Physics Honours Report 1989

A Walk In The Lyman α Forest

David Mar

Except where otherwise acknowledged,
the work presented in this report is my own.

(Signed)

ABSTRACT

The "Lyman α Forest" is the name given to the region of very heavy absorption line density bluewards of a quasar's Lyman α emission peak. This region contains many lines due to Lyman α absorption by clouds of gas distributed at different redshifts along the sightline to the quasar.

A high resolution spectrum of the Lyman α Forest region of the quasar Q2206-199 was analysed in this project. The problem of determining the continuum emission level from the quasar in the presence of so many absorption lines was addressed and a method developed to fit such a continuum. The resulting continuum was seen to imply the existence of a population of weak absorption lines in the spectrum, previously assumed to be noise fluctuations.

The so called "Lyman α clouds", responsible for the absorption may have clustering tendencies, much like galaxies. Published works are divided and inconclusive as to whether the Lyman α clouds are clustered. Possible clustering properties were investigated by calculating a two point correlation function for the clouds along the sightline to Q2206-199. No clustering was found, indicating that the Lyman α clouds are a randomly distributed, intergalactic population of objects.

Who can number the clouds by wisdom
Or who can pour out the bottles of heaven,
When the dust runneth into a mass,
And the clods cleave fast together?
— *Job*, 38:37.

A Walk In The Lyman α Forest

Contents

1. INTRODUCTION

1.1 Quasars as Probes of Intergalactic Space.

Quasars (also called quasi-stellar objects or QSOs) are compact and highly luminous objects at great distances from our solar system. Their size is of the order of a light year or less yet their electromagnetic power output is equivalent to that of hundreds of entire galaxies.

The enormous intrinsic luminosity of quasars enables them to be observed at distances much greater than any other objects. They also have very small angular diameters, unlike other extragalactic objects, as the "quasi-stellar" designation attests. These two properties have meant that, since their discovery in 1963, quasars have not only themselves been studied, but have also proved to be very useful as deep probes of the intergalactic space between us and them. This probing is done by interpreting the absorption features present in their spectra.

A portion of the visible spectrum of a quasar is shown in figure 1.1, in which the light intensity in terms of photon counts is plotted against wavelength. Many of the features present are due to the spectrum of hydrogen.

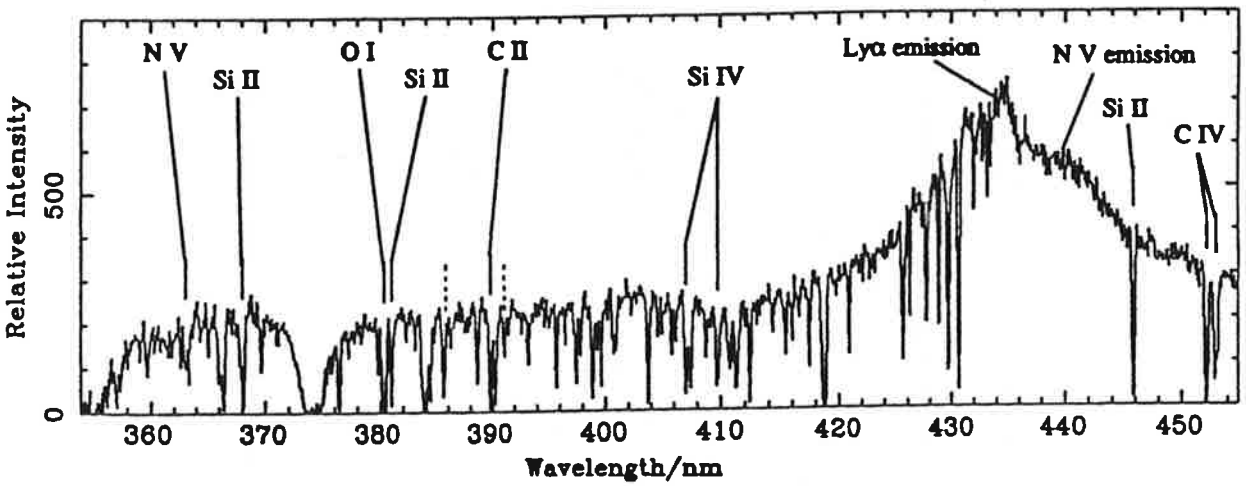


Figure 1.1: Spectrum of quasar Q2206-199 at 0.15 nm resolution.
(Reproduced from Sargent *et al.*, 1988.)

Since hydrogen is the most abundant element in the universe, its spectral lines are usually prominent in astronomical spectra. The Lyman α (Lyα) transition, between the ground and first excited state (*i.e.* a *resonance* line), gives rise to particularly strong emission and absorption. This line has a laboratory wavelength of 121.567 nm, in the near ultraviolet. In general, the most prominent transitions observed in quasar spectra are resonance lines of the most abundant elements. Most of these lines are in the ultraviolet region of the

spectrum and have been observed in absorption by the Copernicus satellite along sightlines through the local interstellar medium.

In quasar spectra, however, these same ultraviolet transitions are seen in the visible region due to a wavelength shift caused by the expansion of the universe[1]. There is overwhelming evidence that the universe has expanded from an early state of great density and temperature (commonly known as the Big Bang). A consequence of this expansion is that the light emitted by distant objects is shifted to longer wavelengths, *i.e.* their light is *redshifted*. This is a true cosmological effect, due to the very expansion of space itself and dependent only on an object's distance from us. It is important to realise that this redshift is not a Doppler effect due to extragalactic objects receding from us. A Doppler shift, either towards the red or the blue, may be superimposed if an object and ourselves are in relative motion. This is small, however, when compared to the cosmological redshifts for objects outside our local group of galaxies.

At such scales, the everyday notion of "distance" becomes less meaningful. The lookback time to an extragalactic object (enforced by the finite speed of light) can be thousands of millions of years. Over the time that the light takes to reach us, the space between us and the object expands. The difficulties encountered in defining distance can be appreciated by looking at a space-time diagram, as shown in figure 1.2.

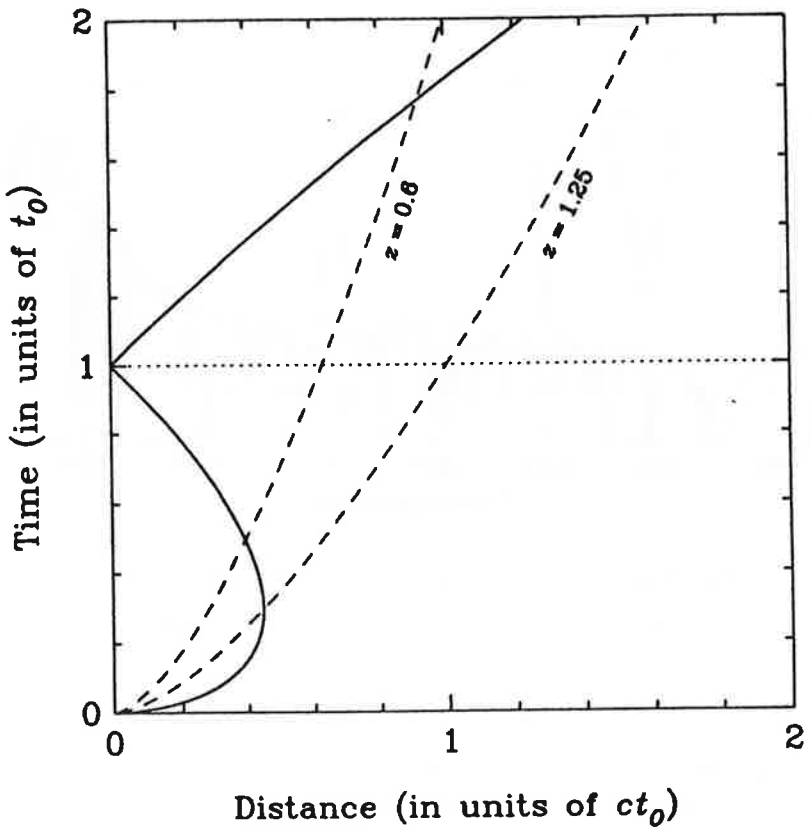


Figure 1.2: Space-time diagram.

[1] The concept of an expanding universe is discussed in appendix A.

In figure 1.2, the present epoch is denoted by the time t_o, from the beginning of the universe at the origin of the diagram. The solid line is the *light cone*, the path that a beam of light would take through space-time. Note that its path is curved, since space expands as the light propagates, and converges to the origin. The dotted lines represent the trajectories of objects of redshifts 0.6 and 1.25. We see such objects on the diagram when and where their trajectories cross our backward light cone. Since distances are measured strictly horizontally on the diagram, it is not clear how the distance between us and an object at high redshift, separated from us by both space and time, should be defined. Merely taking the horizontal displacement on the diagram neglects the fact that the object has by now moved to where its trajectory cuts the horizontal dotted line, yet taking the distance to that point is useless since there is no way in which we can observe the object there.

A definition of an intuitive form of distance is thus not attempted. In its stead, the readily observable parameter of redshift is used, with the understanding that the greater the redshift, the greater the "distance".

As an observational quantity, the redshift, z, is defined for a given spectral line by

$$1 + z = \frac{\lambda_{obs}}{\lambda_{lab}} \qquad (1.1)$$

where λ_{obs} is the observed wavelength and λ_{lab} is the laboratory wavelength of the line. The redshift also relates the scale size of the universe now, R_o, to the scale size of the universe at the epoch of spectral line formation, R, via

$$1 + z = \frac{R_o}{R}. \qquad (1.2)$$

1.2 Explaining Quasar Spectra.

The prominent emission peak in figure 1.1 is the redshifted Lyα line. This is associated with the quasar itself and can be used to determine its *emission redshift*, z_{em}. This redshift assignation to the quasar can be confirmed by comparing wavelengths of other emission lines, including some due to heavier elements (or *metals* in astronomical parlance) such as carbon and silicon, to their laboratory wavelengths. An example in this spectrum is the N V (nitrogen five, or quadruply ionised nitrogen) emission peak indicated near the Lyα emission.

All of the permitted, or large oscillator strength, emission lines in quasar spectra are very broad. This implies a large *velocity dispersion* in the emitting gas. The line is Doppler broadened by the violent motion of matter in the emitting region.

A quasar spectrum also typically contains many narrow absorption lines. These are due to absorption by elements in galaxies and intergalactic gas clouds between us and the quasar. Since such intervening matter is necessarily closer to

us, the absorption lines occur at absorption redshifts $z_{abs} \leq z_{em}$. The extreme narrowness of the lines is the indicator of their origin in intervening objects. An alternate explanation might be that such lines are caused by matter ejected from the quasar at high speed. Any such matter that absorbs light in our sightline would be travelling towards us and so give rise to absorption lines at lower redshift. However, any matter ejected so violently from a quasar would be expected to have a considerable velocity dispersion, leading to absorption lines broader than those seen. Indeed, some quasars show evidence of such broad absorption lines, which are easily recognised. It is also not energetically feasible for a quasar to eject enough matter at high enough velocities to cause all of the lines seen. We can thus be certain that the absorption is due to passively intervening matter. The situation is illustrated in figure 1.3.

