Chapter 7

The Lyman-alpha Forest

Because the line of sight to high-redshift quasars covers a large range in look-back times and several orders of magnitude in the observed column densities, the properties of Ly α forest lines have been studied to see how they change with z and $N_{\rm HI}$ in the hopes of learning more about the distribution of matter and the evolution of structure in the universe. The b values have also been studied because they result directly from the particle motions and physical properties of the absorbing gas. The assumed form of these distributions will be introduced below, before turning to the data themselves.

Given a population of objects with a comoving number density n_o, so that the number density changes with redshift as $n = n_o(1+z)^3$. The number of objects $d\mathcal{N}$ that will be intersected by a line of sight at redshift z is

$$d\mathcal{N} = \sigma(z)\, n(z)\, dl \tag{7.1}$$

where σ is the geometric cross section of each object and dl is the path length in terms of proper displacement, the expression for which depends on the cosmological model (Peebles 1993). Assuming negligible space curvature in a Friedmann cosmology (with density paramer $\Omega_o > 0$ and a zero-valued cosmological constant Λ_o), the number of objects per unit redshift is then

$$\frac{d\mathcal{N}}{dz} = \frac{cn_o\,\sigma(z)}{H_o}\frac{1+z}{\sqrt{1+\Omega_o z}} \tag{7.2}$$

(Rauch 1998). Therefore, if there is no evolution in the number density n_o or cross section σ with redshift, we would expect the number of objects per unit redshift to be

$$\frac{d\mathcal{N}}{dz} \propto \begin{cases} 1+z & \text{for } \Omega_o = 0 \\ (1+z)^{1/2} & \text{for } \Omega_o = 1 \end{cases} \tag{7.3}$$

Traditionally, the redshift distribution of Ly α clouds within a given column-density range has been parametrized by investigators as

$$\frac{d\mathcal{N}}{dz} = A_o(1+z)^\gamma, \tag{7.4}$$

where A_o is a constant. Despite the large spread in the values for γ obtained from different samples of quasar spectra, observations show that the number of absorbers clearly increases more rapidly with redshift than expected for a non-evolving population, which would have $0.5 \leq \gamma \leq 1$.

Similarly, the column-density distribution of Ly α clouds in a given redshift range has been conventionally parametrized by a power law,

$$\frac{d\mathcal{N}}{dN_{\rm HI}} \propto N_{\rm HI}^{-\beta} \tag{7.5}$$

For this form of the equation, the parameter β is taken to be positive to reflect the fact that absorbers with higher column densities are increasingly rare.

7.1 Theoretical models

The usual procedure for determining the values of the parameters in equations (7.4) and (7.5) has been to perform power-law fits to data samples derived from many quasar spectra. But given the poor statistics provided by a single quasar spectrum, no attempt has been made here to do a similar determination from the Ly α forest of HE 0940 alone. Instead, it was thought better to compare the present data to theoretical models constructed from a large observational sample.

Combining equations (7.4) and (7.5) into a single distribution in both column density and redshift yields

$$\frac{\partial^2\mathcal{N}}{\partial z\,\partial N_{\rm HI}} = \mathcal{A}_o\, N_{\rm HI}^{-\beta}\,(1+z)^{\gamma} \tag{7.6}$$

where $\mathcal{A}_o$ is again a constant. Fardal, Giroux and Shull (1998) have recently produced global fits to this distribution for Ly α absorbers. Taking advantage of the better quality, greater redshift coverage and increased overlap in the column-density ranges between the different spectra that have become available in the past few years, the authors have combined the line lists from 7 samples of Ly α forest data with those of 2 studies of Lyman-limit systems. They developed different models for the absorption-line distribution, using maximum-likelihood methods to evaluate three-power-law fits to the data. They also show that the best single-power-law fit describes the data poorly, due to a deficit of clouds with $N_{\rm HI} \sim 10^{15}$ cm^{-2}.

