



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE ASTRONOMÍA

**A MORPHO-KINEMATIC MODEL OF THE HIGH-REDSHIFT CIRCUMGALACTIC
MEDIUM CONSTRAINED BY QUASAR ABSORPTION LINES**

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS, MENCIÓN ASTRONOMÍA

MAGDALENA JESUS HAMEL BRAVO

PROFESOR GUÍA:
Sebastián López Morales

MIEMBROS DE LA COMISIÓN:
Trystyn Berg
Cédric Ledoux
René Méndez Bussard
Nicolás Tejos Salgado

SANTIAGO DE CHILE
2022

RESUMEN DE LA TESIS PARA OPTAR
AL GRADO DE MAGÍSTER EN CIENCIAS, MENCIÓN ASTRONOMÍA
POR: MAGDALENA JESUS HAMEL BRAVO
FECHA: 2022
PROF. GUÍA: SEBASTIÁN LÓPEZ MORALES

UN MODELO MORFO-CINEMÁTICO DEL MEDIO CIRCUMGALÁCTICO A ALTO REDSHIFT CONSTREÑIDO CON ESTADÍSTICA DE LINEAS DE ABSORCIÓN EN QUÁSARES

El medio circumgalactico (CGM) corresponde al gas que rodea una galaxia, pero sigue gravitacionalmente ligado a esta. Este componente es de suma importancia en el campo de evolución galáctica, ya que provee el combustible que lleva a las galaxias a la formación estelar. Las líneas de absorción de Mg II en el espectro de QSOs proveen una forma de rastrear el gas frío en el CGM. En esta tesis propongo un modelo morfo-cinemático para el CGM de un disco poblado de nubes individuales de Mg II que estén en rotación. Dado este modelo, género líneas sintéticas de absorción de Mg II para constreñir el modelo propuesto, usando observaciones reales de líneas de absorción de Mg II en espectro de QSO. Haciendo uso del ancho equivalente y su conocida dependencia con el parámetro de impacto, constreñí los parámetros del modelo usando simulaciones Markov Chain Monte Carlo (MCMC). Los resultados muestran que la distribución de nubes, modelada por el parámetro volumetric filling factor, decrece con la distancia a la galaxia. Específicamente, una función exponencial decreciente con un radio característico de $r_0 = 0.26R_{vir}$ puede explicar la dependencia del ancho equivalente con el parámetro de impacto. Los resultados también muestran que el tamaño de las nubes, c_{size} , se relaciona con el alto del disco, h , de la siguiente forma: $h \approx 6c_{size} + 12$. También, los resultados muestran que el disco tiene que tener una rotación sólida, sin gradiente de velocidad en la componente vertical, para reproducir el scatter observado. En esta tesis también calculé el pixel velocity two-point correlation function para el mejor modelo, para diferentes ángulos azimutales e inclinaciones de la galaxia. Al comparar estos resultados con las observaciones, se puede ver que el modelo puede reproducir la dispersión de velocidad observada en líneas de visión que pasan cerca del eje mayor, donde esperamos encontrar rotación y acreción. Sin embargo, produce dispersiones de velocidad menores a lo observado en líneas de visión cercanas al eje menor, donde esperamos encontrar gas saliendo del CGM. Por lo tanto estos resultados sugieren que al modelo le hace falta una componente de gas saliendo del disco para poder reproducir la dispersión de velocidad observada.

RESUMEN DE LA TESIS PARA OPTAR
AL GRADO DE MAGÍSTER EN CIENCIAS, MENCIÓN ASTRONOMÍA
POR: MAGDALENA JESUS HAMEL BRAVO
FECHA: 2022
PROF. GUÍA: SEBASTIÁN LÓPEZ MORALES

A MORPHO-KINEMATIC MODEL OF THE HIGH-REDSHIFT CIRCUMGALACTIC MEDIUM CONSTRAINED BY QUASAR ABSORPTION LINES

The circumgalactic medium (CGM) is loosely defined as the gas that surrounds galaxies outside the stellar disc, but is still gravitationally bound to the galaxy. The CGM is key to galaxy formation and evolution, since it is the volume where galactic-scale processes meet in what is called the "baryon cycle of galaxies". In summary, cold streams of fresh gas from the intergalactic medium acquire momentum and eventually provide fuel for star formation, while processed gas is expelled out in galactic-scale winds driven by supernova explosions and/or active galactic nucleus (AGN) activity. Mg II absorption lines in the spectra of quasi-stellar objects (QSO) provide a sensitive observational tool to trace the diffuse and cool gas present in the CGM; however, despite a wealth of observational data acquired so far, not much has been proposed on modeling the CGM. In this thesis I propose a morpho-kinematic model of a rotating disk-like CGM populated with individual Mg II clouds. The model generates synthetic Mg II absorption lines that are used to recreate the observed rest-frame equivalent width versus impact parameter relation and to constrain the model via Markov Chain Monte Carlo simulations. The results are as follows: the distribution of clouds, modeled by the volumetric filling factor, decreases as an exponential function with characteristic radius, $r_0 = 0.26R_{vir}$, i.e., $\approx 1/4$ of the virial radius, the cloud size is correlated with the height of the disk, the disc needs to have a solid rotation, with no vertical velocity gradient in order to reproduce the large scatter observed in the equivalent width vs. impact parameter distribution. The best-fit model is used to compute the pixel velocity two point correlation function for different azimuthal angles and inclinations. By comparing these results with observations, it can be seen that the model is able to reproduce the velocity dispersion in sightlines near the major axis, which is where we expect to find accretion or rotation. However the model under-predicts the velocity dispersion along the minor axis where we expect to find outflows. Thus the results suggest that in order to reproduce the observed velocity dispersion along the minor axis, the model is missing an out-flowing component.

Si no estás dispuesta a seguir buscando la luz en los lugares más oscuros sin parar, incluso cuando parece imposible, nunca lo conseguirás.

Amelia Shepherd

Agradecimientos

En primer lugar me gustaría agradecer a mi familia, sobre todo a mi mamá (Pilar Bravo) y a mi papá (Sebastián Hamel), quienes siempre me han apoyado en todos mis sueños. Es realmente un privilegio ser su hija en todo sentido. Muchas gracias por haberme dado todas las herramientas que he necesitado para cumplir mis metas. También quiero agradecer al resto de mi familia, mis hermanos, sobrino y sobrina, abuelas y abuelos, primas y primos, tías y tíos, por siempre entusias-marse con lo que estudio y todos los pasos que doy en mi vida. Estoy muy agradecida de tener una gran y apoyadora familia.

Quiero agradecerle también a mi pololo Lucas quien desde el día 0 me ha acompañado en este Magíster. Es sin duda la persona que más me ha escuchado las alegrías y frustraciones que he tenido en el camino. Gracias por alegrarte conmigo y darme tu apoyo en los momentos difíciles. Sé que eres mi mayor fan y yo soy la tuya.

A todos los amigos que he hecho desde el colegio hasta el Magíster, sin compañeros para estu-diar (y a veces carretiar) no hubiera sido lo mismo toda la experiencia. Gracias a mis amigas del colegio, de scout, a mis amigos de Bachi, mis amigos de la salita y mis amigos del Magíster. Son demasiados así que no alcanzo a poner nombres, pero me voy a asegurar de agradecer en persona, los quiero mucho.

También quiero agradecer a mi profesor guía, Sebastián López, por introducirme en el mundo del medio circumgaláctico. Gracias por todo lo enseñado y las oportunidades que me ha dado para seguir creciendo como astrónoma. También quiero agradecer a mi comisión por los comentarios que hicieron esta tesis mejor. En especial a Trystyn Berg, quien estuvo aconsejándome desde tiempo antes y fue una parte fundamental de que se lograra esta tesis, thank you.

A toda la gente de Arctomo por compartir en reuniones que me han servido mucho tanto par mi trabajo personal como para expandir mi conocimiento a otras partes de la astronomía. Quiero agradecer al instituto francés y a la universidad de Chile por darme la oportunidad de trabajar en Lyon donde completé una parte de esta tesis. Agradecer al Dr. Bouché con quien trabajé en Lyon, por ayudarme a seguir trabajando en este proyecto. También agradecer a la Dra. Hibon con quién realicé una pasantía en la ESO y también me ayudó con parte de esta tesis.

Por último quiero agradecer a mis mascotas, Lara, Gol, Popy y Django. Son lo más lindo de este mundo y siempre me dieron alegrías cuando lo necesitaba.

Table of Content

1. Introduction	1
1.1. Motivation	1
1.2. The circumgalactic medium	2
1.2.1. Baryon cycle	3
1.3. Advances in CGM science used in the present thesis	4
1.3.1. CGM in absorption	4
1.3.2. Mg II absorption data	6
1.3.3. Equivalent width	8
1.3.4. Velocity Two-point correlation function	10
1.4. This Thesis	13
2. Modeled data	14
2.1. The model	14
2.1.1. Model geometry	14
2.1.2. Mg II clouds	18
2.1.3. Cloud kinematics	22
2.2. Absorption line spectra	24
2.3. Synthetic catalogs of spectra	29
2.3.1. Geometrical parameters	29
2.3.2. Mg II clouds	31
2.3.3. Velocity parameters	31
2.3.4. Absorption line parameters	32
3. Model comparison with QSO statistics	35
3.1. Equivalent width vs. impact parameter	35
3.2. Two-point correlation function	39
3.3. Markov Chain Monte Carlo	41
3.3.1. The Likelihood function	42
3.3.2. Priors	43
3.3.3. Computational times	43
4. Results	45
4.1. Filling factor	45
4.2. MCMC results	50
4.2.1. $W_r - d/R_{\text{vir}}$ relations	56
4.2.2. Number of clouds	59
4.3. Two-point correlation function results	61
4.4. Summary of the results	63

5. Discussion and Conclusions	65
5.1. Discussion	65
5.1.1. MgII spatial distribution	65
5.1.2. Kinematics of the CGM	67
5.2. Summary and conclusions	68
5.3. Future work	69
Bibliography	70

List of Tables

1.1.	MAGiiCAT properties. $W_r(2796)(\text{\AA})$: rest frame equivalent width of Mg II(2796). z_{gal} : redshift of the absorbing galaxy. $d(kpc)$: impact parameter of the sightline with respect to the galaxy. M_B : absolute magnitude in the B-band. M_K : absolute magnitude in the K-band. L_B/L_B^* : B-band luminosity. L_K/L_K^* : K-band luminosity. $B - K$: rest-frame color.	8
1.2.	Sub-samples used in Nielsen et al. 2015. α is azimuthal angle and i is galaxy inclination	11
2.1.	Geometrical parameters of the model	17
2.2.	Cloud parameters	22
2.3.	Velocity parameters.	23
2.4.	Atomic constants	25
2.5.	Summary of all parameters necessary to produce a Mg II absorption profile given this model.	27
2.6.	All model parameters and the method used to assign their values in the individual spectra of the synthetic catalog. In blue are the parameters that take values from a distribution in the literature. In yellow are the free parameters of the model.	34
3.1.	Boundaries for uniform priors of the free parameters	43
4.1.	High and low probability values for each parameter of the model according to MCMC	57

List of Figures

1.1.	Artistic representation of the current state of the art on the baryon cycle. In orange we can see the outflowing structures and in pink we can see how some of the outflowing material falls back onto the galaxy. In blue we can see the gas being accreted into the ISM. In purple we see all the diffuse gas that composes the CGM. Figure from Tumlinson et al. (2017)	3
1.2.	Schematic representation of the QSO probing CGM technique. At the top is a representation on how the light of the quasar travels towards the Earth and encounters intervening gas in its way. At the bottom we see the QSO spectra with several absorption resulting from the intervening gas. Most of the redshifted transitions are by neutral hydrogen from the inter-galactic medium (forming the so-called “Lyman-alpha forest”). Some of them, though, are by metal lines, revealing enriched material. Figure by Joe Liske.	5
1.3.	A gallery of $\text{Mg II}(\lambda 2796)$ profiles as a function of rest-frame velocity. The profiles are separated into subsamples (B, C, E and High redshift) and ordered by increasing equivalent width. The rest-frame equivalent widths are labeled in the lower corners of each subpanel and in \AA units. Figure from Churchill & Vogt (2001).	7
1.4.	Schematic picture of the two definitions of equivalent width. On the left there is an absorption profile with the area under the continuum colored with blue, this area would be the equivalent width if the flux is normalized. On the right there is a rectangle with the same area, the width of this rectangle would be the W_r . Picture from COSMOS - The SAO Encyclopedia of Astronomy	9
1.5.	Equivalent width vs. impact parameter for sample in MAGiCAT. Arrows show the upper limits of the sample, solid curve shows the fit and dashed curve shows the 1σ envelope. (a) W_r vs. d for four different mass bins. (b) W_r vs. d for two different mass bins. (c) W_r vs. d/R_{vir} for the same four mass bins as panel (a). Figure from Churchill et al. (2013)	10
1.6.	Pixel-velocity two-point correlation function (TPCF) for different inclinations and azimuthal angles. Panels (a)-(c) show the TPCF for edge-on ($i \geq 57^\circ$) and face-on ($i < 57^\circ$) galaxies with sightlines with all azimuthal angles, sightlines along the minor axis ($\alpha \geq 45^\circ$) and sightline along the major axis ($\alpha < 45^\circ$) respectively. Panels (d)-(f) show the TPCF for sightlines along the major ($\alpha < 45^\circ$) and minor ($\alpha \geq 45^\circ$) axis for galaxies with all inclinations, edge-on ($i \geq 57^\circ$) galaxies and face-on ($i < 57^\circ$) galaxies. All these panels include all galaxy colors, B-K. The solid lines represent the TPCFs and the colored regions are the 1σ bootstrap uncertainties. Each panel includes also a reduced χ^2 value of comparing the two TPCFs, and the degrees of freedom, ν . Figure from Nielsen et al. (2015)	12

2.1.	Model's geometry, light blue represents the CGM disk and darker blue represents the stellar disk. The yellow star and arrow represent the QSO sightline; panel (a), black arrow represents the major axis of the galaxy. d is the impact parameter, which is the projected distance between the center of the galaxy and the sightline, α is the azimuthal angle, which is the angle between the major axis of the galaxy and the line connecting the galaxy centre and the position of the QSO on the sky, h is the thickness of the disk and i is the line-of-sight inclination of the galaxy.	15
2.2.	Coordinate system of the model, light blue represents the CGM disk and darker blue represents the stellar disk. Fig. (a) shows a sketch of the model in the x-y plane or the plane of the disk. x_0 is the x-coordinate where the sightline intercepts the disk. v_{rot} shows the direction of rotation of the stellar disk. Fig. (b) is a sketch of the model in the y-z plane or side view of the disk. y_{max} and y_{min} are the y-coordinates where the sightline enters and leaves the disk; and y_0 is the y-coordinate where the sightline intercepts the mid-plane of the disk.	17
2.3.	Schematic picture of the grid used in the model. The volume of each grid cell is $c_{size} \times c_{size} \times c_{size}$. Red cells are the cells that are crossed by the sightline, R is the 2-d distance from the center of the disk to the center of each cell in the projected plane of the disk and z is the perpendicular distance from the mid-plane of the disk ($z=0$) to the center of each cell.	19
2.4.	Schematic picture of two different $f_v(R/R_{vir})$. (a) Shows a plot of the two different filling factors, a step function $f_v(R/R_{vir})$ (pink) and an exponential $f_v(R/R_{vir})$ (green). (b) Schematic picture of how would the plane of the disk look like with the two different $f_v(R/R_{vir})$. Circles are used to show which cells are filled with Mg II. Black dotted circles show a distance of $R = 0.7R_{vir}$ from the center of the galaxy and blue dotted circles show a distance of $R = R_{vir}$ from the center of the galaxy. Left panel (with pink circles) shows the CGM disk for the step function $f_v(R/R_{vir})$, right panel (with green circles) shows CGM disk for an exponential $f_v(R/R_{vir})$. (c) Schematic picture of how the side of the CGM disk would look like with the two different filling factors. Right picture is for the step function filling factor and left is for the exponential.	21
2.5.	Plot of the circular velocity, in terms of the maximum velocity of rotation, of a cloud with respect to the distance to the mid-plane of the disk. y-axis shows z values, where $z=0$ is the mid-plane of the disk, black dotted line. Color lines show the velocities for three different h_v values.	23
2.6.	Plot of three different profiles that characterize an absorption line. Blue: Gaussian profile, which models the Doppler broadening of the random motion of the gas. Orange: Lorentzian profile that models the pressure broadening. Green: Voigt profile, which is the convolution of the two previous profiles and shows the form of the final absorption profile. All these profiles are normalize so that the area under the curve is equal to 1 and with FWHM = 1	24
2.7.	Plots of different Mg II 2796 absorption profiles calculated with Equation 2.7. Panel (a) shows the effect of changing the column density N and panel (b) shows the effect of changing the Doppler parameter b for optically thick systems.	26
2.8.	Plot to show the effect of adding the absorption lines of different clouds to produce the final absorption profile. In blue are three individual profiles, all with $b = 8\text{km s}^{-1}$, different N and different velocities. In red is the total absorbing profile, resulting from adding the three individual profiles.	27

2.9.	Plot to show the effect of different spectral resolutions. Red is the real spectrum, blue shows how the spectrum would look like with a spectrograph with resolution $R = 45000$, orange shows how the spectra would look like with $R=10000$ and green line shows how the spectrum would look like with a spectrograph with resolution $R = 4000$	28
2.10.	Distribution of $\sin(i)$ used to create the synthetic catalogs. f_y is the distribution of $\sin(i)$ assuming a random distribution for inclinations of disk objects. Figure from Lopez & Jenkins (2012).	30
2.11.	Distribution of impact parameters present in the MAGiiCAT sample.	31
2.12.	Schechter function that characterises the distribution of maximum velocities of rotation for stellar disk galaxies in the low redshift Universe for different inclinations. Figure from Zwaan et al. (2010).	32
2.13.	Data from Churchill et al. (2020) Panel (a): Mg II column density distribution from Voigt profile fitting. (b) Mg II Doppler parameter distribution from Voigt profile fitting.	33
3.1.	$W_r - d$ relation. Blue points show Mg II detections that are included in MAGiiCAT, white points with arrows show upper limits, solid black line shows the best fit for MAGiiCAT data of a fixed form, dashed black line show the 1σ uncertainty of the fit and green line shows the best fit power-law using Chen et al. (2010a) sample. Figure from Nielsen et al. (2013).	36
3.2.	Example of absorption profile produced by the model plotted in the rest-frame velocity scale. Each panel shows a different possible synthetic spectra that can come out of the model at a given impact parameter. Solid lines are the profiles and grey areas represent the kinematic subsystems. Each panel shows the rest-frame equivalent width, these values are in \AA , produced by each absorption profile. It can be seen that smaller d tend to produce larger W_r	38
3.3.	Example of TPCFs for two different samples of spectra. Each sample has five different spectra. The top five panels on the left show the three different spectra in the first sample. They are plotted on a rest-frame velocity scale. First five plots on the right show the three spectra for the second sample, plotted on a rest-frame velocity scale. Bottom panel on the left shows the TPCF for the first sample and bottom panel on the right shows TPCF for the second sample. x-axis in TPCF plots show velocity absolute difference, y-axis shows the probability of having two absorption pixels with a given velocity absolute difference.	40
4.1.	$W_r - d/R_{\text{vir}}$ relation for three different $f_v(R/R_{\text{vir}})$. (a): three different $f_v(R/R_{\text{vir}})$, in blue is the <i>increasing step</i> f_v , in green the <i>decreasing step</i> f_v and in orange the <i>constant step</i> f_v . (b): $W_r - d/R_{\text{vir}}$ relationship for each $f_v(R/R_{\text{vir}})$. Colored areas show the maximum and minimum of W_r in 9 bins of d/R_{vir} , for each $f_v(R/R_{\text{vir}})$, with the respective colors. Grey points and arrows show detections and upper limits in MAGiiCAT respectively. I performed a KS test to compare the models relationship to the observations, the p-values where: $\approx 10^{-20}$ (<i>increasing step</i> f_v), $\approx 10^{-8}$ (<i>decreasing step</i> f_v) and $\approx 10^{-14}$ (<i>constant step</i> f_v). Top of panel (b): histograms showing the distributions of d/R_{vir} for each sample, in the same bins as the W_r areas and in the same colors as panel (a), and in grey, the d/R_{vir} distribution for the MAGiiCAT data.	47

4.2.	<p>$W_r - d/R_{vir}$ relationship for three different exponential $f_v(R/R_{vir})$, with different r_0, panel (a). (b): resulting $W_r - d/R_{vir}$ relations with each $f_v(R/R_{vir})$. Color areas show the maximum and minimum of W_r in 9 bins of d/R_{vir}, for each $f_v(R/R_{vir})$, with the respective colors. Grey points and arrows show detections and upper limits in MAGiiCAT respectively. I performed a KS test to compare the models distribution to the observations, the p-values where: $\approx 10^{-14}$ ($r_0 = 2R_{vir}$), $\approx 10^{-3}$ ($r_0 = \frac{1}{3}R_{vir}$) and $\approx 10^{-16}$ ($\frac{1}{10}R_{vir}$). Top of panel (b): histograms showing the distributions of d/R_{vir} for each sample, in the same bins as the W_r areas and in the same colors as panel (a), and in black, the d/R_{vir} distribution for the MAGiiCAT data.</p>	49
4.3.	<p>Corner plots showing the PDFs obtained with MCMC. Diagonal plots show the marginalized PDF for each free parameter, the median value and the 1σ intervals are showed by dashed lines. At the top of each PDF are the median values and 1σ values. Off-diagonal plots show the 2 dimensional PDFs for each pair of free parameters. Light grey points show where the MCMC walkers accepted values, light grey area, dark grey area and black area show where the 0.84, 0.5 and 0.16 of the sample lay.</p>	51
4.4.	<p>Histogram of acceptance ratios of individual walkers. The x-axis shows the acceptance ratio value and y-axis shows the number of walkers.</p>	52
4.5.	<p>Trace plots for MCMC walkers. It shows the parameter values for each parameter and each walker at each step. Blue lines are walkers with high acceptance ratio, red lines are walkers with low acceptance ratio and green lines show the mean value for each parameter. x-axis is the step number.</p>	53
4.6.	<p>$W_r - d/R_{vir}$ relations for model with high probability for walkers with high acceptance ratio (blue points left panel) and walkers with low acceptance ratio (red points right panel). In both panels black dots and arrows show MAGiiCAT detections and upper limits respectively. At the bottom left corner in each panel, tables show the mean values of parameters for each group of walkers.</p>	54
4.7.	<p>MCMC corner plots using only walkers with high acceptance ratios. Same caption as Figure 4.3 but in red are the 2D PDF for walkers with low acceptance ratios, just for comparison.</p>	55
4.8.	<p>Model $W_r - d/R_{vir}$ relationship for different sets of parameters. In each panel there are grey points and arrows showing MAGiiCAT detections and upper limits, blue points showing models with $\frac{1}{r_0} R_{vir} = 3$ and orange points showing models with $\frac{1}{r_0} R_{vir} = 8$. Panels in the left column (a,b,e and f) show models with $h = 45kpc$ and on the right column panels (c,d,g and h) are models with $h = 5kpc$. The first row of panels (a,b,c and d) show models with $\log(h_v h) = 4$ and bottom row panels (e,f,g and h) show models with $\log(h_v h) = -2$. Finally in each corner of the figure are two panels, all top panel at each corner (a,c,e and g) show models with $c_{size} = 5kpc$ and bottom panels at each corner (b,d,f and h) show models with $c_{size} = 0.1kpc$. In each panel there are two values annotated, in the upper left corner, one in blue, one in orange, those are the p-values for each model in each panel.</p>	58
4.9.	<p>Distribution of number of clouds crossed a sightline. Red histogram shows distribution from Churchill et al. (2020), which comes from Voigt profile fitting to high resolution spectra of random sightlines. Blue, green and orange histograms show the number of clouds for three different models, all of them with $h = 44.45kpc$, but different c_{size}, $c_{size} = 4kpc$, $c_{size} = 0.1kpc$, $c_{size} = 15kpc$, which translate to different number of clouds densities.</p>	60

4.10.	Number of clouds crossed by a single distribution. Red histogram shows distribution from Churchill et al. (2020), which comes from Voigt profile fitting to high resolution spectra. Blue, green and orange histograms show the number of clouds for three different models, with $h = 30kpc$, $h = 60kpc$ and $h = 30kpc$ respectively. All of the models were created with c_{size} values that follow $h = 6c_{size} + 12$	61
4.11.	Pixel two-point correlation function computed from the model with higher probability of reproducing $W_r - d/R_{vir}$ relation, in different sub-samples of galaxy inclination and azimuthal angle. Panel (a) shows all sightlines, (b) shows only sightlines near the minor axis, (c) shows only sightlines near the major axis, (d) shows all galaxies, (e) shows only edge on galaxies and (f) show only face-on galaxies. Light blue represent face-on galaxies, orange represent edge-on galaxies, green represents sightlines near the major axis and pink represents sightlines near the minor axis. Solid lines are the TPCF for the model and colored areas show the 1σ intervals for MAGiiCAT data. . .	62
5.1.	Schematic picture of two different CGM models with high probability. panel (a): model with $h = 30kpc$ and $c_{size} = 3kpc$. panel (b): model with $h = 60kpc$ and $c_{size} = 6kpc$	66

Chapter 1

Introduction

1.1. Motivation

In the Universe objects tend to group in structures, some of them are galaxies. Besides dark matter, galaxies are made of baryonic matter, which, unlike DM, can be detected thanks to its interaction with electromagnetic radiation, either in emission or in absorption. This is the only direct constraint to test theoretical predictions on galaxy formation and evolution; therefore, it is essential to understand galaxy-scale mechanisms involving transfer of baryonic matter to properly interpret observations. Thus, studying such "baryon cycle" makes it possible to better understand the processes leading to star formation, which is fueled by such baryons (Rees & Ostriker, 1977; White & Rees, 1978), and therefore also galaxy evolution.

Beginning in the 1960s, multiple absorption lines by atomic transitions of heavy atoms ("metals") were recognized in the spectra of quasi-stellar objects (hereafter indistinctly called "quasars" or "QSOs") at different redshifts (Bahcall & Spitzer, 1969). Today it is well established that such lines are produced by the interstellar medium of galaxies and/or tenuous gas in the extended halos of intervening galaxies (Bahcall & Spitzer, 1969; Behroozi et al., 2010). The discovery of gas surrounding galaxies outside the interstellar medium (ISM) and surrounded by the intergalactic medium (IGM), forced astronomers to define the notion of circumgalactic medium (CGM).

The CGM is a good place to start looking for answers to some of the questions that are still remaining in the field of galaxy evolution (Tumlinson et al., 2017). For example, the problem of the galactic missing baryons, which arises from the fact that observations of baryonic matter in late type galaxies only account for less than half of the baryons in the early universe (Behroozi et al., 2010; McGaugh et al., 2010). By examining the baryonic gas in the CGM if we could find those missing baryons. This would mean that half of the baryonic matter in galaxies resides in the CGM (Zheng et al., 2015; Peebles et al., 2014).

Another question is how galaxies sustain their star formation, since the gas in their ISM can last for only a small fraction of the time they have been forming stars (Peebles et al., 2014). One possible explanation for this is that there is gas accreting into the ISM and a good place to look for this gas supply is the CGM. In the same line is the question of how galaxies stop forming stars and become passive and a possible explanation relies on the way the baryon supply is shut off from the ISM into the IGM passing through the CGM.

In order to corroborate these theories it is necessary to do an extensive study of the gas in the CGM. During the last years there has been a lot of work in this area, specially with advances in technology. From an observational point of view, there are bigger telescopes capable of detecting fainter objects, higher resolution spectrographs, which are perfect to the study of the CGM, which is very diffuse gas. With these new kind of observations comes the necessity to create new models, capable of explaining them.

However, a theoretical framework is needed to put the wealth of precise observations of the CGM in context. The main motivation for the present thesis comes in this context, in which there seems to be few models to account for the large quantity of data available. Models allow one to understand the processes occurring in the CGM which is fundamental to understand galaxy evolution.

In the following sections I provide a summary of what we already know about the CGM (Section 1.2), ways to observe the CGM (Section 1.3) and the content of the present thesis work (Section 1.4).

1.2. The circumgalactic medium

The CGM is usually defined as the gas located outside the stellar disk but inside the virial radius, where the virial radius corresponds to the radius of a spherical volume which is sufficiently overdense to start collapsing to form a structure (see Figure 1.1) (Tumlinson et al., 2017). Thus, it can be seen as the link between the IGM and the ISM, so if any material exchange occurs between these, it necessarily needs to pass through the CGM. Even though models exist capable of explaining the exchange of gas and stellar mass without defining a CGM (Bouché et al., 2010; Lilly et al., 2013; Dekel & Mandelker, 2014; Somerville & Davé, 2015), nowadays most galaxy evolution theories require some kind of material flow to explain the observations (Voit et al., 2015).

The CGM is known to be multi phase, meaning that different physical conditions (density and temperature) are derived from observations of species with different ionization potentials. These phases can be separated in three main groups:

- The hot CGM: $T > 10^6 K$, highly ionized gas. It has been observed through emission lines (Anderson & Bregman, 2010; Anderson et al., 2013; Bregman et al., 2018; Nicastro et al., 2018) and absorption (Williams et al., 2005; Gupta et al., 2012; Nicastro et al., 2002).
- The cool CGM: $T \sim 10^4 K$, moderately ionized gas. Mainly detected in absorption (Tumlinson et al., 2013; Martin et al., 2012; Chen et al., 2010a), but also in emission (Rubin et al., 2011; Zhang et al., 2018). **This thesis focuses on this phase using observations of Mg II absorption systems towards background quasars.**
- The cold CGM: $T \sim 10^2 K$, mostly neutral gas. It has been observed in absorption (Steidel et al., 2010; Rudie et al., 2012; Werk et al., 2013; Turner et al., 2014)

The CGM contains signatures of gas flow that drives the evolution of galaxies, this is why there has been increasing interest in studying it, both observationally and theoretically, especially with the advent of high-resolution spectrographs on 8-10m class telescopes. However, the spatial and

kinematic structures of the CGM are still a topic of debate in the astronomical community, which are the topics of the present thesis. The next sections address previous works, both on observations and models, and give a general picture of the state of the art on this topic.

