Characterizing the abundance, properties, and kinematics of the cool circumgalactic medium of galaxies in absorption with SDSS DR16

Abhijeet Anand\textsuperscript{1}*, Dylan Nelson\textsuperscript{1,2}, Guinevere Kauffmann\textsuperscript{1}

\textsuperscript{1}Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85741 Garching, Germany
\textsuperscript{2}Universität Heidelberg, Zentrum für Astronomie, Institut für theoretische Astrophysik, Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany

31 March 2021

ABSTRACT

In order to study the circumgalactic medium (CGM) of galaxies we develop an automated pipeline to estimate the optical continuum of quasars and detect intervening metal absorption line systems with a matched kernel convolution technique and adaptive S/N criteria. We process \textasciitilde one million quasars in the latest Data Release 16 (DR16) of the Sloan Digital Sky Survey (SDSS) and compile a large sample of \textasciitilde 160,000 Mg\textsubscript{i}pc\textsubscript{i}pc absorbers, together with \textasciitilde 70,000 Fe\textsubscript{i}pc\textsubscript{i}pc systems, in the redshift range $0.35 < z_{\text{abs}} < 2.3$. Combining these with the SDSS DR16 spectroscopy of \textasciitilde 1\textasciitilde 1 million luminous red galaxies (LRGs) and \textasciitilde 200,000 emission line galaxies (ELGs), we investigate the nature of cold gas absorption at $0.5 < z < 1$. These large samples allow us to characterize the scale dependence of Mg\textsubscript{i}pc\textsubscript{i}pc with greater accuracy than in previous work. We find that there is a strong enhancement of Mg\textsubscript{i}pc\textsubscript{i}pc absorption within 50 kpc of ELGs, and the covering fraction within 0.5$r_{\text{vir}}$ of ELGs is 2-5 times higher than for LRGs. Beyond 50 kpc, there is a sharp decline in Mg\textsubscript{i}pc\textsubscript{i}pc for both kinds of galaxies, indicating a transition to the regime where the CGM is tightly linked with the dark matter halo. The Mg\textsubscript{i}pc\textsubscript{i}pc covering fraction correlates strongly with stellar mass for LRGs, but weakly for ELGs, where covering fractions increase with star formation rate. Our analysis implies that cool circumgalactic gas has a different physical origin for star forming versus quiescent galaxies.

Key words: galaxies: formation – galaxies: evolution – large-scale structure of Universe

1 INTRODUCTION

The gaseous halo surrounding galaxies, known as the circumgalactic medium (CGM), holds important clues to the process of galaxy formation and evolution. The CGM is the main venue for cosmological gas accretion and inflows, as well as galactic outflows produced by feedback processes. These flows pass through the CGM and play a pivotal role in several key processes regulating galaxy formation, such as determining the timescales of gas depletion and star formation (Whitaker et al. 2012), the origin of the observed bimodality and the way in which quenching occurs (Tumlinson et al. 2017; Schiminovich et al. 2010), and the distribution of the baryonic budget as well as the total metal content of the gas surrounding galaxies (Behroozi et al. 2010; Peeples et al. 2014).

The formation of galaxies and their evolution is driven by the gas flows from the CGM (see, Tumlinson et al. 2017, for a review), implying that our understanding of galaxy formation is itself limited by our current understanding of the CGM. However, large sky surveys performed with ground and space-based telescopes such as SDSS, Keck, VLT, and HST have significantly deepened our understanding of the CGM over the past two decades.

In this regard, one of the most powerful tools has been transverse absorption line studies where the CGM is observed in absorption against a bright background source such as a quasar. Different metal absorbers detected at redshifts smaller than the redshift of the background source provide direct observational constraints on the gas flows around galaxies at different epochs (Zhu & Ménard 2013b; Turner et al. 2014; Huang et al. 2021).

Galaxy-absorber pair studies have been at the centre of our understanding into the nature of the CGM since the first discovery of a galaxy-metal absorber pair by Bergeron (1986). More recently, there have been several galaxy-absorber pair studies investigating the physical relationships between galaxies and the CGM, covering different types of galaxies over a broad range of redshifts (Steidel et al. 1994; Churchill et al. 2005; Chen et al. 2010; Nielsen et al. 2013). These studies have found that the amount of gas varies strongly with distance from galaxies, and can depend on galaxy stellar mass, star formation history, color and shape (Bordoloi et al. 2014; Borthakur et al. 2016; Lopez et al. 2018; Lan & Mo 2018; Rubin et al. 2018).

Among all the detected metal lines the most extensively ex-
amined is the Mg II $\lambda\lambda2796$, 2803 doublet. It is one of the strongest absorption features that can be detected by ground based optical telescopes at modest redshifts ($0.3 \leq z \leq 2.5$). Mg II absorbers are tracers of low-ionization cold gas ($\leq 10^4$ K) in the CGM and in the intergalactic medium (IGM). Several Mg II surveys have been performed to constrain the temperature and density profile of low-ionization gas in the CGM along with the statistical properties of weak (EW$_{\text{rest}} < 0.3$ Å) and strong (EW$_{\text{rest}} > 0.3$ Å) Mg II absorbers (Weymann et al. 1979; Tytler et al. 1987; Sargent et al. 1988; Caulet 1989; Steidel & Sargent 1992; Churchill et al. 1999, 2000; York et al. 2006; Quider et al. 2011; Zhu & Ménard 2013b). In addition, C iv $\lambda1548$, 1550 is also commonly probed, tracing $\sim 10^5$ K gas in the CGM around galaxies. Cooksey et al. (2013) compiled a large C iv absorber catalogue detected in SDSS DR7 quasars and quantified the number of C iv absorbers as a function of rest equivalent width. They found a monotonic and significant increase in comoving line density of C iv around galaxies over a large range of redshift, $0 < z < 6$. Chen et al. (2016) compiled a similar C iv absorber catalogue using SDSS DR9 quasars and compared C iv properties with Mg ii systems, finding generally larger velocity dispersion for C iv.

Pieri et al. (2014) stacked several hundred thousand Ly α absorbers at 2.4 $< z <$ 3.1 to probe the CGM at high redshift. They estimated the average neutral hydrogen column densities in those systems and found evidence for metallicities near the solar value. On the other hand, most similar to our present work, Zhu et al. (2014) performed a detailed statistical galaxy-absorber pair study with DR7 Mg ii absorbers (Zhu & Ménard 2013a) and Luminous Red Galaxies (LRGs) from DR11 of SDSS (Dawson et al. 2013). They estimated average cool surface densities as traced by Mg ii around massive galaxies and compared with the expected gas distribution in the parent halo together with gas outside the halo. Extending this analysis, Lan et al. (2014) found that both LRGs and ELGs have high covering fractions of cold gas (~ 1 percent) even at impact parameters of 500 kpc.

Recently, Lan & Mo (2018) have proceeded with a larger dataset of several thousand galaxy-quasar pairs using the Extended baryon Oscillation Spectroscopic Survey (eBOSS; Dawson et al. 2016) in the Sloan Digital Sky Survey IV (SDSS-IV; Blanton et al. 2017) finding an anisotropic metal absorption distribution around emission line galaxies (ELGs). They also observed the amount of cool gas to be different around ELGs and LRGs. There have also been statistical studies investigating the nature of the CGM around star-forming versus quiescent galaxies (Bordoloi et al. 2011; Ménard et al. 2011; Peek et al. 2015; Huang et al. 2016; Lan 2020; Huang et al. 2021). It is observed that the mean covering fraction of Mg ii absorbers varies strongly with galaxy type. Similarly, Zibetti et al. (2005, 2007) performed image stacking to study the photometric properties of Mg ii systems in SDSS and found that weak and strong absorbers are originated in different types of galaxies. Apart from these galaxy centric CGM studies, large spectroscopic surveys have also provided the unprecedented opportunity to perform statistical studies and cross-correlations exclusively with absorbers (Quider et al. 2011; Zhu & Ménard 2013b; Zhu et al. 2014).

The distribution of velocity separations between galaxies and absorbers allows the kinematics of cold gas around galaxies to be investigated (Tremonti et al. 2007). For example, the clustering of Mg ii systems around BOSS LRGs ($M_\star \sim 10^{11.5} M_\odot$) shows an excess of Mg ii up to $R_p = 20$ Mpc, as well as relative velocities of $\Delta v \sim 10000$ km s$^{-1}$ within a projected distance of $\leq 800$ kpc (Kauffmann et al. 2017). The implication is that cool circumgalactic gas can originate in either supernovae or supermassive black hole driven outflows, as well as due to infall and accretion.

Since the first light of Sloan Digital Sky Survey (SDSS; York et al. 2000) more than a million quasar spectra have been observed, and these can be searched for intervening absorber systems. Several detection (both automated and visual) algorithms have been developed to find Mg ii absorption lines in SDSS quasars (Nestor et al. 2005; York et al. 2006; Bouché et al. 2006; Prochter et al. 2006; Lundgren et al. 2009; Quider et al. 2011; Zhu & Ménard 2013a). With the continuous release of ever larger datasets, it is increasingly important to develop efficient automated pipelines to detect absorber systems in background sources, and thereby study gas absorption as a function of galaxy properties.

In this paper, we develop an automated continuum estimation and absorption detection pipeline using the approach of Zhu & Ménard (2013a) as our starting point. We run our pipeline on the full DR16 quasar sample and study the statistical variation of absorbers as well as galaxy-absorber pairs to measure the physical properties of CGM gas as a function of galaxy mass, star formation activity, impact parameter, and redshift.

The paper is divided into five sections: Section 2 introduces the observational data and describes our methods for continuum estimation and automatic absorption detection. We explore the resulting Mg ii absorber catalogue and its statistical properties in Section 3. In Section 4 we analyse the galaxy-centric CGM properties of ELGs and LRGs. Finally we discuss the implications of our findings in Section 5 and summarize the results in Section 6.

2 METHODS

2.1 Quasar catalogue

The latest Data Release 16 (DR16) quasar catalogue$^1$ compiled by Lyke et al. (2020) was released in late July 2020 as part of Value Added catalogue (VAC) of latest SDSS DR16 (Ahumada et al. 2020). Each SDSS data release is cumulative and includes all the objects observed in any previous release. The latest DR16 quasar catalogue contains 750,414 quasars. However, for our analysis we download the spectra of all the objects that are classified as QSOs in SDSS database$^2$. It includes 983,317 objects classified as quasars (QSOs) with $0 < z < 7$ ($\sim 3,000$ QSOs with $z > 4.8$).

In order to create robust continua of quasars we use the previously available quasar catalogue, DR14Q (Pâris et al. 2018; Abolfathi et al. 2018). It contains all quasars that were observed in SDSS-III/III (York et al. 2000; Schneider et al. 2010; Eisenstein et al. 2011; Dawson et al. 2013) and SDSS-IV/eBOSS (Dawson et al. 2016; Blanton et al. 2017) and classified as quasars with SDSS pipeline (Bolton et al. 2012). In order to compile a complete and pure catalogue, pipeline classified quasars were inspected visually to remove failed or uncertain classifications. The completeness and purity within a given target selection are as high as 99.5 percent in the catalogue. This high fidelity sample enables us to construct eigenspectra, as described below.

2.2 Continuum Estimation

The estimation of a robust continuum for a quasar is a crucial step in detecting absorbers in its spectrum. Among the empirical meth-

---

$^1$ https://www.sdss3.org/dr16/algorithms/qso_catalogue/

$^2$ https://dr16.sdss.org/optical/spectrum/search
First, we define a wavelength search window using the quasar intrinsic C\footnotesize{\textsc{iv}} emission line, or the \textit{rest} wavelength of Mg \footnotesize{\textsc{ii}} line. We smooth this residual with a median filter of kernel size = 141 pixels (~8 times the typical width of Mg \footnotesize{\textsc{ii}} lines) to remove intermediate-scale fluctuations. We then derive a second NMF continuum as the product of the first NMF continuum and median filtered residual. We again compute the residual w.r.t. this second NMF continuum and smooth the residual with the median filter of kernel size = 71 (~4 times the typical width of Mg \footnotesize{\textsc{ii}} lines) to remove small-scale fluctuations. The final NMF continuum is then computed by multiplying the second NMF continuum with this median filtered residual.

