Nonuniform Distribution of Galaxies in Very Large Scale

by

Andrzej M. Soltan

Nicolaus Copernicus Astronomical Center, Bartycka 18, 00-716 Warsaw, Poland

Andrzej S. Kudlicki

Warsaw University Observatory, Al. Ujazdowskie 4, 00-478 Warsaw, Poland

Received January 31, 1994

ABSTRACT

Spatial distribution of galaxies identified as intervening objects producing metal absorption lines in quasar spectra is investigated. Autocorrelation function is used to measure fluctuations of the distribution of these galaxies at cosmological distances. To improve statistics several sets of data spread in the literature are analysed simultaneously. Non-homogeneity of the observational material is carefully investigated and selection effects are modelled individually for each data set. After subtraction of effects introduced by observational bias the residual pair distribution exhibits statistically significant excess of pairs with separations up to \( \sim 1000 \) Mpc, indicating the non-homogeneous distribution of matter at that scale. At still larger separations distribution of objects is uniform; for \( 1000 < r < 5000 \) Mpc amplitude of the autocorrelation function is consistent with zero, with typical upper limits averaged over 1000 Mpc bins equal to about 0.015.

Key words: Galaxies: clustering – large scale of Universe

1. Introduction

The problem of sizes of the largest formations of matter existing in the Universe is not solved definitively. Substantial observational material on the galaxy distribution contained in numerous catalogues of galaxies reveals a great variety of structures. Agglomerates of galaxies cover a wide range of sizes and masses from sub-Mpc scales for binary galaxies and small groups to more than 100 Mpc for the Great Wall (Geller and Huchra 1989). Studies of the distribution of galaxies at larger scales are problematical since there are no adequate sets of data covering sufficiently large volume. At these very large scales catalogues of clusters of galaxies have been used to trace the distribution of matter (e.g., Bahcall and Soneira 1982).
On still larger scales, near 1 Gpc and above, only quasars spread out to the adequate distances. Several authors have analysed samples of quasars in respect of the spatial correlations (e.g., Osmer 1981, Iovino and Shaver 1988, Kruszewski 1988). Although investigations of the quasar distribution in these papers are concentrated on the correlations at much smaller scales, some restrictions on the large-scale clustering are obtained. General results can be summarized as follows. Significant ($\sim 3\sigma$) excess of quasar pairs with separations smaller than $20 - 30$ Mpc$^1$ is observed for quasars with low and moderate redshifts ($z < 1.5 - 2.0$). Within statistical uncertainties there is no correlation at any larger scale. Quasars at high redshift do not show detectable correlations both at small and large scales. For the purpose of present investigation it is interesting to get some quantitative estimate for the upper limit of the correlation function for separations above $\sim 100$ Mpc. Fig. 2 of Iovino and Shaver and Fig. 2 of Kruszewski indicate that the correlation function $\xi(r) < \sim 0.05$ for $r > 100$ Mpc, but the authors do not give any numerical values in their texts.

Absorption redshifts in the quasar spectra provide abundant information on the distribution of matter. It is now established that quasar absorption features can be divided into three separate categories: (a) broad absorption line systems which are related to matter ejected from quasars with velocities up to $\sim 0.1c$, (b) low column density Ly-$\alpha$ systems forming the forest of lines shortward of Ly-$\alpha$ emission, and (c) narrow heavy element lines most often associated with high column density Lyman limit systems. Tytler (1987) postulated that classes (b) and (c) belong to the same population.

In the present paper we investigate spatial distribution of objects producing narrow, heavy element absorption lines. Thus, our analysis is limited to the category (c) of the above classification. It is well established that these line systems are produced by gas associated with galaxies situated close to the line of sight towards the quasar. For small redshifts ($< 1$), the hypothesis of intervening galaxies has been directly confirmed by imaging and spectroscopic observations of absorbing galaxies (e.g., Bergeron and Boissé 1991, Bergeron et al. 1992). At higher redshifts, intervening galaxies could not be detected due to obvious reason, however, there is no doubt that absorption systems are produced by intervening objects unrelated to the quasar (e.g., Sargent et al. 1988a and references therein). Lines due to several heavy elements in various ionization states are observed. The most conspicuous and easily identified absorption feature in the quasar spectra is the Mg II 2800 Å doublet for low and moderate redshifts and the C IV 1500 Å doublet for moderate and high redshifts. This redshift dependence obviously results from the observational selection introduced by the atmospheric window. Usually detection of one of these doublets is sufficient to define the absorption system. Variety of absorption characteristics and wide range of equivalent width ratios of lines due

