Clustering of Intergalactic Metals

Bachelor Research Project

Charlotte de Valk & Annemarie Hagenaars
Supervisor: Dr. Joop Schaye
Sterrewacht Leiden
Leiden University
June 22, 2006

Abstract

In this Bachelor Research Project we tested if the Pixel Optical Depth method is able to detect clustering of metal density fluctuations. It tells us about sizes of density fluctuations.

We used Quasar Absorption Line spectra to observe clustering of metals in the Intergalactic Medium. The POD method calculates the optical depth of every pixel in the spectrum. We correlated CIV fluctuations in optical depth as a function of distance measured in velocity differences. Velocity differences are easy to convert to comoving distances.

By making programs for a correlation function we can study the profile of clustering. We compared this profile to that of a correlation function done by another method. Scannapieco et al. (2006) identified CIV lines by eye and correlated them. They only did this for strong metal lines using a probability function. But we looked at the values of the optical depth of the lines.

We found clustering using the POD method and we found another profile then Scannapieco et al. did. We detected clustering up to a scale of 200 km/s ($\approx$ 5 Å); they observed clustering up to 1000 km/s. Using our method results in a clear detection of the doublet. Scannapieco et al. did not totally remove the doublet. Their detection of correlation up to 1000 km/s may be caused by contamination left over after doublet removal.

The POD method is a fast, automatic and statistical way to get information about clustering of IGM metals. It is accurate to study metals in the lowest density regions of the IGM and it is very promising for future research on star and galaxy formation.

Contents

Chapter 1

Introduction

In Astronomy metals are defined as all elements of the periodic system heavier than Hydrogen and Helium. Hydrogen and Helium are the basic elements of the universe. They are formed shortly after the Big Bang. The first metals came to exist when stars started their fusion of Hydrogen into Helium into heavier elements. After a star dies, during supernovae explosions, the metals are ejected into interstellar medium. Then it is, according to models of galaxies, a galactic wind which ejects the metals into the intergalactic medium.

Metals are found even in low density regions of the IGM. The material can tell us about all formative stages in the universe; it contains important information about star and galaxy formation. The different sorts of metals produced, can give an indication of the mass of stars where they came from. It is not clear yet whether massive galaxies or low mass galaxies dominate the enrichment of the IGM. It is important to study the strength of galactic winds which eject the metals into the IGM. It can be a key to a correct model of galaxy formation. Because models of galaxies which do not have feedback processes, like galactic winds causing energy and material loss, form too many stars. Those models do not resemble observed galaxies. The energy and the momentum, that is ejected into interstellar space by feedback processes from stars, are driven into intergalactic space by strong outflows, which also carry the metals produced by the stars (Schaye et al. 2005). In Fig.1.1 you see the galactic winds of galaxy M82.

Metals in the IGM can be observed using Quasar Absorption Line spectra (detection of flux as function of wavelength, see Fig.1.2). Quasars are bright point sources, that emit a continuum flux. Light from the quasar is absorbed by atoms. Observing this absorption can give information about the structure of the IGM.

To find out whether massive galaxies or low mass galaxies dominate the enrichment of the IGM, we should measure the clustering of metals. This can be done by a correlation function. We correlate every pixel as a function of the distance between two pixels, and we look at the value of the pixel optical depth to identify CIV peaks. CIV are relative easy to identify in the best part of the spectrum where there is little contamination by other metal lines.

We measure clustering of CIV lines, but this metal may not represent other metals, like the C atom, the C iii ion or the SiIV ion. It depends on the ionization balance in the metal density fluctuation whether clustering of CIV is a good

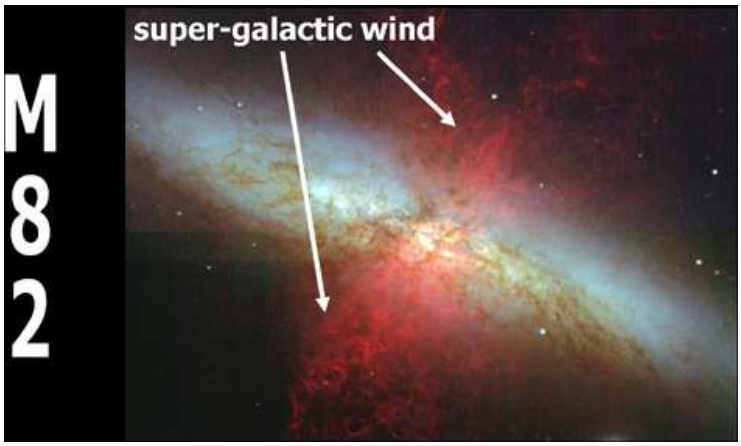


Figure 1.1: *Galaxy M82 ejecting metals by galactic winds. The winds are in red here, because they are radiating at $H\alpha$. The heating is caused by shock waves; hot material is ejected and collides with cold material, that is shock heated. Shock waves appear. The flows radiate because of hydrogen recombination.*

representation of clustering of metals in the IGM. That makes interpreting what we are going to measure more difficult. Ionization balance of heavy elements is sensitive to the physical conditions. With simulations you can constrain gas density, temp, etc. to study what the influence of different ionization balances is on clustering. You need to know the ionization balance in order to convert from C-ions to C-atoms. For further explanation: see Schaye et al. (2005).

If we measure strong clustering, it can be that the metals are recently ejected by massive galaxies. Measuring clustering means that the correlation function is high and stays high for larger distances. We will take velocity (if it is Hubble velocity) as a measure for the size of a system. It is known that big galaxies cluster strongly due to the superposition of density fluctuation waves. On large scale structure the universe can be described by density fluctuation waves, that interfere, resulting in a wave pattern containing clusters of high waves. Strong clustering in IGM metals implies that we are dealing with massive galaxies or with progenitors of massive galaxies.

Instead of Hubble velocity the observed velocities may reflect peculiar velocities within systems, or the velocity of a galactic wind. Some of the quasar spectra we use contain absorption systems, like galaxies or clusters of galaxies. At a certain velocity difference the correlation function peaks due to the system. What you are measuring then is peculiar velocity.

The interpretations of clustering in velocity space are not yet clear. In the 1960s, when quasars were first discovered, also the possibility of detecting the structure of the IGM with so called Quasar Absorption Lines was discovered (see Fig. 1.2). This opened up a lot of research possibilities. The spectra of the QALs show us the absorption lines of the atoms in the line of sight from us to the quasar. In the past these QAL systems have been identified by eye. Voigt profiles help identifying what the physical properties of the density fluctuations are. This identifying by eye has disadvantages. It is subjective. You can only see the bright absorption lines, but you would like to know more about the

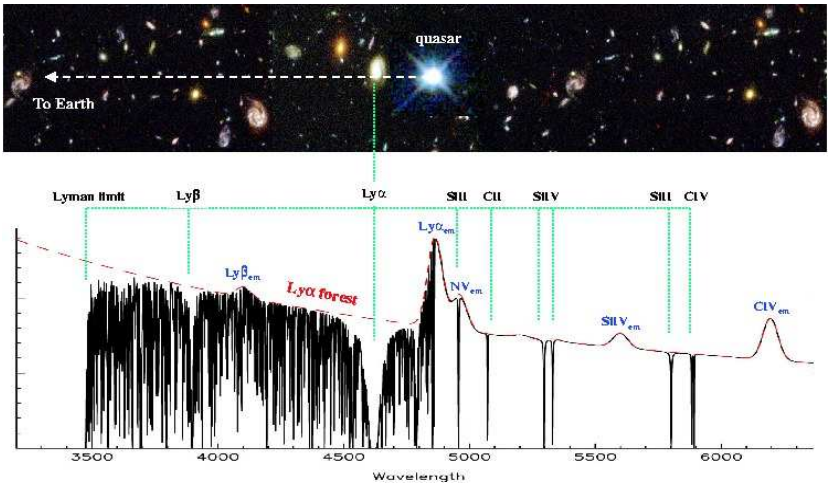


Figure 1.2: *Quasar Absorption line systems. The quasar as a bright distant point source is used to detect what is in the universe between quasar and observer. We are interested in intergalactic clouds. Sometimes you will find in* QAL*'s an absorption system which is gravitationally bound like galaxies or clusters of galaxies.*

weaker ones in the low density IGM. The use of Voigt profiles does not make it easy to correct for contamination. Also it is too insensitive and it takes a lot of time to analyse large data sets. Above all, you can not study absorption that is weak compared to the noise and the contamination.