Since one of the purposes of the paper by Fardal, Giroux and Shull was to explore the He II Gunn-Peterson effect, their models A1 and A2 were designed to respectively minimize and maximize the He II line and continuum opacity relative to that of hydrogen. For our purposes, however, it suffices to note that the authors assumed $N_{\rm HI}$ and z to be statistically independent and required that the column-density distributions be continuous. A slight exception was made for their model A3, which is the same as model A2 up to $N_{\rm HI} = 10^{17}$ cm^{-2} but uses the redshift evolution of Stengler-Larrea et al. (1995) beyond that point. This was done because Lyman-limit systems may evolve with redshift at a different rate than the Ly α forest clouds.

The distributions resulting from models A1 to A3 of Fardal, Giroux and Shull will form the basis of all further comparisons to the present data. Table 7.1 summarizes the parameter values for each of these models, and Figure 7.3 shows the shape of their distributions for the redshift range under consideration.

Model	model region	N_{min} (cm^{-2})	N_{max} (cm^{-2})	$\mathcal{A}_o$	β	γ
A1	1	...	1×10^{14}	9.15×10^5	1.40	2.58
	2	1×10^{14}	1×10^{16}	2.52×10^{12}	1.86	
	3	1×10^{16}	1×10^{19}	2.10×10^2	1.23	
	4	1×10^{19}	1×10^{22}	4.42×10^1	1.16	1.30
A2	1	...	2×10^{14}	1.43×10^7	1.49	2.58
	2	2×10^{14}	2×10^{15}	5.41×10^{12}	1.88	
	3	2×10^{15}	1×10^{19}	2.88×10^6	1.47	
	4	1×10^{19}	1×10^{22}	4.42×10^1	1.16	1.30
A3	1	...	2×10^{14}	1.43×10^7	1.49	2.58
	2	2×10^{14}	2×10^{15}	5.41×10^{12}	1.88	
	3	2×10^{15}	1×10^{17}	2.88×10^6	1.47	
	4	1×10^{17}	1×10^{22}	5.00×10^7	1.50	1.50

Table 7.1: *Parameter values in models A1 to A3 of Fardal, Giroux & Shull (1998).*

7.2 Doppler Parameter and Column Density

The Doppler parameters of Ly α absorption lines are studied in the hopes of deducing information about the temperature and motions of the gas in the clouds. However, as mentioned in section 6.6, none of the simplest assumptions seem to apply. Since the relation between the b value and the temperature is complicated and not yet understood, a discussion about the physical properties of the absorbing gas goes far beyond the scope of this work. Nevertheless, a part of the discussion about low b values is relevant to the data here.

Generally speaking, the new high-resolution studies with the Keck 10-meter telescope appear to be establishing that the median Doppler parameters for Ly α lines at $z \sim 3$ are roughly 30 km s^{-1} (e.g., Kim et al. 1997). The distribution can apparently be described as a Gaussian truncated below a cut-off in b to reflect the apparent dearth of lines below $b \sim 20$ (Hu et al. 1995, Kirkman & Tytler 1997).

Figure 7.1 shows a plot of the Doppler parameter vs. column density for all of the lines between 4928.3 Å and 4944 Åin the spectrum toward HE 0940-1050. Of the 19 identified metal lines, only 11 are evident because 8 of them are doublets and thus have the same $N_{\rm HI}$ and b values. The Ly α lines are represented as crosses, and filled squares show the members of identified metal systems. Both symbols together indicate the 5 Ly α lines that are assumed to belong to metal systems, although only the 2 lines with $N_{\rm HI} = 15.2$ and 16.2 could actually be fitted along with their corresponding metal lines. For the purposes of this graph only, it was assumed that in the heavily blended region from 4644 Å to 4664 Å, only the three Ly α lines with the highest column densities were associated with the C IV absorption seen in the 1.52-meter spectrum.

