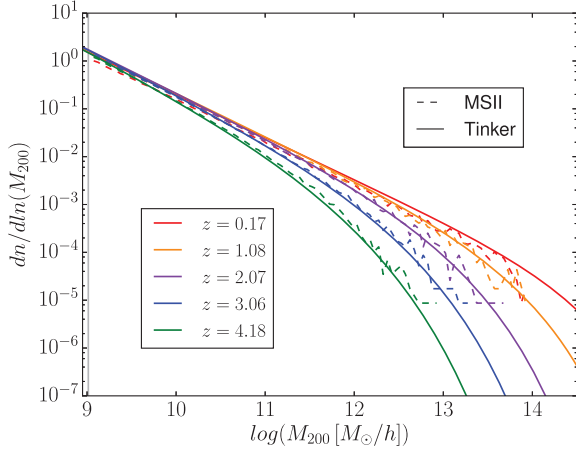
Figure 4. Comparison between the Tinker HMF (solid lines) and that estimated from the MSII (dashed lines), at five different redshifts, listed in the legend.The vertical line marks the resolution limit of the MSII simulation.

Considering the low mass of the haloes treated with the HOD model, we have populated them only with central galaxies, since we do not expect that they host satellites, and distributed them at random positions inside the MSII box. We assign five physical quantities to galaxies in the HOD model: stellar mass($M_\star$), cold gas mass (M_{CG}), scale radius of the gaseous disc ($R_{CG,d}$), scale radius of the stellar disc ($R_{\star,d}$), and abundance ratio [Fe/H]. These quantities are derived extrapolating the scaling relations obtained from our semi-analytic model run on the MSI and MSII. The scaling relations for the first four quantities are shown in Fig. 5 for $z = 2$ (these scaling relations evolve slowly as a function of redshift), while the extrapolation of [Fe/H] is treated in detail in Section 3.3. In the mass regime where MSI and MSII overlap, we observe a nice convergence of the scaling relations (e.g. the difference between galaxy stellar mass of MSI and MSII is less than 10 per cent for haloes with $M_{200} \sim 10^{11}\,M_\odot$). Fig. 5 also shows that the predicted SMHM relation is in good agreement with the observational estimates (Behroozi et al. 2010; Moster et al. 2013; Durkalec et al. 2015).

We have considered two different extrapolations of the predicted scaling relations at each of the snapshots analysed: (i) a linear fit of the median relation obtained for galaxies in MSII (a second-order polynomial for the stellar mass–halo mass relation) and (ii) a flat extrapolation normalized to the value obtained for the smallest haloes in the MSII.

Based on the values extrapolated for $M_\star$, M_{CG}, $R_{\star,d}$, and $R_{CG,d}$, we then estimate the molecular fraction using the empirical relation by Blitz & Rosolowsky (2006) in 21 annuli. For each galaxy in the HOD catalogue, we store the integrated molecular gas fraction in the disc (R_{mol}) that we use to estimate the atomic gas mass. Fig. 6 shows the evolution as a function of redshift of the comoving density of H I and cold gas (solid and dashed lines, respectively) computed considering all DM haloes from the MSI, MSII, and HOD model. The contribution to Ω_{CG} coming from the HOD galaxies becomes non-negligible only at relatively high redshifts (e.g. for $z > 4.5$), and only when considering a flat extrapolation of the scaling relations. The contribution to the cosmic density of neutral hydrogen is dominated by MSI haloes up to redshift $z \simeq 2.3$, when the MSII starts dominating. For the cold gas, the cross-over between the MSI and MSII takes place at $z \gtrsim 3.5$.

We have studied the effect of different halo mass cuts on the cosmic H I content in our simulated universe, finding little differences. In the

following, we adopt the following fiducial cuts: we select haloes with $\log\left(\frac{M_{200}}{M_\odot}\right)$ in the range [8,9.2) from the HOD, haloes with $\log\left(\frac{M_{200}}{M_\odot}\right)$ in the range [11.5, max) from MSI, and haloes from the MSII in the intermediate regime.

3 SIMULATED DLA CATALOGUES

In order to produce samples of simulated DLAs to be compared with observational data we have thrown random lines of sight (LOS) in the volume of the composite simulation described in the previous section. To cover a large halo mass range (10^8–$10^{15}\,M_\odot$), we consider together the galaxies hosted in the DM haloes selected from the MSI, MSII and in those added using the HOD, according to the halo mass cuts described in the previous section. The physical properties assigned to HOD galaxies are derived adopting the flat extrapolation.

We have sub-divided the volume of MSI in 125 sub-boxes of volume equal to that of the MSII box ($L_{box} = 100\,h^{-1}$ Mpc), and constructed at each redshift analysed 125 realizations that differ only for the MSI contribution. In this way, it is possible to investigate the impact of the cosmic variance on the DLA observables considered in this study, for the DM haloes that are well resolved in the MSI (i.e. with $M_{200} > 10^{11}\,M_\odot$). Since cosmic variance is more important for rarer (i.e. more massive) systems, we expect that it does not play an important role for the intermediate mass haloes that are selected from the MSII simulation.

For each simulation snapshot in the redshift range of interest, we throw 100 000 random LOS, parallel to the z-axis, for each of the 125 realizations considered. This provides us with 125 simulated DLA catalogues.

3.1 N_{HI} estimate

For each galactic disc, we assume that the gas density profile follows a double-exponential profile[1]:

$$\rho_{CG}(r, z) = \rho_0\, e^{-r/R_{CG,d}}\, e^{-z/z_0}, \tag{3}$$

where ρ_0 is the normalization of the 3D density profile for the gaseous disc, $R_{CG,d}$, and z_0 are the scale-radius and the scale height of the gaseous disc, respectively.

For the scale height parameter z_0, we apply a linear dependence on the scale radius:

$$z_0 = \frac{R_{CG,d}}{A}, \tag{4}$$

and test two different values of the fudge factor A: =7.3 and =4. The former choice relies on observational relations valid for stellar discs in the local Universe (Kregel, van der Kruit & de Grijs 2002), while the latter is motivated by observations of thicker galactic stellar discs at $z \sim 2$ (Elmegreen et al. 2017).

The H I density profile can be written as

$$\rho_{HI}(r, z) = (1 - f_{mol}(r))\, \rho_{CG}(r, z), \tag{5}$$

where the molecular fraction f_{mol} has been estimated using the BR prescription (described in Section 2.2) in 21 logarithmic radial bins between $r = 0$ and $r = 10\,R_{CG,d}$. When a given LOS intersects a galaxy with a distance (impact parameter) $b \leq 10R_{CG,d}$, the hydrogen column density (N_{HI}) contributed by the galaxy can be estimated by

---

[1]We have tested that assuming an isothermal vertical profile for the gas in the ISM (van der Kruit & Freeman 2011), or alternative vertical profiles suggested in the literature (see Appendix B), does not affect significantly our results.
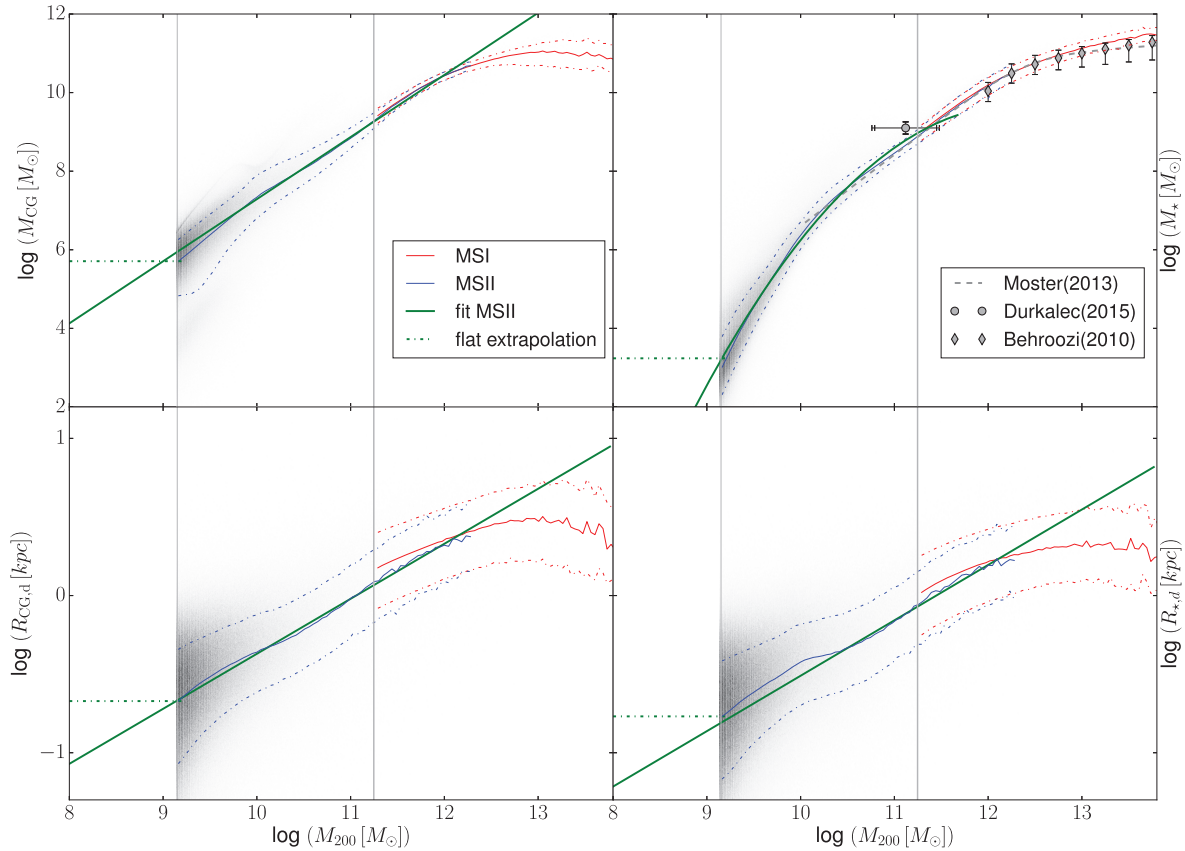
Figure 5. Scaling relation for M_{CG}, $M_\star$, $R_{CG,d}$, and $R_{\star,d}$ as a function of halo mass for central galaxies in MSI (red) and MSII (blue), at redshift $z = 2.07$. The grey colour gradient highlights the number density of MSII central galaxies. The solid and dot–dashed lines show, respectively, the median and percentiles (16th and 84th) of the distributions. The solid green lines show the linear fit for all relations extracted from the MSII, except for the $M_\star$ versus M_{200} relation. In this case, we use a polynomial fit of second order. The dot–dashed lines are flat extrapolations of the scaling relations, normalized to median values corresponding to the lowest halo mass bin resolved in the MSII. In each panel, the vertical solid thick (thin) line shows the resolution limit of the MSI (MSII). For the SMHM relation, we show also observational estimates (Behroozi, Conroy & Wechsler 2010; Durkalec et al. 2015) and the fitting function derived by Moster, Naab & White (2013) for central galaxies.

integrating the H I density profile along the LOS. The value of N_{HI} depends then on the impact parameter and on the inclination of the galactic plane with respect to the LOS.

The assumption that cold gas in model disc galaxies is distributed according to an exponential density profile is in good agreement with observational findings (e.g. Wang et al. 2014).

