Column Density Analysis of Fe II Absorption Lines in the Circumgalactic Medium

Jack Higginson

June 16, 2023

Abstract

In this study, we focus on the column densities of Fe II absorption lines in the CGM. This report represents the first systematic investigation into FeII absorbers at high redshifts. Using spectra from a catalogue of 101 quasars between z = 3.55 and z = 7.09, we find 94 detections of FeII. With FeII, column densities range from approximately log(12.64) to log(15.60). We also find that column density is roughly constant from z = 2.7 to z = 4.2, with a slight downward trend from z = 4.2 to z = 6.2. The methodology adopted in this study involves the application of advanced statistical methods on absorption spectra from quasars. The column densities of Fe II absorbers are estimated using Voigt profile fitting. Due to the complex and non-linear nature of the Voigt profile, a Markov Chain Monte Carlo (MCMC) technique is utilized to sample the posterior distributions of the column density (N), Doppler parameter (b), and velocity centroid (v). Furthermore, extensive MCMC diagnostics, including trace plots, auto-correlation plots, and corner plots, are used to assess chain convergence and the quality of the inferred parameter values. We take extra time to do cover the theory and physical motivations for doing Voigt profile fitting and MCMC simulations.

1 Introduction

The Circumgalactic Medium (CGM) is a vital, yet complex, component of a galaxy that bridges the gap between the interstellar medium (ISM) within galaxies and the intergalactic medium (IGM) that permeates the vast spaces between them. Encompassing the region extending out to the virial radius of a galaxy, the CGM is composed of diffuse gas, and is thought to play a crucial role in the cycle of matter in and out of galaxies, affecting star formation, galaxy evolution, and the structure of the cosmic web. Understanding the structure, composition, and kinematics of the CGM is therefore of paramount importance to astrophysics, yet it remains one of the least explored and understood areas due to its diffuse nature and the technical challenges of observing it. Absorption-line spectroscopy of background sources, such as quasars, provides a powerful tool to probe the physical conditions and chemical composition of the CGM.

In this context, the resonance absorption lines of singly ionized iron (Fe II) and magnesium (Mg II) are particularly interesting. These species are commonly observed in absorption-line studies of the CGM and can serve as tracers of gas enriched by supernova ejecta. Their presence helps determine the relationship between elements produced in the galactic disc and that are being more immediately recycled, in contrast to material that is in-falling from the ISM.

The column density, which quantifies the amount of a particular species along the line of sight, is a key parameter derived from these absorption lines. It provides insight into the density and extent of the gas, and combined with measurements of different species, can shed light on the ionization state, metallicity, and physical conditions of the CGM. In the case of multiple-component absorption systems, the distribution of column densities can provide valuable information about the kinematic structure of the absorbing gas. However, accurately measuring column densities, especially in the presence of line saturation and blending, requires careful analysis and often involves the use of Voigt profile fitting. Furthermore, interpreting the column densities requires understanding of the atomic and radiative processes involved and the broader astrophysical context.

In this report, we present a study of the column densities of Fe II absorption lines in the CGM, using high-resolution quasar spectra. Our aim is to contribute to the understanding of the physical conditions, chemical composition, and kinematics of the CGM, and ultimately, its role in galaxy formation and evolution. We begin in section two by giving a robust overview of the definitions, equations, and physics that fuels this line of enquiry. In section three, we discuss how Voigt profile fitting is done. This includes a brief overview of Markov Chain Monte Carlo (MCMC) simulations, Bayesian inference as well as how to know when a fit is of high quality. Finally, we provide the results of Voigt profile fitting and column density estimates for Fe II for a survey of 101 high redshift QSO's.

2 Methods for calculating Column Density

This is an example of an absorber detection:

