MSci Research Project Report

Quasars Probing Galaxy Clusters: Measuring the Abundance of Neutral Hydrogen Gas in z > 2 Protoclusters

Alexander Woods
Jason Lunn

Project Supervisor:
Nina Hatch

School of Physics and Astronomy
University of Nottingham

May 2020

Abstract

In this project, we have successfully developed and applied our own quasar probing technique in order to study the diffuse and dark areas of protoclusters at redshifts beyond $z = 2$. Using protocluster-quasar data pairs, the foreground galaxy cluster environment can be studied using the line-of-sight spectra from the background quasar, due to photon absorption at the Lyman α transition wavelength. We used a sample of 7,654 quasars coupled to 211 CARLA protocluster targets in order to measure the abundance of neutral Hydrogen in the intracluster medium of galaxy clusters at $z > 2$. The equivalent width of Lyα absorption signals were measured to find the abundance of neutral Hydrogen. The strongest detection we observed is radially within approximately $2.8Mpc$ of the protoclusters central radio galaxy. The column density of the HI clouds we measured within this detection range was found to be $(2.78 \pm 1.66) \times 10^{14} cm^{-2}$, specifically for the $R_{\perp} = [200'', 400'']$ radial bin around protocluster targets. We also found a potential correlation between the foreground protocluster overdensity and the column density of HI within the intracluster medium. A relative column density increase of $(3.15 \pm 3.27) \times 10^{14} cm^{-2}$ was seen when moving from a sample of highly overdense protoclusters to less overdense ones.

Contents

List of Figures

List of Tables

1 Introduction

Understanding the characteristics of protoclusters, the progenitors of galaxy clusters and birth place of galaxies, is vital in establishing a complete picture of galactic and cosmic structure formation in the early Universe [Chiang, R. Overzier, et al., 2013]. They provide a unique test bed for studying different cosmological models, governing the development of large scale structure (LSS) in the Universe [Diener et al., 2015]. Protoclusters are also intrinsically linked to galaxy formation and evolution, where at high redshifts their cosmic volume is orders of magnitude larger than today. Consequently, they affect the cosmic star formation rate in their member galaxies, and the amount of ionising radiation produced by them. As a result, early protocluster evolution may also play a key role in the reionization of the Universe at $z \approx 6$ [Chiang, R. A. Overzier, et al., 2017].

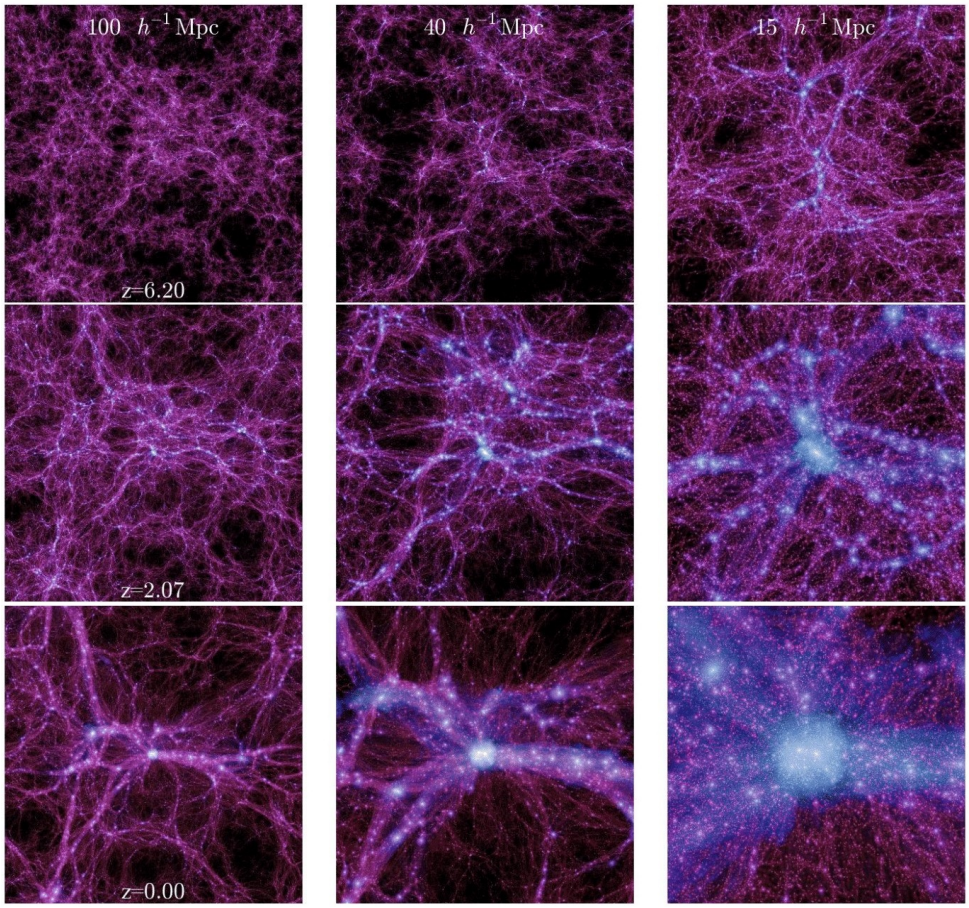

Figure 1: Protocluster evolution over redshift in an N-body simulation, showing gravitationally bound region overdensities collapsing to form the filamentary structure of the cosmic web [Boylan-Kolchin et al., 2009].

Motivated by this, in this report we aim to advance the current methods of studying these structures through a process of quantifying the abundance of neutral Hydrogen gas in these vast gravitationally bound filamentary structures at $z > 2$. With region overdensities of neutral Hydrogen tracing the baryonic and dark matter distributions within the protocluster, we can determine regions that will collapse to form galaxy clusters as $z \to 0$ shown in Figure 1, highlighting protocluster evolution in accordance with ΛCDM cosmology.

1.1 Studying Protoclusters

Studying these vast and diffuse structures at $z > 2$ presents many challenges to conventional observational techniques. Previous studies by Galametz et al. [2012] and Wylezalek et al. [2013] searched for and estimated region overdensities of protoclusters at $1.2 < z < 3.2$ through investigating a sample of radio-loud Active Galactic Nuclei (AGN), which act as beacons for locating protoclusters. Utilising data collected from the IRAC (Infrared Array Camera) on the Spitzer space telescope, counts-in-cell analysis was performed on the spatial distributions of detected IR sources within the field of each radio-loud AGN, selected by colour to ensure they were at the same redshift as the central AGN [Papovich, 2008]. This provided region density estimates for newly discovered and previously confirmed protoclusters, with Figure 2 highlighting one such field at $z = 1.64$.

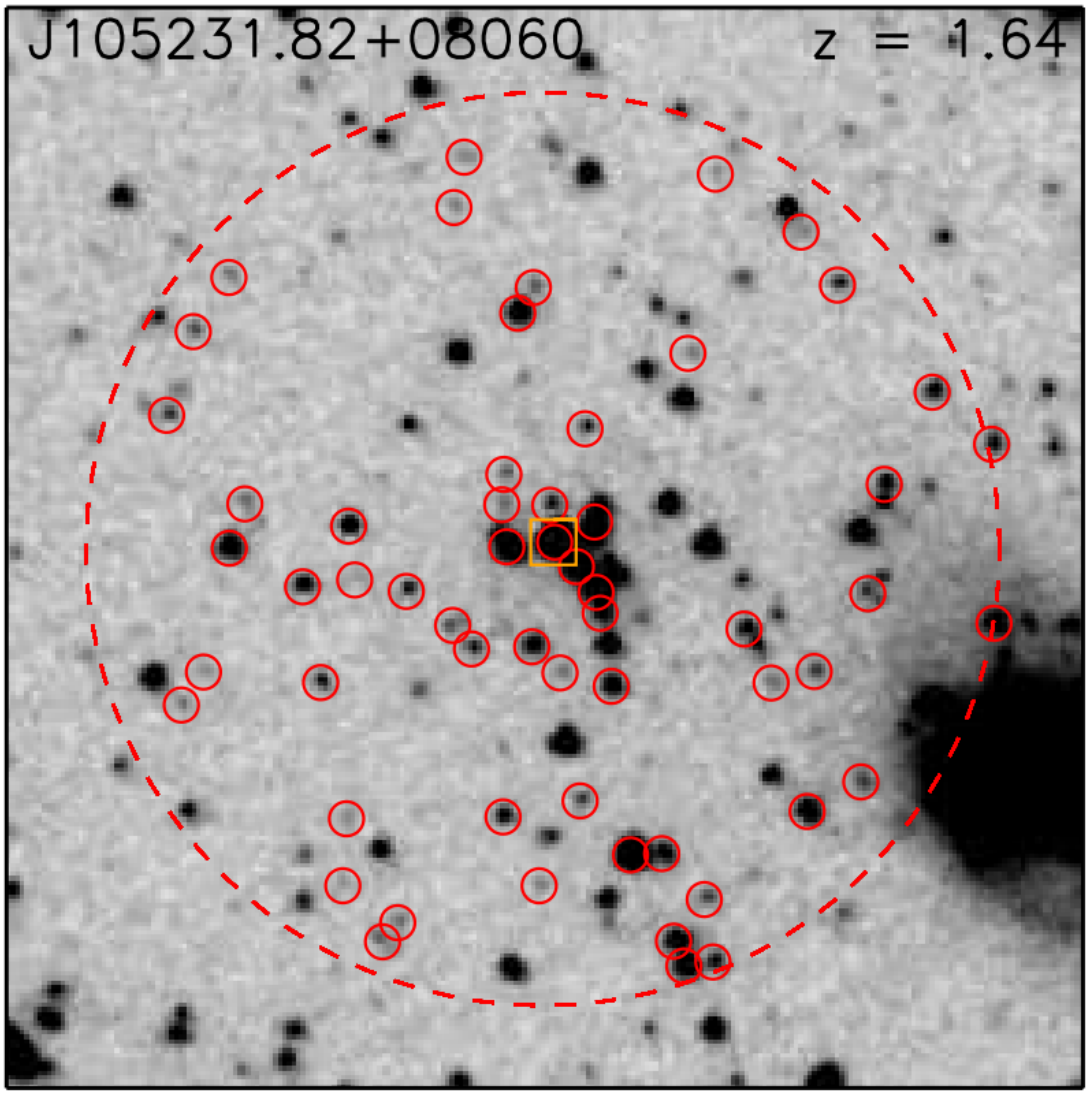

Figure 2: Infrared image of one field in the CARLA Spitzer survey [Wylezalek et al., 2013], showing the central radio-loud AGN (orange) and associate cluster members (red) within a 1 arcmin radius. The large number of red galaxies indicate this field is overdense, meaning this is likely to be a cluster or protocluster.

This statistical approach of utilising cluster member galaxies to estimate dense regions within a field is very effective at identifying protoclusters and overdense regions. However, to study the composition and structure of the cluster and Intracluster Medium (ICM) more precisely, spectroscopic observations are commonly used. The ICM can be observed directly at lower redshifts of z < 2, where the gas is heated sufficiently by shocks during infall, emitting thermal Bremsstrahlung X-ray radiation [R. A. Overzier, 2016; P. Rosati et al., 2004]. Utilisation of this process to study older, more distant and diffuse protoclusters at $z > 2$ becomes impractical due to the cooler ICM [Miller et al., 2019]. These affects culminate in ICM surface brightness dimming proportionally to $(1 + z)^4$, limiting these observations to lower redshift targets [Piero Rosati et al., 2002].

With these inherent limitations imposed on observing the hot gas directly from protoclusters, more novel ways to explore these giant structures must be explored. One path chosen by Mantz et al. [2014] successfully discovered a protocluster at $z = 1.9$ using the Sunyaev-Zeldovich effect. This involves detecting distortions in the CMB caused by inverse Compton scattering by relativistic electrons in the protoclusters ICM. Techniques such as this, which use an externally produced continuum of radiation to explore the characteristics of these distant objects will be explored in this report.

1.2 Quasar Probes

Hennawi et al. [2006] and Prochaska et al. [2013] performed a series of studies on the environments of foreground quasars through spectroscopy, using the continuum emitted by background quasars in the line of sight to the foreground quasar. They were able to quantify excess absorption associated to the foreground quasars environment. Motivated by their research, our project is focused on applying their quasar probing quasars technique to the study of neutral gas within high redshift protoclusters. By far, the greatest appeal of utilising this background quasar probing technique, is that it overcomes many of the difficulties in observing the gas within $z > 2$ protoclusters previously discussed. Moreover, this provides us with the capability to quantify the neutral Hydrogen abundance within these structures, achieved by measuring the HI Lyman α absorption caused by the cluster in the background quasars spectrum. This theoretically provides us with the means to study our faint protocluster targets at these high redshifts, exceeding what is possible with more conventional direct spectroscopic observations.

2 Background and Theory

2.1 The Lyman Alpha Forest

When the line-of-sight emission of a quasar passes through a galaxy cluster containing neutral Hydrogen gas (HI), absorption occurs at the wavelength of the Hydrogen transitions. The Lyman Alpha (Lyα) absorption line occurs for the transition of Hydrogen's single electron from the ground state (n=1) to the next highest energy state (n=2). This absorption corresponds to a wavelength of 121.567 nm or 1215.67 Angstroms (Å), placing it firmly in the ultraviolet band of the E.M. spectrum.

Observations in the UV section of the electromagnetic spectrum are not possible on Earth due to the complete absorption of the photons at those wavelengths by the atmosphere. However, if the source of these photons is far enough away in the Universe then cosmological redshifting will stretch and redshift their wavelength to the visible part of the E.M. spectrum, enabling them to penetrate the atmosphere. This then allows for ground based surveys of these distant quasar spectra. However, when we observe this absorption feature it does not appear as a single narrow line in the spectra of quasars. The light emitted from the quasar can encounter many clouds of neutral Hydrogen positioned at different redshifts along the line-of-sight towards Earth [Faucher-Giguère et al., 2008]. The photons are redshifted as they travel towards us due to the expansion of the Universe. Therefore, each cloud leaves a trace of itself in the form of an absorption line in the quasars spectrum at a different wavelength. This build up of lines in the spectra of quasars is known as the Lyman Alpha Forest [Rauch, 1998].

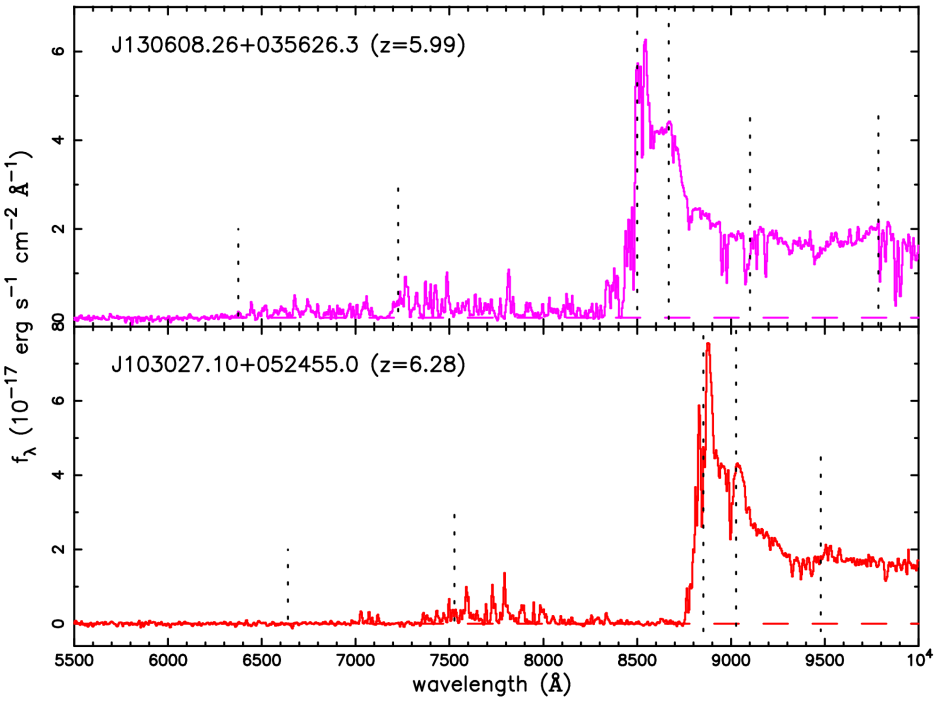

Figure 3: This image from Becker et al. [2001] contains two spectra emitted at different redshift values. Wavelength of the photons is on the x axis and their flux or strength is on the y axis. The top spectra in pink was emitted closer to Earth at $z = 5.99$ compared to the bottom spectra in red which was emitted from a source at $z = 6.28$. The important distinction between these two plots is the presence of a Lyman Alpha forest in the closer source spectrum. The Lyα photons from the source further away have encountered so much HI, that the continuum is mostly absorbed and thus produces a trough at around 8500 Å.

The spectra of relatively nearby quasars will not contain the Lyα forest feature if there are not enough clouds of neutral gas between us and the quasar to form the absorption feature. In the local Universe, most of the Hydrogen is fully ionised so there are few absorption features in low redshift quasars. The spectra of quasars at very high redshifts do not show a Lyα forest feature either. Due to their further distance away, the photons encounter such an abundance of neutral Hydrogen gas at varying redshifts that the absorption lines cover the entire wavelength range of the Lyα forest [Gnedin, 1998]. This results in a feature known as the Gunn-Peterson trough, named for its shape in the spectra of the distant quasars. Figure 3 illustrates a comparison between two spectra with one containing a Lyα forest and the other a Gunn-Peterson trough.

2.2 Quasars Probing Galaxy Clusters

The aim of this project is to detect neutral Hydrogen gas in and around forming galaxy clusters at $z = 2$ and beyond. This will be achieved by looking at how the spectra of background quasars are changed when a galaxy cluster in the foreground intercepts the emitted photons in the line-of-sight. The ICM of $z > 2$ galaxy clusters is not dense or hot enough to emit thermal Bremsstrahlung radiation, therefore most of the gas is not directly observable. The cosmological laboratory setup for this project is key to understanding the method of how we plan to detect the presence of neutral HI gas. In the background sits a quasar emitting electromagnetic radiation which we observe as an emission spectrum. Before reaching the observer here on Earth, the radiation passes through a galaxy cluster containing neutral HI positioned in the foreground.