$^1 H_0 = 50$ km s$^{-1}$ Mpc$^{-1}$ is used throughout the paper and all results quoted from other papers are scaled to this value of $H_0$. 
to different elements reveal substantial diversity among absorbers. However, they share some common characteristics and are usually treated as a single class of objects (e.g., Steidel et al. 1988). The available data do not allow to specify whether "intervening objects" are typical galaxies or form a peculiar class of galaxies distinguishable by gas content, heavy element abundances, size of the gaseous halo etc. (Steidel 1990b, Bechtold and Ellingson 1992). Thus, it cannot be ruled out that these objects exhibit also some specific characteristics of the spatial distribution and are not representative for the general population of galaxies.

A simple statistics applied frequently in the analysis of the clustering of absorbers is the correlation function of redshifts measured in the spectrum of each quasar. This method was applied by many authors (e.g., Young et al. 1982, Sargent et al. 1988a,b, Steidel 1990a, Heisler et al. 1989, Steidel and Sargent 1992, other references therein). The most detailed analysis with quantitative and precise estimates of the correlation function for a wide range of separations is given by Heisler et al. (1989). Results presented in that paper are most relevant for the present investigation. The conclusions reached by other authors are generally similar. Heisler et al. analysed a sample of 229 absorption systems identified by the C IV doublet in spectra of 55 quasars (Sargent et al. 1988). The authors confirm earlier results that there is a strong clustering of absorption systems at separations less than 1000 km/s. They claim also that significant clustering is present on scales up to 10000 km/s. The authors point out that the entire signal at this large scale is due to a collection of 11 absorption redshifts around $z = 1.6$ in the spectrum of a quasar 0237 – 233. Although it is difficult to assign definite statistical uncertainty to the amplitude of the correlation function, reality of the cluster itself is not questioned. It is because the chance coincidence of so many redshifts within narrow band is virtually excluded on statistical ground. Since there is only one such cluster in the sample, one is not allowed to make estimates on the common properties of structures of that kind. The distribution of absorption redshifts in the spectrum of another quasar, 0854+191, produces a second peculiarity in the amplitude of the correlation function. Two widely separated groups of redshifts generate a signal in the correlation function in the largest velocity bin centered on $\sim 50000$ km/s. Removal of these two quasars from the sample substantially changes amplitude of the correlation function and radically affects degree of redshift clustering: sample of 53 spectra does not show any correlations on scales above 1000 km/s. The authors assume that the redshift distribution in the spectrum of 0237 – 233 results from real clustering on scales of 200 Mpc, while feature observed in 0854+191 is purely accidental. This opinion is supported by apparent lack of correlations for all the remaining bins. Inspection of figures and discussion by Heisler et al. lead to a conclusion that uncertainties of the amplitude of large-scale clustering are dominated by statistical noise due to small number of objects in the sample. To reduce fluctuations, larger samples are required. At present a single sample satisfying homogeneous and well-defined criteria similar to those in the Sargent et al. (1988) data is not available.
In order to increase the number of redshifts involved in the analysis we decided to use several data sets scattered in literature. Extension of the analysis to several independent samples substantially increases number of galaxies involved and allows us to reach significant improvement of signal to noise ratio as compared to previous analyses. However, the main advantage of the present investigation results from the fact, that galaxy-galaxy pairs are not restricted to a single quasar spectrum. Limited range of wavelength covered by individual measurement introduces bias against very wide pairs found in one spectrum. Pairs of galaxies identified in spectra of different quasars are not subject to such restriction.

The observational material compiled from many sources is highly non-homogeneous and subject to various selection effects. The main effort done in the present paper is to overcome this evident weakness of the data. In the next Section we describe the available material and define criteria used to construct a relatively large sample satisfying statistical requirements. Analysis of the selected sample is described in Section 3.

2. Observational Material

The available data are highly non-homogeneous. At present, more than 1100 narrow, metal absorption line systems in spectra of ~ 350 quasars are reported in the literature. The list of references on this subject contains more than 300 papers (Junkkarinen et al. 1991, York et al. 1991). Non-homogeneity of the data results from the number of reasons. First, spectra of different resolution and signal-to-noise ratio have been used. This means that equivalent widths of absorption lines cannot be used as a criterion to select the data. Second, different spectral ranges and variety of lines used in the redshift determinations introduce a natural selection for the range of measurable redshifts. And third, the authors apply different criteria to accept the observed lines as a distinct redshift system. All these arguments show that due to the complex set of selection effects it is not feasible to construct from the observational material a statistically complete and numerous sample.