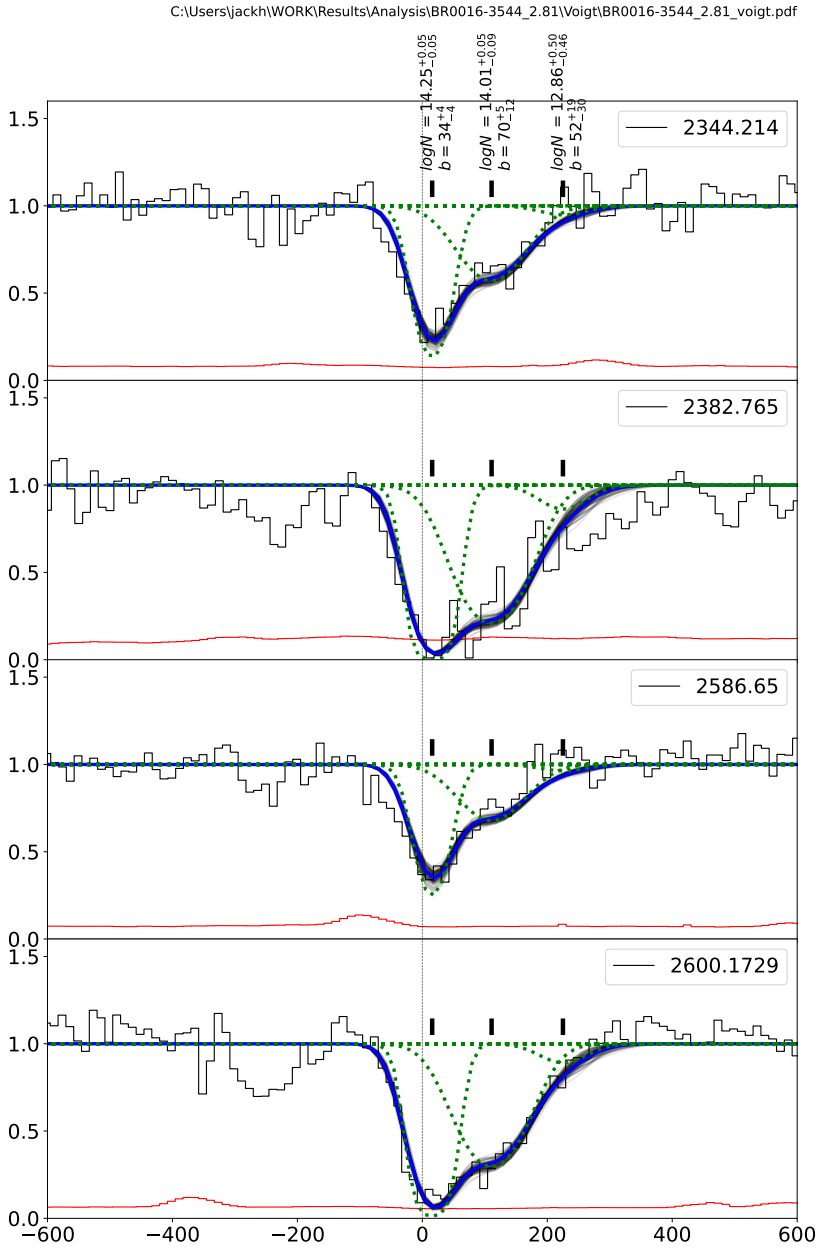

Figure 1: Example spectrum showing an absorption detection for FeII at 2600A

The appropriate fit for spectral lines is the Voigt profile. The Voigt function is a convolution of a Gaussian function (which describes thermal broadening) and a Lorentzian function (which describes natural or collisional broadening).

For the detection shown in the figure above, we can see that it there are multiple components; several Voigt profiles would have to be fit to accurately represent the curve. We can interpret this is there being several "clouds" of the detected element along the path. The velocity (see x-axis) that the Voigt profiles are fit to tells us the magnitude of the velocity component in the direction of our line of sight of the clouds. As we will see later on, there are many other details about these clouds that we can infer.

When trying to fit the Voigt profile to the spectrum, there are three parameters to vary:

- Velocity (v): This determines the center of the Voigt profile. It is called velocity because we usually work in velocity space, with the zero velocity defined to be around the wavelength of the species of interest (for example 2600.2A would be the zero velocity for FeII,2600).
- b-parameter or Doppler parameter (b): This determined the width of the Voigt profile. It takes into account both the thermal and non-thermal motions (see equation below).
- Column density (N). This determines the depth of the Voigt fit (up to a certain point, the intricacies of this are discussed later in this section). It is related to the optical depth by $\tau_\lambda = N * f(\lambda)$ where $f(\lambda)$ is the Voigt profile.

The equation for the b-paramter is,

$$b = \sqrt{2kT/m + b_{nt}} \tag{1}$$

Where k s the Boltzmann constant, T is the temperature of the gas, m is the mass of the absorbing atom or ion, and b_{nt} is the non-thermal velocity dispersion, it is caused by turbulence.

2.1 Optical Depth and Column Density

Optical depth is a measure of the fraction of light of a certain wavelength that passes through a gas. Optical depth at wavelength λ is defined as

$$\tau_\lambda = \int_0^L a_\lambda n \, dx \tag{2}$$

Where we are integrating over a length L, a_λ is the line absorption coefficient describing the absorption of photons by bound-bound atomic transitions, and n is the volume density of the absorbing atoms.

Following this thought, the change in intensity of light I_λ passing through a distance dx of the gas would be,

$$dI_\lambda = -a_\lambda n I_\lambda dx \tag{3}$$

Comparing these two equations, the solution to this is

$$I_\lambda = I_{\lambda,0} e^{-\tau_\lambda} \tag{4}$$

And it follows that

$$\tau_\lambda = N a_\lambda \tag{5}$$

Where N is the column density; the number of absorbing atoms per unit area. Typically, we work in units of cm^{-2}.

The three main approaches to estimating column densities from absorption lines:

- Apparent Optical Depth (AOD) method: This method is relatively straightforward and does not require assumptions about the line profile or number of components. However, it can underestimate N for saturated lines and overestimate N for blended lines. It's also sensitive to noise and errors in the continuum placement.

- Voigt profile fitting: This method is more complex and involves fitting a model line profile (a Voigt profile, which includes both Doppler and natural broadening) to the observed line. It can provide accurate estimates of N, the Doppler parameter (b), and the line center, even for saturated or blended lines. However, it can be computationally intensive and requires good initial estimates of the parameters.

- Curve of Growth (CoG): This method involves plotting the equivalent width (EW) of the line as a function of N and seeing where the line falls on the curve of growth, which has different regions corresponding to different optical depths. The CoG method can provide accurate estimates of N in the optically thin (linear) and optically thick (square root) regimes, but it's more uncertain in the intermediate (logarithmic) regime where the line is partially saturated. Also, CoG can be difficult to apply if multiple components are present.

Each of these methods has its advantages and limitations, and the best method to use can depend on factors like the quality of the data, the degree of line saturation, and the presence of blending. These three methods are discussed in more detail below.

2.2 Apparent Optical Depth (AOD) method

The Apparent Optical Depth (AOD) method, also known as the direct integration method. This method involves directly integrating the apparent optical depth across the absorption line to estimate the column density. It assumes that each velocity pixel in the absorption line profile is independent and optically thin.

The AOD method has the advantage of being relatively simple and not requiring assumptions about the number of components or their line profiles. However, there are some important caveats:

- Saturation: The AOD method can significantly underestimate the column density in the case of saturated lines. This is because once a line is saturated, the optical depth (and therefore the apparent column density) no longer increases proportionally with the true column density. Voigt profile fitting, on the other hand, can account for saturation effects.

- Blended lines: The AOD method assumes that each velocity pixel corresponds to a single component of absorption. If there are multiple overlapping components (as is often the case), the AOD method may overestimate the column density. Voigt profile fitting can account for multiple overlapping components.