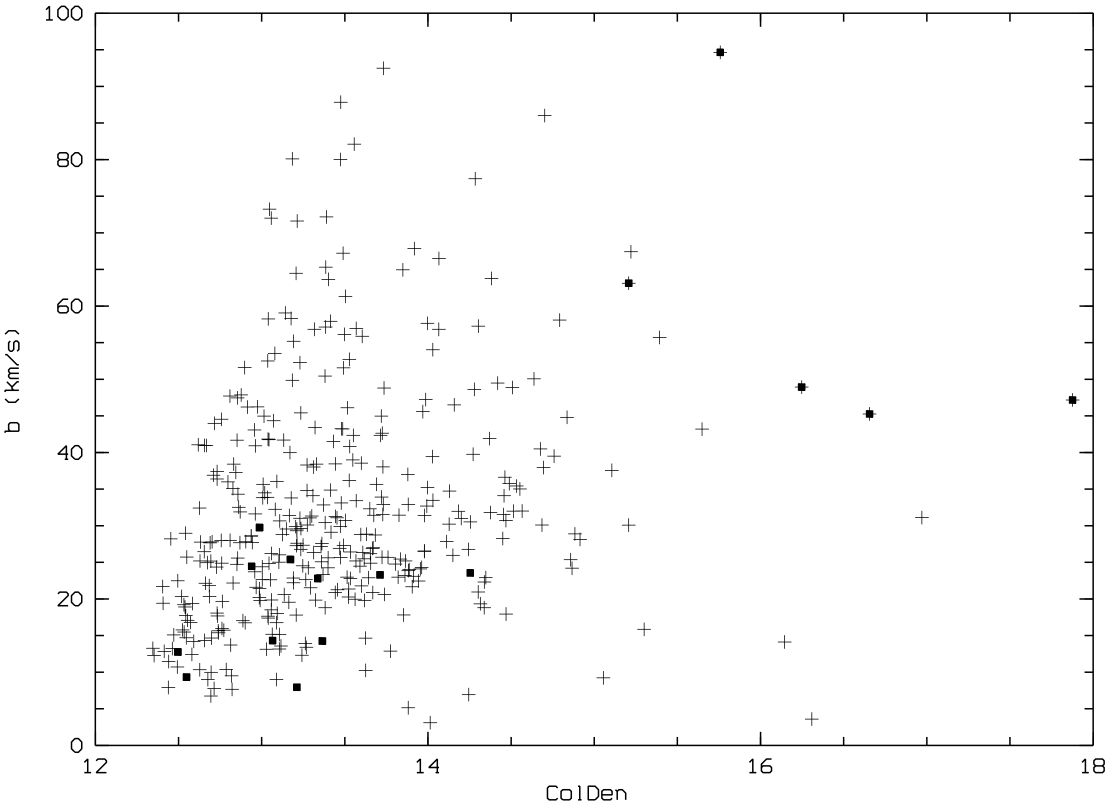

Figure 7.1: *b values vs. column density $N_{\rm HI}$ for all lines between Ly α and Ly β emission.*

Although there is no obvious minimum b cut-off in figure 7.1, this is probably—as is most often the case—due to unidentified metal lines. However, unless or until more metal systems can be identified, it is worth considering whether at least part of the low b values are due to problems with noise.

Pettini et al. (1990) reported a correlation between the b values and the column densities, as well as a high precentage of low $b \lesssim 20$ values; however, both were later shown to be artifacts of the line-finding and fitting procedures. Rauch et al. (1992) performed simulations on echelle spectra with an average FWHM resolution of 23 km s^{-1} (not too different from that of the current spectrum at hand). After fitting simulated spectra with the same programs used on the actual spectrum, they observed the following effects:

1. Broad, shallow lines that went undetected by the line-finding programs.
2. Fits with excessively high b values arising from unresolved blends.
3. Fits with unrealistically low N and b values that were not present in the original sample; these were due to S/N effects.
4. A slight increase in other fitted b values due to blending.

The extremely high b values in the current spectrum are therefore most probably unresolved blends. In some cases, they may be due to the choice of continuum as well; broad shallow lines may be missed by line-finding programs because of noise, but they can also be "created" when the continuum fit is too high. This could be the case, for example, for the line at 4913 Å with $b = 111$.

The current data also seem to confirm that some of the fits of lines with lower column densities tend to have low b values. Among the assumed Ly α lines, 21% have a Doppler parameter of $b < 20$. It is instructive to note that when only lines with $\log N_{\mathrm{HI}} \gtrsim 13$ are considered, the percentage drops to around 10% and stays in that neighborhood for all higher cut-offs that were tried. This is shown, too, in the histogram of the b values in Figure 7.2. The form of the histogram remained basically the same for column densities above $\log N_{\mathrm{HI}} > 13$. This could be because unidentified metal lines will tend to have lower column densities than the average Ly α lines, so that many of them are removed by the cut in N_{HI}. But there also seem to be systematic errors in the profile fitting of lines with smaller column densities. The latter argument would provide an additional reason for the introduction of the completeness limit of $\log N_{min} = 13.33$ discussed in section 7.3.