Earlier, people studied clustering of intergalactic metals. Scannapieco et al. (2006) used a correlation function to detect clustering and found correlation up to velocities of 1000 km/s. They used VLT spectra of 19 quasars. Although their correlation function is calculated in a different way (this will be explained in chapter 5), the shape of the function can be compared to what we have found using a more objective method. They added different quasars to get an average correlation function from which general conclusions can be drawn.

They made a distinction in the quasar's redshift by making a low range and a high range correlation function. Their conclusion is 'that the distribution of intergalactic metals does not appear uniform, nor simply dependent on the local density, but rather it bears the signature of the population from which it came'. Because these results are based upon eye research, they only observed strong metal lines. To verify Scannapieco's findings we need a different method to get the same result.

A new method which is statistical and objective is the Pixel Optical Depth Method. It is first used by Cowie and Songaila (1998) and later further developed by Aguirre and Schaye (2002). Per pixel the value of the optical depth is used to identify metal absorption lines. The POD method is fast and you are using more of the spectrum's information. The method can be applied to heavily contamination regions, and it appears to be more sensitive to metals in gas with low density, even in regions with contamination (Aguirre et al. 2002). If you put everything what you know into a simulation, you can compare carefully your observed data to the model and draw conclusions. For the POD method

there has not been defined a correlation function before. In this project we are going to check whether you still detect clustering and on what scale, using the Pixel Optical Depth method.

Chapter 2

Some Basics

We are doing research with the Pixel Optical Depth Method. This method is based on the capability of working with every pixel of the spectrum. We are looking at the value of the optical depth, because optical depth is proportional to the particle density of the cloud ($\tau \propto n$). Although we do not study cloud densities now, for future work it is proficient to know what the physical properties of intergalactic clouds are. We are using the program GET-OD-Z, Get Optical Depth Redshift written by Anthony Aguirre et al. (2002), to remove bad pixels. For these removals we need the optical depth per pixel.

In this chapter is described how we use this method and how we have to modify the spectrum before we can calculate the correlation function. But first we need to know some facts about the data.

2.1 Observations

Out of 20 available spectra, we have chosen to work with spectra of 6 quasars, see Table 2.1. Here is given respectively, the redshift of the quasar, the redshift range and the minimum λ. λ_{min} is the chosen reference point in the spectrum. It is the minimum λ where we have no contamination by Hydrogen and other metals.

All those spectra were taken with the High Resolution Echelle Spectrometer (HIRES) on the Keck telescope. The signal to noise ratio S/N $\approx$ 100. The

TABLE 2.1: OBSERVED QUASARS.
Given is the redshift of the specific quasar, the range for the best detection of CIV*, and the minimum λ where you should start looking for metals.*

QSO	$z_{\rm qso}$	$z_{\rm min}$	$z_{\rm max}$	$\lambda_{\rm min}$ (Å)
Q1442+101	2.67	2.375	2.947	3644.36
Q1107+485	3.00	2.375	2.947	3644.36
Q1425+604	3.20	2.544	3.144	3736.20
Q1422+230	3.62	2.898	3.552	3645.24
Q1055+461	4.12	3.320	4.033	4586.36
Q2237-061	4.558	3.690	4.451	4933.68

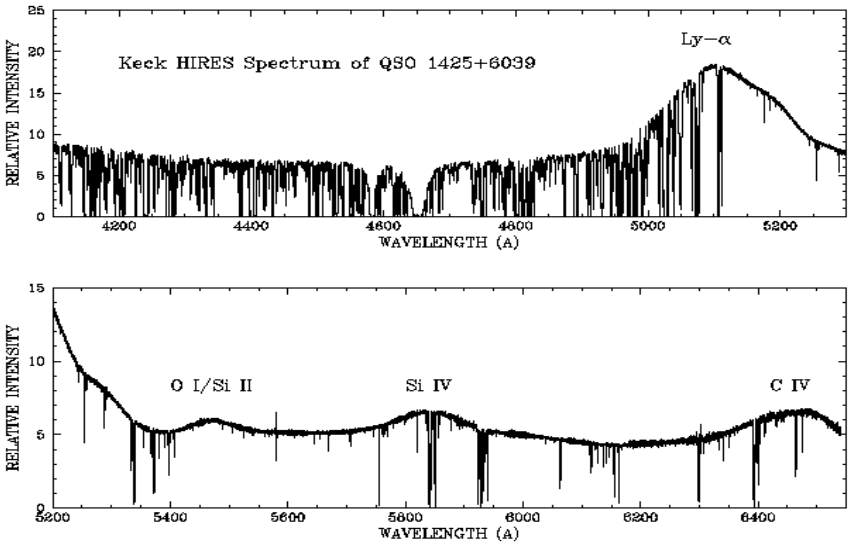


Figure 2.1: *Absorption spectrum of* QSO *1425. The spectrum shows what the best ranges are for the detection of C* IV*, Si* IV *and O* I*. Also it is clear what range is contaminaded by Lyα.*

resolution of the spectra is 6.6 km/s FWHM, what means R $\approx$ 45,500. Resolution is defined as:

$$R = \frac{\lambda}{\Delta\lambda} = \frac{c}{\Delta v} \tag{2.1}$$

The original plan was that we would calculate the correlation function of 16 qso's. We would also use 10 spectra made by UVES of the VLT. But some of these spectra gave weird CF-values for very small velocity differences. We think this is because the VLT spectra were fitted to the continuum in a different way then the Keck spectra. In the end we used the 6 available Keck spectra.

2.2 How to use spectra of the different quasars

In a spectrum of a quasar, flux is measured as a function of wavelength in Å. In the spectra you see lines of the C IV doublet (1548.2041 Å, 1550.7812 Å), Si IV doublet (1393.76018 Å, 1402.77291 Å), Hydrogen multiplet and some other metals (see Fig. 2.1). To find out if there is clustering in a spectrum we cannot use all these lines. We just need one sort of metal line absorption coming from high density areas (gas clouds). Now we have absorption lines coming from more than one sort of metal in one cloud.

We choose to look at the CIV lines. Those lines are not saturated like *Lyα* absorption systems and there is a large field range in the spectrum where there is little contamination by other strong metal lines.

When we observe a CIV line at a certain λ in the absorption spectrum, we can calculate at what redshift this absorption system has to be. Knowing λ_{obs},

λ_{rest}, we can calculate z (a measure for distance):

$$1 + z = \frac{\lambda_{obs}}{\lambda_{rest}} \tag{2.2}$$

We know that CIV is a doublet with a primary (1548.2041 Å) and a secondary (1550.7812 Å) and that the optical depth value in the primary is twice the optical depth value of the secondary. On this fact the CIV lines are detected. There is a pixel with a certain optical depth and when the pixel at a distance of $\Delta\lambda \backsim 2.3$ Å(the doublet spacing) with the half of that optical depth value you know that they belong to the CIV primary and secondary.

We are working with very good data. But still there is some noise which makes it difficult to recognize a CIV doublet. So we will use a number of sigma (nsig) that allows the optical depth of the pixel in the secondary to vary. So the optical depth of the pixel in the secondary must have the value:

$$0.5\tau_p - nsig \cdot \sigma \leqq \tau_s \leqq 0.5\tau_p + nsig \cdot \sigma \tag{2.3}$$

where τ_p is the optical depth of the pixel in the primary and nsig the number of sigma's the value of the second pixel may deviate. Nsig is a kind of measure on how strict you are in identifying CIV lines. If the second pixel satisfies this condition we will see the first pixel as a part of the primary and the second pixel as one of the secondary. If the second pixel does not satisfy this condition, both the pixels will be given a new value, which is the noise of that pixel. This because they are not part of CIV lines.