- Non-Gaussian line profiles: The AOD method does not explicitly account for the shape of the absorption line. If the line has a non-Gaussian profile (e.g., due to multiple overlapping components, microturbulence, etc.), this can introduce errors in the estimated column density. Voigt profile fitting takes into account the detailed shape of the line profile, including both thermal (Gaussian) and collisional (Lorentzian) broadening.

- Errors and Noise: The AOD method is very sensitive to errors in the continuum placement and to noise in the data. Uncertainties in the continuum level or pixel-to-pixel variations due to noise can lead to significant errors in the calculated column densities.

2.3 Curve of Growth

The curve of growth, which describes the relation between the equivalent width (EW) of a spectral line and the column density, typically has three regions.

- Linear Region: At low column densities, the optical depth is low, meaning the line is optically thin. The absorption line is not saturated, so increases in the column density result in proportional increases in the EW. Thus, the EW is linearly proportional to the column density.

- Flat or Logarithmic Region: At intermediate column densities, the line becomes optically thick at the center (the line is saturated), but the wings of the line are still optically thin. Further increases in column density cause the line to broaden and the EW increases slowly, leading to a logarithmic relation between EW and column density.
- Square Root Region: At high column densities, the line is optically thick even in the wings, leading to a 'damping' or square root growth. Further increases in column density cause the line to broaden further, but now the broadening affects the wings of the line, which are less sensitive to changes in column density. The result is that the EW is proportional to the square root of the column density.

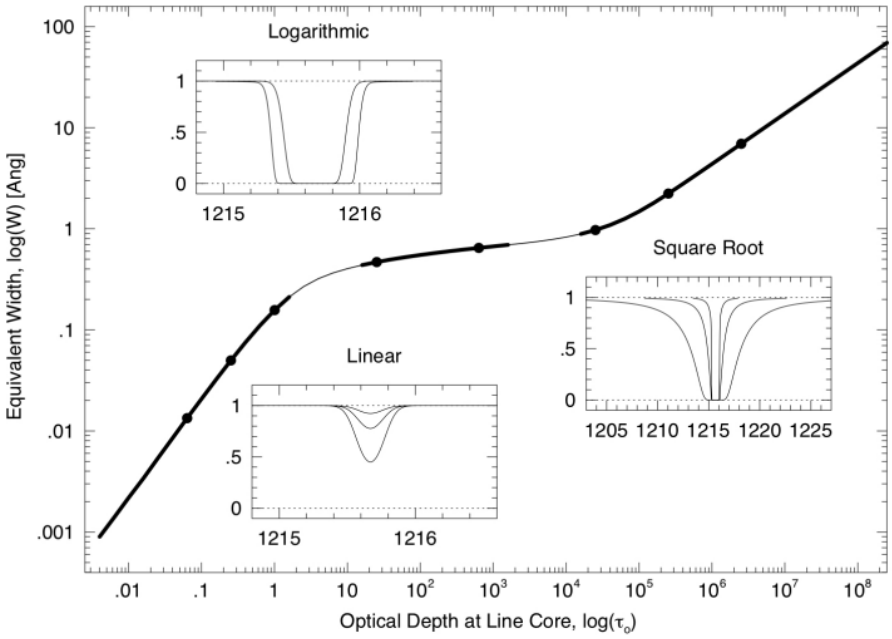

Figure 2: Curve of Growth graph for Lyman-α lines with $b = 30 kms^{-1}$.The three regions, the linear, flat, and damping part of the CoG are shown by thicker curves. (Figure courtesy of Chris Churchill).

The transition to the square root region happens because of the "damping wings" of the Lorentzian line profile, which represent natural broadening. Natural broadening is due to the finite lifetime of the upper energy level involved in the transition, and the uncertainty principle results in an uncertainty in the energy of the transition, which translates to an uncertainty in the frequency, or a broadening of the line.

The Lorentzian profile has extended wings, unlike the Gaussian profile, which drops off more rapidly. Thus, at very high optical depths, the column density is so high that even the distant wings of the line become optically thick, and the broadening of the line is determined by the shape of these wings, which follow a square root dependence.

2.4 Voigt Profile fitting

The Voigt profile is a mathematical function used to describe the shape of spectral lines, which is a convolution of two broadening mechanisms: Doppler broadening and natural (or Lorentzian) broadening. These two broadening mechanisms impact the spectral line in different ways:

- Doppler Broadening: This broadening effect is due to the thermal motion of particles. Atoms or ions moving towards or away from the observer cause a shift in frequency due to the Doppler effect. The resultant line profile is Gaussian in shape. The width of the Gaussian profile is

proportional to the square root of the temperature of the gas and inversely proportional to the square root of the mass of the atom or ion, which leads to the concept of the b-parameter, often used in Voigt profile fitting.

- Natural Broadening (or Lorentzian Broadening): This broadening effect is caused by the natural lifetime of the excited state involved in the transition. Quantum mechanics dictates that states with shorter lifetimes have more uncertain energies, which leads to a broadening of the spectral line. The resulting line profile follows a Lorentzian shape.

In the case where both types of broadening are important, neither a pure Gaussian nor a pure Lorentzian accurately represents the line profile. The Voigt profile is a convolution of these two profiles, and it accurately represents the line profile in this intermediate regime.

During the fitting process, a Voigt profile function is used to model the observed spectral line. This function has several parameters, including the line center (related to the velocity of the gas), the Doppler width (related to the temperature and turbulence of the gas), and the Lorentzian width (related to the natural lifetime of the transition). By adjusting these parameters to find the best match between the model and the observed spectral line, one can infer the physical conditions of the gas causing the absorption.

This is one of the reasons why Voigt profile fitting is commonly used in the analysis of interstellar and intergalactic medium: it provides a wealth of information about the physical state and composition of the gas along the line of sight. And this is the method used in the rest of this report.