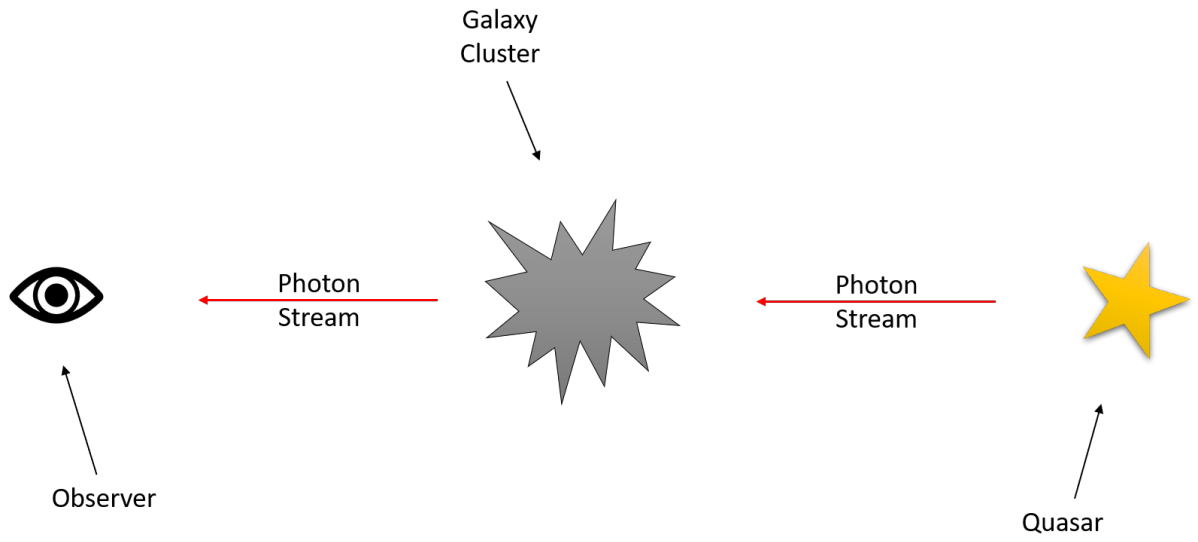

Figure 4: A diagram illustrating the relative positioning of the cosmological bodies being studied. A quasar sits in the background, a galaxy cluster in the foreground and the observer on Earth detects an emission spectrum from the quasar.

Figure 4 above shows a basic diagram of the geometry involved which allows for the detection of any potential neutral gas within galaxy clusters at $z > 2$. The redshift range that we are searching for an absorption signal in is chosen due to the lack of radiation produced from neutral gas in galaxy clusters during this epoch in the Universe. Clusters that are at a lower redshift (less than two) have ionised most of the gas within their ICM which negates the purpose and effectiveness of the methods and aims of this study.

2.3 Quantifying Neutral Hydrogen Abundance

As we have outlined, the Lyα absorption within our setup can be interpreted as the indication of the presence of a neutral Hydrogen cloud. However, we need to be able to quantify any signal we may detect in order to produce any useful analysis. In order to achieve this, we are required to measure the relative strength of any Lyα spectral absorption line we may detect. These spectral lines are generally not ideal narrow wavelength absorptions. They can be smeared out by imperfect instrumentation, therefore we need a system of measurement which is invariant for all qualities of absorption line profiles. The equivalent width, W_λ, is a quantity which allows an absorption feature in a QSO spectrum to be directly measured with this benefit. Measured in wavelength units, W_λ is found by taking a rectangular strip of spectrum given the same area as the absorption line. It is the width that the line would have if the relative intensity of the line were maximum everywhere, instead of spread out in a normal distribution. The equivalent width can be defined as;

$$W_\lambda \equiv \int \frac{F_{\lambda,0} - F_\lambda}{F_{\lambda,0}}\, d\lambda \tag{1}$$

where $F_{\lambda,0}$ is the flux at the continuum level, F_λ is the flux elsewhere across the spectral line. The equivalent width allows us to directly measure our absorption feature and produce a quantitative result. However, we would like to directly quantify the abundance of the HI gas in the cluster. To achieve this, we convert the width of the absorption line to an observable, the column density of neutral Hydrogen gas N_{HI}. Column density is a measure of the amount of matter intervening between an observer and the object being observed, typically measured by the number of Hydrogen atoms per cm^{-2} along the line-of-sight. The relation between W_λ and N_{HI} is complex, and is dependent on the optical thickness of the line being studied. For the purposes of this study, it is sufficient to only be concerned with the optically thin regime. This allows to use the relation below in Equation 2;

$$N_{HI} = 1.84 \times 10^{14}\, W_\lambda\, cm^{-2} \tag{2}$$

This assumption is useful since in the optically thin regime equivalent width is a sensitive measure of the column density for HI, where W_λ grows linearly with N_{HI}. The constant of proportionality relating the two quantities in the above equation is calculated using various relevant and known properties of the atomic transition we are searching for. The wavelength of the spectral line at Lyα = 1216 Å and the oscillator strength, f, of the Lyα equal to 0.4162 are examples of this [H.-W. Lee, 2013]. Most importantly, the relation in Equation 2 is constant for all Lyα absorption lines that will be analysed in this study, assuming we do not include optically thick damped Lyα systems in our data sample. We are now positioned to hunt for HI absorption and calculate the column density of HI gas clouds.

3 Data Selection

For this project, the foreground galaxy cluster targets are taken from the 'Clusters Around Radio-Loud AGN' (CARLA) survey carried out using the Spitzer Space Telescope [Wylezalek et al., 2013]. The spectra for the background quasar probes are taken from the extended Baryon Oscillation Spectroscopic Survey (eBOSS) which is part of the Sloan Digital Sky Survey (SDSS) Data Release 14 (DR14) published by Abolfathi and Aguado [2018]. The foreground targets were chosen this way to provide areas of sky in which we have confidence there is a protocluster object, which we can then compare to the general absorption seen at this redshift range from regions without massive protoclusters.

3.1 Combining Data Sets

As the theory compels us to search for cluster-quasar pairs aligned along the LOS, the first step we take with our data sets is to match objects according to their celestial coordinates. The quasars are treated as point source objects, and we take the protocluster centre to be the position of the radio galaxy. The foreground protoclusters are by nature large and diffuse objects, with the denser central regions on the order of 10Mpc in diameter [Muldrew et al., 2015], and the radio galaxy may not lie exactly in the centre.

The angular diameter distance, D_A, of an object is the ratio of its physical perpendicular size, d, to its angular size on the sky, θ [Peebles, 1993]. If we approximate to a flat curvature Universe, we can relate the angle an object subtends on the sky to its physical diameter as follows;

$$\theta = d \times \frac{(1+z)}{D_C} \tag{3}$$

where θ is the angular size on the sky, d is the objects diameter, z is the redshift of the object and D_C is the co-moving distance of the object from Earth.

Inputting the known quantities into the 'Cosmology Calculator for the World Wide Web' by Wright [2006], we find the angular size of the protoclusters on the sky. The scale given by the calculator is negligibly altered for the spread of redshifts in our protoclusters sample. For the purpose of our study, using the mean protocluster redshift to obtain the sky angle to scale relation is sufficient. We make an informed assumption for the reasonable maximum diameter of a protocluster at this epoch in order to determine the size of the radius on the sky around a protoclusters centre to spatially match background quasars to. For a redshift of $z \approx 2$, and an upper bound protocluster diameter of 25Mpc [Venemans et al., 2007; Yamada et al., 2012], we use a matching tolerance of one degree (3,600 seconds of arc) between our data sets.

Of the original half a million or more quasars surveyed by SDSS included in DR14, only 50,536 were found to be within one degree on the sky of any of the 420 CARLA protoclusters. Combining this with filters to ensure only quasars with a redshift greater than that of their respective cluster, and that only clusters at a redshift greater than two are permissible, we arrive at a total of 7,654 quasars behind 209 protoclusters. The majority of southern hemisphere protoclusters were found to have no related quasars since the SDSS catalogue of objects is collected from a ground based telescope situated in New Mexico, in the northern hemisphere [Abolfathi and Aguado, 2018]. Figure 5 outlines the distribution of the matched pairs, plotting the positions of all objects in both data sets.

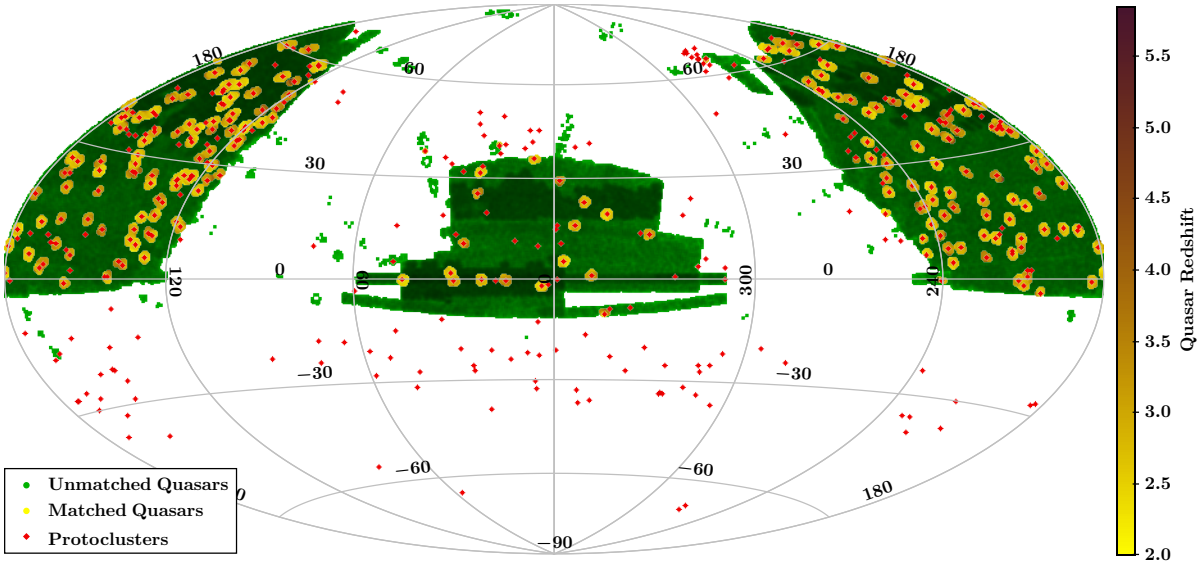

Figure 5: An Aitoff projected sky plot of all quasar-cluster data pairs. The right ascension and declination of each object was used to determine their line-of-sight separation on the sky. The quasi-stellar objects are represented as coloured dots weighted by their redshift as the colour-bar on the right shows. The central radio galaxies in the protoclusters are mapped as the red points.

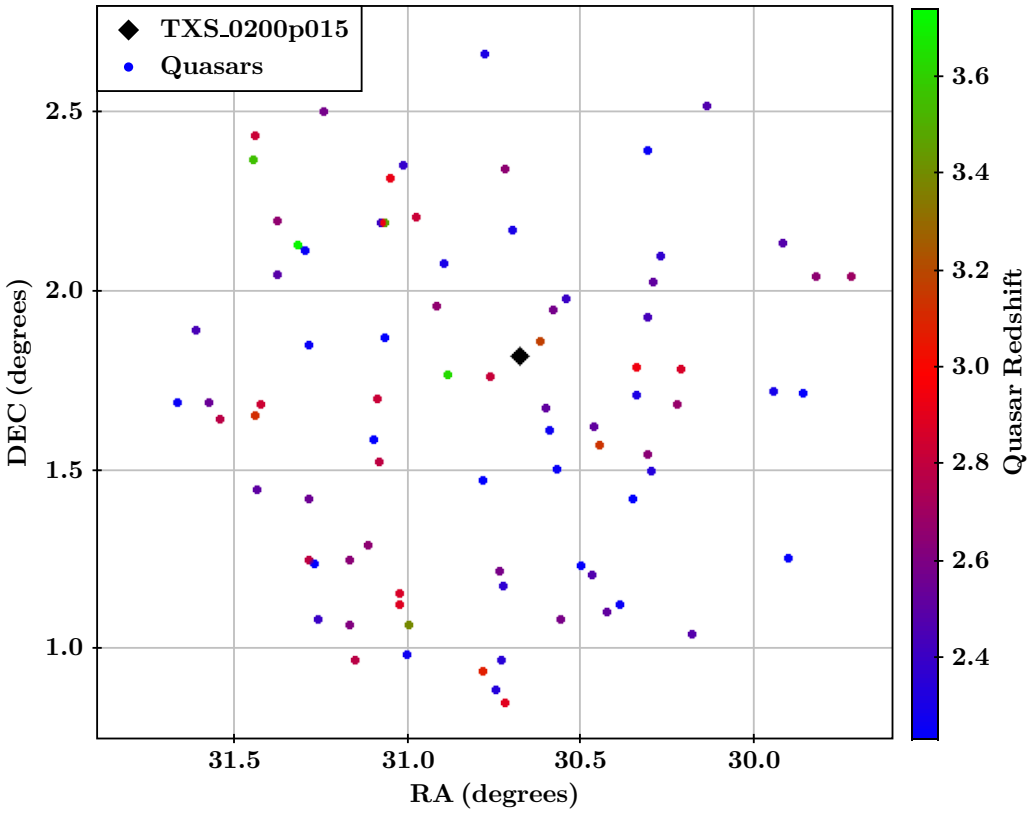

Figure 6: A position plot zoomed in to focus on the protocluster TXS_0200p015 with a central radio galaxy at redshift $z = 2.23$. All quasars are within 1° of the cluster's AGN position. Right ascension is plotted along the x-axis, and declination is plotted along the y-axis. All quasars in the plot are weighted by their redshift denoted by the colour-bar.

Figure 6 shows a positional plot for a single protocluster and all quasars found to be radially within one degree on the sky of the central radio loud AGN. This single protocluster plot demonstrates the typical random spread of quasar position and redshift that is seen for the majority of the data pairs in our matched sample. We perform a simple check on our data set in order to confirm there is no bias in the positions of quasars relative to their respective clusters. The distribution shown in Figure 7 confirms our prediction that the density of objects across the two-dimensional areas of sky follow a $1/r^2$ relation. The square of the separation between each object pair in our matched catalogue follows a flat trend within statistical confidence.

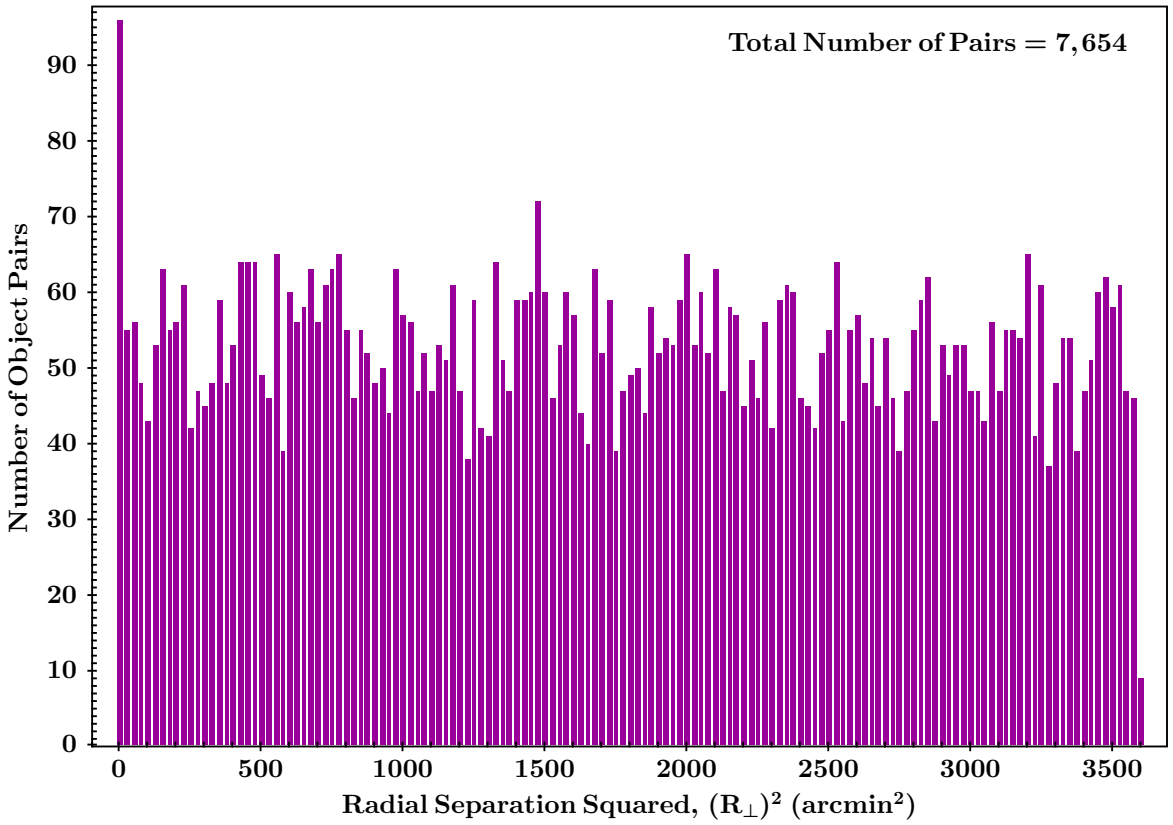

Figure 7: The plotted distribution of the squared separation of our data pairs. As expected, this distribution overall follows a flat trend indicating the number of objects within a particular radius follows a $1/r^2$ law where r represents the radial separation. The large spike in the first bin is expected due to the central radio loud galaxy in the protocluster being counted as a quasar behind the entire cluster. The significant drop of the right-most bin compared to the average is accounted for by the artificial separation cutoff introduced by us when matching our data.

The very first and very last bins in the Figure 7 distribution do not follow the flat trend expected. The over-filled bin at low separations is explained by the inclusion of the central AGN of the protoclusters also being counted as SDSS surveyed quasars in our matched data since it meets both the redshift and separation criteria. These can be omitted by further redshift filtering as outlined below in §3.2. The drop off at the furthest radius is an artefact of our artificial cutoff at a radius of one degree separation.

3.2 Data Filtering

Before attempting to analyse any of the spatially aligned quasar-cluster pairs, certain filters are applied to the data set in order to improve our final results. The first important consideration for our data is the signal to noise ratio of each spectrum. By initially removing relatively low S/N spectra, we eliminate as much noise as possible in the final combined data which in turn increases the likelihood of finding a strong Lyα absorption signal. Figure 8 below shows the logarithmic distribution of S/N ratios calculated by considering the variance in flux over the entire spectrum for each quasar. Since we are only particularly concerned with the Lyman α forest region of each spectrum, we also calculate the S/N ratio of this region as a standalone value to directly compare to the full S/N value. For our calculations, the forest wavelength range is defined as only the flux blueward, i.e. at wavelength values lower than the Lyα peak for each spectrum.