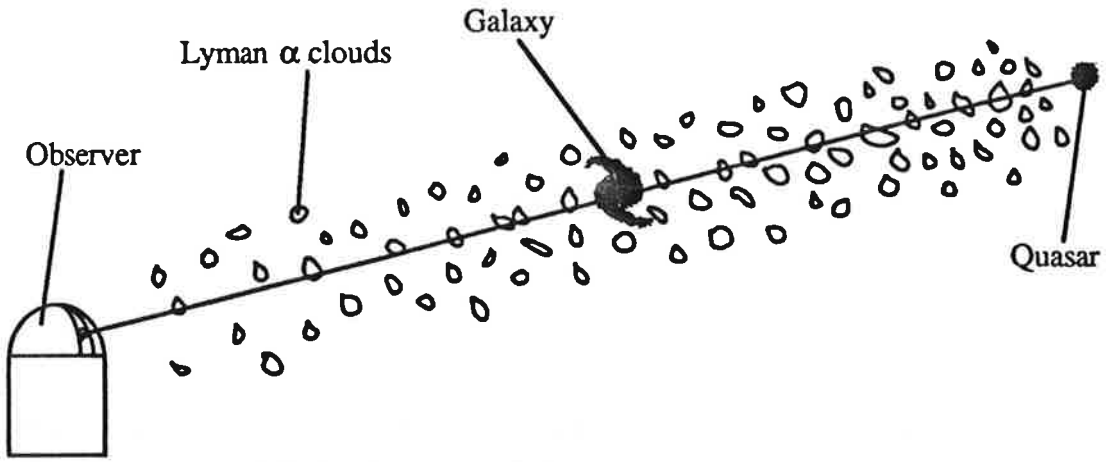


Figure 1.3: Absorption lines in a quasar spectrum are generated by intervening matter.

The intervening objects may be galaxies or proto-galaxies, in which case a *system* of absorption lines at the redshift of the galaxy will be seen. This will include prominent metal lines. Several such systems are seen in many quasar spectra. They are easily identifiable as metal systems because each line corresponds to a known laboratory wavelength redshifted by the same amount as the other lines in the system. A set of lines thus appears with the same relative positions as in a laboratory spectrum. One such system, at a redshift of 1.92, is shown in figure 1.1. The lines are labelled with the element and a Roman numeral which is, by convention, one greater than the ionisation state.

Many lines, however, do not fit into metal systems. In particular, as can be seen in figure 1.1, there are often a great many absorption lines blueward of the Lyα emission peak. Most of these cannot be assigned to metals and there are no such lines redward of Lyα emission. This indicates that the lines are due to Lyα absorption in intervening clouds of hydrogen, the so-called *Lyα clouds*.

1.3 The Lyman α Forest.

The great line density blueward of Lyα emission has led to that region of a quasar spectrum being termed the *Lyα Forest*. The abundance of Lyα lines implies the existence of many discrete clouds of hydrogen at various redshifts along the sightline to the quasar.

Such clouds would also be expected to absorb light via the Lyman β (Lyβ) transition of hydrogen (and, of course, the other lines of the Lyman series) as well as the Lyα transition. Being between the ground and second excited state, Lyβ has a shorter wavelength than Lyα. In spectra that extend far enough blueward to include the quasar's Lyβ emission, Lyβ absorption lines are found at the same redshifts as the Lyα forest lines. Lyman γ lines have also been found at corresponding redshifts in some spectra. We can thus be certain that there are clouds of hydrogen at the redshifts implied by the Lyα forest lines.

Since the Lyα forest lines define specific redshifts for the clouds, we can search for metal lines at those redshifts. None are found. This implies that the Lyα clouds may not have been enriched with heavy elements since the Big Bang. They are very possibly clouds of primordial matter — a sampling of the early composition of the universe.

Understanding the Lyα clouds and their properties is therefore an important goal in the process of understanding the universe as a whole.

1.4 The Project.

This project was concerned with two different aspects of the analysis of quasar spectra.

Firstly, in order to study absorption lines quantitatively it is necessary to normalise a spectrum by fitting a continuum emission level, *i.e.* the intensity level that the spectrum would have if no absorption had occurred. An objective method for doing this, utilising the statistical properties of the photon counting method of accumulating the spectrum, was developed. This was then applied to a high resolution spectrum of the quasar Q2206-199 and compared to other methods currently in use.

Secondly, a study was made of the clustering properties of the Lyα clouds along the sightline to the same quasar. The metal systems evident in quasar spectra exhibit a strong tendency to cluster on scales comparable to clusters of galaxies. Published papers have given only inconclusive results on the clustering of the Lyα clouds (Sargent *et al.*, 1980). Any clustering present could affect our current ideas on the evolution of the universe. This part of the project was intended to further investigate these properties.

The following sections of this report detail the work done in this project. Section 2 contains details of the observations made of the quasar Q2206-199 and the preliminary data reduction. Section 3 describes the problems of fitting a

continuum emission level to a quasar spectrum and outlines the approach developed in this project to fit an objective continuum in the Lyα forest region of Q2206-199. A discussion of the results and the implications of this work is included. Section 4 describes the methods used to study the clustering properties of the Lyα clouds and compares the results obtained with those from work by other researchers. Some possible ramifications of these results are discussed. An overall conclusion is presented in section 5, summarising the main results and their implications, and outlining possibilities for further research.

2. OBSERVATIONS

The spectra of the quasar Q2206-199 studied in this project were recorded at the Anglo-Australian Telescope (AAT) at Siding Spring Observatory, Coonabarabran, New South Wales in 1988, August 14—16. Max Pettini (Anglo-Australian Observatory) and Richard Hunstead made the observations using the University College London Echelle Spectrograph (UCLES) and the Image Photon Counting System (IPCS) at the 3.9 m telescope's Coudé focus.

Nine spectral orders (orders 52 through to 60) from the high resolution spectrograph were recorded, together with spectra of the blank sky near the quasar. It is important to note that the quasar spectra include background counts due to the sky, which must be subtracted before any data analysis is done. To enable wavelength calibration of the spectra, a thorium-argon lamp was used before and after each observation to obtain a spectrum with lines of known wavelength.

After the observations had been made, preliminary reduction of the order 58 spectrum was performed by Pettini. This involved using the Th-Ar spectral lines to assign a wavelength to each spectrum pixel, or bin, in the object-plus-sky and sky spectra. An average sky spectrum was formed from three sky spectra recorded at adjacent spatial positions on the detector to improve the signal-to-noise ratio in the sky spectrum. The resulting sky spectrum was then subtracted from the object-plus-sky spectrum to give an object spectrum with no background sky counts.

Note that the actual photon count levels in each wavelength bin are not distorted by any of these operations. The sky subtraction process preserves the underlying statistics of the photon counts. The procedure of linearising the wavelength scale so that each bin is separated from its neighbours by a constant wavelength increment, which is normally done as a preparation for further study, was purposely not done as it would have resulted in rebinning of the photon counts, distorting their statistical distribution.

The resulting sky-subtracted object spectrum is shown in figure 2.1. The spectrum has a resolution of 0.008 nm, which is at the limits of current optical spectrum resolution (*cf.* the 0.15 nm resolution of figure 1.1). Note that figure 2.1 only covers a small portion of the Lyα forest in figure 1.1, indicated by the vertical dotted lines near 390 nm.

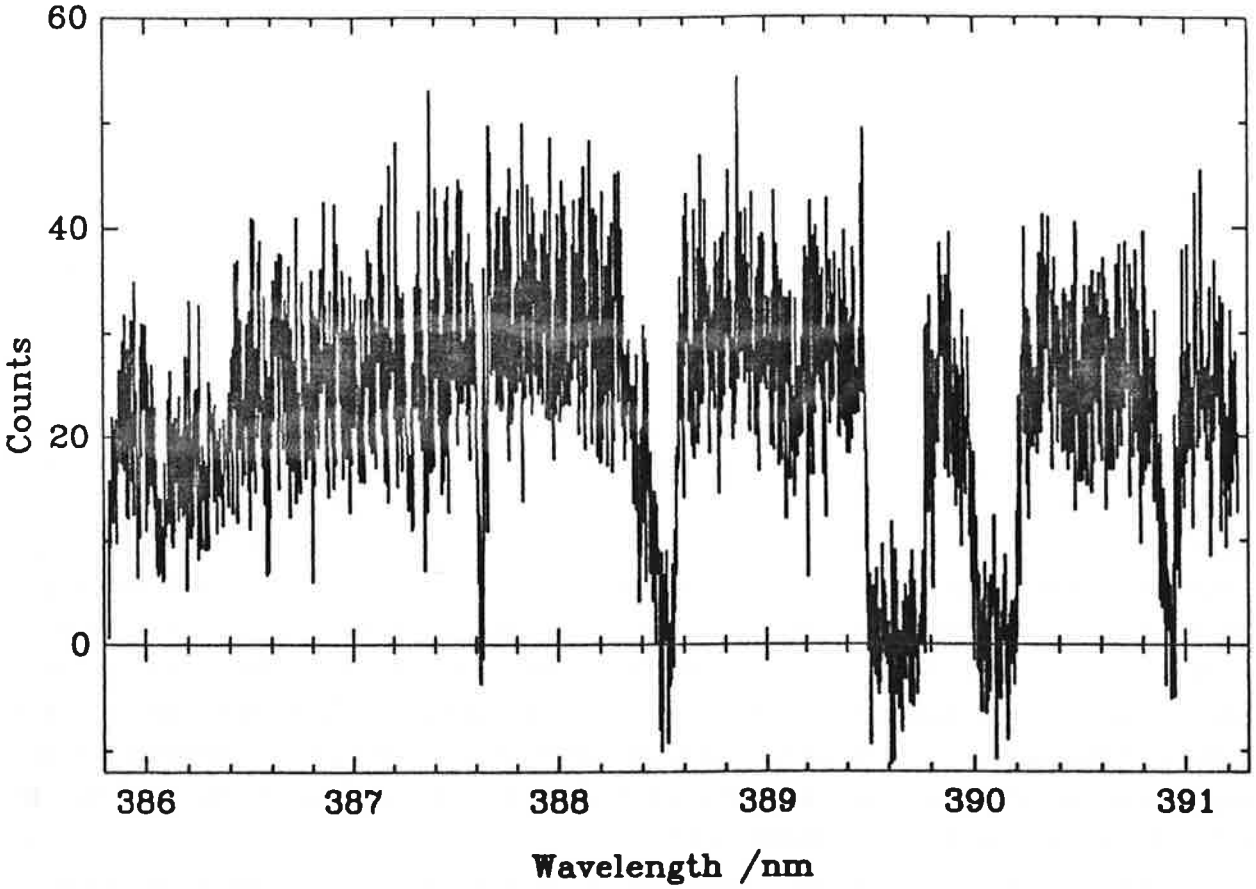


Figure 2.1: Sky-subtracted order 58 echelle spectrum of Q2206-199.

