Stochastic Absorption of the Light of Background Sources due to Intergalactic Neutral Hydrogen

I. Testing different line-number evolution models via the cosmic flux decrement

Thorsten Tepper-García[1]* and Uta Fritze[2]

[1]*Institut für Astrophysik, Georg-August Universität, Friedrich-Hund-Platz 1, D-37077 Göttingen*
[2]*Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK*

arXiv:0705.1242v2 [astro-ph] 9 Nov 2007

ABSTRACT

We test the accuracy of different models of the attenuation of light due to resonant scattering by intergalactic neutral hydrogen by comparing their predictions of the evolution of the mean cosmic flux decrement, D_A, to measurements of this quantity based on observations. To this end, we use data available in the literature and our own measurements of the cosmic flux decrement for 25 quasars in the redshift range $2.71 < z_{em} < 5.41$ taken from the SDSS Data Release 5. In order to perform the measurements of D_A, we fit a power-law to the continuum redward of the Lyα emission line, and extrapolate this fit to region blueward of it, where the flux is severely affected by absorption due to intervening HI absorbers.

We compute, using numerical simulations, the redshift evolution of D_A accounting for the presence of Lyα Forest absorbers and Lyman limit systems randomly distributed along the line-of-sight, and compute its intrinsic scatter at the 1-, 2-, and 3σ level due to fluctuations in the absorber properties (column density, Doppler parameter, redshift) along different lines-of-sight. The numerical simulations consist of Monte Carlo realizations of distributions of the absorber properties constrained from observations.

The results from the models considered here confirm our theoretical expectation that the distribution of D_A at any given redshift be well described by a lognormal distribution function. This implies that the effective optical depth, usually defined as the negative logarithm of the average flux, 1 - D_A, is very accurately Gaussian distributed, in contrast to previous studies. This result is independent to the form of the input distribution functions, and rather insensitive to the presence of high-column density absorbers, such as the Lyman limit systems.

By comparing our and previous measurements of D_A to the outcomes of our simulations, we find an excellent agreement between the observations and the evolution of the *mean* D_A as predicted by one of the models considered in this work. The *observed scatter* in D_A at each redshift, however, cannot be recovered from our simulations. Even though there is evidence for the fact that the lack of agreement between models and observations comes from the combination of heterogeneous measurement sets obtained by different methods, the failure of the models to accurately account for the absorption by intergalactic Lyα absorbing systems and its variation along different lines-of-sight cannot be completely ruled out.

Key words: methods: numerical, intergalactic medium, quasars: absorption lines

1 INTRODUCTION

Since the introduction of the Gunn-Peterson (GP) test by Gunn & Peterson (1965), a detailed knowledge about the physical state of the intergalactic medium (IGM) has been gained from the study of the absorption features identified in the spectra of quasi-stellar objects (QSOs) at restframe wavelengths $\lambda \leqslant 121.5$ nm,

which are now known to be mainly due to resonant scattering by intergalactic neutral hydrogen randomly distributed along the line-of-sight, as first proposed by Lynds (1971). For instance, the null result in the search for a GP trough has been used to rule out the existence of a hot intercloud medium (ICM) (Steidel & Sargent 1987a, Giallongo et al. 1992, 1994), which was thought to confine by pressure the Lyα clouds (Sargent et al. 1980, Ostriker & Ikeuchi 1983). As a result of detailed analyses of the line statistics of the absorbing material, a wealth of information on its clustering (see *e.g.* Ostriker et al. 1988), in particular the existence of voids

---

* Currently at: Institut für Physik, Universität Potsdam, Am Neuen Palais 10, D-14469 Potsdam; E-mail: tepper@astro.physik.uni-potsdam.de

(*e.g.* Pierre et al. 1988, Duncan et al. 1989, Dobrzycki & Bechtold 1991), and the evolution of its number densities, column densities, and Doppler parameter with redshift (*e.g.* Kim et al. 1997) has accumulated over the past years. These results, in combination with the use of state-of-the-art numerical simulations of structure formation based on the currently accepted paradigm of the concordance cosmology (Springel et al. 2005), show that the features seen in absorption against bright background sources arise when the line-of-sight intersects the structures that naturally emerge and evolve with time under the influence of gravitational attraction. Different types of structures such as the filaments present in the *cosmic web*, galactic haloes, and even the discs of primeval galaxies, give rise to distinct absorption features attributed to entities known historically as Lyα Forest clouds, Lyman limit systems (LLSs), and damped Lyα absorbers (DLAs) (see *e.g.* Rauch 1998, Wolfe et al. 2005, for excellent reviews, respectively). Nevertheless, the relation between the observed absorption features and the objects causing them, in particular the correlation between observed damped absorption lines, metal lines (*e.g.* Mg II, O VI) and galaxies–the so-called Absorber-galaxy Connection–is still a matter of debate (see *e.g.* Williams, Shu & Ménard 2005, Part 1). As a consequence of numerous efforts over many years, we now have a better understanding of the origins of the different absorption features observed in QSO spectra. In particular, the notion of discrete, intervening HI absorbing systems randomly distributed along the line-of-sight has been embedded into the more general picture of an evolving *continuous* intergalactic medium with a HI density field that varies in space and time, with its evolution driven mainly by the Hubble expansion, the radiation field of ionising UV sources, and the collapse of structures due to gravity.

1.1 Methods and Input Distributions: A Brief Review

Over decades many people have been working hard towards inferring the physical properties of the intergalactic medium such as its chemical content, density, temperature, etc. (see *e.g.* Kim et al. 2002) by measuring the type of transition, strength, number density, and profiles of absorption lines imprinted in the spectra of QSOs and, more recently, of Gamma-Ray bursts (GRBs) (*e.g.* Lamb & Reichart 2000). There has also been a great effort to quantify the effect of the absorption due to intergalactic neutral hydrogen on the photometric properties of background sources. As a matter of fact, several models have been developed in order to account for this so-called intergalactic attenuation, with different approaches and purposes. Møller & Jakobsen (1990) used Monte Carlo simulations to estimate the amount of absorption at wavelengths shorter than the redshifted HeII λ 30.4 nm line, in order to test the feasibility of the equivalent of the Gunn-Peterson test for intergalactic helium. They found that the absorption as a function of wavelength, averaged over many lines-of-sight, should display together with a characteristic stair-case profile due to the cumulative absorption at the HI resonant wavelengths, an additional characteristic valley-shaped feature (the "Lyman-valley") due to the cumulative effect of the photoionisation of HI by photons with energies $E_\gamma \geqslant hc/\lambda_L$, where $\lambda_L = 91.2$ nm. Later on, in a seminal paper Madau (1995) developed an analytical method to quantify the opacity due to intergalactic HI as a function of redshift, and its effect on the colors of high-redshift galaxies. The underlying assumption of this model is that the observed flux of a source at redshift z is given by the product of the *intrinsic* flux and a transmission factor that accounts for the *mean* absorption as a function of wavelength given in the form $\exp(-\tau_{\rm eff}) \equiv \langle \exp(-\tau) \rangle$ (see

Madau 1995, their equation 3), where the brackets denote the average over an ensemble of random lines-of-sight. The most common application of this model consists in correcting the flux of a synthetic (galaxy-, QSO-) spectrum for intergalactic absorption. This correction is of particular importance at high redshift, where intergalactic HI severely absorbs the light of a background object at restframe wavelengths shorter than 121.6 nm, leading to a substantial reddening of its colour (see *e.g.* Bicker et al. 2004). As the numerous references in the literature attest, the Madau model has become the most widely used attenuation model. However, in a later work Bershady, Charlton & Geoffroy (1999, from now on BCG99) argued that it is not possible to estimate the mean change in the magnitude of a source at a given redshift due to absorption by intergalactic HI along the line-of-sight by multiplying the mean transmission curve of Madau's model with the spectrum of the source and integrating over the corresponding passband, mainly because of the existence of color terms. They suggested that the correct way of accounting for the mean effect of HI absorption on the spectrum of a background source and on its photometric properties is to model first the absorption along many random lines-of-sight, compute the desired photometric quantities for each one of them, and *then* compute the average over the ensemble of lines-of-sight. In other words, they argue that the processes of averaging over many random lines-of-sight and measuring photometric quantities are non-commutative. Indeed, they showed using a Monte Carlo technique that the average magnitudes computed following their approach substantially differ from those computed using Madau's model, even when using the same input distributions for the number of absorbers, their column densities and Doppler parameters. The approach proposed by BCG99 effectively mimics the measurement process that would take place if one would determine *e.g.* the mean observed brightness of a collection of galaxies with different absorber populations along their particular lines-of-sight, but otherwise identical in their intrinsic properties (spectrum, morphology, etc.), and is hence physically meaningful. It turns out that features such as the characteristic stair-case profile and the Lyman-valley cannot possibly be observed in a *single* spectrum, since they arise only by averaging over sufficient numbers of lines-of-sight, a process that has no physical meaning.

In a more recent paper, Meiksin (2006) developed a method to compute the opacity due to intergalactic HI by using hydrodynamical simulations of structure formation in the framework of the concordance ΛCDM cosmology performed by Meiksin & White (2004). Applying their model to compute broad-band magnitudes for different types of objects (*e.g.* starburst galaxies, QSOs of Type I and II), Meiksin (2006) reports differences of $0.5 - 1.0$ mag with respect to Madau (1995)'s model. Despite the different results obtained, this model is similar to Madau (1995)'s model in the sense that it implicitly assumes that the mean opacity of the IGM along a random line-of-sight due to the presence of HI can be accounted for by multiplying a given input spectrum with a mean attenuation curve of the form $\exp(-\tau_{\rm eff})$ and integrating over the corresponding filter function (see Meiksin 2006, their equation 8).

Following BCG99, we state that

$$\int_0^\infty f_\lambda \cdot \langle \exp(-\tau) \rangle \cdot T(\lambda)\, {\rm d}\lambda \neq \left\langle \int_0^\infty f_\lambda \cdot \exp(-\tau) \cdot T(\lambda)\, {\rm d}\lambda \right\rangle, \quad (1)$$

where f_λ is the intrinsic flux, $T(\lambda)$ is the filter transmission function, and the brackets denote the average over all lines-of-sight. We consider that the operation implied by the right-hand side of this expression is the correct way of estimating mean magnitudes of background objects including the effect of the absorption due

to intergalactic HI. This approach is of course not restricted to the computation of mean magnitudes and colors, and can be applied to the estimate of the mean of any photometric quantity. Furthermore, it is also possible to determine not only the mean, but in principle any desired confidence interval around the mean, *e.g.* $\pm \sigma$ range, via the computation of quantiles (see Section 5).

It should be clear that not only the method, but also the input physics is an (even more) crucial ingredient of a particular model that accounts for the intergalactic attenuation, as already shown by BCG99. It is, however, not trivial to test whether using a particular method and a set of input distributions accurately describes the observed effect of the absorption by intergalactic HI on the spectra of background sources. For example, the evolutionary synthesis models of Bicker et al. (2004) that include the correction for intergalactic absorption based on Madau's model match quite well the observations of galaxies in the Hubble Deep Field (Bicker et al. 2004, their Figure 12), since the magnitude differences reported by BCG99 (their Figure 7) for one of their models with respect to Madau's model are in this case of the order of the scatter of the observations around the predicted colors. In other words, even though Madau's and BCG99's approaches are fundamentally different, it is difficult to test the accuracy of their predictions on the basis of a comparison to *e.g.* observed galaxy colors. A quantity that is more sensitive to the absorption due to intergalactic HI is the mean cosmic flux decrement D_A (cf. Section 2). The reason for this is that the restframe wavelength range over which this quantity is measured is typically 10 nm wide, and is hence narrower than typical broadband filters. We thus consider as a primer test that any model that accounts for the absorption due to intergalactic HI should reproduce first of all the observations of this quantity. Hence, through the comparison of their respective predictions to measurements of D_A it should be possible to discriminate between different models which may or may not be appropriate to account for the effect of the intergalactic attenuation on the spectra of background sources. We select for this comparison two of several models introduced by BCG99 to compute the magnitude changes of high-redshift galaxies due to intergalactic absorption, and which are based, respectively, on the input distribution functions of Kim et al. (1997), and the input physics of the Madau (1995) approach. The latter is chosen since, as already mentioned, it is the model most widely used in the literature; the former constitutes, to the best of our knowledge, the most complete set of input distribution functions for the evolution of the Lyα absorbers to date, later expanded but not significantly changed by Kim et al. (2002).