We have also considered the contributions from close galaxies/pairs to each absorption feature. Adopting a FoF-like merging algorithm, we summed all column densities of absorbing systems, intersected by the same LOS, with a maximum velocity offset of $\Delta v \leq 2000$ km s^{-1}. Our merging algorithm works as follows: we first sub-divide the systems along the same LOS into groups of close systems sorted along the z-coordinate. Then, we merge the two nearest systems in each group, estimate the barycentre of the pair, and re-estimate the distance between the first merged system and the other systems in each group. If necessary, we repeat the merging process and re-iterate until there is no other pair to merge.

The estimated fraction of DLAs originated from multiple systems is large (more than 70 per cent at $z = 2$ for the $2M-2R$ model and more than 50 per cent for the fiducial model). However, in most cases one single galaxy contributes significantly more than the others. In particular, if we consider only systems with column density $N_{HI} \geq 10^{17}$ atoms cm^{-2}, in 87 per cent (84 per cent) of the cases more than 80 per cent of the total hydrogen column density comes from one

single galaxy while the cases where the contribution of each single galaxy is less than 50 per cent represent only 1 per cent (0.5 per cent) of the all cases for the $2M-2R$ model (for the fiducial model). Therefore, the distribution of simulated DLA column densities is not significantly affected by the blending of close absorption features.

3.2 Distribution of impact parameters versus N_{HI}

For our model galaxies, we define the impact parameter, b, as the distance between the LOS and the centre of mass of the galaxy hosting the DLA. In observations, b measures the projected distance between the (luminosity) centre of the galaxy and the quasar sightline piercing the cold gas.

In Fig. 7, we show the distribution of impact parameters as a function of the hydrogen column density (N_{HI}), obtained considering DLAs originating from the MSI and MSII haloes, in the redshift range $2 < z < 3$ and with [Fe/H] > -2. Model predictions are compared with observational measurements by Krogager et al. (2017) that cover the same redshift and metallicity range. The data point come partly from the literature and partly from an X-shooter follow-up campaign. The latter is the first sample of DLA counterparts at high redshift associated with a relatively high detection rate (~ 64 per cent), likely due to the adopted DLA pre-selection: EW$_{SiII} > 1$ Å (EW$_{SiII}$:
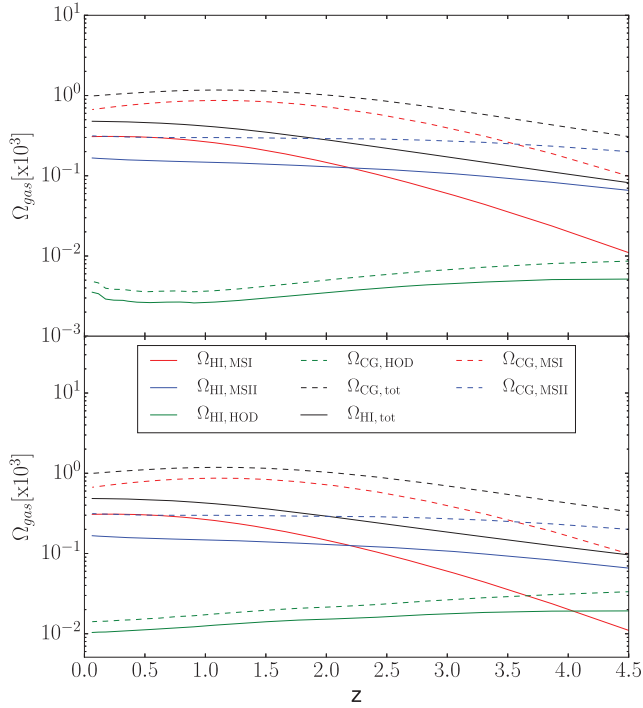
Figure 6. Evolution of the comoving density of atomic hydrogen and cold gas content of model galaxies (H I solid, CG dashed), residing in massive haloes of MSI (red), intermediate-mass haloes of MSII (blue), and low-mass haloes of our HOD extension (green). The black solid (dashed) line shows the total content of H I (CG) in our simulated universe. The top panel shows the contribution coming from HOD galaxies when considering the linear (or second order for the stellar mass–halo mass relation) extrapolation of the scaling relations obtained from the MSII galaxies, while the bottom panel corresponds to the flat extrapolation of the scaling relations.

rest-frame equivalent width of the Si$_{II}$ line, with $\lambda = 1526$ Å) implying large metallicities.

The top panel of Fig. 7 shows results from the run of GAEA described in Section 2.2, our fiducial model, while the bottom panel shows results obtained multiplying by a factor of 2 both the scale radius and the cold gas mass of all model galaxies. In the following, we will refer to this as the $2M-2R$ model.

The largest 99 per cent contour level of the simulated distribution, in both the fiducial and the $2M-2R$ model, encloses all the observed data. In the $2M-2R$ model, all data points fall inside the 95 per cent contour level of the simulated distribution (for MSI haloes), and there is a more clear anticorrelation between impact parameters and column density. For the fiducial model, we find $<b>_{\rm DLA}^{\rm sim} = 8.23$ (3.00) for MSI (MSII). The corresponding values for the $2M-2R$ model are $<b>_{\rm DLA}^{\rm sim} = 14.63$ (5.14). The different mean value of the impact parameters and the different contour levels between DLA originated from the MSI and MSII haloes reflect the dependence of the galactic disc size on the virial radius of the halo where the galaxy resides.

Averaging and weighting over the relative contribution of MSI and MSII, we obtain $<b>_{\rm DLA}^{\rm sim} = 5.53$ and $<b>_{\rm DLA}^{\rm sim} = 10.03$ for the fiducial and the $2M-2R$ model, respectively. Both estimates are in agreement with the one found by Krogager et al. (2017), $<b>_{\rm DLA} = 8.32$, with a slight preference for the $2M-2R$ model.

Extending the sample of observed DLAs towards lower redshift, Rhodin et al. (2018) reports $<b>_{\rm DLA} = 11.1$ kpc. Older work by Rao et al. (2011b) based on low-redshift DLAs counterparts found
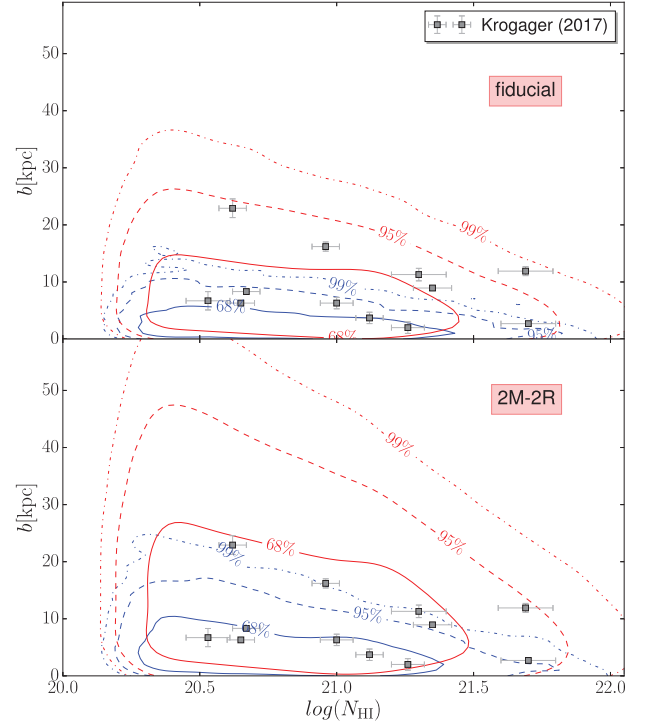
Figure 7. Impact parameter (b) as a function of the hydrogen column density $N_{\rm H I}$, for simulated DLAs in the redshift range $2 < z < 3$, compared to the observations by Krogager et al. (2017, grey symbols). The top panel shows the contour distributions based on the fiducial model, while the bottom panel shows results obtained multiplying by a factor of 2 both the scale radius and the cold gas mass of all model galaxies ($2M-2R$ model, hereafter). Red and blue lines refer to galaxies in the MSI and MSII, respectively, and show different contour levels of the distribution as indicated in the legend. We apply a cut to the metallicity of model DLAs equal to [Fe/H] > −2.0 for consistency with the observational measurements considered.

$<b>_{\rm DLA} = 17.4$ kpc, considering a larger metallicity cut ([Fe/H] > −1). The different observational estimates depend on the adopted DLA pre-selections, on the techniques used to search for DLAs counterparts, and in part also on the expected redshift evolution of galaxy sizes that implies an evolution of the observed range of impact parameters.

It is important to bear in mind that all observations of DLA counterparts are biased against smaller impact parameters, for which it is difficult to detect the DLA counterparts (as discussed in Krogager et al. 2017). The technique adopted by Krogager et al. (2017) likely misses also some counterparts at large impact parameters due to the partial coverage of the FoV by the three long-slits, while DLA systems with very high metallicities are not completely detected because of the dust bias (Khare et al. 2012), which affects the colour selection of QSOs. Therefore, the comparison shown in Fig. 7 should be considered more as qualitative than rigorous.

3.3 Assigning metallicity to DLAs

The GAEA model adopts a detailed chemical enrichment scheme that accounts for the finite lifetime of stars and the non-instantaneous recycling of metals, gas, and energy (De Lucia et al. 2014).

As discussed in previous work, the fiducial model used here is able to reproduce the observed evolution of the correlation between galaxy stellar mass and cold gas metallicity, up to $z \sim 2$ (Hirschmann et al. 2016; Xie et al. 2017). This is an important achievement, met
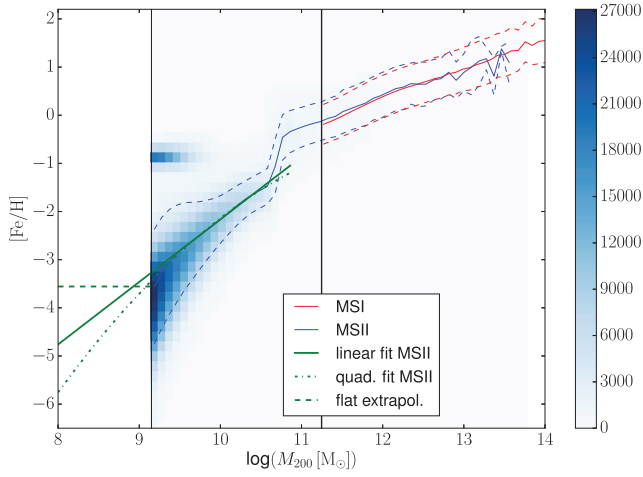
Figure 8. [Fe/H] as function of $\log(M_{200})$ for central galaxies, at $z = 2$. The solid blue (red) line shows the mean relation for MSII (MSI) galaxies. The green solid, dashed, and dot–dashed lines show the extrapolated linear fit to the mean relation measured for the MSII, a flat extrapolation, and an extrapolation based on a quadratic fit to the MSII results. The colour coding quantifies the number density of the MSII central galaxies. The vertical lines show the resolution limits of the MSI and MSII.

by only a few recently published theoretical models (see discussion in Somerville & Davé 2015). Our study offers an additional test to the model.