3. CONTINUUM FITTING

3.1 The Continuum.

In order to interpret the information contained in a quasar spectrum correctly, it is necessary to define a continuum emission level, from which quantitative measurements of the absorption lines can be made. As can be imagined from figures 1.1 and, especially, 2.1, this is not a trivial task.

Many authors have used estimates based on the assumption that most of the intensity fluctuations are due to noise. Correspondingly, they have placed the continuum so as to pass through the centre of the small amplitude fluctuations in intermediate resolution spectra such as that shown in figure 1.1. (Most work on quasar spectra to date has been done at such resolutions.)

There is reason to believe, however, that the true continuum lies higher and that a population of weak absorption lines is lowering its apparent level. This has come from work by Hunstead *et al.* (1986), who advocated the view that the continuum level should be set to skim through the uppermost tips of the spectrum. Steidel and Sargent (1987) have investigated this by taking spectra covering a large range of wavelengths on both sides of Lyα emission for eight quasars. A power law was seen to be a good fit to the continuum level redward of Lyα, as expected from theories of radiation processes. When extrapolated blueward of Lyα emission, the same law was seen to skim the tops of the intensity fluctuations in all of the spectra. Steidel and Sargent commented on the enormous line density blueward of Lyα emission and highlighted the problem of defining a satisfactory continuum level.

An objective method of setting something close to the true continuum is clearly needed. Only one supposedly objective method has previously been used, by Young *et al.* (1979). Both it and the method developed in this project rely on the statistics of photon counting.

3.2 Poisson Statistics.

A source which has a totally featureless spectrum, in the sense that the rate of photon emission per unit wavelength range is independent of wavelength, is called a *flat spectrum* source. Ideally, such a flat spectrum source would produce a perfectly flat response in a detector. However, when photon count levels are low, as they are when quasar observations are made, the statistics of counting individual photons becomes important. The arrival of photons at a detector follows a Poisson probability distribution:

$$P(k) = \frac{e^{-\lambda}\lambda^{k}}{k!} \qquad (3.1)$$

where $P(k)$ is the probability that k photons will arrive per unit time and λ is a constant. The mean and variance are both equal to λ.

An intrinsically flat spectrum is thus expected to give rise to a measured spectrum containing random intensity fluctuations with a variance equal to the mean photon count level. It can be seen that the mean count level can be determined by either a simple averaging calculation, or by estimating the standard deviation in the count levels of the measured spectrum and then squaring the result.

For a quasar spectrum, the presence of absorption lines will imply that these values are not equal. The previously used "objective" method of continuum fitting involved splitting up the spectrum into many small wavelength intervals. The two mean count level estimates were made on each interval and the lowest count rejected iteratively until the estimates agreed with each other. The resulting value was then taken to be the true continuum level on each interval. This method can lead to low continuum estimates because some absorption contaminated count levels will always be included in the final result. Agreement between the mean and variance does not guarantee that the mean defines the true continuum level.

3.3 The Algorithm.

The algorithm used in this project involved estimating the standard deviations of *positive excursions* of the count levels from the continuum by two different methods. Only the positive excursions were used because they could be regarded as pure noise, whilst the negative excursions from the continuum could be contaminated by the presence of the absorption lines. In fact, it was recognised that the positive excursions will also be affected by the presence of absorption lines, so that some of the expected positive excursions will be pulled down to lower levels. This means that any derived mean count levels will be lower limits to the continuum level. The procedure used is described in the following paragraphs and also in the flowchart in figure 3.1.

The spectrum was first split up into many short wavelength intervals, each chosen to be 0.1 nm in length. This value was chosen so that a reasonable number of data bins (35—40) would be included in each interval, allowing useful statistical calculations. Before anything else was done, Chauvenet's criterion for the rejection of suspicious data points was applied to each interval and the points rejected were not included in any further calculations[2].

An estimate of the mean continuum level on each interval was made by first taking two initial estimates, these being zero and the maximum count level in the interval. The estimates were bisected to give a third mean level estimate. Two different calculations of the standard deviation in the count levels were made, assuming that this estimate was the true mean.

---

[2] See Appendix B.

From Poisson statistics, the expected standard deviation in the sky-subtracted spectrum is

$$\sigma_{\text{Poiss}} = \sqrt{m + \frac{4}{3}s} \tag{3.2}$$

where m is the estimate of the mean and s is the mean count level of the sky spectrum over the same wavelength interval[3].

The other calculation of the standard deviation originates from the method of maximum likelihood[4], which gives

$$\sigma_{\text{ML}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2} \tag{3.3}$$

where there are n points at which the count level, x_i, is greater than the mean estimate, m.

By comparing these two estimates of σ, it could be decided in which half of the bisection interval the true mean lay. The other half was discarded and the remaining half bisected again. This process was iterated until the upper and lower estimates of the mean were within a prescribed, small difference, which was set at 0.000 01. A final bisection was taken to be the value of the continuum level in the wavelength interval in question.

This technique was repeated on each wavelength interval to give equally spaced estimates of the continuum level all along the spectrum.

Figure 3.2 shows the results of these calculations, plotted on a smoothed version of the order 58 spectrum. Some degree of digital smoothing is usually applied to raw spectra before presentation to reduce the apparent noise level and highlight the more extended features. The points are shown with 1σ error bars, given by the standard error of the mean:

$$\text{S.E.M.} = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(x_i - m)^2} \tag{3.4}$$

where now n is the total number of count levels x_i and m is the fitted continuum level.

The points are plotted at the centre of each of the 0.1 nm intervals used in the calculations. These intervals begin at 385.8 nm and end at 391.3 nm, covering the full extent of the spectrum.

---

[3] See Appendix C for the derivation of this result.

[4] See Appendix D.

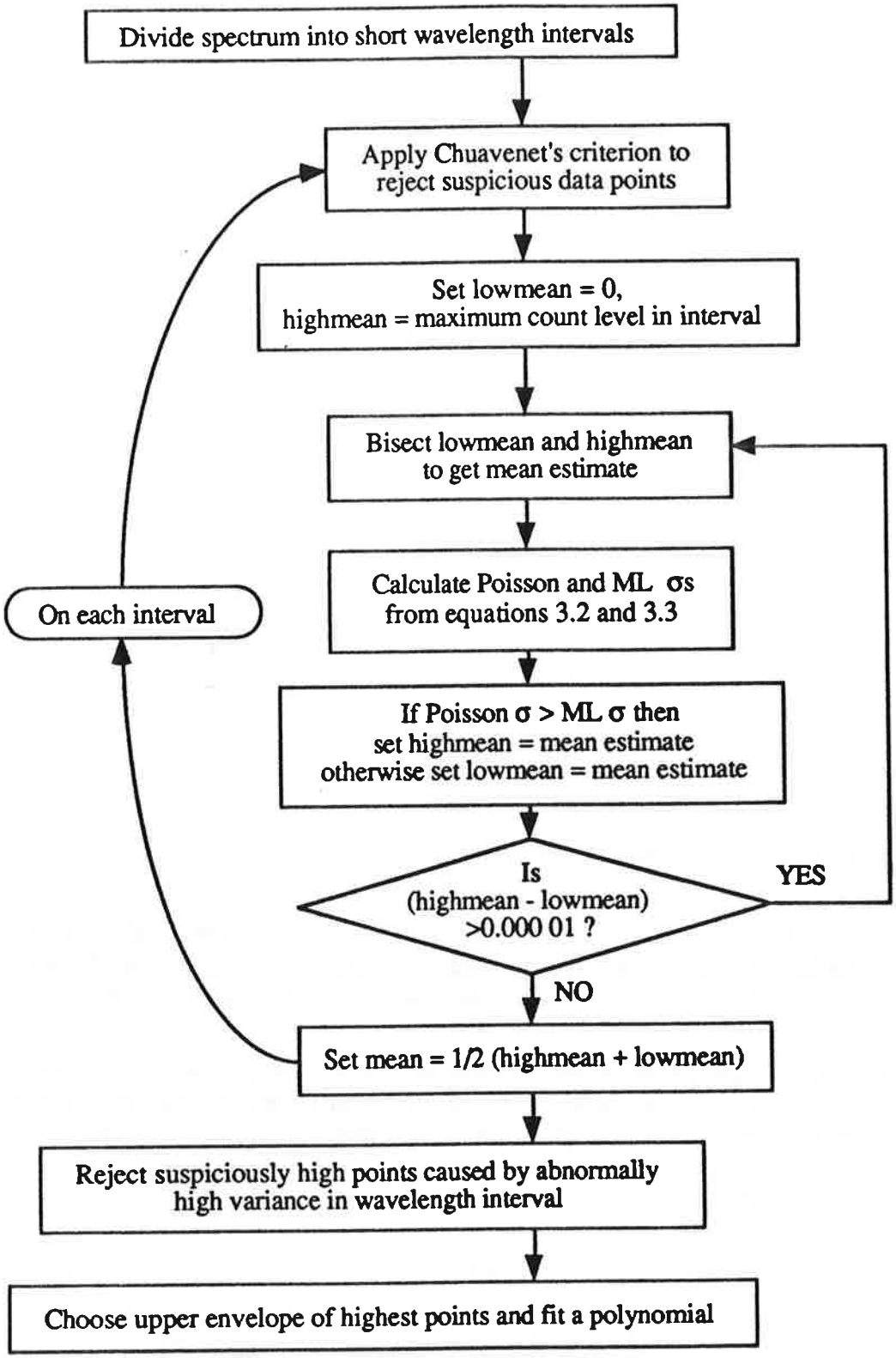


Figure 3.1: Flowchart outlining the continuum fitting procedure.