To sum up: The main goal of this work is to model and analyse the redshift evolution of D_A for different evolution scenarios of the intergalactic neutral hydrogen, conveniently parametrized by input distribution functions of the form of equation (4, Section 2). Hence, the models we consider here differ only by the set of input distribution functions used, but they all equal in method, *i.e.* all of them take advantage of the Monte Carlo technique. We judge the goodness of a particular model by its power to reproduce the observations of D_A in a wide redshift range, and analyse to which extent the theoretical expectations about the properties of D_A are recovered from the measurements.

This work is organised as follows: In Section 2 we briefly recall the concept of the cosmic flux decrement and discuss some issues related to its measurement. In Section 3 we present two different types of models for the intergalactic attenuation, which we use to model the redshift evolution of the cosmic flux decrement.

In Section 4, we describe our measurements of this quantity for a sample of SDSS QSO spectra. Finally, we compare these and previous measurements to the outcomes of each model, and discuss the results of this comparison as well as some other implications of the models for the evolution of D_A in Section 5.

2 THE COSMIC FLUX DECREMENT REVISITED

Before high-resolution (*i.e.* $\Delta\lambda \lesssim 1$ nm), high S/N observation became feasible, the basic spectroscopic technique used to analyse the effect of the absorption due to intergalactic neutral hydrogen on the spectra of background sources was to measure the mean depression of the observed flux relative to the unabsorbed flux –or emission continuum–, a quantity which became to be known as cosmic flux decrement. This quantity, first introduced by Oke & Korycansky (1982), can be defined as a function of redshift by

$$D_A(z) \equiv \frac{1}{\Delta\lambda} \int_{\lambda_1\cdot(1+z)}^{\lambda_2\cdot(1+z)} \left(1 - \frac{f_{obs}(\lambda)}{f_c(\lambda)}\right) d\lambda, \qquad (2)$$

where f_c and f_{obs} are the continuum and the observed fluxes, respectively, and $\Delta\lambda \equiv (1+z)\cdot(\lambda_2 - \lambda_1)$. Formally, the integral is computed in the restframe wavelength range [102.5, 121.6] nm, *i.e.* between the Lyβ and the Lyα emission lines. However, the actual estimate of D_A is usually performed between the restframe wavelengths $\lambda = 105$ nm and $\lambda = 117$ nm –or in an even narrower wavelength interval– in order to avoid contamination by the emission wings of the Lyβ + Ovi and Lyα lines, respectively.

Since D_A is extremely sensitive to f_c, as can be easily seen from the definition (2), an accurate measurement of this quantity demands a reliable estimate of the underlying continuum. Unfortunately, there is no consensus of what the best method to estimate the continuum may be. A popular choice, mainly because of the presence of emission lines redward of the Lyα forest region, consists in fitting a local continuum, most commonly using cubic splines (see *e.g.* Lu et al. 1996) or b-spline functions (see *e.g.* Kirkman et al. 2003, Tytler et al. 2004a,b), searching for regions apparently free of absorption blueward of the Lyα emission line. Other authors prefer to fit a continuum in the region redward of the red wing of the Lyα emission line, and extrapolate it to the region blueward of it (Steidel & Sargent 1987b, Schneider et al. 1989, Cristiani et al. 1993). A widely adopted form for the fitted continuum in this case is a power-law with spectral index α_v which for wavelengths $\lambda > 121.6$ nm takes on values in the range [0.28, 0.99] (Steidel & Sargent 1987b, Vanden Berk et al. 2001, and references therein). The latter method may tend to place the intrinsic continuum level higher than it actually is, thus overestimating the measured values of D_A (cf. Sec. 4.1; see also Tytler et al. 2004a, and references therein); for the former method the opposite is true, in general (see *e.g.* Faucher-Giguere et al. 2007). For either method, there is an uncertainty in the estimate of the continuum, and this is the main drawback of the mean flux depression as a technique to estimate the mean absorption due to neutral hydrogen present in the intergalactic medium (IGM). In an attempt to overcome this problem, Zuo & Lu (1993) used the idea that the cosmic flux decrement effectively measures the *total* equivalent width of all Lyα absorption lines in the chosen wavelength range –if corrected for the contribution of metal absorption lines– to measure this quantity by simply adding up the equivalent widths of lines identified as Lyα absorption lines. Of course, the reliability of this technique highly depends on the correct identification of lines, and misidentification

may lead to larger uncertainties than those associated to the continuum estimate.

It turns out that reliable measurements of D_A are very useful to constrain estimates of fundamental cosmological parameters such as the mean baryon density Ω_b, the UV background intensity (Rauch et al. 1997), the normalization of the power spectrum σ_8, the vacuum-energy density Ω_Λ, and the Hubble parameter H_0 (Tytler et al. 2004a). It is hence of great interest to contrast observations to theoretical models of the evolution of D_A, in which the bias due to the uncertainty in the estimate of the continuum is absent. This has been previously done by different workers (see *e.g.* Giallongo, Gratton & Trevese 1990, Cristiani et al. 1993, Madau 1995), usually obtaining a good agreement with observations. However, there is still a scatter in the observations of this quantity for which it has not been accounted yet in any modelling so far. It is possible that this scatter may well be due to the difference in the methods used by each group to measure D_A and to the different redshift ranges probed. Hence, by taking advantage of the Monte Carlo technique, we assess to which extent the *observed* scatter can be ascribed to the *intrinsic* scatter in D_A due to fluctuations in the properties (number density, column density, Doppler parameter) of the H I absorbers along different lines-of-sight (see Section 3.3).

From the theoretical point of view, assuming that the restframe equivalent width of the absorbers does not evolve with redshift, and that the number density of the absorbing systems evolves like $\propto (1 + z)^\gamma$, it is expected that D_A should evolve with z like

$$D_A(z) \propto (1 + z)^{1+\gamma} \tag{3}$$

where the extra factor $(1 + z)$ comes from the scaling of the equivalent width, as pointed out by Jenkins & Ostriker (1991). Indeed, it has been found empirically that the redshift evolution of D_A can be described by a power law $D_A(z) = A \cdot (1 + z)^{1+\gamma}$ with $A = 6.2 \cdot 10^{-3}$ and $\gamma = 1.75$ (Kirkman et al. 2005), even though other functional forms, *e.g.* an exponential of the form $D_A(z) = D_A^0 \cdot e^{\alpha \cdot (1+z)}$ with $D_A^0 = 0.01$ and $\alpha = 0.75$ (Zhang et al. 1997) also match well the observations. More recently, Kirkman et al. (2007) showed that the observed evolution of D_A with redshift in the range $0 < z < 3.2$ is well described by a broken power-law, even though the significance of the fit is low. In any case, expressions like these are only valid up to a given redshift, since they diverge for $z \to \infty$, whilst D_A converges asymptotically to 1 in this limit or, more precisely, when z approaches the redshift z_{reion} at which reionisation sets on. As will be shown later, the redshift evolution of D_A predicted by our simulations does satisfy this asymptotic behaviour (cf. Section 5). Furthermore, even if the power-law form for the evolution of D_A holds, the index γ in equation (3) should be replaced by $\overline{\gamma}$, where the latter index accounts for the averaged evolution of absorbers of different column densities, which evolve all at different rates. Conversely, estimates of a single γ from D_A measurements assuming a power-law of the form of equation (3), as done by O'Brien et al. (1988), may give a hint on the population of absorbers *dominating* the behaviour of D_A, comparing the estimated γ with the power-law index of the different populations. For instance, according to one of the models considered here, the redshift evolution of D_A will be shown to be dominated by absorbers with column densities $N_{HI} < 10^{17}$ cm^{-2}.

A compilation of D_A measurements, accumulated in the literature over the past twenty years approximately, and which includes our own measurements that extend the redshift range to $z_{em} = 5.41$,

Table 1. Types of absorbers and their corresponding parameters adopted from Madau (1995, their equation 10). Note, however, that they quote $N_{HI} = 2.0 \cdot 10^{12}$ as lowest column density, while we use $N_{HI} = 1.0 \cdot 10^{12}$, in order for the adopted normalisation to be consistent.

N_{HI} [cm^{-2}]	$\mathcal{N}_0$	γ	β
$10^{12} - 1.59 \cdot 10^{17}$	$2.40 \cdot 10^7$	2.46	1.50
$1.59 \cdot 10^{17} - 10^{20}$	$1.90 \cdot 10^8$	0.68	1.50

is shown in Figure 6 (cf. Section 4 and cited references for details on the measurements in this figure).

3   MODELLING THE INTERGALACTIC ATTENUATION

Since the observation of individual sources (galaxies, QSOs, GRBs) necessarily implies observations along different lines-of-sight, it is expected that the stochastic nature of the distribution of the Lyα absorbers, especially of those with the highest column densities, causes a scatter in the observed absorption, even for sources with identical intrinsic spectra. Hence, depending on the absorption along a particular line-of-sight, one would expect different observed values for each measurement of any photometric quantity, for example, the cosmic flux decrement D_A. Performing enough measurements of such a quantity for sources with –ideally– identical SEDs at a fixed redshift, one could in principle estimate its mean and its scatter due to stochastic effects in the absorption by neutral hydrogen in the IGM.

The numerical realisation of this thought experiment is best achieved through Monte Carlo simulations. Following *e.g.* Møller & Jakobsen (1990), Giallongo et al. (1990), Cristiani et al. (1993), and BCG99, we generate thousands ($4 \cdot 10^3$) of lines-of-sight each with a random population of H I absorbers, and compute the absorption along each of them for a given input spectrum at a fixed redshift. The population of each line-of-sight consists of a random number N_{abs} of absorbing systems, each of them characterized by three parameters: its redshift z_{abs}, its column density N_{HI}, and its Doppler parameter $b \equiv \sqrt{2kT/m_H}$, where k is the Boltzmann constant, T is the kinetic temperature of the gas and m_H is the mass of the hydrogen atom.

The redshift and column density characterising each absorber are drawn from a distribution of the form

$$f(N_{HI}, z) = \mathcal{N}_0 \cdot (1 + z)^\gamma \cdot N_{HI}^{-\beta}, \tag{4}$$

where $\mathcal{N}_0$ is a normalization constant. This function defines the 1-dimensional distribution of the H I present in the IGM probed by a random line-of-sight. The number N_{abs} of systems for each line-of-sight is drawn from a Poisson distribution with parameter

$$\langle N_{abs} \rangle = \int_{I_z} \int_{I_{N_{HI}}} f(N_{HI}, z) \, d N_{HI} \, d z, \tag{5}$$

where the integral is carried out over appropriate redshift- and column density intervals I_z and $I_{N_{HI}}$, respectively.
We use different sets of input distributions that include the evolution of both low- and high density absorbers, and that give rise to the following models:

MMC This model relies on the input distributions from Madau (1995, their equation 10) listed in Table 1. Here, the Doppler parameter is kept constant at a value $b = 35.0$ km s^{-1}, which roughly

Table 2. Types of absorbers and their corresponding parameters, adopted from BCG99, their equation 10.

N_{HI} [cm^{-2}]	$\mathcal{N}_0$	γ	β
$10^{12} - 10^{14}$	$3.14 \cdot 10^7$	1.29	1.46
$10^{14} - 1.59 \cdot 10^{17}$	$1.70 \cdot 10^6$	3.10	1.46
$1.59 \cdot 10^{17} - 10^{20}$	$1.90 \cdot 10^8$	0.68	1.50

matches the mean b-value derived by Rauch et al. (1992). This model corresponds to one of the several models presented by BCG99, and is referred to as the *MC-NH* method in their work.

BMC This model matches the best method of BCG99 named *MC-Kim*. The corresponding parameters for the line-density evolution and column density distribution functions are summarized in Table 2. In this model, in contrast to the MMC model, the Doppler parameter for each absorber is drawn from a truncated, redshift-dependent Gaussian distribution of the form

$$P(b) \equiv \Theta(b - b_{tr}) \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(b - \mu)^2\right),$$

where $\Theta(x)$ is the Heaviside function:

$$\Theta(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geqslant 0 \end{cases}$$

and the mean, standard deviation, and truncation value as a function of redshift are given by $\mu(z) = -3.85\,z + 38.9$, $\sigma(z) = -3.85\,z + 20.9$ and $b_{tr}(z) = -6.73\,z + 39.5$, respectively. BCG99 originally used this model to analyse the impact of the intergalactic attenuation on high-redshift galaxy colors in the range $1.75 < z < 5.0$, but we use it in the extended range $0.2 < z < 5.41$. Since our highest redshift limit is not too far away from BCG99's, we may safely assume that the model is valid in our extended redshift range.