As commonly done in DLA studies, we use the iron over hydrogen abundance ratio, $\left[\frac{Fe}{H}\right]$, as a proxy for the metallicity of the gaseous discs of our simulated galaxies. GAEA assumes a uniform distribution of the metals in the different baryonic components. So we can write

$$\left[\frac{Fe}{H}\right] = \log\left(\frac{M_{Fe,d}\,\mu_H}{M_{HI,d}\,\mu_{Fe}}\right) - \log\left(\frac{M_{Fe}\,\mu_H}{M_{HI}\,\mu_{Fe}}\right)_\odot, \qquad (6)$$

where $M_{Fe,d}$ and $M_{HI,d}$ are the masses of Fe and H I in the cold gaseous disc of each galaxy, while μ_{Fe} and μ_{HI} are the corresponding mean atomic weights. [Fe/H]$_\odot$ is the solar abundance ratio that we take from Asplund et al. (2009).

Fig. 8 shows the relation between the abundance ratio [Fe/H] and $\log(M_{200})$, for the central galaxies in the MSI and MSII. There is a good convergence between MSI and MSII in the galaxy mass range $10^{8.5} < M_\star < 10^{10}\,M_\odot$. For the extrapolation of the [Fe/H]–M_{200} relation to galaxies inside the haloes sampled by the HOD, we have used a linear regression in the mass range $10^{9.2} < M_{200} < 10^{10.6}\,[M_\odot]$, i.e. after the step-like feature visible in the figure. This feature arises mainly as a consequence of a specific assumption of our galaxy formation model: for $M_{200} \le 5 \times 10^{10}\,M_\odot$, 95 per cent of the new metals are ejected directly into the hot phase, instead of being mixed with the cold-gas in the ISM, as assumed for more massive haloes. This assumption was motivated by results from hydrodynamical simulations (Mac Low & Ferrara 1999) and was helpful, in previous versions of our models, to reproduce the metal content of satellites in Milky-Way-like haloes (Li, De Lucia & Helmi 2010).

Considering the large $1-\sigma$ scatter of the predicted relation, we have applied it to the extrapolated values of the abundance ratio for HOD galaxies. Once assigned the atomic hydrogen mass to HOD galaxies (following the procedure described in the previous section), we can use the extrapolated abundance ratio to assign an iron mass to each HOD galaxy.

We have also considered the effect due to the presence of a metallicity radial gradient. Specifically, we have assumed a slope consistent with the observational study by Christensen et al. (2014):

$$\Gamma = -0.022\ \text{dex kpc}^{-1}. \qquad (7)$$

Christensen et al. (2014) and Rhodin et al. (2018) found an almost universal metallicity gradient for a sample of DLAs observed in the redshift range (2–3.5). Stott et al. (2014) found at $z \sim 1$ a slight correlation between the metallicity gradient and the sSFR (but see Carton et al. 2018; Ma et al. 2017, for a different view). Here, we test if we are able to recover the observed trends using the simplest assumption of a universal metallicity gradient.

The iron over hydrogen abundance ratio can be estimated for each DLA, for a given impact parameter $b = r$, applying the following formula:

$$\left[\frac{Fe}{H}\right](r) = \log_{10}(Z_{DLA,0}) - \log_{10}(Z_\odot) - \Gamma r, \qquad (8)$$

where

$$Z_{DLA,0} = Z_{DLA}(r=0) = \frac{\mu_{HI}}{\mu_{Fe}}\frac{M_{Fe}(1 - 11\,e^{-10})}{M_{HI}}$$

$$\times \frac{(1 - <f_{mol}>)}{\int_0^{10} dy(1 - f_{mol}(y))}e^{-(1+ln(10)\Gamma r_s)y}$$

and $y = r/r_s$, while $r_s = R_{CG,d}$ and f_{mol} is the molecular fraction.

4 PROPERTIES OF SIMULATED DLAS

In this section, we compare the properties of the DLAs in our simulated universe with those estimated from observational data. For each property derived in this section, we have combined the MSI and MSII simulation as explained in Section 3.

4.1 The column density distribution function

The CDDF is defined as the number of absorbers observed per unit redshift path and column density interval:

$$f(N_{HI}, X)dXdN_{HI} = n_{abs}(N_{HI}, X), \qquad (9)$$

where the absorbing path dX is defined as $dX = \frac{H_0}{H(z)}(1+z)^2 dz$ in terms of the redshift path dz.

The CDDF plays, in absorption line studies, a similarly central role (and provides a similarly 'vague' information) as the luminosity function in galaxy evolution studies. The analytic model of the CDDF proposed by Schaye (2001) (devised for overdensities that cannot self-shield from the UV background), together with results from cosmological simulations (e.g. Altay et al. 2011), indicate that systems of a given column density originate from dramatically different overdensities. Nevertheless, higher column density systems are typically connected to denser gas that, in general and average sense, tends to lie closer to galaxies. It has been argued that the steepest part of the CDDF, made of the densest absorbers, may be particularly sensitive to stellar feedback and stellar evolution (Rosenberg & Schneider 2003; Bird et al. 2014).

Fig. 9, in the upper panel, shows the CDDF derived from our simulated absorbers in the redshift range $2 < z < 3$, for our fiducial combination of halo mass cuts (see Section 2.4) applied to fiducial GAEA model. Model predictions are compared with observational estimates by Noterdaeme et al. (2012).

The figure shows a significant discrepancy between our fiducial model and the observed CDDF, in particular below $\log(N_{HI}) < 21$.
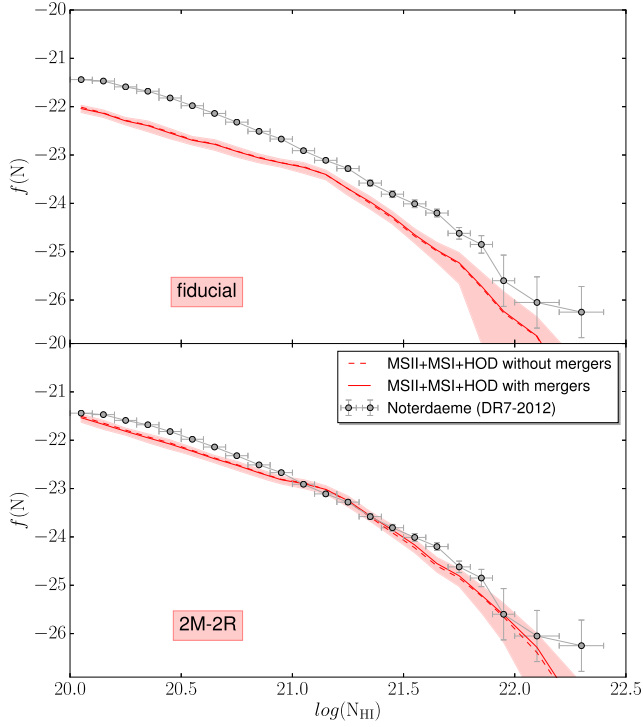
Figure 9. Predicted CDDF in the redshift range $2 < z < 3$. The top panel shows results based on our fiducial model, while the bottom panel shows the results of the model where $R_s = 2R_{s,orig}$ and $M_{CG} = 2M_{CG,orig}$. We estimate the CDDF for each of the 125 realizations described in Section 3, converting into a redshift interval (dz) the length of each LOS, that is equal to $L_{box} = 100\,h^{-1}$ Mpc comoving at all redshifts. The solid red line indicates the average of the CDDFs obtained for all realizations considered in the redshift range of interest ($2 < z < 3$), while the shaded area highlight the $1 - \sigma$ scatter of the distribution.

This discrepancy motivated us to test the dependence of the CDDF on the physical properties of simulated galaxies, and in particular the scale radius of the gaseous disc and the cold gas mass. The lower panel of Fig. 9 shows results obtained multiplying by a factor of 2 both the scale radius and the cold gas mass of all model galaxies (the *2M–2R* model introduced above). Results from this ad hoc modifications are in very good agreement with observations at $20.0 < \log(N_{HI}) < 22.2$, for the redshift range considered. We have verified that the better agreement with observational data is mainly driven by the increase of scale radius that leads to a larger galaxy cross-section (i.e. a larger probability of intersecting model galaxies).

4.2 The cosmic hydrogen density associated with DLAs

The cosmic hydrogen density associated with DLAs can be computed as

$$\Omega_{DLA} = \frac{m_H H_0 \sum_i N_i(H\,I)}{c\rho_c \Delta X}, \tag{10}$$

where ρ_c is the critical density at $z = 0$, m_H is the mass of the hydrogen atom, and the sum is carried out over all systems with $\log N(H\,I) > 20.3$, across a total absorption path-length ΔX.

Fig. 10 compares the redshift evolution of the comoving H I density derived from our simulated DLAs, with the observational estimates from Crighton et al. (2015). To be consistent with the observations, we have corrected the values provided by equation (10) by a factor
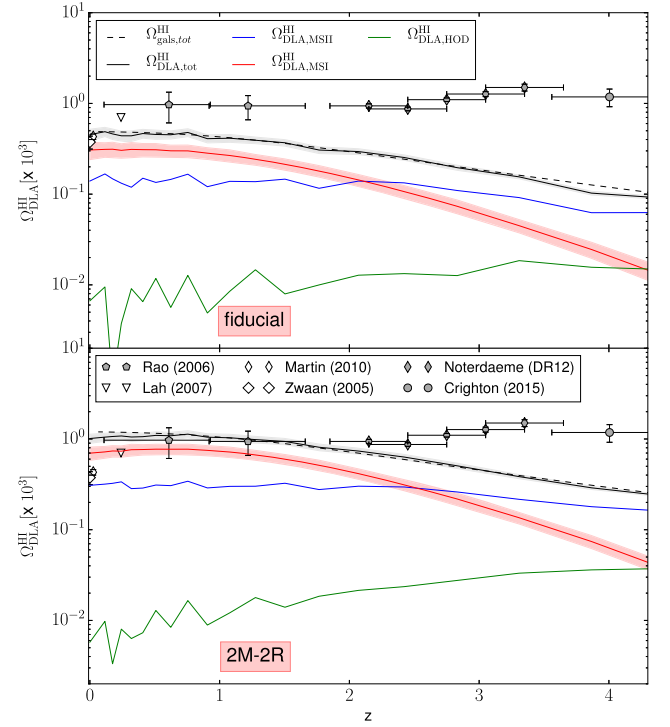

Figure 10. The top (bottom) panel shows the evolution with redshift of Ω_{DLA}^{HI}, in our fiducial (2R–2M) model. We define $\Omega_{DLA}^{HI} = 1.2 \times \Omega_{DLA}$, taking into account the contribution to the comoving H I density of systems with column density lower than the characteristic one of DLAs (Crighton et al. 2015). The solid black line shows the average Ω_{DLA}^{HI} evolution considering the contribution of all individual systems, while the dashed black line refers to the comoving H I density (Ω_{HI}) of all the galaxies in the box. The three solid lines in red, blue, and green refer to the DLAs in the MSI, MSII, and HOD, respectively. Symbols with error bars show observational data points, taken from the literature as detailed in the legend, and expressed in the cosmology used by Crighton et al. (2015).

of 1.2 ($\Omega_{DLA}^{HI} = 1.2 \times \Omega_{DLA}$) that takes into account the contribution to the comoving H I density of absorbers with column density below $N_{HI} = 20.3$ (Crighton et al. 2015). Using this correction, Ω_{DLA}^{HI} turns out to agree remarkably well with Ω_{gals}^{HI} that is derived summing the H I contribution of all model galaxies. This non-trivial result indicates that our model predicts the correct shape for the CDDF (both for the fiducial and the $2M-2R$ model).