If you want to remove the secondary of the doublet, the τ_s will be set to a new value :

$$\tau_{s,new} = \tau_{s,old} - 0.5 \cdot \tau_p. \tag{2.4}$$

It will not be set to the value of the noise, because it is possible that at the pixel of the secondary, there is also another metal line or a primary of CIV and you do not want to remove that one. Then you will lose information. Formula (2.4) is a simple version of what Aguirre et al. (2002) did. They used iterations to remove the doublet.

2.3 From wavelength to comoving distances

In the spectrum, between two lines there is a distance $\Delta\lambda$. And now that we have one metal detection line, the distance between two lines is also a measure for the distance between two corresponding 'clouds' or density fluctuations in the IGM. But the difference in wavelength is not a good size for the distance between clouds, because it is not corrected for the redshift.

What we want is to investigate what the comoving sizes of the intergalactic metal density fluctuations are, and on what scale metals cluster in comoving space coordinates. These comoving coordinates are the easiest way to compare the sizes of clouds at different redshifts, independent of time. In comoving coordinates we take as measure the sizes the clouds would have at z=0. During time metal density clouds are expanding. We want to correct this for Hubble expansion.

In our spectrum it is basically this problem: if you take the same distance in λ at the beginning of the spectrum (low redshift) and the same at the end

of the spectrum (high redshift) it does not mean it is the same real distance in space. The real distance between the corresponding gas clouds with lines at the beginning of the spectrum is larger than the distance between the two clouds corresponding to the lines at the end of the spectrum. This is all due to the Hubble expansion. So we have to correct this by defining a dimension that is a correct size for the distance. Comoving distances can be calculated like this:

$$\Delta r_{comoving} = \frac{\Delta v}{H(z)} \cdot (1+z) = (1+z) \cdot \Delta l_{proper} \tag{2.5}$$

$$H(z) = H_0 \sqrt{\Omega_m \cdot (1+z)^3 + \Omega_\Lambda} \tag{2.6}$$

where z stands for redshift, H(z) is the Hubble expansion at a certain redshift, Δl_{proper} is the absolute distance, $\Omega_m = 0.3$ (matter dominated density parameter), $\Omega_\Lambda = 0.7$ (curvature dominated density parameter) and Δv can be computed by:

$$\frac{\Delta v}{c} = \frac{\Delta z}{1+z} = \frac{\Delta \lambda}{\lambda} \tag{2.7}$$

where c is a constant; the speed of light. From these formula's you can conclude that velocity differences are a good measure for distance, because what you expect to measure is Hubble velocity. This is easy to convert to comoving distances. Above all, velocity is physical interpretable. This will be explained later in this section.

What we should do is convert the whole spectrum from wavelength to velocity, but the formula above is only valid for small wavelength differences. We have to use an integral, and the velocity as a function of the wavelength will become:

$$v = \int_{\lambda_{min}}^{\lambda} c \cdot \frac{\Delta \lambda}{\lambda} = c \cdot ln(\frac{\lambda}{\lambda_{min}}) \tag{2.8}$$

where λ_{min} is the minimum wavelength. This will be a reference point, where you choose velocity to be zero.

Clustering can tell us about sizes of clouds, if we assume we measure Hubble velocity. But we can measure three kinds of velocity differences.

Hubble velocity When all you measure is this velocity you can calculate at what distances metals in the IGM cluster.

Peculiar velocity This kind of velocity we measure when we are going through a system, where the gravitational force is stronger than the Hubble expansion. This system will not expand with Hubble velocity. It can be a galaxy or a cluster of galaxies, where the material rotates around the centre of mass (rotational velocity). Another example of peculiar velocity is when you measure redshift space distortions; high massive parts of the universe expand slower than low massive parts.

Velocity of a galactic wind that ejects metals into the IGM It is thought that, when stars exhaust their fuel for nuclear fusion and they explode, the material is carried out of the galaxy by galactic winds. The strength of the galactic winds is not exactly known, but indications are around 1000 km/s.

It does not mean that we are not content with detection of other velocity differences than Hubble velocity. If we can identify different kinds of velocity and what it causes, we find out more about absorption systems, including intergalactic clouds, galaxies, clusters of galaxies and galactic winds.

Chapter 3

The correlation function

We want to know if it is possible to detect clustering with the Pixel Optical Depth method. If so, maybe we can get some new information on the distribution of metals in the Intergalactic medium. We cannot use the same correlation function that is used by Scannapieco et al. in the research done by eye and with the help of Voigt profiles. He defined a correlation function which is specified on the method he used.

We do not want to use this method, because it has many disadvantages. It is subjective and you can only see the bright absorption lines. It is too insensitive and it takes a lot of time to analyse large data sets. And you cannot correct for contamination.

So we have to define another correlation function that can be used in the combination of the fast, automatic and objective Pixel Optical Depth method. These characteristics of this method are very important when you want to work with simulations. Because in the end you also want correlation functions of simulated data. So you can interpret and understand the real data.

3.1 Definition

A good measure for clustering is the correlation function of the spectrum. The definition of the auto-correlation of a function is as follow:

$$autocorr(g) = corr(g, g) = \int_{-\infty}^{\infty} g(t + d)g(t)dt \tag{3.1}$$

In words: the correlation function is a measure of similarity of two signals. We can correlate the flux, the optical depth, the fluctuation of the flux or the fluctuation of the optical depth. We have chosen to calculate the correlation of the fluctuations in the optical depth (δ_τ).

$$\tau = -ln(\frac{F}{F_c}) \tag{3.2}$$

$$\delta_\tau = \frac{\tau - \bar{\tau}}{\bar{\tau}} \tag{3.3}$$

δ_τ can be positive and negative. It is a relative value and in this way you can easily compare different QSO spectra, because every QSO is at a different

redshift. At a high redshift the IGM is denser. So the spectrum would have a higher optical depth for every pixel than in a spectrum for a quasar at a low redshift. Working with δ_τ will correct for that.

Another advantage is when the spectrum is fitted wrong to the continuum. This will have an effect on the optical depth of every pixel, but not on the δ_τ. We used a program for continuum fitting, CONFIT (Schaye et al. 2003).

The definition of the correlation function we are using is:

$$\xi(\Delta v) = \frac{\int \delta_\tau(v)\delta_\tau(v+\Delta v)dv}{\int dv} = \langle \delta_\tau(v)\delta_\tau(v+\Delta v)\rangle \tag{3.4}$$

But what does this mean in practice. For example we want the correlation for a distance of 10 km/s. For every pixel we take δ_τ of one pixel and δ_τ of a pixel that is 10 km/s separated. We calculate the product of those two pixels. After that, we calculate the average of all products. Now we have the correlation for a Δv of 10 km/s. We did this for every distance from 0 km/s up to 3.000 km/s.

3.2 Results

We want the correlation function for all 6 quasar spectra together. So first we calculated the CF for every single quasar and their error with bootstrap samples. Bootstrap resampling is explained in chapter 5. We divided the results in velocity bins. After that, we calculated the CF value for 0 km/s for every QSO. The CF(0) is equal to the variance divided by the squared average of the optical depth, because:

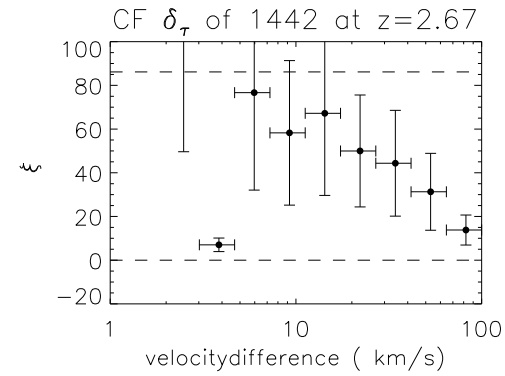
$$\xi(0) = \langle \delta_\tau(v)\delta_\tau(v+0)\rangle = \langle \delta_\tau(v)\delta_\tau(v)\rangle = \langle \delta_\tau(v)^2\rangle = \left\langle (\frac{\tau-\bar{\tau}}{\bar{\tau}})^2 \right\rangle = \frac{var(\tau)}{\bar{\tau}^2} \tag{3.5}$$

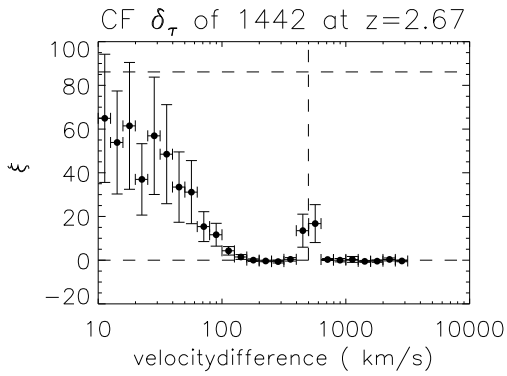
In Fig 3.1 we have the CF of the 6 QSO's with the CF(0) value plotted horizontal. We made per QSO two plots: one for small Δv (1-100 km/s) and one for large Δv(10-3000 km/s). As you can see in these plots, for some QSO's the CF starts with a higher value than others. But for all is valid that they start at some relative high value and when Δv approaches infinity, the CF is zero for all QSO's.