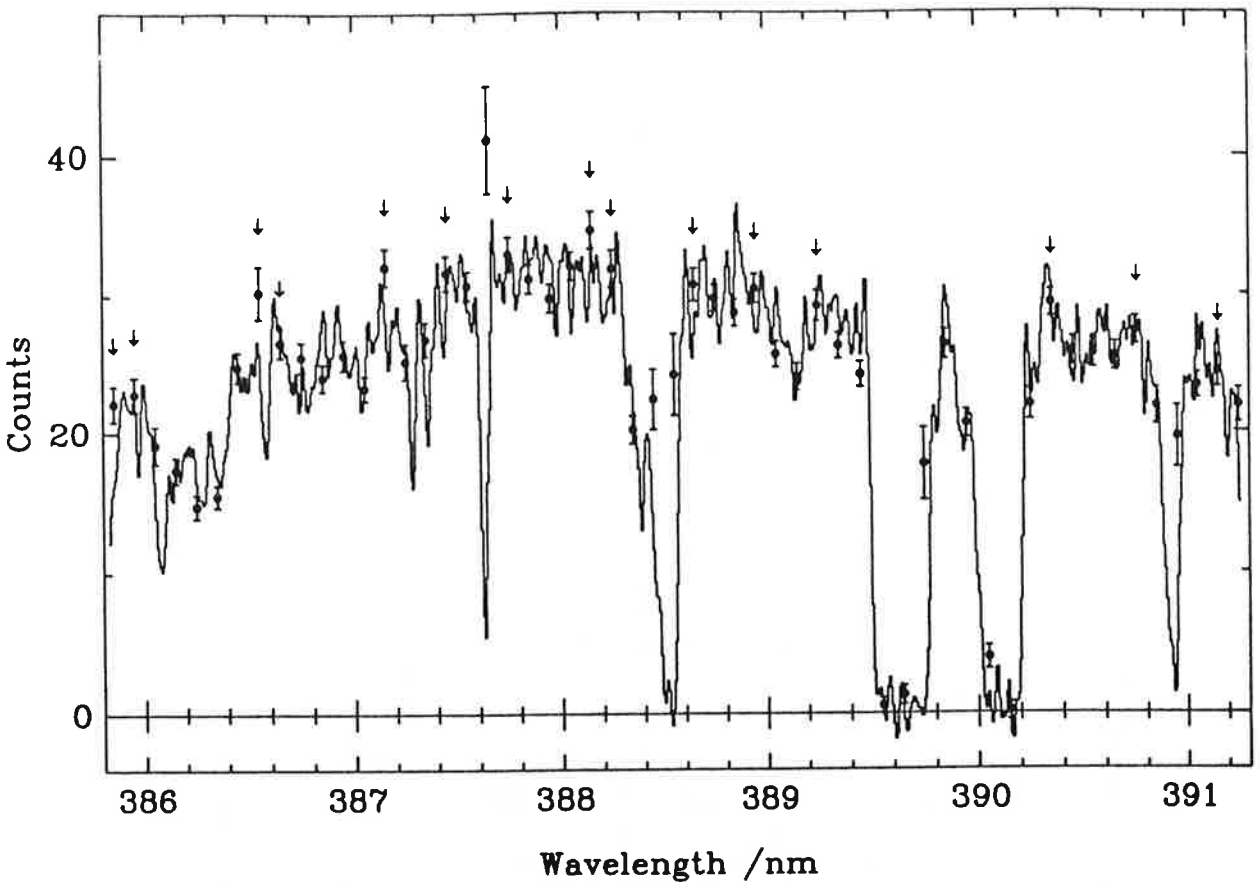


Figure 3.2: Continuum fit points on order 58 spectrum, smoothed to 0.02 nm resolution.

It is important to note that, because of the presence of strong absorption lines and the possible presence of many weak absorption lines, the derived continuum estimates will represent a lower limit to the true continuum level on each interval.

The suspiciously high point at 387.65 nm is due to the presence of both a deep absorption feature and several very high count levels in the same wavelength interval. The artificially high standard deviation results in Chauvenet's criterion not rejecting the abnormally high count levels and so in the assignation of an unjustifiable continuum level. This point was therefore not considered for further fitting.

Once these points had been generated, several of the highest points, indicated by arrows in figure 3.2, were chosen as representing the closest objective values to the true continuum that could be attained. Since the continuum level over a wavelength range as short as 5 nm is expected to be very smooth, the lower points must be regarded as having been depressed by the presence of absorption lines.

A fifth order polynomial was fitted to the selected points by the routine SPIFIT in the astronomical data reduction software package FIGARO. The resulting continuum level is shown in figure 3.3, together with a continuum fit

by Young *et al.*'s method, outlined in section 3.2. A fifth order polynomial was used to fit the highest points generated by this method as well.

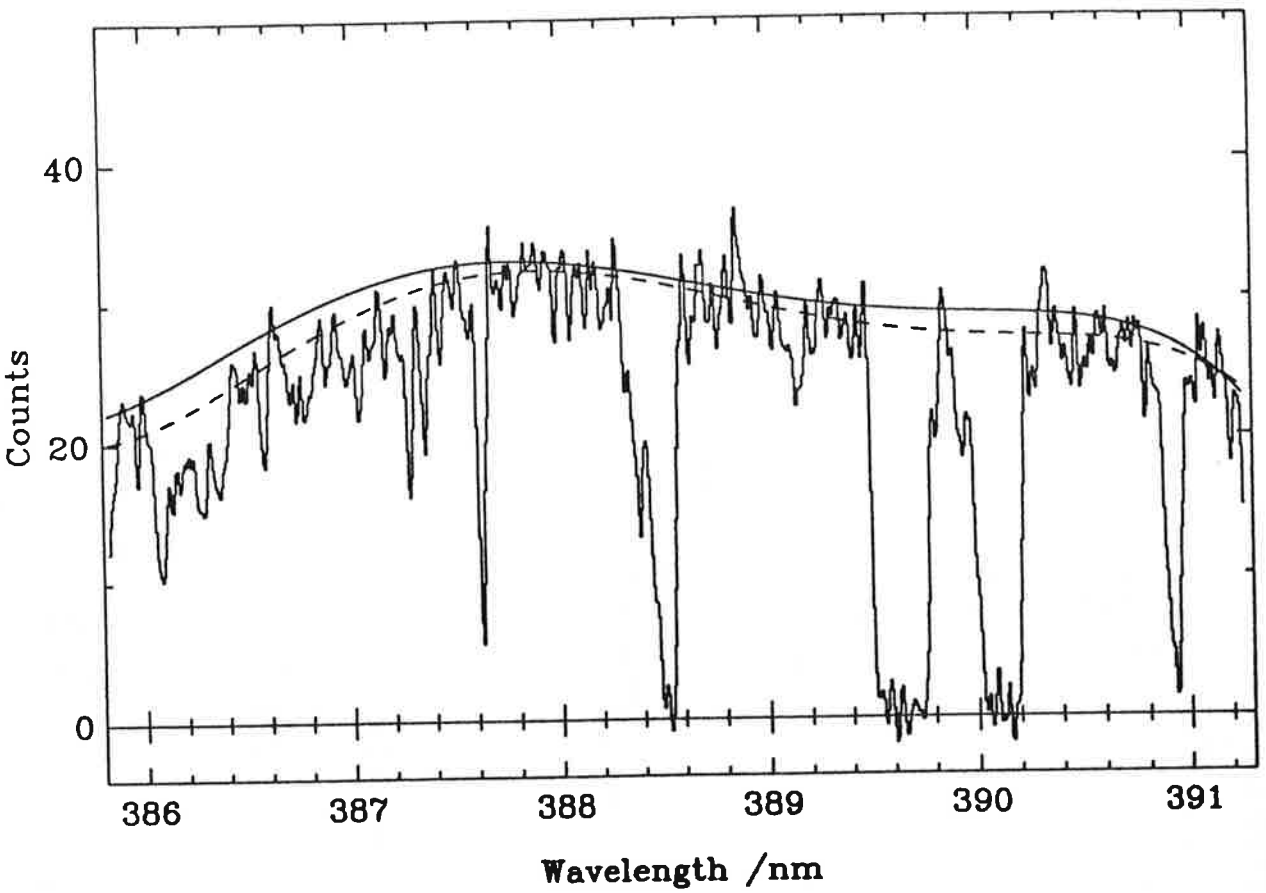


Figure 3.3: Continuum fit on smoothed spectrum.

The solid continuum line is the fit from the procedure developed in this project, whilst the dotted line is that derived from Young *et al.*'s method. This plot shows clearly that, even at the very high resolution of the echelle spectrum, Young *et al.*'s method suffers from a systematic lowering of the fitted continuum because of the presence of the absorption lines.

3.4 Discussion of Result.

In order to compare the continuum fit on the 0.008 nm resolution spectrum to an intermediate resolution spectrum, the spectrum was first divided by the continuum fit and then digitally smoothed to resolutions of 0.02 nm, 0.05 nm and 0.15 nm. The resulting spectra are shown, together with the portion of a real 0.15 nm resolution spectrum covering the same wavelength range, in figure 3.4. The real 0.15 nm spectrum was normalised by applying Young *et al.*'s method over a wavelength range extending beyond the echelle spectrum by ~5 nm on each side. The bin size chosen for this continuum fit was 1 nm, giving each wavelength interval 20 data values.