### 2.3 Automatic Detection of Absorbers

The doublet nature of Mg \footnotesize{\textsc{ii}} absorbs motivates us to develop an automated algorithm to detect its presence. Our approach is broken up into the following steps.

#### 2.3.1 Wavelength Search Window

First, we define a wavelength search window using the quasar intrinsic emission lines. We start from \( \Delta z = 0.018 \) redshifted from the quasar’s C\footnotesize{\textsc{iv}} line, or the blue end of SDSS (~3800 Å), and end at \( \Delta z = 0.03 \) blueshifted from quasar’s Mg \footnotesize{\textsc{ii}} emission line, or the red end of SDSS (~9200 Å). We select these \( \Delta z \) values to detect intervening absorbers rather than quasar associated absorbers. This also avoids misidentification due to the cases where the continuum is not well fitted to the quasar’s intrinsic C\footnotesize{\textsc{iv}} or Mg \footnotesize{\textsc{ii}} emission.

### Table 1. Details of our normalization scheme used in the NMF fitting for DR16 QSOs, using eigenspectra constructed from DR14 QSOs.

| Normalization Wavelength Range (QSO rest-frame) | Eigenspectra Construction Redshift Range (QSO) | Continuum Fitting Range (QSO) |
|-----------------------------------------------|-----------------------------------------------|--------------------------------|
| 4150-4250 Å                                  | \( z < 1.0 \)                                  | \( z < 0.97 \)                  |
| 3020-3100 Å                                  | \( 0.4 < z < 1.8 \)                            | \( 0.97 < z < 1.49 \)            |
| 2150-2250 Å                                  | \( 0.8 < z < 2.8 \)                            | \( 1.49 < z < 2.10 \)            |
| 1420-1500 Å                                  | \( 2.0 < z < 4.8 \)                            | \( 2.10 < z < 4.8 \)             |

### Table 2. Parameters of threshold and weighting scheme. All parameters are set by atomic physics. Note that \( \beta = \alpha/\delta \) and \( \delta = f_{\text{strong}}/f_{\text{weak}} \) and \( f_{\text{strong}} \) and \( f_{\text{weak}} \) denote the oscillator strengths of doublet components. For selecting an optimized \( \alpha \), we show the detailed analysis in Appendix A.

| Parameters | Mg \footnotesize{\textsc{ii}} |
|------------|-------------------------------|
| \( \alpha \) | 2.50                          |
| \( f_{\text{strong}} \) | 0.6123                       |
| \( f_{\text{weak}} \) | 0.3054                       |

| \( \lambda_{\text{prim}} \) | 2799.935 Å                     |
| \( \lambda_{s1} \) | 2792.765 Å                     |
| \( \lambda_{s2} \) | 2807.405 Å                     |
| \( \lambda_{\text{rest}} \) | 7.17 Å                         |
| \( \lambda_{a} \) | 2795.65 Å                      |
| \( \lambda_{b} \) | 2797.05 Å                      |
lines, which could produce spurious absorption dips. We also mask the possible Ca II λλ13934, 3969 lines (due to confusion with Mg ii absorbers at z ~ 0.4) and O i lines (15577 and 16300). In Figure 1 we show an example QSO spectrum in normalized flux (top) and residual (bottom), with the wavelength search window indicated.

2.3.2 Absorber Candidate Selection

Next, we search for potential absorbers by defining a Gaussian kernel that mimics the Mg ii λλ2796, 2803 doublet. We convolve the residual with this kernel and apply a threshold on the convolved array (C_R). To define the threshold we use the local noise estimate σ_C_R. To estimate this local noise for the convolved array, for each pixel we take the noise as the standard deviation of the convolved array within ±100 pixels around that pixel. This adaptive noise approach accounts for the noisy regions of the spectra, particularly at the edges. This convolution generically produces three peaks corresponding to the overlap of the kernel with the Mg ii doublet. The primary peak (λ_prim) corresponds to the case when two Gaussians of the kernel fully overlap with Mg ii lines. The two secondary peaks (λ_1, λ_2) correspond to the case when kernel overlaps just one of the lines. We apply the following thresholds on the primary and secondary peaks to identify potential absorbers, selecting all the pixels that satisfy our threshold criterion. For the threshold parameters see Table 2.

Rule 1: sigma criteria on primary peak:
- select all pixels \(\{1 + z_{abs} = \frac{\lambda_{abs}}{\lambda_{prim}}\}\) where \(C_R \leq \text{median}(C_R) - \alpha \cdot \sigma_{C_R}\). This is a sigma criteria on the convolved residual and median(C_R) implies global median of the convolved array.
- for each pixel with its associated redshift \(z_{abs}\) in \(\{z_{abs}\}\), define the neighbourhood of secondary peaks:
  \[\lambda_{sec1} = \lambda_{prim} \cdot (1 + z_{abs}) \pm 1.3\]
  \[\lambda_{sec2} = \lambda_{prim} \cdot (1 + z_{abs}) \pm 1.3\]
- measure the median residual amplitude for these pixels:
  \[S_{R1} = \text{median}(C_R, \lambda_{sec1})\]
  \[S_{R2} = \text{median}(C_R, \lambda_{sec2})\]
  \[T = \text{median}(C_R) - \beta \cdot \sigma_{C_R}\] where \(\beta = \frac{\delta}{\sigma}\)

Rule 2: sigma criteria on secondary peaks:
- if \(S_{R1} < T\) or \(S_{R2} < T\) then accept \(z_{abs}\)
- else, reject this \(z_{abs}\) as a detected absorption pixel

For our chosen values of \(\alpha\) (primary threshold), \(\beta\) (secondary threshold), and \(\delta = f_{\text{strong}}/f_{\text{weak}}\), where \(f_{\text{strong}}\) and \(f_{\text{weak}}\) are the oscillator strengths of Mg ii 2796 and 2803 lines, see Table 2. Note that the threshold on the secondary peaks is based on the theoretical strength of lines given the oscillator strengths. The threshold \(T\) is defined for the two smaller peaks centered at \(\lambda_{1}\) and \(\lambda_{2}\), which are weaker in the convolved residual.

Finally, note that we apply our thresholds to all individual pixels in the spectrum. The resulting list of accepted pixels will in general contain several contiguous ranges which need to be grouped and considered as one detected absorption feature. To do so we combine contiguous pixels absorbers with \(\Delta z < 0.0026\), corresponding to \(\Delta z_{\text{rest}}, \text{Mg II} = 7.17\,\text{Å}\). For each group we derive the absorber redshift as the mean of the pixel redshifts, weighting by the cube of the corresponding flux residual, such that the highest weight comes from the pixel closest to line center.

To provide the flexibility to detect absorbers of different strengths (rest equivalent widths) we separately run the kernel convolution varying the kernel FWHM from 3-8 pixels. We combine all the absorbers from each run, identifying and discarding duplicates by applying a similar weighted mean approach, weighting each redshift by the cube of the median of residual in the wavelength range \(\lambda_{1} < \frac{\lambda_{rest}}{\sigma_{\text{C_R}}} < \lambda_{b}\) (see Table 2). Note that this grouping is applied only after running the detection pipeline for all widths to select the best location of the absorber.

Rule 3: S/N criteria for final candidate selection:

The matched kernel convolution gives a list of potential Mg ii candidates. To select the final candidates we then use the S/N information of QSO spectra. Therefore, for each potential candidate from this list, we estimate the S/N (12796) and S/N (12803). \(^3\) For the final candidate selection we apply the following criteria. We call this the Mg ii doublet criteria.

\[\frac{S}{N}(12796) > 3 \quad \text{and} \quad \frac{S}{N}(12803) > 2\]

2.3.3 Rejecting False Positives

Given the aspirer candidates that passed the above criteria, we next reject likely false positives. We fit each absorption profile with a double Gaussian function and discard candidates which have peculiar doublet separation. In fitting we allow all six parameters (two amplitudes, two-line centres, and two-line widths) to vary. We keep only those absorbers for which the line separation is within 1.2 Å of the fiducial value (Δλ_rest, Table 2). At this stage we also estimate the mean redshift of the absorbers using the centroid of both lines. Finally, we check for cases where Fe ii lines are incorrectly identified as Mg ii lines. For spectra with more than one potential Mg ii absorber, we check whether its observed wavelength corresponds to any Fe ii λλ2586, 2600 lines, within ±3 Å, corresponding to the redshifts of already identified Mg ii lines. If this is the case, we reject the Mg ii candidate having smaller S/N(12796) than any of the two Fe ii lines as Mg ii is a stronger transition than Fe ii. This selection might also reject a few true cases in which the Fe ii lines are also very strong, however, this is not very common. We show an example of a SDSS QSO spectrum with relatively high S/N ratio (≈ 7.5 pix⁻¹) in Figure 1, highlighting two detected Mg ii absorbers and the locations of the corresponding Fe ii lines.

2.3.4 Rest equivalent widths of Absorbers

With the detected absorbers in hand, we apply a Monte Carlo approach to estimate errors on the fit parameters and ultimately the measured rest equivalent widths. We fit each doublet with a standard Levenberg-Marquardt minimization, first by adding randomly generated noise, using a normal distribution centred at 0 and with standard deviation equal to the mean of the errors on the residual. We also try the alternative approach of adding the true noise of the spectra multiplied by a standard normal distribution. In both cases we repeat this process 200 times. In case of failures, we record all

\(^3\) S/N = \(\sum_{i=p_1}^{p_2} \left(1 - \frac{F_i}{C_i}\right) / \left(\sum_{i=p_1}^{p_2} \sigma^2_i\right)^{1/2}\); where F and C are flux and continuum respectively and \(\sigma^2\) is the corresponding error. p_1 and p_2 are the starting and ending pixels within ±5 pixels from line centre.
parameters and rest equivalent widths as zeros. This guarantees that errors will be large in cases with many failures, if there is a false positive all 200 runs fail and the rest equivalent width is consistent with zero. Finally, we take the median of all 200 runs as the estimate of the rest equivalent width of lines. For errors we take the 16th (p16) and 84th (p84) percentiles and compute sigma as

\[ \sigma = \frac{p84 - p16}{2} \]

Overall we find that these two error estimate methods agree well (shown in Figure B1), also indicating that the majority of absorbers are genuine. For our purposes we use the un-

---

**Figure 1.** Top: the black line shows the normalized observed flux of quasar SDSS J095013.01-002839.0 (S/N = 7.46). The red line is the best-fitting NMF continuum after median filtering. The hatched orange rectangle shows the wavelength search window where the spectrum is searched for intervening Mg II absorbers. The intrinsic C IV, C III and Mg II emission lines of the quasar are shown in blue, green and red respectively. Bottom: the black line shows the residual spectrum defined as the ratio of the observed spectrum to the NMF continuum. The dashed blue line indicates unity. In this case the absorber detection pipeline identifies two intervening Mg II absorbers at \( z_{\text{abs}} = 1.6740, 1.7807 \) (shown in red). The corresponding Fe II lines are shown in blue. Inset shows the convolved array with the thresholds (cyan: primary, blue: secondary) at the location of one of the absorbers. We clearly see three peaks as described in the text.

**Figure 2.** A short schematic of our automated detection pipeline.
certainties obtained from the true noise of the spectra. The typical error in the rest EW\textsubscript{2796} (σ\textsubscript{EW\textsubscript{2796}}) is ~ 0.2 Å.