1.2.1. Baryon cycle

Currently, most theories assume that star-formation in galaxies is regulated by gas flows going in and out at galactic scales (Tumlinson et al., 2017; Schaye et al., 2015; Oppenheimer et al., 2016). Gas flowing into the galaxies provides the material necessary to maintain observed star formation rates, while gas flowing out of galaxies regulates star formation.

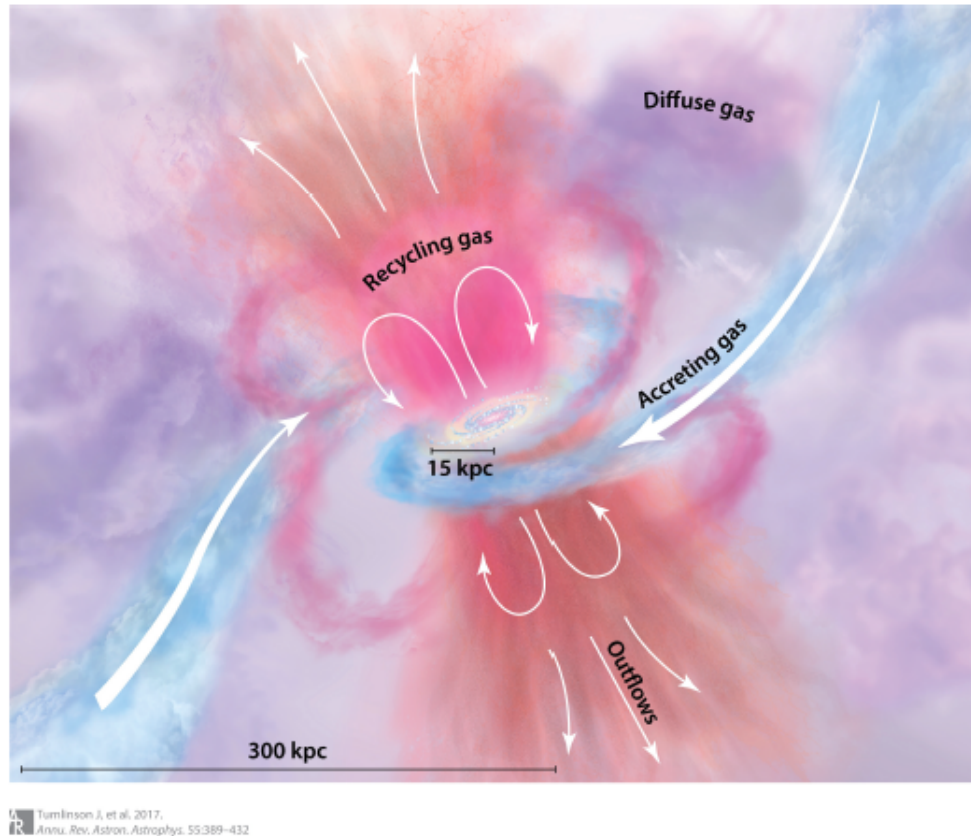


Figure 1.1: Artistic representation of the current state of the art on the baryon cycle. In orange we can see the outflowing structures and in pink we can see how some of the outflowing material falls back onto the galaxy. In blue we can see the gas being accreted into the ISM. In purple we see all the diffuse gas that composes the CGM. Figure from Tumlinson et al. (2017)

The presence of inflows and outflows could lead to a process of gas recycling. This refers to gas that has been expelled from the galaxy by outflows and can turn into inflowing gas. Simulations suggest that this process is fundamental to explain galaxy evolution (Oppenheimer et al., 2010; Ford et al., 2014; Christensen et al., 2016; Muratov et al., 2017) more so, it suggests that at $z=0$, around 60% (Ford et al., 2014) of the gas that forms stars was in the CGM before, suggesting that the matter responsible for forming stars comes from recycled gas in the CGM (Figure 1.1 in pink).

To date, substantial evidence of outflows exists (see Veilleux et al. (2005), for a review). This structure, as the name suggests, is responsible for expelling gas away from the galactic disk and thus regulating star formation. The expelled gas could be responsible for enriching the CGM and even the IGM; without such mechanism it would be impossible to explain the presence of metals in the IGM (Hummels et al., 2013; Ford et al., 2013). Observations suggest that these outflows are preferably located along the minor axis of the galaxies, suggesting a biconical outflow model (Martin et al., 2012; Rubin et al., 2014). Figure 1.1 shows a simplified model of the CGM; it can be seen that outflows, in orange, can reach a large distance from the galaxy.

On the other hand, in order to form stars, galaxies need to get fresh gas from the IGM; this process is called accretion or inflows. Simulations (Kereš et al., 2005) have shown that the morphology of accreted gas occurs in filaments of cold gas, parallel to the galactic disk, which is where star formation occurs. Figure 1.1 shows this structure in blue. The problem with this model has been its limited observational prediction power. The gas is quite diffuse and sometimes can be obscured by outflows. For instance, using Mg II there have been evidence of larger velocity dispersion for edge-on galaxies along the major axis (Rubin et al., 2012; Nielsen et al., 2015), which might be associated with cold accretion along the disk. However, to better understand the process of accretion it is necessary to develop models that constrain other effects, like disk rotation, the focus of this thesis.

The goal of the present thesis is to create a morpho-kinematic model for the cool gas in the CGM. The model consists of a rotating disk, as suggested by some authors (Prochaska & Wolfe, 1997; Haehnelt et al., 1998; Ledoux et al., 1998) and is constrained by observational data (details of which are described in the next section).

1.3. Advances in CGM science used in the present thesis

Emission line maps of the CGM are still challenging to get, because they look for photons directly emitted by the CGM, which has very low densities, $n_H \sim 10^{-2} \text{ cm}^{-3}$ or less, and emission scales as n^2 . However, another technique to observe the CGM is using absorption against bright background sources, like QSOs. This offers some advantages over emission line studies: (1) It is sensitive to very low column densities, $N \simeq 10^{12} \text{ cm}^{-2}$, (2) It can access a wide range of densities, unlike emission line measurements and (3) detection limits do not vary with host galaxy redshift or luminosity.

1.3.1. CGM in absorption

QSO absorption lines provide the majority of information available today to understand the different structures in the CGM (Lanzetta et al., 1995; Chen et al., 1998; Churchill et al., 2000; Nielsen et al., 2013; Prochaska et al., 2017; Chen, 2017). This powerful technique takes advantage of the QSO light travelling to the Earth and being absorbed by the gas present in the intervening CGM of galaxies close in projection (Figure 1.2). Thus, such QSO absorption line technique is free of the usual dimming bias affecting galaxies in emission. Moreover, QSOs are good background sources because of their extreme luminosities, relatively flat spectral slopes, and relatively featureless intrinsic spectra. This said, the main limitation for this technique is that it provides a one-dimensional, projected, pencil-beam measure of the intervening gas, usually limited to one

line-of-sight per galaxy given the rarity of background quasars.

Some efforts have been done to overcome this challenge. One way is to use extended background sources, like background galaxies (Rubin et al., 2010; Steidel et al., 2010; Lee et al., 2014; Cooke & O’Meara, 2015) and more recently gravitational arcs (Lopez et al., 2018, 2020; Tejos et al., 2021). These kinds of studies offer the advantage of having several sightlines per source, which allows one to get a better understanding of the spatial structure of the CGM in a single galaxy. However, with extended sources it is more difficult to get a sufficient SNR to get reliable absorption detections.

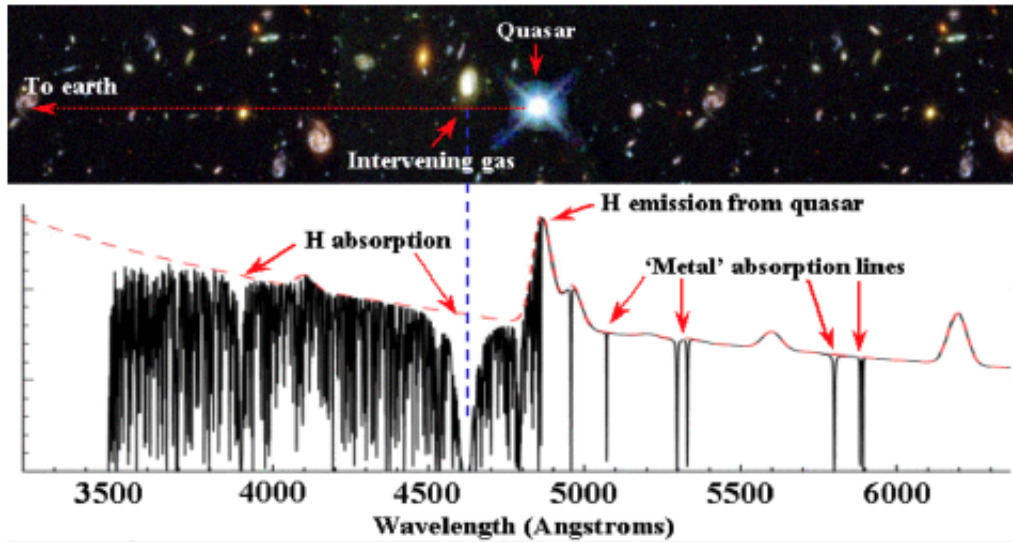


Figure 1.2: Schematic representation of the QSO probing CGM technique. At the top is a representation on how the light of the quasar travels towards the Earth and encounters intervening gas in its way. At the bottom we see the QSO spectra with several absorption resulting from the intervening gas. Most of the redshifted transitions are by neutral hydrogen from the inter-galactic medium (forming the so-called “Lyman-alpha forest”). Some of them, though, are by metal lines, revealing enriched material. Figure by Joe Liske.

Another technique used to study the CGM in absorption is using the absorbing galaxy as background source (so called "down the barrel" absorption). This technique uses the light of the own absorbing galaxy to probe its own halo (Martin, 2005; Kornei et al., 2012; Bordoloi et al., 2011; Rubin et al., 2014; Heckman et al., 2015; Kacprzak et al., 2014). The disadvantage of this method is that the exact position of the absorber is uncertain, it could be at any distance from the center of the galaxy. The same could be said about QSO absorption lines and extended sources, however we can measure the projected distances, which gives us a proxy for real distance.

Thus, every technique to study the CGM has its advantages and disadvantages. In this thesis I make use of the large amount of data available in QSO absorption line studies to constrain a morpho-kinematic model of the CGM. I create synthetic observations of Mg II absorption lines, given the model, and statistically compare these with real Mg II absorption lines in the spectra of QSO.

1.3.2. Mg II absorption data

Mg II is a very commonly used species to trace cool CGM gas (Bergeron & Boissé, 1991; Steidel et al., 1994; Churchill et al., 2005; Chen et al., 2010a; Kacprzak et al., 2011). This species probe photoionized gas with $n_H \approx 10^{-1} g cm^{-3}$ and $T \approx 10^{4.5} K$. Mg II absorption is easily observed in ground based telescopes in the optical range at redshifts $0.1 < z < 2.5$, it has high cosmic abundance and the resonant transitions Mg II 2796 and Mg II 2803 have large oscillator strengths with 2-to-1 ratio, thus they are easy to identify in QSO spectra.

Figure 1.3 shows examples of Mg II($\lambda 2796$) absorption systems at high resolution (Keck/HIRES spectrograph, $R = \lambda/\Delta\lambda = 45,000$) QSO spectra (Churchill & Vogt, 2001). They are plotted in the rest-frame velocity scale, which converts the wavelength scale into a radial velocities according to the Doppler effect around an arbitrary systemic redshift at $v = 0 \text{ km s}^{-1}$. Each absorption system has vertical lines above, indicating the position of individual absorption components, thus each system is composed of several individual absorbers that are at different velocities.

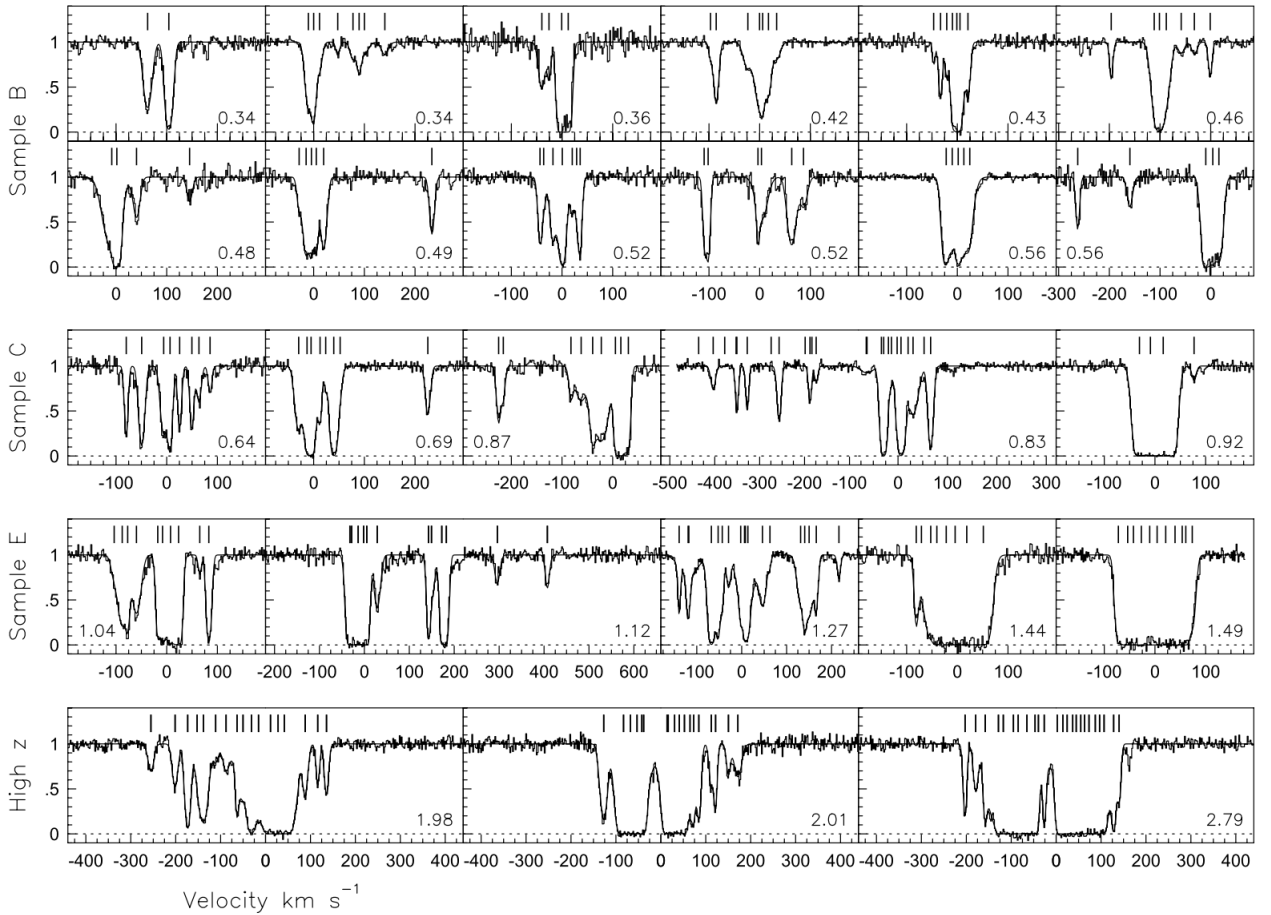


Figure 1.3: A gallery of $\text{Mg II}(\lambda 2796)$ profiles as a function of rest-frame velocity. The profiles are separated into subsamples (B, C, E and High redshift) and ordered by increasing equivalent width. The rest-frame equivalent widths are labeled in the lower corners of each subpanel and in \AA units. Figure from Churchill & Vogt (2001).

By using large surveys of Mg II absorption lines one could statistically study CGM properties. There are two main ways in which Mg II absorption catalogs for high redshift galaxies are built nowadays. The first type of catalogs selects bright QSOs with known Mg II systems first and then looks for foreground galaxies at similar Mg II redshifts, which are potentially responsible for the absorption (Guillemin & Bergeron, 1997; Steidel et al., 1994; Kacprzak et al., 2011). Such studies commonly target the brightest galaxies close in projection to the QSO sightline and therefore may result in biased galaxy populations. The second type of catalogs selects galaxy-QSO pairs without prior knowledge of the presence (or not) of Mg II absorption lines, thus producing unbiased samples (Chen et al., 2010a; Johnson et al., 2015; Huang et al., 2016; Lan & Mo, 2018; Martin et al., 2019).

The Mg II Absorber-Galaxy Catalog (MAGiCAT)

Nielsen et al. (2013) presented a compilation of 182 galaxies (with spectroscopic redshifts) with Mg II absorptions from their CGM within projected distances of 200kpc from background QSOs,

called MAGiiCAT. This work is a compilation of several previous works (Steidel et al., 1994; Churchill et al., 1996; Guillemin & Bergeron, 1997; Steidel et al., 1997; Chen & Tinker, 2008; Barton & Cooke, 2009; Chen et al., 2010b; Kacprzak et al., 2010; Gauthier & Chen, 2011; Kacprzak et al., 2011; Churchill et al., 2013), thus it mixes different selection criteria and observations with a variety of spectral resolutions. The goal of MAGiiCAT is to have a large compilation of Mg II absorption lines, where the host galaxy properties are standardized in the same system. They do so by relying on the published observations and calculate new values according to the newest cosmology. Table 1.1 shows a summary of properties of their sample.

Table 1.1: MAGiiCAT properties. $W_r(2796)(\text{\AA})$: rest frame equivalent width of Mg II(2796). z_{gal} : redshift of the absorbing galaxy. $d(kpc)$: impact parameter of the sightline with respect to the galaxy. M_B : absolute magnitude in the B-band. M_K : absolute magnitude in the K-band. L_B/L_B^* : B-band luminosity. L_K/L_K^* : K-band luminosity. $B - K$: rest-frame color.

Property	Min value	Max value	Mean value	Median value
$W_r(2796)(\text{\AA})$	0.003	4.422	0.629	0.400
z_{gal}	0.072	1.120	0.418	0.359
$d(kpc)$	5.4	193.5	61.1	48.7
M_B	-16.1	-23.1	-20.3	-20.4
M_K	-17.0	-25.3	-21.9	-22.0
L_B/L_B^*	0.017	5.869	0.855	0.611
L_K/L_K^*	0.006	9.712	0.883	0.493
$B - K$	0.04	4.09	1.54	1.48

The MAGiiCAT team have six papers published where they measure different Mg II systems properties and their dependence with galactic properties. In this thesis I use the data published by MAGiiCAT to constrain the rotating disk model of CGM. In particular I use their equivalent width/impact parameters database, their pixel velocity two-point correlation function, and their range of galaxy properties.

1.3.3. Equivalent width

One way to analyze Mg II absorption lines in the spectra of QSO is by measuring the rest frame equivalent width (W_r), which provides a measurement of the strength of absorption, independent of the spectral resolution and redshift. It is defined as the area of an absorption, under the normalized continuum or as the width of a box from 0 to the continuum flux that has the same area as the spectral line divided by $(1 + z_{abs})$ (See figure Figure 1.4). If the absorbing clouds ("velocity components") are resolved, this quantity depends on the number of atoms present in the absorbing gas, which is characterized by the column density (i.e., the "curve of growth"; more on column densities in Section 2.2 where I define how to model the absorption lines). On the other hand, in low-resolution spectra W_r depends strongly on the line-of-sight velocity structure of the (unresolved) individual components (See Figure 1.3).

Several studies show an anti-correlation between W_r and impact parameter (d), i.e., the projected distance from the QSO to the galaxy (Lanzetta & Bowen, 1990; Bergeron & Boissé, 1991; Bouché

et al., 2006; Kacprzak et al., 2008; Chen et al., 2010a; Churchill et al., 2013). However there is a very large scatter in this relationship. The source of this scatter could be due to: galaxy luminosity (Kacprzak et al., 2008; Chen et al., 2010a), stellar mass or specific star formation rate (Chen et al., 2010b), morphology (Kacprzak et al., 2007), or geometry and orientation (Bouché et al., 2012; Kacprzak et al., 2011). In Churchill et al. (2013) they study the dependence of W_r on galaxy halo mass. Using MAGiCAT, they were able to establish that lower mass halos tend to produce lower W_r than higher mass halos. Given that the mass is a big source of scatter in the $W_r - d$ relationship they decided to measure W_r vs. d/R_{vir} , since $R_{vir} \propto M^{1/3}$. They obtained the values of R_{vir} for each galaxy in the sample using halo abundance matching and found out that the scatter was significantly reduced in the $W_r - d/R_{vir}$ plane.

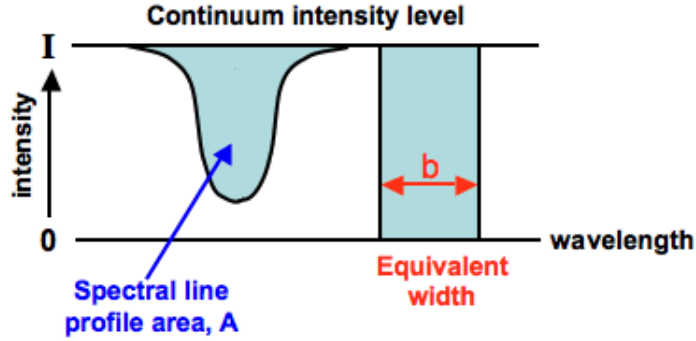


Figure 1.4: Schematic picture of the two definitions of equivalent width. On the left there is an absorption profile with the area under the continuum colored with blue, this area would be the equivalent width if the flux is normalized. On the right there is a rectangle with the same area, the width of this rectangle would be the W_r . Picture from COSMOS - The SAO Encyclopedia of Astronomy

Figure 1.5 (from Churchill et al. (2013)) shows W_r values for the sample in MAGiCAT with respect to the impact parameter, d (panel (a) and (b)) and d/R_{vir} (panel (c)), with different colors for halo mass bins. They performed a two-dimensional Kolmogorov-Smirnov (KS) test on the two samples in panel (b) and ruled out the hypothesis that the two samples came from the same distribution. In panel (c) they plotted the W_r vs. d/R_{vir} relationship and performed a BHB- τ test to find out that W_r is anti-correlated with d/R_{vir} at a 8.9σ significance.

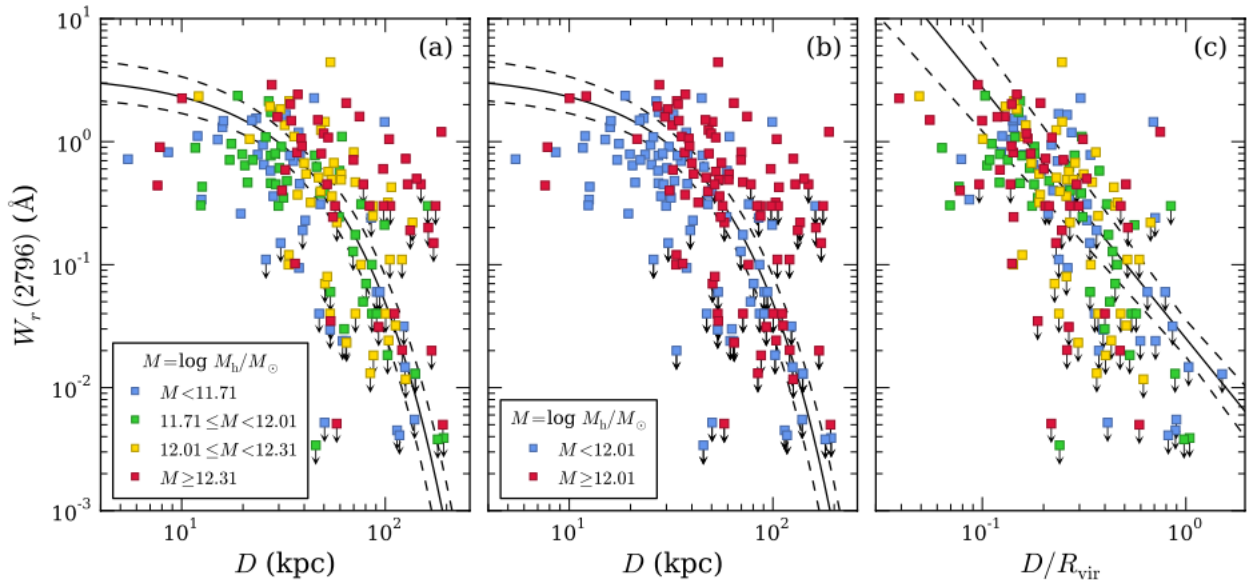


Figure 1.5: Equivalent width vs. impact parameter for sample in MAGiCAT. Arrows show the upper limits of the sample, solid curve shows the fit and dashed curve shows the 1σ envelope. (a) W_r vs. d for four different mass bins. (b) W_r vs. d for two different mass bins. (c) W_r vs. d/R_{vir} for the same four mass bins as panel (a). Figure from Churchill et al. (2013)

In this thesis work I use these results on W_r to constrain the model. The goal is to determine if the simple morpho-kinematic model presented in this work is able to reproduce the same results.

1.3.4. Velocity Two-point correlation function

Due to the limitations in interpreting W_r in terms of the kinematics signature, many works study the kinematics of Mg II absorption lines by calculating the two-point correlation function (TPCF) (Sargent et al., 1988; Petitjean & Bergeron, 1990; Churchill et al., 1996; Churchill & Vogt, 2001; Evans, 2011). Nielsen et al. (2015) uses the pixel velocity TPCF, which is a measurement of the line-of-sight velocity dispersion of the absorbers and is computed by using the velocities of pixels in regions of the spectrum where Mg II absorption is formally detected.

Nielsen et al. (2015) use MAGiCAT data from which there is high resolution spectra available (Keck/HIRES and VLT/UVES) to compute the pixel TPCF. The authors separate that sample in different sub-samples according to azimuthal angle (α) and galaxy inclination (i). Both of these properties are important to define where the absorber is located around the galaxy. By calculating the TPCF for this sub-samples Nielsen et al. (2015) find differences in the velocity dispersion (Figure 1.6) that could give us better understanding of the kinematic structures present in the CGM.

Table 1.2 shows the cut values Nielsen et al. (2015) use to define their sub-samples and the amount of Mg II systems in each of them.

Table 1.2: Sub-samples used in Nielsen et al. 2015. α is azimuthal angle and i is galaxy inclination

Sub-sample	Number of galaxies	Cut value
Face-on	17	$i < 57^\circ$
Edge-on	13	$i \geq 57^\circ$
Major-axis	15	$\alpha < 45^\circ$
Minor-axis	15	$\alpha \geq 45^\circ$
Face-on & Major-axis	10	$i < 57^\circ$ & $\alpha < 45^\circ$
Face-on & Minor-axis	7	$i < 57^\circ$ & $\alpha \geq 45^\circ$
Edge-on & Major-axis	5	$i \geq 57^\circ$ & $\alpha < 45^\circ$
Edge-on & Minor-axis	8	$i \geq 57^\circ$ & $\alpha \geq 45^\circ$

After having the different sub-samples Nielsen et al. (2015) defines the regions of absorption using the methods of Churchill & Vogt (2001). They pull out every velocity pixel in the absorbing region for every spectra in the sub-sample and calculate the absolute velocity difference for each possible pixel velocity pair. They bin up the pixel-velocity separations and normalize each bin by the total number of pixel pairs in the sub-sample, thus the TPCF is a probability distribution. Each bin has a width of 10 km s^{-1} which correspond to roughly one resolution element in UVES.

Figure 1.6 shows the TPCFs for the different sub-samples in Nielsen et al. (2015). In the top row they compare the TPCF for face-on galaxies (blue) and edge-on galaxies (orange) at all azimuthal angles (panel(a)), along the minor axis (panel (b)) and along the major axis (panel (c)). In the bottom row they plot the TPCF for sightlines along the major axis (green) and along the minor axis (pink) for all inclination (panel (d)), for edge-on galaxies (panel (e)) and for face-on galaxies (panel (f)). The TPCF is represented by solid lines and the shaded regions represent the 1σ bootstrap uncertainties for 100 realizations.