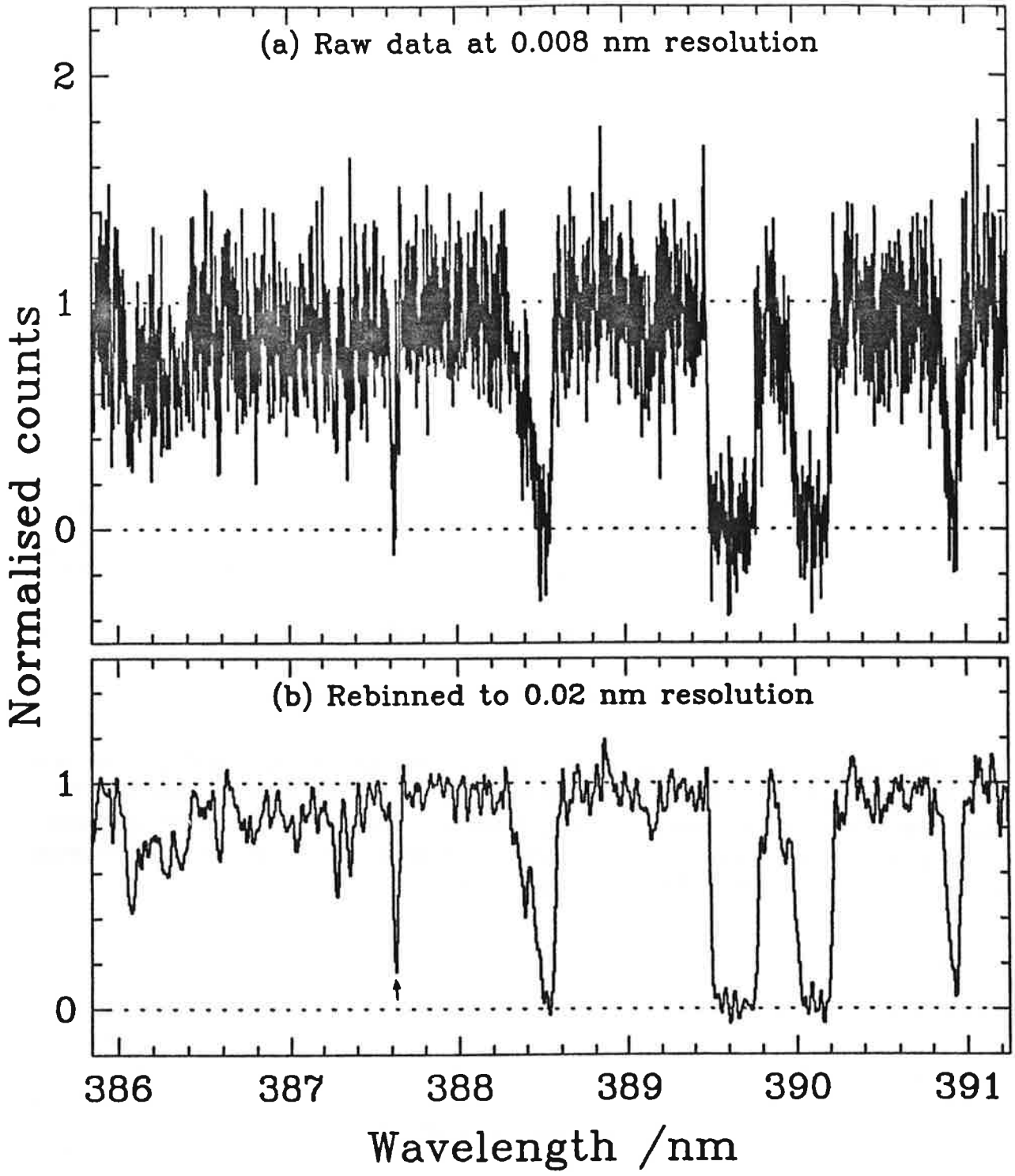
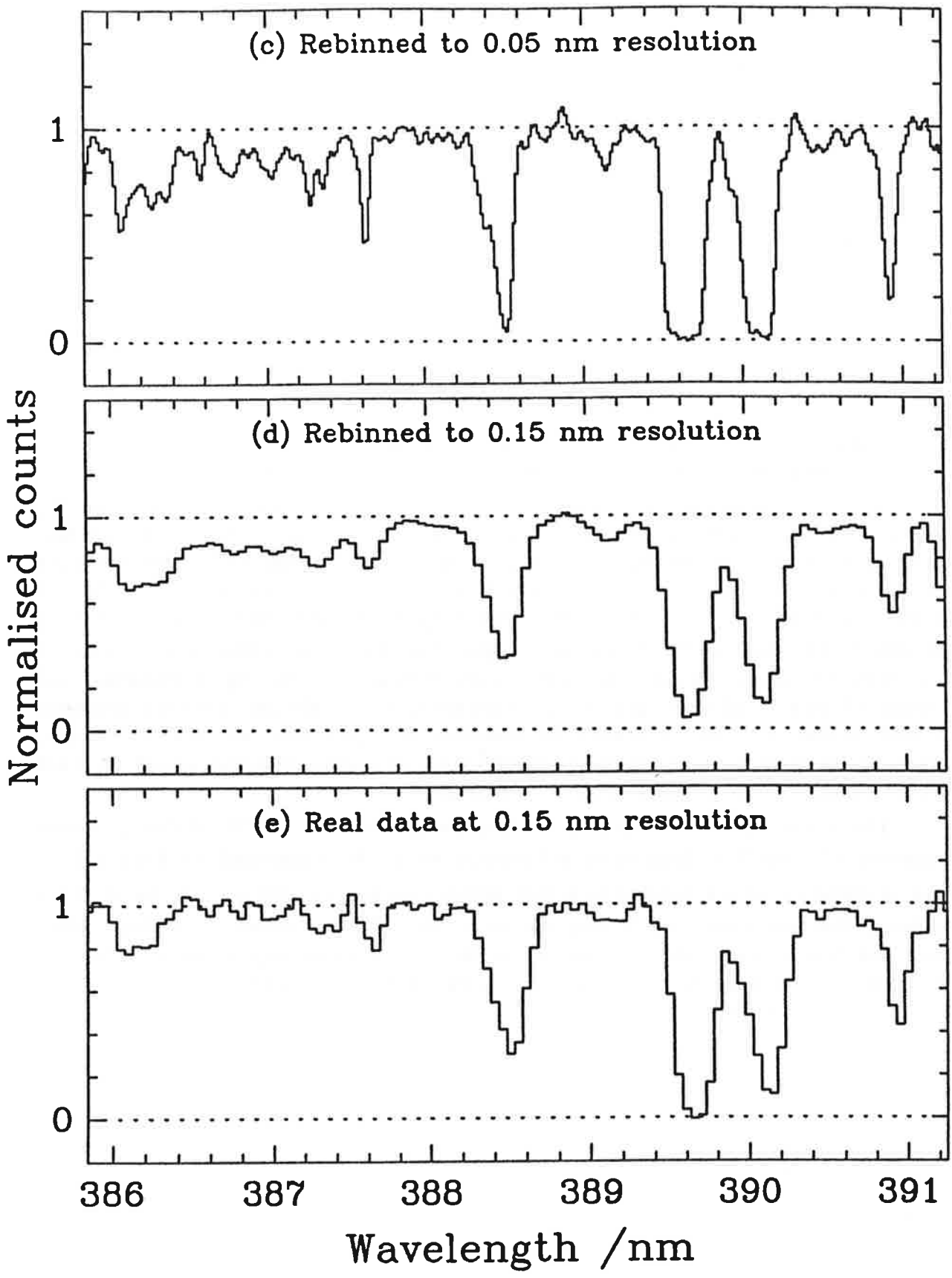


Figure 3.4: (a)—(d) Order 58 spectrum, normalised and smoothed.
(e) Real 0.15 nm data normalised by Young *et al.*'s method.


(c) Rebinned to 0.05 nm resolution
(d) Rebinned to 0.15 nm resolution
(e) Real data at 0.15 nm resolution
Normalised counts
1
0
1
0
1
0
386
387
388
389
390
391
Wavelength /nm

It can be seen from figure 3.4 that the continuum fitting method developed in this project implies that the true continuum lies near the uppermost tips of intermediate resolution spectra. Note particularly that the 0.15 nm resolution smoothed spectrum lies entirely below the fitted continuum level. For obvious reasons, no author has ever published a continuum level lying wholly above a 0.15 nm resolution spectrum. Most would tend to place the continuum as suggested by figure 3.4(e), or even lower. Clearly, what would be assumed to be noise fluctuations in the intermediate resolution spectrum correspond to real features in the high resolution spectrum. Thus many absorption lines can escape detection altogether, such as the arrowed line in figure 3.4(b), which is clearly present but is not at all obvious in figure 3.4(e).

What is more, when the continuum level has been set too low, any absorption lines which are still detected and identified as such will appear to be weaker than they really are. It is possible to use the strength of a line to calculate the *column density* (a two dimensional density along our line of sight) of the absorbing species. A depressed continuum level implies a lower column density. This can lead to apparently low abundance levels of heavy elements in intervening galaxies. Webb (1987) has calculated that a depression of the continuum level by 5% will result in errors of ≥20% in column density values.

There is clearly an advantage to be gained in identifying narrow lines by having spectra at as high a resolution as possible. Even so, without an accurate continuum fit many weak lines, such as those which appear to be present near 388 nm and most easily visible in figure 3.4(b), will be missed. Indeed, although the existence of many very weak lines has long been suspected, this is the first indication that such lines are genuinely present. The population of Lyα clouds may thus be much larger than previously realised, including a considerable number of weakly absorbing (*i.e.* low column density) clouds. Further research on this issue will be necessary to confirm the presence of such systems, requiring more data to be taken at resolutions and signal-to-noise ratios at least as high as the 0.008 nm data used in this project.

There has been considerable discussion (Hunstead, 1987) of the possible evolution of Lyα line density as a function of z. As suggested by figure 1.3, there is some evidence that the Lyα line density increases with increasing z. This suggests that the Lyα clouds may disperse into the intergalactic medium over time, and limits the models proposed for their manner of confinement. Clearly, this question can only be addressed for weak lines if the continuum level can be specified with confidence.

4. CLUSTERING OF LYMAN α CLOUDS

4.1 Introduction.

The clustering properties of the Lyα clouds have been debated many times since the benchmark paper by Sargent *et al.* (1980), which asserted that there was no clustering of the Lyα clouds on any scale along the sightlines to a sample of 5 quasars. Bechtold (1987) produced results implying that this result held for another 4 quasars. These results were, however, taken from intermediate resolution spectra, generally of the order of 0.15 nm. At this resolution, any small scale clustering is likely to be masked by the blending of adjacent lines in the Lyα forest.

On the other hand, Webb (1987), with spectra of resolutions going down to 0.1 nm, claimed strong evidence for clustering at small distance scales. Clearly, these contradictory results at intermediate resolution need to be checked. It is advantageous to go to much higher resolution data because it allows identification of separate lines at much closer wavelength spacings and it is at these small scales that any clustering is most likely to be observed.

The most commonly used technique in searches for clustering of extragalactic objects at various redshifts is the calculation of a statistic known as the two point correlation function. This was the chosen method for this project, since it is simple to implement and lends itself easily to modifications that rectify certain problems (which will be described in section 4.3).

4.2 The Two Point Correlation Function.

Although the term is misleading, we can define a *velocity splitting*, v, between any two objects at different redshifts. This is the velocity difference between the two objects that would be necessary to produce the observed redshifts by the Doppler effect. To a good approximation, the velocity splitting for small redshift differences is given by

$$v = c\,\frac{z_2 - z_1}{1 + (z_2 + z_1)/2} \qquad (4.1)$$

where z_1 and z_2 are the redshifts of the objects and c is the speed of light (Webb, 1987). The velocity splitting is commonly expressed in kilometres per second.

An indication of the clustering properties of objects can be gained from the *two point correlation function*, $\xi(v,\Delta v)$ (Peebles, 1980). This can be defined by

$$\xi(v,\Delta v) = \frac{N_{obs}(v,\Delta v)}{N_{exp}(v,\Delta v)} - 1 \qquad (4.2)$$

where $N_{obs}(v,\Delta v)$ is the number of splittings observed between v and $v+\Delta v$ and

$N_{exp}(v, \Delta v)$ is the number of splittings expected in the same interval from a purely random distribution of objects in redshift (Bechtold, 1987). The correlation function measures the fractional excess of objects in the given velocity splitting range over that of a random distribution. Since the velocity splitting is essentially a difference in redshifts, this excess corresponds to the excess of objects within a certain redshift, or "distance", of each other. A positive correlation function thus indicates a clustering of objects.

This statistic has been used to measure the clustering of galaxies, which show a definite clustering at velocity splittings below $\approx$2000 km.s^{-1}. This is expected because we can observe clusters of galaxies directly, which have separations which would lead to velocity splittings of that order. For Lyα clouds, it is not clear what we can expect.

By listing the wavelengths of all of the detected Lyα lines in the Lyα forest one can obtain, from equation 1.1, a list of the redshifts at which Lyα clouds occur along the line of sight to a particular quasar. Using this and equations (4.1) and (4.2), it is possible to investigate whether or not there is any clustering of the clouds in a radial direction from us.

4.3 The Implementation.

The two point correlation function was calculated for the Lyα lines in the spectrum of Q2206-199. An absorption line list from UCLES orders 52 to 60 was compiled by Linda Smith (UCL). Of the 101 lines detected, 21 were identified as metal lines, leaving 80 lines due to Lyα.

A FORTRAN program was written to calculate the value of $N_{obs}(v, \Delta v)$ from the Lyα line list. The value of $N_{exp}(v, \Delta v)$ was calculated by a Monte Carlo simulation. A random number generating function (RAN1 from the book *Numerical Recipes*) was used to generate 80 random wavelengths in the same wavelength interval as that covered by the UCLES data. Values for $N_{exp}(v, \Delta v)$ were calculated using these random wavelengths. This random procedure was repeated a hundred times and the values for $N_{exp}(v, \Delta v)$ averaged for each velocity splitting bin. This ensured that the final values for $N_{exp}(v, \Delta v)$ were closer to expectation values than single trial random values. This was confirmed by running the procedure with several different random number seeds and finding that the sets of $N_{exp}(v, \Delta v)$ values differed only very slightly.