2.3.5 Fe - confirmed Mg \textsc{ii} catalogue

With our Mg\textsc{ii} absorber catalogue we perform an additional step and attempt to confirm each Mg\textsc{ii} absorber with Fe\textsc{ii} λλ 2586, 2600 by fitting gaussian at the location of Fe\textsc{ii} lines. If the fitting succeeds (parameters are finite and within the boundaries) and the line separation of Fe\textsc{ii} lines is within 1.2 Å of the fiducial value (Δ\textsubscript{rest} = 13.48 Å) we flag this Mg\textsc{ii} absorber as ‘confirmed’ with Fe\textsc{ii}. We also estimate the rest frame equivalent width and errors of the Fe\textsc{ii} λλ2586, 2600 lines using a similar Monte Carlo approach as above. Note that Fe\textsc{ii} is a much weaker transition than Mg\textsc{ii} and the minimum Fe\textsc{ii} strength detectable with our algorithm is EW\textsubscript{2600} ≥ f\textsubscript{FeH}, 2600/5f\textsubscript{MgII}, 2796 - EW\textsubscript{2796}, assuming the same abundance for both Mg and Fe. With the minimum detected EW\textsubscript{2796} ~ 0.2 Å in our catalogue, and f\textsubscript{FeH}, 2600 ~ 0.3, we obtain a theoretical minimum EW\textsubscript{2600} ~ 0.1 Å. In total we have ~ 70,000 Fe\textsc{ii} confirmed Mg\textsc{ii} absorbers (~ 44% of the sample). Finally, we show a schematic in Figure 2 that summarizes the steps of our automated detection pipeline.

3 METAL ABSORBER CATALOGUE

3.1 Properties of Individual Absorbers

Running the detection pipeline on the DR16 quasars we compile our final Mg\textsc{ii} absorber catalogue. In Figure 3 we show the redshift distributions of quasars (light gray shows all DR16 QSOs and dark gray shows QSOs with S/N\textsubscript{QSO} > 2) and detected absorbers (blue). We construct three absorber catalogues: (i) all Mg\textsc{ii} detections (light blue), (ii) strong Mg\textsc{ii} absorbers defined as having EW\textsubscript{2796} > 3σ\textsubscript{EW\textsubscript{2796}} (medium blue), and (iii) ‘Fe\textsc{ii} confirmed’ detections where we detect Fe\textsc{ii} λλ2586, 2600 at the same redshift (dark blue). Within our wavelength search window we find 159,524 Mg\textsc{ii} absorbers which satisfy our detection pipeline. Out of these 69,675 absorbers have also passed our Fe\textsc{ii} confirmation test and 121,989 have EW\textsubscript{2796} > 3σ\textsubscript{EW\textsubscript{2796}} (i.e. S/\textsubscript{N}\textsubscript{2796} > 3).

In the four panels of Figure 4 we present the properties of individual Mg\textsc{ii} absorbers. The topmost panel shows the ratio of total rest equivalent width (EW) of Mg\textsc{ii} to Fe\textsc{ii} using λλ2586, 2600 only as a function of the total rest equivalent width of Mg\textsc{ii} absorption. This quantifies the relative strength of Fe\textsc{ii} absorbers, which also trace K gas around galaxies. We see that most (≥ 80 percent) of the Fe\textsc{ii} confirmed Mg\textsc{ii} absorbers lie within the theoretical limit (assuming abundances to be same for Fe and Mg) for the line strength ratio of 1 (for the saturated case) and ~3 (for the completely unsaturated case). The majority of absorbers regardless of EW have a line ratio of ≥ 1, indicating they are intermediate between saturated and unsaturated. The median line ratio is ~ 2. We also note the large scatter and many outliers above ~ 3, likely indicating cases of non-solar abundance ratios.

The second (from top) panel of Figure 4 shows the doublet ratio of EW\textsubscript{2796} and EW\textsubscript{2803} as a function of EW\textsubscript{2796} for all Mg\textsc{ii} absorbers. The doublet ratio indicates if the lines are saturated or unsaturated and is sensitive to the opacity of the medium. For Mg\textsc{ii} λλ2796, 2803 the theoretical value of the doublet ratio varies between 1 (fully saturated) and 2 (completely unsaturated). We see that most strong absorbers (EW\textsubscript{2796} > 1 Å) have a doublet ratio close to 1, indicating saturation. For weak absorbers (EW\textsubscript{2796} < 1 Å) we see doublet ratio ~ 2 as expected due to low saturation.

The third (from top) panel of Figure 4 shows the fractional errors on the EW\textsubscript{2796} measurement as a function of EW\textsubscript{2796} for all detected Mg\textsc{ii} absorbers. We see that, most (~ 94 percent) of the absorbers have low errors (fractional error < 0.5). Weak absorbers tend to have large errors, making it more problematic to measure their properties. The median EW\textsubscript{2796} is ~ 1.3 Å with typical σ\textsubscript{EW\textsubscript{2796}} ~ 0.2 Å for the sample. The bottom panel of Figure 4 shows the cumulative (red) and differential (blue) distributions of

Table 3. Sample statistics comparison versus SDSS data release (DR). References are: [a] Pâris et al. (2017); [b] Pâris et al. (2018); [c] Ahumada et al. (2020); [d] Zhu & Ménard (2013a) (trimmed case with the same masks as ours); [e] this work. The increased quasar sample enables us to increase the statistics of Mg\textsc{ii} absorption systems significantly (~ 4 times). Note that S/\textsubscript{N}\textsubscript{2796} is defined as EW\textsubscript{2796}/σ\textsubscript{EW\textsubscript{2796}} and S/\textsubscript{N}\textsubscript{2796} implies signal-to-noise of quasar spectra.

| Objects | DR12 | DR14 | DR16 |
|---------|------|------|------|
| QSOs    | 309,411\textsuperscript{a} | 526,356\textsuperscript{b} | 983,317\textsuperscript{c} |
| QSOs (S/N\textsubscript{QSO} > 1) | - | - | 941,939 |
| QSOs (S/N\textsubscript{QSO} > 2) | - | - | 773,594 |
| LRGs    | - | ≤ 1 million | 1,252,722\textsuperscript{d} |
| ELGs    | - | 35,094\textsuperscript{f} | 269,888\textsuperscript{g} |
| Mg\textsc{ii} Absorbers | 39,219\textsuperscript{e} | - | 159,524\textsuperscript{i} |
| Mg\textsc{ii} (S/N\textsubscript{2796} > 1) | 38,327\textsuperscript{e} | - | 158,725\textsuperscript{i} |
| Mg\textsc{ii} (S/N\textsubscript{2796} > 2) | 37,763\textsuperscript{e} | - | 150,236\textsuperscript{i} |
| Mg\textsc{ii} (S/N\textsubscript{2796} > 3) | 33,376\textsuperscript{e} | - | 94,403\textsuperscript{i} |
| Fe\textsc{ii} Absorbers | - | - | 69,675\textsuperscript{e} |
| Mg\textsc{ii} Absorbers | - | - | 158,494\textsuperscript{e} |
| (with S/N\textsubscript{QSO} > 2) | - | - | 69,594\textsuperscript{e} |
3.2 Properties of Absorbers in Stacked Spectra

We stack the residual spectra of quasars in the rest frame of detected Mg ii absorbers to study features which become visible in these composite spectra. The top panel of Figure 5 shows a composite median spectrum, stacking on all Mg ii absorbers. We see several weak metal lines such as Si ii, C iv, Al ii, Al iii, Fe ii (several lines) and Mg i. The clear detection of these weak lines provides strong evidence that the majority of our absorbers are genuine. We have divided our sample into five $E_W_{2796}$ bins to understand the corresponding variation of the strength of metal absorbers, which increases with increasing $E_W_{2796}$.

In spectral regions without absorption features the median residual is flat, indicating that our continuum estimation pipeline works well. In the bottom panel Figure 5 we zoom into the profiles of three prominent metal lines, C iv $\lambda\lambda 1548,1550$ (left), Fe ii $\lambda\lambda 2586,2600$ (middle), Mg ii $\lambda\lambda 2796,2803$ (right). The absorption features are well represented by Gaussian profiles, as expected in a stack due to the central limit theorem. As we move to strong absorber systems the lines become saturated (doubtlet ratio $\sim 1$) and doublet profiles overlap, particularly for C iv. For the strongest $E_W_{2796}$ bin noise begins to be visible in the residual due to the low statistics. As a check we also stack the spectra of all quasars with Mg ii absorbers without Fe ii confirmation (not shown). In this case we also detect the same ensemble of weaker metal transitions.

In Figure 6 we show the rest equivalent widths of visible metal lines in the composite spectra as a function of the total rest equivalent width of Mg ii. We see that Fe ii and C iv lines are the second most prominent features though the absorption strength of Fe ii is slightly higher compared to C iv lines. This is because Fe ii is a stronger transition and also at longer wavelength, Fe ii can be detected in SDSS quasar spectra with redshifts as low as $z \sim 0.5$, while C iv absorbers can only be detected in quasars with $z_{\text{QSO}} \geq 1.5$ due to the wavelength coverage of SDSS. We note that there should be a slight asymmetric redshift path-lengths in order to properly quantify the relative scaling between the lines.

In order to investigate whether our pipeline estimates the consistent rest equivalent widths, we also make a comparison of metal equivalent widths against the empirical scaling relations (as a function of $E_W_{\text{MgII}}$, 2796 and redshift) derived in Lan & Fukugita (2017). We observe that our pipeline yields a consistent trend, though there is a slight discrepancy in some cases.

3.3 Completeness of Detection Algorithm

The search for absorption features in the relatively low S/N quasar spectra from SDSS is challenging, and an accurate characterization of our ability to recover absorbers of different strengths is required. To estimate the completeness of our detection pipeline we therefore implement a Monte Carlo simulation approach. We generate doublet profiles that mimic true absorber and insert these at a random location in a real residual chosen randomly from the set of DR16 QSOs. While inserting we take care of all the masks that we have defined in our detection algorithm. To generate doublets we uniformly...
sample EW$_{2796}$ from 0 ≤ EW$_{2796}$ ≤ 8 Å. We select both a doublet ratio (0.25 < doublet ratio < 4.5) and width (0.34 < σ < 3.5 Å) from gaussian distributions. We then run our detection pipeline on the spectrum with a fake absorber and check for a successful detection. We define success as having satisfied the Mg II doublet criteria and recovered a sufficiently accurate rest equivalent width, EW$_{\text{abs}}$, as the ratio of detected absorbers to legitimate absorbers. As expected, the completeness is higher for strong absorbers and lower for weak absorbers. The bottom right panel shows the completeness (and purity) as a function of z averaged over all EW$_{2796}$, where dips at specific redshifts reflecting excluded wavelength search windows. Overall, the purity of our algorithm (shown in Figure C1) is extremely high, at the cost of moderate completeness, however, completeness goes up significantly for QSOs with high S/N data as shown in the bottom left panel of Figure 7. Because the eBOSS spectra have relatively lower S/N (Dawson et al. 2016), we also include here completeness curves restricted to spectra above minimum S/N thresholds (i.e. S/N > 1.2). As expected we find that it is difficult to detect ab-

Figure 5. Top: Median composite spectra of quasars, stacked in the rest-frame of all DR16 Mg II absorbers. We have divided the absorber sample into five sub-samples with 0 Å < EW$_{2796}$ < 0.5 Å (N$_{abs}$ = 5718), 0.5 Å < EW$_{2796}$ < 1.5 Å (N$_{abs}$ = 90, 317), 1.5 Å < EW$_{2796}$ < 2.5 Å (N$_{abs}$ = 50, 083), 2.5 Å < EW$_{2796}$ < 4 Å (N$_{abs}$ = 12, 869), 4 Å < EW$_{2796}$ < 8 Å (N$_{abs}$ = 537). We have indicated the most prominent metal lines to guide the eye: in stacks we detect weaker transitions including Si II, C IV, Al II, Al III, Mg I and other weak transitions of Fe II. Bottom: Zoom into the median composite spectra to show the structure of three of the most prominent metal lines. For display purposes we have shifted the lines of different EWs by a small amount in the vertical direction in all panels.