3 Voigt profile fitting

3.1 MCMC simulations and Bayesian inference

Markov Chain Monte Carlo (MCMC) simulations are a class of algorithms used to sample from a probability distribution, especially when directly sampling is difficult. These methods are particularly useful in Bayesian statistical inference, where we often want to estimate the posterior distribution of model parameters given observed data. In the context of this report, we are unable to re-sample from the probability distribution without making another observation of the QSOs which is often impractical. The underlying theory of MCMC relies on constructing a Markov chain that has the desired distribution as its equilibrium distribution. The Markov chain is constructed such that, no matter what value(s) it starts from, it will eventually reach a state of equilibrium where the distribution of values does not change from one step to another.

The connection between MCMC and Bayes' theorem comes in when we're talking about Bayesian inference. Bayes' theorem provides a way to update our beliefs about a hypothesis (in this case, the values of the model parameters) in light of new evidence (the observed data). In mathematical terms, Bayes' theorem is often written as:

$$P(\theta|\text{data}) = [P(\text{data}|\theta) * P(\theta]/P(\text{data}) \tag{6}$$

Where $P(\theta|\text{data})$ is the posterior probability of the parameters, θ, given the data. This is what we are trying to estimate. $P(\text{data}|\theta)$ is the likelihood function, it respesents the likelihood of the data given the parameters. $P(\theta)$ is the prior probability of the parameters, which represents our beliefs about the parameters before seeing the data. And $P(\text{data})$ is the evidence, which is the probability of the data under all possible values of the parameters.

In Bayesian inference, we want to estimate the posterior distribution of the parameters, but this often involves complex, high-dimensional integrals that are difficult or impossible to solve analytically. MCMC provides a way to estimate this distribution numerically by generating a sequence of samples that, given enough time, will be representative of the posterior distribution. So in the context of our task, we would use MCMC to draw samples from the posterior distribution of the parameters of our Voigt profile (the column density, line center, and b-parameter) given the observed spectrum. These

samples can then be used to estimate quantities of interest, such as the mean or median of the parameters, their standard deviations (which gives a measure of the uncertainty).

When we use a Markov Chain Monte Carlo (MCMC) method, we start by initializing the parameters at some guess, and then iteratively update the parameters according to the Markov chain transition rule. The goal is for the chain to reach equilibrium, at which point it should be sampling from the desired probability distribution, in this case, the posterior distribution of the parameters given the data. The validity of MCMC relies on the assumption that the equilibrium distribution of the Markov chain matches the "real" distribution we want to sample from. In practice, the way we ensure this is through a property called detailed balance. Detailed balance requires that, in equilibrium, the probability of transitioning from any state i to any state j is the same as the probability of transitioning from state j to state i. Mathematically, this means that:

$$P(i) * P(i \to j) = P(j) * P(j \to i) \tag{7}$$

Where $P(i)$ is the probability of being in state i, $P(i \to j)$ is the probability of transitioning from state i to state j, and similarly for the reverse situation. When the detailed balance condition is satisfied, the Markov chain is guaranteed to have the correct equilibrium distribution.

The other key consideration when using MCMC is convergence — ensuring that the Markov chain has run long enough to reach its equilibrium distribution. Various diagnostic tools exist for assessing convergence, such as autocorrelation plots, and corner plots. It's generally recommended to discard the early part of the chain (known as the "burn-in" period) because it may be influenced by the initial guess and not representative of the equilibrium distribution. The beauty of MCMC methods is that, given enough time, they will converge to the desired distribution regardless of the initial guess. However, a good initial guess can help speed up convergence and reduce the necessary burn-in.

The author finds it pertinent to illustrate how the above theory has been applied to the problem at hand. In our case the spectrum for each quasar was collected only once. The rest of the choices we made can be directly related to Baye's theorem:

The likelihood function: Chi-squared function would be an appropriate choice for the likelihood function as our observational errors are Gaussian. To be precise, the likelihood function would be the exponential of the negative chi-squared divided by 2, because this gives us a proper probability density function (pdf) for a multi-dimensional Gaussian;

$$L = e^{\chi^2/2} \tag{8}$$

Where,

$$\chi^2 = \sum r^2/2\sigma^2 \tag{9}$$

Where r is the flux and σ is the error at each point in the spectrum.

The prior: The prior should encapsulate our prior knowledge or beliefs about the parameters. We do this by setting upper and lower bounds for each parameter, using our intuition which will be based on some physical factors as well as past experience when trying to run MCMC simulations. Setting boundaries for our parameters is a way of encoding uniform or "uninformative" priors, where we are saying that any value within those bounds is equally likely before seeing the data.

The overall setup: Given our likelihood and priors, we then use a MCMC sampler to draw samples from the posterior distribution of our parameters given the data. The MCMC sampler does this by constructing a Markov chain where each step is a proposed set of parameters, and the proposal is accepted or rejected based on the Metropolis-Hastings criterion (or a similar rule, depending on the specific MCMC algorithm). Over time, the chain should converge to the target distribution, which in this case is the posterior distribution.

Model definition: One critical step before all of this is defining the model that we use to fit the

data, which in our case would be a Voigt profile (or multiple Voigt profiles) for each detected spectral line. The parameters of this model (the line centers, widths, and column densities) are what we are inferring with our MCMC analysis.

After the MCMC simulation has been run, we are left with all of the sets of parameter values that were accepted by the Metropolis-Hastings algorithm. These are all values selected from the probability distribution that is representative of the probability distribution the data was collected from. Remember that we only kept results that were collected after the Markov Chain had reached equilibrium (once it had converged). We then calculate the mean or median of these retained sets of values. We can also compute the standard deviation or, even better, the credible intervals (for example, the range that contains 68% or 95% of the samples) to quantify the uncertainty in our estimates. It's noteworthy that these credible intervals can be asymmetric around the median or mean, reflecting the true shape of the posterior distribution.