Ω_{DLA}^{HI} derived from our simulations ('fiducial' model) is, on average, a factor of ~ 2.5 below the observational estimates in the redshift range $0 < z < 2$, and it further decreases to about an order of magnitude below the data at $z = 4$. As discussed in Section 2, the difference is in part due to the fact that our model is tuned to reproduce the HIMF measured in the local Universe (Martin et al. 2010) that gives an estimate of the integrated comoving H I density a factor of ~ 2 lower than that derived from DLA observations. The $2M–2R$ assumption alleviates the discrepancy, at least up to redshift $z \sim 3$, as shown in the lower panel of Fig. 10.

At higher redshift, also the predictions from the $2M-2R$ model exhibit a significant decline, while observations measure little evolution of Ω_{DLA}^{HI} up to $z \sim 5$. This can be due to different reasons: one hypothesis is that the uniform redistribution of the missing hydrogen, applied in the $2M-2R$ model, is limited, since it gives too much gas to the more massive haloes, which already reproduce the observations, and too less to the intermediate/ low mass ones. The other possibility is that the contribution of outflows and/or
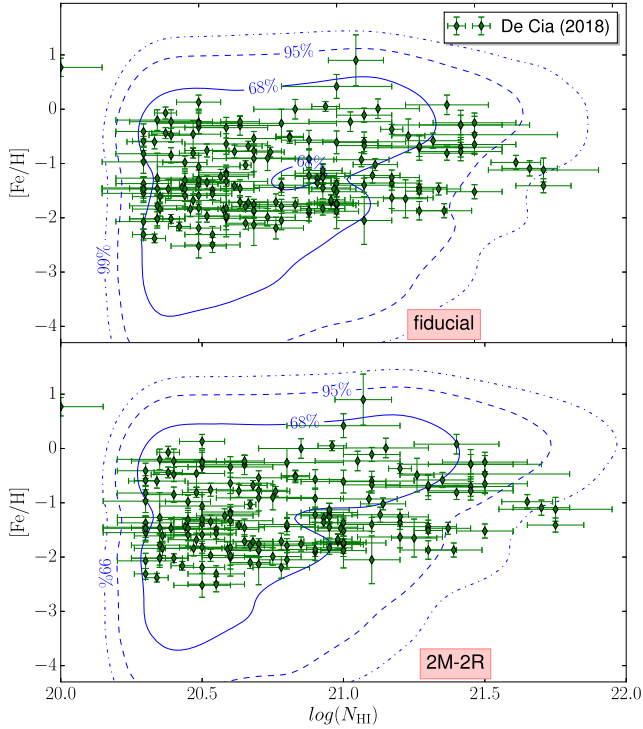
Figure 11. $[\frac{Fe}{H}]$ as a function of $N_{H\,I}$ in the redshift range $2 < z < 3$, with no correction for a metallicity gradient. The top and bottom panels show the metallicity distributions based on the fiducial and $2M–2R$ models, respectively. In both panels, we show the distributions of abundance ratios obtained by stacking the 125 realizations considered.

filamentary structure becomes more significant at higher redshift (e.g. Fumagalli et al. 2011; van de Voort et al. 2012).
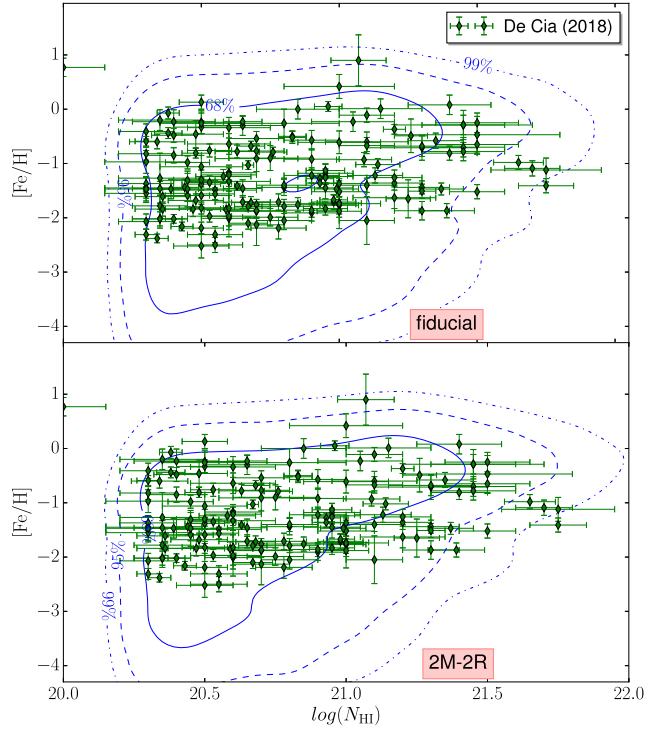
4.3 DLA metallicity

4.3.1 Relation between metallicity and $N_{H\,I}$

We compare the metallicity of our simulated DLAs with observations taking advantage of the catalogue by De Cia et al. (2018) that provides also dust-corrected abundance ratios. As explained earlier, we adopt the iron over hydrogen abundance ratio ([Fe/H]) as a proxy of the metallicity, and we analyse separately the 125 DLA catalogues built (as described in Section 3), in the redshift range $2 < z < 3$.

Figs 11 and 12 show [Fe/H] as a function of $N_{H\,I}$ in the redshift range $2 < z < 3$, with lines of different styles contouring the regions enclosing 68, 95, and 99 per cent of the distribution coming out from the stacking of the 125 DLA catalogues. Green symbols with error bars show observational measurements. Fig. 11 shows results obtained assuming a uniform distribution of the metals in the gaseous disc, while Fig. 12 shows the distribution obtained assuming a universal metallicity gradient (see Section 3.3).

The simulated abundance ratios appear in somewhat better agreement with the data when we consider a metallicity gradient, in particular at larger metallicity values. Our model, however, predicts a not negligible number of low abundance ratios ([Fe/H] < −3.) systems that are not observed (Prochaska & Wolfe 2009).

To make the comparison more quantitative, we carry out a two-dimensional Kolmogorov–Smirnov test to quantify the probability that the simulated and observed distributions are extracted from the same sample (i.e. are consistent).

Figure 12. As in Fig. 11, but applying a correction for the metallicity gradient, based on the fitting formula by Christensen et al. (2014).

The estimate of the *p*-value, namely the probability of obtaining the observed distribution assuming the null hypothesis, is 1.16 × 10⁻⁵ (1.32 × 10⁻⁵) for the $2M−2R$ run, with (without) a correction for the metallicity gradient. The corresponding value for the fiducial model is 1.77 × 10⁻⁷ (5.53 × 10⁻⁷). Therefore, the hypothesis that the observed DLA metallicities come from the same parent population of the simulated DLAs is on average rejected for both models considered, independently of the metallicity gradient applied.

The discrepancy between observed and simulated data is mainly driven by the presence, in our model, of low-metallicity systems that are absent in the observed DLA samples. In addition, our simulated DLAs corresponding to large column densities tend to have an average metallicity that is larger than the observed one. This 'shift' in the average metallicity at higher column densities of simulated systems, with respect to that observed, increases in the $2M–2R$ model.

The excess of low-metallicity systems in our model suggests that the treatment of the chemical enrichment of low-mass haloes (see Fig. 8) may be inadequate and should be revised. The difference in the distributions at larger metallicities is more difficult to explain. It is worth noting that the High A_V Quasar survey (HAQ; Fynbo et al. 2013; Krogager et al. 2015; Zafar et al. 2015) and the extended-HAQ (Krogager et al. 2016) have shown that the traditional quasar selection used in SDSS is biased against reddened quasars. In addition, the work by Noterdaeme et al. (2015) showed that DLAs associated with large column-densities and metallicities are typically found to exhibit a more significant reddening of the background quasar. Therefore, it is plausible that the combined effect of dust and large atomic hydrogen densities cause a dust-bias in DLA observations, preferentially excluding from the observations DLAs hosted in massive, metal-rich and dusty galaxies. In the $2M−2R$ model, the average metallicity is slightly larger than for the fiducial model, due
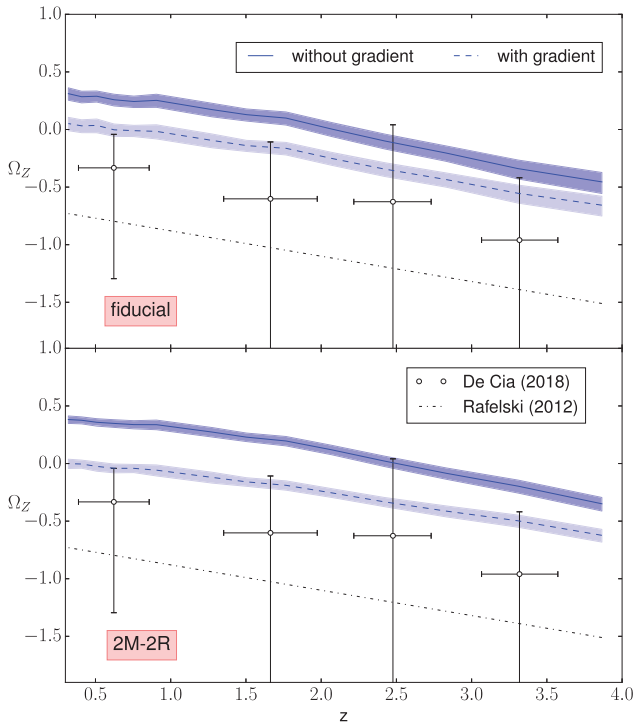
Figure 13. Cosmic metallicity evolution. Solid and dashed blue lines show model predictions without and with a correction for metallicity gradient, respectively. The shaded areas highlight the relative $1 - \sigma$ scatter regions. The top panel shows results for our fiducial GAEA run, while the bottom panel corresponds to the $2M$–$2R$ model.

to the reassignment 'a posteriori' of the scale radius and the mass, which penalizes the low-mass galaxies at intermediate/high column density (see Appendix A).

4.3.2 Cosmic metallicity evolution

As observed by Rafelski et al. (2012), the chemical enrichment of DLAs evolves from about 1 to 10 per cent solar from $z \sim 5$ to today. The work by Rafelski et al. (2012) also revealed a statistically significant decline of the DLA average metallicity with increasing redshift that can be described as $<\Omega_Z> =(-0.26 \pm 0.07)z - (0.59 \pm 0.18)$. This behaviour was confirmed at $z < 4$ by independent measurements (Kulkarni et al. 2007, 2010).

De Cia et al. (2018) also found a similar decrease with redshift, but with a different normalization at low redshift, based on abundance ratios corrected for dust depletion.

We have investigated the evolution of the DLA metallicity by computing the mean cosmic metallicity of simulated DLAs at different redshifts. Following Rafelski et al. (2012), this can be defined, at each redshift, as

$$< \Omega_Z > = \log_{10}\left(\sum_i \frac{10^{[M/H]_i} N_{\mathrm{H\,I},i}}{\sum_i N_{\mathrm{H\,I},i}} \right) \qquad (11)$$

where the index i runs over all DLAs in the redshift bin considered, and $[M/H]_i$ is the adopted metal abundance ratio (in our case [Fe/H]). Fig. 13 shows the mean cosmic metallicity as defined in equation (11) for the redshift range $0.3 < z < 4$, together with the fitting function (black dashed line) found by Rafelski et al. (2012) and the data from De Cia et al. (2018). We consider the latter sample as our reference

Table 1. Median DLA host halo masses predicted by our fiducial and $2M$–$2R$ model.

z	M_{200} (fiducial)	M_{200} ($2M$–$2R$) [10^{11} M$_\odot$]
2.07	$1.50\pm^{0.64}_{0.21}$	$2.67\pm^{0.56}_{0.50}$
2.42	$1.32\pm^{0.14}_{0.23}$	$2.00\pm^{0.38}_{0.43}$
3.06	$0.71\pm^{0.18}_{0.11}$	$0.94\pm^{0.21}_{0.11}$

data sample, since our metal abundances do not account for dust depletion.