Finally, for absorbers in a given EW$_{2796}$ and redshift bin, we estimate the completeness, c(EW$_{2796}$, z) as the ratio of detected absorbers to injected absorbers. Using this estimate we compute, for each absorber with given EW$_{2796}$ and redshift, a completeness corrected effective ‘number of absorbers’ N$_{\text{eff}}$ = 1/c(EW$_{2796}$, z), the count of absorbers our detection method been perfect. By definition, c < 1 and so N$_{\text{eff}}$ ≥ 1. Similarly we estimate the purity as the ratio of detected absorbers to legitimate absorbers (i.e. all the absorbers that pass the Mg II doublet criteria) found by our pipeline. We show the 2D distribution of completeness in EW$_{2796}$ – z space in the top panel of Figure 7.
Figure 6. Rest equivalent width of different metal transitions as a function of rest equivalent width of Mg II 2796 line for the median composite spectra. The solid circles show the measured rest EWs (rEWs) and dashed solid lines show the empirical rEWs scaling as a function of EW_{MgII,2796} from Lan & Fukugita (2017). To compare we use the median redshift of absorbers in each rEW (2796) bin. The shaded region shows the corresponding 5th and 95th percentiles of redshifts in each rEW (2796) bin. We have estimated the rest equivalent widths by fitting suitable gaussian profiles and errors using the bootstrap method.

absorbers primarily in noisy spectra, while our algorithm has excellent performance on high S/N data.

The observed and completeness corrected EW_{2796} distribution for all Mg II absorbers (that are associated with QSOs having S/N_{QSO} > 2, i.e. our “fiducial” sample, defined explicitly in section 4.1, as well as for the Fe II confirmed subset, are shown in Figure 7. The completeness corrected distribution follows an exponential profile, hence the mean value after completeness correction shifts to smaller values as weak absorbers dominate the distribution. The mean EW_{2796} for the observed distribution is ~1.4 Å while for completeness corrected distribution it decreases to ~0.8 Å. We also see that the ‘completeness corrected distribution’ is very high in the lowest bin, which may be due to the small number statistics of absorbers and, such a peak is possibly not significant. We observe that the fraction of Fe II confirmed Mg II absorbers for high rest equivalent widths does not reach unity even for very strong absorbers. This is because Fe II is a lower wavelength transition than Mg II and does not fall within the SDSS wavelength coverage for Mg II absorbers at low redshifts.

4 CONNECTING TO GALAXIES

With absorbers identified across a wide range of redshifts and over a significant fraction of the sky, we now need to connect them to nearby galaxies. Our main resource to obtain a large sample of galaxies with 0.4 < z < 1 are the BOSS and eBOSS programs of SDSS. Namely, emission line galaxies (ELGs) and luminous red galaxies (LRGs), which we use to study the galaxy - Mg II absorber connection. When necessary, we use a Planck-consistent cosmology for our analysis, i.e. \( \Omega_{m,0} = 0.307 \), \( H_0 = 67.7 \text{ km s}^{-1} \text{ Mpc}^{-1} \).

4.1 Fiducial Quasar Sample

As discussed in Section 3.3, our detection pipeline has excellent performance on high S/N data. Therefore, for the galaxy-centric analysis we select only quasars having S/N_{QSO} > 2 and Mg II absorbers associated with them. We make this cut to define the fiducial quasar sample and the corresponding Mg II absorber catalogue because the completeness and purity both are very high for our detection algorithm and we miss intermediate or strong absorbers rarely. For example, at EW_{2796} ~ 2 Å completeness and purity are \( \geq 75 \) percent and \( \geq 90 \) percent respectively as shown in bottom left panel of Figure 7. The high completeness catalogue also brings more confidence in the results as we are not applying big corrections. As presented in Table 3, out of 978,561 quasars 773,594 have S/N_{QSO} > 2, which is \( \sim 80\% \) of the total sample size. We show the histogram of redshifts of quasars with S/N_{QSO} > 2 in Figure 3. We lose quasars uniformly in redshift space. The median S/N_{QSO} of QSOs is \( \sim 7.5 \). In our Mg II catalogue we have \( \sim 1,030 \) absorbers associated with quasars having S/N_{QSO} < 2 and we remove them for consistency. From now onwards we perform all galaxy-centric analysis with this fiducial quasar sample and Mg II absorber catalogue.

4.2 Galaxy Samples

4.2.1 Emission Line Galaxies

The main selection criteria for ELGs in SDSS is a cut in the g − r vs r − z colour-colour diagram and g band magnitude (Raichoor et al. 2017). Their typical stellar mass is M* \( \sim 10^{10.5} M_\odot \) (Raichoor et al. 2017) and they reside in dark matter haloes with M_{halo} \( \sim 10^{12.2} M_\odot \) (Favole et al. 2016).

The characteristic features of ELG spectra are strong gas emission lines such as [O II] \( \lambda \lambda 3727, 3729 \) and H β, due to high star formation activity. Based on [O II] \( \lambda \lambda 3727 \) luminosity and models described in Kennicutt (1998) the star formation rate in ELGs is estimated to be from 1 to 20 M_\odot yr^{-1}. The presence of hot and young stars make ELGs appear blue.

We select the latest DR16 ELG catalogue compiled by Rai-
choor et al. (2021)\textsuperscript{4}, which contains 269,178 objects. The catalogue includes stellar masses estimated by the FAST spectral fitting code (Kriek et al. 2009), with the Bruzual & Charlot (2003) stellar model and Chabrier (2003) initial mass function (IMF). We include only ELGs with reliable redshifts (selected by IMATCH=1 condition). We also apply cuts on stellar mass and redshift to connect with Mg\,II absorbers. The final sample contains 188,323 ELGs at $z > 0.4$ and with $9 < \log M_\star [\text{M}_\odot] < 12$. The typical measurement error in redshift is $\sim 20$ km s\textsuperscript{-1}. The median stellar mass, halo mass and redshift of the ELGs in the final sample are $M_\star \sim 10^{10.4} \text{M}_\odot$, $M_{\text{halo}} \sim 10^{12.1} \text{M}_\odot$ and $z \sim 0.84$, respectively. The ELG catalogue does not include a stellar mass uncertainty for each galaxy, however the typical uncertainty in stellar mass for ELGs is 0.05 dex and the majority are within 0.25 dex (Raichoor et al. 2017, see section 6.3)\textsuperscript{5}. The stellar mass-redshift distribution is shown in the left panel of Figure 9. For this sample we estimate the star-formation rate (SFR) using the scaling relation from Kewley et al. (2004).

\begin{multline}
\text{SFR} \ [M_\odot \text{yr}^{-1}] = \frac{L_{[\text{O} \text{III}]}}{1.52 \times 10^{-41}} \text{ergs s}^{-1} \end{multline}

where $L_{[\text{O} \text{III}]}$ is luminosity measured as $L_{[\text{O} \text{III}]} = F \cdot 4\pi D_L^2$; $F$ is the measured $[\text{O} \text{III}]$ \text{3727} flux and $D_L$ is the luminosity distance. We note that the measured $[\text{O} \text{III}]$ \text{3727} flux ($F$) is not corrected for reddening and the SFR may be underestimated by a factor of a few on average and up to factor of 10 or more for the most massive, star-forming systems. Possible contamination by AGN is also discussed.

\textsuperscript{4} VAC available at: https://data.sdss.org/sas/dr16/eboss/lsst/catalogues/DR16/eBOSS_ELG_full_ALLdata-vDR16.fits

\textsuperscript{5} Hong Guo, private communication.
The cool CGM in SDSS galaxies

4.3 Methods

4.3.1 Covering Fraction of Mg II absorbers

We define \( N_{\text{obs},j}^{\Delta \zeta} \Delta \zeta \) as the number of detected absorbers around the \( j^{\text{th}} \) SDSS galaxy within an annulus \( \Delta \zeta \) satisfying a maximum \( \Delta \zeta \). An example of this is shown in Figure 8.1. The 1D histogram of intrinsic observed (gray) and completeness corrected (red) of \( EW_{\text{Mg II}, 2796} \) for all Mg II absorbers (associated with QSOs having \( S/N_{\text{QSO}} > 2 \)) as well as the Fe II confirmed subset (blue). The completeness corrected distribution follows an exponential profile in rest equivalent width, such that the weakest absorbers are also the most frequent. Our Mg-II catalogue spans 0.2 Å \( EW_{\text{Mg II}, 2796} \) \( \Delta \zeta \) ≤ 8 Å.

4.2.2 Luminous Red Galaxies

The luminous red galaxies in SDSS are observed as described in Padmanabhan et al. (2012). The selection of LRGs is based on cuts on SDSS \( g, r, \) and \( i \) band magnitudes. Due to low star formation activity, a large fraction of these galaxies have no emission lines in their spectra. The presence of old stellar populations makes them look optically red. They are typically more massive than ELGs and have \( M_{*} \sim 10^{11.2} M_{\odot} \) and reside in dark matter haloes with \( M_{\text{halo}} \sim 10^{13.5} M_{\odot} \) on average (White et al. 2011).

For LRGs we take the Wisconsin PCA-based catalogue which contains a total of \( \sim 1,489,670 \) objects. The stellar masses were estimated using a principal component analysis (PCA) method described in Chen et al. (2012), assuming Kroupa (2001) IMF. The authors also added 0.057 dex to adjust to a Chabrier (2003) IMF (see Herrmann et al. 2016). The stellar libraries and single stellar population models were taken from Maraston & Strömbäck (2011). The typical measurement error in redshift is \( \sim 30 \text{ km s}^{-1} \). For our galaxy-absorber correlation analysis we select LRGs having \( z > 0.4 \) with \( 10 < \log M_{*} [M_{\odot}] < 12 \). Our final LRG catalogue contains \( \sim 1,081,329 \) galaxies and the median stellar mass, median halo mass, and redshift \( M_{*} \sim 10^{11.5} M_{\odot}, M_{\text{halo}} \sim 10^{14} M_{\odot} \) and \( z \sim 0.54 \), respectively. The typical uncertainty in stellar mass is 0.16 dex. The stellar mass-redshift distribution is shown in the right panel of Figure 9.

4.3.2 Random Galaxy Samples

To better understand the bias and clustering of Mg II absorbers around galaxies, we compare to the expected average absorption signals for random sky sightlines. To do so, we define 40 random galaxy samples, each equal in size to our fiducial galaxy sample. Starting from true galaxies, we shuffle sky positions as well as redshifts. This shuffling de-correlates the positions and redshifts of the galaxies while preserving the original redshift distribution and sky coverage. We compute all observational measurements for both the true and random galaxy samples.

4.3.3 Bootstrap Error Estimates

To estimate errors we use the bootstrap approach, repeating the absorber/QSO-galaxy pair identification procedure 100 times. In each iteration we select galaxies from the original sample at random, with repetition, to account for Poisson statistics. We take the mean over these 100 iterations to estimate each quantity and take the error as the standard deviation of these samples.

4.3.4 Halo Mass and Virial Radius

We perform analysis requiring two unobservable quantities: halo mass and virial radius. To estimate halo mass for a given stellar mass and redshift we make use of the Stellar Mass-Halo Mass (SMHM) model developed in Behroozi et al. (2010). Given halo mass, the virial radius \( r_{200} \) of the galactic halo is found using the \( \Delta_{c} \) \( \sim 200 \) definition of Bryan & Norman (1998)

\[
r_{200} = r_{\text{vir}} = 211.83 \frac{M_{12}^{1/3}}{M_{12}} \left[ \Omega_{m,0}(1 + z)^{3} + \Omega_{\Lambda,0} \right]^{-1/3} \text{[kpc]}
\]

where \( M_{12} = M_{\text{halo}}/10^{12} [M_{\odot}] \). Note that adopting other SMHM relations would lead to different halo mass and radii estimates.
4.4 Galaxy-Absorber Correlation

Figure 10 shows the positions and redshifts of 1000 objects chosen randomly from the galaxy samples and our DR16 [Mg II] absorber catalogue. The black dots indicate the positions of detected Mg II absorbers and blue and red dots denote ELGs and LRGs, respectively. ELGs clearly extend out to $z \sim 1$ compared to LRGs that extend to $z \sim 0.7$. The map also shows that the Mg II absorber distribution is quite uniform in RA-$z$ space and that many absorbers fall within the volume of the galaxy samples.