MMC without Lyman limit systems In order to asses the impact of the Lyman limit systems on the intergalactic absorption, we introduce this model, which consists of the same input distributions as the MMC model, excluding the systems with column densities $N_{HI} > 1.59 \cdot 10^{17}$ cm^{-2}.

In all the models describe above, the attenuation factor for each absorber is given by $\exp[-\tau(\lambda)]$, where, for the general case, the absorption coefficient $\tau(\lambda)$ can be written as

$$\tau(\lambda) = \tau_{LL}(\lambda) + \sum_{i=2}^{N_{trans}} \tau_i(\lambda). \tag{6}$$

The first term on the right–hand side is the opacity due to the ionisation of neutral hydrogen by photons with wavelengths $\lambda \leqslant \lambda_{LL} \equiv 91.18$ nm. It is given by

$$\tau_{LL}(\lambda) = N_{HI} \cdot \Theta(\lambda_{LL} - \lambda) \cdot g(\lambda) \cdot \sigma_\infty \cdot \left(\frac{\lambda}{\lambda_{LL}}\right)^3, \tag{7}$$

where $\sigma_\infty \equiv 6.3 \cdot 10^{-18}$ cm^2 is the HI photoionisation cross-section, and g is the Gaunt–factor for bound–free transitions[1]. The second term is the sum of the opacities due to resonant scattering at each

transition of the Lyman series[2]. In general, the absorption coefficient for the transition $1 \rightarrow i$ is

$$\tau_i(\lambda) = N_{HI} \cdot \sigma_i \cdot \phi(a_i, x). \tag{8}$$

The cross-section σ_i is a function of the Doppler parameter b, the oscillator strength of the transition f_i, and the resonant wavelength λ_i, and is given by

$$\sigma_i = \frac{\sqrt{\pi}\, e^2}{m_e c^2} \frac{\lambda_i^2}{\Delta\lambda_D}\, f_i, \tag{9}$$

where $\Delta\lambda_D = \lambda_i\, b/c$ is the Doppler broadening, and the variable $x \equiv (\lambda - \lambda_i)/\Delta\lambda_D$ is the distance to the line center in Doppler units. We assume the profile function ϕ of the absorption line to be given by the Voigt-Hjerting function

$$H(a_i, x) \equiv \frac{a_i}{\pi} \int_{-\infty}^{+\infty} \frac{e^{-y^2}}{(x - y)^2 + a_i^2}\, dy. \tag{10}$$

Here, $a_i \equiv \lambda_i^2\, \Gamma_i/(4\pi\Delta\lambda_D)$ is the relative strength of the natural broadening to Doppler broadening for the ith transition, and $y \equiv v/b$ is the kinetic velocity in units of the Doppler parameter. In this work, we neglect the opacity due to the photoionisation term and consider only the first resonant transition, *i.e.* the Lyα transition, since this is the only one of interest in the wavelength range studied here. Furthermore, we use the approximation to H for values of a and column densities characteristic for intergalactic HI of Tepper-García (2006).

3.1 The transmission factor Φ

The cumulative absorption along a random line-of-sight of the flux f_{em} of a source at redshift z_{em} is calculated according to expression

$$f_{obs}(\lambda_{obs}) = \frac{f_{em}(\lambda_{em})}{1 + z_{em}} \cdot \Phi(\lambda_{obs}), \tag{11}$$

where λ_{obs} and λ_{em} are the observed and the emitted wavelengths, respectively. These are related by $\lambda_{obs} = \lambda_{em} \cdot (1 + z_{em})$. The quantity Φ is the transmission factor and is given by

$$\Phi(\lambda) \equiv \prod_{i=1}^{N_{abs}} \exp[-\tau_2(\lambda/(1 + z_i)] = \exp\left[-\sum_{i=1}^{N_{abs}} \tau_2(\lambda/(1 + z_i)\right], \tag{12}$$

where τ is given by equation (8), and z_i is the redshift at the epoch of absorption[3].

Introducing the relation $f_c(\lambda) = f_{em}[\lambda/(1 + z)]/(1 + z)$, it follows from equations (2), (11), and (12) that

$$1 - D_A(z) = \frac{1}{\Delta\lambda} \int_{\lambda_1 \cdot (1+z)}^{\lambda_2 \cdot (1+z)} \Phi(\lambda)\, d\lambda. \tag{13}$$

The right–hand side of this expression is just wavelength-averaged value of Φ at redshift z, and we will denote it by $\overline{\Phi}_z$. Note, however, that this quantity still depends on redshift, as indicated by the subscript. Since $D_A(z)$ and $\overline{\Phi}_z$ differ only by a constant factor, they may be considered as equivalent with respect to their statistics, which will be discussed in the next section.

[1] An extensive tabulation of values for the Gaunt–factor can be found in Karzas (1961).

[2] We adopt the convention that the Lyα transition (from the ground state to the next higher energy level) be identified with i = 2, the Lyβ transition with i = 3, etc. The photoionisation cross-section is thus consistently denoted by σ_∞.

[3] The reader shall bear in mind that N_{HI}, and σ_i and a_i–through the dependence on the Doppler parameter–in equation (8) are different in general for each absorber. However, we do not write this explicitly by *e.g.* introducing a new subscript in order to avoid a cumbersome notation.

3.2 Distribution of D_A

Judging from the dependence of Φ on τ and N_{abs} (see equation 12), it is expected that the transmission factor and consequently D_A are rather complicated random variables. Nevertheless, as is well known from statistics, a random variable **x** that can be expressed as the product of a large number of small, statistically independent factors is distributed lognormally, *i.e.* according to the distribution

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}\,x}\,e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2}, \tag{14}$$

where μ and σ are the mean and the standard deviation of $\ln$ **x**. This expression is equivalent to the statement that a random variable **x** is distributed lognormally, if and only if its logarithm is distributed normally.

The expected value μ' and the standard deviation σ' of a lognormal distributed quantity can be expressed in terms of the parameters μ and σ as

$$\mu' = e^{\mu + \frac{1}{2}\sigma^2}, \tag{15}$$

and

$$\sigma' = \left(e^{\sigma^2} - 1\right)^{1/2}\mu'. \tag{16}$$

From the form of equation (12) we may suspect that the transmission factor is a lognormally distributed variable, since it can be expressed as the product of a large number of statistically independent factors that take on values in the range [0, 1]. The implications of this statement are profound: If Φ is distributed lognormally, so does $\overline{\Phi}_z$ and consequently $D_A(z)$. Furthermore, due to the property of the lognormal distribution stated above, the effective optical depth of the Lyα absorption, usually defined as $\tau_{\rm eff} \equiv -\ln(1 - D_A)$ (see *e.g.* Kim et al. 2001) should obey a Gaussian distribution. This result follows independently from the fact that the total optical depth can be expressed as the sum of the independent contribution of each system, as indicated in equation (12). Thus, for a sufficiently large number of absorbers N_{abs}, and if the optical depth for each absorber has the same mean value $\langle\tau\rangle$ and dispersion $\sigma(\tau)$ at each wavelength (*i.e.* redshift), $\tau_{\rm eff}$ should obey a Gaussian distribution at each redshift, centered at $N_{abs}\langle\tau\rangle$ and with a dispersion $\sqrt{N_{abs}}\,\sigma(\tau)$. These statements are completely independent of the form of evolution of the intergalactic neutral hydrogen, as long as the transmission factor can be expressed in the form of equation (12).

In order to test whether the values of D_A obey a lognormal distribution, we compute D_A using the MMC and BMC models for an ensemble of $N_{\rm LOS} = 4\cdot10^3$ lines-of-sight at a given redshift, and from these values we compute the mean and standard deviation of $\ln D_A$ according to the equations

$$\mu(\ln D_A) \equiv \frac{1}{N_{\rm LOS}}\sum_{i=1}^{N_{\rm LOS}}\ln D_A^i, \tag{17}$$

and

$$\sigma^2(\ln D_A) \equiv \frac{1}{N_{\rm LOS}-1}\sum_{i=1}^{N_{\rm LOS}}\left(\ln D_A^i - \langle\ln D_A\rangle\right)^2, \tag{18}$$

With these parameters, we generate for each redshift a set of normally distributed random numbers using a standard random-number generator, and compare them statistically to the $\ln D_A$ values from our simulations at each given redshift, in order to determine whether the latter are normally distributed. Recall that $\ln D_A$ is distributed normally if and only if D_A obeys a lognormal probability distribution. As a first approach, we choose to compare both
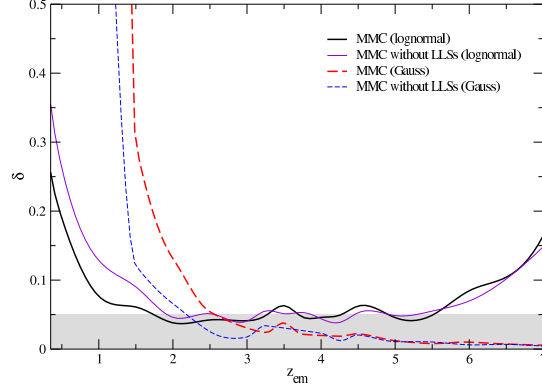

Figure 1. Values of δ for the MMC model assuming a lognormal (heavy solid line) and Gaussian (heavy dashed line) parent distribution. The shaded area indicates the range $\delta \leqslant 0.05$, that corresponds to a probability of $1.9\cdot10^{-5}$ that the data are not drawn from the same parent distribution. Note how the assumption of a Gaussian parent distribution improves with redshift, but that a lognormal distribution is a better assumption for low redshifts. The light solid and light dashed lines correspond to the MMC model without Lyman limit systems (see Section 3.2.1). Note that the approximation to a lognormal distribution, is better when these systems are included (see text for discussion). The fluctuations seen with respect to a perfect smooth curve in each case are due to the random nature of the process at each redshift, and are not significant.

data sets via a χ^2-test similar method, and to this end we compute for each set of values the 50 per cent quantile (median), and the $\pm34.13, \pm43.32, \pm47.72, \pm49.38$, and ±49.87 per cent quantiles around the median. We define the following measure

$$\delta^2 \equiv \sum_{i=1}^{N_q}\left(\frac{Q_i - q_i}{Q_i}\right)^2, \tag{19}$$