Selection of data in the present investigation has been performed as follows. First, we have selected several sub-samples which in the original papers were constructed according to the following criteria: (a) the absorption redshift is measured using Mg II or C IV lines, (b) wavelength range analysed in each spectrum is well-defined and (c) equivalent widths of detected lines are above some specified threshold value. For the computational algorithm used in the present analysis (see below) the equivalent width condition is not crucial, nevertheless, it was applied to eliminate spurious line identifications. Next, we retained for further processing only those samples which contained more than 20 absorption redshifts systems. In Section 3 below we generate simulated quasi-random distributions of absorption redshifts for each sample separately. Parameters used in the simulations are calculated using the real data. This imposes the requirement of minimum number of galaxies in the samples.
Several quasar spectra and absorption redshifts were common to two or more sub-samples. In second step of analysis we have removed all duplicates, and in the final sample each quasar appears exactly in one sub-sample. Obviously parameters defining the individual sub-samples were different and the whole material selected this way still was not homogeneous. But these effects are accounted for by careful modelling of our simulation procedure described in the next Section.

Absorption lines separated by less than $\sim 300(1 + z_{abs})$ km/s in a single quasar spectrum exhibit well-known clustering which is commonly interpreted (e.g., Bahcall and Spitzer 1969) as being due to motions of clouds within galactic halo. One can expect also some correlations between absorption redshift systems at scales comparable to velocity dispersion in clusters of galaxies, i.e., of the order of 1000 km/s. The objective of the present investigation is to measure correlations of the galaxy distribution at the very large scale corresponding to 10 000 km/s and more. From this point of view, close absorption redshift pairs could be identified as a single object. To investigate effects on our analysis of multiple redshifts produced by a single galaxy as well as clusters of galaxies we have introduced several minimum redshift separations ranging from $\delta v = 300(1 + z_{abs})$ to $3000(1 + z_{abs})$ km s$^{-1}$. Absorption redshifts which differed from the quasar emission redshift by less than $10 000(1 + z_{em})$ km/s were excluded from the data.

All the restrictions imposed on the observations reduced drastically the size of usable data. Our final sample contains 377 galaxies for $\delta v = 300(1 + z_{abs})$ km/s and 334 for $\delta v = 3000(1 + z_{abs})$ km/s situated in the line of sight towards 196 quasars. The sample consists of 9 separate sub-samples containing from 7 to 124 galaxies.

Fig. 1. Distribution of quasars on the celestial sphere in equatorial coordinates. The sample contains 196 objects. In simulations only radial coordinates of galaxies are randomized, while spherical coordinates are taken from the observed sample.
Distribution of quasars on the celestial sphere in the whole sample is shown in Fig. 1. In agreement with the expectations the distribution is highly non-random. Similarly, the distribution of number of absorption redshifts detected in each spectrum deviates substantially from the Poissonian distribution. This is due mostly to the varying sensitivity and wavelength range covered by observations, although small scale clustering contributes also to this effect. Histogram of absorption redshifts in the whole sample (Fig. 2) has a maximum around $z = 1.7$ and extends from $\sim 0.2$ up to $\sim 3.6$. However, the average redshift range observed for individual quasar is substantially smaller and in the whole sample $\Delta z \approx 0.8$. These collective characteristics of the sample are critical for the subsequent analysis. Typical distance between two galaxies drawn randomly from the sample is of the order of the Hubble radius, while the maximum separation of absorption redshifts in quasar spectrum corresponds to $\sim 1000$ Mpc.

Fig. 2. Distribution of 334 absorption redshifts in the sample. Redshifts separated by less than $3000(1 + z_{\text{abs}})$ km/s are plotted as a single object.