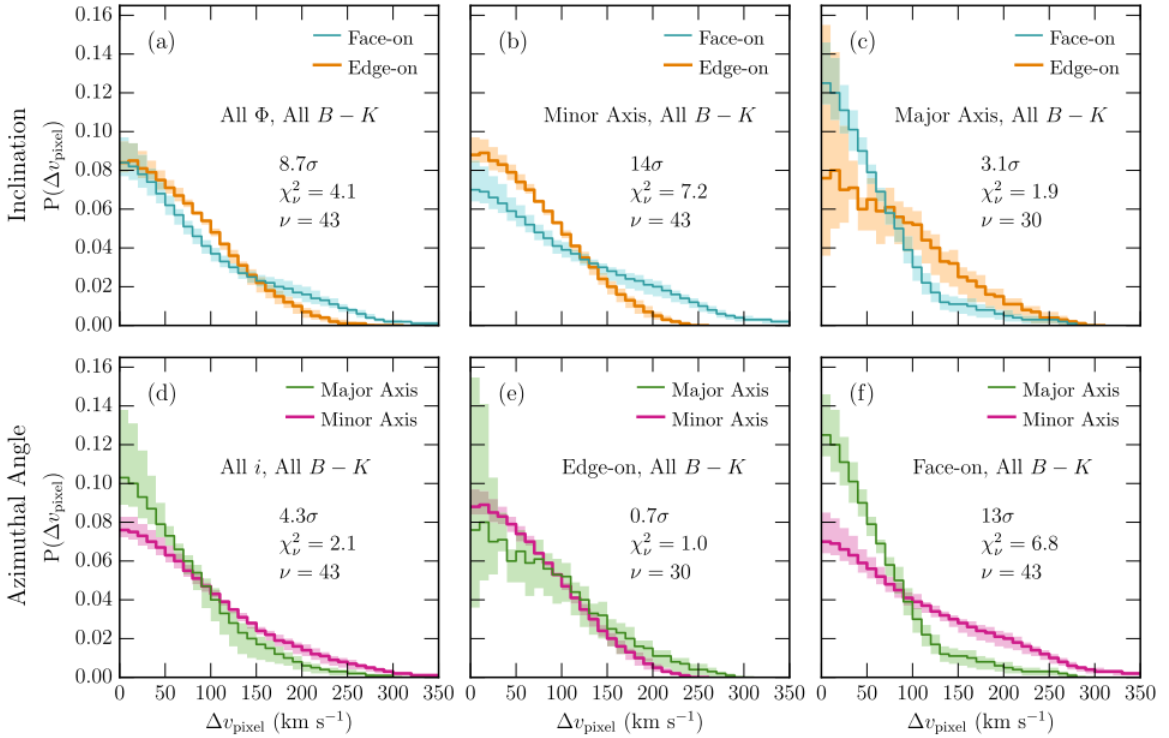


Figure 1.6: Pixel-velocity two-point correlation function (TPCF) for different inclinations and azimuthal angles. Panels (a)-(c) show the TPCF for edge-on ($i \geq 57^\circ$) and face-on ($i < 57^\circ$) galaxies with sightlines with all azimuthal angles, sightlines along the minor axis ($\alpha \geq 45^\circ$) and sightline along the major axis ($\alpha < 45^\circ$) respectively. Panels (d)-(f) show the TPCF for sightlines along the major ($\alpha < 45^\circ$) and minor ($\alpha \geq 45^\circ$) axis for galaxies with all inclinations, edge-on ($i \geq 57^\circ$) galaxies and face-on ($i < 57^\circ$) galaxies. All these panels include all galaxy colors, B-K. The solid lines represent the TPCFs and the colored regions are the 1σ bootstrap uncertainties. Each panel includes also a reduced χ^2 value of comparing the two TPCFs, and the degrees of freedom, ν . Figure from Nielsen et al. (2015)

Nielsen et al. (2015) also study the TPCF for sub-samples of blue and red galaxies, to see how the star formation rate could affect the velocity dispersion and study the column density values in each sample. In this thesis work I only examine the dependencies on α and i . They find a larger velocity dispersion for sightlines along the major axis of edge-on galaxies than for face-on galaxies (Figure 1.6 panel (c)), they associate this with rotating gas whose line of sight velocity is maximized for edge-on galaxies compared to face-on. They suggest that the velocity dispersion is larger for sightlines along the minor axis of face-on galaxies than edge-on (Figure 1.6 panel (b)) because they are observing bi-conical outflows into the flow as opposed to across, which will produce more dispersion.

In summary Nielsen et al. (2015) indicate that the differences in the TPCF of edge-on vs. face-on galaxies with sightlines along the major vs. minor axis are due to two main kinematic structures. One is the rotating material along the disk plane of the galaxy and the other one is outflowing material in the form of bi-conical outflows. The morpho-kinematic model presented in this thesis consists of a single rotating disk with a vertical velocity gradient (See Chapter 2 for a full descrip-

tion of the model) and it does not include any form of outflows.

1.4. This Thesis

In this thesis I present a morpho-kinematic model for the CGM. Using this model I create catalogs of synthetic Mg II absorption lines and compare them with the MAGiCAT observational data. The model consists of an extended and rotating disk of gas that is aligned with the galactic disk of the host galaxy. The extended disk is populated with individual Mg II absorbing clouds having a cloud distribution given by the filling factor. The rotation of such clouds proceeds in the plane of the disk and has a vertical velocity gradient that makes clouds on the edge of the disk rotate slower than clouds in the disk mid-plane. Model parameters are constrained by comparing synthetic and observed equivalent width vs. impact parameter relation via MCMC minimization. In a second stage, the fitted model produces a TPCF that is also compared with MAGiCAT, but in a qualitative fashion. By comparing the models TPCF with the results in Nielsen et al. (2015) I aim to see if the rotating disk model is enough to explain the velocity dispersion we observe.

The goals of this thesis are:

1. Find whether a disk populated with individual Mg II clouds can reproduce QSO absorption line observations.
2. Find whether a single rotating component in the CGM can reproduce the Mg II QSO absorption line statistics.
3. Fit the distribution of Mg II clouds and their filling factor to infer the covering factor of the CGM.
4. Fit other model parameters like cloud size, the height of the disk and the velocity scale factor.

The rest of the thesis is structured as follows: In Chapter 2 the CGM model, the creation of synthetic spectra and the creation of W_r catalogs are described. In Chapter 3 the methods used to constrain the model against QSO statistics are shown. The results are presented in Chapter 4 and later discussed in Chapter 5. This last chapter also contains the main conclusions and an outline of possible future work.

Chapter 2

Modeled data

Currently there are many spectroscopic confirmations of galaxies believed to host Mg II absorption systems (Bahcall & Spitzer, 1969; Bergeron & Boissé, 1991; Lanzetta et al., 1995; Churchill et al., 2000; Werk et al., 2014; Prochaska et al., 2017; Chen, 2017; Dutta et al., 2020) in their CGM. In the current paradigm of galaxy evolution, this gas is believed to flow in and out of galaxies (Tumlinson et al., 2017; Schaye et al., 2015; Oppenheimer et al., 2016) (See Subsection 1.2.1). Some theories exist on how this process occurs, but there is no real consensus, especially in the process of fresh gas flowing into galaxies. One theory establishes that the gas is distributed on a disk that co-rotates with the galaxy (Prochaska & Wolfe, 1997), another one suggests that it has a spherical shape (McDonald & Miralda-Escudé, 1999). In this thesis work I explore the scenario of a rotating disk model, taking advantage of the large amount of Mg II absorption lines in the spectra of QSO sightlines to constrain it.

To constrain the model I create synthetic observational data with groups of different representative absorption lines that can be compared with observational statistics. The model creates several Mg II absorption lines which probe the CGM of different galaxy types, to simulate different properties found in observational catalogs. This chapter is divided into three different sections: a detailed description of the model morphology and kinematics (Section 2.1), the procedure to create individual spectra from the model (Section 2.2), and how to produce groups of representative absorption lines (Section 2.3).

2.1. The model

In this section I describe the morpho-kinematic model proposed in this work. This section is divided into three subsections: the geometry of the model, the filling factor of Mg II and the kinematics of the model.

2.1.1. Model geometry

The model geometry is based on a rotating disk. This choice is supported by several works suggesting that this is the preferred geometry to explain Mg II QSO observations close to galaxies (Steidel et al., 2002; Nielsen et al., 2015; Rubin et al., 2012; Christensen et al., 2016). The rotating disk geometry is also supported by the finding of gas at large impact parameters with velocity

corotating with the stellar disk (Barcons et al., 1995; Ho et al., 2017; Zabl et al., 2019).

Based on Steidel et al. (2002) and Ho et al. (2017), the model’s CGM disk is an extension of the stellar disk of the galaxy which has a circular symmetry in the plane of the galactic disc. When observing the galaxy perpendicular to this plane (face-on) we can see that the stellar disk has a circular form. On the other hand, when observing the galaxy very near the edge of the disk (edge-on), the stellar disk appears elliptical. This effect is defined by the inclination (i), which goes from 0° when the galaxy is face-on, to 90° when it is edge-on. For this model the CGM disk has the same inclination as its host galaxy (See Figure 2.1 and a list with all geometrical parameters in Table 2.1).

Another parameter to take into consideration is the thickness of the CGM disk (h) (see Figure 2.1). Steidel et al. 2002 show that, in order to reproduce velocity dispersions observed in Mg II absorption line systems, the disk needs to be thick ($\approx 30 - 40 kpc$) and/or have a vertical velocity gradient. One of the goals of this thesis is to explore these parameters in more detail and constrain them using the statistics of QSO absorption lines.

Since the objective of this model is to fit it to real QSO absorption line observations, it is also necessary to define a background, point-like source. In the 2D plane of the sky, the QSO sightline is seen like a point and its position relative to the galaxy is defined by two parameters: the *impact parameter* (d) defined as the projected distance from the center of the galaxy to the sightline position, and the *azimuthal angle* (α) which is the angle (measured clockwise) between the major axis of the galaxy and the line connecting the galaxy center and the sightline position (see Figure 2.1 and Table 2.1).

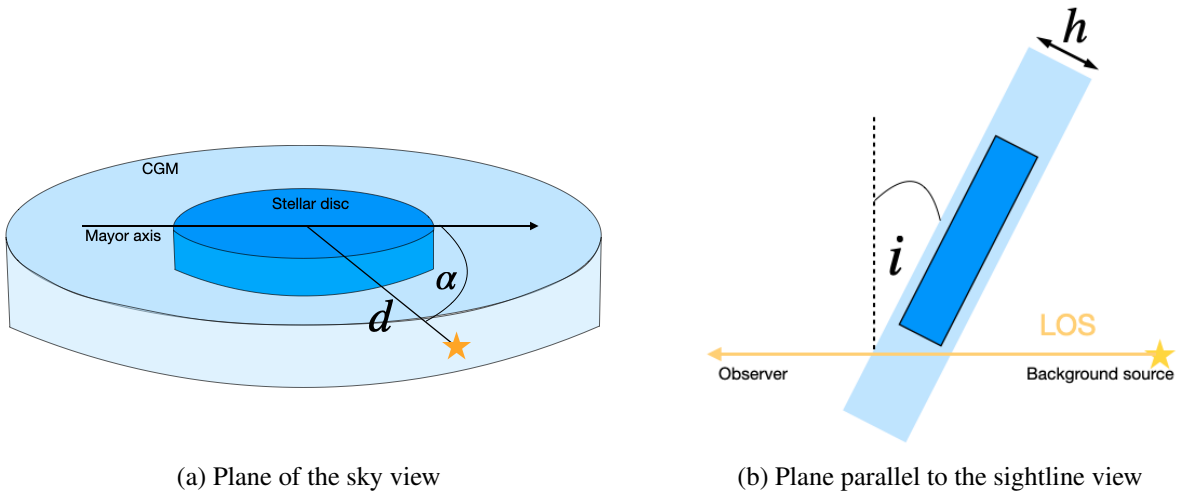


Figure 2.1: Model’s geometry, light blue represents the CGM disk and darker blue represents the stellar disk. The yellow star and arrow represent the QSO sightline; panel (a), black arrow represents the major axis of the galaxy. d is the impact parameter, which is the projected distance between the center of the galaxy and the sightline, α is the azimuthal angle, which is the angle between the major axis of the galaxy and the line connecting the galaxy centre and the position of the QSO on the sky, h is the thickness of the disk and i is the line-of-sight inclination of the galaxy.

To produce spectra that represent QSO sightlines in the 2D plane of the sky (d and α), as a result of sightlines passing through a 3D representation of the CGM, there needs to be a 3D framework to describe exactly where the sightline is crossing the gas, for any given CGM geometry. To this end I use a 3D coordinate system. (Figure 2.2). The point (0,0,0) of this coordinate system is located at the center of the disk. The x-axis is aligned with the mayor axis (but on top of the disk) and the positive side of the x-axis is defined to be the one that has a redshifted velocity. As shown by Figure 2.2 (a), the y-axis lays in the plane of the disk and Figure 2.2 (b) shows that the z-axis is perpendicular to the plane of the disk.

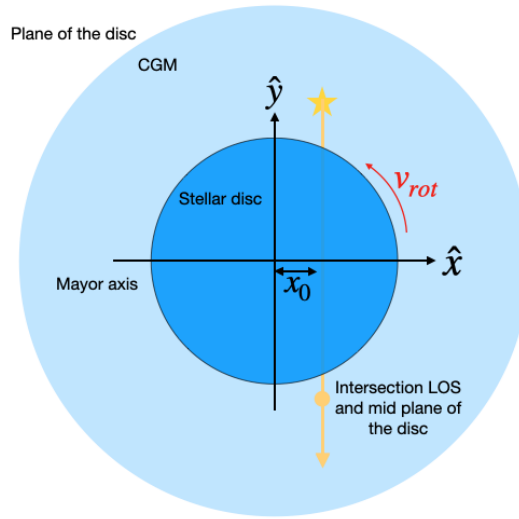
Instead of modelling the entire disk and then creating spectra, the model is only computed where the sightline intercepts the CGM. This saves a wealth of computational time. Given the observable parameters in the plane of the sky, d and α ; the geometrical parameters i and h ; and the coordinate system, it is possible to define the intersection between the sightline and the disk. Since the sightline is perpendicular to the x-axis, by construction, the sightline position varies only along the y-axis and z-axis, and is constant in the x-axis with a value of x_0 (see Figure 2.2) (a). The coordinates (y,z) where the sightline enters and leave the disk can be calculated with the following formulas:

$$y_{min} = y_0 - \frac{h}{2} \tan(i) \quad (2.1)$$

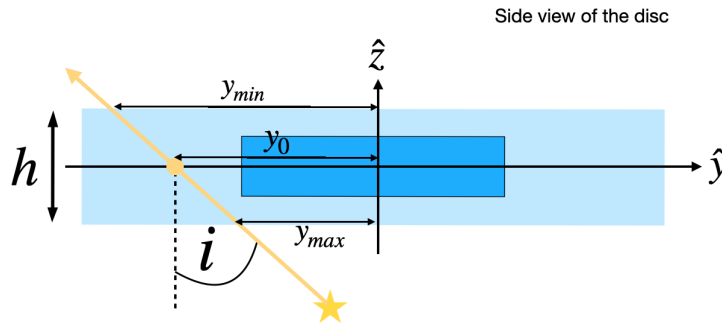
$$y_{max} = y_0 + \frac{h}{2} \tan(i) \quad (2.2)$$

where y_0 is the perpendicular distance from the mid-plane $z = 0$ to the QSO sightline intersection. y_0 can be calculated as:

$$y_0 = \frac{d \sin(\alpha)}{\cos(i)}. \quad (2.3)$$



(a) x-y plane (Plane of the disk)



(b) y-z plane (Plane parallel to the sightline view)

Figure 2.2: Coordinate system of the model, light blue represents the CGM disk and darker blue represents the stellar disk. Fig. (a) shows a sketch of the model in the x-y plane or the plane of the disk. x_0 is the x-coordinate where the sightline intercepts the disk. v_{rot} shows the direction of rotation of the stellar disk. Fig. (b) is a sketch of the model in the y-z plane or side view of the disk. y_{max} and y_{min} are the y-coordinates where the sightline enters and leaves the disk; and y_0 is the y-coordinate where the sightline intercepts the mid-plane of the disk.

In Table 2.1 there is a list of all geometrical parameters that were defined in this section. With this 3D coordinate system in place, it is now possible to insert Mg II clouds into the disk model.

Table 2.1: Geometrical parameters of the model

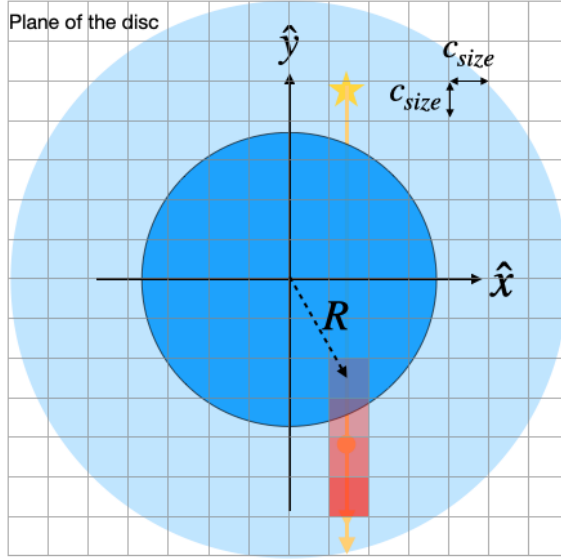
Model parameters	
Geometrical Parameters	Inclination (i)
	Height (h)
	Impact parameter (d)
	Azimuthal angle (α)

2.1.2. Mg II clouds

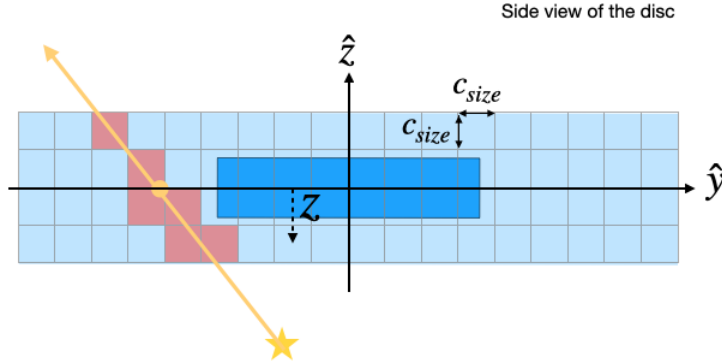
It is known from the literature, specially from absorption line studies (Werk et al., 2013, 2014; Tumlinson et al., 2013), that the cool CGM is not homogeneous, but is in fact composed of individual absorbing gas clouds with different velocities (see Figure 1.3). This has also been observed in the Milky Way and nearby galaxies (see Putman et al. (2012) for a review). A very important parameter to understand how the CGM affects the galaxy is the distribution of these clouds in space. My model assumes a disk morphology and in this section I explain how the gas clouds are distributed inside the CGM disk.

Using the connection between the geometrical parameters and the physical coordinates of the disk (Table 2.1), the Mg II clouds that constitute the CGM need to be included in the model following a method that reproduces the observational structure of the CGM. As proxy for cloud "sizes", I use resolution elements. The disk is separated into a 3D quadrangular grid where each cell is a cube with edge size equal to c_{size} in kpc units. Using Equation 2.1 and Equation 2.2 it is possible to calculate all cells that are crossed by a sightline.

Figure 2.3 shows a schematic picture of how this grid system looks like. In red are all the resolution elements that are probed by the sightline and therefore, where the model is computed. Figure 2.3 (a) is a representation of the grid system in the plane of the disk. From this perspective all cells that are probed by the sightline have x-coordinate equal to x_0 and, depending on their z-coordinate and y-coordinate, they are crossed or not by the line of sight. This effect is shown in the side view of the disk in Figure 2.3 (b).



(a) x-y plane (Plane of the disk)



(b) y-z plane (Plane parallel to the QSO sightline view)

Figure 2.3: Schematic picture of the grid used in the model. The volume of each grid cell is $c_{size} \times c_{size} \times c_{size}$. Red cells are the cells that are crossed by the sightline, R is the 2-d distance from the center of the disk to the center of each cell in the projected plane of the disk and z is the perpendicular distance from the mid-plane of the disk ($z=0$) to the center of each cell.

Observations of QSO sightlines crossing the CGM of galaxies, inform us that not all sightlines that cross near a galaxy produce absorption (Chen et al., 2010b; Huang et al., 2021). This feature is commonly quantify in the literature by the "covering fraction" (f_{cov}), which is defined as the fraction of sightlines that have detected absorption versus the total, within a given impact parameter bin. Given that observations are noisy, to define an absorption there needs to be a minimum detection threshold. Above a certain equivalent width value, the sightline is considered to have an absorption and below it, it is considered to have no absorption. This quantity is purely observational, but it tells us something about the clumpiness of the CGM.

Thus, f_{cov} is not always unity (Chen et al., 2010b; Huang et al., 2021). This implies that the Mg II gas is clumpy, with some volumes devoid of detectable absorption at. given threshold. f_{cov} is a purely observational parameter. The physical parameter that describes clumpiness is the volumetric filling factor (f_v), which is defined as the percentage of volume in the CGM that actually produces

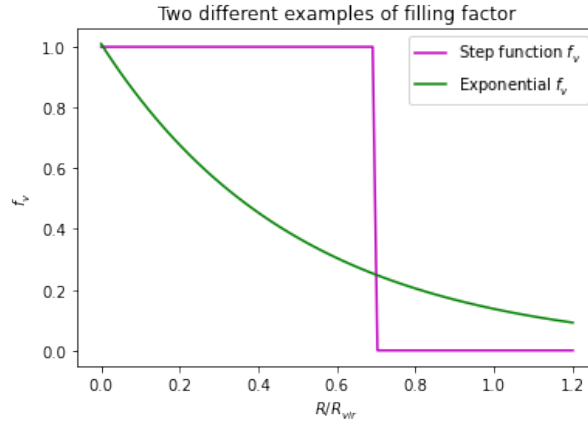
absorption. In some cases a sightline crosses a volume of the CGM filled with Mg II and in other cases it crosses an empty volume, explaining why the observable f_{cov} is not always unity (Steidel et al., 2002).

Observations suggest that f_{cov} depends on the impact parameter. Consequently, the present model assumes that f_v depends on the (3D) distance to the galaxy, R (see Figure 2.3 (a)). In reality, f_v probably depends on the z position as well but this dependence is much harder to test with observations and it would add more free parameters to the model. There also seems to be a relationship between f_{cov} and halo mass of the galaxy (Churchill et al., 2013); therefore, the model f_v depends on R scaled to virial radius in order to compare galaxies of different masses in the model. Thus f_v in the model is a function of R/R_{vir} .

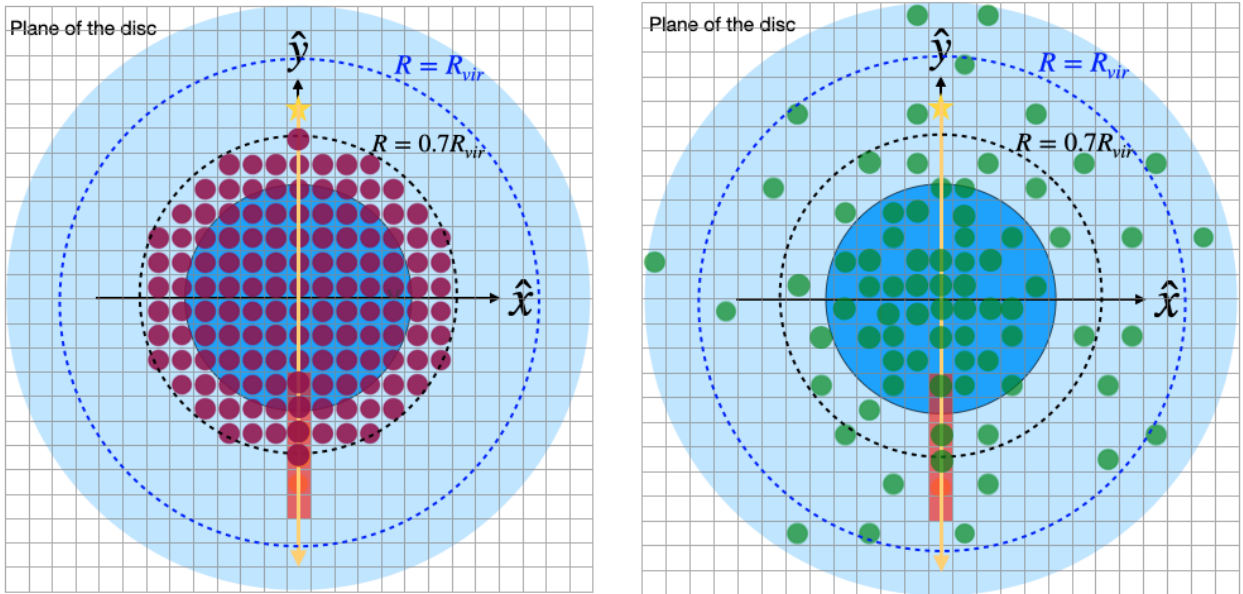
The model allows one to control the physical parameter $f_v(R/R_{vir})$. Looking at Figure 2.3, each of the crossing cells (red cells) have two possible states; filled with Mg II gas or completely empty. The state of each cell depends on f_v , if $f_v = 1$, at a certain R/R_{vir} distance, all cells are filled with Mg II , if $f_v = 0.5$, only half of the cells are filled with Mg II . It is important to mention that since the model is defined in a 3D Cartesian grid, the volume unit in this case is a cylinder of inner radius $R - c_{size}$, outer radius $R + c_{size}$ and height h . So, $f_v(R/R_{vir})$ is actually a probability function that defines the probability of having a Mg II cloud in a cell that is located at R/R_{vir} .

Each of the cells crossed by the sightline is randomly filled with a cloud using $f_v(R/R_{vir})$. This approach of *randomly* filling cells according to a probability function produces different absorption profiles for the same set of parameters in the model, in the same fashion as we do not expect the same absorption profiles in two galaxies of the same mass at the same impact parameter. Thus this method is very useful to constrain a model with QSO statistics, because it adds some randomness that we expect in real observations.

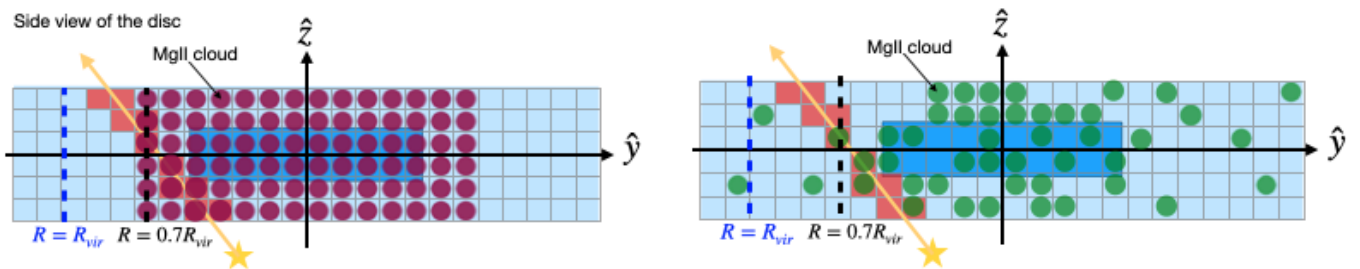
Figure 2.4 shows two examples of how the cells are populated according to different $f_v(R/R_{vir})$ values. Figure 2.4 (a) shows a plot of the two different f_v vs. R/R_{vir} , one is an exponential and the other one is a step function with $f_v = 1$ for $R/R_{vir} \leq 0.7$. Panel (b) shows a schematic picture on how would the plane of the populated disk look like with the two examples of $f_v(R/R_{vir})$ and panel (c) shows a side view of the two different disks. This figure shows two possible distributions for Mg II clouds inside the disk. In Chapter 4 I explore how different distributions result in different statistics for QSOs and I present a function that is capable of reproducing the QSO statistics.



(a) Two different examples of possible filling factors.



(b) Plane of the disk view of two models with different filling factors.



(c) Side view of the disk of two models with different filling factors.

Figure 2.4: Schematic picture of two different $f_v(R/R_{vir})$. (a) Shows a plot of the two different filling factors, a step function $f_v(R/R_{vir})$ (pink) and an exponential $f_v(R/R_{vir})$ (green). (b) Schematic picture of how would the plane of the disk look like with the two different $f_v(R/R_{vir})$. Circles are used to show which cells are filled with Mg II. Black dotted circles show a distance of $R = 0.7R_{vir}$ from the center of the galaxy and blue dotted circles show a distance of $R = R_{vir}$ from the center of the galaxy. Left panel (with pink circles) shows the CGM disk for the step function $f_v(R/R_{vir})$, right panel (with green circles) shows CGM disk for an exponential $f_v(R/R_{vir})$. (c) Schematic picture of how the side of the CGM disk would look like with the two different filling factors. Right picture is for the step function filling factor and left is for the exponential.

Table 2.2 shows a list of the parameters that define the process of placing the individual Mg II clouds in the CGM disk.

Table 2.2: Cloud parameters

Model parameters	
Mg II clouds	Grid size (c_{size})
	Virial radius (R_{vir})
	Filling factor ($f_v(R/R_{vir})$)

2.1.3. Cloud kinematics

Clearly, the velocity field of the individual clouds affects the final Mg II absorption profile. After placing the clouds within the grid, it is necessary to calculate their velocity. If a sightline crosses different clouds with different velocities the final spectra are composed of various absorption lines centered at different wavelengths. If the cloud is moving away from us, the cloud is redshifted with respect to the redshift of the host galaxy, and blueshifted if the cloud is moving towards us. We know from observations that in fact QSO sightlines cross different clouds with different velocities because we can observe different components at different wavelengths (see Figure 1.3).

The kinematic model used in this work is the same as in Steidel et al. (2002). In that model the velocity of the Mg II clouds depends on the measured circular velocity at the mid-plane (v_c) and the distance from the plane $z=0$ to the cloud (z). The rotational velocity of a cloud within the plane of the disk is given by:

$$v_{cir}(R, z) = v_c(R) \exp\left(\frac{-z}{h_v}\right), \quad (2.4)$$

where v_c is the velocity of rotation of the disk and h_v is the velocity scale height.

Observations of rotation curves in nearby galaxies show that the maximum velocity, v_{max} , is already reached at distances of the order of 100 pc away from the center of the galaxy (Sofue & Rubin, 2001). Thus, when probing the CGM (distances $7kpc \leq d \leq 200kpc$), the velocity of rotation v_c is equal to v_{max} . Thus the only parameter that affects the velocity of the clouds is the velocity scale height, h_v , which affects v_{cir} depending on the distance from the mid-plane of the disk to the cloud, z . Figure 2.5 shows v_{cir} in terms of v_{max} as a function of z/h , for three different h_v values. $z = 0$ represents the mid-plane of the CGM disk. It can be seen that clouds at higher distances from the mid-plane of the disk rotate slower and that this decrease is faster for lower values of h_v .