where the Q's are the quantiles determined from the set of random numbers distributed normally, the q's are the corresponding quantiles of the $\ln D_A$ values at a given redshift that result from our simulations, and $N_q = 11$ is the number of quantiles. By definition, δ explicitly gives the absolute deviation of one data set as characterised by these 11 quantiles with respect to the other data set, and thus quantifies the departure of the assumed distribution. Indeed, the smaller the value of δ, the larger the probability that both data sets belong to the same parent distribution. The values of δ computed in this way turn out to lie in the range $(5\cdot10^{-2}, 5\cdot10^{-1})$, which may seem vanishingly small. However, since the distribution of the random variable δ itself is unknown, we determine the significance of the results by generating two sets of normally distributed random numbers with mean and standard deviations distributed uniformly, and comparing them to the values found for our simulated data. We do this for $1.2\cdot10^6$ pairs of sets, and determine from these what is the fraction of realisations with a value of δ smaller than a given value. According to these estimates, there is a probability of $\{1.4\cdot10^{-2}, 3.1\cdot10^{-3}, 1.9\cdot10^{-4}, 2.3\cdot10^{-5}\}$ that a value $\delta \leqslant \{0.50, 0.25, 0.10, 0.05\}$ will be produced by chance, respectively. In Figure 1 we show the values for δ as a function of redshift for the full MMC model, assuming a lognormal parent distribution for D_A (heavy solid line). The result are qualitatively the same for the BMC model and are not shown for clarity. Note
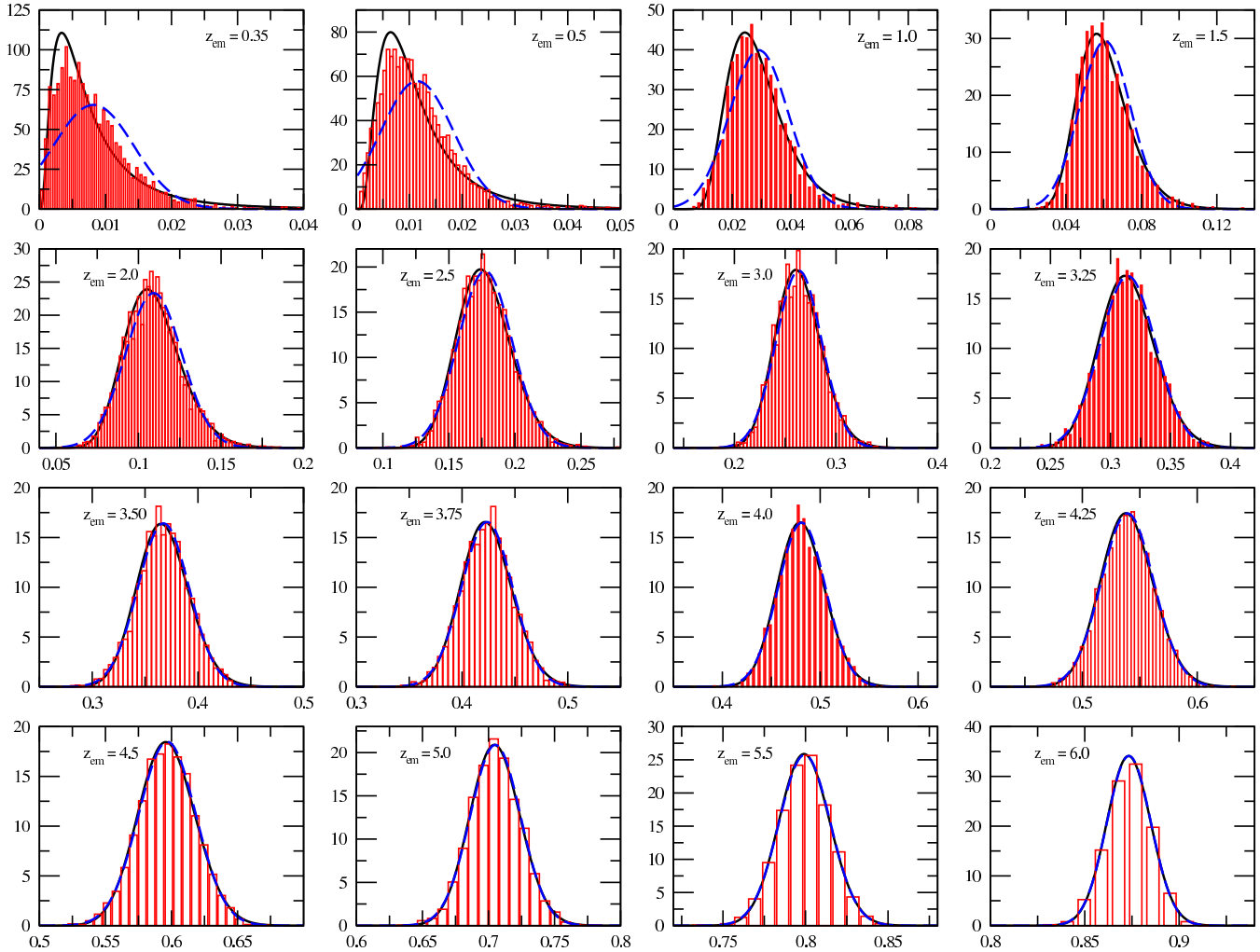
Figure 2. Distribution of the D_A values at a given redshift computed from simulations based on the MMC model (histogram). The distribution is normalised to unit area. The bin size used to compute the histogram at each redshift has been arbitrarily chosen to be given by $max\{D_A(z)\}/100$, and is hence different at each redshift. The binning of the data, *i.e.* the D_A values at each redshift obtained from the simulations, is intended only for display purposes. Shown are also the analytic lognormal distribution (solid line) computed according to equations (14), (17), and (18), and the analytic Gaussian distribution (dashed line) computed using the same equations as before but replacing $\ln D_A$ by D_A. The y-axis indicates the probability of the corresponding D_A value on the x-axis. Note that the binned data have not been used in any form to estimate the parameters of the corresponding probability distribution. Hence, the curves shown are not fits to the binned data, but are computed using only the mean and standard deviation of the raw data. Note the excellent agreement between the data and the analytic lognormal distribution, especially with respect to the skewness and the cut-off at $D_A = 0$. The distribution of the data is not well described by a Gaussian distribution at the lowest redshifts shown. However, the approach of the data distribution to a Gaussian distribution increases with increasing redshift, as expected (see text for details).

that for the whole redshift range shown, $\delta \lesssim 0.5$, and in particular $\delta \lesssim 0.1$ for $z_{em} > 1.5$.

It follows from the Central Limit Theorem that the larger the number of absorbers N_{abs}, the closer the approach of the distribution of Φ to a lognormal distribution, since the distribution of $\ln \Phi$ approaches a Gaussian. Given that N_{abs} increases with redshift, it should be expected that the accuracy with which the distribution of Φ approaches a lognormal distribution also increases with redshift. Furthermore, if the integral in equation (13) is approximated by a sum of the form

$$\overline{\Phi}_z \approx \frac{1}{N} \sum_{i=1}^{N} \Phi_i \Delta \lambda_i,$$

where N is the number of pixels, it is apparent that for a sufficiently large N, and assuming that Φ has the same mean value $\langle \Phi \rangle$ and dis-

persion $\sigma(\Phi)$ at each pixel, the distribution of $\overline{\Phi}_z$ should approach a Gaussian distribution with mean $\langle \Phi \rangle$ and dispersion $\sigma(\Phi)/N$. Since the width of a given restframe wavelength range, *i.e.* the number of pixels also increases with redshift as $(1 + z)$, it is expected that the approximation of the distribution of $\overline{\Phi}_z$ and hence of D_A to a Gaussian becomes better with increasing redshift. We thus have the superposition of two effects: On the one hand, the distribution of Φ at a fixed wavelength approaches a lognormal distribution at any given redshift, with the accuracy increasing with redshift. On the other hand, the distribution of $\overline{\Phi}_z$ and $D_A(z)$ approaches a Gaussian distribution with increasing redshift. The net result should be that $\overline{\Phi}_z$ and D_A are distributed lognormally at low redshifts, and that their distribution approaches a Gaussian for higher redshifts. Since D_A asymptotically converges to unity for very high redshifts, its distribution at these redshifts is expected to be highly peaked

around the mean value. This is naturally given by the fact that the dispersion of $\overline{\Phi}_z$ scales like $\sigma(\Phi)/N$ around $\langle\Phi\rangle$, and that N increases with redshift, as stated above.

In order to test this expectation, we compute the analytical lognormal (Gaussian) probability distribution of D_A as given by equation (14) with the mean and standard deviation of $\ln D_A$ (D_A) computed using equations (17) and (18) from the ensemble of values that result from our simulations at each redshift. The comparison of the analytic distribution thus obtained to the distribution of simulated D_A values is shown for the full MMC model in Figure 2. The results for the BMC model are qualitatively the same and are hence not shown. Note the excellent agreement at all redshifts between the lognormal probability distribution and the distribution of the D_A values resulting from our simulations. We want to emphasize that the curves shown are not fits to the binned data, but are computed using only the mean and standard deviation of the unbinned data. At lower redshift, the agreement at the lower cut-off, *i.e.* at $D_A = 0$ is worth mentioning. It is remarkable that this cut-off, which is physically given by the fact that absorption as measured by D_A cannot take on values smaller than zero, arises in a natural way due solely to the fact that D_A is distributed lognormally. Note, in contrast, that a Gaussian distribution does not satisfactorily describes the distribution of the data at these low redshifts; in particular, the Gaussian distribution does not display the sharp cut-off at $D_A = 0$. Nevertheless, the description of the data by a Gaussian distribution becomes better with increasing redshift, as expected, and also the corresponding Gaussian distribution becomes narrower at every increasing redshift (taking into account the change in the scale of the *x*-axis from the first to the last panel). Moreover, note that the analytic lognormal and Gaussian distribution become visually indistinguishable from each other at $z_{em} > 3.0$. In order to assess quantitatively the differences between these distributions with respect to the distribution of the data, we compute again the value of δ at each redshift, assuming now that the D_A values obtained from the MMC model are drawn from a Gaussian parent distribution. We compare these values to the corresponding values computed before for an assumed lognormal parent distribution. This comparison is shown in Figure 1. As can be seen, the values of δ for both distributions are low at all redshifts, and the difference between them at $z_{em} \gtrsim 2$ is negligible small. This explains why the lognormal and Gaussian distribution functions shown in Figure 2 are practically indistinguishable from each other. However, note that the values of δ for a assumed Gaussian parent distribution become vanishingly small with increasing redshift, eventually becoming smaller than the corresponding values for the lognormal distribution. Besides, it can be see also that the δ values for a assumed lognormal distribution rise again towards high redshifts, implying that the distribution of the data at these redshifts is no longer well described by a lognormal distribution. This confirms our statement made above, that D_A is expected to be distributed lognormally at low redshifts and normally at higher redshifts. The redshift at which the transition from a lognormal to a Gaussian distribution takes place may depend on the particular set of input distributions used. For redshifts where the distribution of D_A is well approximated by a lognormal distribution, the optical depth should be distributed normally. This result is however not in agreement with the results from *e.g.* Madau (1995, his Figure 1), Meiksin & White (2004, their Figure B1), or Bernardi et al. (2003, their Appendix C), where the distribution of the mean flux, *i.e.* the averaged transmission factor $\overline{\Phi}_z$ is claimed to be highly inconsistent with a Gaussian distribution. It is usually assumed that the presence of the high-density Lyman limit systems
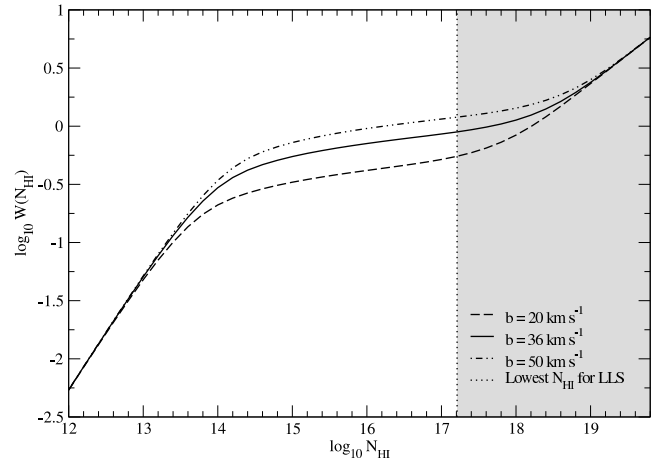

Figure 3. Curve of growth of the Lyα absorption line, for three typical values of the Doppler parameter. The shaded region corresponds to the column densities characteristic to Lyman limit systems.

is responsible for this behaviour, and we next make use of our simulations to address this question.

3.2.1 The Effect of the Lyman limit systems (LLSs)

In order to quantify the effect that the LLSs have on the distribution of D_A, we compare, following the analysis of Section 3.2, the results of the simulations for the MMC model with and without the optically thick LLSs. We find that D_A is lognormally distributed as well with a high confidence for the case where the LLSs are excluded, as can in Figure 1, where the values for δ computed at each given redshift are shown. Again, the results for the BMC model are qualitatively the same and are omitted here. Note that the values of δ for the MMC model without LLSs are larger at redshifts $z_{em} \lesssim 3$ than those for the full MMC model. This is due to the fact that the absence of the optically thick LLSs enhances the probability of the attenuation factor $\exp(-\tau)$ to be closer to unity at a given wavelength along a random line-of-sight. This means effectively that the number of factors in equation (12) are reduced. Hence, when the LLSs are absent, the approach to a lognormal distribution should be worse with respect to the case where the LLSs are included. Furthermore, this effect should be enhanced towards lower redshifts, for which the number of factors, *i.e.* of absorbers decreases as $\propto (1+z)^\gamma$. Note that the trend is the opposite in the case of a assumed Gaussian parent distribution.

We hence find when computing the evolution of D_A with the MMC model with and without LLSs that the predictions for the evolution of D_A with redshift for the full MMC model are practically indistinguishable from the results of the MMC model without LLSs, which implies that the effect of these systems on the total absorption is small. Furthermore, it turns out that the distribution of D_A is lognormal irrespective of the presence of LLSs. All these results point to the fact that the LLS have a negligible impact on the evolution of D_A, as long as the input distributions used here correctly describe the number density evolution of the absorbers. This would confirm similar results already found by Desjacques et al. (2007). Also, as pointed out by McDonald et al. (2005), the picture would not be fundamentally different if higher column density systems such as damped Lyα systems were present. We will come back to this point in Section

5.