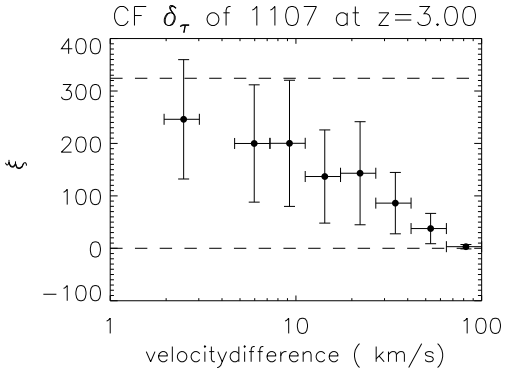
Every QSO has a higher correlation at 500 km/s. This is because of the doublet. We did not remove the secondary of the doublet and there is always a correlation between the primary and secondary. The spacing between the two lines is $\Delta\lambda \backsim 2.3$ Åand using equation (2.7) it is a velocity difference of 497 km/s. So there is a higher correlation at this Δv. If we would remove the secondary of the doublet, this higher value in the CF is probably gone.

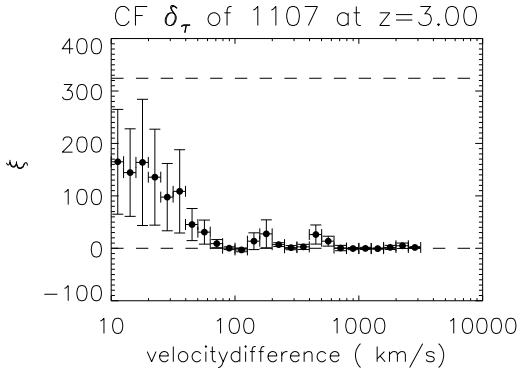
In the CF of the quasars 1107, 1425, 1422, 1055 there are also higher correlations at some other Δv. These are detections of peculiar velocities of some systems, like galaxies.

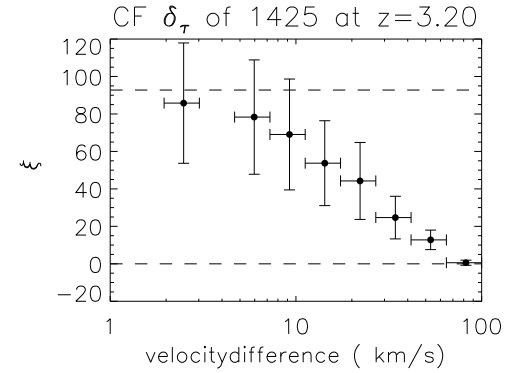
If we want to make one CF of those six QSO's we cannot just take the average of those six. Because for some quasars the CF will start at low values and some at high values. The same is valid for the CF values at larger Δv. The average CF will be flatter than the CF for every single quasar and will gives us a wrong picture of the shape of the total of the six QSO's. To keep the shape we scaled the CF for every QSO up to the CF of QSO 1425, so that their value at 0 km/s (CF(0)) would be the same for every QSO.

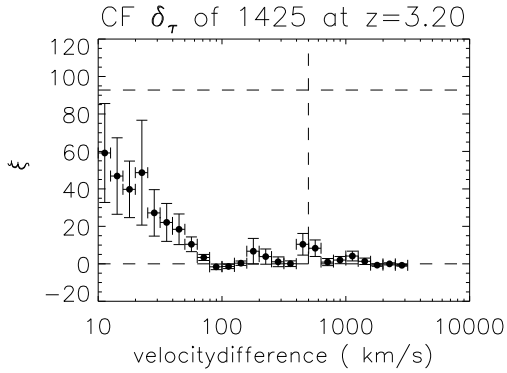

CF δ_τ of 1442 at z=2.67
100
80
60
40
20
0
−20
ξ
1
10
100
velocitydifference (km/s)

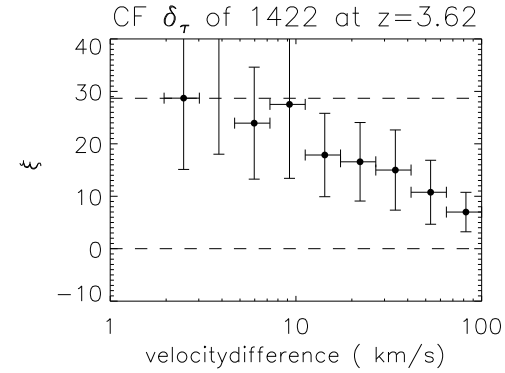

CF δ_τ of 1442 at z=2.67
100
80
60
40
20
0
−20
ξ
10
100
1000
10000
velocitydifference (km/s)

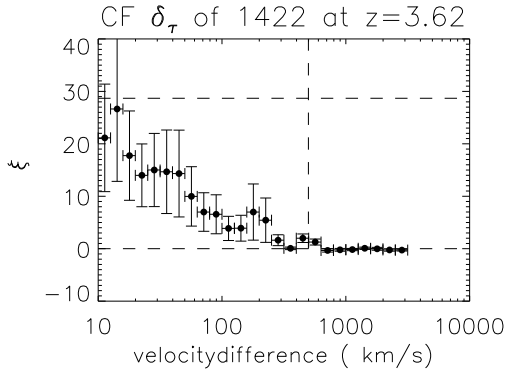

CF δ_τ of 1107 at z=3.00
400
300
200
100
0
−100
ξ
1
10
100
velocitydifference (km/s)


CF δ_τ of 1107 at z=3.00
400
300
200
100
0
−100
ξ
10
100
1000
10000
velocitydifference (km/s)


CF δ_τ of 1425 at z=3.20
120
100
80
60
40
20
0
−20
ξ
1
10
100
velocitydifference (km/s)


CF δ_τ of 1425 at z=3.20
120
100
80
60
40
20
0
−20
ξ
10
100
1000
10000
velocitydifference (km/s)


CF δ_τ of 1422 at z=3.62
40
30
20
10
0
−10
ξ
1
10
100
velocitydifference (km/s)


CF δ_τ of 1422 at z=3.62
40
30
20
10
0
−10
ξ
10
100
1000
10000
velocitydifference (km/s)