3.2 Determining convergence

Both trace plots and auto-correlation plots are diagnostic tools used in MCMC simulations to assess the quality of the chains, but they serve different purposes. Trace plots are plots of the sequence of samples drawn from the Markov chain against the iteration number. They provide a visual way to check for convergence of the chains. Once the chain has converged the plot will oscillate around a constant mean without any trend. Auto-correlation plots are plots that show the correlation of the Markov chain with itself after a different number of iterations. If the auto-correlation drops off quickly, that's a good sign and suggests our chain is exploring the parameter space well.

Corner plots are another common tool used in the analysis of MCMC results, especially when dealing with multi-dimensional parameter space. They provide a way to visualize both one-dimensional and two-dimensional marginal distributions for the parameters. A corner plot is a grid of plots with as many rows and columns as there are parameters in our MCMC analysis. The diagonal elements of this grid show the histograms of the distributions of each individual parameter. This can be used to visually estimate properties such as the median, mode, and credible intervals of the parameters, as mentioned above. The off-diagonal elements of the grid show scatter plots of each pair of parameters. This gives a sense of the correlation different parameters. For example, if two parameters are highly correlated, we will see a scatter plot that shows a clear trend or line; if they are not correlated, the scatter plot would look like a cloud of points with no discernible trend. For convergence we are looking for off-diagonal plots that have concentric rings. Several examples are given below.

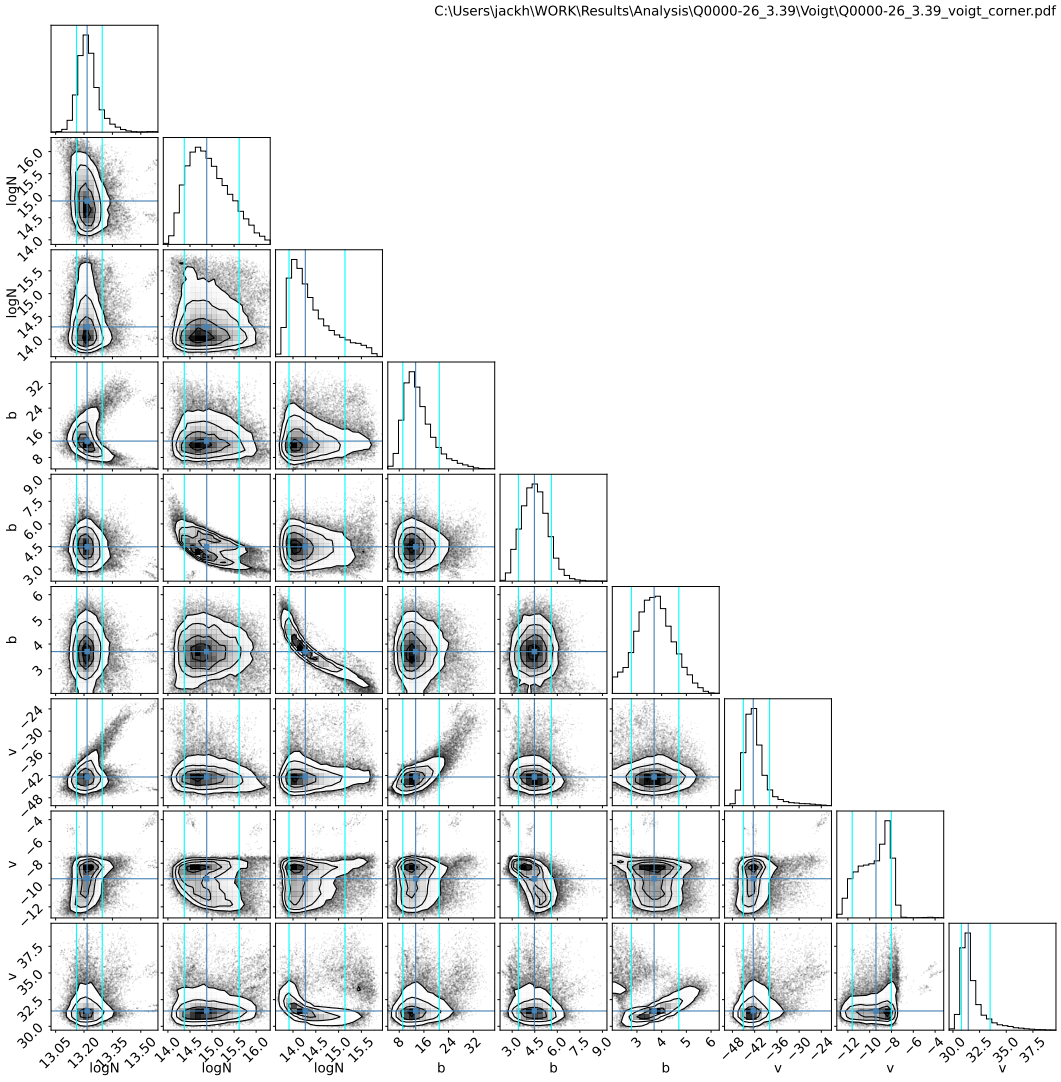

Figure 3: Example of a good corner plot. There are concentric circles, this has simulation has converged well and the errors for the parameter values are low.