We think that, as long as the distribution functions realistically describe the evolution of the Lyα absorbers, the LLSs *cannot* truly have a great impact neither on the absorption as measured by D_A, nor on its statistics because of the following reasons: 1) They are scarce, and even more compared to the thinner Lyα forest systems (*e.g.* of the order of 5 LLSs on average along a random line-of-sight compared to approximately 10^3 Lyα forest absorbers out to $z = 3.0$, according to the MMC model); 2) Their contribution to the absorption is due solely to the absorption at the resonant wavelength of the Lyα line, and the equivalent width of the Lyα absorption line arising in systems with column densities $N_{HI} = 10^{17.21-18}$ cm^{-2} is not too different from that of the systems with column densities $N_{HI} = 10^{14-15}$ cm^{-2} (cf. the curve-of-growth in Figure 3), which are by far more numerous. In order to estimate the relative contribution of each population at a given redshift, we weight the equivalent width $W(N, b)$ of the Lyα absorption line as a function of the column density N and Doppler parameter b with the column density distribution, and compute the ratio

$$\varrho(z_{em}; b) \equiv \frac{\langle W \rangle_{Ly\alpha}}{\langle W \rangle_{LLS}}, \qquad (20)$$

where

$$\langle W \rangle_i \equiv \int_0^{z_{em}} \int_{N_{min}}^{N_{max}} (1 + z) \cdot W_0(N, b) f_i(N, z) \, dN \, dz, \qquad (21)$$

Here, f_i is the distribution function of population $i \in \{Ly\alpha, LLS\}$, $W_0(N, b)$ is the rest equivalent width of a Lyα absorption line for a column density N and Doppler parameter b, and N_{min} and N_{max} are the column density limits that define each population, respectively. For a reasonable value for the Doppler parameter of 36 km s^{-1}, and the input distributions of the MMC model, we find that $\varrho >> 1$ at all redshifts. The result is qualitatively the same for the BMC model. Hence, It follows from these models that the Lyα forest systems dominate the absorption over the optically thicker Lyman limit systems at all epochs. This would explain at least in part why the difference between the predictions for D_A from the MMC model with and without LLSs, is small. Also, it is consistent with the theoretical expectation that the distribution of D_A should not be far from lognormal or Gaussian, with or without LLSs, since this only depends on the fact that the absorption factor be expressed in the form of equation (12), and this is truly independent of the form of the input distributions, as stated previously.

3.3 Scatter in D_A

We expect the intrinsic scatter in the absorption due to cosmic variance to be strongest at lines-of-sight of middle length. At low redshifts, both the number of thin Lyα forest clouds and thick Lyman limit systems is small, and the addition of a few more does not change dramatically the amount of absorption. However, the mean number of Lyα forest clouds and LLS increases with z, and thus the probability of encountering more or less systems than average increases as well. Correspondingly, the absorption increases and so does its scatter. At even higher redshifts, the number of Lyα forest systems increases so dramatically and the absorption is so severe that the addition of more systems does not make any difference neither to the absorption nor to the scatter. Thus, we should expect the stochastic effect, *i.e.* the scatter in absorption, to peak at some intermediate redshift z_{int}.

We compute the intrinsic scatter at the σ level for the values
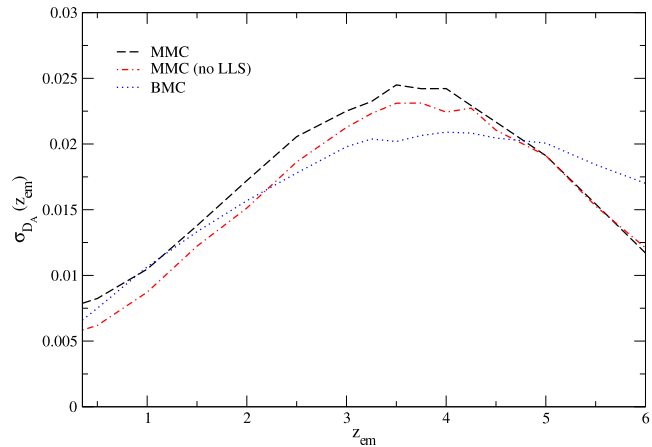

Figure 4. Evolution of the intrinsic scatter of D_A due to the stochastic nature of the absorption in the intergalactic medium for the two competing models MMC (dashed line) and BMC (dotted line). The dot-dashed line corresponds to the MMC without Lyman limit systems (see Section 3.2.1 for details). Note that the behaviour of $\sigma(D_A)$ is qualitatively the same for all three models. As expected, the amount of scatter for the MMC model is less in the case where the Lyman limit systems are absent.

of D_A at each given redshift obtained from our simulations using equations (15), (16), (17), and (18). The result is shown in Figure 4, where we show the evolution of the scatter in D_A with redshift for the MMC model with and without LLSs. We also include for comparison the result from the BMC model. It can be seen that, irrespective of the model, the scatter peaks at a intermediate redshift between $z_{em} \approx 3.5$ and $z_{em} \approx 4.0$. Note that the peak is significant, since it represents an increase in the scatter of 2.5 times with respect to its value at $z_{em} \approx 1$. It is interesting that this result had also been found by Zuo (1993, their Figure 2), who using a semi-analytic approach and different input distributions, reported that $\sigma(D_A)$ is largest at redshifts near 3.7. Thus, the qualitative behaviour of the intrinsic scatter of D_A shown in Figure 4 may be an unavoidable feature of this observable, which could explain at least in part the large scatter in the measurements of D_A seen in the same redshift interval (cf. Figure 6). As will be shown in Section 5, however, there is a disagreement at the 3σ level between the *amplitude* of the scatter in the observations at these redshifts and that predicted by the models.

When comparing models that only differ by the presence of the optically thick Lyman limit systems, we find that the scatter in D_A is larger at any given redshift when the LLS are present, as expected. None the less, the absolute value of the scatter does not differ significantly between the situation where these systems are present and where they are absent. Thus, the net effect of the LLSs is to enlarge the intrinsic scatter in the absorption, contributing only marginally to the mean value of the absorption itself. However, note that for redshifts $z_{em} \gtrsim 4.0$, the amount of scatter at the 1σ level becomes indistinguishable between both scenarios, owing to the fact that the thinner Lyα forest systems are overwhelmingly more numerous at these redshifts.

The results stated above are consistent with the previous conclusion obtained from our simulations that the optically thick LLSs cannot have a great impact on the absorption as measured by D_A and on its scatter.

4   OUR D_A MEASUREMENTS

We want to compare the predictions for the evolution of D_A that result from the models described above to observations. For this purpose, we use previous measurements of D_A reported in the literature, and we perform ourselves new measurements of this quantity, using QSO spectra from the SDSS Data Release 5 (Adelman-McCarthy et al. 2007), thus extending the redshift range of the measurements to $z_{em} = 5.41$.

   Our selection procedure of sources suitable for this purpose was as follows: Since the wavelength range available from SDSS (DR5) is $\lambda \in [380, 920]$ nm, and we measure the continuum depression D_A in the restframe wavelength interval $\lambda \in [105, 117]$ nm, the redshift of our sample is restricted to $z_{min} \geqslant 380/105 - 1 = 2.62$. We choose $z_{min} = 2.7$ as our lowest redshift in order to avoid the low S/N at the blue end of the spectrograph. A simple query for high-z quasi-stellar objects on the SDSS SkyServer Spectroscopic Query Form with this restriction alone returns around 2400 spectra. From this first selection, we rejected those objects for which the redshift was either not measured, the measurement had failed, or the measured photometric and spectroscopic redshifts were inconsistent with each other. We binned the quasars in redshift intervals of $\Delta z = 0.1$ and selected for each redshift bin the spectrum with the highest S/N, leaving us with 28 sources, from which we removed three further objects due to low data quality. The resulting sample is listed in Table 3.

4.1   Continuum fit

The continuum of a quasar is often assumed to be of the form $f_\nu = f_\nu^0 \cdot \nu^{-\alpha_\nu}$ (see *e.g.* Steidel 1987b, Laor et al. 1997), or equivalently, $f_\lambda = f_\lambda^0 \cdot \lambda^{-\alpha_\lambda}$, where both indices are related by $\alpha_\lambda = 2 - \alpha_\nu$, and $f_\lambda^0 = f_\nu^0 \cdot c^{1-\alpha_\nu}$, and c is the speed of light. Empirical evidence in favour of the assumption of an underlying continuum in form of a power-law has been given by *e.g.* O'Brien et al. (1988), who derived an average power-law index of $\alpha_\nu = 2.36$ for $\lambda \in [80, 121.6)$ nm and $\alpha_\nu = 0.67$ for $\lambda \in [121.6, 190]$ nm. This is consistent with the result already found by Neugebauer et al. (1979) that the spectral index α_λ varies over large wavelength ranges. Zheng et al. (1997) found by constructing a composite QSO spectrum from 284 HST FOS spectra that a single power-law describes well the continuum for wavelengths between 105 and 220 nm with $\alpha_\nu = 0.99 \pm 0.05$, but that the continuum steepens (flattens in λ) significantly for $\lambda \leqslant 105$ nm ($\alpha_\nu = 1.96 \pm 0.15$). Similarly, Telfer et al. (2002) reported, using a sample nearly twice as large as the Zheng et al.'s sample, that the continuum in the extreme UV region between $\lambda = 50$ nm and $\lambda = 120$ nm and in the near UV between $\lambda = 120$ nm and $\lambda = 300$ nm is well described by a single power-law with $\alpha_\nu = 1.76 \pm 0.12$ and $\alpha_\nu = 0.69 \pm 0.06$, respectively. Furthermore, by constructing a composite QSO spectrum from a homogeneous sample of over 2200 SDSS QSOs, Vanden Berk et al. (2001) concluded that the continuum in the restframe wavelength range $121.6 < \lambda < 500$ nm can be very well modeled by a single power-law either in wavelength or frequency, with $\alpha_\nu = 0.44 \pm 0.1$. However, they did not estimate a spectral index for wavelengths $\lambda < 121.6$ nm. While it is the consensus that the optical/NIR continuum of quasar spectra is due to synchrotron emission (Whiting et al. 2001), the physical origin of the quasar UV/optical continuum is not yet established. On the hand hand, the observed soft X-ray/UV/optical spectral shape of quasars is found to be consistent with free-free emission of optically thin gas at temperatures $10^5 - 10^6$ K (Barvainis 1993), which is well described

by a power-law. On the other hand, it has been argued that the emission from a geometrically thin, optically thick accretion disk also reproduces well the observed quasar SED in this spectral region (Kawaguchi et al. 2001). Either model, however, is consistent with an underlying continuum in the form of a power-law. Alternatively, as already mentioned in Section 2, other authors prefer to fit quasar continua locally using spline functions (see *e.g.* Lu et al. 1996, Kirkman et al. 2003, Tytler et al. 2004a,b) searching for regions apparently free of absorption blueward of the Lyα emission line.

   Since the Lyα forest region is severely absorbed due to intervening HI systems, especially for high-z QSOs, and a local fit to the continuum in this region is difficult, with the uncertainty increasing with redshift (see *e.g.* Faucher-Giguere et al. 2007), we choose to estimate the continuum of our selected sources in the Lyα forest region by fitting a power-law to the QSO spectrum redward of the Lyα emission line and extrapolating it for $\lambda \leqslant 121.567$ nm. As long as the assumption of the underlying power-law holds, this approach has the advantage that the continuum estimate is completely independent of the spectral resolution and S/N in the Lyα forest region. Nevertheless, this method may lead to an systematic overestimate of the continuum flux and thus to an corresponding overestimate of D_A. It follows from the results quoted above that the quasar's intrinsic continuum in spectral regions below and above the Lyα emission line may both be well described by power-laws, but with potentially different spectral indices. More specifically, the spectrum in the region below 121.6 nm (rest-frame) may be steeper in frequency, and hence flatter in wavelength[4], with respect to the continuum in the region above 121.6 nm (see *e.g.* Telfer et al. 2002, their Figure 4). Thus, fitting a continuum in the form of a power-law at wavelengths $\lambda > 121.6$ nm and extrapolating it towards $\lambda < 121.6$ nm may well lead to an overestimate of the continuum level in that region, as first pointed out by Seljak et al. (2003), according to whom the D_A may be overestimated by at least 0.05 at $z_{em} = 2.72$. We will take this systematic bias into account when estimating the uncertainty in our D_A measurements (cf. Section 4.2).