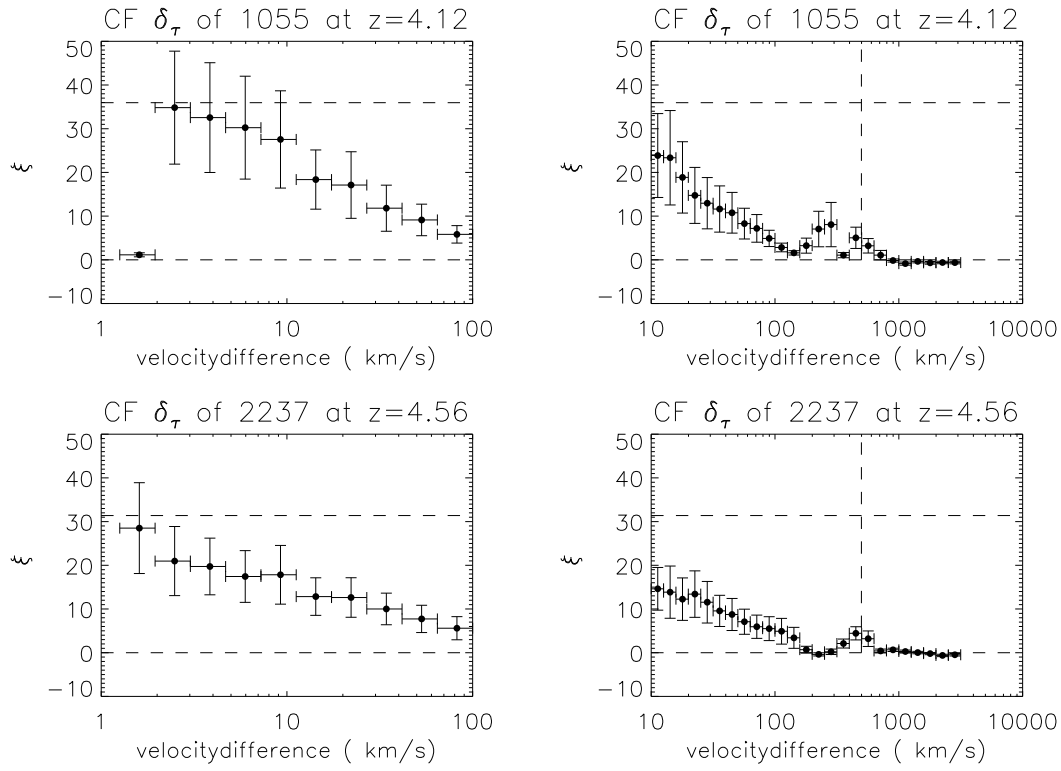


Figure 3.1: *Correlation function δ_τ of 6 single* QSO*'s. Every* CF *starts with a high value and drops off to zero. The* CF*(0) value and the zero line are plotted horizontal. On large scales there is at some Δv a higher correlation . Always at 500 km/s (the doubletspacing, marked with the vertical line). At other Δv it is a system and the Δv is the peculiar velocity of that system.*

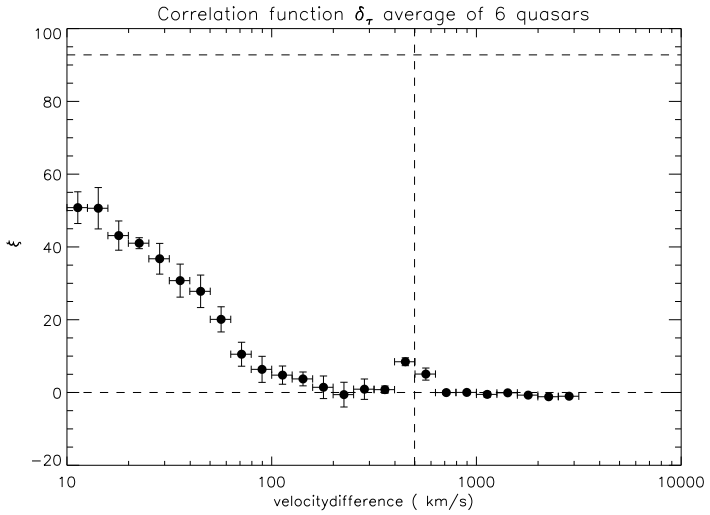


Figure 3.2: *Correlation function δ_τ of the weighted average of the 6* QSO*'s. All* CF*'s were scaled to the* CF *of* QSO *1425. The horizontal line is the* CF *value at 0 km/s, the startvalue. The* CF *starts with a high value and convergates to zero. At 500 km/s (vertical line) there is higher correlation due to the doublet.*

The fact that the correlation function for some QSO's starts at high values and some at low values, is because we are working with different quasars at different redshifts. In Fig. 3.1 the QSO's are ordened by different redshifts. Scannapieco et al. found out that the CF depends on redshift. And we can also see in Fig. 3.1 that the CF for the QSO's at a low redshift starts with a higher correlation than high red shifted. This is because at a higher redshift the fluctuation in the densities are less than at a lower redshift.

So the metal density fluctuations are stronger for the quasars at a lower redshift. To correct for that, we have to scale all the CF to one level. Now we have to keep in mind that the CF for the total of 6 QSO's is not important for its absolute values, but mostly for its shape. For the absolute values, you need the CF of a single QSO and that tells you more about the spectrum you are working with, for example detection of intervening systems. To calculate the CF of all 6 quasars we used a weighted average, defined this way:

$$\bar{\xi}(\Delta v) = \frac{\sum \xi(\Delta v)\frac{1}{\sigma^2}}{\sum \frac{1}{\sigma^2}} \tag{3.6}$$

For every velocity bin the strength of participation of a quasar is $\frac{1}{\sigma^2}$. So the quasar with a large error will have a small contribution to the average. The error for the average value is calculated by making a bootstrap resampling per velocity bin of the 6 QSO's and the standard deviation of the distribution is our error. You can read in Chapter 4 about how we used the bootstrap resampling for this error calculation.

The correlation function of the 6 quasars is shown in Fig.3.2. We plotted a linear version of the correlation of the CIV lines. The CF starts with a high value and falls off to zero. The correlation is gone at a velocity difference of 200

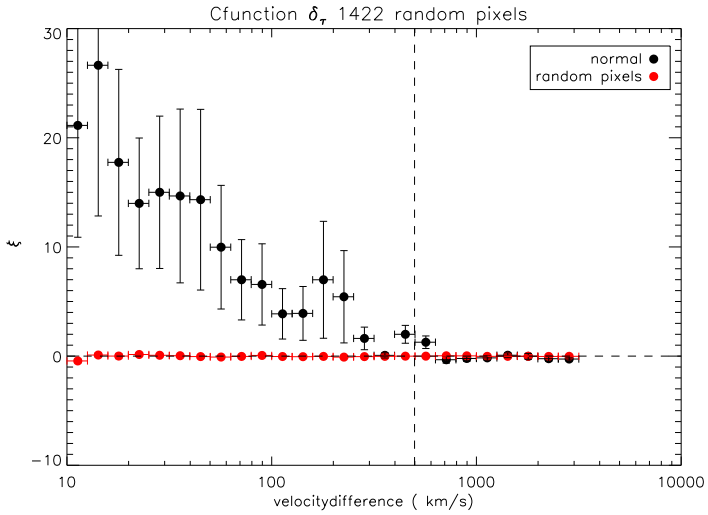


Figure 3.3: *Correlation function δ_τ random pixels. There is no correlation at all in the spectrum of the random pixels.*

km/s. But there is again a higher value at 500 km/s. This is due to the doublet of CIV.

It seems that we can detect clustering. The CF is higher for small Δv and converges to zero for large Δv. But to be sure we need a spectrum where there is no clustering at all. We have not been able to work with simulations, so to get a unclustered spectrum, we choose the spectrum of quasar 1422 and we gave all the pixels a random new place in the spectrum. Again we calculated the correlation and as you can see in Fig. 3.3 the CF is almost flat with a value around zero. There is no correlation at all.

The shape of the CF we had before is totaly gone. So there is no clustering at all, like we expected. We also checked by taking random chunks. We divided the spectrum in chunks of 1 Å. And we randomly picked out the chunks and made a new spectrum with the same size as the original. The chunksize of 1 Åis in velocity $\sim 40 km/s$.

In Fig. 3.4 we can see that up to a velocity difference of 40 km/s it almost has the shape of the normal CF of quasar 1422 and for higher velocities the correlation is gone. Only at some larger velocities the correlation comes back at the doublet spacing, which is to explain. We made a new spectrum and the real CIV lines will not be identified as CIV, because their spacing is gone. And now these lines are contamination and will be removed. But every pair of lines in the new spectrum with the same spacing as the doublet of CIV will be identified as CIV. Although these are not real CIV lines. The lines that now are been detected as CIV are not the strong real CIV lines, but noise or really weak lines that are misidentified. But these misidentified lines do have an influence on the CF. And the CF will still show a higher correlation at the doublet spacing. The chance that these new identified lines are the real CIV lines is small, because the chunk size is less than the doublet spacing and the chunks are used randomly in the new spectrum.