When we apply a correction for the metallicity radial gradient, the mean evolution of the cosmic metallicity of our simulated DLAs is in agreement with the data by De Cia et al. (2018) within the errors, although model predictions tend to give always higher median values than the median of the data, at all redshifts considered. This is expected because, as discussed above, observations likely miss most of the DLAs at high column densities with large metallicity.

4.4 DLA host halo masses

The typical range of halo masses hosting DLAs is still an open question, albeit the low detection rate of DLA counter-parts in optical follow-up observations suggest that DLAs are most likely associated with faint galaxies and therefore reside in small haloes (Fynbo, Møller & Warren 1999; Krogager et al. 2017). If we adopt the same argument used in abundance matching studies (e.g. Conroy, Wechsler & Kravtsov 2006), the results by Fynbo et al. (2008) can be translated into a typical DLA host halo mass of $M_{200} < 10^{11}$ M$_\odot$. This is, however, in tension with more recent observational work based on DLA kinematics and clustering. The distribution of velocity widths measured from low ionization metal lines shows a prominent tail at high velocities, which suggests the existence of a population of large discs hosting DLAs (Bird et al. 2015). Moreover, the recent cross-power spectrum analysis by Font-Ribera et al. (2012), based on the BOSS survey, provides an estimate of the linear bias of the observed DLAs ($b_{\mathrm{DLA}} = 2.17 \pm 0.20$), suggesting a typical host halo mass $\sim 10^{12}$ M$_\odot$. This analysis has been updated by Pérez-Ràfols et al. (2018a), who found a linear bias of $b_{\mathrm{DLA}} = 2.00 \pm 0.19$, only slightly lower than the clustering amplitude measured for Lyman Break Galaxies (see Cooke et al. 2006), and no dependence of the bias value on redshift or column density. This bias value implies $M_{\mathrm{host}}^{\mathrm{DLA}} \geq 10^{11}$ M$_\odot$, that is larger than the one typically predicted by some simulations and semi-analytic models (Pontzen et al. 2008; Barnes, Garel & Kacprzak 2014; Padmanabhan, Refregier & Amara 2017). The median typical DLA host halo masses, found in the redshift range $2 < z < 3$ from our model, are listed in Table 1 and are in agreement with observational results by Pérez-Ràfols et al. (2018a).

Recently, Pérez-Ràfols et al. (2018b) have shown that the bias of DLAs exhibits a dependence on metallicity in line with preliminary observational results (Neeleman et al. 2013; Christensen et al. 2014) and the expectation that more metal-rich DLAs are associated with more massive galaxies. In our model, we also see a variation of the average metal content of DLAs hosted in haloes of different masses, and this can be explained as a consequence of the relation between the gas metallicity in galaxies and the host halo mass (see Fig. 8).

5 DISCUSSION

In this work, we have analysed the properties of DLAs by taking advantage of a semi-analytic model of galaxy formation and

evolution (GAEA; presented in Hirschmann et al. 2016; Xie et al. 2017) coupled to two large cosmological *N*-body simulations: the Millennium (MSI) and Millennium II (MSII). In order to estimate the possible contribution from haloes that are below the resolution of our simulations, we have used a simple HOD approach by placing, at random positions within the simulated box, a number of haloes with mass distribution consistent with analytic formulations tuned on *N*-body simulations (Tinker et al. 2008). Our model assumes that all atomic hydrogen is associated with the gaseous disc of galaxies, i.e. there is no contribution from filamentary regions or extraplanar gas. Our simulated DLAs catalogues are then built by throwing a large number (100 000) of random lines of sight along the *z*-direction of 125 simulated boxes, obtained combining the simulations available and complemented with HOD extrapolation (as described in Section 3).

Our fiducial model predicts a CDDF with the correct shape but offset low with respect to observational measurements by Noterdaeme et al. (2012). This affects the predicted values of the cosmic hydrogen density in DLAs ($\Omega_{\mathrm{DLA}}^{\mathrm{HI}}$), that is a factor ~ 2.5 lower than observational estimates at $0 < z \leq 2$, and even more at higher redshift. Up to $z < 3$, the disagreement with data can be overcome by increasing the radius of the gaseous disc and the gas mass by a factor of ~ 2 (our $2M{-}2R$ model). As for the DLA metallicity distribution in the redshift range $2 < z < 3$, our model predicts an excess of low metallicity DLA systems, while the average cosmic DLA metallicity (Ω_Z), weighted over N_{HI}, follows the same redshift evolution as observational measurements but it is slightly higher than observed values. The predicted Ω_Z becomes compatible with observations, within the uncertainties, once we account for a modest radial metallicity gradient.

Below, we discuss our results in relation with independent recent studies, and point out possible developments/improvements of the adopted physical model that can bring model results in better agreement with observational measurements.

5.1 Comparison with the literature

In the last 20 yr, a number of theoretical studies, either using a semi-analytic approach (Lagos et al. 2011, 2014; Berry et al. 2014; Kim et al. 2015) or hydrodynamical simulations (Nagamine et al. 2007; Pontzen et al. 2008; Tescari et al. 2009; Altay et al. 2011; Cen 2012), have focused on the evolution of the atomic hydrogen content of the Universe.

In the framework of this paper, it is particularly interesting to discuss our results in relation to the analysis by Berry et al. (2014, 2016), also based on semi-analytic models and focused on the predicted properties of DLAs. In their work, Berry et al. (2014) use variations (see their table 1) of the semi-analytic model published in Somerville et al. (2008, see also Popping, Somerville & Trager 2014; Somerville, Popping & Trager 2015), including different prescriptions for the partition of cold gas in atomic and molecular hydrogen, and alternative assumptions for the sizes of gaseous discs. Our model and the one used by Berry et al. (2014) differ significantly for the numerical implementation and for the prescriptions adopted for modelling various physical processes. Popping et al. (2014) and Xie et al. (2017) show that both models are able to reproduce the evolution of disc sizes (both stellar and gaseous) up to $z \sim 2$, for galaxies more massive than $10^9\,\mathrm{M}_\odot$. Fig. 2 of Berry et al. (2014) shows that none of the model variants they considered reproduces well the local HI mass function, while our fiducial model is tuned to reproduce this observational constraint.

Both our fiducial run and the reference disc model used in Berry et al. (2014) underpredict the CDDF of DLAs. Berry et al. (2014) find a better agreement by increasing the cold gas specific angular momentum with respect to what assumed in their reference model. This leads to larger gaseous discs, but also to a significantly worse agreement with the HI galaxy mass function in the local Universe (see their fig. 2). This is consistent with our findings that a model where we arbitrarily multiply by a factor of 2 both the scale radii and HI masses of model galaxies better reproduces the observed CDDF. Ours is an 'ad hoc' solution, and it remains to be demonstrated that plausible modifications of the modelled physical processes can lead to such solution without (significantly) affecting the agreement shown between model predictions and observational data in the local Universe. We will come back to this issue in the next section, in the framework of possible developments of the GAEA model.

It should be noted that also our $2M{-}2R$ model that reproduces the observed column density distribution of DLAs for $z < 3$, predicts a decline of $\Omega_{\mathrm{DLA}}^{\mathrm{HI}}$ at higher redshift. This is in disagreement with observational measurements and consistent with what found by Berry et al. (2014, 2016). This decline is driven by an underestimation of the CDDF for $\log(N_{\mathrm{HI}}) < 21$. The behaviour is not shared by hydrodynamical simulations that typically not underestimate the CDDF for $\log(N_{\mathrm{HI}}) < 21$ and find no evolution or even a moderate increase of $\Omega_{\mathrm{DLA}}^{\mathrm{HI}}$ (Cen 2012; van de Voort et al. 2011; Altay et al. 2011) in better agreement with observational measurements.

The different behaviour at high redshift, predicted by semi-analytic models and hydrodynamical simulations, could be at least in part explained by an increasing contribution to the DLAs cross-section of filamentary structures and outflows/inflows at higher redshift (van de Voort et al. 2011; Fumagalli et al. 2011; Cen 2012). In addition, simulations predict that at $z \sim 3$ the haloes that contribute most to the CDDF for $\log(N_{\mathrm{HI}}) < 21$ are the ones in the mass range $10^9 < M_{200} < 10^{10}\,\mathrm{M}_\odot$ (Tescari et al. 2009; Rahmati et al. 2013), while in our model the major contribution comes from haloes in the mass range $10^{10} < M_{200} < 10^{12}\,\mathrm{M}_\odot$. Since more massive haloes are less numerous at higher redshift, the difference in the typical DLA host halo mass at $\log(N_{\mathrm{HI}}) < 21$ could partially explain the decline of the $\Omega_{\mathrm{DLA}}^{\mathrm{HI}}$ in our model. Another concern is related to the possible contribution of haloes that are below the resolution of our simulations. In order to understand to what extent low-mass haloes contribute to the HI comoving density, we have estimated the contribution of haloes with mass $10^8\,\mathrm{M}_\odot < M_{200} < 10^{9.2}\,\mathrm{M}_\odot$ resorting to a simple HOD model (see Section 2 for details). Our results indicate that these low-mass haloes represent a negligible contribution to the column density distribution in the redshift range of interest. The average covering fraction of HI in different haloes is influenced also by the interplay between the UV background and the gas density in the galactic discs. At the column density typical of DLA systems, the gas is self-shielded by the ionizing photons of the UV background, then mostly neutral. Our semi-analytical model does not include a specific treatment for the self-shielding but this effect is taken into account implicitly through the adoption of the BR prescription for the cold gas partitioning (Blitz & Rosolowsky 2006). Albeit the Blitz & Rosolowsky (2006) prescription is based on observations of local galaxies, we are confident in applying this prescription to all redshifts, since we have demonstrated that it provides very similar results to alternative parametrizations based e.g. on hydrodynamical simulations that account explicitly for self-shielding (Xie et al. 2017).

As discussed in the previous sections, both our fiducial and $2M{-}2R$ models predict an excess of low-metallicity DLAs that are

not present in observational samples, also in the dust-corrected DLA abundance ratio catalogue by De Cia et al. (2018). Results based on the model by Somerville et al. (2015) appear in better agreement with the observed metallicity distribution of DLAs (see fig. 10 in Berry et al. 2014). This difference is likely due to the different treatment adopted for the metal enrichment. In particular, the Somerville model assumes an instantaneous recycling approximation and sets a metallicity floor for the hot gas in low-mass haloes (the haloes with $M_{vir} \leq 10^9 \, M_\odot$ are set to have a hot gas metallicity equal to $10^{-3} Z_{\rm solar}$). Our model instead does not assume pre-enrichment of gas in low-mass haloes and includes a detailed chemical enrichment scheme that accounts for the non-instantaneous recycling of gas and metals (De Lucia et al. 2014). In addition, as discussed above, we assume that the 95 per cent of newly synthesized metals is directly injected into the hot gas phase in low-mass haloes (Hirschmann et al. 2016), which contributes to delay the chemical enrichment of low-mass systems. Fig. 16 of Somerville et al. (2015) shows that their model predicts almost no evolution with redshift of the mass–metallicity relation, while our model predicts an increasing normalization at lower redshift. The different redshift evolution together with the different slope (less steep for low-mass galaxies in the case of the Somerville model) could lead to gas metallicities, for galaxies in the mass range $10^7 < M_\star < 10^8$, that are larger in the Somerville model than in ours. This could also contribute to the different predictions obtained for the metallicity distribution of DLAs, in particular at low metallicities.