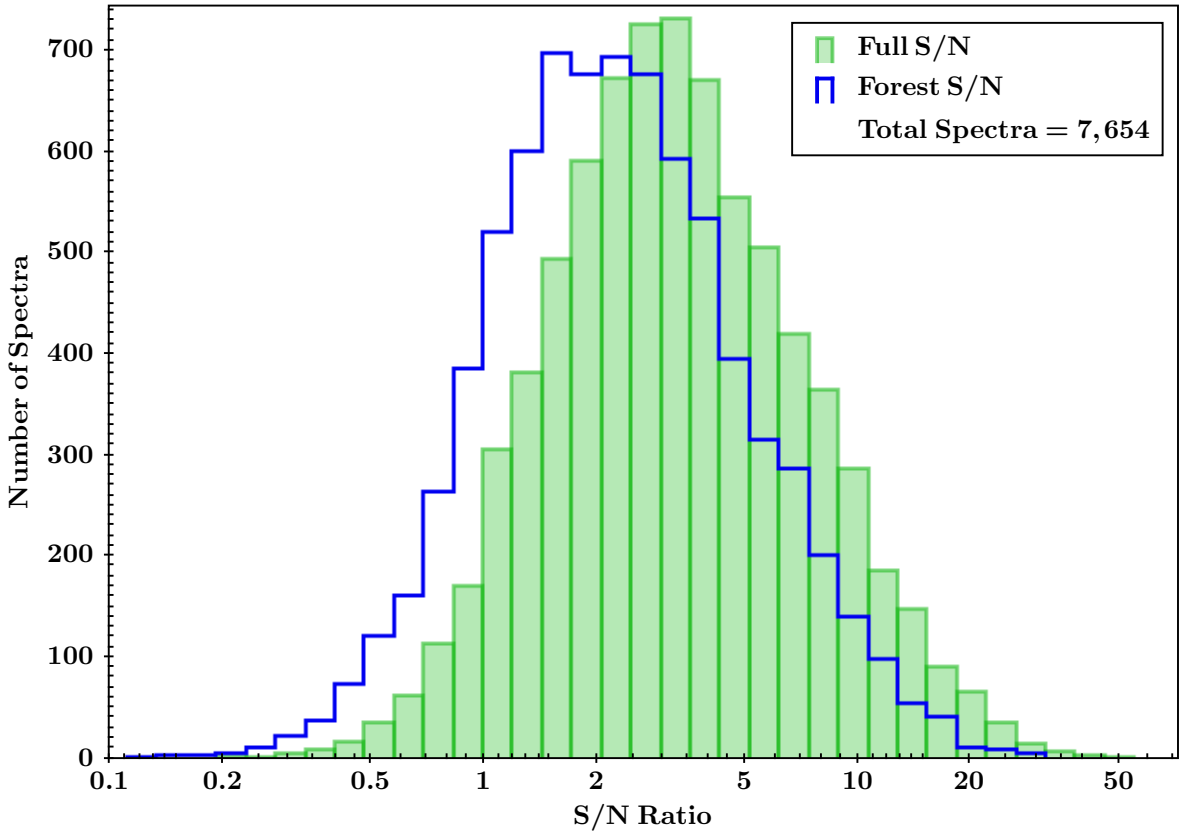

Figure 8: The logarithmic distributions of both the signal-to-noise calculations for the entire spectrum (green full bars) and only the wavelength range containing the Lyα forest (blue outline) for all spectra in our data set. Both show normal Gaussian distributions as expected for the noise, however, the centre of the normal distribution is at a lower S/N ratio when only considering the region of the spectra which is most important to our investigation.

The relative shift between the two signal-to-noise distributions shown in Figure 8 highlights an important distinction when filtering our data. Firstly, when choosing a lower bound cutoff for the S/N ratio, the number of spectra removed will differ which can have a substantial impact on any final results if not enough data is available to draw any solid conclusions. Secondly, there is no guarantee a 'good' S/N for an entire spectrum correlates to a 'good' S/N for only the Lyα forest region. This can introduce more noisy data into our results in the areas we are hoping to detect a signal. The Lyα forest region in each of the spectra is the main feature we are interested

in studying, therefore, we will use only the forest S/N ratio we have calculated to filter our data set. We initially choose a forest signal-to-noise cut at a value of 4. This is done to balance removing a large amount of noisy data yet also leaving enough spectra in our sample to enable us to confidently analyse and interpret our results.

Initially, the only redshift conditions applied to our set of protocluster-quasar pairs were to ensure all clusters are at a distance of at least $z > 2$, and also logically all quasars are at a redshift greater than their respective cluster. We take a further step with the relative redshift between protocluster-quasar pairs, in order to ensure the wavelength of the Lyα absorption we are hunting for at this epoch in the Universe does not overlap with the wavelength of the Lyα emission peak from the quasar.

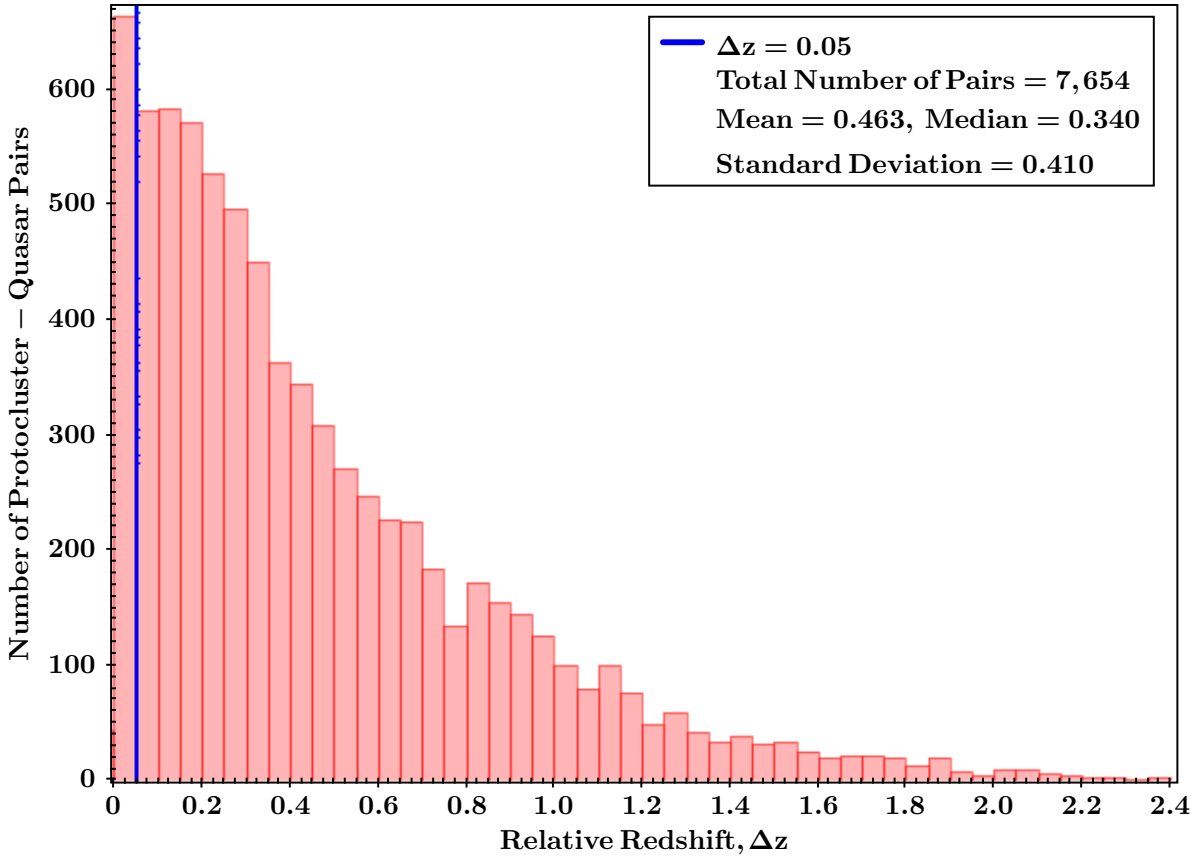

Figure 9: A histogram of the distribution of redshift difference between a protocluster and quasar in each pairing. The vertical blue line marks the lower bound cutoff for relative z between our pairs.

Figure 9 above shows the distribution of z difference. A cut at $\Delta z = 0.05$ is applied, removing $\sim 10\%$ of our matched data set from further processing. This cut is made to ensure a quasar does not reside within the protocluster itself, such as the central radio loud AGN as revealed by the distribution in Figure 7. These quasars must be omitted as their absorption would occur near the Lyα emission line, meaning protocluster absorption could not be discerned from absorption caused by gas in the quasars local environment [Prochaska et al., 2013].

A final filter central to our results analysis involves an important property of the protoclusters. The amount of galaxies combined with the approximated size measured by Wylezalek et al. [2013] allowed them to calculate an overdensity value for each protocluster. The overdensity variable is a confidence measure of the distribution of the density of objects on the sky. Overdensity is quantified in the number of standard deviations away from the mean density of objects over the entire sky. A higher value of overdensity indicates a higher probability that a protocluster will coalesce to form an established galaxy cluster as it evolves. The cumulative distribution of the overdensity values of all the protoclusters in our sample is shown below in Figure 11.

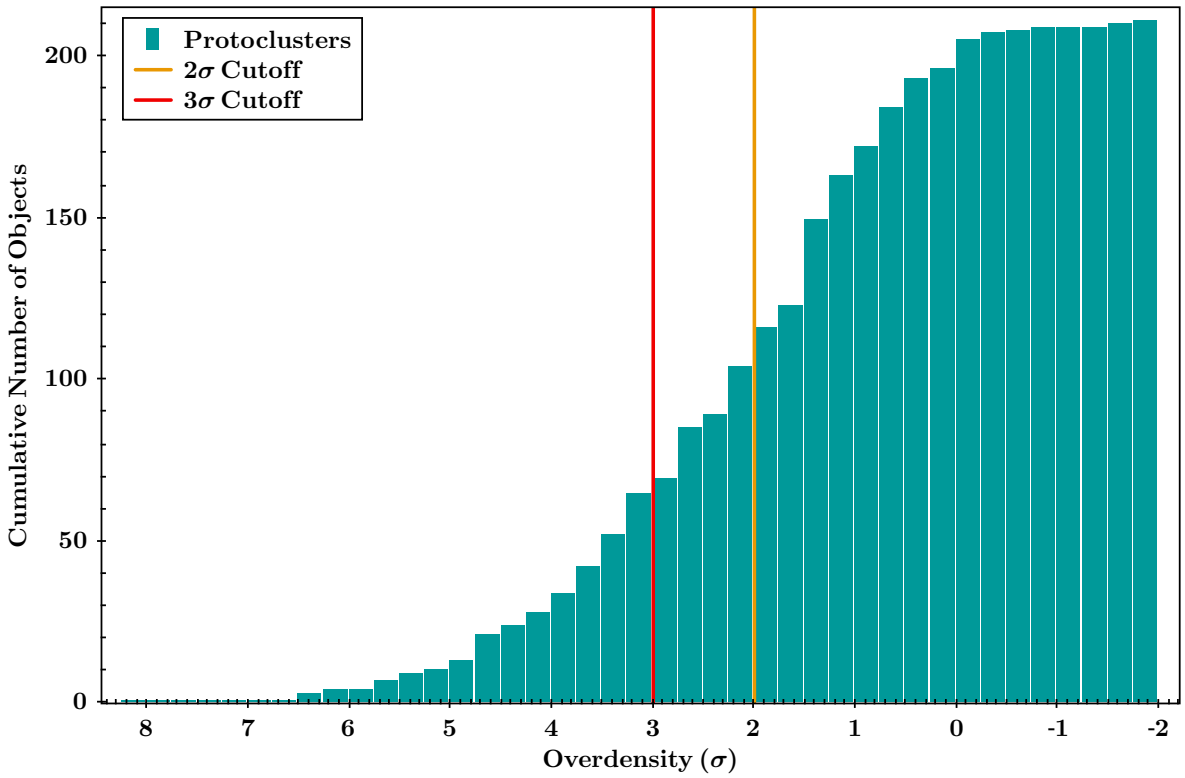

Figure 10: A cumulative histogram of the overdensity distribution for the CARLA protocluster objects. Plotted cumulatively to highlight the number of clusters removed with each confidence level cutoff taken. For the 2σ confidence interval (orange line) approximately half of all CARLA targets are taken out of consideration. Only around a quarter of the CARLA targets have a high overdensity above the 3σ confidence interval denoted by the red line.

Considerations for the overdensity of clusters will become important for part of the analysis of our final results. The number of galaxies associated with a protocluster and its total area on a patch of sky are used to calculate the overdensity value. Less overdense clusters are not automatically removed from our sample before processing, however the relation between Lyα absorption and protocluster density will be explored and analysed as part of our set of results.

3.3 Quasar Control Sample

For this experiment, we must also estimate how much Lyα absorption occurs in the intergalactic medium (IGM) on top of the specific absorption we are hoping to detect from the protoclusters. To this end, we will be using spectra from quasars which are not in the same line-of-sight as any of the CARLA protoclusters which are being studied in this project. The full eBOSS catalogue contains approximately 500,000 confirmed QSO objects. Any QSO targets

at a lower redshift than the closest protocluster target were automatically removed to improve the accuracy of the control sample. We expect generic Lyα absorption from sources besides the protocluster HI gas we are attempting to detect. This control measure allows us to directly compare any signal we find with the baseline signal created as a form of background noise from the Universe.

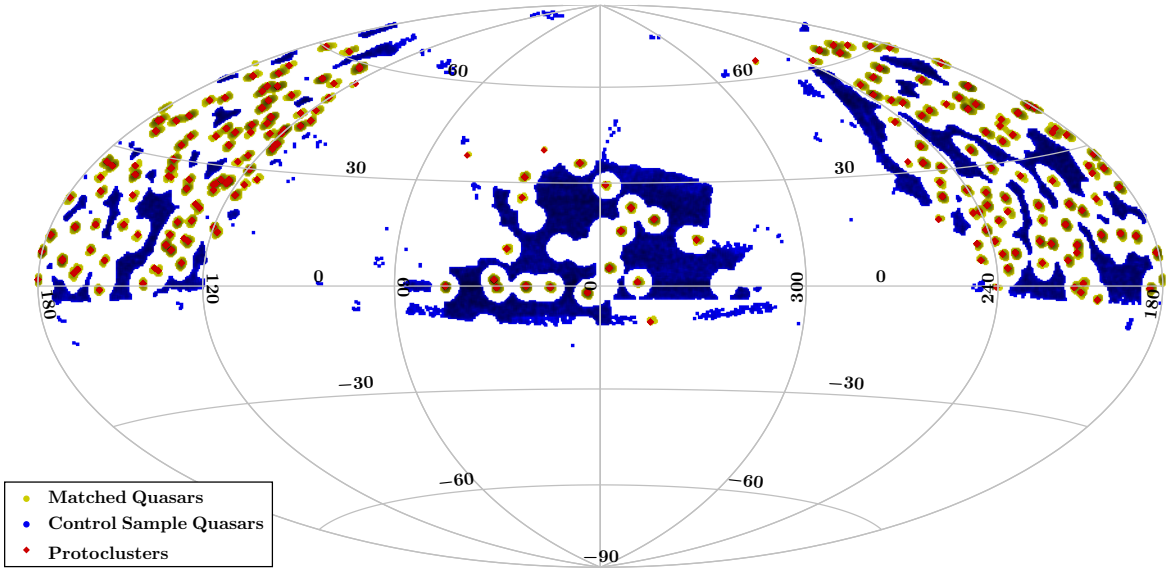

Figure 11: This positional plot shows the areas of the sky the control sample covers (blue dots). The protoclusters (red points) and their respective quasars (yellow dots) are also plotted to demonstrate the distance between our data pairs and the control data.

In order to balance the requirement of filtering any QSO's in a patch of sky occupied by a CARLA target and needing a sensible yet useful data set size for the control sample, we only used quasar spectra found to be at least 4.5° radially separated on the sky from any protoclusters in our data set. This distance was chosen as it guarantees no potential absorption from any CARLA targets, yet leaves a substantial amount of control spectra available in our control sample.

3.4 Continuum Control Sample

The second control measure we employed was a control set of continuum estimations, allowing us to compare results obtained using our continuum normalisation method outlined in §4.1, to another method which is well established and proven to yield viable results. We selected the sample provided by K.-G. Lee et al. [2013], based upon the earlier SDSS Data Release 9, from which 54,468 quasar spectra were suitable for Lyα forest studies. K.-G. Lee et al. [2013] performed optimisation and continuum estimations for each QSO spectrum, as outlined fully in their paper. This data set was used extensively by Prochaska et al. [2013] and Cai et al. [2017], the primary inspiration for our own study. Hence, utilising this data provides an ideal test bed for our own procedure, along with having a significant number of equivalent QSO spectra within our sample based upon the later SDSS Data Release 14. We find 2,162 matches between their data, and our own sample of 7,654 QSO sight lines. When performing measurements with this alternative sample, an equivalent control set was also derived from the 52,306 unmatched QSO in the data set provided by K.-G. Lee et al. [2013]. Allowing measurements made using this separate continuum estimation to be compared to an analogously produced control. This was obtained in the same manner as used for our sample outlined in §3.3, giving us a total of 29,194 equivalent control quasars, distributed similarly to our control as shown in Figure 11.

4 Spectral Analysis Methodology

Upon successfully obtaining our data set of CARLA targets and associate background sight-line quasars filtered accordingly, we performed the following spectral analysis. Due to the large proportion of low S/N spectra in our sample, coupled with relatively low resolution spectra obtained from SDSS. Individual spectral analysis is sufficiently unlikely to reveal the subtle absorption caused by our diffuse target protoclusters. Hence, with the aim of detecting the strongest possible absorption feature, composite spectra were produced. This allowed for several of the uncertainties, such as random unrelated IGM absorption, redshift error in the QSOs and foreground protocluster as well as sky noise [Prochaska et al., 2013] in the spectra to be significantly reduced through averaging.

4.1 Continuum Normalisation

The first processing step applied to all the quasar spectra in our sample is a continuum normalisation procedure, removing the emission profile of each quasar. This process is required to determine H I Lyα absorption by our target at $\lambda_{Ly\alpha}^{CARLA} = 1215.67$Å in the cluster's rest frame. Moreover, this is a mandatory step when combining spectra, ensuring the resultant composite is not disproportionately biased by individual spectra. Figure 12 highlights the variance in flux continuum level between various quasars within our data set. The continuum normalised flux is defined as $\tilde{F} \equiv \frac{F}{C}$ where $F(\lambda)$ is the observed flux and $C(\lambda)$ the estimated continuum level at wavelength λ [Faucher-Giguère et al., 2008].

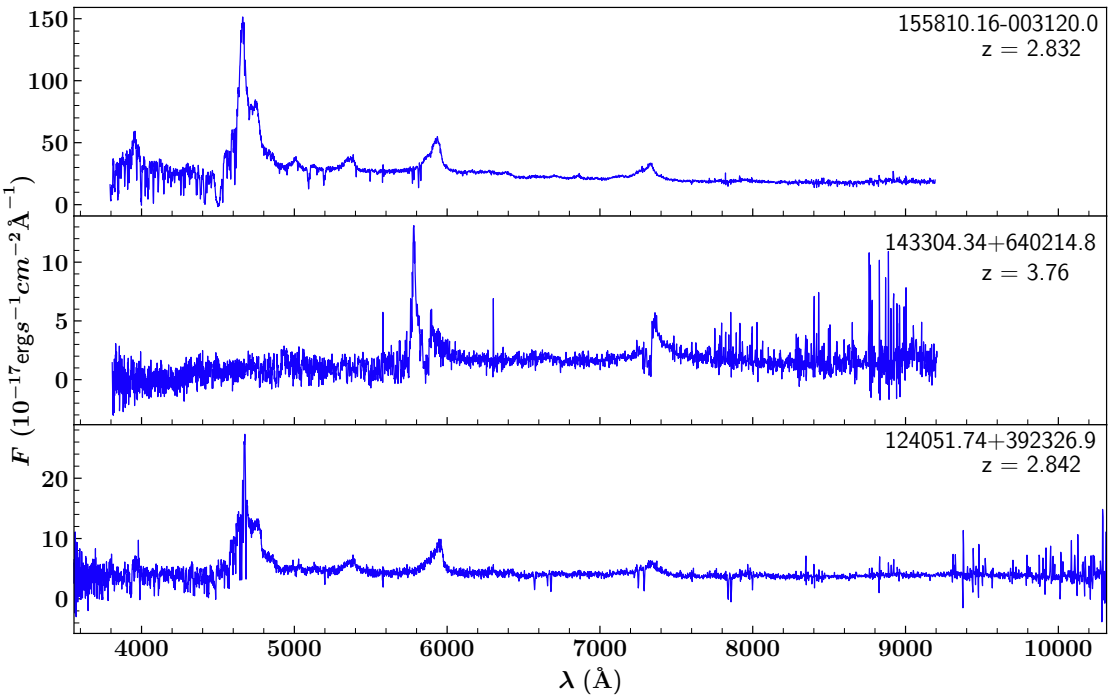

Figure 12: Raw quasar spectra with flux F plotted against wavelength λ in the observed frame, labelled with corresponding SDSS names and redshift values evidencing the variance in continuum level between the quasars in our sample.