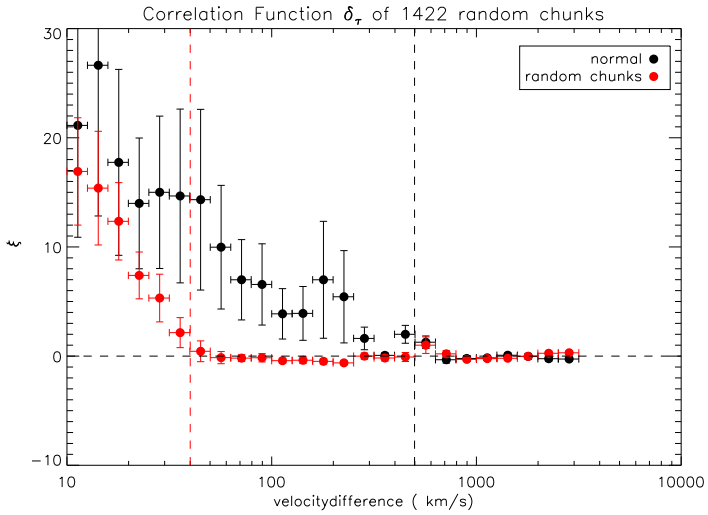


Figure 3.4: *Correlation function δ_τ chunk size 1 Å(40 km/s). The CF of the spectrum with random chunks. Before the 40 km/s (red vertical line) the CF is almost the same as the original one. But after the chunksize the Correlation is zero. So after 40 km/s we don't detect any clustering .*

But we can say for sure that the correlation for velocity differences less than the chunk size did not change much. That also is expected. Because inside the chunks the spectrum did not change. And the small change is due to the fact that pixels at the edge of a chunk will now make a pair with a different pixel than before. And their correlation with the rest of the spectrum will change for small Δv. For velocity differences greater than the chunk size the CF stops abrupt with the decline and will be close to zero for almost every velocity bin.

These tests show, that when there is no clustering at all in the spectrum, we really do not detect clustering anymore. This shows that our program works and that we really detect clustering, when there is clustering.

We also did some other tests. The first one is by removing the doublet of the spectrum. Then the correlation at the velocity difference of 500 km/s should also be gone. The result of this test is to see in Fig. 3.5. At the doublet spacing the correlation function has lower values, but there is still some correlation left. This is probably caused by a not totally correct removal of the doublet. We mean by this that the secondary of the doublet will be removed, but for some pixels there is still some optical depth left over. The reason why after doublet removal we still see higher correlation at the doublet spacing is, because you are working with noisy spectra and that makes the detection of lines and doublet removal not perfect. With higher noise the GET-OD-Z program will make more misinterprets in detecting CIV lines.

It seems that we are calculating the clustering of the CIV lines, but the spectrum contains all kinds of metal lines. And as we saw, the doublet removal is not perfect. Probably the same thing is valid for the removal of other metal lines. And the residues can have an influence on the correlation function. So how do we know we are really measuring the clustering of CIV lines and not the

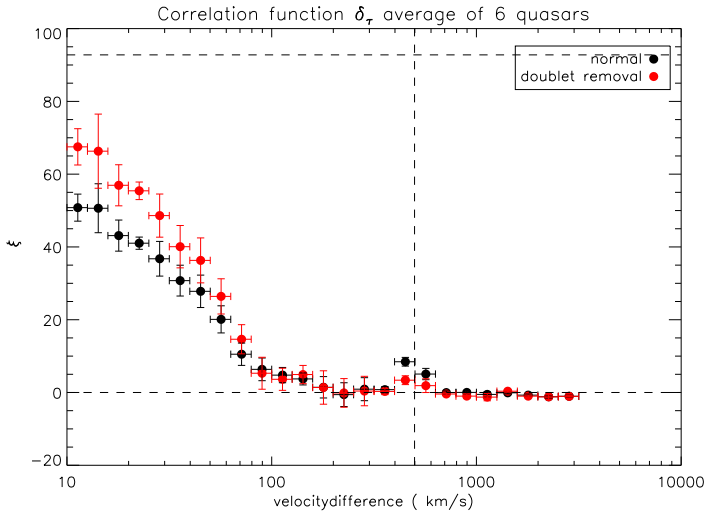


Figure 3.5: *In red the correlation function δ_τ with a removed doublet. All* CF*'s were scaled to the* CF *of* QSO *1425. The* CF *is almost the same as the original one, but has a higher correlation for Δv below the 90 km/s and a lower correlation at the doublet spacing. The* CF *is not zero there, but this is due to the not perfectly removed secondary of the doublet.*

clustering of the residues of some other metal?

We made a test by using the wrong wavelengths for the CIV lines. We changed the doublet spacing, so the program did not recognize the real CIV lines as the metal we were looking for and removed the lines. We made sure that we choose a new doublet spacing that is not equal to the spacing of another metal. So the correlation must be less. It will not be totally gone, due to the fact that there will always be parts of lines left in the spectrum of other metal lines and CIV lines. Because CIV will only be removed if they are much stronger than the noise. If they are not, they will be seen as noise and left behind in the spectrum.

Fig 3.6 shows the result. We chose a new doublet spacing that is 10 Ålarger than the old one. So now the doublet spacing is $\sim 2300 km/s$ instead of the $500 km/s$. The CF has a lower value for small velocity differences than the original one. But they are the same for Δv larger then 90 km/s. At the point where there is almost no correlation anymore. The fact that the CF is different for small Δv, indicates that we are really detecting CIV lines. If we would measure the correlation of the residues, then the CF would be the same for the right and wrong wavelength, because in both cases you are not measuring the distance between the CIV lines, but between noise and residues.

In equation (2.3) we are working with nsig, the number of sigma's. In all the previous results we used nsig = 1. But how would the CF change if we would work with a nsig = 3?. When we are working with a higher nsig, you are less strict on the value of the pixels in the secondary and allow more lines to be CIV lines.

In Fig. 3.7 you can see that for the correlation it makes a small difference. In

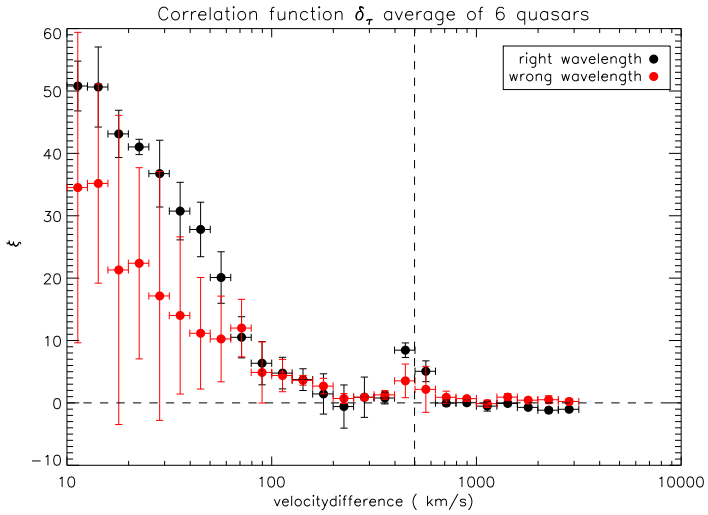


Figure 3.6: *Correlation function δ_τ of* CIV *with the right and the wrong doublet spacing, that is 10 Ålarger than the old one. All* CF*'s were scaled to the* CF *of* QSO *1425. For the wrong doublet spacing the correlation is lower at small Δv than the original one. Because of this we can say that in the original* CF *we are really measuring the correlation of C*IV *and not of the residus from other metal lines.*

general the correlation is a little higher for nsig = 1. This because with nsig=1 you are removing more lines and the average optical depth of your spectrum is less. This makes the δ_τ value of the pixel who are left behind higher and your values of the correlation is a little higher. The correlation ends for both cases at the same Δv.