We begin with the excess mean number of Mg II absorbers per kpc$^2$ (hereafter, excess mean surface density), i.e. the surface density around true galaxies divided by the surface density around the random galaxy sample, as a function of projected distance ($D_{\text{proj}}$). Figure 11 shows the result for Mg II absorbers with $E_W > 0.4$ Å.

For other EW bins the trends are similar, with mean surface density decreasing for stronger absorbers, due to their relative scarcity. For ELGs the excess surface density rises to a maximum at $\sim 50 - 80$ kpc and then declines with projected distance, however, for LRGs, the trend is consistent with decreasing as a function of distance. The decreasing trend of Mg II surface density with projected distance around LRGs was also identified with a different stacking based methodology (Zhu et al. 2014), as well as with individual galaxy-absorber pairs in the COS-LRG survey (Zahedy et al. 2019). We also see an enhancement of Mg II absorbers around ELGs relative to LRGs below $D_{\text{proj}} \sim 100$ kpc by a factor of $2 - 4$. For these two samples the mean surface density is consistent within the error bars beyond $D_{\text{proj}} \sim 100$ kpc.

The excess mean surface density of absorbers converges to the random expectation (unity) only for $D_{\text{proj}} \gtrsim 15$ Mpc for both LRGs, and ELGs, as shown in the inset. To understand how excess surface density depends on redshift of galaxies we further divided ELGs and LRGs into redshift bins and estimate the excess surface density around each. As shown in the bottom panel of Figure 11 we find that within a given sample the excess surface density is slightly higher for low-$z$ galaxies below $D_{\text{proj}} \sim 100$ kpc. They all converge to random values at $D_{\text{proj}} \gtrsim 10$ Mpc.

Next, we estimate the covering fraction, $f_c$ ($d_1 < D_{\text{proj}} < d_2$), of Mg II absorbers around galaxies. We show the variation of $f_c$ as a function of projected distance in Figure 12. The covering fraction decreases with projected distance and varies strongly with galaxy type. In each $E_W > 0.4$ Å bin, ELGs have 2-5 times higher covering fraction than LRGs below $D_{\text{proj}} < 50$ kpc. For example, in the $E_W > 0.4$ Å bin, ELGs have a Mg II covering fraction of $f_c \gtrsim 50 - 70$ percent compared to $f_c \lesssim 15$ percent for LRGs. At large distances ($D_{\text{proj}} > 100$ kpc) there is no significant variation with galaxy type. Several previous studies have reported similar trends using different samples and analysis (Lan et al. 2014; Nielsen et al. 2013; Lovegrove & Simcoe 2011). The covering fraction converges to the random expectation for $D_{\text{proj}} \gtrsim 10$ Mpc.

In addition to the trend with galaxy type, the covering fraction also depends strongly with the strength of absorbers (note the
vertical limits vary with each panel). There is a clear anti correlation between covering fraction and absorber strength. This can be attributed to the rarity of strong absorbers around galaxies, such that absorption with $EW_{2796} > 2$ Å is roughly an order of magnitude less frequent in the inner halo. Particularly for LRGs, however, there are low number statistics and the measurement errors are high. Nonetheless, the strong enhancement of $f_c$ seen within 50 kpc is clearly present for all rest equivalent widths. As we discuss below, this small component implies an association to the ongoing star formation activity of ELGs and the resultant galactic-scale outflows.

We also investigate the cumulative covering fraction $f_c(d < D_{proj})$, and the result as a function of projected distance is shown in Figure 13. The trend reflects what we have previously seen with the differential covering fraction: star-forming galaxies (ELGs) have 2-4 times higher cumulative values than passive galaxies (LRGs) up to $D_{proj} < 100$ kpc. The measurement errors are also smaller as the

Figure 11. Excess mean surface density of absorbers ($d_1 < D_{proj} < d_2$): the mean number of absorbers per galaxy per kpc$^2$ for Mg ii absorbers with $EW_{2796} > 0.4$ Å divided by the corresponding value around the random galaxy sample. Top: The blue and red squares denote ELGs and LRGs, respectively. The corresponding dashed black line denotes unity, i.e. the expectation for random sightlines. The inset shows that the values converge to their random expectation only at $D_{proj} \geq 10$ Mpc. Bottom: Excess mean surface density of absorbers around ELGs and LRGs in two redshift bins. The surface density of absorbers increases towards higher redshift at larger distances.

Figure 12. Differential covering fraction of Mg ii absorbers in three $EW_{2796}$ bins (shown at the top of each panel). The blue and red squares denote ELGs and LRGs, respectively. The corresponding dashed color lines show the covering fractions expected around the random samples. The ELG sample has a pronounced excess relative to the LRGs within $D_{proj} \lesssim 50$ kpc.
cumulative estimation significantly enhances the statistics. We now clearly see a difference in $f_c$ between the two samples which extends up to ~ 200 kpc. This partially reflects the different physical sizes of the gaseous (and dark) haloes hosting LRGs versus ELGs. For LRGs, 100 kpc corresponds to $\lesssim 0.2r_{\text{vir}}$ (near the central galaxy), and gas accreted from halo or IGM may not reach into these central regions (Huang et al. 2016). However, for ELGs, this corresponds to $\sim 0.3r_{\text{vir}}$ where metal rich gas ejected by powerful galactic outflows or winds can be deposited (Muratov et al. 2017; Nelson et al. 2019; Mitchell et al. 2020).

### 4.5 Dependence on Stellar and Halo Mass

To understand how the covering fraction evolves with the size of the dark matter halo we normalize the projected distances by the virial radii of galaxies. In Figure 14 we show the cumulative Mg II covering fraction in stellar mass bins (different for ELGs and LRGs). We see a weak anti-correlation between the stellar mass of ELGs and the covering fraction of Mg II absorbers when normalizing by $r_{\text{vir}}$. In contrast, LRGs show a strong stellar mass dependence in $f_c$. Higher mass passive galaxies have systematically lower cumulative Mg II covering fractions at all distances. For example, LRGs with $M_* < 10^{11} M_{\odot}$ have cumulative Mg II covering fractions up to ~ 25 percent compared to ~ 3 percent around more massive LRGs ($M_* > 10^{11.5} M_{\odot}$) below $D_{\text{proj}} < 0.3r_{\text{vir}}$.

The reason for this strong stellar mass dependence may be two-fold. First, the tight correlation between stellar and halo mass implies that the size, and total mass, of the circumgalactic medium increases rapidly for more massive galaxies. Second, gas is also thermalized to higher temperatures in more massive haloes, which would naturally inhibit the formation of cooler gas phases. In this regime, heating by the virial shock produces long-term ‘hot-mode’ growth, which is the dominant accretion mode for LRG-type galaxies (Binboim & Dekel 2003; Kerer et al. 2005; Nelson et al. 2013).

We show the stellar mass dependence of covering fractions versus physical kpc in the insets of Figure 14. Here we see only a weak stellar mass dependence of $f_c$ on physical separation in kpc, and predominantly at large distances, where massive galaxies have higher covering fractions. The lack of a strong trend at fixed physical distance suggests that the signature above is largely driven by the $r_{\text{vir}}$ normalization, i.e. the increasing size of more massive haloes.

To better explore these mass trends, we derive the cumulative Mg II covering fraction at $D_{\text{proj}} \leq r_{\text{vir}}$, a rough outer boundary of the CGM, as a function of stellar mass. The result is shown in the top panel of Figure 15, where red and blue markers represent LRGs and ELGs, respectively. For ELGs we see a decreasing trend of Mg II covering fraction from ~ 11 percent for $M_* \sim 10^{10} M_{\odot}$ to ~ 9 percent for $M_* \sim 10^{11} M_{\odot}$. For LRGs we see a similar decreasing trend of covering fraction with stellar mass from ~ 3 percent for low mass galaxies to just ~ 1 percent for massive galaxies. Qualitatively this shows a clear picture: more massive galaxies host less cool gas, on average, in their circumgalactic media.

In the bottom panel of Figure 15 we show the cumulative total Mg II absorption EW per absorber within $D_{\text{proj}} \leq r_{\text{vir}}$ (see section 4.3.1) as a function of stellar mass. We see an increasing trend in absorption strength with stellar mass for ELGs, albeit with
**Figure 15.** Top: $\text{Mg} \, \text{II}$ covering fraction at $D_{\text{proj}} \leq r_{\text{vir}}$ around ELGs (blue) and LRGs (red) as a function of stellar mass. Both samples show the same signal: more massive galaxies have lower covering fractions of cool gas in their CGM. Bottom: Total $\text{Mg} \, \text{II}$ rest equivalent width per absorber within $D_{\text{proj}} \leq r_{\text{vir}}$ as a function of stellar mass. The absorption strength increases with the stellar mass of galaxies for both galaxy types. In both panels we show absorbers with $\text{EW}_{2796} > 0.4$ Å only as the other two bins show qualitatively similar results.

The cumulative absorption EW varies from $\sim 0.5$ Å below $M_*$ $\sim 10^{10}$ $M_\odot$ to $\sim 0.6$ Å around $M_*$ $\sim 10^{11}$ $M_\odot$. For LRGs, the trend is much stronger and values vary from 0.3 Å to 0.8 Å from the least to most massive galaxies. Broadly, this result implies that either $\text{Mg} \, \text{II}$ absorbers have larger EWs around more massive galaxies, or that sightlines intersect a larger number of individual absorbers in more massive haloes. A similar positive correlation between equivalent width of $\text{Mg} \, \text{II}$ absorbers and stellar mass was observed in star-forming galaxies (Bradshaw et al. 2013).

In the top panel of Figure 16 we show this same median total rest equivalent width per absorber, i.e. $\text{EW}_{2796} + \text{EW}_{2803}$, cumulative as a function of $d \leq D_{\text{proj}}$. At small distances (less than 100 kpc) the absorption strength declines with projected distance. For LRGs at $D_{\text{proj}} < 50$ kpc, we see an enhancement up to 3 Å, compared to $\leq 0.8$ Å, at larger radii. ELGs show a similar behavior, but have systematically lower rest equivalent widths, up to $D_{\text{proj}} \leq 400$ kpc, remaining slightly above the LRGs at larger distances. This is principally due to the significant difference in the average redshift between the two galaxy samples and the reason could be possibly related to the different gas properties at different epochs. To understand the possible redshift contribution to this difference we divide the ELGs and LRGs into low and high-$z$ bins (splitting at the medians). We clearly see a slightly higher EWs around high-$z$ galaxies at distances $D_{\text{proj}} \geq 1$ Mpc. We show the redshift evolution of EW in the bottom panel of Figure 16 for both ELGs and LRGs. The weak positive correlation between total EW and redshift is clearly visible for a given galaxy sample.

Nevertheless, we can still see in each redshift interval the same characteristic scale dependence of the total rest equivalent width per absorber: a strong decrease to a tentative minimum value at projected radii of $\sim 150$ kpc, followed by a slight rise to value that remains constant beyond 1 Mpc. This feature is more clear for ELGs than LRGs, and may be related to gas inflow processes onto dark matter haloes, which act to shock-heating in falling gas at the boundary of the halo.
4.6 Dependence on Star Formation Rate

A key property of galaxies that may affect the distribution of cold gas traced by Mg II absorbers is the star formation rate (SFR; Lan & Mo 2018; Rubin et al. 2018). The ELG catalogue includes measurements for [O iii] λ3727 flux which traces the star formation activity in ELGs for $z > 0.4$ (Kennicutt 1998). Despite its simplicity and neglect of dust attenuation effects, this SFR value can still be used to perform qualitative studies.