In addition to the expected errors in the b values of weaker lines, the spectrum was also checked for unrealistically low b values at higher column densities. The empirical relation from Kirkman and Tytler (1997)[1]

$$b_{min} = 14 + 4 \times \log \left[\frac{N_{\mathrm{HI}}}{10^{12.5}} \right] \ \mathrm{km\ s^{-1}} \tag{7.7}$$

was used only as a rough guide, since their data had a much higher resolution and signal-to-noise ratio. Nevertheless, when it was postulated that the b cut-off seen in Keck spectra among lines with $\log N_{\mathrm{HI}} > 12.5$ might only appear in the current spectrum among lines with $\log N_{\mathrm{HI}} \gtrsim 13.3$, only 12 offending lines were found below the empirical limit; and only 7 of them were beyond the range allowed by their errors.

The two lines at 4619.4 Å and 4751 Å may be unidentified Si III lines belonging to the system at $z \approx 2.83$ with unresolved C IV lines (see section 6.7). However, no acceptable fits of the heavily saturated Ly α region could be obtained under this assumption, so the original determinations were left unchanged. The interesting point

[1] Hu et al. (1997) obtain a similar, slightly lower straight-line boundary in their plot of b vs. N_{HI}.

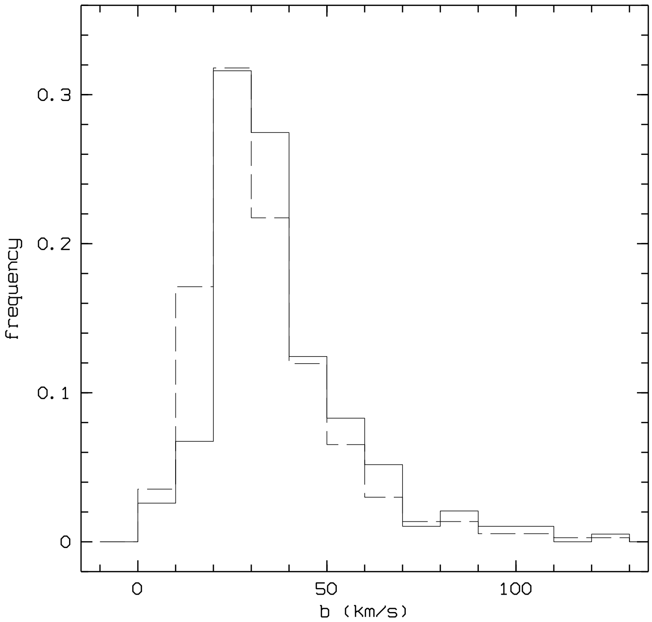

Figure 7.2: *Histogram of b values for the entire Lyα sample (dotted line) and for Lyα lines with* $\log N_{\mathrm{HI}} > 13.3$ *(solid line).*

here is that four of the five remaining lines—and six of the original 12 lines with suspiciously low b values—all come from the same region between 4317 Å and 4331 Å.

It is not clear whether the lines in this region could be part of a some complex metal system, or whether lines the low b values are a result of over-fitting. The standard procedure is to fit each line complex with the very fewest number of components needed to achieve a certain reduced χ^2 or probability of random deviation. Though this criterion was kept in mind while doing the fits with FITLYMAN, no specific limit had been fixed before starting the analysis; so it could be that the method was not followed strictly and that additional components were sometimes added to improve the χ^2 of the fits. Over-fitting could also have resulted from the fact that the initial fit was done by hand with Xvoigt. Because the number of fit components was determined by eye during that manual fitting process, maybe some of the initial fits had already contained too many components to begin with.

An automated line-fitting program using the smallest number of components per fit would have certainly yielded more consistent results. But it would also have been possible to mimic the steps by using FITLYMAN without Xvoigt, following a strict systematic approach in which the initial fit for every region began with a single component, regardless of how many lines it contained. Unfortunately, the fits had not been done using this more time-consuming method and it is not clear how much over-fitting there might have been or how much this would have affected the results. On the other hand, the region from 4317 Å to 4331 Å looked like perhaps it could be simply a noise problem rather than a fitting problem. In any case—whether noise, overfitting or metal lines—it was deemed safer to exclude this particular region from the final sample.