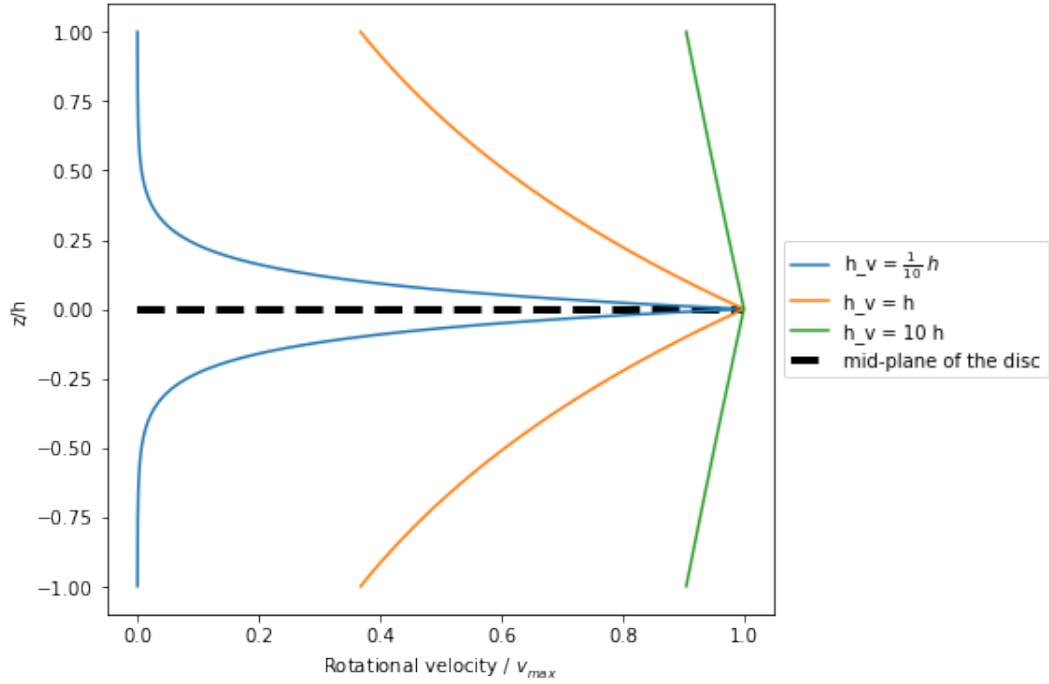


Figure 2.5: Plot of the circular velocity, in terms of the maximum velocity of rotation, of a cloud with respect to the distance to the mid-plane of the disk. y-axis shows z values, where $z=0$ is the mid-plane of the disk, black dotted line. Color lines show the velocities for three different h_v values.

Equation 2.4 defines the real velocity of the clouds in the plane of the CGM disk. In order to generate synthetic spectra the velocities of the clouds need to be projected along the sightline. Using Equation 2.4 and the transformations defined in Section 2.1 the projected velocity along the sightline (v_{LOS}) for a cloud is given by:

$$v_{LOS} = \frac{v_{max} \sin(i)}{\sqrt{1 + (y/d)^2}} \exp\left(\frac{-|y - y_0|}{h_v \tan(i)}\right) \quad (2.5)$$

where i is the inclination of the disk, y is the position of the cloud in the y -axis, d is the impact parameter of the QSO sightline, y_0 is the position in the y -axis where the sightline crosses the mid-plane (Equation 2.3) and h_v is the velocity scale height.

Table 2.3 shows a list with all the parameters that define the kinematics of the clouds. Given all parameters defined so far, it is possible to create the synthetic spectra.

Table 2.3: Velocity parameters.

Model parameters	
Cloud kinematics	maximum velocity of rotation (v_{max})
	velocity scale height (h_v)

2.2. Absorption line spectra

This section illustrates the process of creating synthetic spectra given the model described in the previous section.

An atomic absorption line is described by a Voigt profile. This profile is the result of two broadening mechanisms working together, Doppler broadening and pressure broadening, the final profile is a convolution of both (Mitchell, 1971). The Doppler broadening is modeled by a Gaussian profile and is produced by the random motion of particles that depends on the gas temperature and the velocity of individual atoms or molecules in the gas. The parameter that characterizes this effect in the final spectra is the Doppler parameter (b) and it is measured in km s^{-1} . On the other hand the final spectra also are affected by the quantity of absorbing material, this is called pressure broadening and it is modeled as a Lorentzian profile. This profile is parameterized by the column density (N), which is the number of absorbers per unit area, with unit cm^{-2} .

Figure 2.6 shows a Gaussian profile, a Lorentzian profile and a Voigt profile, all normalized to unity area and having $\text{FWHM} = 1$. The Gaussian profile shows a narrower shape than the Lorentzian and the Voigt profile is the convolution of both.

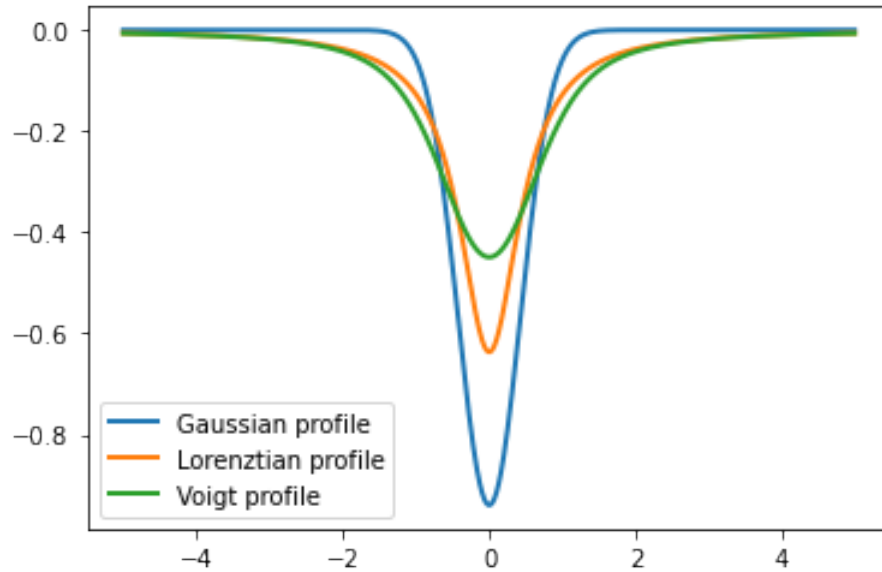


Figure 2.6: Plot of three different profiles that characterize an absorption line. Blue: Gaussian profile, which models the Doppler broadening of the random motion of the gas. Orange: Lorentzian profile that models the pressure broadening. Green: Voigt profile, which is the convolution of the two previous profiles and shows the form of the final absorption profile. All these profiles are normalized so that the area under the curve is equal to 1 and with $\text{FWHM} = 1$

In this thesis the procedure to generate the actual absorption profiles, is based on the software BayesVP for Voigt profile fitting (Liang & Kravtsov, 2017). There are three parameters that affect the shape of the absorption, N , b , and the redshift, z_{abs} , which indicates the line centroid. z_{abs} is usually a combination of two different redshifts, $z_{cosmological}$ and $z_{doppler}$. The first one is the shift produced in the wavelength of the absorption due to the elongation of the wave traveling through

an expanding space. On the other hand, $z_{doppler}$ depends on the motion of the object, in this case it shows the effect of v_{LOS} (Equation 2.5), which is the velocity of the individual clouds. In this model I assume $z_{cos} = 0$ and take into account $z_{doppler}$ only.

The spectrum can be normalized, which means that the flux is re-scaled so that the continuum (wavelengths at which no absorption is detected) the flux is equal to 1. All synthetic spectra in the model are normalized. The final flux is the exponential of the negative optical depth (τ):

$$F(\lambda|N, b, z_{abs}) = e^{-\tau(\lambda)} \quad (2.6)$$

where τ is modeled as a Voigt profile and is calculated as follow:

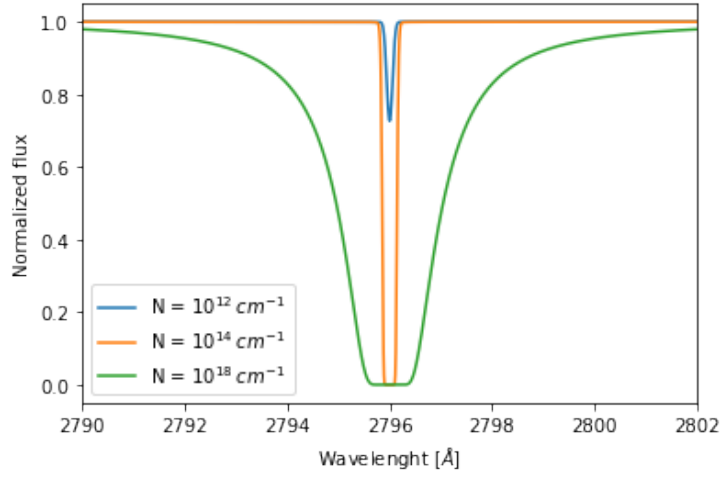
$$\tau(\lambda|N, b, z_{abs}) = N\sigma_0 f_{osc} \Phi(\lambda|b, z_{abs}) \quad (2.7)$$

where N is the column density, σ_0 is the cross section ($\sigma_0 = \frac{\sqrt{\pi}e^2}{m_e c^2}$, where e is the charge of the electron, m_e is the mass of the electron and c is the speed of light), f_{osc} is the oscillator strength (Morton, 2003), Φ is the Voigt function, λ is the wavelength, b is the Doppler parameter and z_{abs} is the Doppler redshift. This work focuses only on the Mg II 2796 transition, thus some of the parameters needed to produce the absorption are constant for all the individual clouds. Table 2.4 shows a list of all the atomic constants and their value.

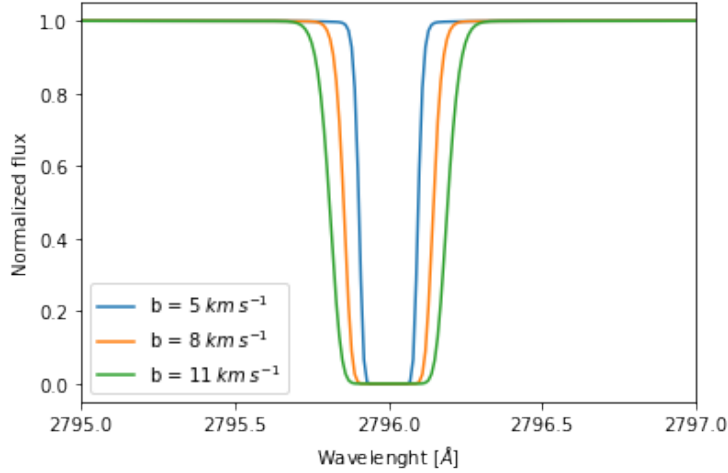
Table 2.4: Atomic constants

	Value	Units
e	1.602×10^{-19}	C
m_e	9.109×10^{-31}	kg
c	299792.45	km s ⁻¹
λ_0 (Mg II 2796)	2796.352	Å
f_{osc} (Mg II 2796)	0.6155	dimensionless
$\Gamma \times 10^8$ (Mg II 2796)	2.612	sec ⁻¹

With Equation 2.7 I can calculate the flux produced by each cloud crossed by the sightline in a range of λ , this range is chosen so that all the absorption is sampled. In Figure 2.7 there is an example of different Mg II 2796 absorbing profile. The aim of this figure is to show the effect of the parameters N and b in the final absorption. Figure 2.7 (a) shows three different profiles, all with $b = 7$ km s⁻¹, centered at 2796Å and three different N . Figure 2.7 (b) shows three different profiles with $N = 10^{14}$ cm⁻¹ and three different b .



(a) Effect of having three different column densities N . All absorption lines have a Doppler parameter of $b = 8 \text{ km s}^{-1}$ and are centered at 2796 \AA .



(b) Effect of having three different Doppler parameters b . All absorption lines have a column density of $N = 10^{14} \text{ cm}^{-1}$ and are centered at 2796 \AA .

Figure 2.7: Plots of different Mg II 2796 absorption profiles calculated with Equation 2.7. Panel (a) shows the effect of changing the column density N and panel (b) shows the effect of changing the Doppler parameter b for optically thick systems.

Table 2.5 shows a summary of all parameters of the model that were explained in last Section 2.1, plus the free parameters to produce the absorption profile. It is important to mention that z_{abs} is calculated from the velocity of the sightline (Equation 2.5) this is why it does not appear in this table.

One sightline can cross several clouds when intersecting the CGM (see Subsection 2.1.2); consequently, this produces several different Voigt profiles. Thus a final spectra will be composed of several individual absorption lines. Figure 2.8 shows an example of a 3-components (3 clouds) spectrum. The individual cloud parameters are indicated in the legend. In the figure the flux is shown as a function of velocity, where $v = 0$ is arbitrarily centred on the second component.

Table 2.5: Summary of all parameters necessary to produce a Mg II absorption profile given this model.

	Parameter
Geometrical parameters	Inclination (i)
	Height (h)
	Impact parameter (d)
	Azimuthal angle (α)
Mg II clouds	Grid size (c_{size})
	Virial radius (R_{vir})
	Filling factor ($f_v(R/R_{vir})$)
Cloud kinematics	Maximum velocity (v_{max})
	Velocity scale height (h_v)
Absorbing parameters	Column density (N)
	Doppler parameter (b)

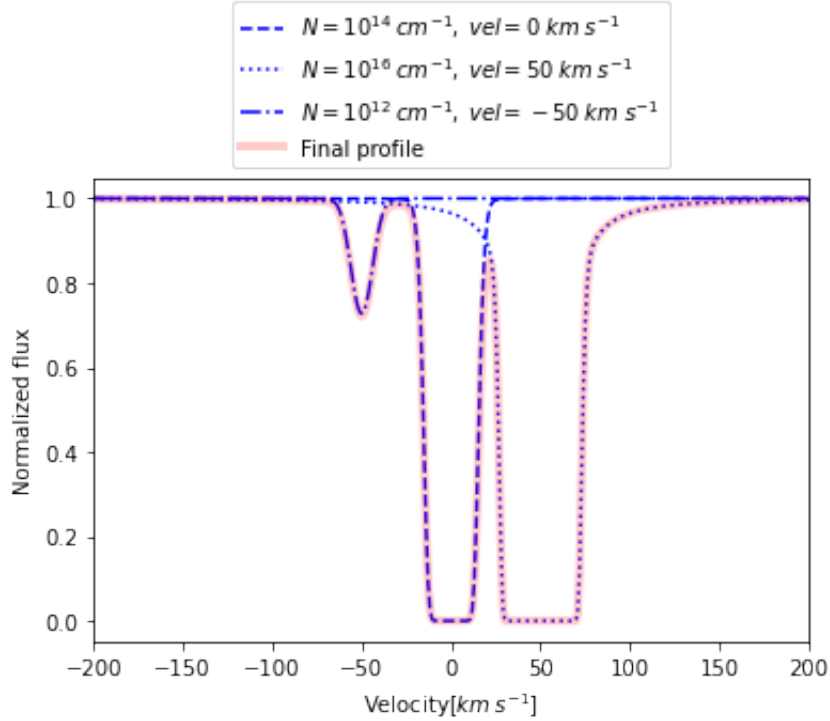


Figure 2.8: Plot to show the effect of adding the absorption lines of different clouds to produce the final absorption profile. In blue are three individual profiles, all with $b = 8 \text{ km s}^{-1}$, different N and different velocities. In red is the total absorbing profile, resulting from adding the three individual profiles.

Finally, to compare synthetic and observed spectra one needs to take into account the instrument profile of the observations. The so called "Line Spread Function" (LSF) is a model of the spectrograph's response to a monochromatic light source. It is defined as the (observed) spectrum of an infinitely thin spectral line ("real" spectrum). The width $\Delta\lambda$ of the LSF is called the spectral resolution, and it can be expressed in wavelength or velocity units. The dimensionless quantity $\lambda/\Delta\lambda$

is called the "resolving power" (R). Thus, the net effect of a spectrograph is to broaden all spectral features due to the finite resolution of the instrument. This is usually modeled as a convolution between the real signal and the LSF. A commonly used LSF is a Gaussian with kernel = $\Delta\lambda$.

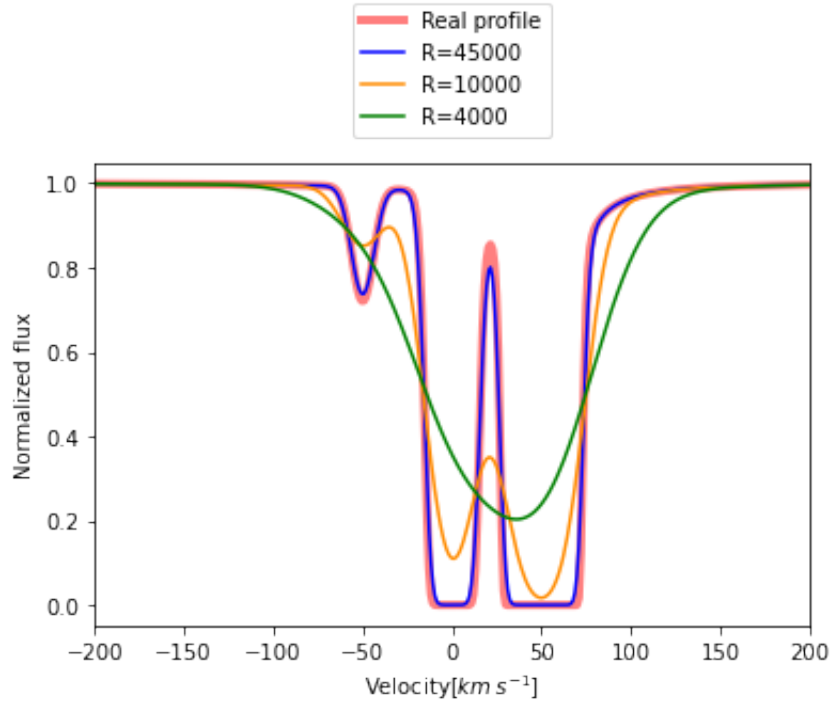


Figure 2.9: Plot to show the effect of different spectral resolutions. Red is the real spectrum, blue shows how the spectrum would look like with a spectrograph with resolution $R = 45000$, orange shows how the spectra would look like with $R=10000$ and green line shows how the spectrum would look like with a spectrograph with resolution $R = 4000$.

Figure 2.9 shows the effect of different spectral resolutions. The red line is the real spectrum (as Figure 2.9). The blue line is the observed spectrum at high resolution ($R = 45\,000$), orange is with medium resolution ($R = 10\,000$) and the green line is the low resolution version ($R = 4\,000$). It can be seen that for high resolution the individual absorption profiles can still be resolved. However, in the low resolution spectrograph, the absorption system looks like a single very broad absorbing profile, therefore unresolved. During this thesis work, every spectra I produce with the model has $R = 45\,000$, which is the spectral resolution of the UVES instrument.

So far I have explained the process of creating the individual spectrum of a QSO sightline crossing the CGM and defined all the parameters of the model. In the next section I explain the procedure of assigning a value to those parameters in order to create realistic synthetic catalogs to be compared to the data.

2.3. Synthetic catalogs of spectra

To fit model parameters, MAGiiCAT is used, hereafter "the Mg II Absorber-Galaxy Catalog". This is one of the most complete, but heterogeneous, Mg II catalogs available in the literature, including systems associated to a wide range of galaxy types and properties. Such diversity can be used to fit more realistic models.

MAGiiCAT is a compilation of 182 isolated (meaning no other spectroscopically identified galaxy within a projected distance of 100kpc and a sightline velocity separation of 500 km s⁻¹) galaxies with associated Mg II absorption lines in the spectra of background quasars. The Absorber-galaxy pairs in MAGiiCAT are a compilation of Mg II absorbers found in other works (see Table 1 in Nielsen et al. (2013)), thus there is a mix of different spectral resolutions. All these galaxies have a spectroscopic redshift between $0.07 \leq z \leq 1.1$ and a quasar at an impact parameter of $d \leq 200kpc$. (For more properties of the MAGiiCATs sample see Table 1.1)

The model has 4 free parameters ($f_v(R/R_{vir})$, c_{size} , h and h_v) (Table 2.5) and galaxy parameters that are chosen according to MAGiiCATs ranges. To fit the free parameters one needs to produce synthetic catalogs that are comparable to the real catalogs. When running the model several times to produce many individual spectra (details in Section 2.2) it is necessary to assign a value to each of the parameters in the model (See Table 2.5 for a summary of the parameters) each time the model runs, to simulate the different kinds of galaxies that exist in the universe. Some parameters are observable, meaning that it is possible to obtain them from real observations. For those kinds of data I assume a distribution from the literature. The other parameters are specific to this model, thus it is not possible to observe them, and I consider them as free parameters of the model. In Table 2.6 there is a list with all the observable parameters and their assumed distribution and all the free parameters.

Based on the model description in Section 2.1, there are four groups of parameters and in this section I summarize previous observations that guide the choice of parameters distributions when creating my synthetic catalogues. This section is divided as: galaxy geometrical parameters (Subsection 2.3.1), Mg II clouds (Subsection 2.3.2), velocity parameters (Subsection 2.3.3) and absorption parameters (Subsection 2.3.4).

2.3.1. Geometrical parameters

This group of parameters are the ones that define first, the position of the QSO sightline with respect to the galaxy and secondly, the geometry of the disk model. The inclination determines how the sightline intercepts the disk, the impact parameter and azimuthal angle determine the position of the QSO sightline with respect to the disk and the height of the disk is a parameter particular to the chosen geometry.

For the inclination (i) of the disk, which is considered here to be the same as the host galaxy, there is no information available, because there is no HST data for the sample in MAGiiCAT. To circumvent this problem, one can assume that the distribution of inclinations in MAGiiCAT is random. Random disks in the sky are distributed according to $\sin(i)$ and the distribution of $\sin(i)$, f_y , with $y = \sin(i)$ is (Ho & Turner, 2011; Lopez & Jenkins, 2012):

$$f_y(y)dy = \frac{y}{\sqrt{1-y^2}} \quad (2.8)$$

To create the synthetic catalogs I use f_y to generate random values in each iteration of the model, to sample from a realistic distribution of random inclinations. This distribution implies that disks with higher inclinations, i.e., close to edge-on, are more probable than face-on disks. Figure 2.10 shows a plot of the used distribution for $\sin(i)$.

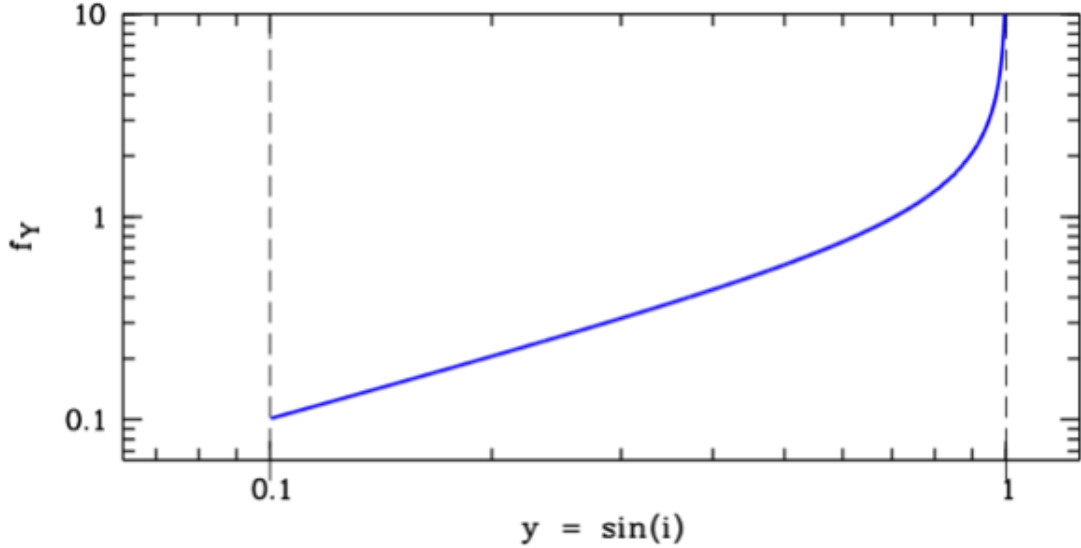


Figure 2.10: Distribution of $\sin(i)$ used to create the synthetic catalogs. f_y is the distribution of $\sin(i)$ assuming a random distribution for inclinations of disk objects. Figure from Lopez & Jenkins (2012).

To create impact parameters, I randomly draw them from the distribution provided by MAGiCAT (See Figure 2.11). This ensures that the synthetic data has the same distribution as the observations, which is a better approach than simply assuming a uniform distribution in the sky. Figure 2.11 shows an histogram with all the impact parameters in MAGiCAT. Data for the azimuthal angles of their sources was not published, however in MAGiCAT V they state that there is no bias in the azimuthal angle distribution of their sample, so it seems safe to assume a uniform distribution for α .

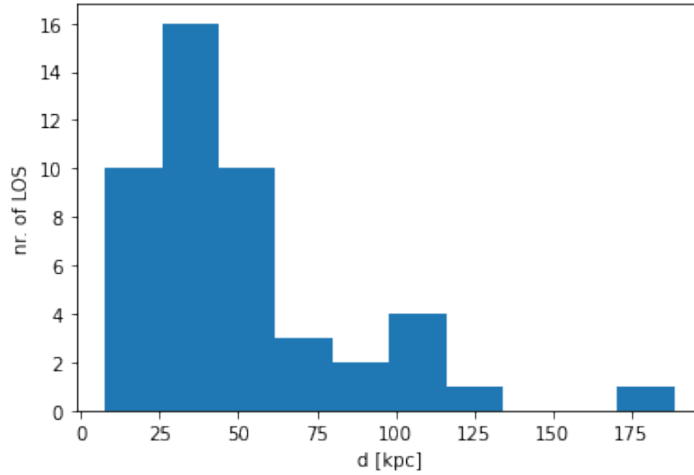


Figure 2.11: Distribution of impact parameters present in the MAGiiCAT sample.

2.3.2. Mg II clouds

The filling factor defines how the individual Mg II clouds are distributed in the gaseous disk. The grid size, c_{size} , determines how many clouds are placed in the model CGM. A large c_{size} means fewer clouds and a very fine c_{size} means more clouds. On the other hand, the virial radius is a characteristic of the host galaxy, which is an input for the filling factor. The filling factor is a very important parameter as it defines the density of clouds within the CGM. In this section I explain in detail how to assign values to these parameters when creating the synthetic catalogs.

The grid size is specific to this model so c_{size} is left as a free parameter when creating the synthetic catalogs. The filling factor is a function, that for this work depends on R and R_{vir} (See Subsection 2.1.2). This function could have any shape in principle. There is very little literature on this function (Stern et al., 2016); thus, here it is considered a free parameter. In Section 4.1 there is an analysis on the shape of $f_v(R/R_{vir})$.

To estimate a distribution of virial radii I assume a spherical collapse model, which assumes that structures in the Universe form from the collapse of high density dark matter regions. They form the so called virialized structures or dark matter halos. The dark matter halo radii follows (Mo et al., 1998):

$$R_{vir} = 0.1v_{max}H(z)^{-1}, \quad (2.9)$$

where $H(z)$ is the Hubble parameter at redshift z and is calculated using a Λ CDM cosmology with the following cosmological parameters: $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$. Here $z = 0.66$ is the median redshift in MAGiiCAT. The values of v_{max} vary from galaxy to galaxy; in the next section there is an explanation on how to obtain this value.

2.3.3. Velocity parameters

Two parameters characterize the kinematics of gaseous disks: the maximum velocity of rotation of the disk and the velocity scale height (See Equation 2.4).

v_{max} is not the same for all galaxies. To make sure the model is sampling a realistic distribution of galaxy masses, I draw v_{max} randomly from the distribution in Zwaan et al. (2010). These authors measure the distribution of rotational velocities and fit it with a Schechter function (Figure 2.12). Given this distribution each sample in the synthetic catalog gets assigned a random v_{max} . With this value I also calculate R_{vir} using Equation 2.9.

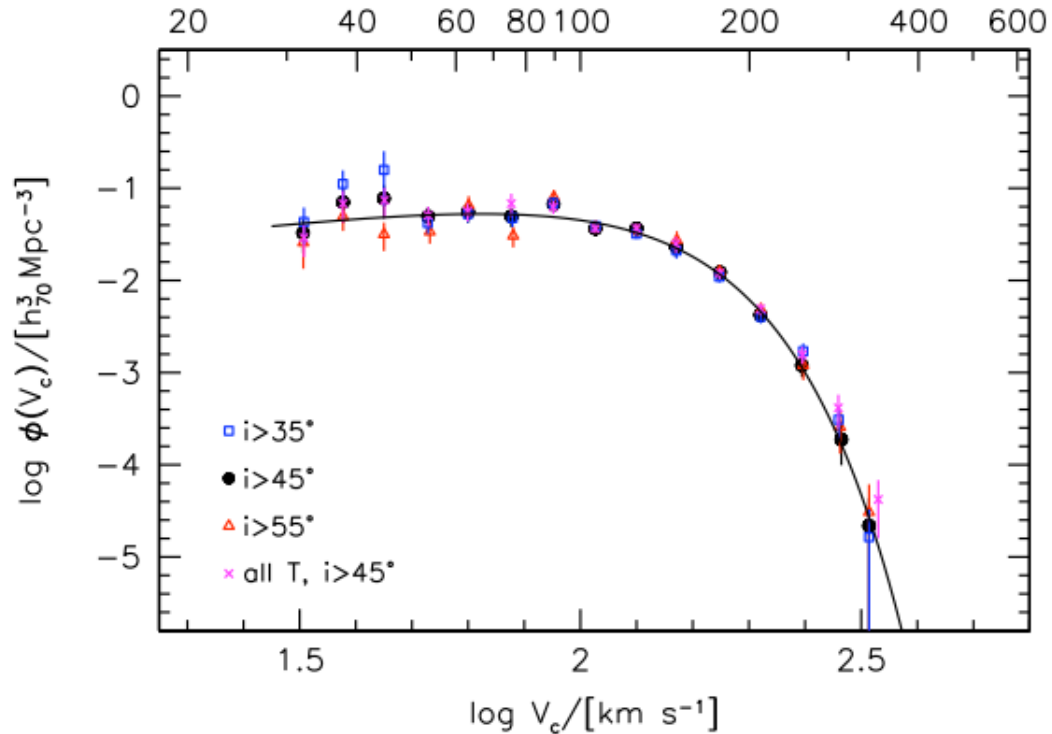


Figure 2.12: Schechter function that characterises the distribution of maximum velocities of rotation for stellar disk galaxies in the low redshift Universe for different inclinations. Figure from Zwaan et al. (2010).