So there is a small change in the CF when we are using a higher nsig. So you have to be careful in choosing your nsig value. When it is to high you will use lines that are not CIV. And with a very low value, you can lose lines. The nsig value has an effect on the correlation function.

The results show that we can detect clustering of CIV lines with a correlation function. We also can detect the doublet. But we are not be able to remove the doublet perfectly. So we have to keep in mind that a high correlation at Δv of 500 km/s is probably due to this problem and no real clustering. The influence of the nsig is not that big. A higher nsig value will lower the correlation values a little, but has no influence on the Δv, where the correlation stops.

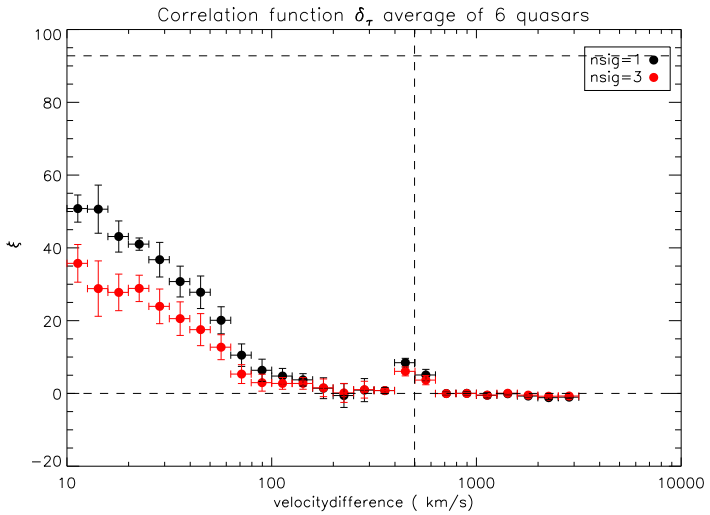


Figure 3.7: *Correlation function δ_τ with different values of nsig. All* CF*'s were scaled to the* CF *of* QSO *1425. There is a small difference in both the* CF*'s. For nsig=3 the correlation is a little lower, but the correlation stops at the same Δv. So a difference in nsig has small influence on the result when you are studying clustering of metal lines.*

Chapter 4

Error calculation using Bootstrap Resampling

Setting up a bootstrap resampling was the most difficult part of our research. We had to intergrate it into our calculation of the correlation function. We used bootstrap resampling in two ways. The 'intern' and 'extern' bootstrap resampling as we call it. It refers to the calculation of errors for the CF for one spectrum or for the spectra of all quasars.

First we will give you a general explanation. And after that, the specific application for our research. For calculation of the error on the correlation function we devide our spectrum in chunks of 5 Å. We need to replace the general X's by chunks of our spectrum.

4.1 Mathematical definition

Let us have a random sample $X_1, X_2, ..., X_n$, where the n X's are the data points. We can calculate the mean of the sample by:

$$\langle X_n \rangle = \frac{X_1 + X_2 + ... + X_n}{n} \tag{4.1}$$

Now we want to bootstrap this realization to calculate, for example, the error in the mean. We generate a bootstrap random sample out of the sample above; we take randomly n times an X and every X may be selected more than once. Then we calculate the mean of the new sample, the bootstrapped average:

$$\langle X_n^* \rangle = \frac{X_1^* + X_2^* + ... + X_n^*}{n} \tag{4.2}$$

The idea is now to take a bootstrap sample like this one a number of times. The more samples you take, the better you approximate the distribution of $\langle X_n \rangle$. If you then calculate the mean of every bootstrap resampling and you take the standard deviation of this distribution (all the bootstrapped averages), you can estimate the error σ on the mean of your original distribution.

4.2 Application

Error calculation for one quasar spectrum To do the bootstrap resampling we divide the spectrum in chunks of 5 Å. The chunks should be greater than the width of the CIV absorption lines. We expect to detect clustering on a scale up to 5 Å($\approx 200km/s$, see formula (2.7), because we can observe by eye in the spectra that absorption line systems can have this size.

Idea is to randomize these chunks a great number of times, where every chunk may be selected more than once (like the dataset of X's). We are not really creating new spectra, but we remember for every resampling how often every pixel of the original spectrum is in the bootstrapped one. Of course pixels stay in chunks together. What you get is a matrix of number of bootstraps and number of pixels in every spectrum. We took n=100 for number of bootstrap realizations.

Calculating the products for the correlation function (see formula (equation:cf)) needs to be done before we randomize the chunks. Otherwize our velocity range is also randomized, and then we cannot find the right pixel anymore at the right velocity difference. We have to remember what pixel with what pixel forms a product and after that, we do bootstrap resampling.

So after we have calculated all the products of the correlation function per velocity bin, we immediately calculate the error per bin by taking the standard deviation from the bootstrap results of each bin. The result per bootstrap is the mean bin value of the correlation function per velocity bin.

The errors we get on the correlation function are overestimated, because the chunks are in the bootstrap treated as independent random variables. But they are dependent. The chunks are correlated. You clearly see in Fig. 3.1 that when you fit a line through the correlation data points 90 percent is on that line.

For more detail check our program cfdeltatau.pro (see Appendix A1).

Total error calculation for more than one quasar spectrum We would like to know what the errors are, when we calculate the correlation function of more than one quasar. Just taking the average is not optimal. We have to use a weighted average by taking the errors of every velocity bin of each spectrum we use. The errors are calculated as explained above and we use formula (3.6) for weighted averages.

Values that have relatively big errors will weigh less in the total value of the correlation function. The total error per velocity bin for all quasars together could have been computed in the same way. But then you are also dealing with the problem of the overestimated errors. But because we now have more spectra, we can take as errors those computed out of a bootstrap resampling of the 6 quasars we used.

We bootstrapped the 6 spectra: instead of choosing chunks of 5 Åas X's, we now take the spectra as X's, so we have a random sample of 6 spectra. Spectra can be taken more than once per sampling. From these bootstrap results we computed the results per velocity bin and took the standard deviation. These error bars are plotted in the figures where the correlation function of 6 spectra is computed. For more detail check our program cfsumcalculation.pro (see Appendix A2).

Chapter 5

Voigt vs. POD
A comparison with previous work

Before the POD method was developed, research has been done on clustering of metal lines by eye and fitting them with the help of Voigt profiles. We would like to make a comparison between our results and their results. Do we get the same shape of correlation? What are the differences?

We compare our results to those of Scannapieco et al. (2006), who did research using 19 quasar spectra from the VLT, taken with the UVES instrument. The resolution of their spectra is high (R $\simeq$ 45,000) almost as high as ours (R $\approx$ 6.6 km/s $\approx$ 45,500 see 2.1) and have a signal to noise from 60-100 per pixel. Not much of a difference with our data, that has a signal to noise of $\approx$ 100.

From the 19 quasars they used 619 CIV lines to find out if these lines cluster or not. These lines were fitted by Voigt profiles so they could measure their column density N and their width b. They measured the distance of the line pairs and divided them into velocity bins. For the correlation value per velocity bin per quasar they used:

$$\xi^{\ell}(v_k) + 1 = \frac{n_k^{\ell}}{\langle n_k^{\ell} \rangle}, \tag{5.1}$$

where n_k^{ℓ} is the number of line pairs with a velocity difference corresponding to a bin k, and $\langle n_k^{\ell} \rangle$ is the average number of such pairs that would be found in the redshift interval covered by QSO ℓ (see Scannapieco et al. 2006).

Per velocity bin the correlation is the number of line pairs that fit in that bin, divided by the number you would expect when the lines would be randomly distributed. So they calculated the probability you can find a line in a certain velocity bin. They also devided their quasar samples into a low and high redshift range. They found that the correlation value depends on the redshift of the quasar.