7.3 Column-Density Distribution

In addition to the line blending which affects the distribution of b values, intermediate-resolution spectra (and high-resolution spectra to a lesser extent) will also suffer from the problem of line blanketing. Line blanketing occurs when a weak line remains unnoticed because it lies within the absorption profile of a stronger line. This results in fewer recognized lines with low column densities, and leads to a fall-off on the lower end of the column-density distribution.

The correction for line blanketing—called the completeness correction by Hu et al. (1997)—cannot be done without simulations in which many synthetic spectra with the assumed b and $N_{\rm HI}$ distributions are fitted to see what percentage of lines at certain column densities will go undetected. Without the resources to do such extensive tests (though the process with FITLYMAN is described nicely by Giallongo et al. 1996), the somewhat unfortunate alternative is to simply determine a lower limit above which it is assumed that not many lines are lost to blanketing effects, and to consider only lines above this limit. The resulting sample is then said to be "complete."

The completeness limit for the data at hand was determined with the help of the models of Fardal, Giroux and Shull (Table 7.1). In order to compare actual line counts to the models, a concrete range $[z_{min}, z_{max}]$ and $[N_{min}, N_{max}]$ must be specified, so that equation (7.6) can be integrated to yield the number of absorbers $\mathcal{N}$ expected in that interval:

$$\mathcal{N} = \frac{\mathcal{A}_o}{(1-\beta)(1+\gamma)} \left[z_{max}^{\;1+\gamma} - z_{min}^{\;1+\gamma} \right] \left[N_{max}^{\;1-\beta} - N_{min}^{\;1-\beta} \right] \qquad (7.8)$$

Figure 7.3 shows this comparison for column-density bins of $\Delta N_{\rm HI} = 0.33$, in which the different parameter values for the various power-law regimes had to be taken into account. Since the model parameters do not change with redshift, the entire spectrum between Ly α and Ly β emission could be used for this graph.

Even when the number counts in the data lie above those of the model predictions, the fall-off at $N_{\rm HI} \approx 10^{13}$ is clearly visible. It is also encouraging to see that this is roughly the same column density at which the b distribution assumes the more truncated shape expected from the Keck results. In order to avoid taking the lower limit of the third bin, in which the transition to significant line blanketing seems to occur, a slightly more conservative cut-off of $\log N_{min} = 13.33$ was chosen.

Figure 7.3 also illustrates why the models were scaled before being used as the basis for the Monte Carlo programs described in section 8.5. Especially when the lowest bins are ignored, the predicted counts per bin are clearly lower than the number of lines actually seen in the data. It was therefore decided to scale each model by the amount necessary to make the predictions match the number of actual counts in the column density range $13.33 \leq \log N_{\mathrm{HI}} \leq 16.0$. The upper limit was the one for which the Voigt-profile fits were judged to be reliable.

For the 115 lines observed within this column-density and redshift range, model A1 predicted 87.05 counts and models A2 and A3 predicted 84.57 counts; so in all of the calculations below, the parameter $\mathcal{A}_o$ has been increased by a multiplicative factor of 1.32 in model A1 and a factor of 1.36 in models A2 and A3. This step improved the fit between the models and the data significantly: the χ^2 for model A1 above $\log N_{\mathrm{HI}} \geq 13.33$ dropped from 21.7 to 11.1, from 19.2 to 7.8 for model A2, and from 18.0 to 7.0 for model A3. The scaled fits are shown in Figure 7.4.