   The power-law that we fit to each spectrum is of the form

$$f_c(\lambda; \alpha_\lambda) = f_\lambda^0 \cdot (\lambda + \lambda_0)^{-\alpha_\lambda} . \tag{22}$$

where the flux amplitude f_λ^0, the wavelength off-set λ_0, and the spectral index α_λ are the parameters to be determined. We fit the continuum in the wavelength range $[130 \cdot (1 + z_{em}), 900]$ nm, in order to avoid the red emission wing of the Lyα line and the red end of the spectrograph, respectively. Note that the available wavelength range decreases with redshift, and this may introduce an increasing uncertainty in the fitted continuum. The fit parameters and their uncertainties were obtained with help of the IDL task CURVEFIT. It turns out that the continuum fit is rather insensitive to the uncertainties in the flux amplitude and the wavelength off-set, and extremely sensitive to the uncertainty in the spectral index. Because of this and for simplicity, in our further analysis we neglect the error in the first two parameters and consider only the uncertainty in the spectral index to be of relevance[5]. An example of a QSO spectrum and its corresponding fit are shown in Figure 5. The spectral index and its uncertainty for each source are listed in Table 3 (columns three and four, respectively). The uncertainty in the spectral index

---

[4] Recall that $\alpha_\lambda = 2 - \alpha_\nu$, which implies $d\alpha_\lambda = -d\alpha_\nu$.

[5] The full list of fit parameters for each source and their uncertainty are available in machine-readable form at http://astro.physik.uni-goettingen.de/~tepper/da/fitparam.txt

Table 3. QSO sample selected from the SDSS DR5. The first four columns include the object designation[a], the emission redshift as quoted in the SDSS DR5 catalog, the spectral index used for fitting the continuum, and its uncertainty, respectively. The last three columns list the measurement of D_A and its total uncertainty, and the uncertainty due to different error sources, respectively (see text for details.)

Object	z_{em}	α_λ	$\sigma_{fit}(\alpha_\lambda)$	D_A	$\sigma_{fit}(D_A)$	$\sigma_{sys}(D_A)$
SDSS J115538.60+053050.5	2.712	-1.213	0.002	$0.390^{+0.010}_{-0.040}$	$^{+0.010}_{-0.010}$	0.039
SDSS J112107.99+513005.4	2.843	-1.256	0.003	$0.281^{+0.015}_{-0.032}$	$^{+0.015}_{-0.016}$	0.028
SDSS J010619.24+004823.3	2.882	-1.170	0.003	$0.294^{+0.018}_{-0.035}$	$^{+0.018}_{-0.018}$	0.029
SDSS J075618.13+410408.5	2.956	-1.320	0.007	$0.329^{+0.038}_{-0.052}$	$^{+0.038}_{-0.040}$	0.033
SDSS J164219.89+445124.0	3.125	-1.550	0.001	$0.349^{+0.005}_{-0.035}$	$^{+0.005}_{-0.005}$	0.035
SDSS J004054.65-091526.8	3.185	-1.278	0.000	$0.459^{+0.000}_{-0.046}$	$^{+0.000}_{-0.000}$	0.046
SDSS J124306.55+530522.1	3.317	-1.255	0.012	$0.426^{+0.057}_{-0.077}$	$^{+0.057}_{-0.064}$	0.043
SDSS J083122.57+404623.4	3.365	-1.230	0.003	$0.261^{+0.017}_{-0.031}$	$^{+0.017}_{-0.017}$	0.026
SDSS J085343.32+370402.3	3.475	-1.115	0.005	$0.527^{+0.019}_{-0.056}$	$^{+0.019}_{-0.020}$	0.053
SDSS J093523.32+411518.7	3.566	-1.407	0.007	$0.396^{+0.036}_{-0.055}$	$^{+0.036}_{-0.039}$	0.040
SDSS J094349.65+095400.9	3.713	-1.199	0.014	$0.529^{+0.057}_{-0.083}$	$^{+0.057}_{-0.064}$	0.053
SDSS J023137.64-072854.5	3.750	-1.269	0.008	$0.462^{+0.037}_{-0.061}$	$^{+0.037}_{-0.039}$	0.046
SDSS J144717.97+040112.4	3.931	-1.303	0.016	$0.525^{+0.062}_{-0.089}$	$^{+0.062}_{-0.072}$	0.052
SDSS J162331.15+481842.1	3.990	-1.113	0.006	$0.494^{+0.022}_{-0.055}$	$^{+0.022}_{-0.023}$	0.049
SDSS J014049.18-083942.5	4.112	-1.181	0.010	$0.488^{+0.041}_{-0.066}$	$^{+0.041}_{-0.044}$	0.049
SDSS J234150.01+144905.9	4.155	-1.094	0.010	$0.558^{+0.035}_{-0.067}$	$^{+0.035}_{-0.038}$	0.056
SDSS J081240.68+320808.6	4.332	-1.074	0.028	$0.517^{+0.100}_{-0.137}$	$^{+0.100}_{-0.127}$	0.052
SDSS J103601.03+500831.8	4.449	-1.149	0.012	$0.612^{+0.039}_{-0.075}$	$^{+0.039}_{-0.043}$	0.061
SDSS J162626.50+275132.4	4.580	-1.187	0.023	$0.681^{+0.057}_{-0.097}$	$^{+0.057}_{-0.069}$	0.068
SDSS J005006.35+005319.2	4.663	-1.204	0.018	$0.693^{+0.043}_{-0.085}$	$^{+0.043}_{-0.049}$	0.069
SDSS J083914.14+485125.7	4.885	-1.350	0.030	$0.737^{+0.057}_{-0.104}$	$^{+0.057}_{-0.073}$	0.074
SDSS J163950.52+434003.7	4.976	-1.321	0.024	$0.694^{+0.056}_{-0.098}$	$^{+0.056}_{-0.069}$	0.069
SDSS J233446.40-090812.3	5.107	-1.332	0.000	$0.735^{+0.000}_{-0.073}$	$^{+0.000}_{-0.000}$	0.073
SDSS J101447.18+430030.1	5.275	-1.220	0.029	$0.758^{+0.058}_{-0.107}$	$^{+0.058}_{-0.076}$	0.076
SDSS J142123.98+463317.8	5.414	-1.321	0.080	$0.825^{+0.087}_{-0.192}$	$^{+0.087}_{-0.173}$	0.082

[a] The designation of each object meets the IAU nomenclature, as required. For details on the official SDSS designation of an object, please consult www.sdss.org/dr5/coverage/IAU.html

quoted in column four and denoted by $\sigma_{fit}(\alpha_\lambda)$ includes the formal error inherent to the fitting process, which is due to the error in the detected flux, the removal of broad emission lines in the region $\lambda \geqslant 121.6$ nm – which makes the wavelength range become patchy –, and the systematic uncertainty due to the decreasing available wavelength range with redshift mentioned above. The systematic effect of the latter in particular can clearly be seen on the overall trend for the uncertainty in the spectral index to increase with redshift.

4.2 Measurement of D_A

For each QSO spectrum, we compute the *total* absorption pixel by pixel in the restframe range $\lambda \in [105, 117]$ nm according to the expression[6]

$$D_A(z_{em}; \alpha_\lambda) \equiv 1 - \frac{1}{N_{pix}} \sum_{i=1}^{N_{pix}} \frac{f_{obs}(\lambda_i)}{f_c(\lambda_i; \alpha_\lambda)}, \qquad (23)$$

where N_{pix} is the total number of pixels between $\lambda_1 = 105 \cdot (1 + z_{em})$ nm and $\lambda_2 = 117 \cdot (1 + z_{em})$ nm. The lower limit of this range is chosen as to avoid the wings of the Lyβ + Ovɪ emission lines, while the upper limits is such that the blue wing of the Lyα emission line is excluded in the computation of

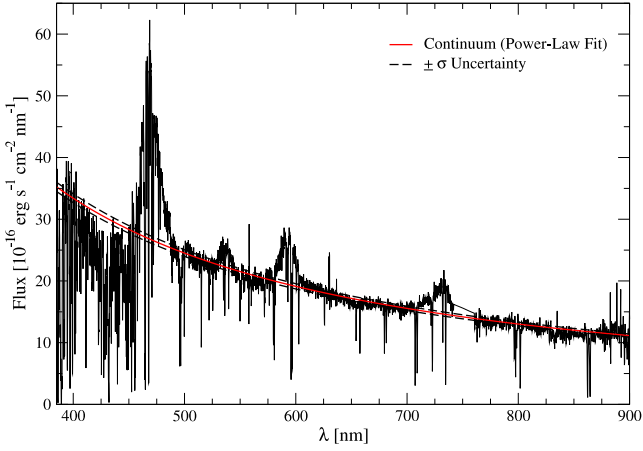[6] This is just the discrete version of equation (2).

Figure 5. Spectrum of the QSO SDSS J112107.99+513005.4 at $z_{em} = 2.843$. The heavy solid and heavy dashed line indicate the power-law fit to the continuum and its uncertainty, respectively. The corresponding spectral index is $\alpha_\lambda = 1.256 \pm 0.003$.

D_A. The measurement of D_A for each source together with its corresponding uncertainty are listed in Table 3 (column five).

There are two independent sources of uncertainty in the measurement of D_A: 1) The error in the spectral index inherent to the continuum fitting, *i.e.* $\sigma_{fit}(\alpha_\lambda)$, and 2) the possibility that the continuum level may have been systematically overestimated due to the change in slope of the power-law at $\lambda = 121.6$ nm (cf. Section 4.1). We shall refer to the uncertainty in D_A due to the first and second sources $\sigma_{fit}(D_A)$ and $\sigma_{sys}(D_A)$, respectively. The total uncertainty in the measurement of D_A –quoted in column five of Table 3– is then assumed to be given by $\sigma^2(D_A) = \sigma_{fit}^2(D_A) + \sigma_{sys}^2(D_A)$.

As a consistency check, we adopt two different methods to estimate $\sigma_{fit}(D_A)$. The first method assumes that the $\pm\sigma$ range for each measurement is given by $D_A[z_{em}; \alpha_\lambda \pm \sigma_{fit}(\alpha_\lambda)] - D_A[z_{em}; \alpha_\lambda]$, respectively, for the corresponding values of $\sigma_{fit}(\alpha_\lambda)$ listed in Table 3. The second method is based on error propagation, according to which the error $\sigma(f)$ in the estimate of a quantity $f(x_i)$, which depends on n independent random variables $\{x_i\}$, each with an uncertainty $\sigma(x_i)$, is given by

$$\sigma^2(f) = \sum_{i=1}^{n} \sigma^2(x_i) \cdot \left(\frac{\partial f}{\partial x_i}\right)^2 . \tag{24}$$

In the case of our measurements of D_A, only the uncertainty in the index α_λ is relevant, and hence equation (24) becomes, with the appropriate notation,

$$\sigma_{fit}(D_A) \equiv \sigma_{fit}(\alpha_\lambda) \cdot \frac{1}{N_{pix}} \sum_{i=1}^{N_{pix}} \ln(\lambda_i + \lambda_0) \frac{f_{obs}(\lambda_i)}{f_c(\lambda_i; \alpha_\lambda)} , \tag{25}$$

where we have used equations (22) and (23). Note that we do not include the term due to the error in the detected flux f_{obs}, since this is already included in $\sigma_{fit}(\alpha_\lambda)$. The error computed according to the first method is asymmetric, as expected, while the error computed using equation (25) approximately (in some cases exactly) corresponds to the arithmetic mean of the former. Since the error estimates computed according to both these methods are consistent with each other, we may use either to estimate the uncertainty in D_A due to the uncertainty in the continuum fit; we choose to use

the values computed according to method 1, which are listed in Table 3 (column six).