For the typical DLA host halo mass, predictions from our model (both for the fiducial and the $2M-2R$ run) are similar to those by the Somerville model. For a mean redshift $z = 2.3$ (taking all DLAs with $1.97 \leq z \leq 2.6$), we find a median DLA host halo mass equal to $M_{\rm host}^{\rm DLA} = 1.55 \times 10^{11} \, M_\odot$ for the fiducial model and $M_{\rm host}^{\rm DLA} = 2.28 \times 10^{11} \, M_\odot$ for the $2M-2R$ model, in agreement with observational estimates by Pérez-Ràfols et al. (2018a).

5.2 Model developments

Our results suggest possible avenues to improve the agreement between the predictions of the GAEA model and observational data of DLAs: (i) increasing the H I content of model galaxies; (ii) increasing the sizes of gaseous discs; and (iii) modifying the treatment for the metal enrichment of low-mass haloes. In this section, we discuss plausible implementations that can bring the model in this direction. In future work, we intend to explore these suggestions in more detail.

The H I content of model galaxies depends on the assumed prescription for cold gas partitioning. In our model, the molecular to atomic hydrogen ratio is slightly larger than what observed in the local Universe by xGASS and xCOLDGASS (Saintonge et al. 2017; Catinella et al. 2018). This is shown and discussed in a forthcoming paper (Xie et al. 2020). Naively, one could think that a lower molecular fraction can be obtained by simply increasing the star formation efficiency: stars are formed from molecular gas and larger star formation rates should lead to consume more molecular hydrogen. The situation is, however, complicated by the strong self-regulation between star formation and stellar feedback that makes model results not very sensitive to the star formation law adopted (Xie et al. 2017, and references therein). In addition, simple modifications of model parameters would generally require a retuning of the model to restore the agreement with the main observables used as constraints (in our case the H I mass function).

Another possible reason for the too low H I masses of galaxies in intermediate mass haloes is the prescription adopted for reionization.

Our model assumes an 'early' reionization (with starting redshift $z_0 = 15$ and completed by $z_r \sim 11$) that is inconsistent with recent Planck results (Planck Collaboration et al. 2018). The reionization feedback is implemented through a 'filtering mass' whose evolution is described by the analytic fitting function introduced in Kravtsov, Gnedin & Klypin (2004) (based on the simulation results by Gnedin 2000). Adopting a time-line for reionization more in agreement with recent results, we expect a filtering mass lower by an order of magnitude with respect to the one assumed in our model for $z > 5$. Besides, Okamoto, Gao & Theuns (2008) showed that the parametrization of the filtering mass presented in Gnedin (2000), based on low-resolution simulations, might overestimate by up to one order of magnitude (at $z = 0$) the characteristic mass where photoionization feedback becomes effective in reducing the baryon fraction (see their fig. 6), independently of the assumed reionization history.

The size of the H I galactic discs in our model is determined by the evolution of the specific angular momentum of the cold gas. Xie et al. (2020) have significantly updated the treatment of the angular momentum, leading to both larger gaseous discs and larger H I masses, in the direction of the $2M-2R$ model considered in previous sections. It is worth noting that also the SFR sizes predicted by our fiducial model tend to be smaller than observational estimates (Xie et al. 2017), and also this disagreement is relieved with the larger discs obtained with the updated angular momentum scheme (as shown in Xie et al. 2020).

Finally, the excess of low-metallicity DLAs, with respect to observational measurements, can possibly be solved by modifying the fraction of metals that are injected directly into the hot gas component in low-mass haloes. This will likely affect also the cooling times (the cooling function is very sensitive to the metallicity of the hot gas), leading to a lower gas accretion rate on to galaxies, but could be compensated by the above-described modifications concerning the reionization scenario.

6 SUMMARY

In this work, we have investigated the properties of DLAs taking advantage of the state-of-the-art semi-analytical model GAEA (our fiducial model is the BR run, described in Xie et al. 2017). We have used model outputs obtained by considering two large cosmological simulations: the Millennium (MSI; Springel et al. 2005) and the Millennium II simulation (MSII; Boylan-Kolchin et al. 2009), with higher resolution but smaller box. We consider also the contribution of DM haloes below the resolution of the simulation adopting a simple HOD approach and populating the box with isolated DM haloes in the mass range $10^8 < M_{200} < 10^9 \, M_\odot$.

From the comparison of the GAEA model predictions with DLA observations in the redshift range $2 < z < 3$, we find that the fiducial GAEA model reproduces the overall shape of the CDDF, but predicts a CDDF and an $\Omega_{\rm DLA}^{\rm H I}$ that are offset systematically below the observational measurements. The agreement with observations is significantly improved, at least up to $z \sim 3$, increasing 'a posteriori' both the H I masses and the gaseous disc radii of model galaxies by a factor of 2 (we have referred to this as the $2M-2R$ model in the text). At higher redshift ($z > 3$), our predicted $\Omega_{\rm DLA}^{\rm H I}$ decreases in both the model versions considered, in disagreement with observational measurements and differently from what happens in hydrodynamical simulations.

Our analysis of the relative contribution to the DLA comoving H I density of simulated DM haloes, in bins of virial mass, highlights that DM haloes with $M_{200} < 10^9 \, M_\odot$ (under the resolution of the

adopted *N*-body simulations) do not give a significative contribution to $\Omega_{\mathrm{H\,I}}$ up to $z \sim 4$, in the framework of our study.

Our model predicts a population of DLAs with very low abundance ratios ([Fe/H] < -2.5), not in agreement with the observed metallicity floor (Storrie-Lombardi & Wolfe 2000; Prochaska & Wolfe 2009; Rafelski et al. 2012), suggesting to explore the possibility of a modification of our prescription for the metal ejection in low-mass haloes. At the same time, the simulated DLA with high column density has an average metallicity larger than the observed one, leading to a consequent relatively high normalization of the values of the cosmic DLA metallicity (Ω_Z). This discrepancy is mitigated when a correction accounting for the metallicity gradient in galaxies, based on the fitting formula by Christensen et al. (2014), is applied.

Our model predicts a median DLA halo host mass of $\sim 10^{11}$ M$_\odot$, in agreement with the results of the work by Pérez-Ràfols et al. (2018a) on the DLA Ly α cross-correlation analysis.

The predicted DLA impact parameters have a distribution which is also in agreement with the estimates derived by Krogager et al. (2017), in particular in the case of the model with larger galactic discs ($2M{-}2R$ model).

The picture emerging from the present analysis, which includes a detailed comparison with the similar work by Berry et al. (2014, 2016), suggests possible improvements of the physical prescriptions of our model, in particular the H I content of galaxies, the sizes of gaseous discs and the metal enrichment of low-mass haloes.

ACKNOWLEDGEMENTS

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author, Serafina Di Gioia.

REFERENCES

Altay G., Theuns T., Schaye J., Booth C. M., Dalla Vecchia C., 2013, MNRAS, 436, 2689

Altay G., Theuns T., Schaye J., Crighton N. H. M., Dalla Vecchia C., 2011, ApJ, 737, L37

Arinyo-i-Prats A., Mas-Ribas L., Miralda-Escudé J., Pérez-Ràfols I., Noterdaeme P., 2018, MNRAS, 481, 3921

Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481

Barnes L. A., Garel T., Kacprzak G. G., 2014, PASP, 126, 969

Barnes L. A., Haehnelt M. G., 2009, MNRAS, 397, 511

Barteldrees A., Dettmar R. J., 1994, A&AS, 103, 475

Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379

Bennett C. L. et al., 2013, ApJS, 208, 20

Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587

Berry M., Somerville R. S., Gawiser E., Maller A. H., Popping G., Trager S. C., 2016, MNRAS, 458, 531

Berry M., Somerville R. S., Haas M. R., Gawiser E., Maller A., Popping G., Trager S. C., 2014, MNRAS, 441, 939

Bird S., Haehnelt M., Neeleman M., Genel S., Vogelsberger M., Hernquist L., 2015, MNRAS, 447, 1834

Bird S., Vogelsberger M., Haehnelt M., Sijacki D., Genel S., Torrey P., Springel V., Hernquist L., 2014, MNRAS, 445, 2313

Blitz L., Rosolowsky E., 2006, ApJ, 650, 933

Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150

Carton D. et al., 2018, MNRAS, 478, 4293

Catinella B. et al., 2010, MNRAS, 403, 683

Catinella B. et al., 2013, MNRAS, 436, 34

Catinella B. et al., 2018, MNRAS, 476, 875

Cen R., 2012, ApJ, 748, 121

Chen H.-W., Lanzetta K. M., 2003, ApJ, 597, 706

Christensen L., Møller P., Fynbo J. P. U., Zafar T., 2014, MNRAS, 445, 225

Christensen L., Møller P., Rhodin N. H. P., Heintz K. E., Fynbo J. P. U., 2019, MNRAS, 489, 2270

Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201

Conselice C. J., Mortlock A., Bluck A. F. L., Grützbauch R., Duncan K., 2013, MNRAS, 430, 1051

Cooke J., Wolfe A. M., Gawiser E., Prochaska J. X., 2006, ApJ, 652, 994

Crighton N. H. M. et al., 2015, MNRAS, 452, 217

De Cia A., Ledoux C., Petitjean P., Savaglio S., 2018, A&A, 611, A76

De Lucia G., Blaizot J., 2007, MNRAS, 375, 2

De Lucia G., Tornatore L., Frenk C. S., Helmi A., Navarro J. F., White S. D. M., 2014, MNRAS, 445, 970

Durkalec A. et al., 2015, A&A, 576, L7

Eisenstein D. J. et al., 2011, AJ, 142, 72

Elmegreen B. G., Elmegreen D. M., Tompkins B., Jenks L. G., 2017, ApJ, 847, 14

Font-Ribera A. et al., 2012, J. Cosmol. Astropart. Phys., 2012, 059

Fraternali F., 2014, in Feltzing S., Zhao G., Walton N. A., Whitelock P., eds, Proc. IAU Symp. Vol. 298, Setting the Scene for Gaia and LAMOST. p. 228

Fumagalli M., O'Meara J. M., Prochaska J. X., Kanekar N., 2010, MNRAS, 408, 362

Fumagalli M., O'Meara J. M., Prochaska J. X., Rafelski M., Kanekar N., 2015, MNRAS, 446, 3178

Fumagalli M., Prochaska J. X., Kasen D., Dekel A., Ceverino D., Primack J. R., 2011, MNRAS, 418, 1796

Fynbo J. P. U., Krogager J. K., Venemans B., Noterdaeme P., Vestergaard M., Møller P., Ledoux C., Geier S., 2013, ApJS, 204, 6

Fynbo J. P. U., Prochaska J. X., Sommer-Larsen J., Dessauges-Zavadsky M., Møller P., 2008, ApJ, 683, 321