4.2 Continuum Fitting Algorithm

The continuum estimation algorithm we developed for this task had to meet a few critical criteria. Firstly, it must yield physical results closely matching the unabsorbed level within the Lyα forest, without sampling absorption features, attenuating any potential signal. Secondly, due to the number of quasars $\approx 36,000$ in our combined control and LOS matched sample, we required the process to be relatively computationally efficient. Lastly, the process must be capable of handling spectra with anomalous features so to not bias composite spectral stacks in subsequent analysis.

The resultant process developed to extract C from the observed flux F, utilised statistical measures within a sliding window to sample a set of points representing the continuum. Additionally, to ensure emission lines listed in Table 1 were correctly sampled, markers were placed at corresponding λ in observed frame wavelength of $\lambda_{obs} = \lambda_{rest}(1 + z)$ where z is the quasars redshift.

	OVI	Lyα	NV	OI	CII	SiIV+OIV	CIV	HeII	OIII
λ_{rest} (Å)	1033.30	1215.67	1239.42	1305.53	1335.52	1399.8	1545.86	1640.4	1665.85

Table 1: Table of first 9 emission lines used in continuum fitting around the important Lyα forest region, used to ensure the derived continuum physically represents the emission profile of the source quasar, with the data provided by Abolfathi and Aguado [2018].

The algorithm required two main input parameters, each guiding the final continuum fit produced with these being window size and number of markers per window. Additionally, a set of statistical measures used to select the continuum markers required definition. The flow diagram in Figure 13 outlines how the parameters were used within the fitting algorithm, as well as outlining the stages in which the sliding window process was used. The sliding window was passed over the range of the entire spectrum, selecting marker points based on the flux distribution of the data within the window. Once all the markers had been located and stored, the sliding window was moved to sample the next section of the spectrum, repeating the same analysis until markers were located where necessary over the entire spectral range.

In order to provide the best fit within the Lyα forest, different parameters were chosen for $\lambda < \lambda_{Ly\alpha}$ and $\lambda > \lambda_{Ly\alpha}$ wavelength ranges. Maximising the goodness of fit in the Lyα forest was of paramount importance, as this ensured no absorption features were sampled into the continuum, minimising signal degradation within resultant composite spectra.

With this, two concurrent sets of markers where produced in each window region, calculated with different statistical measures each providing different continuum estimations. The first of which dubbed `MID` set marker points at $F(\lambda)$ values between specified upper and lower limits about the 50th percentile of the data in the window, e.g between the 45th and 65th or 40th and 60th percentiles. The second set of markers which we will refer to here as `MAX` differs slightly to `MID` in that only a single value was selected, in this instance from the upper quartile of $F(\lambda)$ within the window. The reason for this difference being that noise, especially within the Lyα forest had a greater presence within the upper quartile of data and sampling too many points within each window skewed the continua too high.

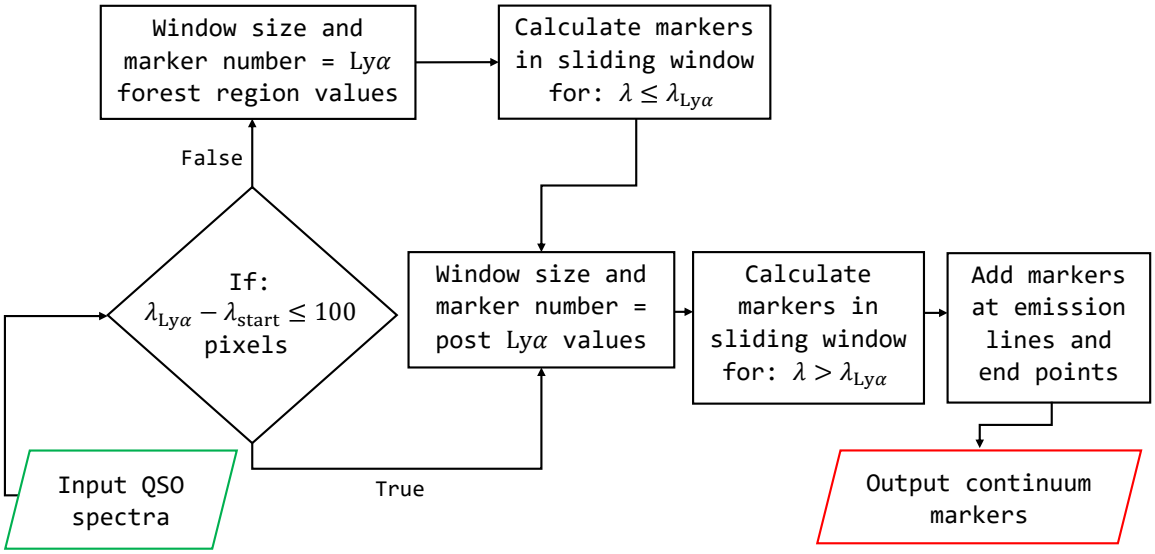

Figure 13: Flow diagram outlining the selection of the markers used to estimate the continuum level in each spectrum. Noting $\lambda_{Ly\alpha} - \lambda_{start}$ represents the number of pixels between the Lyα and the first λ value, checking the Lyα forest is sufficiently long enough to warrant individual treatment. Noting that when the conditional if statement returns true, the entire spectral range is still sampled, placing marker points in the short Lyα forest using post Lyα values throughout the whole spectrum.

The exact values of these parameters were tuned by processing numerous different spectra. This was partially achieved by visually inspecting the results to ensure physical continua were produced for each spectrum. Once the marker points had been calculated producing two sets of data points for estimating the continuum, each was linearly interpolated over the wavelength range of the spectrum giving a C value at all wavelengths. Finally, to yield a more natural result and reduce the effect of any outlying marker points, a Savitzky-Golay filter with a cubic polynomial fit was applied [Savitzky and Golay, 1964]. Likewise, with the sliding window used to obtain the marker points, a different filter window size was used for the Lyα forest and remaining parts of the spectrum. A value for each section, chosen to be double that of the sliding window for both regions was deemed to sufficiently smooth the continuum estimations.

Figure 14 shows both `MID` and `MAX` continuum estimates obtained, along with the markers used to derive them for the example spectra shown in Figure 12. From simple examination of just this small sample of quasar spectra, the competency of each variant in achieving both a physical continuum estimation, whilst minimising any fitting to absorption features is clearly visible. By comparing the top panels of Figure 14, it can be noted the `MID` method strongly fits to the absorption feature at 4500Å whereas `MAX` does not. Alternately, for this same quasar the goodness of fit shortly after the Lyα peak at $\approx$ 4600Å by `MID` exceeds that of `MAX`. Achieving a close fit to the NV emission line at $\approx$ 4700Å shortly after Lyα, With the `MAX` method showing a significant overestimate. For this reason, a new composite method was created using the `MAX` method during the Lyα forest region and `MID` throughout the rest of the spectrum for $\lambda > \lambda_{Ly\alpha}$, utilising each method optimally.

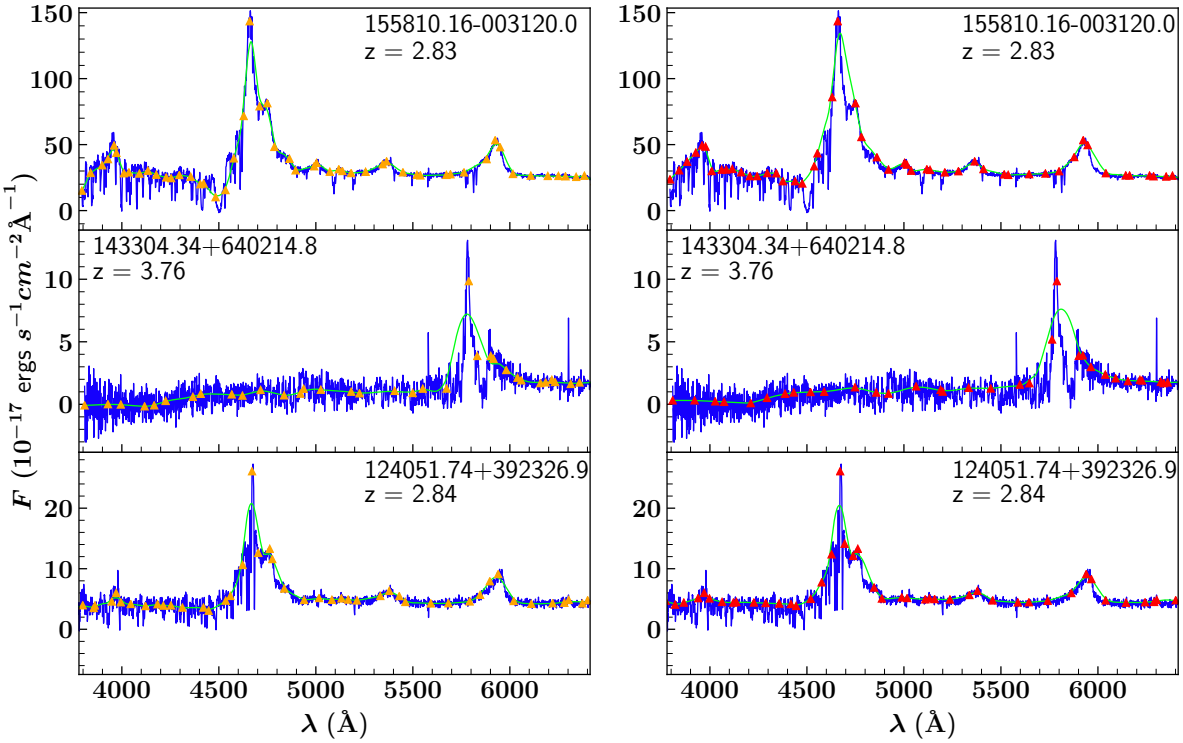

Figure 14: Continuum estimations (green) for the same QSO selected in Figure 12. Also given are the marker points `MID` (orange) and `MAX` (red) obtained by each method, shown the in respective left and right panels for each QSO. Illustrating how the continuum was determined and highlighting the different proficiency's in fitting by both variants over different parts of the spectra.

To highlight the benefit of combining the two estimations in this way, Figure 15 shows the resultant continuum normalised flux $\tilde{F}$ in relative flux units produced by each method for the quasar in the top panels of Figures 12 and 14. As mentioned previously, the deficiency of the `MID` method in correctly estimating the continuum level for the absorption feature present at 4500Å is emphasised in the resultant normalised flux. The underestimation of C causes the flux excess around the absorption feature clearly visible for this method, but absent in the spectrum normalised by `MAX`. Also noteworthy, is the excessive feature present at 4700Å produced by overestimation in the continuum level by `MAX`, alongside its absence in that produced by the `MID` method. This result again reiterates the motives put forward for the combined `MAX` and `MID` method, lacking both of these errors in continuum estimation as seen in the bottom panel of Figure 15. Consequently, this successful composite method was solely used throughout the rest of this study. This provides us with a versatile method for continuum estimation and normalisation, with its efficacy in relation to obtained results and alternate continuum estimation methods fully discussed within §5.

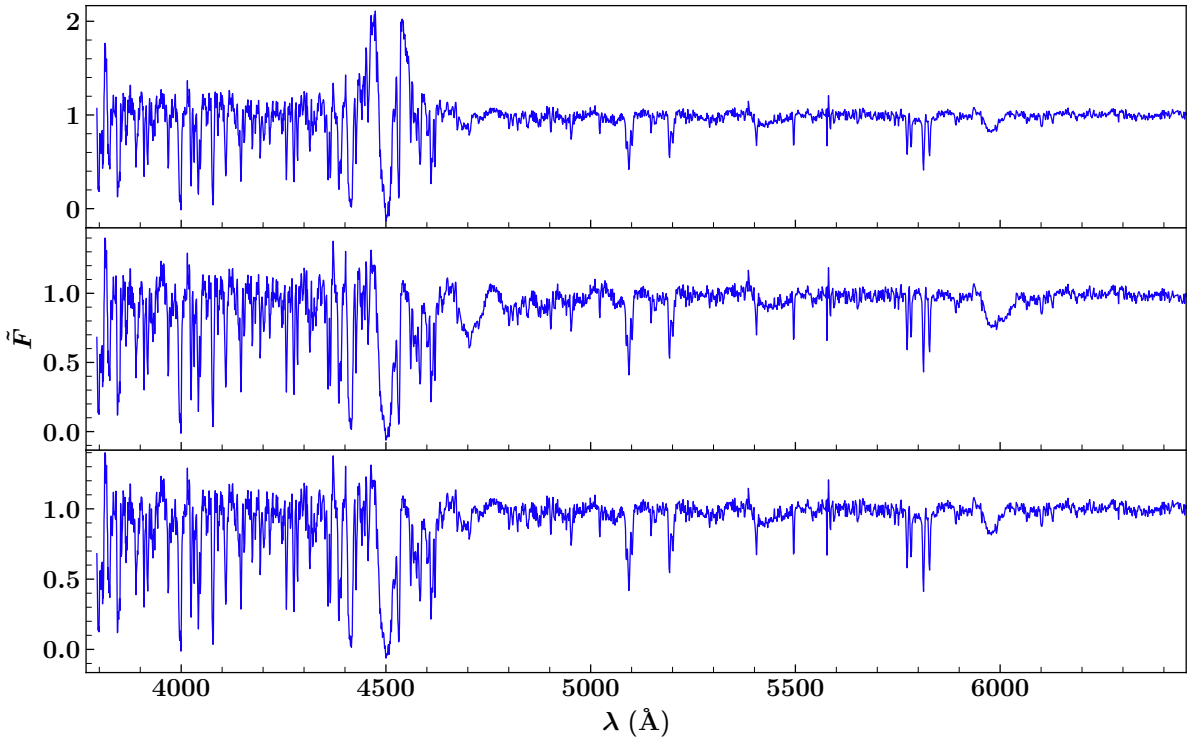

Figure 15: Continuum normalised Flux $\tilde{F}$ in the observed frame for the quasar 155810.16-003120.0, previously shown in the top panel of Figures 12 and 14. These plots show the result of the `MID` (upper panel), `MAX` (middle panel) and composite (lower panel) continuum estimations, comparatively highlighting the benefits of combining the two methods as mentioned.

Before any further analysis of the SDSS spectra was performed, a few optimisation procedures were carried out in order to mitigate sources of error propagating throughout the successive stages in our analysis. Visible in most spectra within our sample is significant noise at the blue end of the spectrum [K.-G. Lee et al., 2013], illustrated in the bottom panel of Figure 12. Hence, to reduce this in our selected sample, each spectrum was cut to the conservative wavelength range of $1030\text{Å} < \lambda_{CARLA} < 1600\text{Å}$ in the CARLA rest frame if it exceeded this range. This covered the important Lyα forest region, along with extending past Lyα to contain the other absorption features associated to the transitions listed in Table 1. Additionally, it improved the efficiency by significantly reducing the data size of each spectrum used.

4.3 Spectrum Stacking

Performed correctly, composite spectrum analysis will enable increased confidence in the detection of protocluster absorption through averaging out many of the sources of uncertainty mentioned at the start of this section. Prochaska et al. [2013] successfully evidence the success of this technique, obtaining composite spectra for sight line QSO's within different radii of their target yielding a strong absorption detection. A composite spectrum is created by shifting its constituent continuum normalised spectra $\tilde{F}(\lambda_{obs})$ in the observed frame, to the rest frame of the potential absorber. In our case the expected absorption is to occur at each foreground CARLA target located by its central radio-loud AGN. Hence, the associated background LOS quasars for each foreground CARLA target (see Figure 5) must be shifted to the reference frame of this foreground AGN via the relation;

$$\lambda_{CARLA} = \frac{\lambda_{obs}}{(1 + z_{CARLA})} \tag{4}$$

where z_{CARLA} is the redshift of the foreground CARLA AGN. In order to then combine the sub-pixel shifted spectra, each spectrum was linearly interpolated to the same number of data points and padded to account for discrepancies in the original spectral range of the input spectra. Finally the aligned and equally sized spectra were combined within each pixel bin through an average which omitted any padded values in the output. Both mean and median composites were to be used, with Figure 16 showing an example stack before and after combination illustrating the procedure.

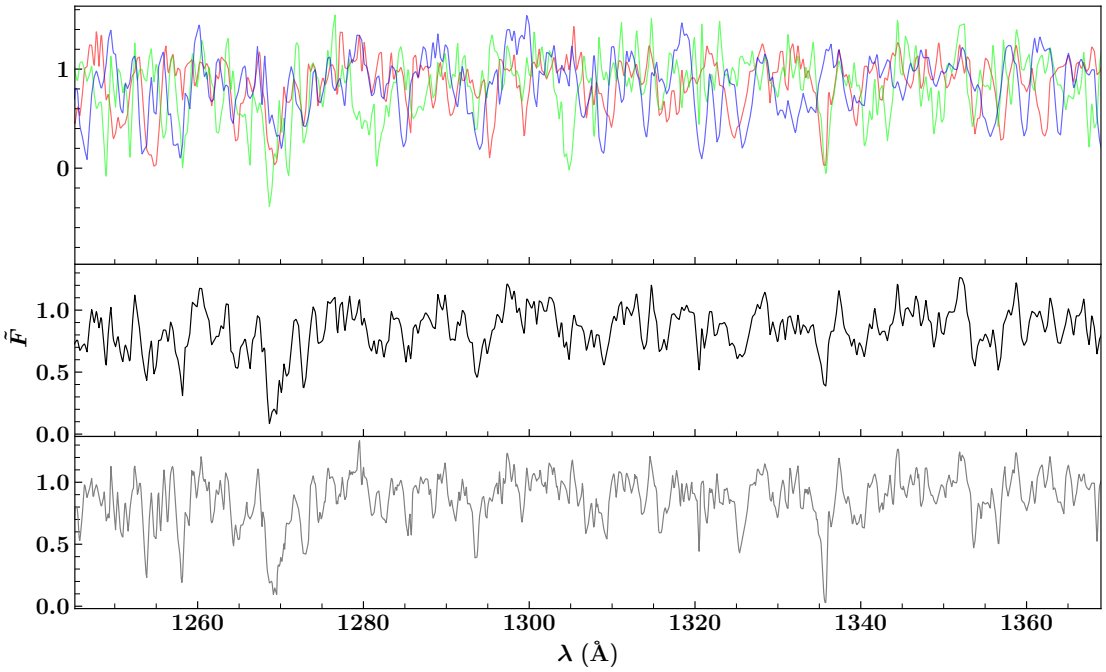

Figure 16: Example stack for three quasars all in the sight line of the CARLA protocluster target J092058.46+444154.0, where each QSO spectrum has been rest frame shifted as outlined in Equation 4 to the redshift $z_{CARLA} = 2.19$. The top panel shows the three quasars (red, green and blue) with continuum normalised flux prior to stacking, and the middle and lower panel composites showing the mean (black) and median (grey) flux produced respectively. We note that the prominent feature, present in all three quasars at ≈ 1270Å for example, is successfully combined and thus visible in both the mean and median composite outputs.