It is not clear based on the results found in the literature (cf. Section 4.1) whether the continuum blueward of Lyα may *always* be overestimated when assuming it to be described by an extrapolation of the power-law redward of Lyα. However, we will assume that this is the case in order to allow for the possibility that our D_A measurements may have been systematically overestimated. In order to compute the corresponding uncertainty in D_A, we make the following assumptions: First, the continuum in the region blueward of Lyα shall be described by a power-law as well, but with a smaller spectral index – *i.e.* the continuum is *flatter* – with respect to the region redward of Lyα. Since it is not possible to fit a power-law in this regime for the data we use, we assume that both power-laws (above and below Lyα) are normalised at $\lambda = 121.6$ nm. Furthermore, and for simplicity, we take the difference in the spectral indices below and above Lyα to be equal 1 (which corresponds roughly to the findings in the literature quoted in Section 4.1). Under these assumptions, the value of D_A for our sample turns out to be overestimated on average by less than ten per cent. This value is slightly lower for our measurement of D_A at $z_{em} = 2.71$ than the correction estimated by Seljak et al. (2003). However, they also remark that the exact value of the bias depends on the particular method applied to extrapolate the continuum in the Lyα Forest region. We hence make the rather conservative assumption that each measurement of D_A has an additional uncertainty $\sigma_{sys}(D_A)$ of ten per cent, and add this uncertainty only to the *low* error bound, since we are assuming that the continuum, and hence D_A, may have been *over*estimated. We could in principle correct our D_A estimates by shifting all measurements by ten per cent towards a lower value, but since it is not clear whether the continuum has been overestimated for *every single* measurement, we prefer to express this bias as an uncertainty. The uncertainty $\sigma_{sys}(D_A)$ for each measurement is listed in the last column of Table 3.

Note that we do not correct our measurements for contamination of metal lines. However, this should not introduce a large error, since their contribution is small. For example, Tytler et al. (2004a) find that they contribute by 2.3 ± 0.5 per cent to the total absorption at $z = 1.9$. The validity of this assumption will be tested in the next section by the comparison of D_A-measurements to the results of our simulations, in which the absorption due to metal lines is not included.

5 RESULTS & DISCUSSION

5.1 Observations vs. Models

We compute the evolution of D_A in the redshift interval $0.35 < z_{em} < 6.0$ using the models MMC (with and without Lyman limit systems) and BMC presented in the Section 3, which include the effect of different populations of absorbers, *i.e.* Lyα forest clouds and Lyman limit absorbers. For comparison, we include also the values of D_A computed using Meiksin (2006)'s model (see Section 1.1). We refer to this model as MTC. For the MMC (with and without LLS) and BMC models, we simulate an ensemble of $N_{LOS} = 4 \cdot 10^3$ lines-of-sight at fixed redshift, and compute for each of them the flux decrement D_A according to equation (23). In this way we get for each given redshift an ensemble of an equal number of D_A values for each model, from which we estimate the 50 per cent quantile (median), and the $\pm 34.13, \pm 43.32, \pm 47.72, \pm 49.38$, and ± 49.87 per cent quantiles around the median, which correspond to the

$\pm 1, \pm 1.5, \pm 2, \pm 2.5$ and $\pm 3\,\sigma$ ranges. We do not compute the mean and σ ranges in the standard way, since the distribution of D_A is strictly speaking unknown a priori. However, according to the results from Section 3, the distribution of D_A is not too far from a lognormal or even a Gaussian distribution; hence, the identification of mean with median, of $\pm\sigma$ range with the ± 34.13 quantile around the median, and so on, is justified. Since the MTC model only provides a (mean) transmission function at each redshift, we compute for this model one single value for D_A at each redshift by numerically integrating the corresponding transmission function in the restframe wavelength range $[105, 117]$ nm.

We compare our simulations to measurements of D_A accumulated in the literature over the past two decades, performed with different methods and approaches. This compilation is by no means intended to be complete. The reason for choosing these measurements is mainly that they were performed in more or less mutually exclusive redshift ranges which all together cover the range $0 < z_{em} < 4$. Our measurements extend this redshift range out to $z_{em} = 5.41$. Therefore, these previous measurements together with our own allow a comparison to the models presented here over a wide redshift range, but with the caveat that using measurements from different groups may introduce an artificial bias in the evolution of D_A due to the differences in the particular method used and the different redshift ranges probed in each case, as previously mentioned. The literature data and our measurements are shown together with our model calculations in Figure 6 (see cited references for details on the corresponding measurements).

As can be seen in this figure, the predictions for the evolution of the mean D_A from all three models, MMC, BMC, and MTC, are practically indistinguishable from each other when compared to observations for redshifts $z_{em} \lesssim 3$. We do not include in this figure the predictions for D_A based on the MMC model without LLSs, since the difference between this and the full MMC model is negligible. At higher redshifts, the values of D_A obtained from the MMC and MTC models match the observations quite well. Note that in spite of the fundamentally different approaches and input physics of these models, their results are very similar. As a result, it is not possible yet to discriminate between these models due to the uncertainty and the strong scatter in the observations, especially around $z_{em} \approx 3.5$ (cf. Section 3.3). In contrast to the MMC and the MTC models, the predictions from the BMC model strongly deviate from the measurements of D_A at $z_{em} > 3$, *underestimating* the absorption. Since the MMC and BMC models differ only in terms of the input distributions and not in terms of the method, this result suggests that the evolution of the Lyα absorbers implied by the input distributions of the BMC model is slower than expected from the evolution of D_A (cf. equation 3). If used to compute the magnitude changes for high-redshift galaxies due to intergalactic absorption as done by *e.g.* BCG99, the difference in the predicted evolution of D_A between the MMC and BMC models corresponds to a difference of slightly more than 0.6 mag in the predicted magnitude change for *e.g.* an Sd-type galaxy –with a constant Star Formation Rate– at $z_{em} = 4.0$ in the F450W filter. The corresponding difference between the MMC and MTC model amounts to less than 0.3 mag.

It is worth mentioning at this point that, in contrast to previous models (*e.g.* Zuo 1993), we explicitly avoid normalising in any way the used distributions to match the observed D_A at some given redshift, or manipulating the models whatsoever to satisfy any other restriction. We simply take the distributions as reported in the literature, where they were determined directly by means of line statistics by the authors and references therein. We mention this in order
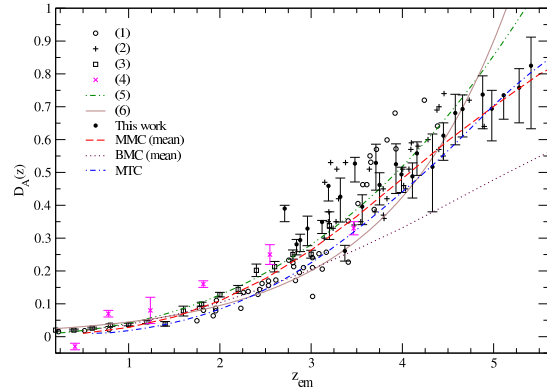

Figure 6. Evolution of the mean D_A computed according to the MMC (dashed line), BMC (dotted line), and MTC (dot–dashed line) models, compared to observations performed over the past twenty years by different groups and different methods: (1) Zuo & Lu (1993). (2) Schneider et al. (1991), (3) Kirkman et al. (2007), (4) O'Brien et al. (1988). The data points display 1σ error bars. Note that, in spite of the heterogeneity of the approaches to measure D_A, its mean evolution as computed using the MMC and MTC models matches well the observations over the entire redshift range shown, while they disagree strongly with the BMC model at redshifts $z_{em} > 3.0$. Below this redshift, the models are practically indistinguishable from each other. For completeness, we include the empirical fits to the evolution of D_A from: (5) Zhang et al. (1997), and (6) Kirkman et al. (2005).

to emphasize the rather good agreement between the observations and the evolution of the mean D_A predicted by all three models at $z_{em} < 3$, and by the MMC model at higher redshifts, further reinforced by the similarity between the results of the MTC and the MMC models, which are completely independent from each other.

We include in Figure 6 the empirical fits of Zhang et al. (1997) of the form $D_A(z) = D_A^0\,\mathrm{e}^{\alpha(1+z)}$ with $D_A^0 = 0.01$ and $\alpha = 0.75$ and of Kirkman et al. (2005) of the form $D_A(z) = A(1+z)^\gamma$ with $A = 0.0062$ and $\gamma = 2.75$. These empirical fits match the observations quite well at $z_{em} \lesssim 4.5$, and may be useful as a rough estimate of the mean absorption at those redshifts. However, as stated before, they should be taken with caution, especially at high redshifts. While D_A asymptotically converges to unity as the redshift approaches the epoch of reionisation, z_{reion}, these empirical fits diverge. In contrast, the predicted evolution of D_A from all models described above does satisfy the expected and observed asymptotic behaviour. The rate of convergence to this limiting value certainly depends on the particular set of input distributions used, as shown by the comparison between the MMC to the BMC model. In general, the stronger the number density evolution, the faster the convergence. In other words, different input distribution functions imply different values for z_{reion}, and this information may be used as a further constraint on the accuracy of a particular set of input distributions. Extending the computations of the evolution of D_A with the MMC model to redshifts $z_{em} > 6.0$, it turns out that D_A is almost unity at $z_{em} \approx 7.0$. This value is slightly higher than the value of 6.5 quoted by Fan et al. (2002) for the epoch of reionisation.

The compilation of D_A measurements shown in Figure 6 displays a large scatter at all redshifts, in particular at redshifts $z_{em} > 3$. Previously, concerns were raised about the possibility that the amplitude of the intrinsic scatter due to cosmic variance may
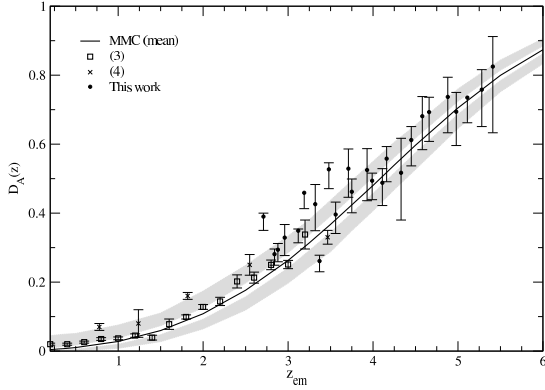
Figure 7. Evolution of the mean D_A according to the MMC model (solid line) in the range $0.2 < z_{em} < 6$. Here we show again our measurements of D_A and those from O'Brien et al. (1988), together with the most recent measurements from Kirkman et al. (2007). The data points display 1σ error bars. The white and shaded areas around the solid line indicate the intrinsic scatter at the ± 1 and $\pm 3\sigma$ level, respectively, due to variation in the absorption from one random line-of-sight to another.

be enhanced by combining measurements performed with different methods and using heterogeneous data samples. Hence, we select a subset from our compilation following the criteria stated below, and compare this observations to the scatter as predicted by our simulations. The results of this comparison are summarised in Figure 7. Here, we show the redshift evolution of D_A as computed from the MMC model. We include the 50 per cent quantile, *i.e.* the median, and the ± 34.13, and ± 49.87 per cent quantiles, which correspond to the ± 1-, and $\pm 3\sigma$ ranges around the median at each redshift. Again, the difference between the predictions of the MMC with and without LLSs is negligible, and thus we show only the results for the full MMC model. We do not include the BMC model in this analysis in view of the disagreement with the observations at redshifts $z_{em} > 3$. Our new data subset consists of our own measurements together with the measurements of O'Brien et al. (1988), who use power-law fits to estimate the continuum level of the quasars, and of the measurements from Kirkman et al. (2007), who fit the continuum locally using b-splines. We select these measurements for the following reasons: O'Brien et al.'s data allow a more direct comparison to our data due to the similarity in method to estimate D_A; Kirkman et al. (2007)'s data are the most recent and accurate measurements of D_A available to date. Furthermore, these data sets individually cover the largest redshift ranges, and together they constitute the smallest set of individual measurements fully spanning the redshift range $0.2 < z_{em} < 5.41$ with some overlap.

Within the quoted uncertainty, the observations are well matched by the evolution of D_A as predicted by the MMC model in the redshift range $0.2 < z_{em} < 5.41$, with the exception of some outliers around $z_{em} \approx 3$, where the scatter of our observations alone as well as the scatter in the full sample of measurements is large, as can be seen in Figure 6. It is apparent from Figure 7 that the largest scatter in the observed values of D_A at a given redshift cannot be ascribed to the *intrinsic* scatter in the absorption due to variations from one line-of-sight to another as computed with the MMC model, not even at the $\pm 3\sigma$ level. Nevertheless, note that this is only the case for O'Brien et al. (1988)'s and our

measurements, while Kirkman et al. (2007)'s measurements are all nicely contained within the $\pm 3\sigma$ envelope of the model. This suggests that the observed scatter –or at least part of it– is not real, but only an artifact introduced by combining observations based on different methods to measure D_A. This assumption is supported by the small amplitude of the scatter in the distribution of high-redshift galaxy colors already found by BCG99 (cf. their Figure 6) for a variety of input distribution functions. Also, the fact that Kirkman et al. (2007)'s measurements are well matched by the MMC model points to the fact that the particular method chosen by O'Brien et al. (1988) and ourselves to estimate the continuum in the Lyα Forest region does introduce a bias in the measurements of D_A. On the one hand, there is very likely a bias do to the already discussed systematic overestimate of the continuum level. Furthermore, neither O'Brien et al. (1988) nor we corrected our corresponding measurements for metal-line absorption. These facts together would explain in part why precisely our and their measurements are higher on average than the median value of D_A predicted by the MMC model, as can be seen in Figure 7. Even though its effect is expected to be small, including the absorption of metal lines should eventually increase the agreement between the results from the models and the observations.