Two complications made this method of generating $N_{exp}(v, \Delta v)$ values preferable to doing what would otherwise have been a simple theoretical expectation value calculation. Firstly, the UCLES data consisted of different spectral orders of about 5 nm of spectral coverage separated by gaps of about 2 nm. The resulting wavelength gaps in the line list were compensated for by simply requiring that no randomly generated wavelength fall in a region not covered by the UCLES data.

The second complication arose from the finite width of the Lyα lines. Two lines which are closer than a certain wavelength difference apart will appear blended together and be difficult to resolve, ultimately appearing as only a single line. In the UCLES data, no lines closer than 0.04 nm could be resolved. This

wavelength separation corresponds to a velocity splitting of ≈20 km.s^{-1}. This was compensated for by rejecting any randomly generated wavelength within 0.04 nm of the previously generated wavelengths and generating another wavelength in its stead.

Other authors have commented that the two point correlation function becomes unreliable at small velocity splittings due to the presence of line blending (Webb, 1987). This is because they have assumed in their correlation function calculations that arbitrarily close lines could be resolved. This results in an artificial lowering of $\xi(v,\Delta v)$ at small splittings. It has become customary for authors to remove the first one or two velocity splitting bins from their plots of $\xi(v,\Delta v)$ because of this. The method adopted here eliminates this problem and allows the resulting plot of $\xi(v,\Delta v)$ to be considered reliable down to the first bin.

4.4 The Result.

The calculated two point correlation function for the Q2206-199 Lyα clouds is shown in figure 4.1, with an inset showing a magnification of the region below 3000 km.s^{-1} velocity splitting.

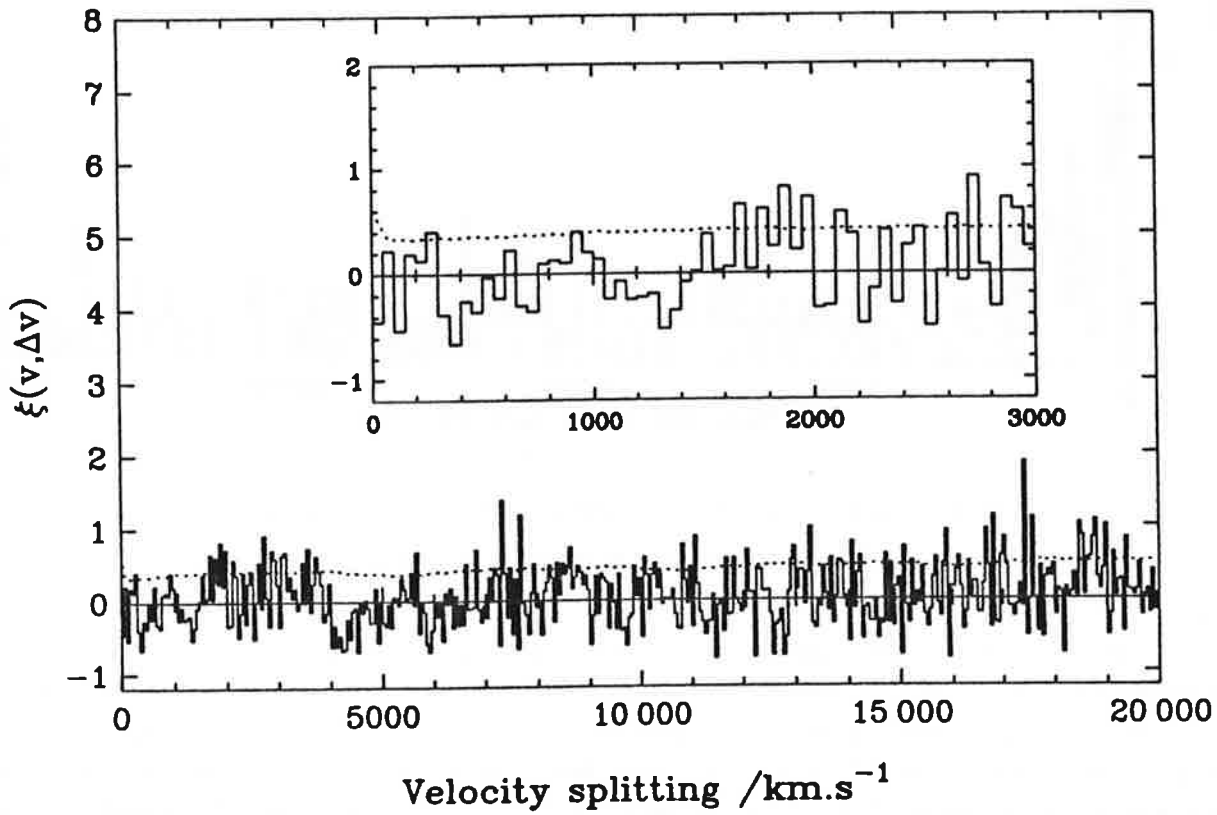


Figure 4.1: Two Point Correlation Function of Q2206-199 Lyα Clouds.

The histogram is formed by plotting the value of $\xi(v,\Delta v)$ at all velocity splittings in the bins between v and $v+\Delta v$. Here the bin size is 50 km.s^{-1}. The dashed line shows the 1σ level for deviations from a random cloud distribution. Clearly there are no significant deviations present at any velocity splitting.

To appreciate the significance of this result, one must contrast it to results obtained for other line populations. A strikingly different result has been obtained in studies of C IV (carbon four, or triply ionised carbon) lines, where a strong tendency for clustering at scales below $\Delta v \approx 600$ km.s^{-1} is found (Sargent *et al.*, 1988). A reproduction of the C IV result by Sargent *et al.* is shown in figure 4.2.

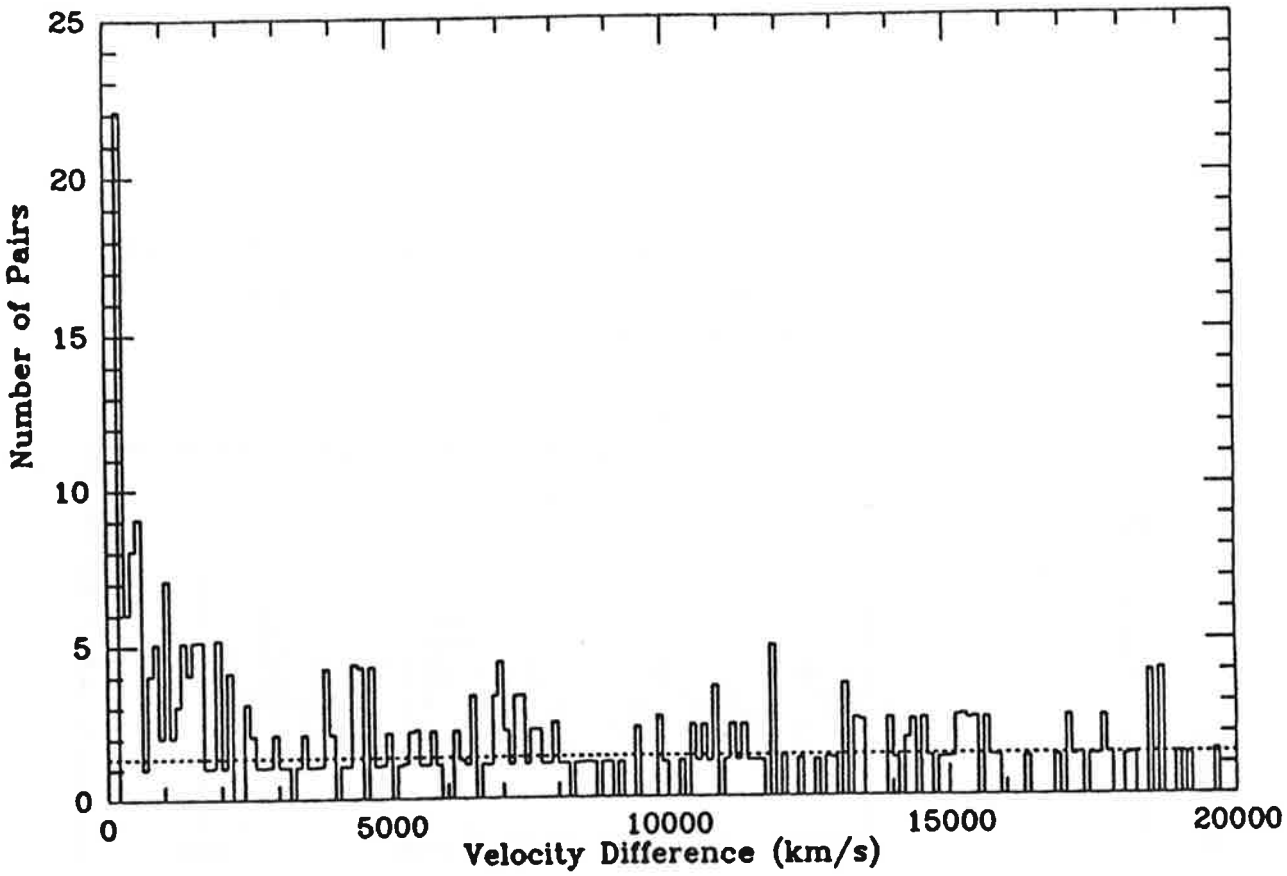


Figure 4.2: Two Point Correlation For C IV Systems
(Reproduced from Sargent *et al.*, 1988).

The figure shows the sum of the correlation functions for 229 C IV systems in a total of 55 quasar spectra (C IV lines are not nearly as numerous as Lyα forest lines), with a velocity splitting bin size of 100 km.s^{-1}. The first two bins have been removed from this plot because of the finite line width effect explained in section 4.3. The vertical axis labelling is not in terms of the correlation function because Sargent *et al.* have used a different normalisation technique (which was not fully explained in the paper). The form of the correlation function is still clear, however.

As a test of the computer program coding, Sargent *et al.*'s result was duplicated in this project, using the original line lists from Sargent *et al.*'s paper. The technique detailed in section 4.3 was used on the C IV systems in each of the 55 quasars and the resulting correlation functions summed in each bin. The bin size used was 100 km.s^{-1}, equal to the bin size in figure 4.2. The result, with 1σ significance level for deviations from a random distribution plotted as a dashed line, is shown in figure 4.3, with the inset showing a magnification of the region below 3000 km.s^{-1} velocity splitting.