Finally, the velocity scale height, h_v , which relates the maximum velocity to the height of the disk, is left as a free parameter.

2.3.4. Absorption line parameters

Not all Mg II clouds have the same column density and Doppler parameter. These parameters are drawn randomly from the observed distributions in Churchill et al. (2020). Such distributions result from fitting Mg II Voigt profiles in 249 high resolution quasar spectra (from Keck/HIRES and VLT/UVES) having a total of 2989 different Mg II components. Figure 2.13 (a) shows a plot with the N distribution obtained in Churchill et al. (2020). Figure 2.13 (b) shows a plot with the b distribution obtained in Churchill et al. (2020).

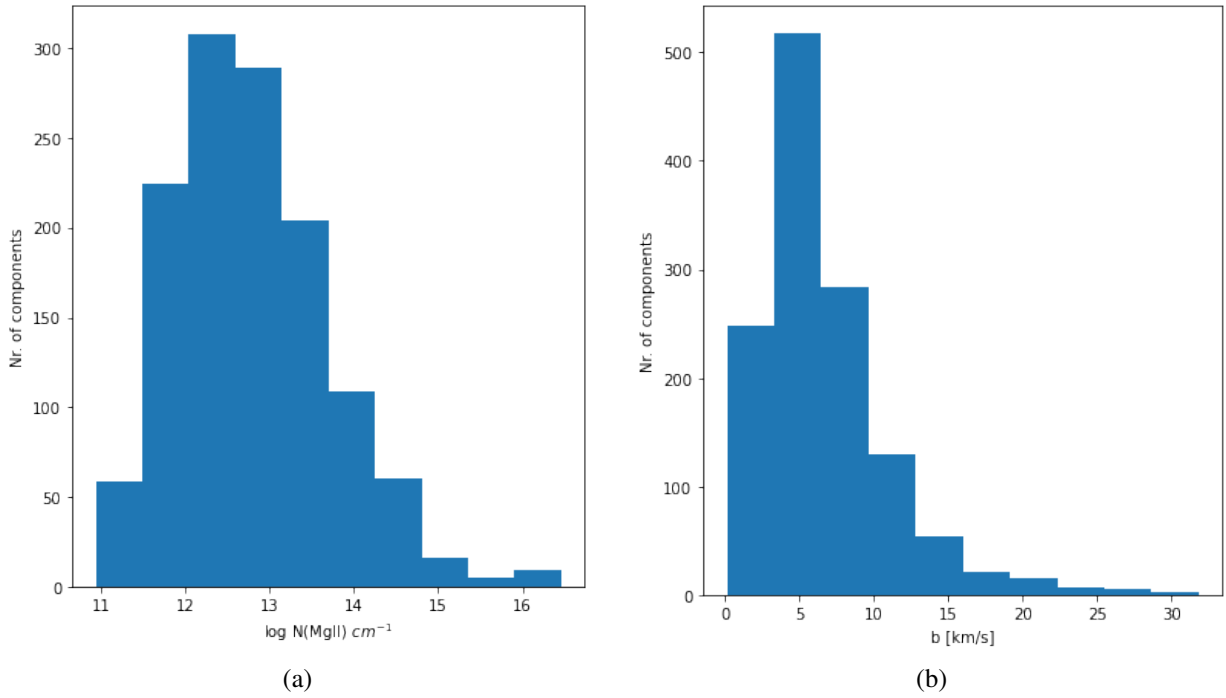


Figure 2.13: Data from Churchill et al. (2020) Panel (a): Mg II column density distribution from Voigt profile fitting. (b) Mg II Doppler parameter distribution from Voigt profile fitting.

It is important to mention that the distributions in Figure 2.13 result from every single Mg II absorbing profile found in high resolution QSO spectra. I cannot confirm that all those absorption systems came from the CGM of a single galaxy. Also, I cannot discard that there is a dependence of N and/or b with some galactic property. Thus, this model assumes that the N and b distributions are the same regardless of galaxy types and through the entire CGM disk, which is not necessary correct.

Table 2.6 shows a list with all the parameters of the model, in yellow are all free parameters and in blue the ones that are drawn from assumed distributions. In the last column there is the method used to assign those values.

Table 2.6: All model parameters and the method used to assign their values in the individual spectra of the synthetic catalog. In blue are the parameters that take values from a distribution in the literature. In yellow are the free parameters of the model.

	Parameter	Synthetic catalog value
Geometrical parameters	Inclination (i)	Distribution of disk objects inclinations (Ho & Turner 2011)
	Height (h)	Free parameter
	Impact parameter (d)	Distribution from MAGiCAT (See Figure 2.10)
	Azimuthal angle (α)	Assume homogeneous distribution
Cloud parameters	Cloud size (c_{size})	Free parameter
	Volumetric filling factor (f_v)	Free parameter
	Virial radius (R_{vir})	Calculated from v_{max} assuming a spherical collapse model (See Section 2.3.2)
Velocity parameters	Maximum velocity (v_{max})	Distribution function of rotational velocities (Zwaan et al. 2009)
	Velocity scale height (h_v)	Free parameter
Absorption parameters	Column density (N)	Distribution from Churchill et al. 2020 (See Figure 2.12)
	Doppler parameter (b)	Distribution from Churchill et al. 2020 (See Figure 2.12)

Chapter 3

Model comparison with QSO statistics

The procedure detailed in Chapter 2 allows me to produce catalogs of synthetic Mg II absorption lines, given rotating disk model for the CGM, with properties similar to MAGiiCAT. In this Chapter I explain how to constrain the model using MAGiiCAT data through two different diagnoses: the equivalent width vs. impact parameter relation (Section 3.1) and the pixel velocity two-point correlation function (Section 3.2). To find the posterior distributions of free parameters I use Markov Chain Monte Carlo (MCMC) (Section 3.3).

3.1. Equivalent width vs. impact parameter

Several studies observe an anti-correlation between W_T and impact parameter d (See Subsection 1.3.3), meaning that sightlines crossing at larger distances from the host galaxy tend to produce weaker Mg II absorption lines. Figure 3.1 shows a sample of several Mg II detections and W_T upper limits as a function of impact parameter. The points show data in MAGiiCAT and the two lines show fits using two different catalogs, black line is for MAGiiCAT and green line is for Chen et al. (2010a).

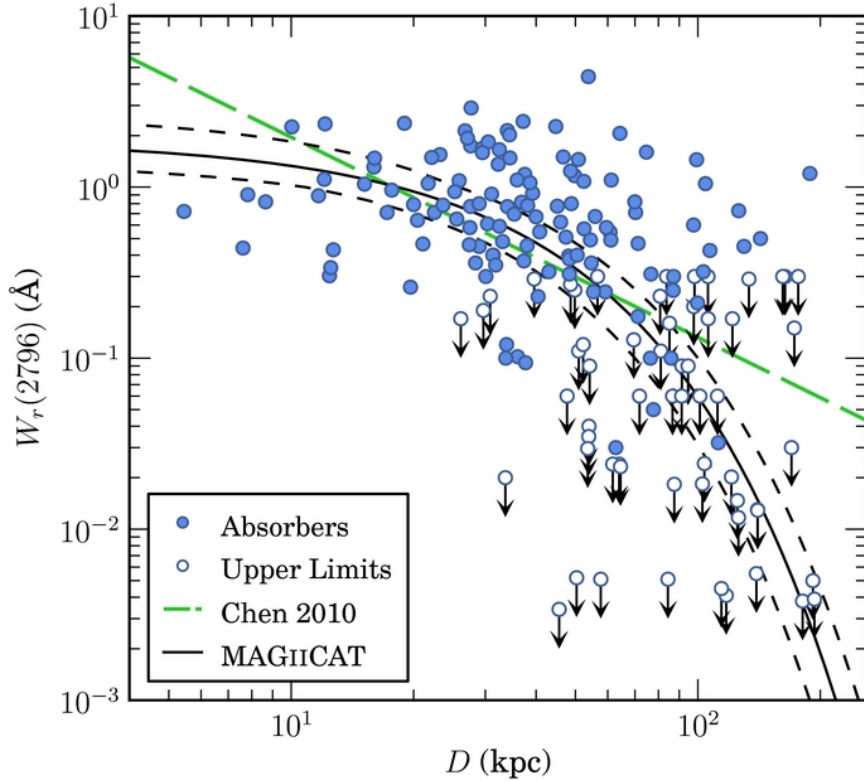


Figure 3.1: $W_r - d$ relation. Blue points show Mg II detections that are included in MAGiiCAT, white points with arrows show upper limits, solid black line shows the best fit for MAGiiCAT data of a fixed form, dashed black line show the 1σ uncertainty of the fit and green line shows the best fit power-law using Chen et al. (2010a) sample. Figure from Nielsen et al. (2013).

To get the W_r value for a synthetic spectrum, first I need to select the range of the spectra where absorption is formally detected, called kinematic subsystems. Since the spectra in my model are noise-free and already normalized, the kinematic subsystem is in the range of λ where the flux is < 1 (the grey regions in Figure 3.2). To calculate the area under the absorption I just sum the value of 1-flux in each spectral pixel and multiply it by the pixel size in \AA , which in this case is selected to match UVES resolution, although it does not matter because W_r is independent of spectral resolution. I do not need to divide these value by $(1+z)$, because the synthetic spectra are already at rest-frame, which means are already independent of redshift.

Figure 3.2 shows ten examples of synthetic spectra. The black lines represent the fluxes and the grey areas show the kinematic subsystems. All of these absorption profiles are generated using the same model parameters, but with different impact parameters (d). Each panel of the plot shows the value of d and W_r . The spectra are generated in the rest-frame, thus they are centered at the rest-frame wavelength of $\text{Mg II}(2796)$. The spectra look realistic (Figure 1.3), except for the fact that they are noise-free.

The total W_r depends on the actual number of clouds and intrinsic kinematics along the sight-lines. Considering the model, the number of clouds can depend on many parameters; the inclination, the height (h), the position of the QSO sightline with respect to the galaxy, the grid size (c_{size}), the virial radius and the filling factor ($f_v(R/R_{vir})$). Among these parameters the ones that

are not taken from a fixed distribution, hence they can vary from model to model, are: h , c_{size} and $f_v(R/R_{vir})$. The kinematics can vary from model to model by changing the velocity scale height (h_v).

In summary all model free parameters affect W_r . Even more, they could affect the model's $W_r - d/R_{vir}$ relation. I use this relationship as a diagnostics for the model, i.e., the model relation is compared with the observational relation.

The observational $W_r - d/R_{vir}$ obtained from MAGiCAT, shows a dependence of W_r with d/R_{vir} , where siglines that are closer to the galaxy tend to show larger W_r . This observed feature is used to constrain the model using MCMC, more on this technique and how to apply it to this specific problem in Section 3.3.

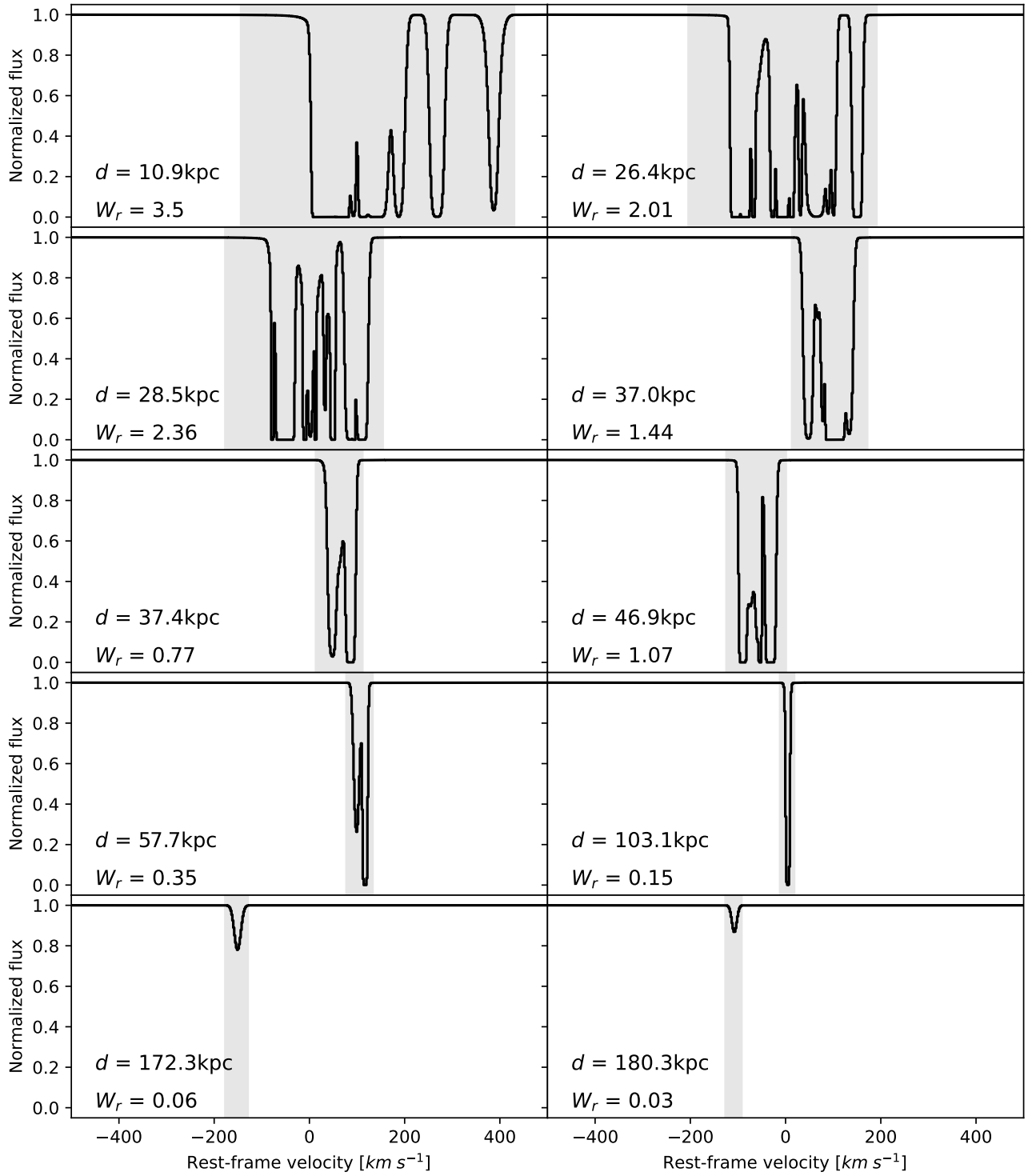


Figure 3.2: Example of absorption profile produced by the model plotted in the rest-frame velocity scale. Each panel shows a different possible synthetic spectra that can come out of the model at a given impact parameter. Solid lines are the profiles and grey areas represent the kinematic subsystems. Each panel shows the rest-frame equivalent width, these values are in \AA , produced by each absorption profile. It can be seen that smaller d tend to produce larger W_r .

3.2. Two-point correlation function

The number of sightlines Nielsen et al. (2015) have for each sub-sample they use to calculate the TPCF is variable. The larger sub-sample consists of 17 spectra (See Table 1.2). To make sure that the models produce enough data to show a trend in each sub-sample I generate 30 sightlines per sub-sample. This is almost twice the largest sample in MAGiCAT, which ensures that the model TPCF shows a realistic trend for each sub-sample.

Once all the sub-groups are created, I can compute the TPCF for each of the sub-samples. To calculate the TPCF for a given sub-sample first I need to obtain the velocities of all pixels where Mg II absorption is formally detected (kinematic subsystems) for each spectra in the sub-sample. For the synthetic spectra, the kinematic sub-systems are defined as the range of wavelengths or velocities where the flux is < 1 (grey areas in Figure 3.2).

With the kinematic sub-systems defined for the synthetic spectra I pull out every velocity pixel in the kinematic sub-system for every spectra in a sub-sample. Then I calculate the absolute difference between every possible pixel pairs to get Δv_{pixel} . I bin up these pixel velocity separations in bins of 10 km s^{-1} (same as in Nielsen et al. (2015)) and normalize each bin by the number of pixels pairs in the sub-sample, for comparison between sub-samples. If the TPCF shows a larger probability towards larger Δv_{pixel} for a given sub-sample of spectral lines, it means that such sub-sample has a larger velocity dispersion.

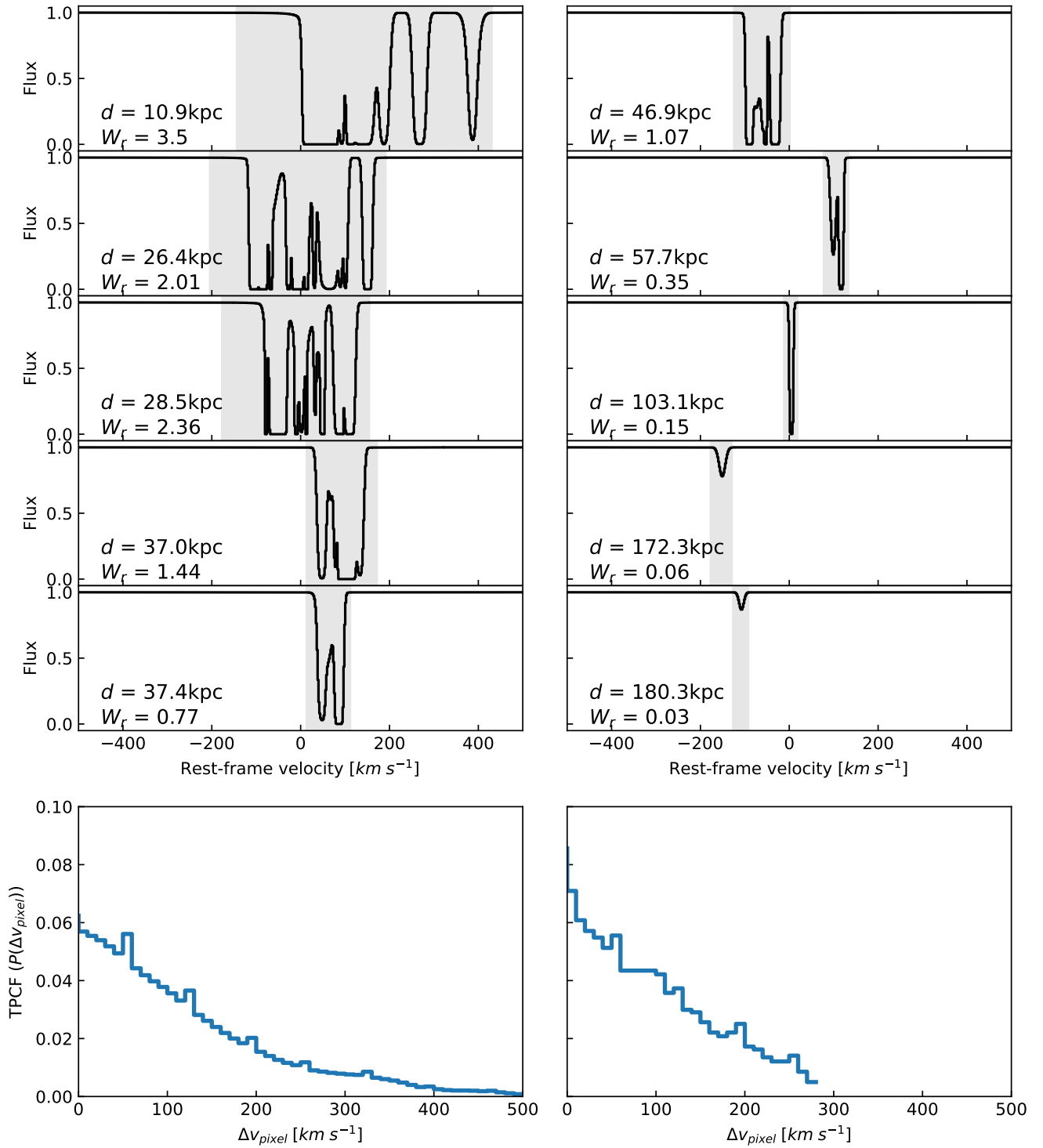


Figure 3.3: Example of TPCFs for two different samples of spectra. Each sample has five different spectra. The top five panels on the left show the three different spectra in the first sample. They are plotted on a rest-frame velocity scale. First five plots on the right show the three spectra for the second sample, plotted on a rest-frame velocity scale. Bottom panel on the left shows the TPCF for the first sample and bottom panel on the right shows TPCF for the second sample. x-axis in TPCF plots show velocity absolute difference, y-axis shows the probability of having two absorption pixels with a given velocity absolute difference.

Figure 3.3 shows an example of two TPCF for two different sub-samples. Using the sample of model spectra showed in Figure 3.2 I separated it in two sub-samples, one with $d < 50kpc$ and another with $d > 50kpc$. Figure 3.3 shows the two sub-samples and the resulting TPCF for each. It can be seen that for $d < 50kpc$ the TPCF shows a wider distribution towards larger Δv_{pixel} values, thus that sample has a larger velocity dispersion.

The same is done for the synthetic spectral sample resulting from the best fit model according to $W_r - d/R_{vir}$ relations for the same sub-samples as in Nielsen et al. (2015) (See Table 1.2). The idea is to see for which sub-samples the best fit model is able to reproduce the observed velocity dispersion. These results are shown in Section 4.4.

3.3. Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) is a very used tool in Astronomy and other fields to do comparisons between models and observations. In this thesis work I use MCMC to find a range of free parameter values of the rotating disk model (the height of the disk (h), the grid size (c_{size}), the filling factor ($f_v(R/R_{vir})$) and the velocity scale height (h_v)), that can reproduce QSO statistics, given MAGiCAT data.

MCMC makes use of the Bayes theorem to find the joint posterior distribution (PDF), which describes the probability of having an event A occur once event B has already occurred:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \quad (3.1)$$

where $P(A|B)$ is the posterior distribution, $P(B|A)$ is the likelihood, $P(A)$ is the prior distribution of A and $P(B)$ is the probability that B has occurred.

If A is the model, and B is the data, the PDF tells us the probability of the model given the measured data. The likelihood (\mathcal{L}) is a way to measure how well does the model reproduce the observations, the prior is the probability of the model in the absence of any new data and $P(B)$ is the probability of the data, which is called the evidence. For model inference, Bayes theorem can be written as:

$$PDF \propto \mathcal{L} \times Prior, \quad (3.2)$$

where the $P(B)$ term is encoded in the proportionality, due to the fact that the PDF equation has to equal 1 (as we are determining a probability).

MCMC uses walkers that move through the parameter space to efficiently map the PDF. At each step it evaluates \mathcal{L} and decides how to keep moving depending on the value, in such a way that the probability to be at a point is proportional to the PDF. After the run it multiplies the resulting probability with the *Prior* to get the final PDF. There are several ways in which MCMC can handle the process of proposing a new position. The most common one is called Metropolis-Hastings algorithm, in which the new position is draw from a Gaussian distribution centered at the old position.

The way in which MCMC works is the following:

1. Select a random initial position in the model parameters space, θ .
2. For i in number of steps, repeat:
 - (a) Draw new candidate, θ^* , using Metropolis-Hasting algorithm
 - (b) Calculate $\mathcal{L}(\theta^*)$
 - (c) Calculate $a = \mathcal{L}(\theta^*)/\mathcal{L}(\theta)$
 - (d) if $a \geq 1$ accept θ^* and set $\theta = \theta^*$
if $0 < a < 1$: accept θ^* and set $\theta = \theta^*$ with probability a
reject θ^* and set $\theta = \theta$ with probability $1 - a$

This process of proposing a new position and moving (or not) towards the new position is repeated several times, allowing the walker to move to the set of parameters that better fit the data. In MCMC one can use several walkers at the same time to sample from the same PDF. In this way one can choose different initial values for θ and assure that a larger the space of parameters is covered. The walkers save in an array all the values of accepted θ_s so when all walkers have taken all the steps it is possible to plot the sample of θ_s that the walkers produce. By combining the results of every walkers it is possible to plot the corner plots, to show the marginalized PDF of every parameter to see the best fit and the uncertainties and the 2D PDFs to look for degeneracy between parameters in the model. In the next sections I discuss how to use MCMC to fit the model presented in this thesis using the $W_r - d/R_{\text{vir}}$ relationship.

3.3.1. The Likelihood function

The first step to run MCMC is to compute the model, which returns several synthetic MgII spectra depending on the free parameters (the procedure to do so is explained in Chapter 2). MAGiiCAT consist of a sample of 182 QSO spectra, thus I produce 200 synthetic spectra in each iteration of the model, to apparently match the observational sample size. In MAGiiCAT they use only absorbing lines with $W_r \geq 0.003\text{\AA}$, so after producing a model sample I need to select absorption lines that produce $W_r \geq 0.003\text{\AA}$ before comparing with MAGiiCAT data.

After creating the model catalog I calculate the models $W_r - d/R_{\text{vir}}$ relationship with the methodology described in Section 1.3. After this, the model is ready to reproduce observables that can be compared to MAGiiCAT observations using MCMC.

I need to define a likelihood function (\mathcal{L}), which gives the probability of the models $W_r - d/R_{\text{vir}}$ relationship given the data $W_r - d/R_{\text{vir}}$ relationship. A common way to compare two distributions is the Kolmogorov-Smirnov (KS) test, which assesses whether two samples come from the same underlying distribution. This test is usually used for 1 dimensional data, however, there is an implementation for 2 dimensional distributions (Fasano & Franceschini, 1987).

The aim of this test is to answer the question of what the probability that two samples come from the same probability distribution is. The null hypothesis of this test is that both samples have identical $W_r - d/R_{\text{vir}}$ relation. This test returns a p -value, small p -values can be interpreted as

the two samples being significantly different, larger p -values means that the two samples are not significantly different (Press, W.H. et al. 2007, Numerical Recipes, section 14.8).

To deal with upper limits in MAGiiCAT sample, I decided to do a bootstrap of the observations when calculating the likelihood. The way I do this is the following: I produce a model, I take the observations and change all the upper limit values randomly with a uniform distribution from 0 to the upper limit value 1000 times, I calculate the p -value from a 2D KS test for all 1000 different observations compared to the model and finally take as likelihood the mean of p -values.

Given that p -values are usually very small, one can use the log of the likelihood function, with no loss of generality, as maximizing the log of a function is the same as maximizing the function. Thus the likelihood function is:

$$\mathcal{L} = \ln(\overline{p\text{-value}}), \quad (3.3)$$

where the line over p -value indicates the mean of 1000 iterations with different values for observational W_r upper limits.

3.3.2. Priors

To properly run MCMC a prior distribution is needed that gives the information we already know about the model parameters. Since I do not have any prior information on the free parameters I use a uniform distribution with some boundaries for all of the four free parameters. The boundaries are selected so that the model makes physical sense and to save computational time.

The filling factor is one of the free parameters of the model and it is a function that depends on R/R_{vir} . In principle $f_v(R/R_{vir})$ could have any possible shape, in Section 4.1 I conclude that an exponential function with characteristic radius r_0 is a good fit. To run MCMC I use r_0 as a free parameter.

Table 3.1 shows the boundaries for the uniform priors for each of the free parameters of the model:

Table 3.1: Boundaries for uniform priors of the free parameters

	Lower boundary	Upper boundary
r_0	$\frac{1}{10}R_{vir}$	$2R_{vir}$
c_{size}	$0.01kpc$	$10kpc$
h	$1kpc$	$70kpc$
h_v	$10^{-3}h$	10^5h

3.3.3. Computational times

One of the biggest difficulties of performing an MCMC with this model is the large computational times needed to compute the model. Each time the model is computed, it produces 200 synthetic sightlines. For this thesis work I run the models in a computer from the Arc-Tomo col-

laboration, located in the department of astronomy in Universidad de Chile. This computer has an Intel Xeon Silver 4210R Processor (2.40GHz, 20 nodes) with a total RAM of 125 GB.

Depending on the model parameters it will take more or less time to produce the model. For example, a model with very fine grid size, c_{size} , and large heights, h , will take longer to compute than a model with big c_{size} and small h . Given the prior ranges (Table 3.1) and the used computer, the model with the smallest c_{size} and largest h takes 10 minutes and 32.52 seconds to run. The model with largest c_{size} and smallest h takes 0.147161 seconds to run.

MCMC requires to compute the model each time the walkers takes a step. Ideally one would like to have many walkers and many steps to make sure that the sample converges to the PDF, but this would require to evaluate the model many times. A good compromise I found between seen results and having doable computational times is 200 walkers with 8000 steps each, which took almost five days to run. The results of the MCMC run are presented in Section 4.2.

Chapter 4

Results

In this section the results of comparing the model (Chapter 2) to the data are presented. The goal is to find a set of free parameters for the model (See Table 2.6) that can reproduce observable features present in the QSO catalogs, from MAGiiCAT. First the W_r vs. d/R_{vir} relation is calculated to find a general shape for $f_v(R/R_{vir})$ that is consistent with the observational data (Section 4.1).