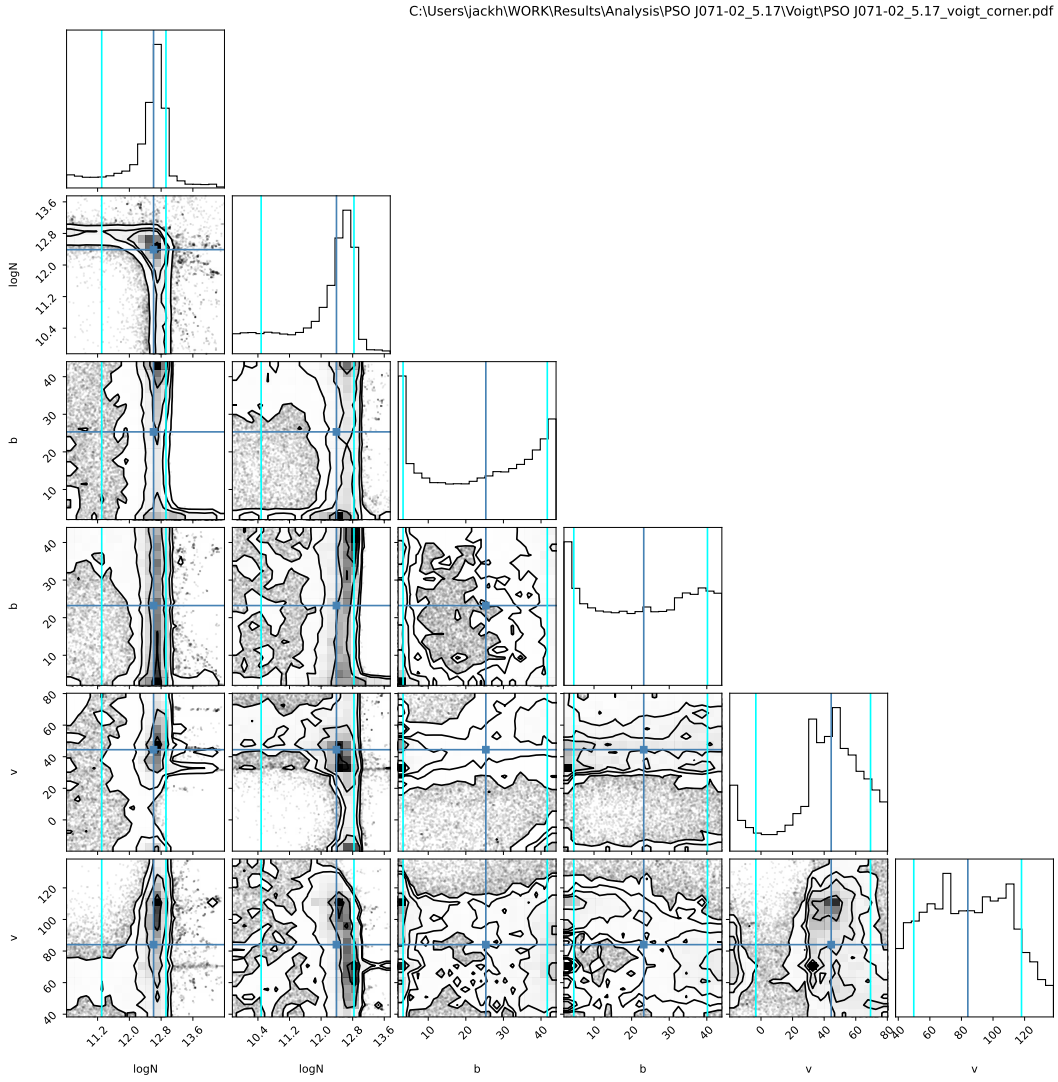

Figure 4: Example of a bad corner plot. There are no rings, the MCMC simulation appears to be sampling parameter space without converging.

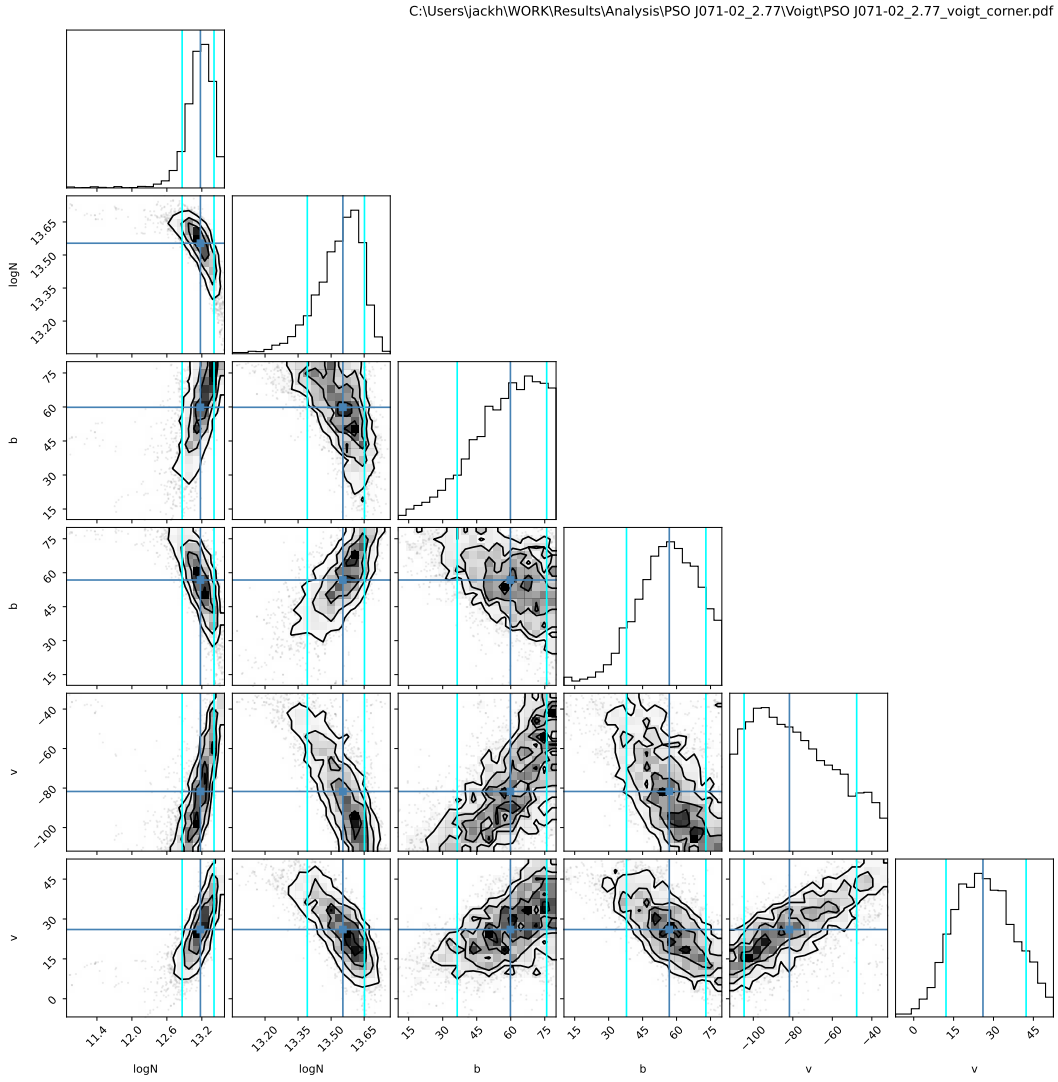

Figure 5: Example of a degenerate corner plot. There are some rings but some of the parameters have columns, this indicates that the solutions are degenerate.