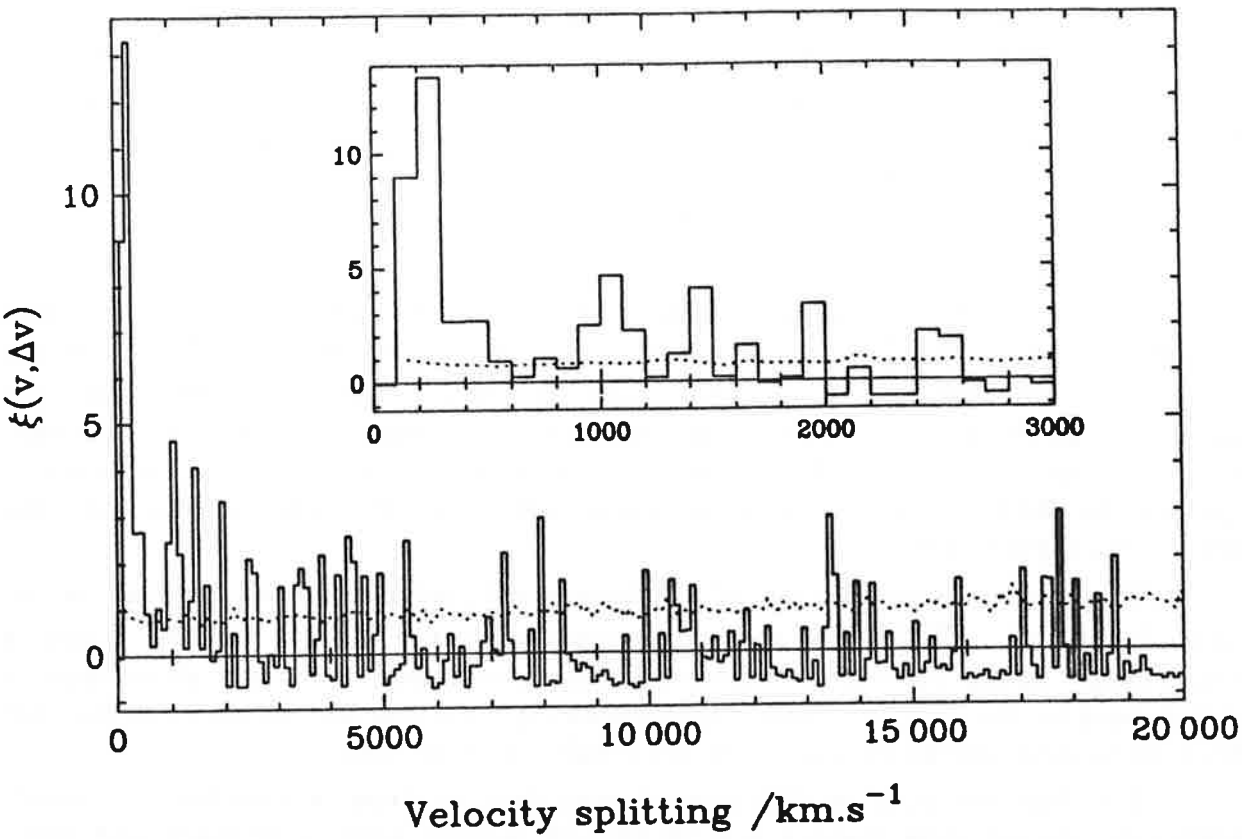


Figure 4.3: Recalculated C IV Correlation Function.

For this calculation, the limit on line proximity before being unresolvably blended was taken to be 0.15 nm. This figure was estimated by finding the closest pairs of resolved lines in the line lists and taking a lower limit. It is greater than the 0.04 nm used in the Lyα calculation because Sargent *et al.*'s spectra were taken at much lower resolution — generally 0.15 nm.

The result arrived at by this recalculation closely resembles that of Sargent *et al.* Minor differences have resulted from the different normalisations. The good agreement with figure 4.2 shows that the program was behaving as expected.

4.5 Discussion.

The C IV result implies that the objects responsible for the metal absorption systems are strongly clustered on velocity scales of 600 km.s^{-1} or less. This corresponds to the scale of clusters of galaxies. It can thus be inferred that the C IV systems arise from absorption in galaxies. Indeed, this is to be expected because galaxies are the homes of the only objects which produce such heavy elements — stars.

By contrast, the Lyα clouds have been reported as having either no significant clustering (Sargent *et al.*, 1980) or only weak clustering (Webb, 1987). Webb has presented data which apparently implies clustering at scales of 200 km.s^{-1} or less at a 3.2σ level in a sample of 9 quasars. Q2206-199 was one of the quasars in Webb's sample and he claimed small velocity splitting correlations at a significance level of 2σ for it. Webb's two point correlation functions for Lyα clouds in the spectrum of Q2206-199 and for his total sample of 9 quasars are shown in figure 4.4.

Figure 4.4(a) is shown with a non-standard normalisation and has 75 km.s^{-1} bins, with the first beginning at 50 km.s^{-1}. Figure 4.4(b) is a standard correlation function, plotted as $1+\xi$, with 60 km.s^{-1} bins, again with the first beginning at 50 km.s^{-1}. Webb's spectra were at resolutions of 0.1—0.2 nm. (*N.B.* Figures 4.4(a) and (b) appear to show significance levels less than those claimed by Webb, but he may have integrated his result over several bins for presentation purposes.)

This project's conclusion of no significant correlations whatsoever in the Lyα forest of Q2206-199 seems contradictory to Webb's result. Such a discrepancy in the correlation functions almost certainly has come about due to differences in the line lists used. Unfortunately, the line list Webb used has not been published and so no checks of his result could be made.

The line list used in this project was derived from a spectrum of much higher resolution than Webb's and so should contain both more lines and more closely spaced lines. From this knowledge, it seems that Webb's result may have suffered from selection effects in the line list. Sargent *et al.* (1980) produced results indicating a complete lack of clustering in the Lyα clouds along the sightlines to 5 quasars. Webb says that there is strong evidence of weak clustering. The result of this project shows that his claims must be investigated more fully.

Webb has also pointed out that the binning process employed in the calculation of the two point correlation function may limit its capability to detect weak correlations at low velocity splittings, due to the loss of information that accompanies data binning. If this is so, then the entire process of searching for correlations must be examined in greater detail and the effort intensified. Of course, no matter what resolution one observes at, there is always the possibility of further populations of weaker, undetected lines. A detailed and comprehensive study of Lyα clouds in many quasars and at high resolutions and signal-to-noise

ratios is needed to resolve the problem of determining the Lyα clouds' clustering properties.

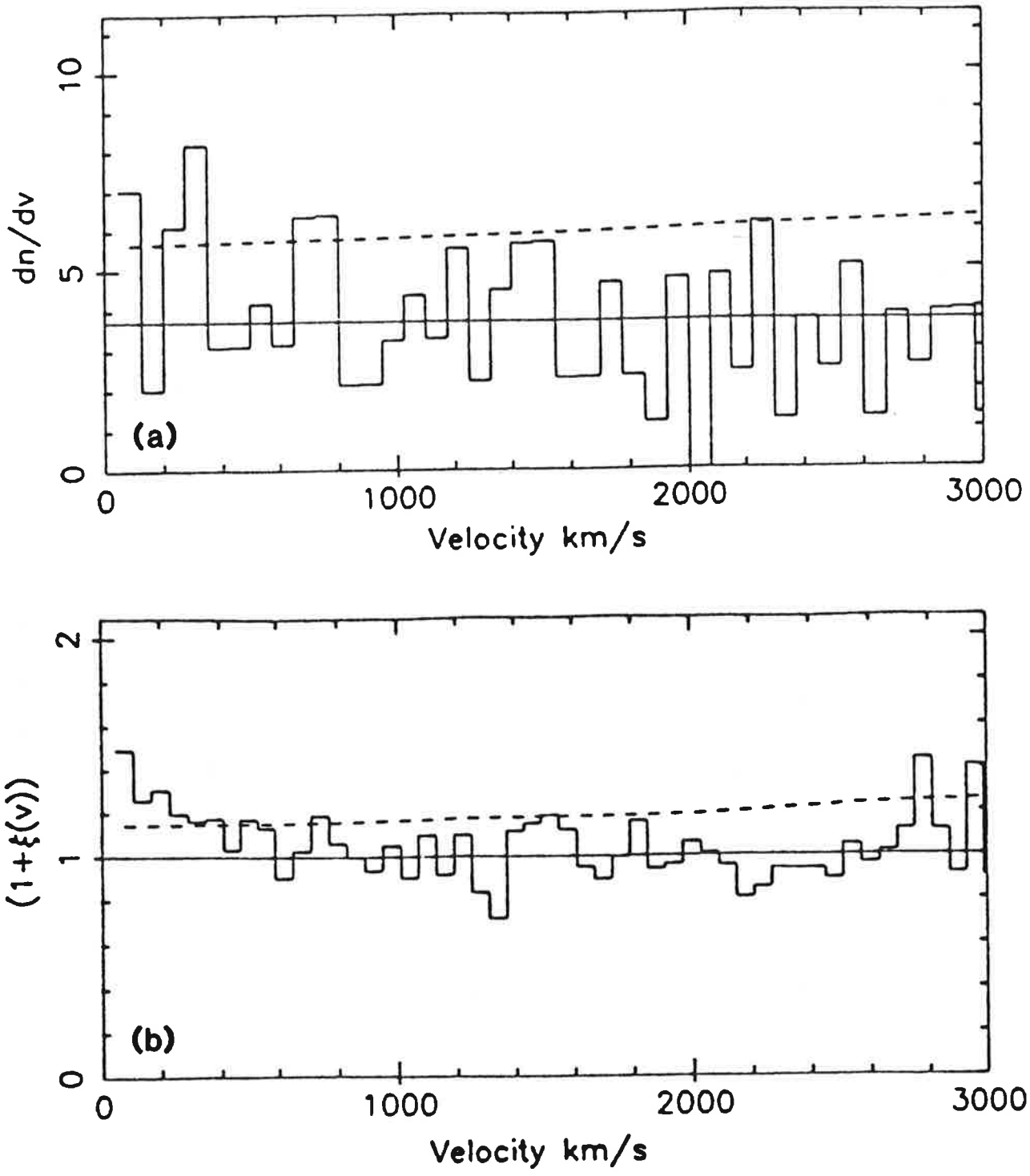


Figure 4.4: (a) Lyman α Two Point Correlation Function for Q2206-199.
(b) Lyman α Two Point Correlation Function for a 9 quasar sample.
(Reproduced from Webb, 1987).

If it is true that the Lyα clouds have no tendency to cluster, then it must be concluded that they have no associations with galaxies at all. A purely random distribution throughout space and the lack of associated heavy element systems implies a population of primordial objects — relics of the early universe. Further study of these objects, including discovering whether or not clustering does exist to some extent, will be helpful in establishing limits to possible models of the evolution of the universe.

5. CONCLUSION

This project was concerned with the analysis of a high resolution spectrum of the quasar Q2206-199.

A procedure for fitting a continuum level to spectral regions of high absorption line density was developed. This relied on photon counting statistics to enable a lower limit to the true continuum to be estimated on many small wavelength intervals. The highest points were chosen and a polynomial was fitted through them to define the continuum level.

A comparison of the normalised order 58 echelle spectrum to a lower resolution spectrum of the same region was made. At smoothings of the echelle spectrum to resolutions between its original 0.008 nm and the intermediate resolution spectrum's 0.15 nm, it was clear that several narrow absorption lines existed in the spectrum but that these would very possibly be missed if one was inspecting only the 0.15 nm data.