Then I show the results of the MCMC run. To do this I chose the W_r vs. d/R_{vir} relation as a diagnostic. I use MCMC to find the posterior distribution of free parameters (Section 4.2). I use the TPCF as a way to get a better understanding of the best parameters I get from the $W_r - d/R_{vir}$ fitting by plotting the models TPCF and compare them to the observational TPCF (Section 4.3). At the end I provide a summary of all results (Section 4.4).

4.1. Filling factor

In principle, the filling factor, $f_v(R/R_{vir})$, could have any form (see Subsection 2.1.2 for more details on how $f_v(R/R_{vir})$ is defined). To avoid large computational times when fitting the four free parameters of the model in this section I explore four simple functions for $f_v(R/R_{vir})$ to see which general form fits better the data and use that as an input for the fitting.

The methodology used is a simple visual inspection of the model against the data. I produce some synthetic catalogs with different $f_v(R/R_{vir})$ functions, measure the W_r vs. d/R_{vir} relation, with the procedure explained in Section 3.1, and plot them against the W_r vs d/R_{vir} relation of MAGiiCAT. The idea is to see how the shape of $f_v(R/R_{vir})$ affects the model relationship of W_r vs. d/R_{vir} and if I can discard some $f_v(R/R_{vir})$ functions and find one that reproduces the observational relationship. First I tried three very simple functions. An increasing step function (Equation 4.1), a decreasing step function (Equation 4.2) and a constant function (Equation 4.3):

$$\text{Increasing step } f_v(R/R_{vir}) = \begin{cases} 0 & \text{if } R/R_{vir} \leq 0.7 \\ 1 & \text{if } R/R_{vir} > 0.7 \end{cases} \quad (4.1)$$

$$\text{Decreasing step } f_v(R/R_{vir}) = \begin{cases} 1 & \text{if } R/R_{vir} \leq 0.7 \\ 0 & \text{if } R/R_{vir} > 0.7 \end{cases} \quad (4.2)$$

$$\text{Constant } f_v(R/R_{vir}) = 1 \quad (4.3)$$

The *Increasing step* f_v (top panel of Figure 4.1 (a)) has a donut shape (completely empty from $R=0$ to $R=0.7R_{vir}$ and full of MgII clouds for $R > 0.7R_{vir}$), the *Decreasing step* f_v (middle panel of Figure 4.1 (a)) has a disk shape (filled with clouds until $R = 0.7R_{vir}$) and the *Constant* f_v (bottom panel of Figure 4.1 (a)) is a completely solid disk (filled with clouds at every R value).

Three different catalogs of spectra are created, one for each function, each having 200 spectral systems and the same values for height (h), cloud size (c_{size}) and velocity scale height (h_v). The values for these three free parameters were chosen to have reasonable values (according to MCMC results showed in Section 4.2), $h=45$ kpc, $c_{size}=5$ kpc and $h_v = 10^4 h$. Each of the three catalogs have a different $f_v(R/R_{vir})$, described before.

Figure 4.1 shows the W_r vs. d/R_{vir} relation for each of the catalogs with different $f_v(R/R_{vir})$. Panel (a) shows the three different $f_v(R/R_{vir})$ used to produce the catalogs in three different colors. Panel (b) shows at the top an histogram with the distribution of $\log(d/R_{vir})$ for each of the three catalogs, in the same colors as in panel (a) and in grey is the $\log(d/R_{vir})$ histogram for the observations. At the bottom of the histogram there is a plot of W_r vs. d/R_{vir} in logarithmic scale for both axes. The grey dots are W_r detections in MAGiiCAT and the grey arrows are the upper limits. The three different colored areas correspond to the maximum and minimum values of W_r in 9 different bins of d/R_{vir} for each of the synthetic catalogs. The plotted squares, diamonds and triangles correspond to the mean values of W_r in the same 9 bins. Each color in the W_r plot correspond to the model with the same color in panel (a) of Figure 4.1.

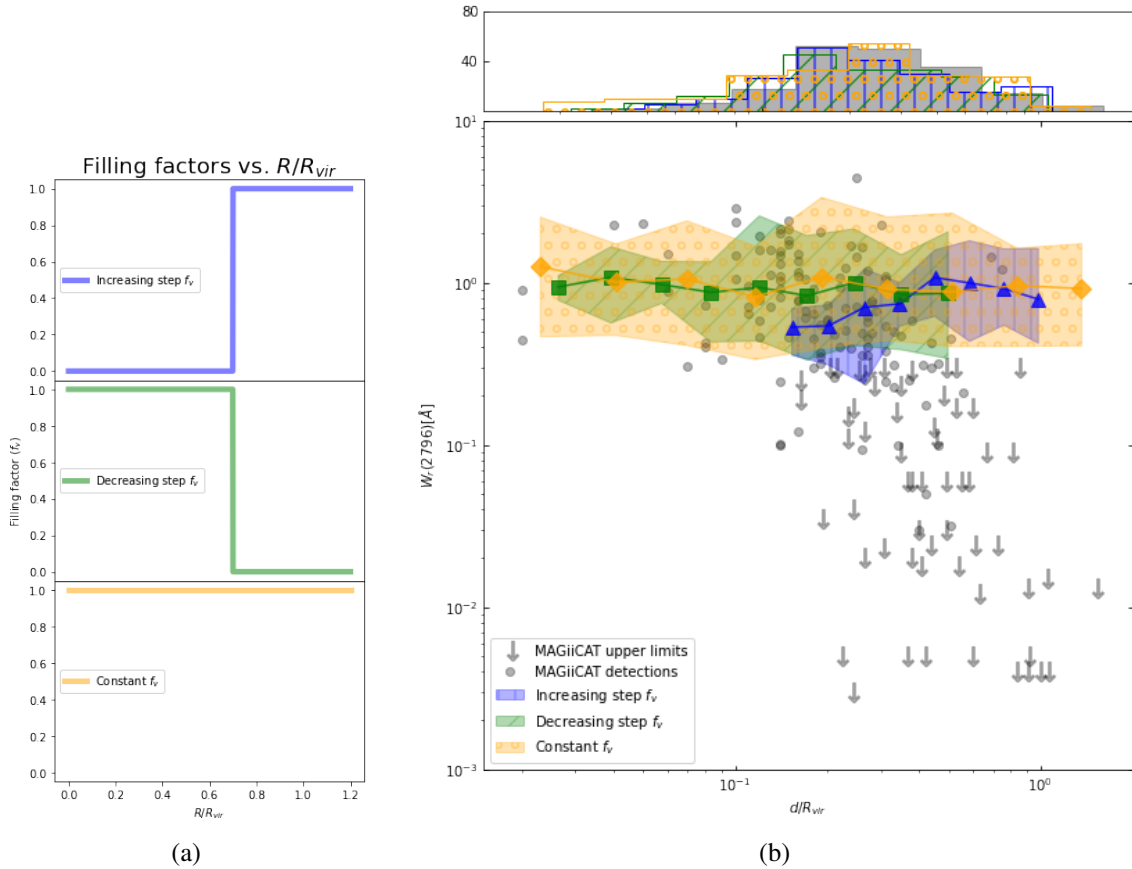


Figure 4.1: $W_r - d/R_{vir}$ relation for three different $f_v(R/R_{vir})$. (a): three different $f_v(R/R_{vir})$, in blue is the *increasing step* f_v , in green the *decreasing step* f_v and in orange the *constant step* f_v . (b): $W_r - d/R_{vir}$ relationship for each $f_v(R/R_{vir})$. Colored areas show the maximum and minimum of W_r in 9 bins of d/R_{vir} , for each $f_v(R/R_{vir})$, with the respective colors. Grey points and arrows show detections and upper limits in MAGiCAT respectively. I performed a KS test to compare the models relationship to the observations, the p-values where: $\approx 10^{-20}$ (*increasing step* f_v), $\approx 10^{-8}$ (*decreasing step* f_v) and $\approx 10^{-14}$ (*constant step* f_v). Top of panel (b): histograms showing the distributions of d/R_{vir} for each sample, in the same bins as the W_r areas and in the same colors as panel (a), and in grey, the d/R_{vir} distribution for the MAGiCAT data.

From Figure 4.1 we can see how $f_v(R/R_{vir})$ affects the synthetic $W_r - d/R_{vir}$ relation. The increasing step $f_v(R/R_{vir})$ or "donut shape" CGM does not produce any absorption at small d/R_{vir} , as seen in the blue area in panel (b) of Figure 4.1, even though there are sightlines at these distances (as shown by the blue histogram at the top). This occurs because by definition this $f_v(R/R_{vir})$ models a disk with no Mg II very close to the galaxy, so it make sense that sightlines crossing near the galaxy do not produce any absorption lines. We observe absorption lines for sightlines very near to the galaxy, therefore, a "donut shape" does not seem to be a good guess for the $f_v(R/R_{vir})$ function, which is also shown by the low p-value.

The other two models produce a better representation of the $W_r - d/R_{vir}$ relation. Both are similar in that they contain clouds in the middle of the disk and show scatter. However, the decreasing step f_v shows a larger scatter towards larger d/R_{vir} than the constant function, and a slight decrease

of the values, which seems to be more consistent with observations, as shown by the larger p-value. This trend stops at some $d < R_{vir}$, after that no absorption is produced. The constant function shows absorption up to much larger d/R_{vir} values, however the values of W_r do not seem to show a decrease as the observations do, which is why it has a lower p-value.

From Figure 4.1, one sees that if $f_v(R/R_{vir})$ has an increasing, decreasing or constant shape, that translates to the $W_r(d)$ relation in the same way. Also, if $f_v(R/R_{vir})$ falls down abruptly (to 0), at some point this is reflected in the W_r values as zero-absorption for some d/R_{vir} values. Since the observations show a decreasing trend of W_r with d/R_{vir} I propose a $f_v(R/R_{vir})$ that has a decreasing form. Also, observations show no gap in the $W_r - d/R_{vir}$ relation, at least until $d \simeq 1.2R_{vir}$, so I propose a smooth function, with no abrupt steps. In conclusion, a good candidate for $f_v(R/R_{vir})$ can be an exponential:

$$f_v(R/R_{vir}) = A e^{-\frac{1}{r_0} \frac{R}{R_{vir}}}, \quad (4.4)$$

where A is a normalization constant and r_0 is the characteristic radius, which is the R/R_{vir} value at which the exponent of the exponential is -1. I assume that $f_v(R/R_{vir})$ is 1 at $R=0$. The first result of the present thesis work is that this f_v can in fact reproduce the observational $W_r - d/R_{vir}$ relationship.

All other free parameters of the model are the same as the ones used to produce data in Figure 4.1, $h = 45$ kpc, $c_{size} = 5$ kpc and $h_v = 10^4 h$ (reasonable values according to MCMC results shown in Section 4.2). I try $r_0 = 2R_{vir}, \frac{1}{3}R_{vir}, \frac{1}{10}R_{vir}$. These values define how steep the exponential function is, Figure 4.2 (a) shows a plot of these three functions.

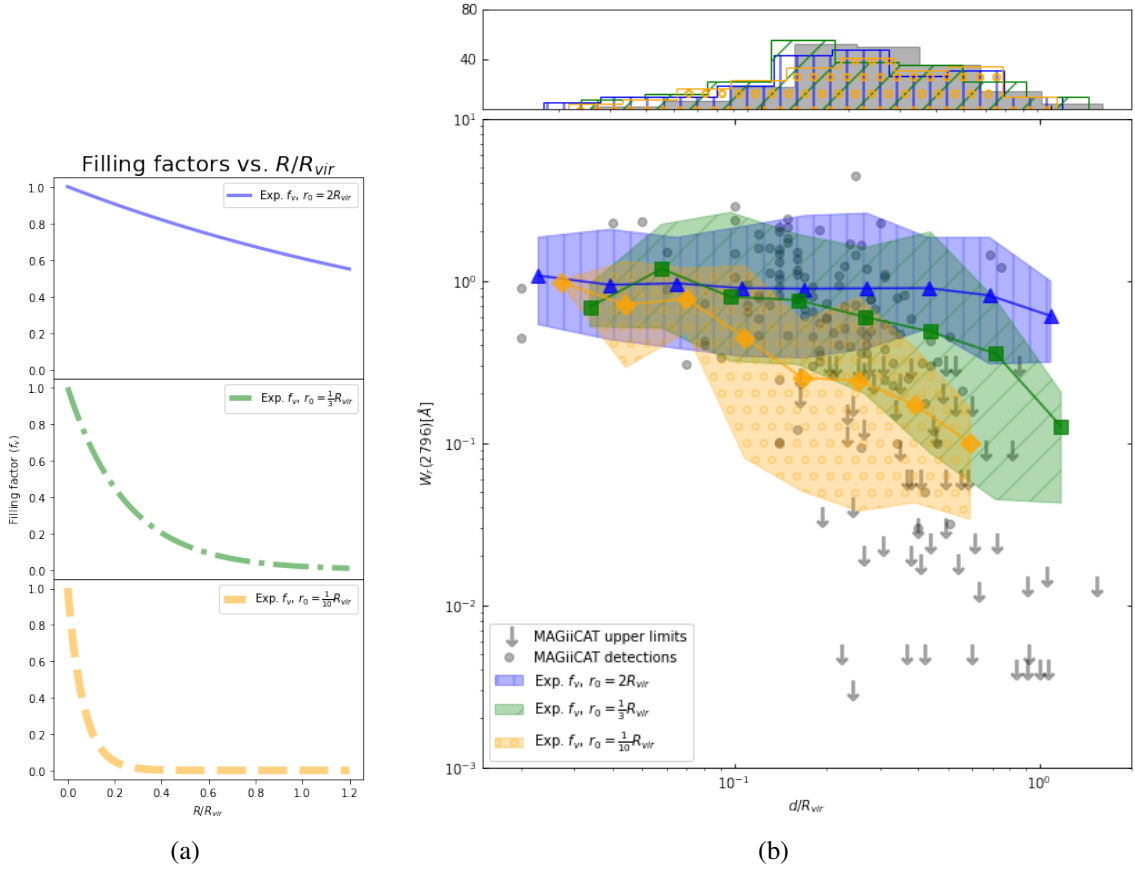


Figure 4.2: $W_r - d/R_{vir}$ relationship for three different exponential $f_v(R/R_{vir})$, with different r_0 , panel (a). (b): resulting $W_r - d/R_{vir}$ relations with each $f_v(R/R_{vir})$. Color areas show the maximum and minimum of W_r in 9 bins of d/R_{vir} , for each $f_v(R/R_{vir})$, with the respective colors. Grey points and arrows show detections and upper limits in MAGiCAT respectively. I performed a KS test to compare the models distribution to the observations, the p-values where: $\approx 10^{-14}$ ($r_0 = 2R_{vir}$), $\approx 10^{-3}$ ($r_0 = \frac{1}{3}R_{vir}$) and $\approx 10^{-16}$ ($\frac{1}{10}R_{vir}$). Top of panel (b): histograms showing the distributions of d/R_{vir} for each sample, in the same bins as the W_r areas and in the same colors as panel (a), and in black, the d/R_{vir} distribution for the MAGiCAT data.

Figure 4.2 shows the W_r vs. d/R_{vir} relation for three different exponential models for $f_v(R/R_{vir})$. Panel (a) shows $f_v(R/R_{vir})$, the blue one has a $r_0 = 2R_{vir}$, the green plot shows an exponential with $r_0 = \frac{1}{3}R_{vir}$ and the orange one has a $r_0 = \frac{1}{10}R_{vir}$. Panel (b) of Figure 4.2 shows the W_r vs. d/R_{vir} for the three different $f_v(R/R_{vir})$. The areas represent the minimum and maximum values of W_r in ten different bins of d/R_{vir} , the triangles, squares and diamonds represent the mean values in the same bins.

Figure 4.2 shows that the steeper $f_v(R/R_{vir})$ is, the steeper the W_r vs. d/R_{vir} relationship is. The sample plotted in green has the largest p-values, indicating that it is the one more similar to the observations. From Figure 4.2 I conclude that an exponential $f_v(R/R_{vir})$ (Equation 4.4) could reproduce the observational relationship better than a *step* f_v or a *constant* f_v , as seen from larger p-values. With the correct value of r_0 and the correct combination of free parameters I could find a model that reproduces the observations. In the next sections I explore how the model is affected by

all the free parameters assuming an exponential $f_v(R/R_{vir})$.

4.2. MCMC results

Using MCMC I got the posterior distribution of all four free parameters of the model, for better visualisation of the results I define new variables as input for the model: $\frac{1}{r_0}R_{vir}$, the grid size (c_{size}) in kpc, the height of the disk (h) in kpc and $\log(h_v h)$.

For this model I use MCMC and compare the model and observational $W_r - d/R_{vir}$. The likelihood function (estimator) is a KS test between the model and the observations (See Subsection 3.3.1 for more details on how this is done). The model $W_r - d/R_{vir}$ relations are computed from catalogs of 200 synthetic spectra and the observational relation comes from 182 Mg II systems with $W_r > 0.003\text{\AA}$ presented in MAGiCAT. By comparing these two relationships the marginalized PDF for each free parameter is obtained and also, the 2 dimensional PDF for each pair of parameters is computed. The marginalized PDF gives the best fit parameters and their uncertainty. The PDFs also provide an idea of possible degeneracy between parameters.

Figure 4.3 shows MCMC corner plots for 200 walkers and 8000 steps. Diagonal panels show the marginalized PDF for each of the free parameters, the middle vertical line in each PDF plot shows the median value of the distribution and the two lines by the sides indicate the 1σ (68%) intervals of the distributions. At the top of each marginalized PDF is the mean value for each parameter with the 1σ intervals. The off-diagonal plots show the 2D PDF for each pair of parameters. Very light grey points represent positions in the 2D parameter space where the walkers stepped into at some point of the run. The light grey, dark grey and black areas represent where the 0.84, 0.5 and 0.16 percent of the sample is, respectively.

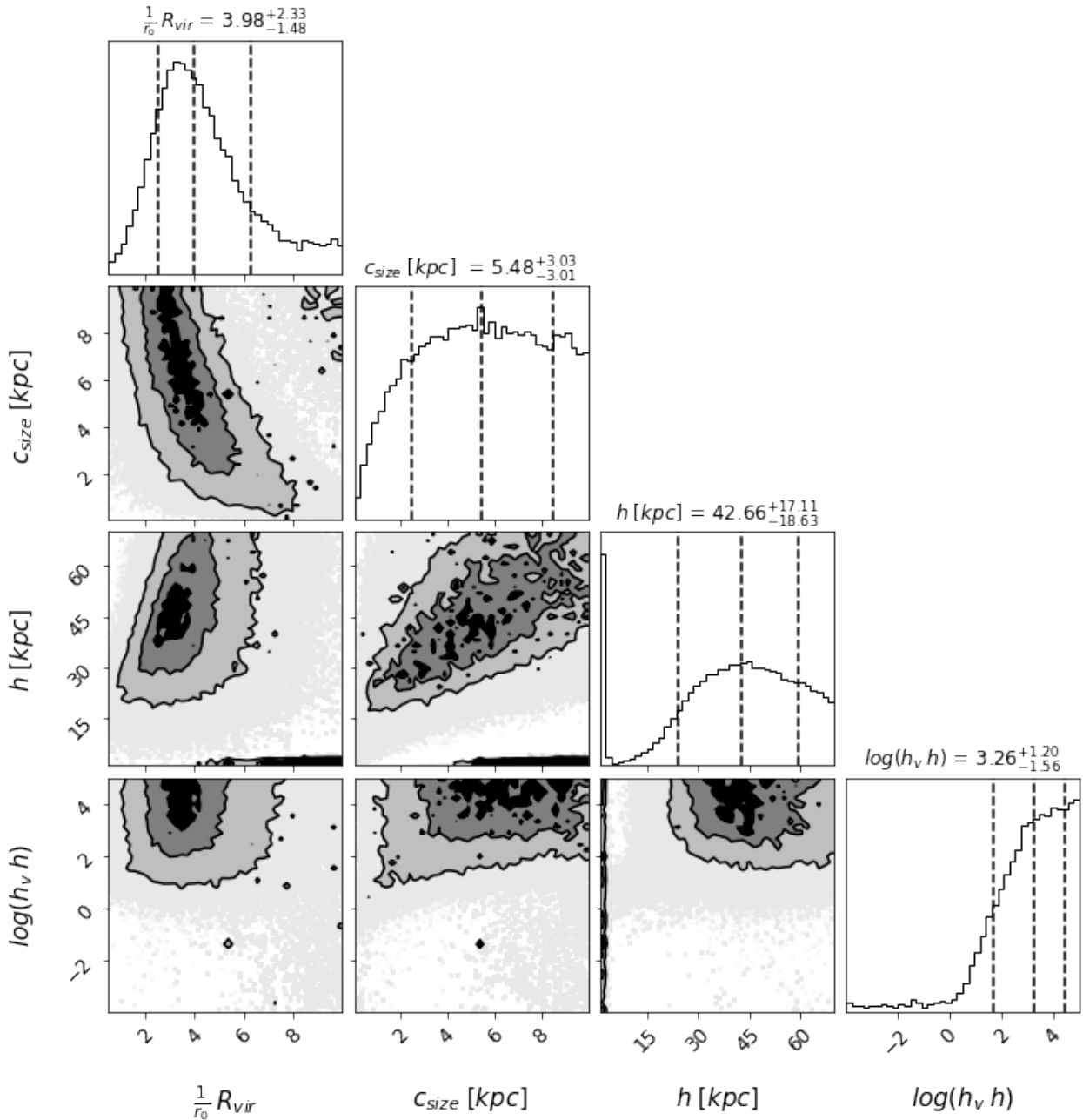


Figure 4.3: Corner plots showing the PDFs obtained with MCMC. Diagonal plots show the marginalized PDF for each free parameter, the median value and the 1σ intervals are showed by dashed lines. At the top of each PDF are the median values and 1σ values. Off-diagonal plots show the 2 dimensional PDFs for each pair of free parameters. Light grey points show where the MCMC walkers accepted values, light grey area, dark grey area and black area show where the 0.84, 0.5 and 0.16 of the sample lay.

The corner plots show some degeneracy between $\frac{1}{r_0} R_{vir}$ and c_{size} ; and between c_{size} and h (the samples lay over large areas). Well behaved fits would show contours that are concentric circles, with low deviations. The $\frac{1}{r_0} R_{vir}$ vs. c_{size} PDF has an elongated form, suggesting correlation

between these two parameters. The same can be seen in c_{size} vs. h , where the data allows models with low c_{size} and low h or models with large c_{size} and big h .

This trend could be due to the data not being able to fully constrain the model or due to poor performance of MCMC. One way to analyze how well MCMC performed is to look at the acceptance ratio. This value is the number of times the algorithm accepted a proposed new position divided by the total number of proposals. If MCMC is accepting almost all proposals, then the steps taken are too small and the algorithm is not exploring the entire parameter space. On the contrary if it is accepting almost no new proposal, the walkers are not converging to the best parameter space. A good number for the acceptance ratio is between 0.215 and 0.5 (Roberts & Rosenthal, 2001).

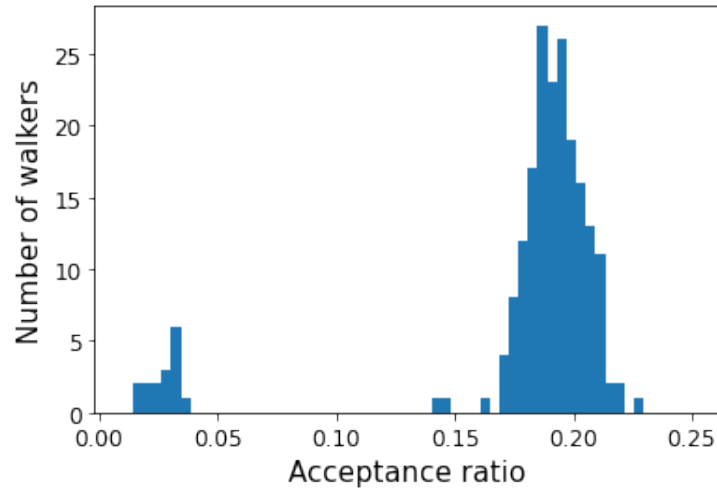


Figure 4.4: Histogram of acceptance ratios of individual walkers. The x-axis shows the acceptance ratio value and y-axis shows the number of walkers.

Figure 4.4 shows an histogram with the distribution of acceptance ratios in the present MCMC run. Most of the walkers have acceptance ratios between 0.13 and 0.23. However there is a smaller group of walkers that have very small acceptance ratios, between 0 and 0.05. One way to figure out why there are two groups of walkers with so different acceptance ratios is by looking at the trace plots, which are plots of the parameter values as a function of steps.

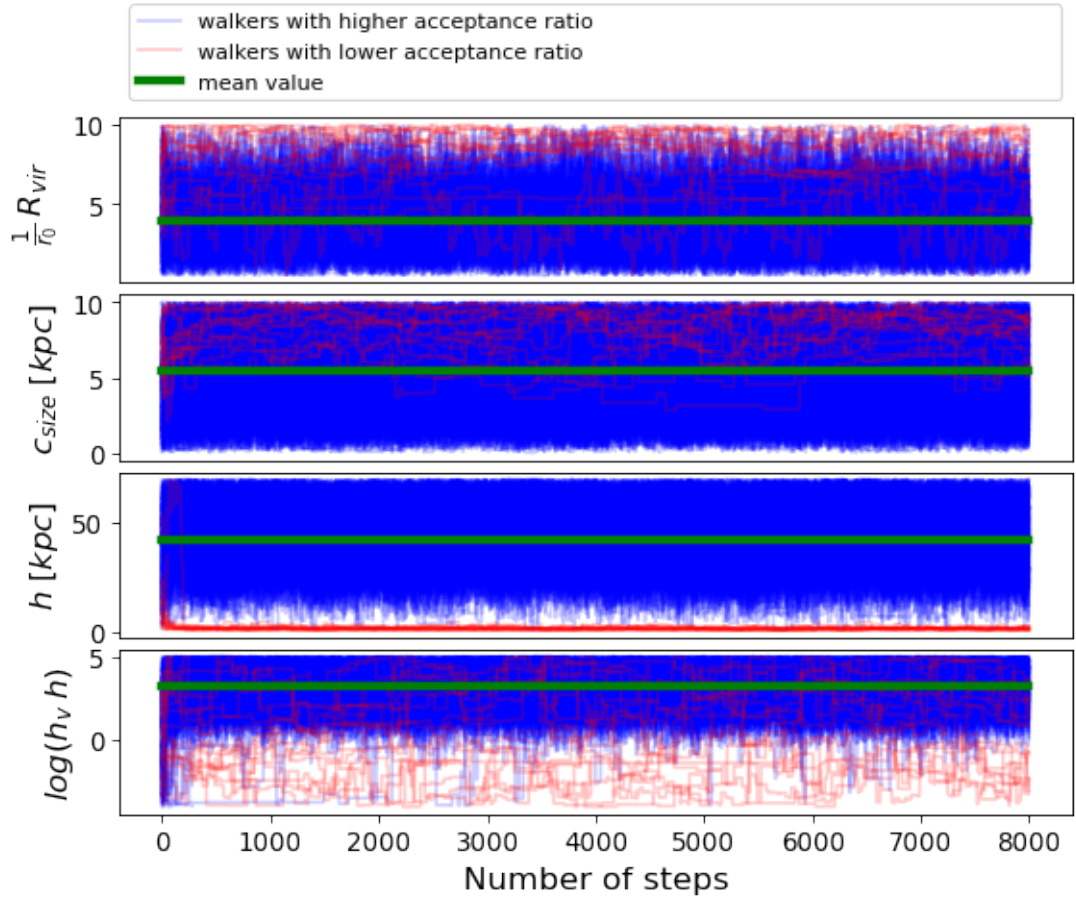


Figure 4.5: Trace plots for MCMC walkers. It shows the parameter values for each parameter and each walker at each step. Blue lines are walkers with high acceptance ratio, red lines are walkers with low acceptance ratio and green lines show the mean value for each parameter. x-axis is the step number.

Figure 4.5 shows the values of the four free parameters that each walker moves to, as a function of number of steps. Blue lines are the walkers with higher acceptance ratio and red lines are walkers with lower acceptance ratio. Green lines show the mean values of the marginalized PDF for each parameter. Ideally the walkers should move around the mean values. This Figure suggest that walkers with low acceptance ratios moved in a different parameter region as walkers with higher acceptance ratios. This is very clear specially in the plot for h (third plot top to bottom), where we can see that walkers plotted in red move toward very low values of h and blue walkers move around the mean value of h .

The difference between low acceptance rate walkers and high acceptance rate walkers is also clear in the other parameters. In the trace plot for $\frac{1}{r_0} R_{vir}$ we can see that red lines move to higher values, which is not the case for blue walkers. The same can be seen in the c_{size} trace plot, where low acceptance walkers tend to go to high values, while high acceptance walkers move though the space around the mean value. The contrary happens in the $\log(h_v h)$ trace plot, where red walkers move in the entire space while blue walkers seem to converge to large values, around the mean value.