Fynbo J. P. U. et al., 2010, MNRAS, 408, 2128

Fynbo J. P. U. et al., 2011, MNRAS, 413, 2481

Fynbo J. U., Møller P., Warren S. J., 1999, MNRAS, 305, 849

Gardner J. P., Katz N., Hernquist L., Weinberg D. H., 1997, ApJ, 484, 31

Gardner J. P., Katz N., Hernquist L., Weinberg D. H., 2001, ApJ, 559, 131

Giovanelli R. et al., 2005, AJ, 130, 2598

Gnedin N. Y., 2000, ApJ, 542, 535

Guo Q., White S., Angulo R. E., Henriques B., Lemson G., Boylan-Kolchin M., Thomas P., Short C., 2013, MNRAS, 428, 1351

Haehnelt M. G., Steinmetz M., Rauch M., 1998, ApJ, 495, 647

Hassan S., Finlator K., Davé R., Churchill C. W., Prochaska J. X., 2020, MNRAS, 492, 2835

Hirschmann M., De Lucia G., Fontanot F., 2016, MNRAS, 461, 1760

Khare P., Vanden Berk D., York D. G., Lundgren B., Kulkarni V. P., 2012, MNRAS, 419, 1028

Kim H.-S., Wyithe J. S. B., Power C., Park J., Lagos C. d. P., Baugh C. M., 2015, MNRAS, 453, 2315

Kravtsov A. V., Gnedin O. Y., Klypin A. A., 2004, ApJ, 609, 482

Kregel M., van der Kruit P. C., de Grijs R., 2002, MNRAS, 334, 646

Krogager J. K., Møller P., Fynbo J. P., Noterdaeme P., 2017, MNRAS, 469, 2959

Krogager J. K. et al., 2015, ApJS, 217, 5

Krogager J. K. et al., 2016, ApJ, 832, 49

Kulkarni V. P., Khare P., Péroux C., York D. G., Lauroesch J. T., Meiring J. D., 2007, ApJ, 661, 88

Kulkarni V. P., Khare P., Som D., Meiring J., York D. G., Péroux C., Lauroesch J. T., 2010, New Astron., 15, 735

Lagos C. D. P., Baugh C. M., Lacey C. G., Benson A. J., Kim H.-S., Power C., 2011, MNRAS, 418, 1649

Lagos C. D. P., Baugh C. M., Zwaan M. A., Lacey C. G., Gonzalez-Perez V., Power C., Swinbank A. M., van Kampen E., 2014, MNRAS, 440, 920
Lah P. et al., 2007, MNRAS, 376, 1357
Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2782
Li Y.-S., De Lucia G., Helmi A., 2010, MNRAS, 401, 2036
Mac Low M.-M., Ferrara A., 1999, ApJ, 513, 142
Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, ApJ, 723, 1359
Ma X., Hopkins P. F., Feldmann R., Torrey P., Faucher-Giguère C.-A., Kereš D., 2017, MNRAS, 466, 4780
Meyer M. J. et al., 2004, MNRAS, 350, 1195
Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121
Møller P., Christensen L., 2020, MNRAS, 492, 4805
Møller P., Warren S. J., 1998, MNRAS, 299, 661
Møller P., Warren S. J., Fall S. M., Fynbo J. U., Jakobsen P., 2002, ApJ, 574, 51
Nagamine K., Springel V., Hernquist L., 2004, MNRAS, 348, 421
Nagamine K., Wolfe A. M., Hernquist L., Springel V., 2007, ApJ, 660, 945
Narayan C. A., Jog C. J., 2002, A&A, 394, 89
Neeleman M., Kanekar N., Prochaska J. X., Rafelski M. A., Carilli C. L., 2019, ApJ, 870, L19
Neeleman M., Wolfe A. M., Prochaska J. X., Rafelski M., 2013, ApJ, 769, 54
Noterdaeme P., Srianand R., Rahmani H., Petitjean P., Pâris I., Ledoux C., Gupta N., López S., 2015, A&A, 577, A24
Noterdaeme P. et al., 2012, A&A, 547, L1
Okamoto T., Gao L., Theuns T., 2008, MNRAS, 390, 920
Padmanabhan H., Refregier A., Amara A., 2017, MNRAS, 469, 2323
Peebles P. J. E., 1984, Hierarchical Clustering. Springer Netherlands, Dordrecht, p. 405
Peebles P. J. E., McCrea W. H., 1981, J. Br. Astron. Assoc., 91, 411
Planck Collaboration I, 2016, A&A, 594, A1
Planck Collaboration VI, 2018, preprint (arXiv:1807.06209)
Pontzen A. et al., 2008, MNRAS, 390, 1349
Popping G., Somerville R. S., Trager S. C., 2014, MNRAS, 442, 2398
Prochaska J. X., Castro S., Djorgovski S. G., 2003, ApJS, 148, 317
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
Pérez-Ràfols I., Miralda-Escudé J., Arinyo-i-Prats A., Font-Ribera A., Mas-Ribas L., 2018b, MNRAS, 480, 4702
Pérez-Ràfols I. et al., 2018a, MNRAS, 473, 3019
Péroux C., Bouché N., Kulkarni V. P., York D. G., Vladilo G., 2011, MNRAS, 410, 2237
Péroux C., McMahon R. G., Storrie-Lombardi L. J., Irwin M. J., 2003, MNRAS, 346, 1103
Péroux C. et al., 2019, MNRAS, 485, 1595
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., Pawlik A. H., Raičević M., Schaye J., 2013, MNRAS, 430, 2427
Rahmati A., Schaye J., 2014, MNRAS, 438, 529
Rao S. M., Belfort-Mihalyi M., Turnshek D. A., Monier E. M., Nestor D. B., Quider A., 2011a, MNRAS, 416, 1215
Rao S. M., Belfort-Mihalyi M., Turnshek D. A., Monier E. M., Nestor D. B., Quider A., 2011b, MNRAS, 416, 1215
Rao S. M., Turnshek D. A., Nestor D. B., 2006, ApJ, 636, 610
Razoumov A. O., 2009, ApJ, 707, 738
Rhodin N. H. P., Agertz O., Christensen L., Renaud F., Uldall Fynbo J. P., 2019, MNRAS, 488, 3634
Rhodin N. H. P., Christensen L., Moller P., Zafar T., Fynbo J. P. U., 2018, A&A, 618, A129
Rice W., Merrill K. M., Gatley I., Gillett F. C., 1996, AJ, 112, 114
Rosenberg J. L., Schneider S. E., 2003, ApJ, 585, 256

Saintonge A. et al., 2017, ApJS, 233, 22
Schaye J., 2001, ApJ, 559, 507
Schneider D. P. et al., 2010, AJ, 139, 2360
Silk J., Mamon G. A., 2012, Res. Astron. Astrophys., 12, 917
Somerville R. S., Davé R., 2015, ARA&A, 53, 51
Somerville R. S., Popping G., Trager S. C., 2015, MNRAS, 453, 4337
Somerville R. S. et al., 2008, ApJ, 672, 776
Sorini D., Oñorbe J., Hennawi J. F., Lukić Z., 2018, ApJ, 859, 125
Spinelli M., Zoldan A., De Lucia G., Xie L., Viel M., 2020, MNRAS, 493, 5434
Spitzer Lyman J., 1942, ApJ, 95, 329
Spring E. F., Michałowski M. J., 2017, MNRAS, 471, L101
Springel V., Hernquist L., 2003, MNRAS, 339, 289
Springel V. et al., 2005, Nature, 435, 629
Srianand R., Hussain T., Noterdaeme P., Petitjean P., Krühler T., Japelj J., Pâris I., Kashikawa N., 2016, MNRAS, 460, 634
Storrie-Lombardi L. J., Wolfe A. M., 2000, ApJ, 543, 552
Stott J. P. et al., 2014, MNRAS, 443, 2695
Tescari E., Viel M., Tornatore L., Borgani S., 2009, MNRAS, 397, 411
Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709
van der Kruit P. C., Freeman K. C., 2011, ARA&A, 49, 301
van der Kruit P. C., Searle L., 1982, A&A, 110, 61
van de Voort F., Schaye J., Altay G., Theuns T., 2012, MNRAS, 421, 2809
van de Voort F., Schaye J., Booth C. M., Haas M. R., Dalla Vecchia C., 2011, MNRAS, 414, 2458
Villaescusa-Navarro F. et al., 2018, ApJ, 866, 135
Wang J. et al., 2014, MNRAS, 441, 2159
Wang W.-H., Kanekar N., Prochaska J. X., 2015, MNRAS, 448, 2832
Wang Y., 2008, J. Cosmol. Astropart. Phys., 2008, 021
Whitney A., Conselice C. J., Bhatawdekar R., Duncan K., 2019, ApJ, 887, 113
Xie L., De Lucia G., Hirschmann M., Fontanot F., 2020, preprint (arXiv:2003.12757)
Xie L., De Lucia G., Hirschmann M., Fontanot F., Zoldan A., 2017, MNRAS, 469, 968
Zafar T., Péroux C., Popping A., Milliard B., Deharveng J. M., Frank S., 2013, A&A, 556, A141
Zafar T. et al., 2015, A&A, 584, A100
Zoldan A., De Lucia G., Xie L., Fontanot F., Hirschmann M., 2017, MNRAS, 465, 2236
Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, MNRAS, 359, L30

APPENDIX A: EVOLUTION OF THE DLA STATISTICS WITH REDSHIFT AND IN DIFFERENT MASS BINS

Observations indicate a negligible evolution of the CDDF as a function of redshift (Noterdaeme et al. 2012), while other theoretical studies based on hydrodynamical simulations (Rahmati et al. 2013) found little evolution of the low column density end, with the slope becoming steeper at higher redshift. In contrast with observations, our model predicts a moderate evolution of the CDDF, in particular of the low column density end, that flattens at lower z. To understand the origin of this evolution, it can be useful to investigate how different DLA host halo masses are distributed in different column density bins.

Fig. A1 shows the predicted CDDF at three different redshifts ($z =$ 2.83, 2.42, 2.07), with the dashed lines highlighting the contribution of haloes of different mass, and the bottom (top) panel showing the results of our $2M-2R$ (fiducial) model.