To accompany each composite spectrum produced, an analogous control stack is required in order to discern any absorption features related to our targets from random unrelated features caused by IGM absorption. By using quasars within the control sample we collected as outlined in §3.3, we created a control stack for each different LOS composite spectrum produced. Using the distribution of quasars and the redshift relation to their sight line CARLA target to define each associated control stack produced. The quasars in the LOS composite stack were binned by redshift into 0.05 intervals, from which a proportionate sample of quasars from the equivalent redshift bin were selected from our control data set. These were selected randomly within each bin from across the sky to ensure no region specific features were combined. This method for selecting the appropriate control sample of quasars for each LOS composite made sure the control was distributed equivalently, limiting the number of variables between each stack. Ultimately, this improves our confidence that any detections we make can be attributed to our targets, and not natural arbitrary absorption within the IGM present throughout the Universe. Similarly to our own data set, control stacks were produced for the continuum control sample [K.-G. Lee et al., 2013], obtained by the same processes as discussed previously. Examples illustrating the matching QSO distributions for the specified composite and associate control stack are given in Figure 18, along with the statistics for each distribution in Table 2.

To provide additional insight into the relationship between any potential features and the targeted protocluster candidates, the composite spectra are displayed in terms of relative velocity δv such that;

$$\delta v = c\frac{\lambda - \lambda_{Ly\alpha}}{\lambda_{Ly\alpha}} \tag{5}$$

where c is the speed of light, λ is the CARLA rest frame wavelength and $\lambda_{Ly\alpha}$ the wavelength of Lyα absorption at 1215.67Å ($\delta v = 0$ km s^{-1}). Additionally the resultant spectra were then binned by relative velocity, with 100km s^{-1} intervals to reduce remnant noise in the final composites.

5 Results

Implementing the series of tools outlined within previous sections, we performed a range of varying investigations on our quasar-protocluster data set in order to the uncover potential absorption features related to our targets. We proceeded first with a crude stacking test. This initial attempt involved stacking the entire library of protocluster-quasar sight line pairs, searching for net absorption related to our targets in QSO spectra over the whole 1° on the sky (≈ 25 Mpc).

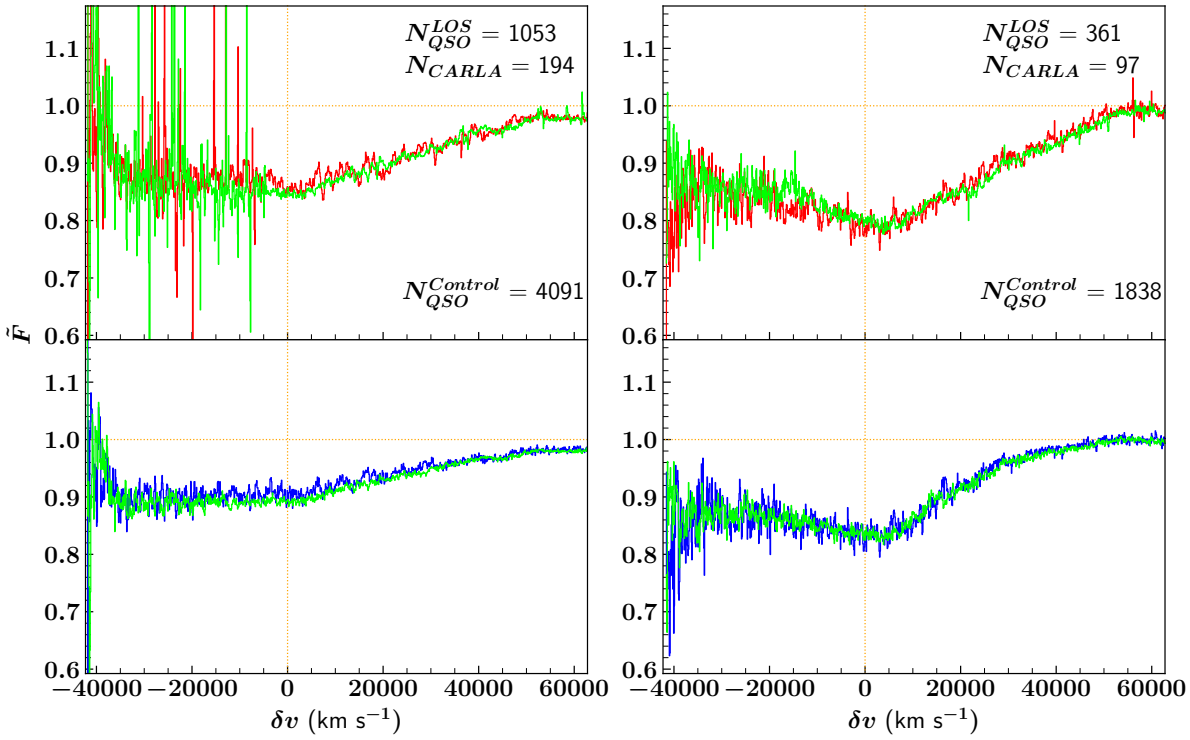

Figure 17: Mean (red) and Median (blue) composite spectra with corresponding control stacks (green), produced from stacking all the LOS quasar-protocluster CARLA targets within our data set. With the left-hand plots showing the results with spectra continuum normalised by our method, and the right plots produced with K.-G. Lee et al. [2013] normalised spectra. The vertical orange line at $\delta v = 0$ highlights the expected absorption line centre, and horizontal orange line at $\tilde{F} = 1$ represents unabsorbed level. This clearly illustrates the presence of broad absorption feature in our LOS sample and control - visible in both data sets. This feature covering the range of the Lyα forest can be attributed to net absorption in the IGM, aligned to the clusters reference frame. Note that the difference in noise between control and the LOS sample is due to the larger numbers of QSO used in each control measure. Also visible is the effect of spectrum optimisation performed by K.-G. Lee et al. [2013], seen in the difference in the amounts of noise between each mean composite (not visible in the median plots).

Figure 17 illustrates the resultant composite spectra produced by stacking our continuum normalised spectra and the equivalent matches within the continuum control sample. Also given is the corresponding control stack as outlined previously, with Figure 18 showing the similarly distributed quasar redshifts for the line of sight composite and control stacks for each data set. Table 2 contains the associated statistics for each redshift histogram, with S/N distribution values also given for reference.

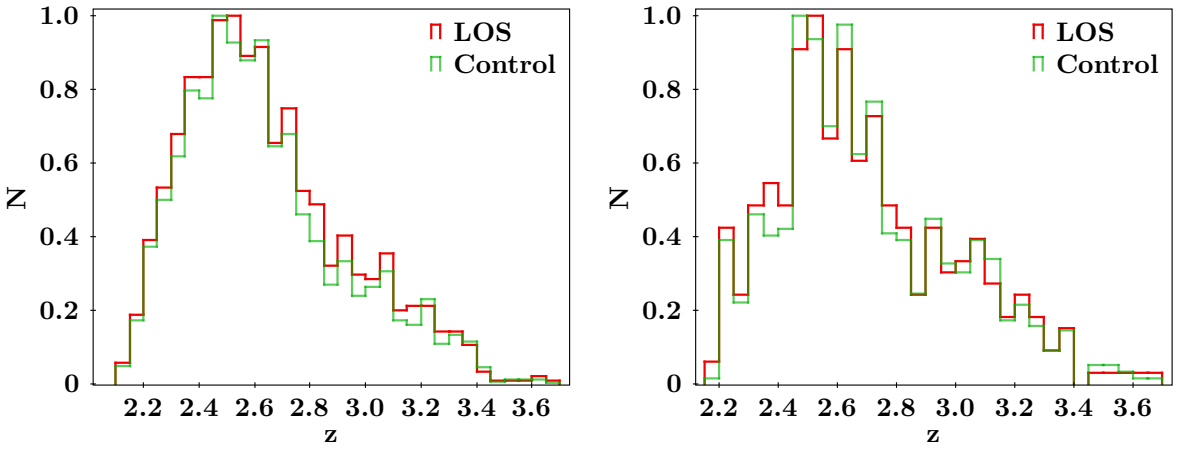

Figure 18: Normalised histograms by max bin count, showing the redshift distributions of the quasars spectra used to produce the composite stacks shown in Figure 17. The distributions for both our data set (left) and K.-G. Lee et al. [2013] (right), clearly show the level of similarity achieved between the control and LOS stacks with relevant statistics given in Table 2.

	$\langle z_{QSO} \rangle$	$\sigma_{z_{QSO}}$ [a]	$\langle S/N \rangle$ [b]	$\sigma_{S/N}$ [a]	N_{QSO} [c]	$\langle z_{CARLA} \rangle$	$\sigma_{z_{CARLA}}$ [a]	N_{CARLA} [c]
Our SDSS and CARLA Data Set								
LOS	2.645	0.294	6.817	3.073	1053	2.348	0.26	194
Control	2.637	0.291	8.133	4.479	4091			
Continuum Control Data Set [K.-G. Lee et al., 2013]								
LOS	2.693	0.303	6.756	2.979	361	2.403	0.273	97
Control	2.704	0.299	8.0236	4.434	1838			

[a] σ here indicates the associated standard deviation
[b] Mean S/N for the quasar spectra
[c] N denotes the total number of objects in the composite

Table 2: Associated statistics for the redshift distributions highlighted in Figure 18, further illustrating the strong similarities between the quasars in the LOS and control stacks. The mean quasar redshift, $\langle z_{QSO} \rangle$, and the standard deviation of the mean QSO redshift, $\sigma_{z_{QSO}}$, clearly show this relation. The control stacks contain quasar spectra which have no associated line-of-sight protocluster, hence no related data to display.

Evident in both the mean (red) and median (blue) composites of Figure 17 is a significant broad absorption feature. We can directly attribute this to a net combination of random absorbers, throughout the IGM along the LOS for each quasar in the stack. Showing absorption over the entire Lyα forest range as we would expect, with the maximal absorption near $\delta v = 0$ indicated by the orange vertical line. We can confidently make this assertion by pointing out that an almost identical absorption feature is visible within the corresponding control (green) stacks produced. Providing additional confidence that this feature can be attributed to absorption in the IGM throughout the Universe, and is not a local feature caused by our LOS protocluster targets.

Hence, we can conclude that this preliminary result yields a measure of the constant IGM absorption present throughout our sample, and does not show any feature directly relatable to our targets within the LOS sample. However, the reason for not observing absorption caused by our target can be attributed to the large number of spectra used for this initial test. This is because not all sight line QSO in our LOS sample exhibit absorption features from our target with a strength exceeding that of random absorption due to the IGM, or they have a sufficiently low S/N such that any signal is lost to noise. This means that for the QSO that do exhibit absorption, we are essentially diminishing this signal by saturating it with too many noisy spectra, resulting in a loss of signal when combining them together. In subsequent results, this point is fully explored with many examples of small composite stacks confirming this idea.

The final result we can take away from this preliminary test, is the proficiency of our continuum fitting normalisation procedure. It yielded comparable results to our control set, showing only slightly weaker absorption as well as enhanced noise seen at large negative δv values in the mean (red) plot. We deemed these slight deficiencies acceptable, as the additional noise does not appear near the region we expect to find absorption by our target at $\delta v = 0$. Additionally we do not see this noise within the median (blue) plot, meaning only a few outlying spectra with excess noise are causing this issue and thus will have minimal impact on different smaller composite samples. The weakened absorption seen when comparing each sample was also deemed to not impair our ability to confidently identify an absorption feature. Similarly weakened measures can be seen in the control, meaning this factor will not bias our neutral Hydrogen detections when measured relatively to this. Additionally, the number of QSO in each differs, and so identical results should not be expected when considering variances in the random velocity dispersion of IGM absorption present in each spectra. Hence, with the validity of our method analysed and confirmed, we will proceed with this for the remainder of our research, with the benefit of having a larger number of QSO sight lines within our sample to study.

5.1 Radially Binned Stacking

Because of the apparent oversaturation exhibited in the initial test for stacking our entire LOS library, we concluded that we must select the QSO spectra carefully in order to maximise the number of those with potential absorption features (that are detectable above the noise in each spectrum) in our results. Consequently, to reduce the number of spectra in the composite and increase the chance of making a detection, we obtained a series of composite spectra with quasars within a specific radii of the central AGN protocluster target. The initial test results shown in Figure 17 were obtained for all QSO and CARLA targets with angular separations on the sky of $0 < R_{\perp} < 3600$ arcseconds as shown in Figure 7. We now will examine the inner and theoretically denser region of the protoclusters within $R_{\perp} < 1000''$ or ≈ 7 Mpc, maximising the number and strength of any corresponding absorption signal in the LOS quasar spectra. In order to ensure signal is not saturated as outlined, and also locate the densest regions within the protocluster, we obtained composite spectra in 200 arcsecond bins from $0''$ to $1000''$. corresponding to ≈ 1.4 Mpc concentric rings around the central CARLA protocluster target.

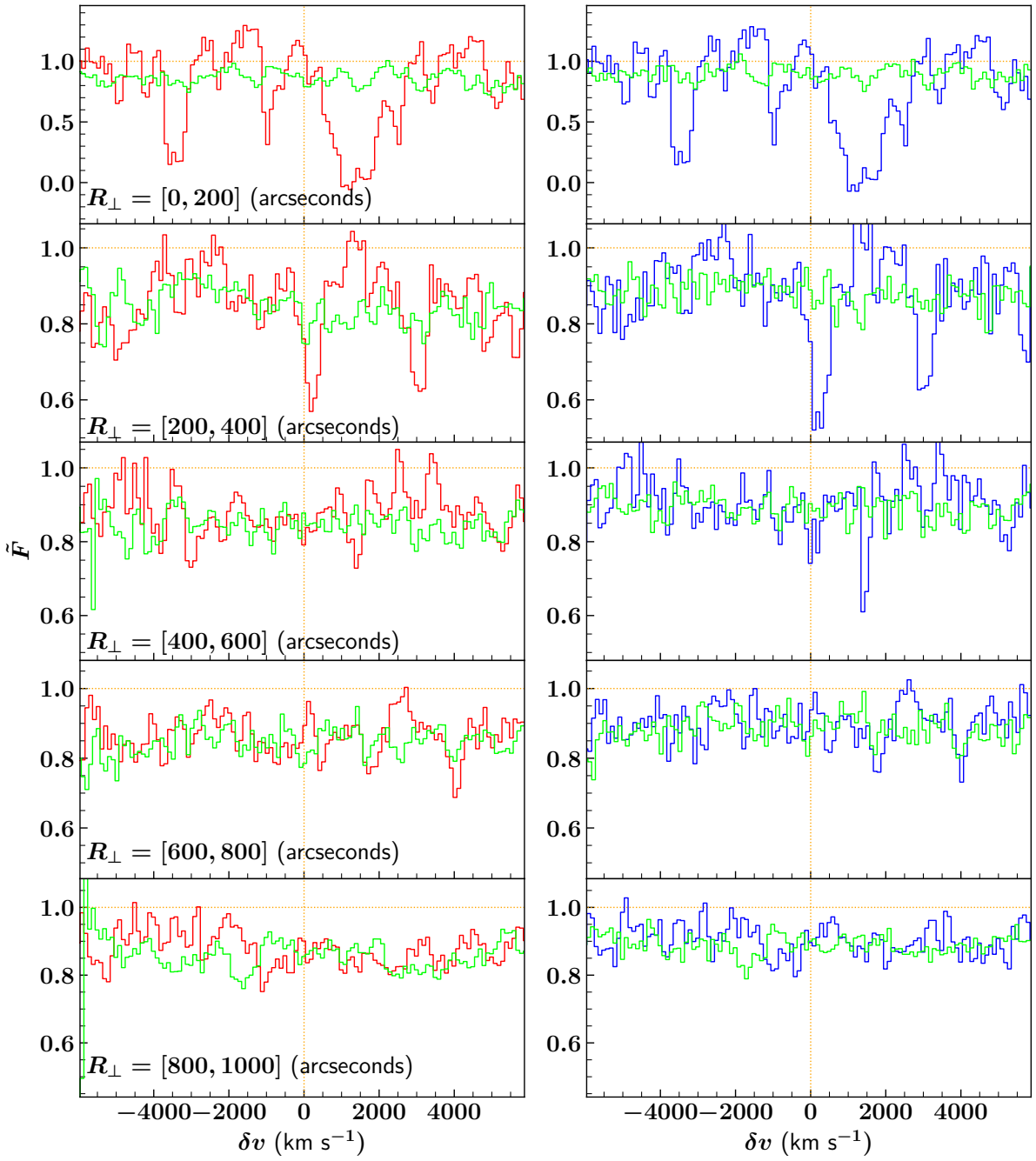

Figure 19: Radial binned composite spectra, with mean (red) on the left and median (blue) on the right for each 200 arcsecond interval from $0''$ to $1000''$ as indicated within each plot. Here we can see the strongest absorptions for the $R_\perp = [0'', 200'']$ and $R_\perp = [200'', 400'']$ composites, compared to their equivalent control absorption level (green). The remaining higher radii composites exhibit no significant absorption above the control level, with the narrow feature at $\delta v \approx 1500$ (km s^{-1}) in the median composite for $R_\perp = [400'', 600'']$ the only notable feature in these larger radii spectra. The number of CARLA targets as well as quasars stacked producing each composite are displayed in Table 3. The vertical orange line at $\delta v = 0$ (km s^{-1}) indicates where we expect Lyα absorption by our target to take place, with the horizontal orange line at $\tilde{F} = 1$ representing the unabsorbed continuum level.

The resultant composite spectra for each $R_\perp$ range displayed in Figure 19, with the corresponding number of quasars and CARLA targets for each stack is given in Table 3. Visible in the mean and median plots corresponding to the $R_\perp = [0'', 200'']$ composite is a strong absorption feature centred at $\delta v \approx 1500$ (km s^{-1}). This radial bin only contains one LOS quasar-protocluster pair, with all other QSO's within this angular radii range not meeting the necessary redshift and filtering requirements. As a result, any conclusions we may draw from the analysis of this single spectrum will be difficult to interpret, as we can only attribute the feature to a single quasar sight-line. Another critical aspect of this result we must consider is the large δv offset of this absorption. In particular, how this feature relates to the possible sources of errors and physical processes that may cause this offset, if this feature is to be associated to the cluster. The first possible cause of this misalignment are errors in the quasar redshift provided by SDSS, introducing an offset in δv when shifting to the cluster rest frame. Bolton et al. [2012] outline the pipeline redshift determination procedure used by SDSS for the quasars in our sample, giving typical errors for $2.0 < z_{QSO} < 3.0$ normally distributed around ≈ 30 (km s^{-1}). Furthermore, we have no guarantee that the AGN to which the quasars spectra are aligned to are located within the actual centre of the cluster. As a result, further offset due to uncertainty in the AGN position, which could vary over the size of the protocluster, may also be introduced. However, these alone are insufficient to account for the discrepancy in δv.