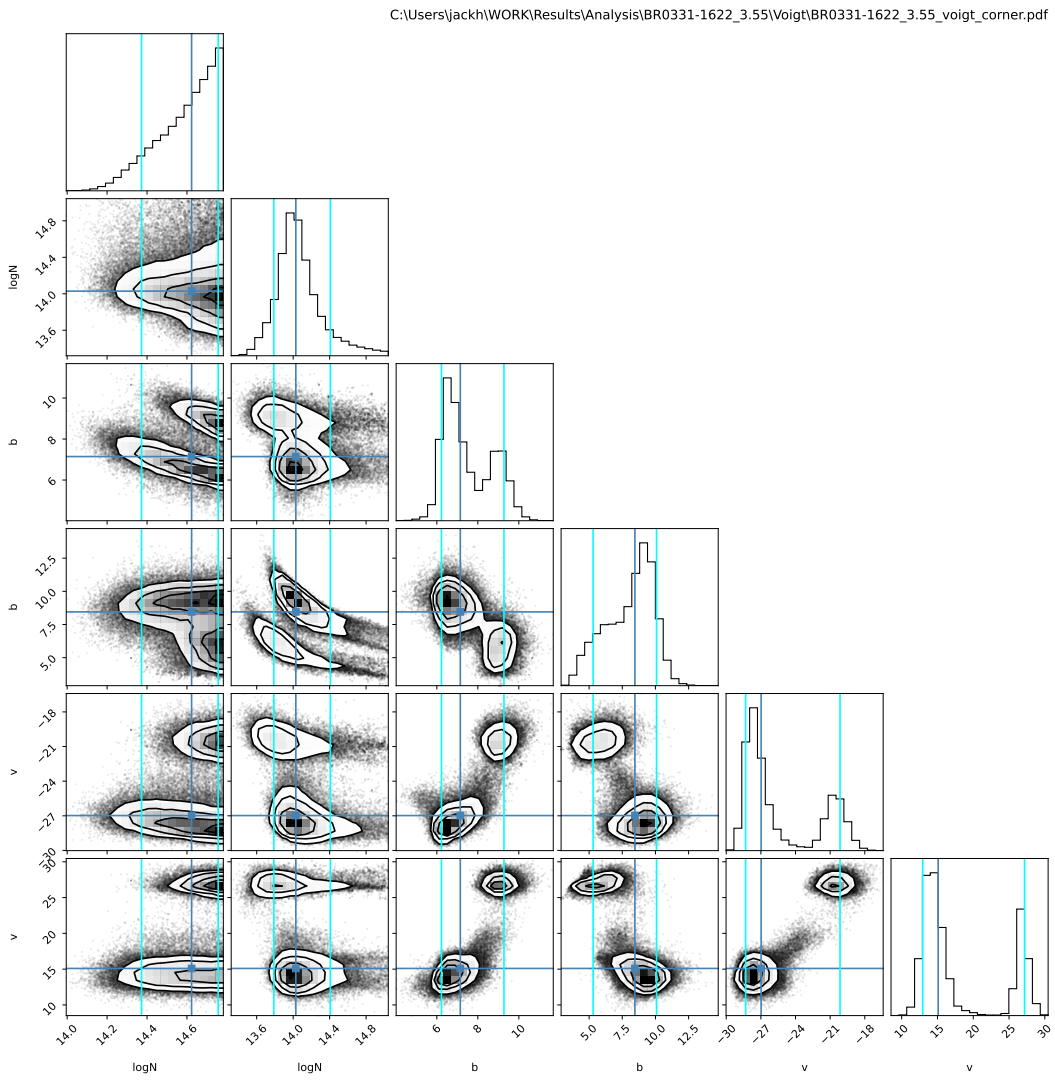

Figure 6: Example of a doubly degenerate corner plot. It appears there are two solutions that are valid for one of the Voigt profiles. In this case it would be prudent to redo the simulation with two Voigt profiles to replaces this one.

4 Results

From our catalogue of 101 high redshift QSO's, we found 94 detections of FeII. From there we performed Voigt profile fitting using MCMC simulations to calculate the velocity centroid(s), b-parameter, and column density, N. The methods use run parallel to the discussion and theory above. The code that we used was *rbvfit* (https://github.com/rongmon/rbvfit.git) by Rongmon Bordoloi which utilizes the emcee python package.

Of the 94 absorbers, 13 had had errors with a 68% credibility interval greater than 0.8 dex. And of these fits with larger dex, only 3 have poor convergence. The reason for the poor convergence is still being investigates, however there are several contributing factors that are believed to be at play. For instance, 2 of the fits have significant contamination on some of the lines. As in, because there are 4 lines to do the Voigt profile fitting for (FeII, 2344.2, 2382.8, 2586.7, 2600.2) and there are contaminations for some of these lines, they have different shapes so there is large error on the components at that particular velocity. This leads the code to retain samples from the probability distribution for both solutions which, in term, increases the uncertainty.