They took all quasars together and computed the average like this:

$$\xi(v_k) + 1 = \frac{\sum_{\ell} n_k^{\ell}}{\sum_{\ell} \langle n_k^{\ell} \rangle}, \tag{5.2}$$

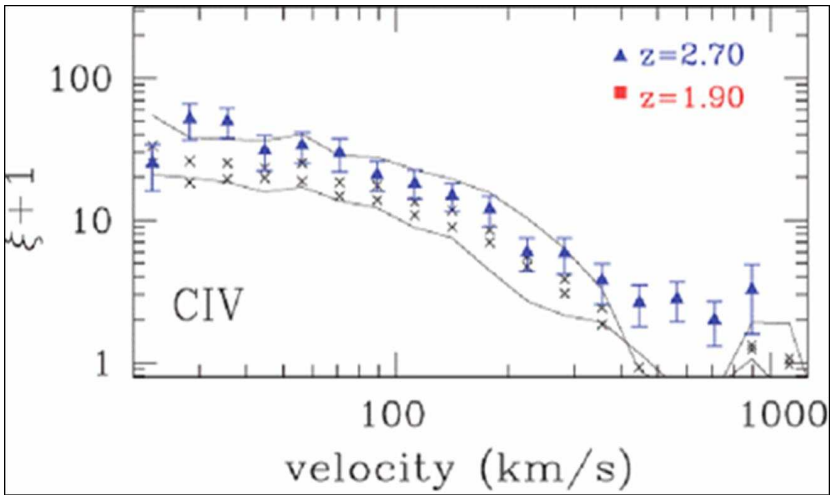


Figure 5.1: *Correlation function made by Scannapieco et al. (2006) (the blue triangles). Made with 19* QSO *spectra and they find a higher correlation at low* Δv *and their* CF *will be zero from a* Δv *of 1000 km/s . There is no clear doublet detection. Figure taken from Scannapieco et al. (2006).*

Per quasar the correlation is normalized. After the normalization of all 19 QSO's, they computed the sum of these QAL's. Scannapieco's result can be found in Fig. 5.1.

The blue triangles in the plots are their results. They detect correlation up to 1000 km/s. The crosses in the plot are results from Boksenberg et al. (2003). They found correlation up to $\approx 500 km/s$.

If we want to make a comparison to our work, we have to adjust our result to their results. So we changed our velocity bins and we also used $\xi + 1$. That plot is shown in Fig. 5.2. Our correlation function has a steeper decline after the elbow.

It is difficult to compare, because Scannapieco et al. calculated the correlation function in a different way. We must not look at the values of the CF, but only at the shape and at the velocity where the correlation stops.

We think that the most important difference is, that they detect clustering up to 1000 km/s and we up to 200 km/s. Scannapieco et al. removed the doublet, but they probably did not do this totally. This can be the reason, why they detected correlation a higher velocity differences. The rest of the doublet may be smeared out over a range of hunderds of km/s. We do not know for sure. So further research is needed, with both methods, because the two methods are giving a different result and we do not know why.

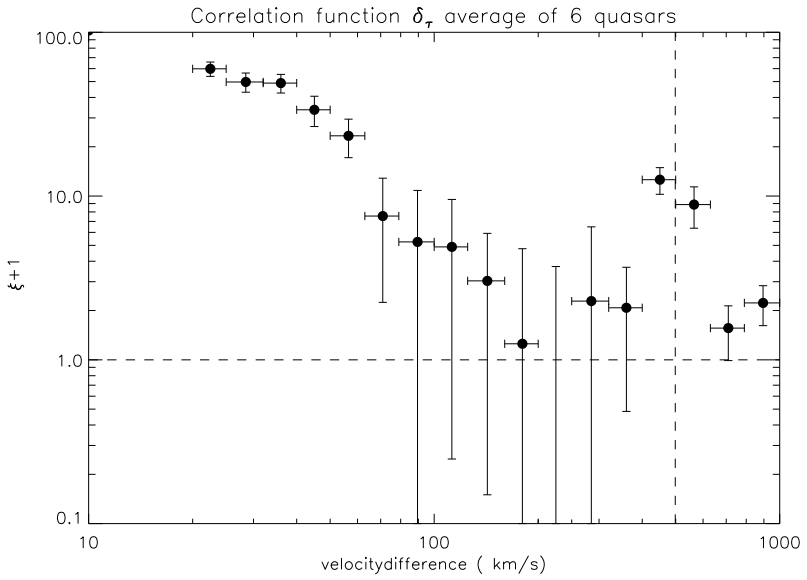


Figure 5.2: *The same result we had in Fig. 3.2 , but devided in the same bins as Scannapieco et al. and we also used* $\xi + 1$.

Chapter 6

Conclusions

Our research has shown that the Pixel Optical Depth method is a good and reliable method to get information about the clustering of metals in the intergalactic medium. In the plots we can clearly see that there is clustering up to 200 km/s and around the doublet spacing. You get a clear indication of the doublet of CIV. The correlation function shows us that there is enough information that can be interpreted.

The clustering up to 200 km/s tells us about the sizes of clouds, if we assume it is Hubble velocity. Using formula (2.5) and formula (2.6) we can compute our limit in the absolute size of the cloud. If it is correct, average sizes of intergalactic clouds are of order hundreds of thousands light years. For a redshift of z = 3 a cloud with a upper limit of 200 km/s is Δl_{proper} 0.64 Mpc and $\Delta r_{comoving}$ is 2.56 Mpc, what means 2,086 thousand light years in proper distance ($1Mpc = 10^6 pc = 3.26 \cdot 10^6 ly$) and 8,358 thousand light years in comoving distance. The correlation function gives us a maximum size for metal density fluctuations, a measure for the scale of clustering.

There is a lot of work to do, because we only tested the method. Good interpretations can be made by comparing the results to simulations. Unfortunately we did not have enough time to do that part of the research. But we will leave that to others.

How to use simulations, in short: simulate a part of the large scale structure of the universe, put galaxies and intergalactic gas in it. The trick is that you should exactly know what you put into your simulation. You can vary the distribution of metals in the simulation. To find out what is the most reliable model, let metals cluster around massive galaxies. Then you should calculate the correlation function and compare it to that of your quasar or to the total of all quasars. Then let your metals cluster around low mass galaxies. You can also distribute them randomly through the universe and see what happens, if you calculate the correlation function.

Finally you hope to get your simulated correlation function the same as that of your quasars. Then you have a good model for large scale structure and you will have new information about the intergalactic medium. You also should find out what sort of velocity differences you are dealing with.

Making simulations and putting different parameters in and out will take a lot of time. Maybe it will take another couple of months. So it can be a good subject for bachelor research projects of next year.

References

Aguirre, A., Schaye, J., & Theuns, T. 2002, ApJ, 576, 1

Aguirre, A., Schaye, J., Kim, T.-S., Theuns, T., Rauch, M., & Sargent, W. L. W. 2004, ApJ, 602, 38

Boksenberg, A., Sargent, W. L. W., & Rauch, M. 2003, ASP Conf. Ser. 297: Star Formation Through Time, 297, 447

Charlton, J., Churchill, C., & Murdin, P. 2000, Encyclopedia of Astronomy and Astrophysics,

Cowie, L. L., & Songaila, A. 1998, nat, 394, 44

Dekking, F.M. et al., 2005, A Modern Introduction to Probability and Statistics, 251

Efstathiou, G.,Schaye, J., & Theuns, T. 2000, Astronomy, physics and chemistry of H_3^+, 358, 2049

Madau, P. 2000, ArXiv Astrophysics e-prints, arXiv:astro-ph/0005106

Scannapieco, E., Pichon, C., Aracil, B., Petitjean, P., Thacker, R. J., Pogosyan, D., Bergeron, J., & Couchman, H. M. P. 2006, mnras, 365, 615

Schaye, J., Aguirre, A., Kim, T.-S., Theuns, T., Rauch, M., & Sargent, W. L. W. 2003, ApJ, 596, 768

Schaye, J., & Aguirre, A. 2005, IAU Symposium, 228, 557

Theuns, T. 2005, IAU Colloq. 199: Probing Galaxies through Quasar Absorption Lines, 185

Appendix A

Programs

A.1 The correlation function of fluctuations in optical depth

A.2 How to calculate the weighted average of more quasar spectra?