By far the most probable cause of a large offset such as this is due to peculiar motion within the protocluster. Venemans et al. [2007] characterised the velocity dispersion of radio emitting galaxies within many different protoclusters, finding typical distributions with a mean $\delta v = 0$ (km s^{-1}) as expected when considering the average peculiar motion within the cluster. However within these distributions, it was not uncommon to find member galaxies with $\delta v = \pm 2000$ (km s^{-1}). Hence, as we only have a single QSO spectrum, and thus are not seeing the net combination of absorptions which should average to $\delta v = 0$ (km s^{-1}), we must conclude this discrepancy is acceptable and within the errors of our study.

Radial bin (arcseconds)	N_{QSO}^{LOS}	$N_{QSO}^{Control}$	N_{CARLA}
$R_\perp = [0, 200]$	1	17	1
$R_\perp = [200, 400]$	8	62	8
$R_\perp = [400, 600]$	12	90	11
$R_\perp = [600, 800]$	20	69	18
$R_\perp = [800, 1000]$	30	120	26

Table 3: Table of radially binned composite spectra shown in Figure 19, with quasar and CARLA protocluster data for each radial bin. Also given, for reference, is the number of quasars within the control for each radial bin, with each quantity indicated within the column header.

The radial bin with $R_\perp = [200'', 400'']$ also shows a significant absorption feature, this time only a few hundred km s^{-1} from $\delta v = 0$ and well within the errors previously described. Furthermore, as stated in Table 3 this composite consists of 8 sight-line QSO, greatly improving our confidence in this result being associated to our targets. Comparing this result to that of $R_\perp = [0'', 200'']$ we can see the optical depth from the control IGM absorbtion level (green) is weaker for the $R_\perp = [200'', 400'']$, which can be attributed to the smaller radial bin having only a single quasar. Whereas the $R_\perp = [200'', 400'']$ radial bin is composed of 8 QSO spectra, of which not all may show absorption and so diminish the signal when averaging as outlined previously. However, also present is a second feature at $\delta v \approx 3000$ (km s^{-1}) with a weaker signal than the main feature near $\delta v = 0$. This secondary feature is not an absorption related to another transition as listed in Table 1, and due to its large relative velocity offset, we therefore cannot be certain this is a secondary absorption related to our cluster either. Therefore, the most probable

origin for this signal is random IGM absorption nearby to the protocluster. In §5.2 we attempt to clarify the nature of this feature by increasing the number of LOS quasars in the composite. Lastly, the results for the composites produced at larger radii with $R_\perp$ spanning from 400 to 1000 arcseconds all exhibit no significant absorption features above the control level, aside from the weak narrow absorption present at $\delta v \approx 1500$ in $R_\perp = [400'', 600'']$ median composite. As this feature is narrow it likely stems from only a single spectrum, and thus corresponds to random IGM absorption within an individual QSO spectrum in the stack. Furthermore, due to the composite consisting of 12 sight line quasars, such a large δv is unlikely when considering the average peculiar motion of gas randomly sampled throughout the target protoclusters. Thus, we can ultimately conclude that we detect gas no further than $400''$ radially on the sky from our target.

5.2 S/N Adjusted Composites

In order to confirm the validity of our strongest absorption result for the $R_\perp = [200'', 400'']$ composite, additional analysis was performed for this subset of data with a larger sample of quasars obtained by adjusting the S/N filtering. Increasing the number of sight-line QSO within each composite should increase the signal strength, as well as reduce the amplitude of any random IGM absorption. Thus, by lowering the conservative S/N cutoff set at 4 throughout this study we can increase the number of quasars in $R_\perp = [200'', 400'']$ composite. However, care must be taken in doing this to ensure that the additional lower S/N QSO are not overly noisy and weaken our absorption feature already present.

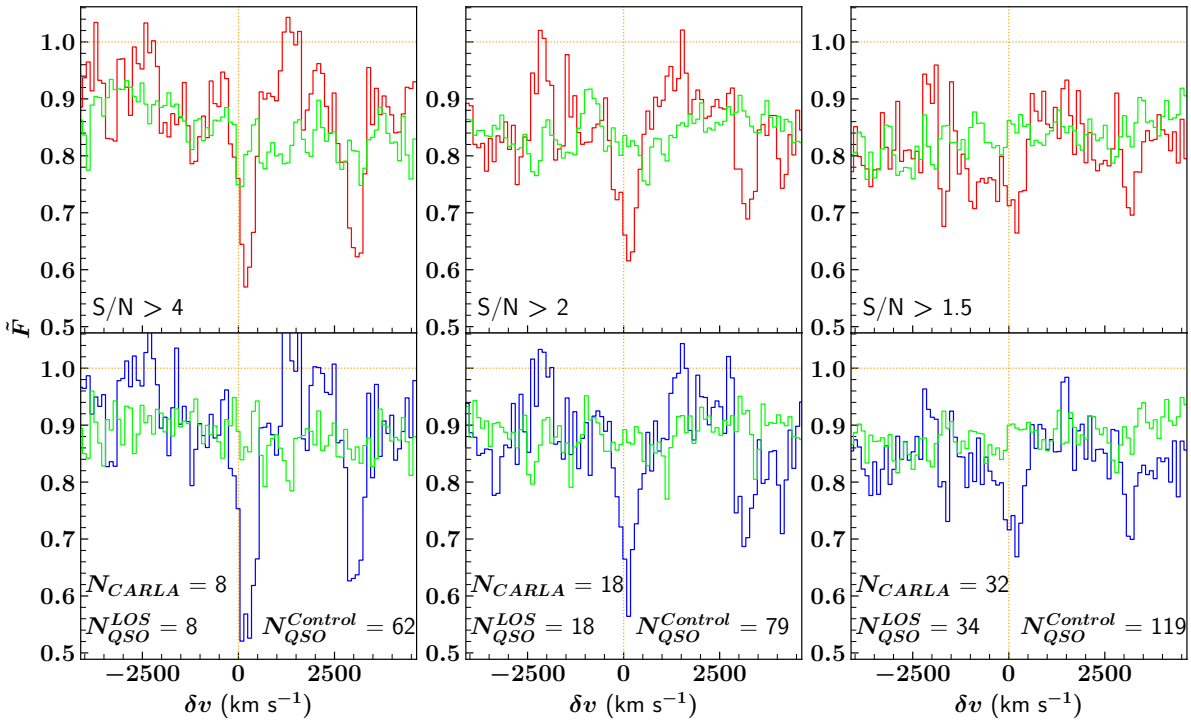

Figure 20: Mean (red) and median (blue) composites for the $R_\perp = [200'', 400'']$ radial bin, showing the resultant impact of stacking additional lower signal to noise QSO spectra. With the S/N > 4 result as previously obtained in Figure 19, compared to the composites formed with additional QSO with S/N > 2 (middle) and S/N > 1.5 (right). Also provided is the control sample (green) calculated for each, with the number of LOS and control quasars also provided, along with the number of CARLA targets sampled by each composite.

Figure 20 outlines the resultant composites produced for three different S/N limits for the $R_\perp = [200'', 400'']$ radial bin. Reducing the S/N to 2 more than doubled the number of LOS quasars to from 8 to 18 within this bin. The net effect of these additional lower S/N quasars slightly decreased the depth of absorption, seen visibly when comparing to the S/N > 4 result. These lower resolution spectra did slightly broaden the feature, centring it slightly more on $\delta v = 0$ (km s^{-1}). Although the amplitude of the signal was slightly weakened, it is important to note that the relative strength of the secondary feature at $\delta v \approx 3000$ (km s^{-1}) decreased more relative to the main absorption. This would indicate that this feature is not related to our targets. However, without having further LOS quasars we are unable to draw any definitive conclusions on this. In an attempt to further categorise this secondary feature, a further S/N cut at 1.5 was made providing an additional 14 quasars. As highlighted in Figure 20 the additional noisy quasar spectra significantly diminish the signal, emphasising the limits of varying S/N to increase the number of quasars. From this we can conclude that fine-tuning of the S/N cutoff can potentially increase signal strength, along with improving our confidence in attributing absorption features to our targets. Hence, with further spectra of sufficient quality we may be able to classify this secondary feature fully, however due to the limitations of our data set we do not have the means to do so.

5.3 Neutral Hydrogen Abundance Measurements

Upon obtaining composite spectra that show clear absorption features in excess of our control sample at $\lambda_{Ly\alpha}^{CARLA} = 1215.67$Å, we can proceed with calculating the neutral Hydrogen abundance as outlined in §2.3. We approximate the absorption features as Gaussian profiles in order to calculate their area, measuring the absorption feature from the control level. We do this in order to omit the affects of IGM absorption from our neutral Hydrogen abundance measurements, with the control providing a good approximation of this. To reduce the effects of any random oscillations in the control, a linear function spanning ± 4000 (km s^{-1}) was fitted to the control level, ensuring an accurate measure for the spectral range around each feature was obtained. This is shown as the pink dotted line in Figure 21, highlighting how the Gaussian profile was determined from the control level. When selecting the correct feature to fit, only those within ± 1500 (km s^{-1}) were taken into consideration Prochaska et al., 2013. As previously discussed, peculiar motion of gas clouds within protoclusters can have relative velocities of up to 2000 (km s^{-1}) [Venemans et al., 2007]. This is the dominant cause of any offset of our feature from rest. Based on this, we applied a conservative limit of ± 1500 (km s^{-1}) for selecting features for analysis, in order to increase certainty that they are related to our targets.

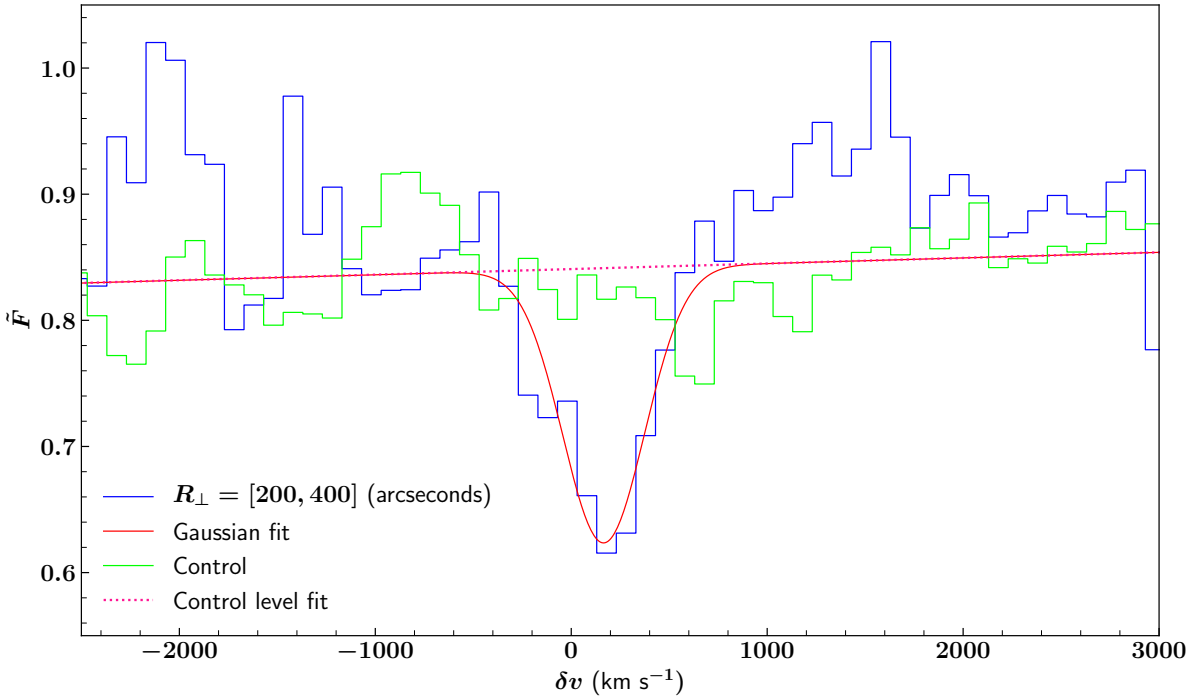

Figure 21: A plot which highlights the Gaussian fitting (red) to the absorption feature (blue) for the $R_\perp = [200'', 400'']$ composite stack. The Gaussian absorption profile was measured from a straight line (dotted pink) fitted to the control level (green) to ensure only absorption related to our target was sampled. From this, the equivalent width and neutral Hydrogen column density N_{HI} was calculated as outlined in Equation 1, providing an estimate for the neutral HI gas abundance in our protocluster sample.

The equivalent width, W_λ, of the absorption feature is calculated according to Equation 1 which finds the area of the Gaussian profile relative to the control level fit. The column density of neutral Hydrogen gas, N_{HI}, is then calculated from this width via Equation 2. To provide limits on the amount of Hydrogen present for null detection's, as we found for the radial bins from $400''$ to $1000''$, we performed the same fitting procedure to the strongest discernible feature within ± 1500 (km s^{-1}) of our CARLA targets. These values were then used to infer an upper limit on what we could detect with our approach. Figure 22 shows the absorption feature fit in each of the $200''$ radial bins. The mean average for all stacks is used when fitting the Gaussian absorption profile to calculate the equivalent width, and will be used solely throughout the remainder of our results to maintain consistency. The error on the equivalent width values were calculated from the 1σ confidence interval of the Gaussian fitting parameters, estimated through a non-linear least squares fitting procedure. These were then used to find the area error of the Gaussian fit, from which an associated error for equivalent width and column density can be calculated using the same equations as outlined previously.

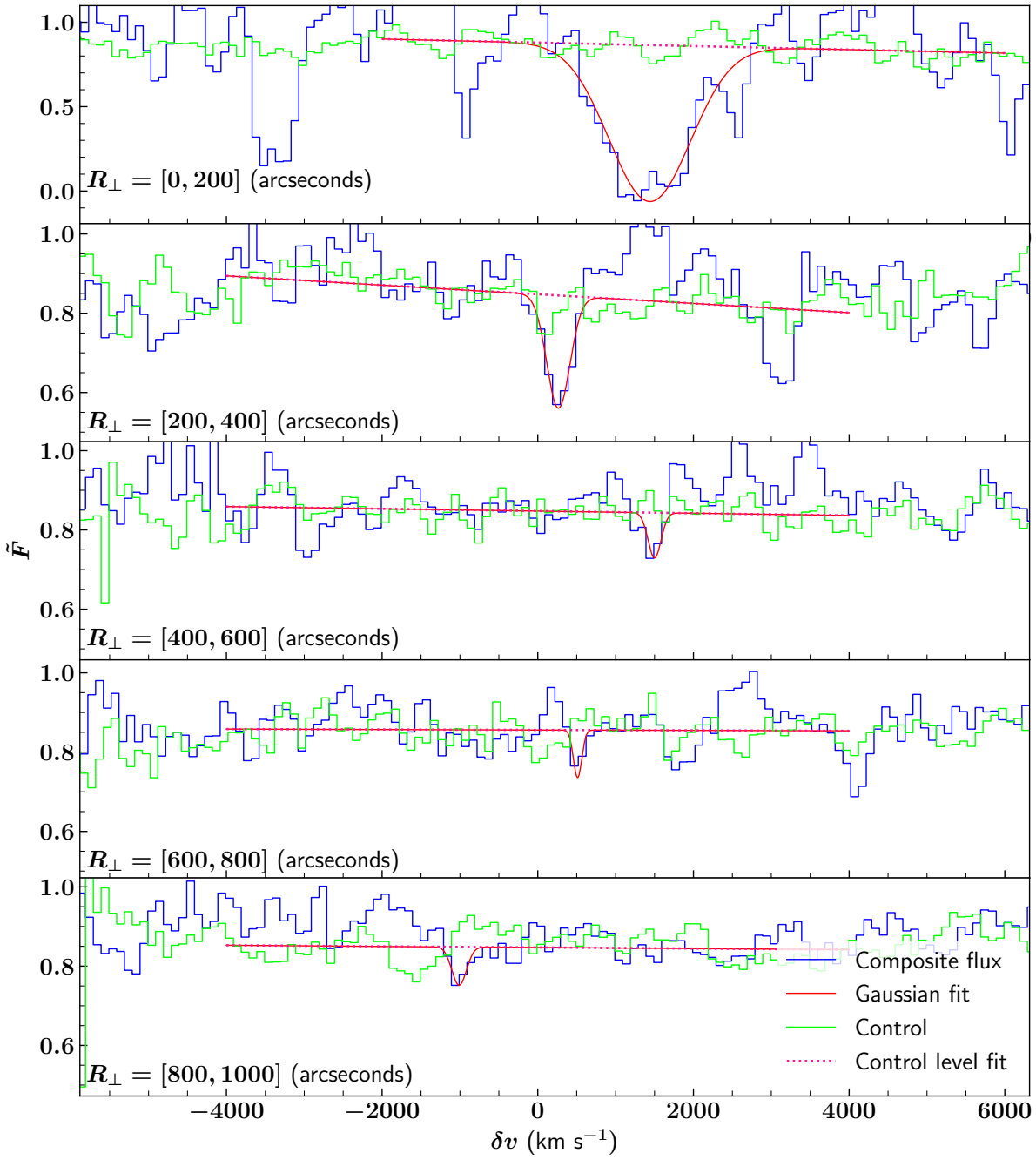

Figure 22: Resultant absorption line fitting for the radially binned composite spectra as previously shown in Figure 19. Shown here is the Gaussian profile and continuum level fits used to quantify the neutral Hydrogen abundance indicated by each feature, as outlined in Figure 21 for the $R_\perp = [200'', 400'']$ composite. Note that for last 3 radial bins from $400''$ to $1000''$, all show no significant absorption attributable to our target. Hence to provide a upper limit on the amount of gas at these radii, fitting was also performed to the most significant feature visible within $\delta v = \pm 1500$ (km s^{-1}) in order to provide these bounding values at these larger radii from our target. The resultant equivalent width values and neutral Hydrogen column densities for each radial bin are given in Table 4.