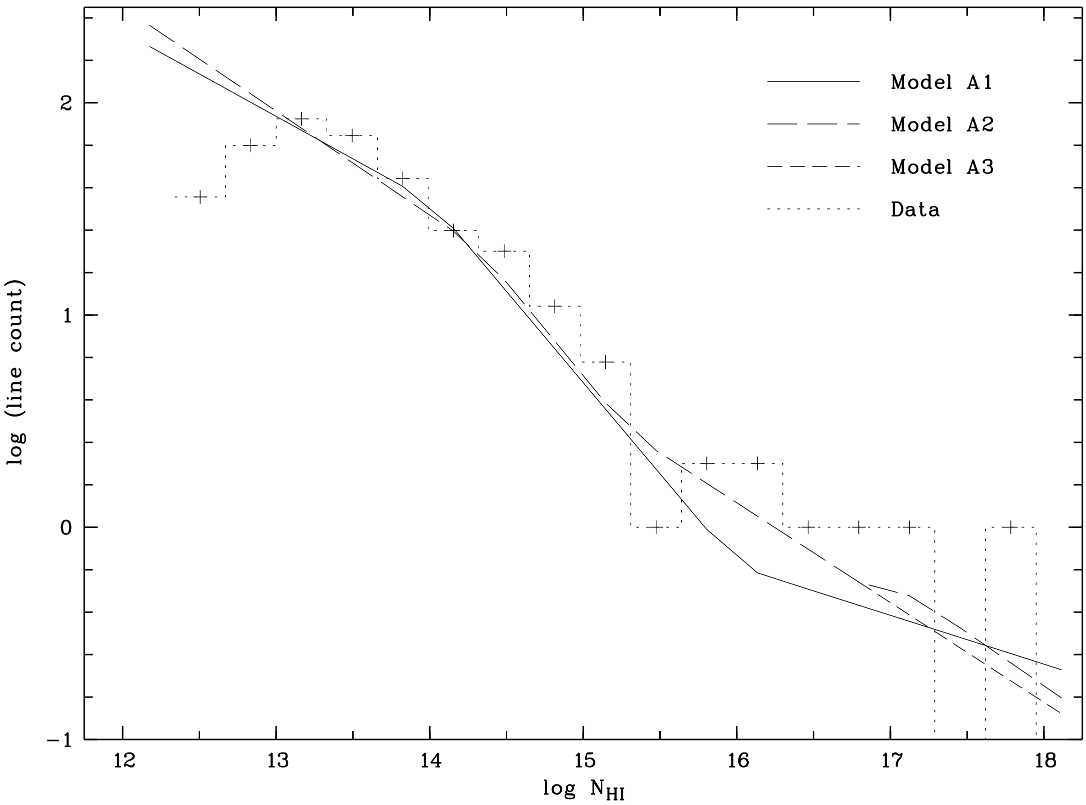

Figure 7.3: *Observed and predicted column-density distributions for Lyα lines, shown with bins of 0.33 in log N_{HI}. Line counts are over the entire redshift interval* $2.420565 < z < 3.054$.

7.4 The Data Sample

To summarize the steps until now: 19 out of the 387 lines in Figure 7.1 had been identified as metal lines and therefore removed from consideration in all subsequent steps. The b and $N_{\rm HI}$ distributions of the remaining 368 Ly α lines were then examined and used to determine a completeness limit of $\log N_{min} = 13.33$, which is noticeably higher than the maximum detection limit of $\log N = 12.5$ for that part of the spectrum. Since lines with low column densities are much more numerous than stronger lines, this last cut-off reduced the available sample to 185 Ly α lines.

Two additional cuts were made at this point. The "noisy" region from 4317 Å to 4331 Å was excluded for reasons discussed in section 7.2. The highly saturated region described in section 6.7 was also excluded, primarily because the heavy line blanketing there made it difficult to determine the position or even the number of lines with certainty.[2] The resulting gaps of $\Delta z = 0.0115$ and $\Delta z = 0.0176$ in the z distribution were taken into account in every step of the analysis described below.

Finally, because the column densities of saturated lines show severe systematic effects in FITLYMAN when not enough continuum is present on both sides, the upper limit for the column density was chosen to be $\log N_{min} = 17.0$. But since the only line with a column density above that had already been excluded as part of the heavily saturated region, this was more of a formal limit. This cut-off did, however, have the advantage of reducing the number of theoretical Ly α forest models that had to be considered. Models A2 and A3 were identical in this range, but because model A3 incorporated the results from Lyman-limit systems and produced a better agreement to the data in section 7.3, these two equivalents will be referred to as model A3 in the discussions below.