In the top panel of Figure 17 we show the dependence of cumulative and differential Mg II covering fraction on star formation rate for ELGs. We divide the galaxies into two SFR bins: low (SFR < 10 $M_\odot$ yr$^{-1}$) and high (SFR > 10 $M_\odot$ yr$^{-1}$). We observe a strong variation of covering fraction with star-formation activity. Galaxies with high star-formation rate have 2-3 times higher cumulative covering fraction than their low SFR counterparts, all the way up to $D_{\text{proj}} \approx 400$ kpc, close to the $r_{\text{vir}}$ of ELGs. One the other hand the differential covering fraction (shown in the inset) is 2-5 times higher below ~ 50 kpc in galaxies with high star-formation rate.

In the bottom panel we show the cumulative covering fraction at $D_{\text{proj}} \leq r_{\text{vir}}$ as a function of SFR. We again see a strong positive correlation with SFR, though the highest SFR bin has large uncertainties. This increasing trend of covering fraction with SFR indicates that SF activity in ELGs plays an important role in enriching the cold gas in their CGM. We also point out that for massive star-forming galaxies a possible source of influence is the central active galactic nucleus (AGN). We have not tried to exclude AGN from our samples. The galactic outflows powered by AGNs can significantly enrich the metal abundance in the CGM of massive star-forming galaxies (Veilleux et al. 2005), and future work will focus on disentangling the roles of star formation versus AGN activity.

4.7 The Relative Kinematics of Galaxies and Absorbers

With robust spectroscopic information for both the absorbing gas clouds and parent galaxies we can study the line-of-sight velocities to constrain the relative motion between the two. For this purpose we estimate the line-of-sight velocity separation, $\Delta v = c \Delta z/(1 + z)$, where $c$ is the speed of light and $z$ is the galaxy redshift.

In the left panel of Figure 18 we show the distribution of $\Delta v$ for Mg II absorber-galaxy pairs, separating ELGs (blue) from LRGs (red). Here we show absorbers with EW_{2796} > 0.4 Å within 10 kpc < $D_{\text{proj}}$ < 400 kpc from the galaxy. The velocity distribution of Mg II absorbers around ELGs (shown in blue solid line) is well characterized by a single gaussian with mean ($\langle \Delta v \rangle = -11$ (±8) km s$^{-1}$ and dispersion, $\sigma_v = 135$ km s$^{-1}$ (shown in the dashed blue line). The velocity distribution for LRGs (shown in the red solid line) in the central region is best-fit by a single gaussian of mean ($\langle \Delta v \rangle = -9$ (±10) km s$^{-1}$ and dispersion, $\sigma_v = 200$ km s$^{-1}$ (shown in the dashed red line). The mean velocity difference is in agreement with previous studies such as Huang et al. (2021), even though the samples and methods are significantly different. This further suggests that our analysis yields consistent trends. The typical dark matter halo velocity dispersion ($\langle \Delta v_{\text{200}} \rangle$) for ELGs is ~ 140 km s$^{-1}$ and ~ 350 km s$^{-1}$ for LRGs (Elahi et al. 2018). This implies $\sigma_{\text{ELG}} > \sigma_{\text{ELG, halo}}$ while $\sigma_{\text{LRG}} < 0.6 \sigma_{\text{LRG, halo}}$. While the cool CGM is consistent with virial motion around star-forming galaxies, it is sub-virial for massive quiescent systems.

Visual inspection shows that a single gaussian is not a good representation of the tails of these two $\Delta v$ distributions. A multi-component gaussian has been shown to better characterize the velocity distribution around LRGs (Huang et al. 2016; Chen 2017). We therefore fit a double gaussian (shown with the dashed black line) to the LRG distribution. The best-fit means are consistent with zero, while the two velocity dispersions represent a narrow component with $\sigma_{\text{narrow}} \sim 180$ km s$^{-1}$, describing the majority of pairs (~ 85 percent, central distribution), together with a broad component with $\sigma_{\text{broad}} \sim 1100$ km s$^{-1}$ describing the tails. We discuss possible interpretations in Section 5.

In the right panel of Figure 18 we show the relative velocity separation of galaxy-absorber pairs as a function of projected distance from the galaxy. The median values of the distribution in each radial bin are shown in colored circles for ELGs (blue) and LRGs (red). The vertical bars show the 5th to 95th percentiles of the distribution in each radial bin. The corresponding dashed color lines show the decreasing escape velocity for a dark matter halo mass, estimated using the average masses of the galaxies. The shaded gray region shows the increasing Hubble flow at the average redshift of the galaxy samples.

The median values are always within ~ 100 km s$^{-1}$ and well below the escape velocity of the dark matter halos. There is a small positive offset for ELGs, possibly due to gas moving away...
Table 4. Line-of-sight velocity separation between galaxy-absorber pairs (within 10 kpc < \(D_{\text{proj}}\) < 400 kpc) as a function of galaxy stellar mass. \(\langle \Delta v \rangle\) and \(\sigma\) are the mean and standard deviation of each distribution. \(\Delta v_{10}, \Delta v_{50}, \Delta v_{90}\) are 10th, 50th and 90th percentiles of the velocity separations.

|       | \(\log M_\star\) [M_\odot] | \(\langle \log M_\star \rangle\) [M_\odot] | \(\sigma_{\log M_\star}\) [M_\odot] | \(\langle \Delta v \rangle\) [kms^{-1}] | \(\sigma\) [kms^{-1}] | \(\Delta v_{10}\) [kms^{-1}] | \(\Delta v_{50}\) [kms^{-1}] | \(\Delta v_{90}\) [kms^{-1}] |
|-------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| ELGs: | Full Sample | 10.4 | 0.33 | -55 | 543 | -768 | 28 | 325 |
| LRGs: | Full Sample | 11.4 | 0.33 | 51 | 529 | -462 | 33 | 584 |
| ELGs: | [9, 10] | 9.8 | 0.2 | 151 | 341 | -68 | 52 | 513 |
|       | [10, 10.5] | 10.3 | 0.14 | -47 | 531 | -760 | 39 | 360 |
|       | [10.5, 11] | 10.7 | 0.15 | -120 | 576 | -1100 | 17 | 278 |
| LRGs: | [10, 11] | 10.5 | 0.24 | 122 | 394 | -100 | 85 | 286 |
|       | [11, 11.5] | 11.3 | 0.12 | 28 | 547 | -455 | -34 | 607 |
|       | [11.5, 12] | 11.7 | 0.12 | 66 | 528 | -498 | 125 | 578 |

from ELGs towards the observer powered by strong galactic outflows, while gas behind the galaxy is detected less frequently due to obscuration effects. At small distances, \(D_{\text{proj}} < 50\) kpc, the distributions are narrow and the vast majority of galaxy-absorber pairs have low relative velocities. At larger distances \(D_{\text{proj}} > 200\) kpc the tails of the distributions imply a small fraction of absorbing clouds with velocities higher than the escape velocity of dark matter halo for both ELGs and LRGs.

To study the dependence of gas kinematics on the stellar mass of galaxies we divide the galaxy samples into stellar mass bins and estimate the mean (\(\langle \Delta v \rangle\)) and standard deviation (\(\sigma\)) as well as the 10th, 50th, and 90th percentiles of the velocity distributions within 10 kpc < \(D_{\text{proj}}\) < 400 kpc. The values are compiled in Table 4. For both ELGs and LRGs, we see an increasing trend of gas velocity dispersion \(\sigma\) with stellar mass. For ELGs, this can be attributed to the positive correlation between SFR and stellar mass, and consequently stronger galactic outflows. However, for LRGs, this is more likely related to a trend of gas accretion rate with halo mass, either from the IGM or from satellite galaxies and their interaction with pre-existing halo gas.

Finally, we directly assess how our measured dispersion values depend on halo mass. The result is shown in Figure 19. For simplicity, the velocity dispersion is estimated by fitting a single gaussian in each halo mass range for both ELGs and LRGs. We clearly see that the gas motion around LRGs is suppressed (0.4 – 0.55 \(\sigma_{\text{halo}}\), as indicated by the dot-dashed red lines) over a large range of halo masses \(M_{\text{halo}} \sim 10^{13} – 10^{14.5}\) M_\odot). This is in contrast to ELGs (blue points), which exhibit dispersions similar to \(\sigma_{\text{halo}}\), as indicated by the dashed blue line. The trend for ELGs is also visible up to \(M_{\text{halo}} \sim 10^{13}\) M_\odot. We contrast against measurements from previous studies (Nielsen et al. 2015; Lan & Mo 2018), which we here extend to a larger mass range – this comparison is discussed in the next section.
Figure 19. The line-of-sight velocity dispersion, $\sigma_{\text{gas}}$, of galaxy-absorber pairs, within $10 \text{kpc} < D_{\text{proj}} < 400 \text{kpc}$, as a function of halo mass and galaxy types. The dark blue and red dots represent ELGs and LRGs, respectively. The dashed blue line shows the expected velocity dispersion as a function of halo mass of ELGs. The two dash-dotted red lines show the 0.5 and 4.0 times the dark matter velocity dispersion of LRGs. We also show the comparison with previous measurements from Nielsen et al. (2015) in violet and pink squares, and Lan & Mo (2018) in purple and orange-red triangles. The motion of gas around LRGs is suppressed relative to the expected velocity dispersion given their dark matter halo masses.

5 DISCUSSION: THE NATURE OF COLD GAS AROUND GALAXIES AT 0.5 < Z < 1

5.1 Comparison with Previous Studies

5.1.1 Mg II Covering Fraction

We now make a comparison between our new results and previous studies. First, we quantitatively compare our covering fractions with values obtained in Lan et al. (2014) for EW$_{2796} > 1$ Å and Lan & Mo (2018) for EW$_{2796} > 0.4$ Å, shown in Figure 20. Our analysis extends over larger spatial scales due to the increased statistics offered in DR16, but within the same scales probed, the results agree well within the error bars. We observe that for EW$_{2796} > 0.4$ Å, $f_c$ slightly lower than Lan & Mo (2018) below $D_{\text{proj}} \sim 50$ kpc. However, a similar covering fraction dependence with galaxy type is present. The minor discrepancy is likely due to the non-trivial differences in methodologies and analysis choices.

For EW$_{2796} > 0.4$ Å Lan & Mo (2018) estimated covering fraction using a stacking analysis. They assumed that the average Mg II absorption is dominated by absorbers with EW$_{2796} > 0.4$ Å and can be expressed as the product of covering fraction and the average value of EW$_{2796}$ estimated from the intrinsic distribution of individual absorbers toward random quasar sightlines. Lan et al. (2014) estimated the excess in the number of galaxies found within a given impact parameter w.r.t reference quasars as opposed to random positions. Then by counting galaxies around the selected quasars they estimated the denominator in the covering fraction expression. The error bars were estimated with Poisson statistics.

The uncertainties on Lan & Mo (2018) values are smaller than ours even though the absorber sample is smaller. This is due to the small errors on the average Mg II strength in stacked spectra. For EW$_{2796} > 1$ Å absorbers we agree well with Lan et al. (2014) at $D_{\text{proj}} < 100$ kpc even though our methods and resulting absorber catalogues are substantially different. For instance, Lan et al. (2014) used the Zhu & Ménard (2013a) Mg II catalogue (based on SDSS DR7 quasars) which has different completeness characteristics than our catalogue. We also note that the Lan & Mo (2018) results show less clear evidence for a sudden decline in covering fraction at a projected radius of 50 kpc.

5.1.2 Trends with Galaxy Properties

A detailed quantitative comparison of the stellar mass dependence of the Mg II covering fraction with results from previous studies lies beyond the scope of this paper. The study of Lan (2020) of the stellar mass dependence of covering fraction indicates that it is important to study these trends in relatively narrow redshift intervals, because there are relatively strong evolutionary trends in covering fraction with redshift at fixed mass. When this is accomplished, Lan (2020) find that after normalizing the projected distance by the virial radii the covering fraction of absorbers with rest equivalent widths (EW$_{2796}$) greater than 1 Å shows very little dependence on stellar mass for star-forming galaxies, but decreases with stellar mass for
the passive galaxy population. The weak dependence on mass that we find for our ELG sample is in agreement with these findings.