Radial Bin (arcseconds)	δv_{abs} (km s^{-1})	Amplitude	σ_{fit} (km s^{-1})	$W_{Ly\alpha}$ (Å)	N_{HI} (cm^{-2})
$R_\perp = [0, 200]$	$+1439 \pm 56$	0.93 ± 0.09	524 ± 56	5.33 ± 1.12	$(9.80 \pm 2.06) \times 10^{14}$
$R_\perp = [200, 400]$	$+264 \pm 42$	0.28 ± 0.07	149 ± 42	1.51 ± 0.90	$(2.78 \pm 1.66) \times 10^{14}$
$R_\perp = [400, 600]$	$+1500 \pm 62$	0.11 ± 0.08	83 ± 67	0.85 ± 1.71	$(1.56 \pm 3.15) \times 10^{14}$
$R_\perp = [600, 800]$	$+512 \pm 45$	0.12 ± 0.22	53 ± 112	0.54 ± 4.31	$(9.99 \pm 79.3) \times 10^{13}$
$R_\perp = [800, 1000]$	-1010 ± 59	0.10 ± 0.05	94 ± 60	0.95 ± 1.48	$(1.75 \pm 2.72) \times 10^{14}$

Table 4: Table of results indicating the measured neutral HI gas column density N_{HI} within different angular radii of the central radio-loud source, signposting the location of our protocluster targets. With the three radii from 400″ to 1000″ evidencing the upper limit of the amount of gas we can attribute to our targets at these radii. δv_{abs} is the relative velocity shift of the absorption from rest in the cluster reference frame (mean value of the Gaussian), and $W_{Ly\alpha}$ is the equivalent width of the Lyα absorption feature. Also given are the fitting parameters and errors for each, where σ_{fit} is the standard deviation of the Gaussian fit. Note that the large errors associated to the $W_{Ly\alpha}$ values for 600″ to 1000″ are as a result of fitting a Gaussian to features with little Gaussianity, as visible in Figure 22.

For the stacks separated into 200″ radial bins, the column density of HI gas is reported in Table 4. The data for the three largest radial bins is considerably weaker than the two smallest bins. We posit that the features found in the 0″ to 200″ and 200″ to 400″ bins are the protocluster Lyα absorption detection's we are searching for. The other radial bins have features too similar to noise meaning we cannot discern a signal from the background noise from the IGM. The lack of any detections at larger radii is not unexpected, and is likely due to the average size of protoclusters being within the order of this radius on the sky at this redshift. Hence, finding a signal only within the two innermost radii is in agreement with what we would expect for typical protocluster sizes at $z = 2$ [Venemans et al., 2007].

5.4 Comparative Overdensity Stacking

To further probe our strongest result, we analyse the influence of the overdensity associated to the quasar spectra from the foreground protocluster. The overdensity value allows us to test our detection against varying protocluster environments. A higher overdensity indicates more galaxies within a protocluster which could alter the distribution and gas contents of the intracluster medium. Figure 23 shows the $R_\perp = [200, 400]$ radial bin stack split at the 2σ overdensity cutoff, with a forest signal-to-noise cut made at a value of two. By dividing the spectra within the bin into two separate stacks based on the protocluster overdensity, we can analyse how the strength of the signal varies due to the protocluster environment.

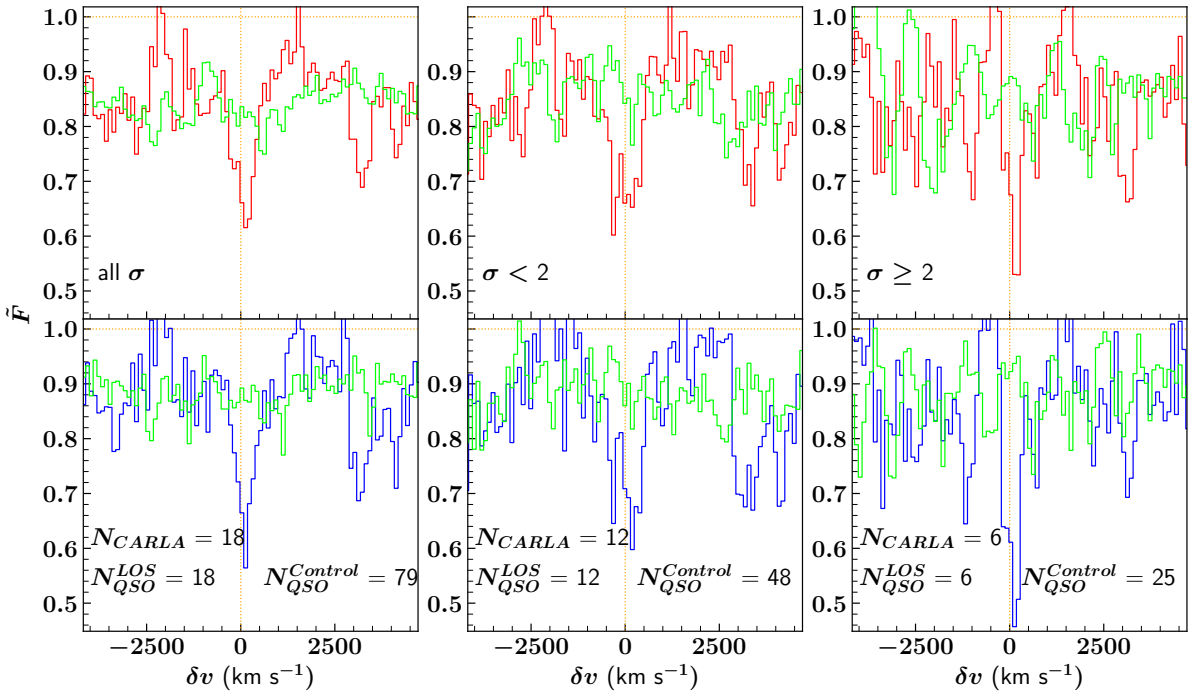

Figure 23: Spectra composites for the $R_\perp = [200, 400]$ (arcseconds) radial bin with S/N > 2 divided into three stacks based on the associated protocluster overdensity. The top and bottom stacks compare two averages, mean (red) and median (blue). The left stacks show the radial bin with no overdensity filtering as shown in Figure 20. The central stacks show only QSO spectra which are situated behind protoclusters with an overdensity value less than two. The stacks on the right show only spectra which are in the background of clusters with an overdensity greater than or equal to two.

The averaged stacks in Figure 23 show how splitting the composite by overdensity affects the signal strength as well as the noise. The $\sigma \geq 2$ stack displays the largest amount of noisy variance around the continuum level in both the LOS and control spectra stacks. This can be explained by the low number of spectra present in the stack which prevents the continuum from averaging out further around the rest frame Lyα line. The $\sigma < 2$ stack in the middle of Figure 23 contains twice the amount of spectra compared to the higher overdensity stack. The variation in QSO number for the overdensity comparison could account for the relative change in noise seen between the two stacks. At such a low total stacked LOS spectra count, this relative noise shift could also originate from more noisy spectra coming through by chance in the higher overdensity bin. This biased spread of stronger and weaker signal spectra between the overdensity bins is potentially an indication of the differing environments in higher density protoclusters, however with such a relatively low amount of data to infer from, a clear conclusion can not be immediately drawn. In order to discern which overdensity environment contains the larger amount of neutral Hydrogen, we must fit the absorption feature as before. The results of this are shown below in Figure 24 and Table 5.

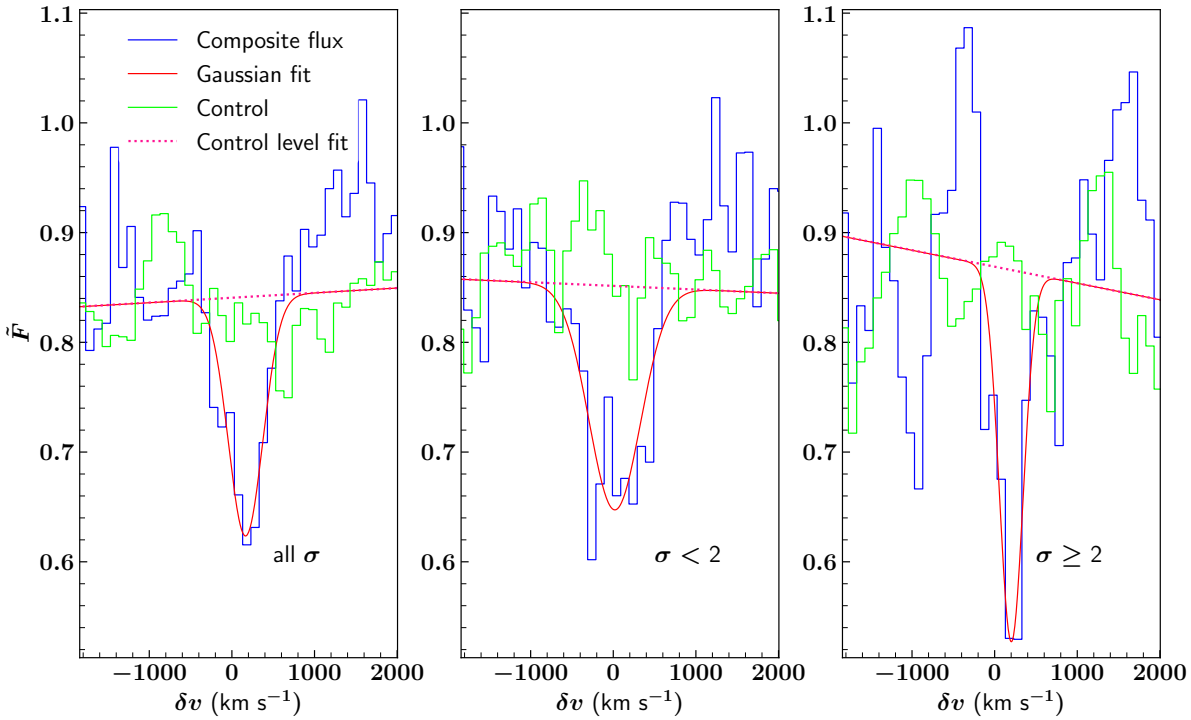

Figure 24: The absorption features from the median average stacks from Figure 23 are fit by a normal distribution in order to calculate their equivalent widths. The higher overdensity stack plotted on the right has a deeper relative flux and Gaussian fit, however it is much narrower than the lower overdensity stack plotted in the centre.

The spectral absorption line in the central stack for $\sigma < 2$ has a wider dip in normalised flux at δv_{abs} implying a greater amount of absorption of neutral Hydrogen occurs in protoclusters that contain a lower density of galaxies. Our fitting analysis shows this to be true. The column density of HI gas increases with respect to lower protocluster overdensity, as shown in Table 5. A relative column density increase of $(3.15 \pm 3.27) \times 10^{14}\text{cm}^{-2}$ is seen when moving from the sample of $\sigma \geq 2$ protoclusters to the $\sigma < 2$ overdense ones.

Overdensity Range	δv_{abs} (km s^{-1})	Amplitude	σ_{fit} (km s^{-1})	$W_{Ly\alpha}$ (Å)	N_{HI} (cm^{-2})
all σ	165.7 ± 49	0.22 ± 0.04	208 ± 49	2.12 ± 1.02	$(3.90 \pm 1.88) \times 10^{14}$
$\sigma < 2$	18 ± 75	0.20 ± 0.04	312 ± 75	3.17 ± 1.58	$(5.83 \pm 2.90) \times 10^{14}$
$\sigma \geq 2$	202 ± 39	0.34 ± 0.08	143 ± 39	1.46 ± 0.82	$(2.68 \pm 1.52) \times 10^{14}$

Table 5: Tabulated data quantifying the absorption profiles for the stacks split by protocluster overdensity, plotted in Figures 23 and 24. The relative velocity at the central region of the absorption, δv_{abs} (mean of the Gaussian), is included along with the equivalent width and column density measured from the fitting calculations as outlined previously. The fitting parameters and associated errors are also given as outlined in Table 4.

The presence of more absorption in less overdense protoclusters could be due to less ionisation of the neutral Hydrogen gas because of a lower presence of radiation emitting galaxies within these clusters. For this specific set of criteria, it is hard to interpret our result with confidence due to the low sample size of quasar spectra and the relatively noisy distribution for more overdense protoclusters when directly compared to the less overdense stack.

6 Discussion of Results

With the initial aim of investigating protoclusters at high redshifts, in this project we have developed and tested our own quasar probing technique. We have successfully quantified the abundance of neutral Hydrogen gas in these structures, giving us the ability to draw a range of conclusions from our results. Firstly, we confirmed the proficiency of our continuum fitting algorithm, with the initial test shown in Figure 17. Ideally each quasar spectrum would have been visually inspected to examine the resultant continuum fit, however due to the tens of thousands of spectra within our LOS and control samples this would not be practical. Instead, the first test of our method was performed in parallel to a control produced by K.-G. Lee et al. [2013], which uses principal component analysis and a mean-flux regulation technique to yield continuum estimates in strong agreement to data provided by Faucher-Giguère et al. [2008].

With results from K.-G. Lee et al. [2013] used in many other similar studies to our own [Cai et al., 2017; Mukae et al., 2019; Prochaska et al., 2013], conformity to their data as a control gave us confidence in our procedure within the context of our study. The only minor differences in the results produced by our method were an increase in noise in the blue end of our spectra, as well as a slightly weaker absorption detections. K.-G. Lee et al. [2013] implemented a set of additional processes to mitigate these effects such as removing sky noise, masking of known bad pixels, as well as removing Damped Lyα (DLA) absorption features from each spectra. All of these processes could have been implemented into our algorithm, however, no issues stemming from any of the sources of error could be found within our results. Moreover, with alternative solutions to this problem proposed with a similar operating principal to our own [Meyer et al., 2019], we can take confidence in the results produced by own method.

Many of the challenges we faced were brought about by the natural limitations of the quasars spectral data. Solely using SDSS data meant that for the most part, we had to regularly use lower resolution and S/N spectra for our analysis. Routinely using spectra with S/N > 2 within our spectral region of interest, a factor of 4 lower than the similar limit imposed by Prochaska et al. [2013]. This meant many potential features within our quasar sample were simply unusable, reducing our LOS data set by almost a half (see Figure 8). To bypass this, many other studies simply used the SDSS data as a means to identify potential LOS absorption systems due to the large survey size. These studies then performed follow up analysis with high precision spectroscopic surveys which used larger ground based telescopes such as Keck, the Giant Magellan Telescope, Gemini, and the Large Binocular Telescope to retrieve high resolution spectra for absorption line analysis [Cai et al., 2017; Mukae et al., 2019; Prochaska et al., 2013; Shull, 1995]. Similar measures would further our own study, providing vast improvements on the low resolution and S/N limitations of our own data.

To help address most of these inherent limitations in our background sight-line quasar data set, composite spectra analysis helped improve our confidence in the results we collected when utilised in the correct manner. Our strongest result obtained for the composite spectra was produced for QSO with a radial separation between $R_{\perp} = [200'', 400'']$ from the central CARLA targets. This result highlights the huge benefit of this technique when working with low resolution, noisy spectra. The technique reduces the strength of noise and random IGM absorption that is unrelated to our protocluster targets, that would otherwise make identification of the target absorption impossible. With this made clear by examining the individual uncombined spectra as evidenced for reference within the appendix in Figure 25. This process also decreased the other sources of uncertainty via averaging, such as discrepancies in quasar redshift, and the peculiar motion of gas within the cluster itself as outlined within the results.

The improvements to the quality of our detection's, which were made due to stacking the quasar spectra together, are clearly seen when comparing between each S/N composite produced for the $R_\perp = [200'', 400'']$ radial bin. This analysis showed the averaging of various absorption features when comparing the S/N > 4 composite to that of S/N > 2, with the absorption in the latter broadening and centring further about $\delta v = 0$. We can further see this effect in the equivalent width values calculated for each, within Tables 4 and 5 respectively. With the S/N $>$ 4 having a smaller $W_{Ly\alpha} = 1.51$ (Å) compared to $W_{Ly\alpha} = 2.12$ (Å) for S/N > 2. This variance in the amount of gas measured highlights the importance of having the largest possible sample of LOS quasars to provide the best possible assessment of the target absorption. We were limited by this on multiple occasions, with this being one of many such examples. The $R_\perp = [0'', 200'']$ radial bin only contained a single usable quasar spectrum for analysis, severely reducing our ability to confidently attribute the strong absorption seen to the protocluster targeted.

Despite these few caveats, on the whole we were able to generate measurements and draw conclusions from our stacking results. The radial binning study, looking at potentially the densest inner most region of protoclusters for $R_\perp \leq 1000''$ ($\approx$ 7Mpc) scale indicated the presence of gas for the innermost $400''$ ($\approx$ 2.8Mpc). We found no such signals further out than this, as shown in Figures 19 and 22, imposing a size limit on the scale of the absorbing neutral gas content of the cluster. This region corresponds to $\approx$ 2.8Mpc radially from the central CARLA AGN. Comparing this to the size and scale of protoclusters studied by Venemans et al. [2007] and Diener et al. [2015] we find this to agree with the typical scales for protoclusters at $z = 2$. Confidence on this result would be improved by having a larger QSO sample, especially for the underpopulated inner most radial bin containing only one analysable quasar. This result does, however, illustrate the possibility of bounding the size, structure and density through this radial binning process. With more quasars, the size and therefore precision of this approach could be vastly improved to outline finer detail within each protocluster [Mukae et al., 2019].

The values found for the neutral Hydrogen column density in protoclusters in this project are higher than those for quasars found by Prochaska et al. [2013]. This suggests there is more HI gas within the intracluster medium of protoclusters compared to the local environment around quasi-stellar objects. This is interesting to note as these clouds of gas are expected to be abundant around quasars, but the quantity within a protoclusters ICM is less well-defined. The technique outlined in this report is important for this reason, as it provides a deep field probing tool without the need for expensive long exposure space based observations.

The accuracy of our Gaussian fitting method to calculate the equivalent width of spectral lines, and their related column density, could have been improved by using a more sophisticated Voigt profile calculation as outlined by Liang and Kravtsov [2017]. However, this technique is beyond the required scope of this study. Also noteworthy is the method used to obtain errors on our $W_{Ly\alpha}$ and column density measurements. Comparatively, we could have followed the error analysis method by used Prochaska et al. [2013], who fitted bootstrap realisations in order obtain errors on their measured equivalent width values. The accuracy improvement to be gained from using these alternate techniques are minimal, especially when compared to other improvements which could be made to this study to reduce errors, such as including a larger sample of higher resolution quasar spectra. To achieve such an upgrade, we suggest the methodology outlined in this report be extended to focus on a single known protocluster target. Picking a specific cluster target with a large sample of high resolution QSO spectra, such as those outlined by Shull [1995]. These could then be used to gather specific information on a single CARLA target, for example. Currently our method focuses on averaged results for many protoclusters, but performing enough spectroscopy on quasar targets in more specific fields could further strengthen the cluster parameters found in this study.

An important feature of our final column density results was the connection made between protocluster overdensity and neutral Hydrogen abundance. The overdensity results in table 5 show we find a weaker signal for HI gas in protoclusters with a denser amount of galaxies within them. This is potentially explained by galactic radiation ionising the intracluster medium, and therefore reducing the abundance of neutral Hydrogen seen within the cluster [Villaescusa-Navarro et al., 2016]. However, we interpret our result as an indication rather than a solid conformation of this phenomenon. This is due to the low sample size available when hunting in this very specific environment compared to the generic absorption searches we perform otherwise. There is still a chance that with more high resolution spectra behind the more overdense protoclusters, we would find a stronger absorption and our explanation would be moot. As stated before, more confident results from using the method in this study can only come about with larger amounts of high S/N spectroscopic quasar data.