There is thus rather strong evidence that most of the amplitude in the observed scatter in D_A displayed in Figures 6 and 7 is introduced by combining measurements performed on heterogeneous data samples using different methods. If real, however, the observed scatter would indicate that the models, in particular the MMC model, cannot account for the variation of D_A among different lines-of-sight, even though it reproduces well its *mean* redshift evolution.

According to the results from our simulations (cf. Section 3.2), it is expected that D_A as predicted by the MMC model be lognormally distributed with a high confidence at redshifts $z_{em} \lesssim 6$, and consequently, that $\ln D_A$ and hence the optical depth of H$_I$ should obey a Gaussian distribution. However, this result is not supported by the observations given the poor agreement between the scatter predicted by the MMC models and the observed distribution of D_A at a given redshift. On the other hand, the expectation that D_A should be either lognormally or Gaussian distributed is based solely on the fact that the transmission factor be written in the form of equation (12), which is certainly independent of the particular model accounting for the evolution of the Lyα absorbers. A different set of input distribution functions may have an effect on the mean value and the spread of the distribution of D_A at each redshift, but not the on the form of the distribution. It may be thus possible that a different (more accurate) set of input distribution functions of the absorber properties may help to reconcile the lack of agreement between the *amplitude* of the predicted and the observed scatter, but even in this case the disagreement regarding the *form* of the distribution when compared to previous studies (*e.g.* Madau 1995, Bernardi et al. 2003, Tytler et al. 2004a, Meiksin & White 2004) remains.

Even though the amplitude of the observed and simulated scatter do not quite match, it is interesting to note that the redshift at which the intrinsic scatter in evolution of D_A peaks (cf. Figure 4) roughly coincides with the redshift at which the observations show a strong scatter (cf. Figure 6). This redshift matches the redshift at which Bernardi et al. (2003) reported a local decrement in the evolution of the Lyα optical depth τ_{eff}, recently confirmed by Faucher-Giguere et al. (2007), which has been interpreted by Bernardi et al. as a signature of the reionisation of He II. Note, however, that at this redshift the D_A measurements lie predominantly

above the mean value as given by the models (cf. Figure 6), and hence the trend suggested by these observations is exactly the opposite from that found by Bernardi et al. (2003), since a decrease in τ_{eff} implies a decrease in D_A. This strongly indicates that the difference in method and the heterogeneity of the samples of the different surveys does introduce an artificial bias in the observed evolution of D_A, in particular in its scatter.

We finally want to highlight the following curiosity: By taking a close look at Figures 4 and 7 it becomes apparent that the maximum in the evolution of $\sigma(D_A)$ roughly coincides with the point of inflection of the curve that describes the expected redshift evolution of the mean D_A. Mathematically, this would imply that the intrinsic scatter of D_A is proportional to the rate of change of the mean D_A with redshift, *i.e.*

$$\sigma(D_A) \propto \frac{\partial \langle D_A \rangle}{\partial z} . \tag{26}$$

Indeed, if one computes numerically the derivative of $\langle D_A \rangle$ with respect to z, it turns out that it qualitatively matches the evolution of $\sigma(D_A)$, up to a scale transformation. A physical interpretation may be gained in light of equation (3): A rapid evolution of the absorbers, quantified by the parameter γ, implies a stronger evolution of D_A with z. Since the evolution is however different in general for different lines-of-sight, this in turn implies a larger variation in the absorption from line-of-sight to line-of-sight, and thus a larger value of $\sigma(D_A)$. This is mathematically consistent with the fact that the scatter given by the derivative of equation (3) with respect to z is proportional to γ.

6 SUMMARY & CONCLUSIONS

(i) We measured the cosmic flux decrement D_A and its uncertainty due to statistical and systematic errors for 25 QSOs of the SDSS DR5 catalog in the redshift range $2.71 \leqslant z_{em} \leqslant 5.41$.

(ii) We modeled the redshift evolution of the mean D_A and its distribution at each given redshift in a Monte Carlo fashion, adopting two among the various models presented by Bershady et al. (1999). We found that the predictions of the MMC model for the evolution of the mean D_A reproduce well the observations in the range $0.2 < z_{em} < 5.41$, in contrast to the BMC model. We conclude from this that the underlying input distributions of the BMC model may not accurately account for the evolution of the Lyα absorption on the IGM. Hence, estimates of the impact of the intergalactic attenuation on the photometric properties of high-redshift galaxies using this particular model should be taken with caution. Incidentally, by showing the rather good agreement between one particular set of simulations and the data, we show the power of the relatively simple approach used here to model the effect of intergalactic absorption, as compared to models based on hydrodynamical simulations which are by far more complex.

(iii) Through the comparison between the MMC and BMC models, we showed that different input distribution functions may have very different predictions regarding the evolution of D_A, a quantity which is intimately related –and hence very sensitive– to the amount of HI present along the line-of-sight. Therefore, we state that any model of the intergalactic absorption should first of all reproduce the observed evolution of D_A and its distribution before it is used to *e.g.* correct synthetic or observed spectra for intergalactic absorption.

(iv) The results from our simulations suggest that the distribution of D_A at a given redshift be well described by a lognormal distribution at low redshifts and by a Gaussian distribution at high redshifts, in agreement with our theoretical expectation based on the fact that the absorption is mathematically expressed as the product of small, statistically independent factors. This result implies that at redshifts where D_A is distributed lognormally, the effective optical depth of the intergalactic HI should obey a normal distribution, contrary to the results of previous studies. However, this result should be taken with caution in light of the fact that the models, in particular the MMC model, cannot reproduce the amplitude of the observed scatter in D_A.

(v) We argue that most of the observed scatter in D_A is introduced by combining measurements based on different methods. Nevertheless, the reason for the lack of agreement between the scatter predicted by the MMC model and the observed distribution of D_A is not ultimately settled. Because of this and until this discrepancy is clarified, we may warn about using any attenuation model based on the input distributions of the MMC model to estimate magnitude changes in the spectra of background sources.

(vi) A larger, homogeneous sample of accurate measurements of D_A over a wide redshift range is needed in order to allow for a more faithful comparison to models, and in particular, to determine the intrinsic *form* and *amplitude* of the distribution of D_A as a function of redshift.

---

[7] The SDSS Web Site is http://www.sdss.org/.

Naval Observatory, and the University of Washington.

REFERENCES

Adelman-McCarthy J. K. et al., 2007, ApJS, 172, 634
Barvainis R., 1993, ApJ, 412, 513
Bernardi M. et al., 2003, AJ, 125, 32
Bershady M. A., Charlton J. C., Geoffroy J. M., 1999, ApJ, 518, 103
Bicker J., Fritze-v. Alvensleben U., Möller C. S., Fricke K. J., 2004, A&A, 413, 37
Cristiani S., Giallongo E., Buson L. M., Gouiffes C., La Franca F., 1993, A&A, 268, 86
Desjacques V., Nusser A., Sheth R. K., 2007, MNRAS, 374, 206
Dobrzycki A., Bechtold J., 1991, ApJ, 377, L69
Duncan R. C., Ostriker J. P., Bajtlik S., 1989, ApJ, 345, 39
Fan X., Narayanan V. K., Strauss M. A., White R. L., Becker R. H., Pentericci L., Rix H.-W., 2002, AJ, 123, 1247
Faucher-Giguere C. ., Prochaska J. X., Lidz A., Hernquist L., Zaldarriaga M., 2007, ArXiv e-prints, 709
Giallongo E., Cristiani S., Trevese D., 1992, ApJ, 398, L9
Giallongo E., D'Odorico S., Fontana A., McMahon R. G., Savaglio S., Cristiani S., Molaro P., Trevese D., 1994, ApJ, 425, L1
Giallongo E., Gratton R., Trevese D., 1990, MNRAS, 244, 450
Gunn J. E., Peterson B. A., 1965, ApJ, 142, 1633
Jenkins E. B., Ostriker J. P., 1991, ApJ, 376, 33
Karzas W. J., Latter R., 1961, ApJS, 6, 167
Kawaguchi T., Shimura T., Mineshige S., 2001, ApJ, 546, 966
Kim T.-S., Carswell R. F., Cristiani S., D'Odorico S., Giallongo E., 2002, MNRAS, 335, 555
Kim T.-S., Cristiani S., D'Odorico S., 2001, A&A, 373, 757
Kim T.-S., Cristiani S., D'Odorico S., 2002, A&A, 383, 747
Kim T.-S., Hu E. M., Cowie L. L., Songaila A., 1997, AJ, 114, 1
Kirkman D., Tytler D., Lubin D., Charlton J., 2007, MNRAS, pp 131–+
Kirkman D. et al., 2005, MNRAS, 360, 1373
Kirkman D., Tytler D., Suzuki N., O'Meara J. M., Lubin D., 2003, ApJS, 149, 1
Lamb D. Q., Reichart D. E., 2000, ApJ, 536, 1
Laor A., Fiore F., Elvis M., Wilkes B. J., McDowell J. C., 1997, ApJ, 477, 93
Lu L., Sargent W. L. W., Womble D. S., Takada-Hidai M., 1996, ApJ, 472, 509
Lynds R., 1971, ApJ, 164, L73+
Madau P., 1995, ApJ, 441, 18
McDonald P., Seljak U., Cen R., Bode P., Ostriker J. P., 2005, MNRAS, 360, 1471
Meiksin A., 2006, MNRAS, 365, 807
Meiksin A., White M., 2004, MNRAS, 350, 1107
Møller P., Jakobsen P., 1990, A&A, 228, 299
Neugebauer G., Oke J. B., Becklin E. E., Matthews K., 1979, ApJ, 230, 79
O'Brien P. T., Wilson R., Gondhalekar P. M., 1988, MNRAS, 233, 801
Oke J. B., Korycansky D. G., 1982, ApJ, 255, 11
Ostriker J. P., Bajtlik S., Duncan R. C., 1988, ApJ, 327, L35
Ostriker J. P., Ikeuchi S., 1983, ApJ, 268, L63
Pierre M., Shaver P. A., Iovino A., 1988, A&A, 197, L3
Rauch M., 1998, ARA&A, 36, 267
Rauch M., Carswell R. F., Chaffee F. H., Foltz C. B., Webb J. K., Weymann R. J., Bechtold J., Green R. F., 1992, ApJ, 390, 387
Rauch M. et al., 1997, ApJ, 489, 7
Sargent W. L. W., Young P. J., Boksenberg A., Tytler D., 1980, ApJS, 42, 41
Schneider D. P., Schmidt M., Gunn J. E., 1989, AJ, 98, 1507
Schneider D. P., Schmidt M., Gunn J. E., 1991, AJ, 101, 2004
Seljak U., McDonald P., Makarov A., 2003, MNRAS, 342, L79
Springel V. et al., 2005, Nature, 435, 629
Steidel C. C., Sargent W. L. W., 1987a, ApJ, 318, L11
Steidel C. C., Sargent W. L. W., 1987b, ApJ, 313, 171
Telfer R. C., Zheng W., Kriss G. A., Davidsen A. F., 2002, ApJ, 565, 773
Tepper-García T., 2006, MNRAS, 369, 2025
Tytler D. et al., 2004a, ApJ, 617, 1
Tytler D., O'Meara J. M., Suzuki N., Kirkman D., Lubin D., Orin A., 2004b, AJ, 128, 1058
Vanden Berk D. E. et al., 2001, AJ, 122, 549
Whiting M. T., Webster R. L., Francis P. J., 2001, MNRAS, 323, 718
Williams P., Shu C.-G., Ménard B., eds,, 2005, Probing Galaxies through Quasar Absorption Lines
Wolfe A. M., Gawiser E., Prochaska J. X., 2005, ARA&A, 43, 861
Zhang Y., Anninos P., Norman M. L., Meiksin A., 1997, ApJ, 485, 496
Zheng W., Kriss G. A., Telfer R. C., Grimes J. P., Davidsen A. F., 1997, ApJ, 475, 469
Zuo L., 1993, A&A, 278, 343
Zuo L., Lu L., 1993, ApJ, 418, 601