3. Statistical Analysis

We define the autocorrelation function (ACF) of the galaxy distribution in a standard way:

$$\xi(r) = \frac{N_{\text{obs}}}{N_{\text{rnd}}} - 1,$$

(1)
where $N_{\text{obs}}$ denotes the number of observed pairs with separation $r$ and $N_{\text{rnd}}$ – number of pairs expected in the unclustered data. Separations used here are comoving distances calculated for the standard Friedman flat model ($\Lambda = 0$ and $q_0 = 0.5$). For these cosmological parameters relation between redshift $z$ and distance $d$ to the source is given by:

$$d = 2\frac{c}{H_0} \left(1 - \frac{1}{\sqrt{1 + z}}\right),$$

where all symbols have their usual meaning. Although different choice of cosmological parameters would change the absolute scale of separations within the sample objects, it would not affect significantly our estimates of the correlation function.

The number of actually observed pairs $N_{\text{obs}}$ was calculated directly for the whole sample of 377 galaxies. The number of pairs expected for the uncorrelated distribution has been obtained using Monte Carlo technique. Since our sample is subject to a complex constraints simulations of random distributions has been carried out in a way preserving various selection effects.

Spherical coordinates of all the objects in simulated samples were taken from the real data. This is because the nonuniform distribution of quasars in our sample on the celestial sphere results solely from observational limitations and it is not related to the real inhomogeneities of the distribution of matter in space. The simulated data were generated separately for each sub-sample. Although within the individual sub-sample all the absorption lines used in redshift measurements had nominal equivalent widths above specified threshold, distribution of the number of absorption redshifts per unit length of the quasar spectrum in some sub-samples was non-Poissonian. Since the number of galaxies detected in the line of sight toward a quasar is highly sensitive to the equivalent width threshold, small errors in the width measurement introduce variations of the number of catalogued redshifts. Thus the number of galaxies in a sample distributed along the particular line of sight cannot be used in our analysis as evidence of fluctuations in the density of galaxies.

For the each sub-sample the observed redshift distribution was used to create simulated distributions using the Maximum Likelihood method. The actual distribution was approximated by a smooth function:

$$n(z) = n_0(1 + z)^\beta.$$  

(3)

Single parameter $\beta$ was determined separately for all the sub-samples. In the calculations it was taken into account that for each quasar the allowed redshift range is different. Then, separately for each quasar random redshifts were drawn from the distribution given by Eq. (3). Both the number of generated redshifts and the redshift range were equal to their respective parameters in the real data. Statistical characteristics of the original and simulated redshift distributions have
been compared in each sub-sample to search for systematic differences between the data and simulations. We have not noticed any evidences for the potential discrepancies. In particular, both the average redshifts and the second moment of the respective distributions have been found indistinguishable in statistical sense.

After the simulated catalogue containing, similarly to the original data, 9 subsamples had been created, the number of pairs for the relevant range of separations was calculated in the same way as for the observed data. The whole simulation procedure was repeated 50 times and the average number of pairs computed from those 50 randomized catalogues was used as estimator of the \(N_{\text{rnd}}\) in Eq. (1). One should note that our randomization scheme is conservative in the sense that it transfers some intrinsic correlations present in the real data into the simulated samples. Because the signal in the ACF is defined as an excess of the number of pairs in the sample relative to the number of pairs in simulations, the clustering amplitude obtained with the present algorithm is underestimated.

Results of the calculations are presented in Fig. 3. Upper panel shows the ACF in 200 Mpc bins. Error bars correspond to the rms scatter of the number of pairs \(N_{\text{rnd}}\) obtained in the simulated catalogues. In the lower panel the same data are shown in the 1000 Mpc binning. For the 200 Mpc binning the ACF is dominated by statistical noise with signal-to-noise ratio below 2 in all the bins. Results shown in Fig. 3 refer to the input data where two redshifts in a quasar spectrum separated less than \(\delta v = 3000(1 + z_{\text{abs}})\) km/s were treated as a single object. This effectively eliminated tight galaxy pairs and removed strong signal in the ACF at the lowest separations. We have performed calculations also with different values of \(\delta v\). For \(\delta v = 300(1 + z_{\text{abs}})\) km/s small scale redshift clustering produces strong signal in the ACF: \(\xi(r < 100 \text{ Mpc}) = 0.23 \pm 0.08\). At larger distances the results are practically insensitive to \(\delta v\).