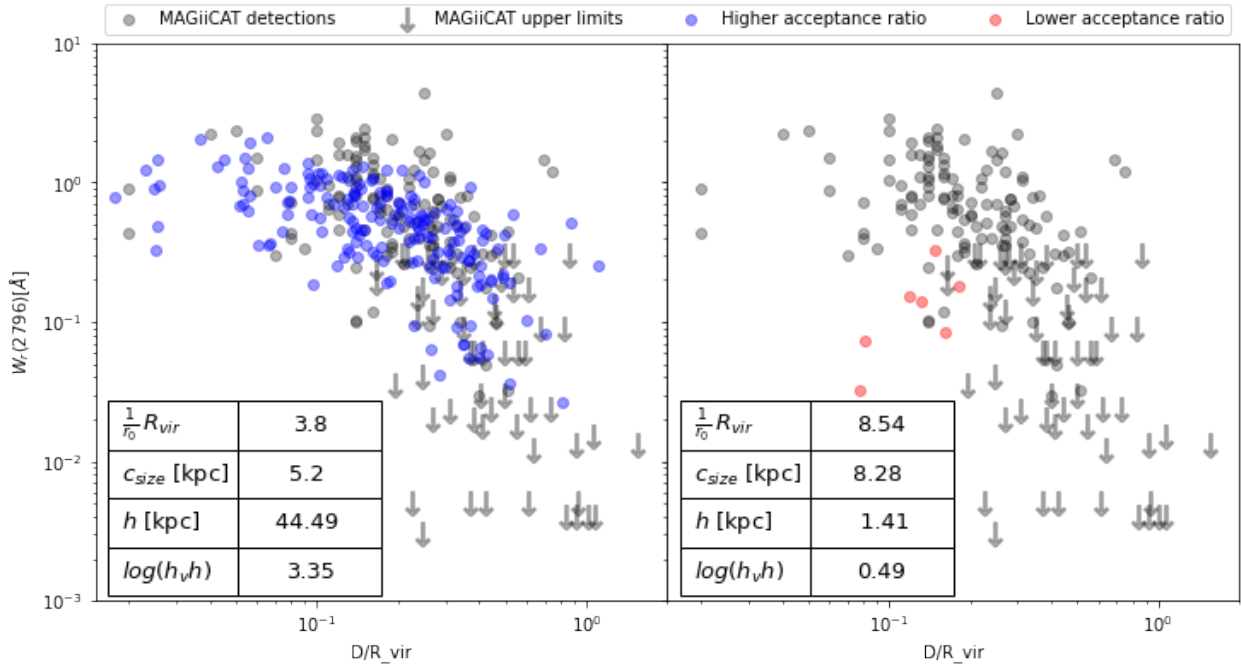


Figure 4.6: $W_r - d/R_{vir}$ relations for model with high probability for walkers with high acceptance ratio (blue points left panel) and walkers with low acceptance ratio (red points right panel). In both panels black dots and arrows show MAGiCAT detections and upper limits respectively. At the bottom left corner in each panel, tables show the mean values of parameters for each group of walkers.

Figure 4.6 shows the $W_r - d/R_{vir}$ relation for the mean values of free parameters for the walkers with high acceptance ratio (blue points) and for walkers with low acceptance ratio (red points), plotted against the values of MAGiCAT (black points). In the left panel we see that the $W_r - d/R_{vir}$ relation for the model with mean values of the high acceptance ratio walkers produce W_r with a distribution very similar to the data. On the right panel we see that low acceptance ratio models produce many unabsorbed sightlines, hence the lack of points in the plot.

It makes sense that parameters on the right panel produce very few absorption lines. The value of $\frac{1}{r_0} R_{vir}$ informs how rapidly the number of clouds decreases with R and c_{size} is a proxy for cloud size, the bigger the clouds, fewer clouds can be placed inside the CGM. Thus a model with very large clouds and a rapidly decreasing $f_v(R/R_{vir})$ produces CGM disks with very few Mg II clouds inside. Thus producing models with very few (7 out of 200) sightlines with $W_r > 0.003\text{\AA}$.

Why do some walkers move to these regions of the parameter space where models produce so few sightlines with absorption lines? This is because of the likelihood function used to compute MCMC. Using a KS test, this statistical test needs a large amount of data (at least 30 points) to be considered a good comparison between distributions. The power of the test is the chance of having false negatives, the lower the number of samples, the worst is the power of the test. The null hypothesis in KS test is that two samples come from the same distribution, thus if the test fails to reject the null hypothesis, for small MAGiCAT, that tells little, due to the high false negative rate.

For this reason, walkers that converge to values of the parameters that produce very few sightlines with $W_r > 0.003\text{\AA}$ are not trustworthy. Looking at the right panel of Figure 4.6, even with the low amount of samples, those sets of parameters do not seem to produce a distribution similar to the observations. Hence, it should be safe to exclude walkers with low acceptance ratios from the MCMC run. Figure 4.7 shows the MCMC corner plots considering only the good 184 walkers. In red are the 2D PDFs produced by the low acceptance ratio walkers, just for comparison.

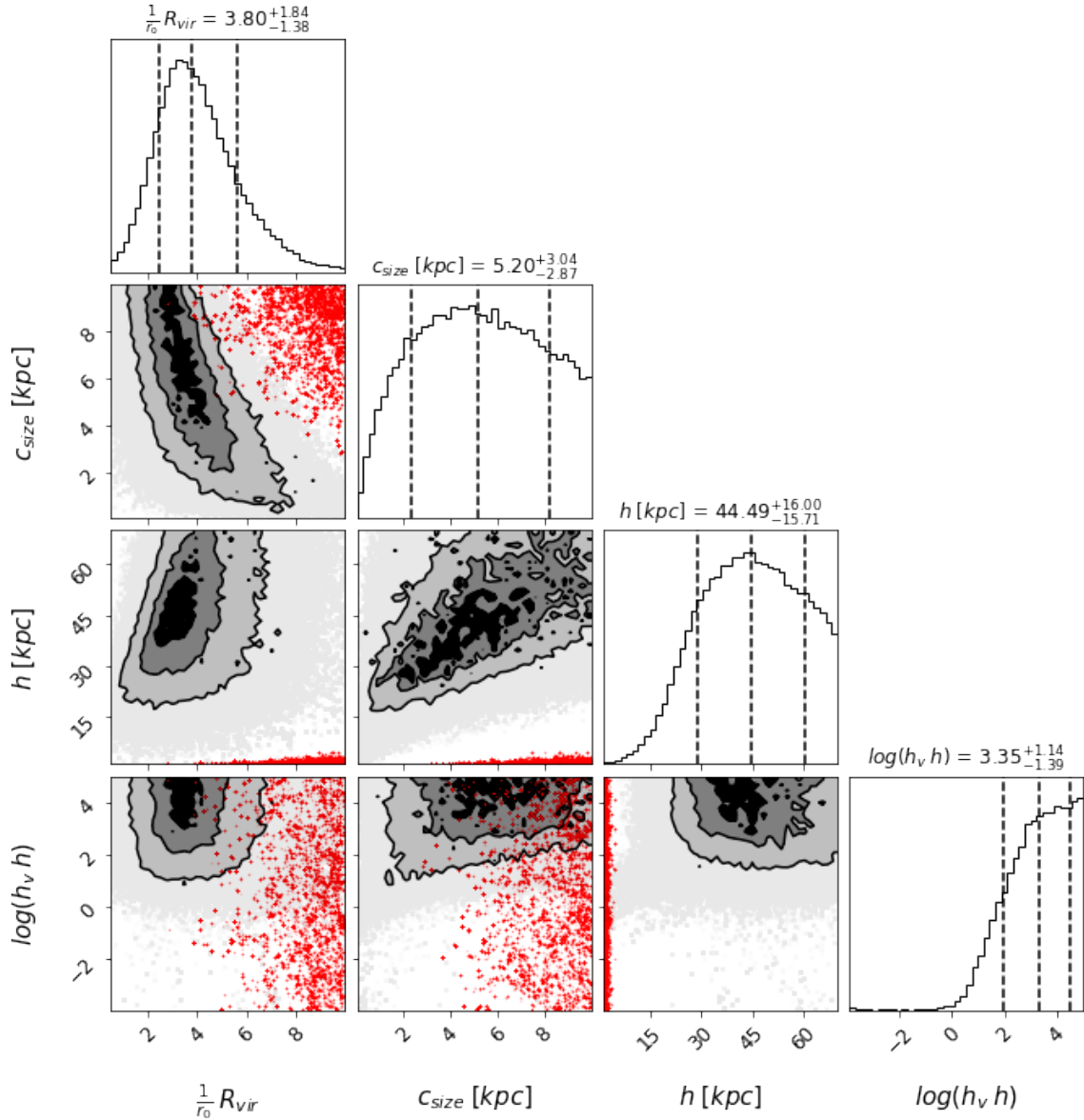


Figure 4.7: MCMC corner plots using only walkers with high acceptance ratios. Same caption as Figure 4.3 but in red are the 2D PDF for walkers with low acceptance ratios, just for comparison.

By excluding bad walkers, I can observe the same tendencies in the marginalized PDFs, but they are more defined, meaning high probability values are even more probable. This is specially noticeable in the h PDF, where in Figure 4.3 there is a very high peak of probability for very low h values, while in Figure 4.7 that later peak completely disappears.

Even clean corner plots show degeneracy between model parameters. Degeneracies are specially noticeable in the $\frac{1}{r_0}R_{vir}$ vs c_{size} PDF and the h vs. c_{size} PDF. I already established that $\frac{1}{r_0}R_{vir}$ and c_{size} relate to the number of clouds inside the CGM disk, and so does h . The thicker the disk, the more clouds of a certain size can be placed inside. The degeneracies suggest that the data is actually constraining the number of individual Mg II clouds inside the CGM disk.

From the 2D PDF of h vs. c_{size} it can be seen that models with $h = 30kpc$ & $c_{size} = 3kpc$ are as probable as models with $h = 60kpc$ & $c_{size} = 8kpc$. This suggests a linear relationship between these two parameters (with a big scatter, as indicated by the large area in the 2D PDF). The relationship suggest that $h \approx 6c_{size} + 12$. In Chapter 5 I discuss what this implies about the number of individual clouds.

For the h_v parameter the results in Figure 4.7 show that regardless of the other three free parameters the model needs to have a very large value of h_v to reproduce the observations. This parameter tells us about the gradient in rotational velocity with the height of the disk (Equation 2.4). If h_v is too small, the velocity decreases very fast with the height, if it is too large it decreases quite slowly. The results show that in order to reproduce MAGiiCAT observations, the model needs to have $\log(h_v h) > 2$. This means that, if the maximum velocity of rotation of the disk is v_{max} , which occurs at the middle of the disk ($z = 0$), at the maximum height of the disk, which is $z = \frac{h}{2}$, the velocity is $0.995v_{max}$. Thus the velocity decreases only a 0.5%, which is almost nothing. Thus, the results suggest that a disk with constant rotation has a higher probability of reproducing the observations.

Summarizing, MCMC corner plots are useful to visualize if the data is able to constrain the model and if so, what areas of the parameter space reproduce the observations. In this case, using MCMC, I was able to find degeneracies between parameters, which indicate that the model is actually constraining the number of clouds in the CGM. I was also able to find out that the model should have little to no gradient in rotational velocity in order to reproduce the data. In Chapter 5 I discuss why this happens, but before, to better understand the model, I need to look into the $W_r - d/R_{vir}$ relation produced with different set of parameters. This allows us to understand the physical implications of the MCMC results and why they show what they show.

4.2.1. $W_r - d/R_{vir}$ relations

Why do some sets of parameters have higher probabilities than others, as shown by the corner plots (Figure 4.7) ? To address this effect I choose a high probability value and a low probability value for each of the parameters of the model and produce $W_r - d/R_{vir}$ relations with all possible combinations of them. Since I have four free parameters and I take two values for each and produce models with each possible combination of them, I have 16 different models. By plotting the resulting distributions one can better understand the MCMC results.

According to MCMC results, Table 4.1 shows a high and a low probability value for each free parameter.

Table 4.1: High and low probability values for each parameter of the model according to MCMC

Free paramater	High probability	Low probability
$\frac{1}{r_0} R_{vir}$	3	8
c_{size}	5kpc	0.1kpc
h	45 kpc	5kpc
$\log(h_v h)$	4	-2

Figure 4.8 shows the $W_r - d/R_{vir}$ relation for all 16 possible models that can be produced from a combination of high and low probability values. Every plot shows observational W_r values and upper limits (in grey), for comparison with the model relationship. Blue points indicate the values for models generated with high probability $\frac{1}{r_0} R_{vir}$, orange points show model with low probability $\frac{1}{r_0} R_{vir}$.

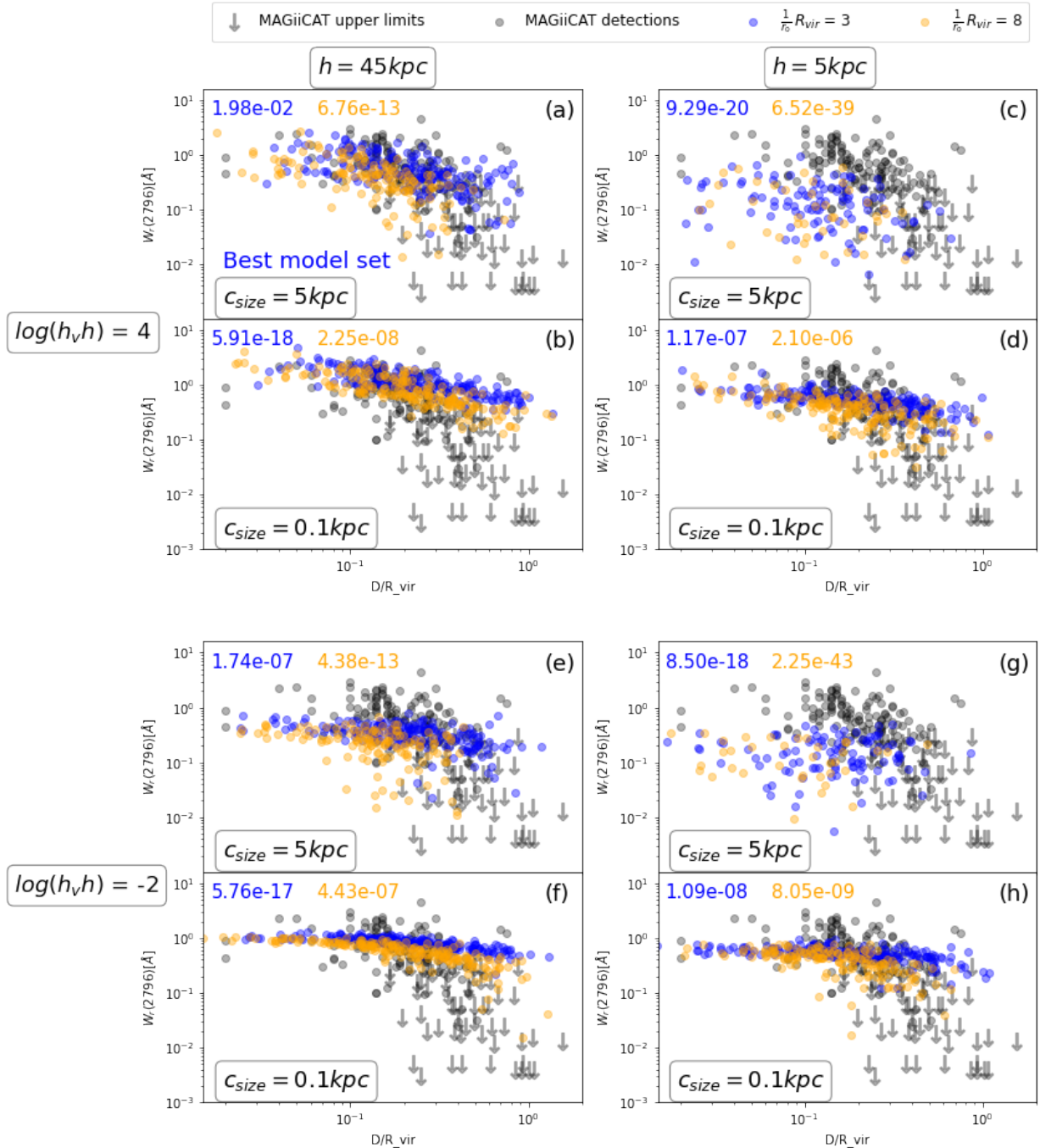


Figure 4.8: Model $W_r - d/R_{vir}$ relationship for different sets of parameters. In each panel there are grey points and arrows showing MAGiCAT detections and upper limits, blue points showing models with $\frac{1}{r_0} R_{vir} = 3$ and orange points showing models with $\frac{1}{r_0} R_{vir} = 8$. Panels in the left column (a,b,e and f) show models with $h = 45kpc$ and on the right column panels (c,d,g and h) are models with $h = 5kpc$. The first row of panels (a,b,c and d) show models with $\log(h_v, h) = 4$ and bottom row panels (e,f,g and h) show models with $\log(h_v, h) = -2$. Finally in each corner of the figure are two panels, all top panel at each corner (a,c,e and g) show models with $c_{size} = 5kpc$ and bottom panels at each corner (b,d,f and h) show models with $c_{size} = 0.1kpc$. In each panel there are two values annotated, in the upper left corner, one in blue, one in orange, those are the p -values for each model in each panel.

These are three main takeaways from Figure 4.8:

1. **Exponential filling factor:** Models with large $\frac{1}{r_0} R_{vir}$ values (orange points) produce a faster decrease of W_r with d/R_{vir} than models with small $\frac{1}{r_0} R_{vir}$ values (blue points), regardless of the other parameters. If f_v decreases too fast with R/R_{vir} , the model $W_r - d/R_{vir}$ relation also decreases too fast, which is not consistent observations.
2. **Height and cloud size:** If c_{size} is too small compared with h (panel (b) and (f)), the model fails to reproduce small W_r values. If the clouds are too large compared to h (panel (c) and (g)) the model fails to reproduce large W_r values. This is because small clouds imply many clouds in the sightline and thus always larger W_r . MCMC results show that there needs to be a relationship of $h \approx 6c_{size} + 12$ in order to reproduce observational $W_r - d/R_{vir}$ relation.
3. **Velocity scale height:** Models with small $\log(h_v/h)$ values fail to reproduce the large W_r values we see in observations (panels (e), (f), (g) and (h)). The disk needs to have a constant rotation (panel (a)) in order to reproduce $W_r - d/R_{vir}$ relation. If the velocity decreases with the height of the disk, too many clouds have low (and similar) velocities and do not add to the final W_r and are not able to reproduce the large MAGiiCAT values.

These results gives an idea on how $W_r - d/R_{vir}$ is affected by the individual model parameters, how the MCMC results should be understood and how robust the present CGM model is. The interconnection between parameters h and c_{size} , required to match the observations, are the second result of this thesis. In Chapter 5 I discuss these results and later summarize the main conclusions.

4.2.2. Number of clouds

MCMC results show a large degeneracy between c_{size} and h , where the relationship that needs to occur is $h = 6c_{size} + 12$, with a large scatter as shown by the 2D PDF (Figure 4.7). c_{size} is related to the number of clouds, where small c_{size} allows for more clouds to be placed inside the CGM disk and larger c_{size} translate to fewer clouds. h also relates to the number of clouds, higher h means more volume for clouds to be placed inside the CGM disk and lower h translates to thinner disks, thus less volume to place clouds.

This degeneracy suggests that the data is constraining the number of individual MgII clouds, instead of c_{size} and h . Churchill et al. (2020) measure the distribution of number of different Voigt profiles they could fit in 422 different high resolution Mg II absorption systems. This number allows them to measure the number of individual Mg II clouds that are crossed by a single sightline.

Figure 4.9 shows in red the number of clouds distribution measured in Churchill et al. (2020) in random sightlines. It can be seen that the individual sightlines cross between 1 and 40 individual clouds, where fewer clouds are more probable than many clouds, except for the first bin. I generated three different models with the best fit $\frac{1}{r_0} R_{vir} = 3.8$, $h_v = 10^4 h$ and height, $h = 45$ kpc, but different c_{size} values: 5.5 kpc (blue), 0.1 kpc (green) and 15 kpc (orange). Figure 4.9 shows the distribution of number of clouds for each of the models in the respective colors. Blue is the best fit c_{size} , green is a small c_{size} and orange is a big c_{size} .

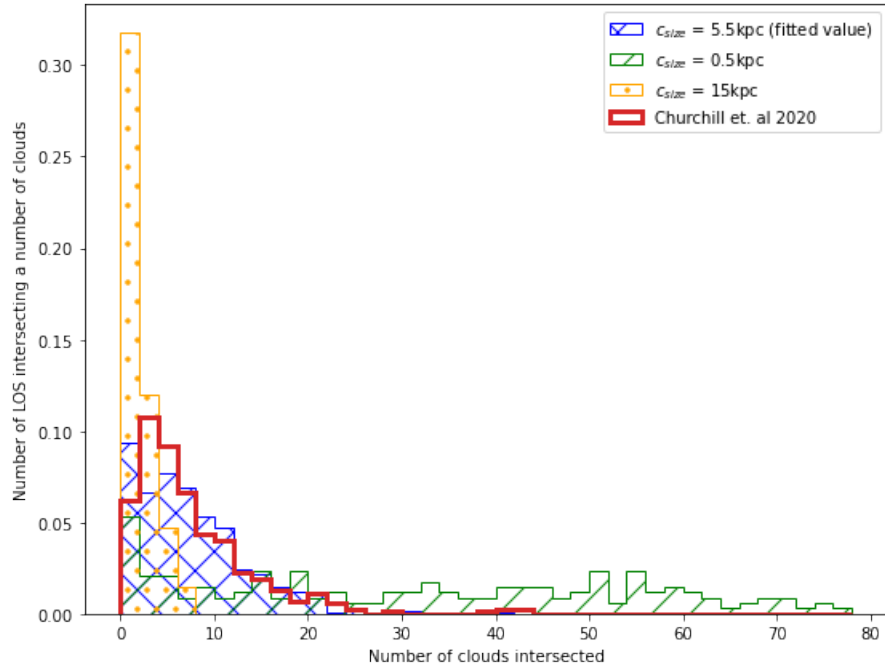


Figure 4.9: Distribution of number of clouds crossed a sightline. Red histogram shows distribution from Churchill et al. (2020), which comes from Voigt profile fitting to high resolution spectra of random sightlines. Blue, green and orange histograms show the number of clouds for three different models, all of them with $h = 44.45 \text{ kpc}$, but different c_{size} , $c_{size} = 4 \text{ kpc}$, $c_{size} = 0.1 \text{ kpc}$, $c_{size} = 15 \text{ kpc}$, which translate to different number of clouds densities.

It can be seen, in Figure 4.9, that the fitted model produces a number of clouds distribution very similar to Churchill et al. (2020). The model with small clouds produces a distribution with larger number of clouds than what we observe, thus confirming that small clouds translate to more clouds than we expect. The opposite happens with the model with bigger clouds, the number of clouds distribution shows fewer clouds than what is expected from observations, thus confirming that big clouds translate to a small number of clouds.

According to MCMC results models with different c_{size} values that follow the relationship $h = 6c_{size} + 12$ (with a big scatter), are a good fit to observations. Figure 4.10 shows again in red the number of clouds distribution from Churchill et al. (2020). I generated three different models with different h values: 30 kpc (blue), 60 kpc (green) and 90 kpc (orange). All of the models have c_{size} values that follow the fitted relationship.

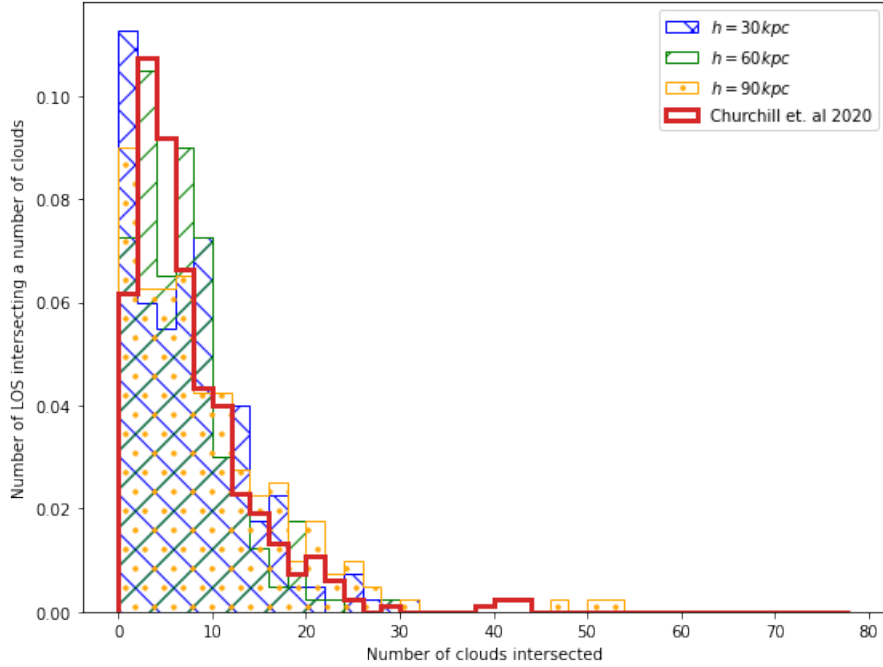


Figure 4.10: Number of clouds crossed by a single distribution. Red histogram shows distribution from Churchill et al. (2020), which comes from Voigt profile fitting to high resolution spectra. Blue, green and orange histograms show the number of clouds for three different models, with $h = 30kpc$, $h = 60kpc$ and $h = 30kpc$ respectively. All of the models were created with c_{size} values that follow $h = 6c_{size} + 12$.

It can be seen, in Figure 4.10, that all three models produce a number of clouds distribution similar to Churchill et al. (2020), further confirming that the data is constraining the number of individual clouds, thus we observe a big degeneracy between h and c_{size} . The results suggest that the sightlines need to intersect between 0 and 40 individual clouds to reproduce the data. Therefore, this test provides a second confirmation for the best fit model by comparing the results to Churchill et al. (2020) results.

4.3. Two-point correlation function results

As has been shown, the $W_r - d/R_{vir}$ relation is a useful observable to constrain the morphology and kinematics of the CGM. However, it has some limitations. The data include a large number of sightlines crossing different types of galaxies, having a variety of geometries. To better constrain CGM structures, it seems convenient to select sub-samples according to azimuthal angle and inclination.

Nielsen et al. (2013) use the TPCF to study the velocity dispersion found in different regions of the CGM (see Subsection 1.3.4 for a detailed description of Nielsen et al. (2013)). Those authors compute the TPCF for different sub-samples of sightlines: near the galaxy major axis, near the galaxy minor axis, for face-on galaxies, for edge-on galaxies, and also for all possible combinations between the two groups of α and inclinations (Table 1.2). In this chapter, a comparison is made between the present CGM model and MAGiCAT TPCF for different sub-samples. The

only downside of this comparison is that the TPCF requires high-resolution spectra, making the MAGiCAT TPCF dataset (17 sightlines for the largest sub-sample) much smaller than that for the $W_r - d/R_{\text{vir}}$ data.

Despite the smaller sample, Nielsen et al. (2013) are able to find trends (See Figure 1.6). Sightlines near the major axis have a larger velocity dispersion for edge-on galaxies, while sightlines near the minor axis show larger velocity dispersion for face-on galaxies. Nielsen et al. (2013) suggest two different components in the CGM to explain these results: a rotational component in the plane of the disk, and an outflowing component. The model presented in this thesis consists only of a rotational component in the plane of the disk.

The model TPCF is calculated (see Section 3.2 for the same sub-samples as in Nielsen et al. (2013) (See Table 1.2) and using the best-fit parameters found in the previous section. The idea is to see if by reproducing the $W_r - d/R_{\text{vir}}$ the model can also reproduce the TPCF. The rationale is the following: if the best fit model can reproduce the TPCF for all sub-samples, it can constrain a single rotational component, which could explain the observed velocity dispersion, without the need of an outflowing component. If it does not reproduce the observed velocity dispersion, the model is incomplete.

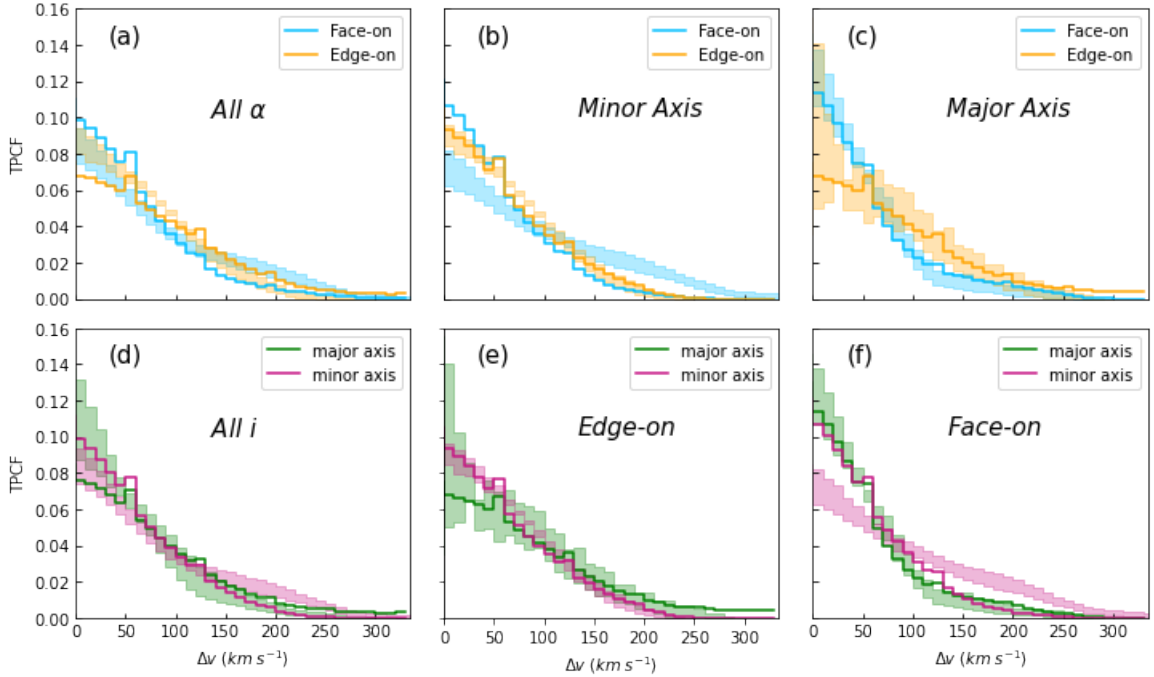


Figure 4.11: Pixel two-point correlation function computed from the model with higher probability of reproducing $W_r - d/R_{\text{vir}}$ relation, in different sub-samples of galaxy inclination and azimuthal angle. Panel (a) shows all sightlines, (b) shows only sightlines near the minor axis, (c) shows only sightlines near the major axis, (d) shows all galaxies, (e) shows only edge on galaxies and (f) show only face-on galaxies. Light blue represent face-on galaxies, orange represent edge-on galaxies, green represents sightlines near the major axis and pink represents sightlines near the minor axis. Solid lines are the TPCF for the model and colored areas show the 1σ intervals for MAGiCAT data.