The continuum level fitted to the echelle spectrum was seen to be higher than the level set by a previously used method, and certainly higher than any author would have felt it safe to set a level on an intermediate resolution spectrum in a publication. This has led to the discovery of a possible population of very weak absorption lines which, until now, had been taken to be mere noise in the spectrum. More observations at high resolution and subsequent analysis will be required to look further into this possibility.

The very height of the fitted continuum level is also of importance. The result obtained in this project shows that continuum levels in the Lyα forest region of quasar spectra should be set higher than they conventionally are. This is of consequence when estimates are to be made of elemental abundances from line strengths.

The clustering of the Lyα clouds along the sightline to Q2206-199 was investigated by a calculation of the two point correlation function for the clouds. The result showed that there is no tendency for the clouds to cluster, their distribution being consistent with a random distribution along the sightline. This result, although consistent with the results of some other authors, contradicts the conclusions of Webb (1987).

Webb's result may have been caused by selection effects enforced by the intermediate resolution of his spectrum. This is not to say that the results of this project are unquestionably correct. There will always be undetected lines in any spectrum, but this problem can be addressed by taking more data at high resolution and with long integration times to achieve high signal-to-noise ratios.

The results of this project throw light on the problems associated with the interpretation of quasar spectra. Although a possible discovery of many weak absorption lines was made, it has become clear that much more observing time will be required in order to disentangle the full story of the Lyα forest.

6. ACKNOWLEDGEMENTS

The following people deserve eternal gratitude for making this project: (a) possible, (b) easier, (c) more enjoyable.

Richard Hunstead, my supervisor, without whose help and advice you would not be reading this now. He also made it possible for me to visit the Anglo-Australian Telescope, where I did more sight-seeing than observing (but I had fun anyway) and for which I am grateful.

Ann Burgess and Haida Liang, my office mates, who made life in 569 interesting, entertaining and sometimes even informative.

Alan Roy, the "alternative" office mate, who kindly did a last minute proof-read, and who had the nerve to blatantly suggest that I should subtly split an infinitive. (To really annoy Dick.)

Amani Ahmed and George Discala, work experience students, who helped with some really, really tedious and boring stuff. Amani compiled an exhaustive, cross-referenced list of Lyα and metal lines from Linda Smith's data and multiplied 68 wavelengths by 1.0016 when I was checking the coding of my two point correlation function program. George typed every last one of the 229 C IV system redshifts from Sargent *et al.* into computer files, separated by quasar name. If these two don't deserve thanks, then I don't know who does.

Everyone in the Astrophysics Department who helped me in the endless struggle to get the MicroVAX to do what I wanted it to do.

Kim Lester, who made so many comments and suggestions throughout the year that one of them must have been useful somewhere.

Louise Scott, who was the friend in need.

7. APPENDICES

A. The Expansion of the universe.

In order to fathom a concept commonly expressed as vaguely as "the expansion of the universe", it helps to consider a situation such as that shown in figure 8.1. Particles close together are bound to each other by various forces (gravity on a large scale, electromagnetic and nuclear forces on smaller scales), at what we might call "particle distance". Space itself, however, is essentially free to expand with time; so that what could be called a constant "space distance" is continually growing with respect to the "particle distance". Since we usually measure distance by comparison to the size of objects, we see "particle distance" and are oblivious to "space distance". Objects at very large distances feel negligible forces from each other and more readily follow the growth of the "space distance" between them, resulting in an apparent increase in "particle distance". This effect dominates for objects as far from us as galaxies outside our local cluster of galaxies and we interpret the apparent recession of all such objects from us as the expansion of the universe.

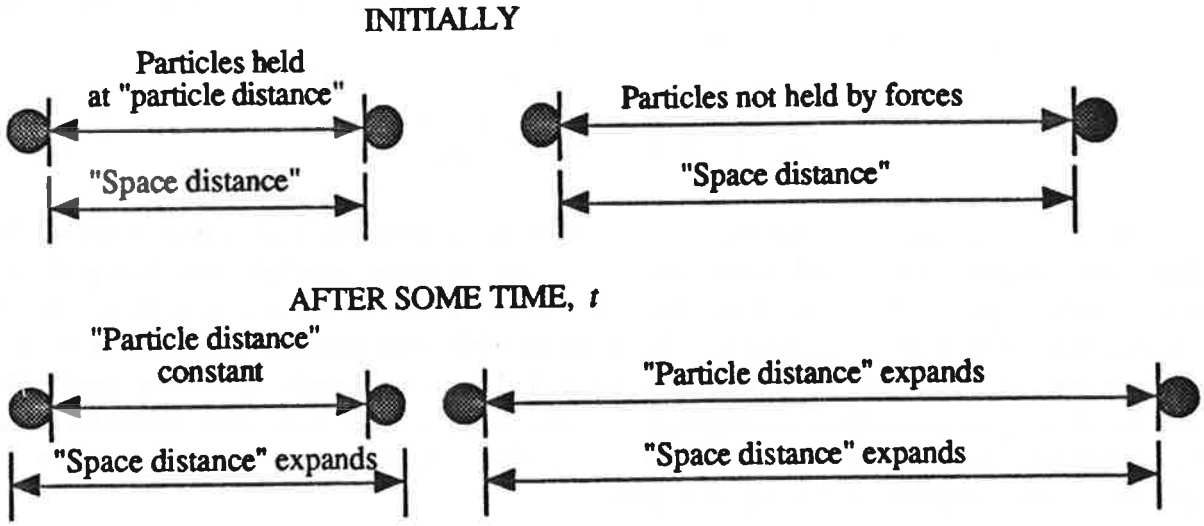


Figure 8.1: The expansion of the universe.

B. Chauvenet's Criterion for the Rejection of Suspicious Data Points.

In any sample of data which should be distributed about some mean, one can calculate the number of standard deviations for which the expectation value for the number of data points at least as far from the mean is 0.5. Chauvenet's criterion states that any data points lying outside this range should be rejected as suspicious deviations form the mean.

For data which follow a normal probability distribution,

$$\Phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \tag{8.1}$$

with mean μ and standard deviation σ, the probability of a data value lying outside the range $\mu - t\sigma$ to $\mu + t\sigma$ for some number t is

$$P(t) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-t}^{t} \exp\left(-\frac{x^2}{2}\right) dx. \tag{8.2}$$

The expectation value for the number of points outside this range is this probability multiplied by the number of data points, N. Setting the expectation value equal to 0.5 gives that the number, t, of standard deviations from the mean outside which points should be rejected satisfies

$$\frac{1}{\sqrt{2\pi}} \int_{0}^{t} \exp\left(-\frac{x^2}{2}\right) dx = \frac{2N-1}{4N}. \tag{8.3}$$

Since equation (8.3) cannot be solved analytically for t as a function of N, values of t were calculated from tables of the normal probability integral for several values of N. A logarithmic function was found to fit the resulting values to better than 4% over a range of N from 10 to 100 data points. At around N=40, which is where the value of N lies for the 0.1 nm wavelength intervals used, the error in this approximation is negligible. The logarithmic function derived from this analysis was used to calculate t in the computer program written to implement the continuum fitting procedure.

Using the value of t thus calculated, all points in the wavelength interval with count levels lying outside the range $\mu - t\sigma$ to $\mu + t\sigma$ were rejected.

C. Derivation of Equation (3.2).

Neglecting the presence of absorption lines, for Poissonian statistics the variance in the object-plus-sky spectrum is $\sigma_{obj+sky}^2=m+s$. A single sky spectrum has a variance $\sigma_{sky}^2=s$.

When three sky spectra at adjacent spatial positions on the detector are averaged, the variance in the resulting spectrum is the original variance divided by 3: $\sigma_{av}^2=s/3$.

The sky-subtracted spectrum has a variance given by the sum of the variances of the object-plus-sky and the averaged sky spectra:

$$\sigma^2 = \sigma_{obj+sky}^2 + \sigma_{av}^2 = m + s + s/3. \qquad (8.4)$$

Equation (3.2) follows directly from equation (8.4).

D. Method of Maximum Likelihood.

The method of maximum likelihood can be used to estimate the most probable value for a statistical parameter from a finite data set. The method involves maximising a quantity known as the *log likelihood* with respect to the parameter being estimated.

In the case of a normal distribution with known mean μ, the log likelihood as a function of the standard deviation σ is

$$\ln L(\sigma) = \sum_{i=1}^{n} \left[\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2 \right]. \qquad (8.5)$$

Differentiating with respect to σ gives

$$\frac{\partial \ln L(\sigma)}{\partial \sigma} = \frac{-n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n} (x_i - \mu)^2 \qquad (8.6)$$

and setting this equal to zero gives equation (3.3).

The variance in the estimate of σ can also be obtained from the log likelihood, being equal to the negative of the inverse of the second derivative of $\ln L(\sigma)$ at the value of σ given by equation (3.3). This evaluates to

$$\text{var}\,\sigma = \frac{1}{2n^2}\sum_{i=1}^{n} (x_i - \mu)^2 \qquad (8.7)$$

and was used in the continuum fitting program to determine the error bars shown in figure 3.2.

8. REFERENCES

Bechtold, J. 1987, in *High Redshift and Primeval Galaxies*, ed. J. Bergeron, D. Kunth, B. Rocca-Volmerange and J. Tran Thanh Van, Editions Frontières, 397.

Harrison, E. R. 1981, *Cosmology*, Cambridge University Press.

Hunstead, R. W. 1987, in *QSO Absorption Lines: Probing the Universe*, ed. J. C. Blades, D. A. Turnshek, C. A. Norman, Cambridge University Press, 71.

Hunstead, R. W., Murdoch, H. S., Peterson, B. A., Blades, J. C., Jauncey, D. L., Wright, A. E., Pettini, M. and Savage, A. 1986, *Ap. J.*, **305**, 496.

Mood, A. M., Grayhill, F. A. and Boes, D. C. 1974, *Introduction to the Theory of Statistics*, McGraw-Hill, 276-286.

Peebles, P. J. E. 1980, *The Large-Scale Structure of the Universe*, Princeton University Press.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. 1986, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press.

Robinson, R. D., Diego, F., Fish, A. C., Lupton, W. F., Pettini, M. and Walker, D. D. 1989, *The UCL Echelle Spectrograph*, Anglo-Australian Observatory.

Sargent, W. L. W., Boksenberg, A. and Steidel, C. C. 1988, *Ap. J. (Suppl.)*, **68**, 539.

Sargent, W. L. W., Young, P. J., Boksenberg, A. and Tytler, D. 1980, *Ap. J. (Suppl.)*, **42**, 41.

Steidel, C. C. and Sargent, W. L. W. 1987, *Ap. J.*, **313**, 171.

Taylor, J. R. 1982, *An Introduction to Error Analysis*, University Science Books.

Webb, J. K. 1987, Ph. D. Thesis, University of Cambridge.

Young, P. J., Sargent, W. L. W., Boksenberg, A., Carswell, R. F. and Whelan, J. A. J. 1979, *Ap. J.*, **229**, 891.