To further our investigation of the CARLA targets using our method, we would require many more usable quasar spectra. This limitation hampered our ability throughout the report to study any single particular CARLA target. However, this is not a limitation of our technique but a constraint imposed by our restricted single SDSS quasar data set. Despite this, we have still effectively shown that quasars can probe the environments of potential CARLA protocluster targets, through a broader investigation of their characteristics as a group. This study has yielded valuable insights about their size and composition through their neutral Hydrogen gas content. Simply having more or higher quality background quasar spectra for each target would further our ability to classify these CARLA targets, without the need for follow up observations with more advanced telescopes [Noirot et al., 2018]. Furthermore, with possible future SDSS data releases providing additional quasar targets, the quasars probing galaxy (proto)clusters technique has the potential to become a invaluable tool in future.

7 Conclusions

Ultimately, we have successfully developed and applied our own quasar probing technique to study protoclusters at relatively high redshifts. We detect Lyα absorption due to neutral Hydrogen gas clouds within the intracluster medium of $z > 2$ protoclusters. The strongest of these detections is radially within approximately 2.8Mpc of the protoclusters central radio galaxy, which equates to a radius of 400 seconds of arc on the sky at this redshift. The column density of the HI clouds we measure within this strong detection range is calculated to be $(2.78 \pm 1.66) \times 10^{14}\text{cm}^{-2}$, specifically for the $R_{\perp} = [200'', 400'']$ radial bin. We found a potential correlation between the foreground protocluster overdensity and the column density of HI within the intracluster medium, due to the ionisation of neutral Hydrogen from galaxies which are members of the protoclusters. The column density increases by $(3.15 \pm 3.27) \times 10^{14}\text{cm}^{-2}$ when changing from a sample of protoclusters with an overdensity value greater than two to those with a value less than two. Although some uncertainties remain in our findings, mainly brought about by the inherent limitations of the data sample used within this study, we are confident in the Lyα absorption detections we have made. The aim of probing potential protocluster targets found by the CARLA survey in order to reveal their size and quantify the abundance of neutral Hydrogen gas clouds within them has been achieved.

References

Abolfathi, B., & Aguado, D. (2018). The fourteenth data release of the Sloan Digital Sky Survey: First spectroscopic data from the extended baryon oscillation spectroscopic survey and from the second phase of the apache point observatory galactic evolution experiment, *235*(2), arxiv 1707.09322, 42. https://doi.org/10.3847/1538-4365/aa9e8a

Becker, R. H., Fan, X., White, R. L., Strauss, M. A., Narayanan, V. K., Lupton, R. H., Gunn, J. E., Annis, J., Bahcall, N. A., Brinkmann, J., Connolly, A. J., Csabai, I., Czarapata, P. C., Doi, M., Heckman, T. M., Hennessy, G. S., Ivezic, Z., Knapp, G. R., Lamb, D. Q., ... York, D. G. (2001). Evidence for Reionization at z approximately 6: Detection of a Gunn-Peterson Trough in a z=6.28 Quasar. *The Astronomical Journal*, *122*(6), arxiv astro-ph/0108097, 2850–2857. https://doi.org/10.1086/324231

Bolton, A. S., Schlegel, D. J., Aubourg, E., Bailey, S., Bhardwaj, V., Brownstein, J. R., Burles, S., Chen, Y.-M., Dawson, K., Eisenstein, D. J., Gunn, J. E., Knapp, G. R., Loomis, C. P., Lupton, R. H., Maraston, C., Muna, D., Myers, A. D., Olmstead, M. D., Padmanabhan, N., ... Wood-Vasey, W. M. (2012). Spectral Classification and Redshift Measurement for the SDSS-III Baryon Oscillation Spectroscopic Survey. *The Astronomical Journal*, *144*(5), arxiv 1207.7326, 144. https://doi.org/10.1088/0004-6256/144/5/144

Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., & Lemson, G. (2009). Resolving cosmic structure formation with the Millennium-II Simulation, *398*(3), arxiv 0903.3041, 1150–1164. https://doi.org/10.1111/j.1365-2966.2009.15191.x

Cai, Z., Fan, X., Bian, F., Zabludoff, A., Yang, Y., Prochaska, J. X., McGreer, I., Zheng, Z.-Y., Kashikawa, N., Wang, R., Frye, B., Green, R., & Jiang, L. (2017). Mapping the most massive overdensities through hydrogen (MAMMOTH). II. Discovery of the extremely massive overdensity BOSS1441 at z = 2.32, *839*(2), arxiv 1609.02913, 131. https://doi.org/10.3847/1538-4357/aa6a1a

Chiang, Y.-K., Overzier, R. A., Gebhardt, K., & Henriques, B. (2017). Galaxy Protoclusters as Drivers of Cosmic Star Formation History in the First 2 Gyr. *The Astrophysical Journal*, *844*(2), L23. https://doi.org/10.3847/2041-8213/aa7e7b

Chiang, Y.-K., Overzier, R., & Gebhardt, K. (2013). Ancient light from young cosmic cities: Physical and observational signatures of galaxy proto-clusters, *779*(2), arxiv 1310.2938, 127. https://doi.org/10.1088/0004-637X/779/2/127

Diener, C., Lilly, S. J., Ledoux, C., Zamorani, G., Bolzonella, M., Murphy, D. N. A., Capak, P., Ilbert, O., & McCracken, H. (2015). A PROTOCLUSTER AT z = 2.45. *The Astrophysical Journal*, *802*(1), 31. https://doi.org/10.1088/0004-637X/802/1/31

Faucher-Giguère, C.-A., Prochaska, J. X., Lidz, A., Hernquist, L., & Zaldarriaga, M. (2008). A Direct Precision Measurement of the Intergalactic Lyα Opacity at 2 <= z <= 4.2. *The Astrophysical Journal*, *681*, 831–855. https://doi.org/10.1086/588648

Galametz, A., Stern, D., De Breuck, C., Hatch, N., Mayo, J., Miley, G., ro Rettura, A., Seymour, N., Stanford, S. A., & Vernet, J. (2012). The mid-infrared environments of high-redshift radio galaxies, *749*(2), arxiv 1202.4489, 169. https://doi.org/10.1088/0004-637X/749/2/169

Gnedin, N. Y. (1998). Probing the universe with the Lyman-alpha forest: II. The column density distribution. *Monthly Notices of the Royal Astronomical Society*, *299*(2), arxiv astro-ph/9706286, 392–402. https://doi.org/10.1046/j.1365-8711.1998.01755.x

Hennawi, J. F., Prochaska, J. X., Burles, S., Strauss, M. A., Richards, G. T., Schlegel, D. J., Fan, X., Schneider, D. P., Zakamska, N. L., Oguri, M., Gunn, J. E., Lupton, R. H., Brinkmann, J., & Brunner, R. J. (2006). Quasars Probing Quasars I: Optically Thick Absorbers Near Luminous Quasars. *The Astrophysical Journal*, *651*(1), arxiv astro-ph/0603742, 61–83. https://doi.org/10.1086/507069

Lee, H.-W. (2013). Asymmetric Absorption Profiles of Ly$alpha$ and Lyβ in Damped Lyman Alpha Systems. *The Astrophysical Journal*, *772*(2), arxiv 1306.0181, 123. https://doi.org/10.1088/0004-637X/772/2/123

Lee, K.-G., Bailey, S., Bartsch, L. E., Carithers, W., Dawson, K. S., Kirkby, D., Lundgren, B., Margala, D., Palanque-Delabrouille, N., Pieri, M. M., Schlegel, D. J., Weinberg, D. H., Yeche, C., Aubourg, E., Bautista, J., Bizyaev, D., Blomqvist, M., Bolton, A. S., Borde, A., ... Weaver, B. A. (2013). The BOSS Lyman-alpha Forest Sample from SDSS Data Release 9. *The Astronomical Journal*, *145*(3), arxiv 1211.5146, 69. https://doi.org/10.1088/0004-6256/145/3/69

Liang, C., & Kravtsov, A. (2017). BayesVP: A Bayesian Voigt profile fitting package. *arXiv:1710.09852 [astro-ph]*, arxiv 1710.09852.

Mantz, A. B., Abdulla, Z., Carlstrom, J. E., Greer, C. H., Leitch, E. M., Marrone, D. P., Muchovej, S., Adami, C., Birkinshaw, M., Bremer, M., Clerc, N., Giles, P., Horellou, C., Maughan, B., Pacaud, F., Pierre, M., & Willis, J. (2014). The XXL survey. V. Detection of the sunyaev-zel'dovich effect of the redshift 1.9 galaxy cluster XLSSU J021744.1-034536 with CARMA, *794*(2), arxiv 1401.2087, 157. https://doi.org/10.1088/0004-637X/794/2/157

Meyer, R. A., Bosman, S. E. I., Kakiichi, K., & Ellis, R. S. (2019). The role of galaxies and AGN in reionizing the IGM - II: Metal-tracing the faint sources of reionization at $5\lesssim z\lesssim6$. *Monthly Notices of the Royal Astronomical Society*, *483*(1), arxiv 1807.07899, 19–37. https://doi.org/10.1093/mnras/sty2954

Miller, J. S. A., Bolton, J. S., & Hatch, N. (2019). Searching for the shadows of giants: Characterizing protoclusters with line of sight Lyman-α absorption, *489*(4), arxiv 1909.02513, 5381–5397. https://doi.org/10.1093/mnras/stz2504

Mukae, S., Ouchi, M., Cai, Z., Lee, K.-G., Prochaska, J. X., Cantalupo, S., Zheng, Z., Nagamine, K., Suzuki, N., Silverman, J. D., Misawa, T., Inoue, A. K., Hennawi, J. F., Matsuda, Y., Mawatari, K., Sugahara, Y., Kojima, T., Ono, Y., Shibuya, T., ... Kakuma, R. (2019). 3D Distribution Map of HI Gas and Galaxies Around an Enormous Lyα Nebula and Three QSOs at $z=2.3$ Revealed by the HI Tomographic Mapping Technique. *arXiv e-prints*, *1910*, arXiv:1910.02962.

Muldrew, S. I., Hatch, N. A., & Cooke, E. A. (2015). What are Protoclusters? – Defining High Redshift Galaxy Clusters and Protoclusters. *Monthly Notices of the Royal Astronomical Society*, *452*(3), arxiv 1506.08835, 2528–2539. https://doi.org/10.1093/mnras/stv1449

Noirot, G., Stern, D., Mei, S., Wylezalek, D., Cooke, E. A., De Breuck, C., Galametz, A., Hatch, N. A., Vernet, J., Brodwin, M., Eisenhardt, P., Gonzalez, A. H., Jarvis, M., Rettura, A., Seymour, N., & Stanford, S. A. (2018). HST Grism Confirmation of 16 Structures at $1.4 < z < 2.8$ from the Clusters Around Radio-Loud AGN (CARLA) Survey. *The Astrophysical Journal*, *859*, 38. https://doi.org/10.3847/1538-4357/aabadb

Overzier, R. A. (2016). The realm of the galaxy protoclusters. A review, *24*(1), arxiv 1610.05201, 14. https://doi.org/10.1007/s00159-016-0100-3

Papovich, C. (2008). The angular clustering of distant galaxy clusters, *676*(1), arxiv 0712.1819, 206–217. https://doi.org/10.1086/527665

Peebles, P. J. E. (1993). Principles of Physical Cosmology. *Principles of Physical Cosmology by P.J.E. Peebles. Princeton University Press, 1993. ISBN: 978-0-691-01933-8.*

Prochaska, J. X., Hennawi, J. F., Lee, K.-G., Cantalupo, S., Bovy, J., Djorgovski, S. G., Ellison, S. L., Lau, M. W., Martin, C. L., Myers, A., Rubin, K. H. R., & Simcoe, R. A. (2013). Quasars Probing Quasars VI. Excess HI Absorption Within One Proper Mpc of z˜2 Quasars. *The Astrophysical Journal*, *776*(2), arxiv 1308.6222, 136. https://doi.org/10.1088/0004-637X/776/2/136

Rauch, M. (1998). The lyman alpha forest in the spectra of QSOs, *36*arxiv astro-ph/9806286, 267–316. https://doi.org/10.1146/annurev.astro.36.1.267

Rosati, P. [P.], Tozzi, P., Ettori, S., Mainieri, V., Demarco, R., Stanford, S. A., Lidman, C., Nonino, M., Borgani, S., Della Ceca, R., Eisenhardt, P., Holden, B. P., & Norman, C. (2004). Chandra and XMM-Newton observations of RDCS 1252.9-2927, a massive cluster at z=1.24, *127*(1), arxiv astro-ph/0309546, 230–238. https://doi.org/10.1086/379857

Rosati, P. [Piero], Borgani, S., & Norman, C. (2002). The evolution of x-ray clusters of galaxies, *40*arxiv astro-ph/0209035, 539–577. https://doi.org/10.1146/annurev.astro.40.120401.150547

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, *36*(8), 1627–1639. https://doi.org/10.1021/ac60214a047

Shull, J. M. (1995). High-Resolution Spectroscopy of Quasars and Quasar Absorption-Line Systems. *Publications of the Astronomical Society of the Pacific*, *107*, 1007. https://doi.org/10.1086/133653

Venemans, B. P., Röttgering, H. J. A., Miley, G. K., van Breugel, W. J. M., De Breuck, C., Kurk, J. D., Pentericci, L., Stanford, S. A., Overzier, R. A., Croft, S., & Ford, H. (2007). Protoclusters associated with $z > 2$ radio galaxies: I. Characteristics of high redshift protoclusters. *Astronomy & Astrophysics*, *461*(3), 823–845. https://doi.org/10.1051/0004-6361:20053941

Villaescusa-Navarro, F., Planelles, S., Borgani, S., Viel, M., Rasia, E., Murante, G., Dolag, K., Steinborn, L. K., Biffi, V., Beck, A. M., & Ragone-Figueroa, C. (2016). Neutral hydrogen in galaxy clusters: Impact of AGN feedback and implications for intensity mapping. *Monthly Notices of the Royal Astronomical Society*, *456*(4), arxiv 1510.04277, 3553–3570. https://doi.org/10.1093/mnras/stv2904

Wright, E. L. (2006). A Cosmology Calculator for the World Wide Web. *Publications of the Astronomical Society of the Pacific*, *118*, 1711–1715. https://doi.org/10.1086/510102

Wylezalek, D., Galametz, A., Stern, D., Vernet, J., De Breuck, C., Seymour, N., Brodwin, M., Eisenhardt, P. M., Gonzalez, A. H., Hatch, N., Jarvis, M., Rettura, A., Stanford, S. A., & Stevens, J. A. (2013). Galaxy Clusters around radio-loud AGN at $1.3 < z < 3.2$ as seen by Spitzer. *The Astrophysical Journal*, *769*(1), arxiv 1304.0770, 79. https://doi.org/10.1088/0004-637X/769/1/79

Yamada, T., Nakamura, Y., Matsuda, Y., Hayashino, T., Yamauchi, R., Morimoto, N., Kousai, K., & Umemura, M. (2012). PANORAMIC SURVEY OF Lyα EMITTERS AT $z = 3.1$. *The Astronomical Journal*, *143*(4), 79. https://doi.org/10.1088/0004-6256/143/4/79

Appendix

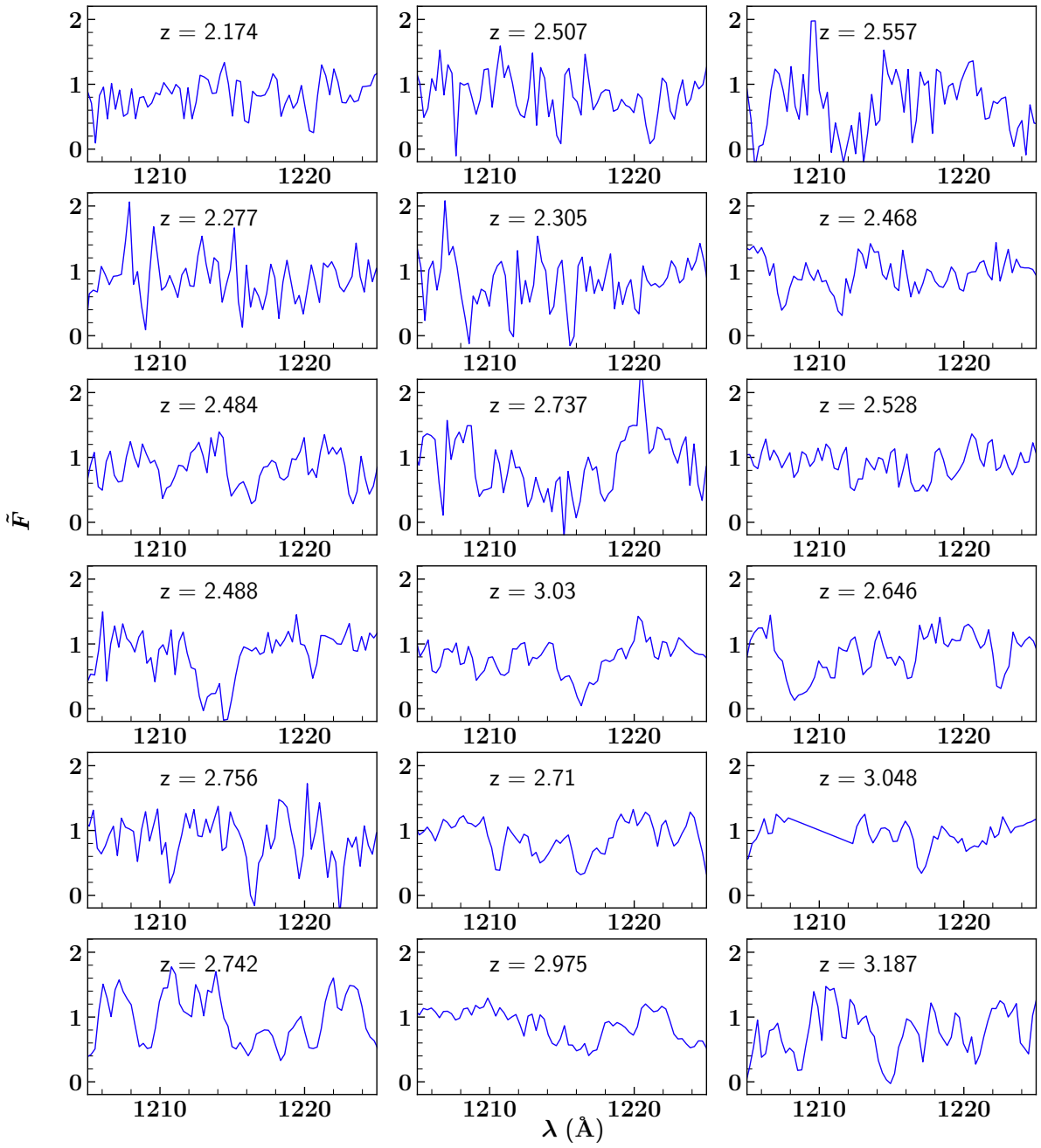

Figure 25: For reference, the continuum normalised spectra shifted to the corresponding LOS CARLA rest frame for our strongest absorption detection in the $R_\perp = [200'', 400'']$ radial bin (S/N > 2), shown in Figure 20 are provided here as an example to highlight the data used to form the composite stacks used throughout the report. The quasar redshift is also given. Each plot shows the relative flux $\tilde{F}$ against wavelength λ (Å).