It should be noted that in past studies, an effort was made to exclude the Ly α lines belonging to metal systems from the Ly α forest samples used to study the proximity effect. Because metal systems seemed to show different composition and clustering properties, they were assumed to belong to a different parent population of absorbers. However, recent studies with high resolution and signal-to-noise ratios have found associated C IV absorption for all of the high-redshift Ly α clouds with column densities $\log N_{\rm HI} > 15.0$ and for 60% of the clouds with column densities $14.5 < \log N_{\rm HI} < 15.0$ (Cowie et al. 1995, Tytler et al. 1995). Since the detection limit for metal absorption is still signal-to-noise-limited, it has even been argued (Lu et al. 1996 and references therein) that *all* Ly α clouds could conceivably contain metals. In any case, since the distinction seems to be weakening, the Ly α lines corresponding to metal systems identified in the spectrum of HE 0940have been kept in the Ly α sample. And since only two such lines are present, their influence on the results should be negligible anyway.

Now that the final sample of 169 Ly α lines had been selected, their redshift distributions were studied to see whether a proximity effect could indeed be detected in the spectrum of HE 0940.

[2]Slightly different limits of 4642 Å to 4663Å were used in the end, to match an existing bin boundary.

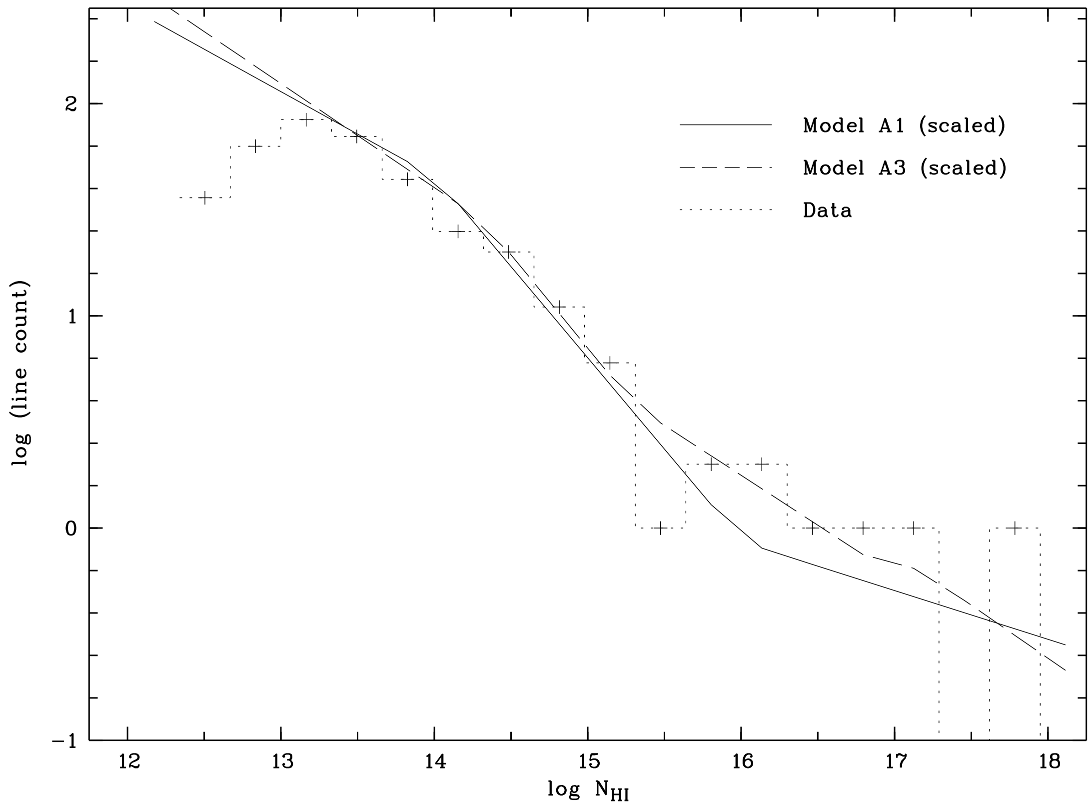

Figure 7.4: *Column-density distribution of models A1 and A3, scaled to the data over the column-density interval* $13.33 < \log N_{\rm HI} < 16.0$ *and the redshift interval* $2.420565 < z < 3.054$. *These scaled models were the basis of the simulations in chapters 8 and 9.*