Haloes in the mass bin $10^{11}\,\mathrm{M_\odot} \leq M_{200} < 10^{12}\,\mathrm{M_\odot}$ represent the major contribution to the CDDF at all column densities
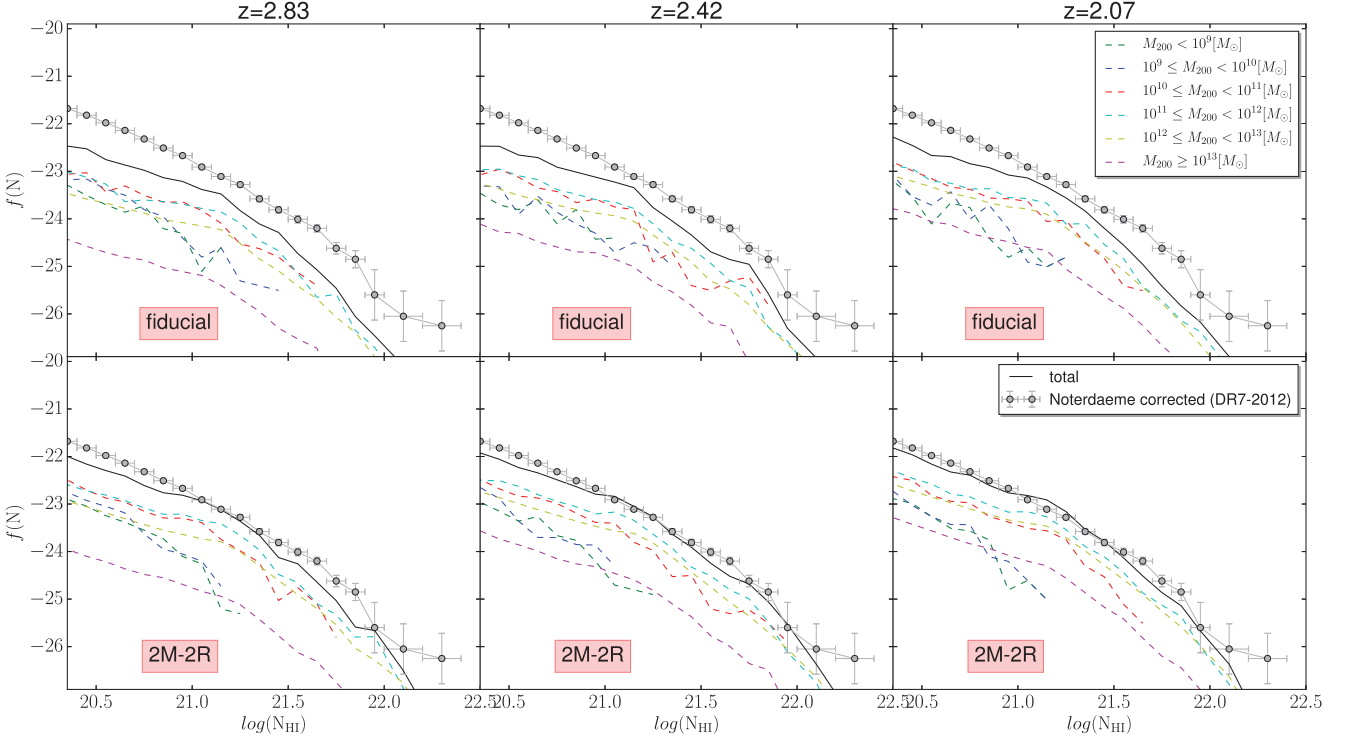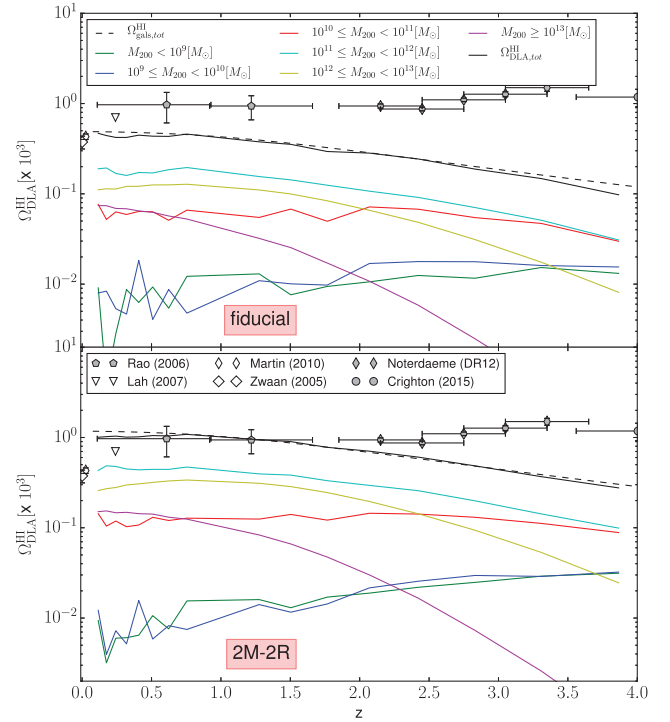
Figure A1. Evolution with redshift of the predicted CDDF and its dependence on the DLA host halo masses. The top panels show results from our fiducial model, while the bottom panels show the corresponding results from the 2M–2R model. The black solid lines show the total CDDF, while the dashed coloured lines show the average contribution to the CDDF of dark matter haloes in different virial mass bins, as indicated in the legend.

– a contribution that decreases at higher redshift, as expected in a hierarchical scenario. The second major contribution come from haloes in the mass bin $10^{10}\,\mathrm{M_\odot} \leq M_{200} < 10^{11}\,\mathrm{M_\odot}$ for intermediate/low DLA column densities and from haloes in the mass bin $10^{12}\,\mathrm{M_\odot} \leq M_{200} < 10^{13}\,\mathrm{M_\odot}$ for high DLA column densities. Haloes with $M_{200} < 10^{10}\,\mathrm{M_\odot}$ start to contribute significantly for $\log(N_{\rm H\,I}) < 20.7$ around $z \sim 2.8$ and moving to higher redshift their relative contribute to all column densities increases.

In the 2M–2R model, the contribution to the CDDF of the haloes with $M_{200} > 10^{11}\,\mathrm{M_\odot}$ increases at all column densities, while the contribution of low-mass haloes ($M_{200} < 10^{10}\,\mathrm{M_\odot}$) increases only at low column densities (for $N_{\rm H\,I} < 21$) and that of the intermediate mass ($10^{10}\,\mathrm{M_\odot} < M_{200} < 10^{11}\,\mathrm{M_\odot}$) haloes increases for intermediate column densities (up to $N_{\rm H\,I} = 1.4$), with respect to the fiducial model. This could be explained remembering that the H I surface density scales linearly with the mass and as the inverse of the square radius of the galactic disc. Moreover, since the cross-section increases quadratically with the galaxy scale radius, a larger number of discs are intersected by our l.o.s. in the 2M–2R model, in particular at larger halo masses. Therefore, the predicted CDDF from this model is in better agreement with observational measurements.

In Fig. A2 we show, for the redshift range $0 < z < 4$, the contributions to the predicted $\Omega_{\rm DLA}^{\rm H\,I}$ of haloes in different mass bins, with logarithmic bin size $\Delta \log(M_{200}/\mathrm{M_\odot}) = 1$. The haloes that contribute more at all redshifts are those in the mass bin $10^{11}\,\mathrm{M_\odot} \leq M_{200} < 10^{12}\,\mathrm{M_\odot}$. The second largest contribution is provided by the mass bin $10^{12}\,\mathrm{M_\odot} \leq M_{200} < 10^{13}\,\mathrm{M_\odot}$ up to $z = 2.5$ and by the mass bin $10^{10}\,\mathrm{M_\odot} \leq M_{200} < 10^{11}\,\mathrm{M_\odot}$ for $z > 2.5$. It is worth noting that the contributions of the two lowest mass bins are very similar and

Figure A2. Evolution with redshift of the predicted comoving H I density in DLAs ($\Omega_{\rm H\,I}^{\rm DLA}$) and its dependence on the DLA host halo masses. The top panel shows results from our fiducial model, while the bottom panel shows the corresponding results from the 2M–2R model. The black solid lines show the total $\Omega_{\rm H\,I}^{\rm DLA}$, while the dashed coloured lines show the average contribution to $\Omega_{\rm H\,I}^{\rm DLA}$ of dark matter haloes in different virial mass bins, as indicated in the legend.

both represent less than 10 per cent of the total $\Omega_{\mathrm{DLA}}^{\mathrm{H\,I}}$ in the entire redshift range considered.

APPENDIX B: INFLUENCE OF DIFFERENT GAS VERTICAL DENSITY PROFILES ON THE ESTIMATED DLA PROPERTIES

The distribution of H I gas detected through the 21 cm line is fairly flat and uniform (Leroy et al. 2008), with a scale length much larger than stellar disc one. The work by Narayan & Jog (2002) suggests that the vertical structure of the gaseous disc is sensitive to the gravity of all galactic components, i.e. stars, dark matter, and gas. Under the assumption of an isothermal distribution, one expects that the gaseous/stellar vertical density profile is described by the function sech^2, as shown theoretically by Spitzer (1942) and confirmed by some observations (van der Kruit & Searle 1982). However, more recent observational studies have found that the observed vertical distribution for gas and stars in galaxies is steeper than the one predicted by an isothermal distribution, and it is well-approximated by an exponential or a sech function, especially close to the galactic mid-plane (Barteldrees & Dettmar 1994; Rice et al. 1996).

We have considered the effect on the DLA CDDF of assuming a different vertical density profile for the model galaxies.

We assume four different density profiles for the gas in the galactic disc: the 'classic' double-exponential

$$\rho_{\mathrm{CG}}(r, z) = \rho_0 \, e^{-r/R_s} \, e^{-z/z_0}$$

and three additional profiles, described by the formula presented in van der Kruit & Freeman (2011):

$$\rho_{\mathrm{CG}}(r, z) = \rho_0 \, e^{-r/R_s} \operatorname{sech}^{\frac{2}{n}}\left(\frac{n\,z}{2z_0}\right)$$

with $n = 1$, 2, and 4, respectively.

In Fig B1, the predicted average CDDF in the redshift range $2 < z < 3$ is shown. The solid line refers to the exponential vertical profile, while the star-dashed, dot–dashed, and dashed line refer to the function presented in van der Kruit & Freeman (2011), respectively, with $n = 1$, 2, 4.

Fig. B1 highlights that the four different density profiles lead to differences in the CDDF only in the high column density regime. And for large n-values the class of functions presented by van der Kruit & Freeman (2011) give very similar results to the exponential density profile for all the column densities considered.
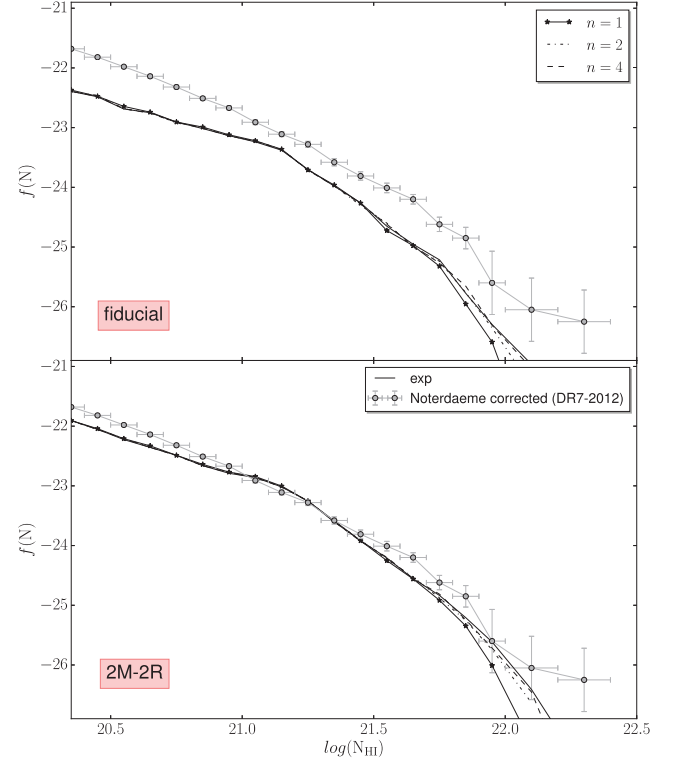
Figure B1. Average CDDF in the redshift range $2 < z < 3$, compared to the data (grey dots) by Noterdaeme et al. (2012). The top (bottom) panel shows the results from the fiducial ($2M-2R$) model. The black lines describe the total CDDF, while the other lines show the average contribution to the CDDF assuming different density profiles, as described in the legend.

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