Another cause of higher uncertainty in the fitting could come from the relatively low resolution and high noise for some of the spectra. This creates ambiguity when knowing how many components to add to the fit. There is always a trade off between having the more components which will hopefully be more reflective of the actual absorber, and having less components which will yield a significantly lower uncertainty in the column density value. These issues will be resolved in the forthcoming iterations of this report.

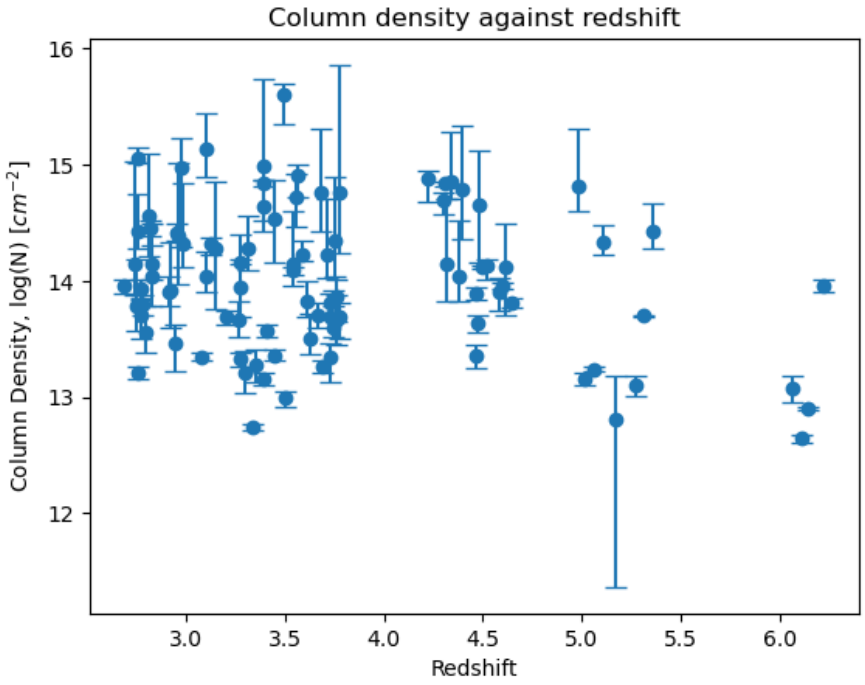

Figure 7: Column density against redshift. Error bars represent the 68% credibility interval.

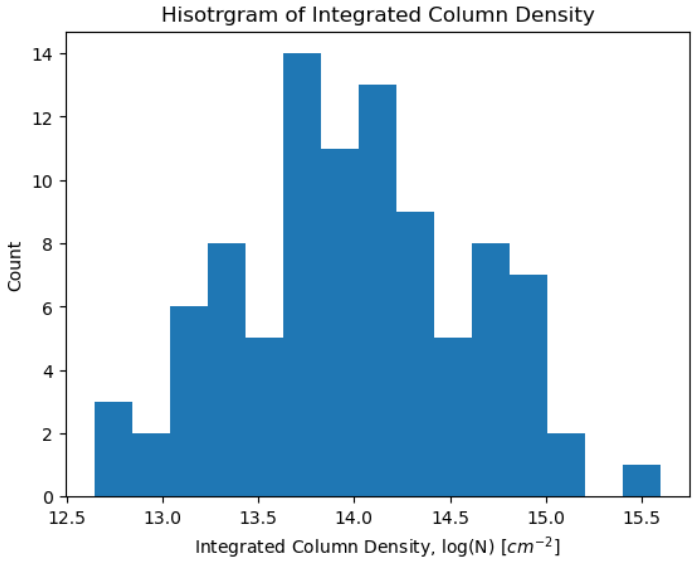

Figure 8: Histogram of Integrated Column Density.

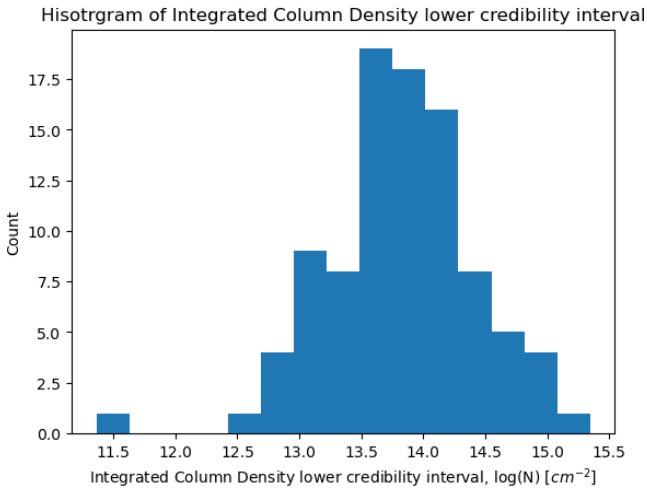

Figure 9: Histogram of Integrated Column Density lower credibility interval.

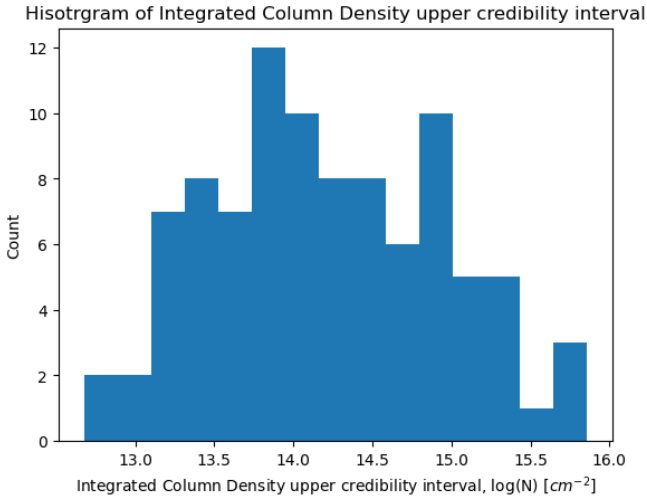

Figure 10: Histogram of Integrated Column Density upper credibility interval.