For LRGs we see a strong anti-correlation of covering fraction with stellar mass after normalizing projected distance by the virial radius, such that \( f_c \) is lower for more massive galaxies (bottom panel of Figure 14). These results are also in agreement with Lan (2020) for strong absorbers (EW_{2796} > 1 \, \text{Å}), who found that the covering fraction (on projected scale) varies weakly with stellar mass as \( f_c \propto M_*^{0.3} \), and varies from 2 percent from low mass to 3 percent for high mass LRGs at \( D_{\text{proj}} \geq 100 \, \text{kpc} \). For similar stellar masses, we find that the covering fraction varies from 3 percent to 7 percent. Note that our LRGs are more massive than DESI passive galaxies.

5.1.3 The Kinematics of Galaxy-Absorber Pairs

By connecting galaxy-absorber pairs in velocity space we estimated the ensemble velocity distributions of gas clouds associated with galaxies. Using single Gaussian profiles, we have characterized the velocity dispersion for absorbers (Figure 19). Using the line width at half-maximum intensity in stacked spectra, similar velocity dispersion values were obtained in Lan & Mo (2018). This work also reported a large velocity dispersion comparable to the dark matter halo velocity dispersion for ELGs and sub-virial gas motion (\( \sim 0.5 \sigma_{\text{sub}} \)) around LRGs. The sub-virial velocity of cold gas was also observed by Huang et al. (2016) for both passive LRGs and [O\text{II}] emitting LRGs (see also Chen 2017). Furthermore, the velocity dispersion scaling is in rough agreement with Nielsen et al. (2015, 2016) who characterized the galaxy-absorber relative velocity around blue and red galaxies using high-resolution spectra of quasars. The existence of broad components in the velocity distributions of Mg\text{II} absorbers around massive red galaxies is in agreement with the results of Kaufmann et al. (2017), who proposed that the large velocity separation Mg\text{II} absorbers trace gas that has been pushed out of the dark matter haloes by multiple episodes of AGN-driven mechanical feedback acting over long time-scales.

5.1.4 The main new results in this paper

Thanks to the larger quasar and galaxy samples that are included in our study, we are able to characterize the scale dependence of Mg\text{II} absorber covering fraction and rest equivalent width with greater accuracy. We find that the Mg\text{II} covering fraction for both LRGs and ELGs declines very rapidly at a projected radius of 50 kpc. Our results are consistent with previous findings, but the greater S/N of our measurements allows us to better characterize the sharpness of the break, indicative of a rather sudden transition between the regime where physical properties of the CGM are regulated by the intergalactic medium which can be enriched by metals ejected via galactic outflows, stellar winds, or supernovae in star-forming galaxies. The infall velocity of gas accreting from the IGM can be smaller or similar to the velocities expected from gravitational free-fall. In addition, cold gas can be in virial equilibrium with orbiting halo gas e.g. rotating with the halo orbital velocity. Alternatively, the metals forming inside stars can be thrown out of the galaxies by powerful supernovae with velocities up to \( \sim 500 – 1000 \, \text{km s}^{-1} \) (Dalla Vecchia & Schaye 2008; Sharma & Nath 2013). The powerful outflows from supermassive black holes can likewise eject gas out to \( \sim \) tens of kpc with velocities exceeding \( \sim 3000 \, \text{km s}^{-1} \) (Circosta et al. 2018; Perrotta et al. 2019; Nelson et al. 2019).

The high velocity dispersion around ELGs suggests that the origin of cold gas traced by Mg\text{II} absorbers around ELGs is likely due to powerful galactic outflows. Such powerful outflows around star-forming galaxies have been observed (Steidel et al. 2010; Bordoloi et al. 2014; Rubin et al. 2014; Zhu et al. 2015). This conclusion is also supported by the particularly strong enhancement of the Mg\text{II} covering fraction within \( D_{\text{proj}} < 50 \, \text{kpc} \). Using higher fidelity imaging, Lan & Mo (2018) found an enhancement of Mg\text{II} absorption along the minor axes of ELGs, which also supports the outflow scenario. In contrast, accreted gas from the halo or satellite galaxies is predicted to align along the major axes of galaxies (Péroux et al. 2020). Finally, we also see a strong positive correlation between the SFR of ELGs and covering fractions (differential values) (Figure 17), evidence that star driven outflows play a pivotal role in metal enrichment at \( \leq 100 \, \text{kpc} \) or \( D_{\text{proj}} \leq r_{\text{vir}} \) in star-forming galaxies.

On the other hand, the suppression of gas motion around LRGs indicates that the absorbing gas associated with them is gravitationally bound and unlikely to have originated from galactic outflows. This is further supported by the low star formation activity in LRGs and negligible contribution of stellar activity to the abundance of metals around them (Afruni et al. 2019). A similar suppression of gas velocity has been found for Ly\text{α} and O\text{VI} gas around COS-halo galaxies (Tumlinson et al. 2011, 2013). The interaction between cold and hot gas inside the halo also plays an important role in deciding the fate of cold gas around LRGs. As pointed in Huang et al. (2016), the drag exerted by the hot CGM can slow down the gas motion around LRGs and, cool clumps of gas can fall towards the central galaxy due to orbital decay. However, the evaporation time is significantly smaller than the infall time and most cool clumps would therefore evaporate before reaching the galaxy, or small distances (Zahedy et al. 2019). The observed existence of cold gas around LRGs implies that there is some form of balance between heating and cooling inside the halo such that the cold gas is routinely formed and destroyed in the halo (Sharma et al. 2012; Voit 2018). The origin, and survival, of such clouds of cool gas remains an unsolved theoretical question (Schneider et al. 2018; Nelson et al. 2020; Das et al. 2021).

Furthermore, our analysis also shows that a non-negligible fraction (\( \sim 15 \% \)) of LRGs have high-velocity gas clouds associated with them. As it is unlikely that galactic outflows powered by stellar activity could account for this population, one possible explanation could be AGN-driven outflows. The origin of cool gas around LRGs is undoubtedly therefore diverse and results from a combination of processes including accretion from the IGM, ram-stripping of the gas from satellite galaxies, mass loss from massive halo stars, and metal-enrichment by AGN driven outflows (Bordoloi et al. 2011; Huang et al. 2016; Lan & Mo 2018).

Combining our results with previous measurements (Figure 19), we see that the different behaviour of the gas motion around
star-forming versus quiescent galaxies supports the picture that the origin of cold gas is fundamentally different for star-forming versus passive galaxies.

6 SUMMARY OF CONCLUSIONS

In this work we have developed a fully automated quasar continuum estimator and absorption line detection pipeline. We have run the pipeline on the SDSS DR16 quasar sample and compiled the largest Mg ii metal absorber catalogue to date. Our main findings are:

- Our Mg ii catalogue contains 160,000 Mg ii absorbers (0.36 < z < 2.3) based on the SDSS DR16 final quasar sample. The median redshift and Mg ii absorption strength of the catalogue are z ~ 1.14 and EW2796 ~ 1.3 Å, respectively. Fe ii lines are also commonly detected and have smaller rest equivalent widths than Mg ii as they have weaker oscillator strengths.
- Stacking quasar spectra in the rest-frame of Mg ii absorbers, we detect the presence of other weak metal lines including Si i, C iv, Al ii, Al ii, Mg i and Fe ii. We study their properties as a function of Mg ii line strength.
- The measured doublet ratio of Mg ii absorbers shows that most lie within the theoretical limit of the Mg ii doublet ratio, i.e. 1 (saturated case) and 2 (unsaturated case). Strong absorbers are almost always saturated.
- To investigate the completeness of our detection pipeline we simulate ~ 33 million fake absorbers. Completeness is a strong function of both rest equivalent width and redshift of absorbers. Our method is naturally less complete for weak absorbers (EW2796 < 0.5 Å) and reasonably complete for strong absorbers. Our detection pipeline performs excellent on quasars with relatively high S/N. On the other hand purity is also very high (~ 95%) for our catalogue.

In the second half of the paper we connect Mg ii absorbers to the latest SDSS DR16 catalogue of ~ 1.3 million galaxies, divided into star-forming, emission line galaxies (ELGs) and quiescent, luminous red galaxies (LRGs). We use these samples to study the incidence and properties of metal absorption in the circumgalactic medium and nearby intergalactic medium. We investigate the absorber-galaxy correlation based on several properties of galaxies to understand the physics which most affect the properties of gas around galaxies. Our main results are:

- The mean surface density of Mg ii absorbers is larger than the expected random background for both ELGs and LRGs. At large distance, it reaches the expectation for random sightlines at a distance (DPproj ~ 15 Mpc) for both ELGs and LRGs. The mean surface density is also a strong function of galaxy type within DPproj < 50 kpc, with ELGs showing a stronger enhancement in the inner halo.
- The covering fraction of Mg ii absorbers varies strongly with galaxy type. ELGs have 2 - 4 times higher covering fractions than LRGs within DPproj < 50 kpc, regardless of EW2796 strength. The covering fraction decreases with projected distance and decreases strongly with increasing rest equivalent width (EW). For DPproj < 400 kpc, the covering fraction is larger than the expected random background and converges to this value only at DPproj ≥ 10 Mpc. A similar trend is visible for the cumulative covering fraction, which is clearly enhanced in star-forming versus passive galaxies up to DPproj ≤ 200 kpc.
- The average rest equivalent width of Mg ii absorption per absorber shows a clear enhancement close to galaxies, DPproj < 50 kpc for both ELGs and LRGs. LRGs always have systematically higher values than ELGs. There is a weak evolution of EW with galaxy redshift, such that higher z galaxies have slightly higher EWs at all DPproj ≥ 1 Mpc. In all redshift bins, the Mg ii average rest equivalent width declines to the field value beyond ~ 150 kpc.
- When normalizing projected distances by the virial radii of galaxies we find that the covering fraction varies strongly with galaxy type. There is a weak stellar mass trend for ELGs, but a strong stellar mass trend for LRGs, whereby the covering fraction decreases with increasing stellar mass, consistent with previous studies. The trend is visible even at DPproj ~ 2r200.
- Splitting ELGs by star-formation activity we find that Mg ii covering fraction is positively correlated with SFR, supporing the idea of a galactic outflow origin for the cold gas around ELGs.
- We trace the kinematics of galaxy-absorber pairs with the distribution of line-of-sight relative velocity Δv. Around ELGs, this is well-characterized by a single gaussian profile, while LRGs require a second, broad component with σv ~ 1100 km s⁻¹ to capture high-velocity tails, likely indicative of AGN-driven outflows. The velocity dispersion is higher around LRGs than ELGs. The majority of the absorbers close to galaxies (DPproj < 100) kpc have low velocities and are likely gravitationally bound.
- The velocity dispersion of cool CGM gas increases with stellar (and halo) mass for both samples. The gas motions around ELGs are similar to the expected dark matter halo velocity dispersion, while gas motions around LRGs are suppressed and closer to ~ 0.5σDM,halo. The different properties and, moreover, trends of Mg ii absorption with galactic properties between ELGs and LRGs implies that cool circumgalactic gas around star-forming versus quiescent galaxies has a fundamentally different physical origin.

In this study we have benefited from the cross-correlation of a large Mg ii absorber catalogue with the similarly expansive spectroscopic galaxy samples of SDSS. However, in understanding the galaxy-CGM connection we were limited in the physical properties of galaxies available. The ground-based wavelength limitations also restricted the available redshift range to connect galaxies and absorbers. Finally, available SDSS imaging does not generally allow fine-grained galactic morphologies to be inferred. However, such a morphology based analysis can be performed with the recent data from the Dark Energy Survey Instrument (DESI) Legacy Imaging Survey (Dey et al. 2019).

In the future, other large imaging surveys such as the Large Synoptic Survey Telescope (LSST) at the Rubin Observatory will provide enormous datasets of galaxies, up to higher redshifts, and with high quality imaging. Together with upcoming large spectroscopic galaxy surveys such as PFS on the Subaru telescope (Tamura et al. 2016), statistical analyses of the circumgalactic medium will be an ever more powerful tool to understand the formation and evolution of galaxies across cosmic time.