Figure 4.11 shows the model TPCF for the same sub-samples as in Nielsen et al. (2013). The x-axis shows the pixel velocity absolute difference bins and y-axis shows the PDF at each velocity bin. Different colors in each panel represent different sub-samples. Solid lines are the model TPCFs for each sub-sample, where the model was computed with the MCMC best-fit parameters, and colored areas show the 1σ intervals of MAGiCAT's TPCFs in each sub-sample.

Visual inspection of Figure 4.11 shows in which sub-sample the best fit model performs best. If the TPCF shows a larger distribution towards larger values of Δv , that sub-sample has a larger velocity dispersion. This visual comparison shows that:

1. Panel (a): The model produces smaller velocity dispersion in face-on galaxies compared with edge-on galaxies. However, observations show that face-on galaxies tend to have larger velocity dispersion.
2. Panel (b): For sightlines near the minor axis, the model produces very similar velocity dispersion in edge-on galaxies and face-on galaxies. However, observations show that for sightlines near the minor axis, face-on galaxies produce larger velocity dispersion.
3. Panel (c): For sightlines near the major axis, the model is able to reproduce the observations. Sightlines that cross edge-on galaxies have larger velocity dispersion than face-on galaxies.
4. Panel (d): For all galaxies, the model produces larger velocity dispersion in sightlines near the major axis, this is the opposite in observations, where sightlines near the minor axis produce larger velocity dispersion.
5. Panel (e): For edge-on galaxies, the model is able to reproduce the observational TPCF in sightlines near the major axis and in sightlines near the minor axis.
6. Panel (f): For face-on galaxies, the model produce very similar velocity dispersion in sightlines near the major and minor axis. However, observations show a larger velocity dispersion in sightlines near the minor axis.

In summary, the present CGM model is able to reproduce the TPCF for some of the sub-samples, specifically sightlines near the major axis, for edge-on and face-on galaxies, but not all of them. It is important to mention that this is true only for the MCMC best-fit parameters. I did not perform any analysis on other values for the free parameters of the model, because of large computational times (Subsection 3.3.3) thus I do not know if any set of parameters would be able to reproduce all the sub-samples. However, the results shown in Figure 4.11 do have some physical implications for the model. These implications are discussed in Section 5.

4.4. Summary of the results

After having compared the present CGM model and the observational Mg II MAGiCAT data, the results are as follows:

Filling factor exploration:

1. The volume filling factor of Mg II clouds in the $z \sim 1$ CGM, $f_v(R/R_{vir})$, needs to be a smoothly decreasing function (Figure 4.1).

2. A decreasing exponential for $f_v(R/R_{vir})$ produces $W_r - d/R_{vir}$ model data that is qualitatively consistent with the observations (Figure 4.2).

MCMC simulations:

1. The median of the parameters PDF of a rotating CGM model, given $W_r - d/R_{vir}$ in the MAGiCAT data, are: the inverse of the characteristic radius of the exponential f_v in terms of R_{vir} ($\frac{1}{r_0} R_{vir} = 3.80^{+1.84}_{-1.38}$), the size of the individual Mg II clouds in kpc ($c_{size} = 5.2^{+3.04}_{-2.83}$), the height of the disk in kpc ($h = 44.49^{+16.00}_{-15.71}$) and the logarithm of the velocity scale height in terms of the height of the disk ($\log(h_v/h) = 3.35^{+1.14}_{-1.39}$) (Figure 4.7).
2. h and c_{size} are correlated, suggesting that the data is constraining the number of clouds inside the disk, rather than h and c_{size} . The correlation that best reproduces the observations is $h = 6c_{size} + 12$ (Figure 4.7).
3. The fact that $h_v \gg h$ indicates that a disk with constant rotation best reproduces the data.
4. Although not a model parameter, the best-fit model predicts a distribution of the number of clouds crossed by the sightlines, which is consistent with the observations (Figure 4.10).

Two-point correlation function:

1. The MCMC "best-fit" model, given the $W_r - d/R_{vir}$ data, cannot reproduce the TPCF of MAGiCAT sightlines near the minor axis (Figure 4.11 panel(b)).
2. The best-fit model can reproduce the TPCF of MAGiCAT sightlines along the major axis (Figure 4.11 panel(c)).

These results are further discussed in Chapter 5.

Chapter 5

Discussion and Conclusions

In this section I discuss the results of fitting an extended rotating disk model for the cool CGM to real data, and its implications for our understanding of the CGM. First I discuss all the results presented in Chapter 4 (Section 5.1), then I present the main conclusions (Section 5.2) and lastly I propose possible future work (Section 5.3).

5.1. Discussion

5.1.1. MgII spatial distribution

I have shown that, if the CGM has a disk morphology filled with individual Mg II clouds, the cloud number needs to smoothly decrease with distance from the galaxy (see Figure 5.1). I assumed an exponential f_v which does reproduce observations (see Figure 3.2), although other functions are possible (Stern et al., 2016).

The result on $\frac{1}{r_0}R_{vir}$ implies that the characteristic radius for the exponential f_v , $r_0 = 0.26^{+0.15}_{-0.09}R_{vir}$. Thus the mean value of f_v (for sightlines crossing between $0.02R_{vir} < d < 1.51R_{vir}$ (which is the MAGiCAT sample range) goes from $f_v = 16\% - 38\%$. On the other hand, the COS-Halos survey (Werk et al., 2014) find a median value of $f_v = 11^{+15}_{-9}\%$ for Mg II at impact parameters $d \leq 160kpc$ to 44 galaxies, which is consistent with my results.

$r_0 = 0.26R_{vir}$ suggests that the gas fraction bearing Mg II decreases to 63% at $R=0.26 R_{vir}$, assuming 100% at $R = 0$. Thus, beyond $0.26 R_{vir}$ Mg II is very diffuse. Both models shown in Figure 5.1, have 28% of the Mg II clouds within $R < 0.26R_{vir}$ and 91% of the gas within $R < R_{vir}$. Hydrodynamical simulations show that accretion occurs in cold streams that acquire angular momentum similar to the host galaxy (Stewart, 2017) and that these streams could be seen inside one virial radius. Thus some of the rotating gas in the present model could be expected to come from accretion.

Using extended background sources Rubin et al. (2018) find that the scale length over which $W_r(2796)$ does not vary (i.e., the “coherence scale” of Mg II) is $l_A > 1.9kpc$. Cosmological simulations that predict the presence of inflows also suggests coherence scales of ≈ 2 kpc (Stewart et al. 2011, 2013). The results in this thesis further confirm these results, where the best fit model indicates that cloud sizes are ≈ 3.8 kpc.

However, the model also shows degeneracy between c_{size} and h . This occurs because the data are rather constraining the number of clouds. The fitted relationship produces number of clouds crossed by the sightlines similar to the one observed in Churchill et al. (2020). Figure 4.9 shows the distribution of number of clouds for three fitted models and for Churchill et al. (2020) observations. It can be seen that the sightlines need to intersect between 0 and 30 individual clouds to reproduce the data. Given this diagnosis I have a second confirmation of this range by comparing the results to Churchill et al. (2020). However, this comparison needs to be taken with caution, because even though the number of fitted Voigt profiles can give us a proxy of the number of clouds intersected by a sightline, it is not necessary the same as the model, because observations are limited due to blending. Thus observations actually provide a lower limit on the real number of clouds in the line-of-sight.

A typical MAGiCAT galaxy has $R_{vir} \approx 200$ kpc. Figure 5.1 shows a plot of how two typical CGM disks would look like, for galaxies with $R_{vir} = 200$ kpc, with the fitted parameters. Panel (a) shows a CGM disk of $h = 30$ kpc and panel (b) shows a CGM disk of $h = 60$ kpc. Both of these models are reproduced using the fitted f_v . Upper panels show a view in the plane of the disk and bottom panels show a side view of the disk. Light blue areas represent the R_{vir} and h of the disk. Blue points represent the individual Mg II. Just for displaying purposes, the individual clouds are shown as spheres (the model uses a rectangular grid). It can be seen how there is a big density of clouds near the center of the disk, that gets smaller far away, for both models.

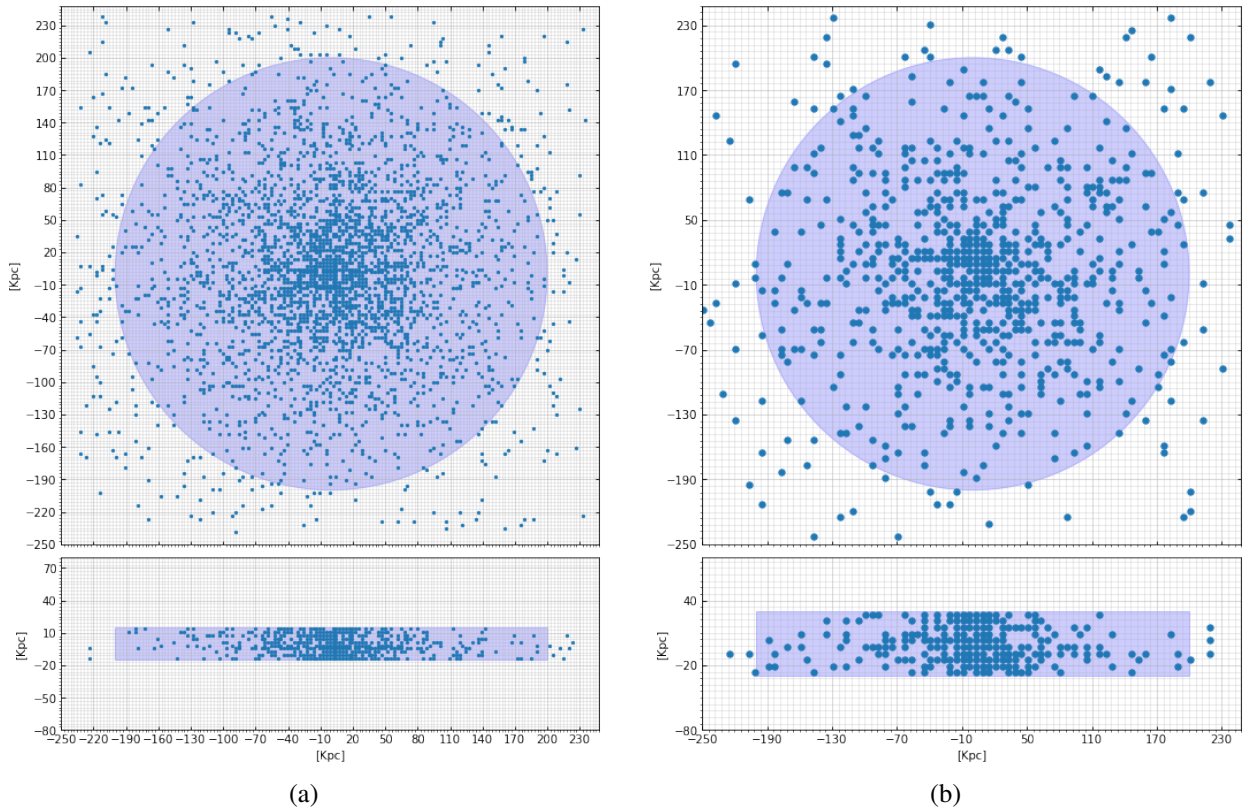


Figure 5.1: Schematic picture of two different CGM models with high probability. panel (a): model with $h = 30$ kpc and $c_{size} = 3$ kpc. panel (b): model with $h = 60$ kpc and $c_{size} = 6$ kpc

It is important to mention that there are some disadvantages of using MAGiCAT data. Some of the absorption systems in their sample come from absorption selected samples, which can lead to biases over-representing sightlines with strong W_r . Thus the fitted distribution of Mg II clouds could also be biased.

5.1.2. Kinematics of the CGM

The kinematic model presented in this work is inspired by the kinematic model presented in Steidel et al. (2002) (to explain individual QSO sightlines, not to draw statistical conclusions like the present work). Steidel et al. (2002) propose a rotating disk with a vertical velocity gradient, the same used in this work. Their results show that in order to reproduce the observed velocity dispersion the disks need to be extremely thick and/or have a rapidly decreasing rotational velocity in their vertical component. As shown above, this work, using $W_r - d/R_{\text{vir}}$ relationship as a diagnosis, MCMC minimization favors a disk with constant rotation.

I analyzed velocity dispersion by comparing the model and observational TPCF (Figure 4.11). It can be seen that the best fit model reproduces Nielsen et al. (2015) results for sightlines near the major axis but not for sightlines near the minor axis. Some observations (see Veilleux et al. (2005)) suggest the presence of gas flowing out of the galaxy in a bi-conical way. Nielsen et al. (2015) conclude that sightlines near the major axis are probing rotating gas, as proposed by this thesis, while sightlines near the minor axis are probing an outflowing component in the CGM, which was not a part of the proposed kinematic model.

Previous works found rotating material is usually probed in edge-on galaxies with sightlines near the major axis (Steidel et al., 2002; Kacprzak et al., 2010; Rubin et al., 2012; Bouché et al., 2012). Figure 4.11 panel (c) shows the TPCF for sightlines near the major axis for face-on galaxies (blue) and edge-on galaxies (blue), for the model and for observations. It can be seen that both sub-samples produce TPCF similar to the observations. The velocity dispersion is greater for edge-on galaxies than for face-on galaxies. Nielsen et al. (2015) state that this may be due to rotating gas whose line of sight velocity is maximized in edge-on inclinations, while the vertical velocity dispersions in the disks of galaxies are small. This agrees with the model, which consists in rotating gas and from MCMC results we obtained that the vertical velocity dispersion is very small or non existing.

Thus, the best fit model for a rotating disk, using only $W_r - d/R_{\text{vir}}$, is able to reproduce the results that suggest rotation. This gives a second constraint on the model and leads one to further confirm that the model can reproduce observational data. This implies that a rotating thick disk, aligned with the galaxy disk, populated with individual Mg II clouds whose number density decreases exponentially is a possible model for the CGM. These results apply to the "best-fit" parameters, it cannot be discarded that other parameters could also reproduce the observations for sightlines near the major axis.

Even though the model can reproduce the TPCF for sightlines near the major axis it fails at reproducing other sub-samples. Figure 4.11 (b) shows the TPCF for sightlines near the minor axis, for face-on and edge-on galaxies for MAGiCAT sample and the model. Observations show a larger velocity dispersion in face-on galaxies, Nielsen et al. (2015) associate this with outflows.

When looking at face-on galaxies they observe through the outflows, thus the velocity dispersion is greater. As showed by the models TPCFs, a single rotating disk cannot reproduce these results everywhere, it produces very similar velocity dispersion for edge-on and face-on galaxies.

These results show that even though the model can reproduce observations that suggest rotation, a single rotating disk is not sufficient to also explain the different velocity dispersions found along the minor axis of galaxies with different inclinations. This suggests that the model could be missing a kinematic component, maybe bi-conical outflows as the observations suggest. This analysis was done only using the best fit parameters of the model, thus I cannot discard that other sets of parameters could reproduce these observations.

5.2. Summary and conclusions

In this MSc thesis I have proposed a morpho-kinematic model for the cool CGM. The model consists of a single rotating disk populated with individual Mg II clouds, which rotates and has a vertical velocity gradient. In total, the model has 4 free parameters: the characteristic radius of the exponential filling factor, r_0 ; the grid size, which is a proxy for cloud size, c_{size} ; the height of the disk, h ; and the velocity scale height h_c . To properly compare model and observations the model enables synthetic Mg II absorption line spectra at different impact parameters and azimuthal angles and for different halo masses.

To constrain the model I have used QSO absorption line data presented in MAGiiCAT (Nielsen et al., 2013). By taking advantage of 182 observed W_r -vs- d points and the two-point-correlation function of different sub-samples in MAGiiCAT I created synthetic catalogs with similar characteristics. An MCMC minimization scheme led to best-fit values for: filling factor ($f_v(R/R_{vir})$), grid size (c_{size}), height of the disk (h) and velocity scale height (h_v). On the other hand, the observed TPCF was used to discuss these parameters in a qualitative fashion.

The main conclusions are:

1. The model is able to provide realistic QSO spectra, at least judging from visual comparison. However each model can take up to 10 minutes to compute on a 20 nodes CPU (Subsection 3.3.3). These times can eventually get considerably long and prevent running MCMC with a large number of walkers and steps.
2. MCMC results show that a disk morphology for the CGM is capable of explaining the $W_r - d/R_{vir}$ relation we see in QSO observations.
3. A single rotational component can explain observations of $W_r - d/R_{vir}$ relation only if the vertical velocity gradient is negligible, meaning that a disk with constant rotational velocity is the best model representation given the MAGiiCAT data.
4. The distribution of Mg II clouds (or the number of clouds per unit volume) needs to decrease smoothly with distance to the galaxy. The disk filling factor can be an exponential with $r_0 = 0.26R_{vir}$.
5. The observational number of clouds that is crossed by individual sightlines is reproduced from the best fit model. Results show that sightlines cross between 1 and 30 MgII clouds, which

means that there needs to be a relationship between the height of the disk and the grid size, which is: $h = 6c_{size} + 12$.

6. Even though the model is able to reproduce the $W_r - d/R_{vir}$ statistics in QSO sightlines, the velocity dispersion, showed by the TPCF, is not totally explained by this model. For geometries where we expect to find rotation, the model explains the observational TPCF. For geometries where outflows are expected the model single rotation disk model cannot explain the observed large velocity dispersion.

5.3. Future work

The available QSO absorption line statistics offers plenty of opportunities to test CGM models like the one presented here. In this particular case, I have shown that more constrains are necessary to completely accept or reject the model. For instance, adding the TPCF as an extra constraint within the MCMC run would perhaps lead to parameters able to reproduce both the $W_r - d/R_{vir}$ relation and the TPCF. The extra constraint could perhaps solve the degeneracies we observe in the current fit (although another possible way to constrain the model without degeneracies would be to make c_{size} depend on h and/or R_{vir} , which was not tested here).

The present MCMC results use only one observed relation, $W_r - d/R_{vir}$. This relation combines every possible sightline. Since the proposed methodology allows one to create several synthetic spectra from a model, one could use more than one constraint. For example one could use W_r measurements for different azimuthal angles or inclinations, or one could use the observable covering fraction, or the two-point correlation function, or the distribution of equivalent widths. This is proposed for future work.

Of course another possible continuation of this work is trying alternative models. Ho et al. (2017) proposed a model similar to the one used in this work but with an extra inflow velocity in the clouds. Using the methodology in this thesis one could constrain the inflow velocity of the clouds. Another possible model is a CGM with bi-conical outflows, as suggested by the present TPCF results and other works.

Bibliography

- Anderson, M. E. & Bregman, J. N. 2010, , 714, 320
- Anderson, M. E., Bregman, J. N., & Dai, X. 2013, , 762, 106
- Bahcall, J. N. & Spitzer, Lyman, J. 1969, , 156, L63
- Barcons, X., Lanzettat, K. M., & Webb, J. K. 1995, , 376, 321
- Barton, E. J. & Cooke, J. 2009, , 138, 1817
- Behroozi, P. S., Conroy, C., & Wechsler, R. H. 2010, , 717, 379
- Bergeron, J. & Boissé, P. 1991, , 243, 344
- Bordoloi, R., Lilly, S. J., Knobel, C., et al. 2011, , 743, 10
- Bouché, N., Dekel, A., Genzel, R., et al. 2010, , 718, 1001
- Bouché, N., Hohensee, W., Vargas, R., et al. 2012, , 426, 801
- Bouché, N., Murphy, M. T., Péroux, C., Csabai, I., & Wild, V. 2006, , 371, 495
- Bregman, J. N., Anderson, M. E., Miller, M. J., et al. 2018, , 862, 3
- Chen, H.-W. 2017, in *Astrophysics and Space Science Library*, Vol. 430, *Gas Accretion onto Galaxies*, ed. A. Fox & R. Davé, 167
- Chen, H.-W., Helsby, J. E., Gauthier, J.-R., et al. 2010a, , 714, 1521
- Chen, H.-W., Lanzetta, K. M., Webb, J. K., & Barcons, X. 1998, , 498, 77
- Chen, H.-W. & Tinker, J. L. 2008, , 687, 745
- Chen, H.-W., Wild, V., Tinker, J. L., et al. 2010b, , 724, L176
- Christensen, C. R., Davé, R., Governato, F., et al. 2016, , 824, 57
- Churchill, C. W., Evans, J. L., Stemmock, B., et al. 2020, , 904, 28
- Churchill, C. W., Kacprzak, G. G., & Steidel, C. C. 2005, in *IAU Colloq. 199: Probing Galaxies through Quasar Absorption Lines*, ed. P. Williams, C.-G. Shu, & B. Menard, 24–41
- Churchill, C. W., Mellon, R. R., Charlton, J. C., et al. 2000, , 130, 91
- Churchill, C. W., Nielsen, N. M., Kacprzak, G. G., & Trujillo-Gomez, S. 2013, , 763, L42

Churchill, C. W., Steidel, C. C., & Vogt, S. S. 1996, , 471, 164

Churchill, C. W. & Vogt, S. S. 2001, , 122, 679

Cooke, J. & O'Meara, J. M. 2015, , 812, L27

Dekel, A. & Mandelker, N. 2014, , 444, 2071

Dutta, R., Fumagalli, M., Fossati, M., et al. 2020, , 499, 5022

Evans, J. L. 2011, PhD thesis, New Mexico State University

Fasano, G. & Franceschini, A. 1987, , 225, 155

Ford, A. B., Davé, R., Oppenheimer, B. D., et al. 2014, , 444, 1260

Ford, A. B., Oppenheimer, B. D., Davé, R., et al. 2013, , 432, 89

Gauthier, J.-R. & Chen, H.-W. 2011, , 418, 2730

Guillemin, P. & Bergeron, J. 1997, , 328, 499

Gupta, A., Mathur, S., Krongold, Y., Nicastro, F., & Galeazzi, M. 2012, , 756, L8

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

Heckman, T. M., Alexandroff, R. M., Borthakur, S., Overzier, R., & Leitherer, C. 2015, , 809, 147

Ho, S. & Turner, E. L. 2011, , 739, 26

Ho, S. H., Martin, C. L., Kacprzak, G. G., & Churchill, C. W. 2017, , 835, 267

Huang, Y.-H., Chen, H.-W., Johnson, S. D., & Weiner, B. J. 2016, , 455, 1713

Huang, Y.-H., Chen, H.-W., Shectman, S. A., et al. 2021, , 502, 4743

Hummels, C. B., Bryan, G. L., Smith, B. D., & Turk, M. J. 2013, , 430, 1548

Johnson, S. D., Chen, H.-W., & Mulchaey, J. S. 2015, , 452, 2553

Kacprzak, G. G., Churchill, C. W., Barton, E. J., & Cooke, J. 2011, , 733, 105

Kacprzak, G. G., Churchill, C. W., Ceverino, D., et al. 2010, , 711, 533

Kacprzak, G. G., Churchill, C. W., Steidel, C. C., & Murphy, M. T. 2008, , 135, 922

Kacprzak, G. G., Churchill, C. W., Steidel, C. C., Murphy, M. T., & Evans, J. L. 2007, , 662, 909

Kacprzak, G. G., Martin, C. L., Bouché, N., et al. 2014, , 792, L12

Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, , 363, 2

Kornei, K. A., Shapley, A. E., Martin, C. L., et al. 2012, , 758, 135

Lan, T.-W. & Mo, H. 2018, , 866, 36

Lanzetta, K. M. & Bowen, D. 1990, , 357, 321

Lanzetta, K. M., Wolfe, A. M., & Turnshek, D. A. 1995, , 440, 435

Ledoux, C., Petitjean, P., Bergeron, J., Wampler, E. J., & Srianand, R. 1998, , 337, 51

Lee, K.-G., Hennawi, J. F., Stark, C., et al. 2014, , 795, L12

Liang, C. & Kravtsov, A. 2017, arXiv e-prints, arXiv:1710.09852

Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, , 772, 119

Lopez, S. & Jenkins, J. S. 2012, , 756, 177

Lopez, S., Tejos, N., Barrientos, L. F., et al. 2020, , 491, 4442

Lopez, S., Tejos, N., Ledoux, C., et al. 2018, , 554, 493

Martin, C. L. 2005, in *Astronomical Society of the Pacific Conference Series*, Vol. 331, *Extra-Planar Gas*, ed. R. Braun, 305

Martin, C. L., Ho, S. H., Kacprzak, G. G., & Churchill, C. W. 2019, , 878, 84

Martin, C. L., Shapley, A. E., Coil, A. L., et al. 2012, , 760, 127

McDonald, P. & Miralda-Escudé, J. 1999, , 518, 24

McGaugh, S. S., Schombert, J. M., de Blok, W. J. G., & Zagursky, M. J. 2010, , 708, L14

Mitchell, A. C. G. 1971, *Resonance radiation and excited atoms* (Cambridge University Press)

Mo, H. J., Mao, S., & White, S. D. M. 1998, , 295, 319

Morton, D. C. 2003, , 149, 205

Muratov, A. L., Kereš, D., Faucher-Giguère, C.-A., et al. 2017, , 468, 4170

Nicastro, F., Kaastra, J., Krongold, Y., et al. 2018, , 558, 406

Nicastro, F., Zezas, A., Drake, J., et al. 2002, , 573, 157

Nielsen, N. M., Churchill, C. W., Kacprzak, G. G., & Murphy, M. T. 2013, , 776, 114

Nielsen, N. M., Churchill, C. W., Kacprzak, G. G., Murphy, M. T., & Evans, J. L. 2015, , 812, 83

Oppenheimer, B. D., Crain, R. A., Schaye, J., et al. 2016, , 460, 2157

Oppenheimer, B. D., Davé, R., Kereš, D., et al. 2010, , 406, 2325

Peebles, M. S., Werk, J. K., Tumlinson, J., et al. 2014, , 786, 54

Petitjean, P. & Bergeron, J. 1990, , 231, 309

Prochaska, J. X., Werk, J. K., Worseck, G., et al. 2017, , 837, 169

Prochaska, J. X. & Wolfe, A. M. 1997, , 487, 73

Putman, M. E., Peek, J. E. G., & Jounge, M. R. 2012, , 50, 491

Rees, M. J. & Ostriker, J. P. 1977, , 179, 541

Roberts, G. O. & Rosenthal, J. S. 2001, *Statistical Science*, 16, 351

Rubin, K. H. R., Diamond-Stanic, A. M., Coil, A. L., Crighton, N. H. M., & Stewart, K. R. 2018, , 868, 142

Rubin, K. H. R., Prochaska, J. X., Koo, D. C., & Phillips, A. C. 2012, , 747, L26

Rubin, K. H. R., Prochaska, J. X., Koo, D. C., et al. 2014, , 794, 156

Rubin, K. H. R., Prochaska, J. X., Ménard, B., et al. 2011, , 728, 55

Rubin, K. H. R., Weiner, B. J., Koo, D. C., et al. 2010, , 719, 1503

Rudie, G. C., Steidel, C. C., Trainor, R. F., et al. 2012, , 750, 67

Sargent, W. L. W., Steidel, C. C., & Boksenberg, A. 1988, , 334, 22

Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, , 446, 521

Sofue, Y. & Rubin, V. 2001, , 39, 137

Somerville, R. S. & Davé, R. 2015, , 53, 51

Steidel, C. C., Dickinson, M., Meyer, D. M., Adelberger, K. L., & Sembach, K. R. 1997, , 480, 568

Steidel, C. C., Dickinson, M., & Persson, S. E. 1994, , 437, L75

Steidel, C. C., Erb, D. K., Shapley, A. E., et al. 2010, , 717, 289

Steidel, C. C., Kollmeier, J. A., Shapley, A. E., et al. 2002, , 570, 526

Stern, J., Hennawi, J. F., Prochaska, J. X., & Werk, J. K. 2016, , 830, 87

Stewart, K. R. 2017, in *Astrophysics and Space Science Library*, Vol. 430, *Gas Accretion onto Galaxies*, ed. A. Fox & R. Davé, 249

Tejos, N., López, S., Ledoux, C., et al. 2021, , 507, 663

Tumlinson, J., Peebles, M. S., & Werk, J. K. 2017, , 55, 389

Tumlinson, J., Thom, C., Werk, J. K., et al. 2013, , 777, 59

Turner, M. L., Schaye, J., Steidel, C. C., Rudie, G. C., & Strom, A. L. 2014, , 445, 794

Veilleux, S., Cecil, G., & Bland-Hawthorn, J. 2005, , 43, 769

Voit, G. M., Donahue, M., Bryan, G. L., & McDonald, M. 2015, , 519, 203

Werk, J. K., Prochaska, J. X., Thom, C., et al. 2013, , 204, 17

Werk, J. K., Prochaska, J. X., Tumlinson, J., et al. 2014, , 792, 8

White, S. D. M. & Rees, M. J. 1978, , 183, 341

Williams, R. J., Mathur, S., Nicastro, F., et al. 2005, , 631, 856

Zabl, J., Bouché, N. F., Schroetter, I., et al. 2019, , 485, 1961

Zhang, H., Zaritsky, D., Werk, J., & Behroozi, P. 2018, , 866, L4

Zheng, Y., Putman, M. E., Peek, J. E. G., & Jounge, M. R. 2015, , 807, 103

Zwaan, M. A., Meyer, M. J., & Staveley-Smith, L. 2010, , 403, 1969