Using wide 1000 Mpc bins reduced substantially the noise in ACF measurements. For separation above 1000 Mpc galaxy positions are uncorrelated and the 1\(\sigma\) (3\(\sigma\)) upper limit for the ACF is of the order of 0.015 (0.05). This represents new stringent limit on the fluctuations of the galaxy distribution in the largest scale. At separations below 1000 Mpc the picture is more confusing. Statistically significant correlations are detected at scales roughly an order of magnitude greater than the largest agglomerates of matter known at present. The amplitude of the ACF \(\xi(r < 1000 \text{ Mpc}) = (6.8 \pm 2.2) \cdot 10^{-2}\). The ACF variations plotted with narrow 200 Mpc bins show that most of the positive signal is produced not in the low separations \((r < 500 \text{ Mpc})\), but in the upper part of the 1000 Mpc bin. However, due to large uncertainties, the data are consistent with roughly stable positive ACF in the entire 1000 Mpc bin.

It should be stressed that these uncertainties are at the level of 0.08 or below for the 200 Mpc bins and the overall results are substantially more accurate than in any previous work on this subject.
Fig. 3. The ACF of quasar metal line absorption redshifts. Two panels are for different bin sizes: upper – 200 Mpc, lower – 1000 Mpc. In both panels a dashed horizontal line is drawn for uncorrelated distribution ($\xi \equiv 0$).
4. Discussion

Two main results of the present investigation are the following: (a) the 1σ upper limit for the ACF at separations larger than $\sim 1000$ Mpc is $\sim 0.015$ (Fig. 3), and (b) significant correlations ($\xi \approx 0.07$) on scales up to $\sim 1000$ Mpc are detected.

Relatively numerous sample of galaxies used in the present investigation reduced statistical uncertainties substantially below uncertainties existing in many previous works. Conclusion (a) is consistent with various investigation published to date, but because of better statistics, our constraints on the amplitude of clustering on the Gpc scale are much more stringent.

Potentially more interesting result, viz. positive correlations on scales up to $\sim 1$ Gpc is also formally consistent with lack of the signal detection reported by Heisler et al. Extremely poor statistics in the Heisler et al. investigation at the relevant separations produces uncertainties in the ACF of the order of 1. Other authors who investigated the quasar distribution (e.g., Iovino and Shaver 1988, Kruszewski 1988) concentrated on smaller separations but their results on the ACF at separations around 1 Gpc look consistent with the present detection. Also estimates by Osmer (1981) of the quasar distribution on scales of $100 - 3000$ Mpc which are subject to large uncertainties are not in conflict with our measurements.

Amplitude of the density fluctuations of the galaxy distribution implicated by the positive signal in the ACF is $\delta \rho/\rho \approx 0.26$. Unfortunately, present data on the distribution of galaxies on Gpc scale are not sufficiently accurate to verify our result. Nevertheless, automatic galaxy counts at faint magnitudes in the selected areas offer a possibility to investigate this problem in the near future. Some estimates of the fluctuations of the matter distribution on scales comparable to those investigated here but in earlier epoch can be derived from the fluctuations of the cosmic microwave background (CMB). Smoot et al. (1992) have detected structure with characteristic amplitude of $\delta T/T \approx 6 \cdot 10^{-6}$ on angular scales of $\sim 7^\circ$ which roughly corresponds to the characteristic comoving scale obtained in the present work. Since the relative fluctuations of the matter density at the recombination era are about equal to the temperature fluctuations and the density fluctuations (for $q_0 = 0.5$) increase linearly with the expansion parameter $a \sim (1 + z)^{-1}$ the present amplitude of matter density corresponding to the CMB $7^\circ$ structure is $\delta \rho/\rho \approx 0.006$. This is a factor of 40 below the inhomogeneities reported in this paper. It implies that effect measured using quasar absorption redshifts does not describe overall fluctuations of the matter density in the universe or that growth of density fluctuations is not adequately described by gravitational instability. We are not going to explore the latter possibility and concentrate on other explanations of this discrepancy. First, although the ACF amplitude is detected at 3σ significance level, one could not rule out a conjecture that our measurement is a result of specific statistical fluctuation in the data. In this case the signal reported here should be considered as the upper limit of the ACF rather than the genuine detection.
Second, it is likely that objects producing absorption lines in quasar spectra are not good tracers of the general galaxy distribution, i.e., due to some unknown reason catalogued absorption redshifts are less uniformly distributed than ordinary galaxies. Third, there is a possibility that mechanism of the biased galaxy formation operates in an extreme form also on Gpc scale, and that the galaxy space density is by a substantial factor amplified in areas of only slightly increased total matter density.

Acknowledgements This work has been supported by KBN grant 2 P304 009 04.

REFERENCES