DATA AVAILABILITY

Data directly related to this publication and its figures is available on request from the corresponding author. This work is based on data which is publicly available in its entirety as part of DR16 of the Sloan Digital Sky Survey (SDSS; www.sdss.org).

ACKNOWLEDGEMENTS

We thank the anonymous referee for a very constructive and insightful report that has helped us improve the quality of the paper. We
also thank Ting-Wen Lan, Hong Guo and Guangtun Ben Zhu for discussions and assistance with various aspects of this paper. AA would also like to thank his colleagues Mohammadreza Ayromlou, Periklis Okalidis and Ilkham Galilullin for insightful discussions. We make use of Python libraries such as NUMPY, SCIPY and MANTOLOGY for data analysis and visualization.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org. SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics | Harvard & Smithsonian (CfA), the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

REFERENCES

Abolfathi B., et al., 2018, ApJS, 235, 42
Afruni A., Fraternali F., Pezzulli G., 2019, A&A, 625, A11
Ahumada R., et al., 2020, ApJ, 249, 3
Behroozi Y., Dekel A., 2003, MNRAS, 345, 349
Blanton M. R., Roweis S., 2007, AJ, 133, 734
Bollinger S., et al., 2014, ApJ, 794, 130
Bolton A. S., et al., 2012, AJ, 144, 144
Bordoloi R., et al., 2006, MNRAS, 371, 495
Bradshaw E. J., et al., 2013, MNRAS, 433, 194
Brown G. D., et al., 2016, MNRAS, 456, 2080
Cautel A., 1989, ApJ, 334, 100
Chabrier G., 2003, PASP, 115, 763
Chen H.-W., Helsby J. E., Gauthier J.-R., Shectman S. A., Thompson I. B., 2017, The Circumgalactic Medium in Massive Halos. Springer International Publishing. p. 167, doi:10.1007/978-3-319-52512-9_8
Chen H.-W., Helsby J. E., Gauthier J.-R., Shectman S. A., Thompson I. B., 2015, ApJ, 811, 171
Churchill C. W., Rigby J. R., Charlton J. C., Vogt S. S., 1999, ApJS, 120, 51
Churchill C. W., Mellon R. R., Charlton J. C., Jannuzi B. T., Kirhakos S., Steidel C. C., Schneider D. P., 2000, ApJS, 130, 91
Churchill C. W., Kacprzak G. G., Steidel C. C., 2005, in Williams P., Shu C.-G., Menard B., eds, IAU Colloq. 199: Probing Galaxies through Quasar Absorption Lines. pp 24–41 (arXiv:astro-ph/0504392). doi:10.1017/S1743921305002401
Circosta C., et al., 2018, A&A, 620, A82
Cooksey K. L., Kao M. M., Simcoe R. A., O’Meara J. M., Prochaska J. X., 2013, ApJ, 763, 37
Dalla Vecchia C., Schaye J., 2008, MNRAS, 387, 1431
Daw K. S., Choudhury P. P., Sharma P., 2021, MNRAS, 502, 4743
Dawson K. S., et al., 2013, AJ, 145, 10
Dawson K. S., et al., 2016, AJ, 151, 44
Dey A., et al., 2019, AJ, 157, 168
Eisenstein D. J., et al., 2011, AJ, 142, 72
Elahi P. J., Power C., Lagos C. d. P., Poulton R., Robotham A. S. G., 2018, MNRAS, 477, 616
Favole G., et al., 2016, MNRAS, 461, 3421
Herrmann K. A., Hunter D. A., Zhang H.-X., Elmegreen B. G., 2016, AJ, 152, 177
Huang Y.-H., Chen H.-W., Johnson S. D., Wechsler R. H., 2010, MNRAS, 405, 1713
Huang Y.-H., Chen H.-W., Shectman S. A., Johnson S. D., Zahedy F. S., Helmbold J. E., Gauthier J.-R., Thompson I. B., 2021, MNRAS, 502, 4743
Kaufmann G., Nelson D., Ménard B., Zhu G., 2017, MNRAS, 468, 3737
Kennicutt Robert C. J., 1998, ARA&A, 36, 189
Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2
Kewley L. J., Geller M. J., Jansen R. A., 2004, AJ, 127, 2002
Kriek M., van Dokkum P. G., Labbé I., Franx M., Illingworth G. D., Marchesi D., Quadri R. F., 2009, ApJ, 700, 221
Kroupa P., 2001, MNRAS, 322, 231
Lan T.-W., 2020, ApJ, 897, 97
Lan T.-W., Fukugita M., 2017, ApJ, 850, 156
Lan T.-W., Mo H., 2018, ApJ, 866, 36
Lee D. R., Seung H. S., 1999, Nature, 401, 788
Lopez S., et al., 2018, A&A, 613, A51
Lovegrove E., Simcoe R. A., 2011, ApJ, 740, 30
Lyke B. W., et al., 2020, ApJS, 250, 8
Maoraton C., Strömbäck G., 2011, MNRAS, 418, 2785
Ménard B., Wild V., Nestor D., Quider A., Zibetti S., Rao S., Turnshek D., 2011, MNRAS, 417, 801
Mitchell P. D., Schaye J., Bower R. G., Crain R. A., 2020, MNRAS, 494, 3971
Muratov A. L., et al., 2017, MNRAS, 468, 4170
Nelson D., Vogelsberger M., Genel S., Sijacki D., Kereš D., Springel V., Hernquist L., 2013, MNRAS, 429, 3353
Perrotta S., Hamann F., Zakamska N. L., Alexander R. M., Rupke D., 2015, ApJ, 813, 7
Peeples M. S., Werk J. K., Tumlinson J., Oppenheimer B. D., Prochaska J. X., 2015, ApJ, 818, 171
Pâris I., et al., 2017, MNRAS, 477, 616
Pâris I., et al., 2018, A&A, 613, A51
Peebles M. S., Weyl R., Winer L., 1962, MNRAS, 455, 1713
Pieri M. M., et al., 2014, MNRAS, 441, 1771
Pieri M. M., et al., 2014, MNRAS, 441, 1718
Prieur M. et al., 2014, MNRAS, 441, 1718
Prochter G. E., Prochaska J. X., Burles S. M., 2006, ApJ, 639, 766

The cool CGM in SDSS galaxies 21
APPENDIX A: RECEIVER OPERATING
CHARACTERISTIC (ROC) ANALYSIS FOR THE
PARAMETERS

To select an optimized value for the $\alpha$ parameter (see Section 2.3), we make use of the available DR12 Mg $\alpha$ catalogue compiled by Zhu & Ménard (2013a) to test our detection pipeline. The reference catalogue has 76,148 Mg $\alpha$ absorbers detected in 47,065 quasars. After applying all the masks as described in the section 2.3, we end up with 39,219 Mg $\alpha$ absorbers in 25,716 quasars. Then we run our complete detection pipeline on this set of quasars for several values of $\alpha$ and estimate the True Positive Rate (TPR) and False Positive Rate (FPR) to analyse the ROC curve to choose an optimized value for $\alpha$. To estimate the TPR and FPR we only use Mg $\alpha$ absorbers with EW$_{2796}$ > 0.1 Å ($N = 38,313$). We define TPR and FPR as follows:

A1 Adaptive SNR approach: Mg $\alpha$ doublet criteria

In this approach (using Rule 1, 2, and 3 described in section 2.3) we define TPR as the ratio of absorbers that passed our Mg $\alpha$ criteria to the total number of Zhu absorbers. FPR is defined as the ratio of all legitimate absorbers that passed our Mg $\alpha$ criteria, but did not match with Zhu absorbers, to the maximum of total Zhu absorbers and all absorbers detected with our pipeline.

We take the maximum of these two quantities to account for the case when our pipeline finds several absorbers that are not matched with the Zhu & Ménard (2013a) catalogue. Figure A1 show the ROC curve estimated for different $\alpha$ values in adaptive S/N approach. We clearly see that applying the adaptive S/N condition significantly reduces the false cases. For example at $\alpha = 2.5$ the corresponding TPR and FPR are ~ 0.8 and ~ 0.02, respectively. Therefore, we take advantage of the S/N of the spectra to reduce the incidence of false cases. As expected for larger values of $\alpha$, the TPR and FPR are both low while, increasing for smaller $\alpha$ values.

Finally, for our current catalogue, we conservatively choose $\alpha = 2.5$ because our ROC analysis is based on quasars (from the DR12 absorber catalogue only) with possibly high S/N compared to the entire DR16 quasar set used in the current study. By choosing a slightly higher $\alpha$ value we select relatively strong potential absorbers at the convolution step thus reducing the pipeline runtime significantly. However, we miss a good fraction of absorbers as shown in the completeness analysis (see section 3.3, Figure 7). On the other hand, this guarantees a very high purity for our catalogue as shown in Figure C1.

APPENDIX B: REST EQUIVALENT WIDTHS

As described in section 2.3.4 we measured the rest equivalent widths and the uncertainties of Mg $\alpha$ doublet by adding true noise from the background.
MgII Absorbers

EW (with random noise) [Å]

EW (with true noise) [Å]

ΔEW2796 [Å]

EW2796, Our [Å]

EW2796, Zhu [Å]

EWrand = 1.00 EWtrue + 0.0048

Figure A2. Comparison with Zhu & Ménard (2013a) DR12 Mg ii catalogue, for $\alpha = 2.5$. Top: EW2796, Zhu vs. EW2796, our, showing good agreement, though our EWs are slightly higher (as visible in best-fit dashed line shown in black). Bottom: Difference between these two values as a function of EW2796, our. The majority of the absorbers have $|\Delta EW| \leq 0.2$ Å and the typical error between our values versus Zhu is $\sim 0.16$ Å.

spectra as well as purely random noise to the absorption feature. We find that rest EWs and errors measured with these two approaches agree well. This further supports the fact that a very high fraction of our absorbers are genuine. The comparison of these two approaches is shown in Figure B1. We find that the difference is similar to the typical uncertainty in the measurement. The uncertainties measured from adding random noise are slightly ($\sim 5\%$) higher than the errors measured by adding true noise from the spectra. However, the overall match is very good, and we use measurements based on true spectral noise for our study.

APPENDIX C: PURITY ANALYSIS OF THE PIPELINE

In order to quantify the quality of our catalogue we also estimate the purity of simulated absorbers detected in our pipeline. As described in Section 3.3, we estimate the purity in each $EW - z$ bin as the ratio of detected absorbers to the total absorbers that pass our Mg ii criteria.

We show the 2D purity distribution of simulated absorbers as a function of rest EW and redshift of the absorber in Figure C1. We clearly see that the purity of detected absorbers is very high in almost every bin for $EW_{2796} > 0.4$ Å systems. For very weak absorbers ($EW_{2796} < 0.3$ Å) the purity is low because the corresponding completeness is also low, and possibly the pipeline finds false cases due to noisy features. The overall purity of our detection pipeline is $\sim 95\%$. Note that in Figure C1 we only show the purity for absorbers detected in QSOs having $S/N_{QSO} > 2$.

Figure B1. Comparison of rest EW and error measurements with random and true noise. Top: Measured rest $EW_{2796}$ (by adding random noise) as a function of rest $EW_{2796}$ (by adding true noise). The corresponding smaller panel shows the difference between two as a function of rest $EW_{2796}$ (by adding true noise). The agreement is very good and this further supports that measurements are robust. Bottom: Corresponding EW errors measured using these approaches. They match very well though error from random noise approach is slightly higher (as visible in best-fit dashed line shown in black).
Figure C1. Purity function $p(EW_{2796}, z)$ for QSOs with $S/N_{QSO} > 2.2$ distribution as a function of rest equivalent width and redshift. The purity is very high for detected absorbers. For very weak absorbers ($EW_{2796} < 0.3 \text{ Å}$) the purity is lower because the completeness is similarly low, and the pipeline is more likely to identify false cases due to noisy features.