Correlations and structures in modern galaxy redshift surveys

© **Nikolay L. Vasilyev** [1,2]

[1] Sobolev Astronomical Institute, St. Petersburg University, St. Petersburg, Russia
[2] Email: n.l.vasilyev@gmail.com

Abstract: The new generation galaxy redshift surveys which became available over the last five years are first to provide three-dimensional maps containing hundreds of thousands of the local galaxies. These new data allow researchers to extend the range of scales available for correlation analysis to more than 100 Mpc/h. Correlations in the volume limited samples extracted from the latest available versions of 2dF and SDSS galaxy catalogs are characterized using the conditional density and the reduced two-point correlation function. For the conditional density a power-law behavior is detected in the range of scales from about 0.5 Mpc/h to 30 Mpc/h, having an exponent $\gamma = 0.8 \pm 0.2$ for the 2dF and $\gamma = 0.9 \pm 0.1$ for the SDSS. Extending analysis to the larger scales using the SDSS data shows an evidence of possible homogeneity detection with the conditional density going towards a constant value; reliability of this evidence is then discussed. The samples are then tested for a presence of the super-large structures or systematic biases which may affect correlation analysis using simple methods of the differential number counts and the radial density distribution.

1. Introduction

Statistical characterization of the large scale galaxy structures is one of the most important problems of the modern observational cosmology. A detailed understanding of the galaxy clustering and the physics of the structure formation process are crucial in the process of building of the self-consistent model of the evolving Universe. Although it is believed nowadays that the Universe is dominated by a non-baryonic dark matter of an unknown kind, the visible matter distribution is still the only direct way available to study the local extra-galactic structures. The cosmological principle postulating the homogeneity of the matter distribution at large scales is one of the basic assumptions of the standard (Friedmann-Lemaitre-Robertson-Walker) solution, [1]. However recently discovered evidences of super-large structures (see e.g., [2] and [3]) and voids (such as described in [4] and [5]) both in visible and dark matter (see for [3] details) give rise to a question: can the scale at which visible matter distribution becomes statistically uniform be clearly identified?

The problem of galaxy clustering has been actively discussed during several last years, especially in relation to two modern galaxy surveys: the Two degree Field Galaxy Redshift Survey (2dFGRS, [6]) and the Sloan Digital Sky Survey (SDSS, [7]). These new generation three-dimensional surveys represent a significant improvement of our knowledge of the local Universe properties: the numbers of accurately measured redshifts have grown from several thousands at maximum in previous surveys to more than half of a million in the latest releases of the SDSS catalog. Moreover accurate redshift determinations and the multi-band photometry allow one a precise characterization of many parameters and effects (e.g. K-corrections) which were poorly constrained up to a few years ago. However it should be clearly understood that to study the structures at large scales without making a priori assumptions it is not enough for the survey to be deep along the line of sight: a large solid angle is also required. Currently only the latest versions of the SDSS data partially satisfy these requirement, nevertheless the final release of the SDSS shall provide a large contiguous angular sky region (about ¼ of the celestial sphere) in 2009.

When the statistical properties of the large scale structure are considered, the first question is usually about the two-point correlation properties. The two-point correlation analysis methods most commonly used to study galaxy correlations are the reduced two-point correlation function $\xi(r)$, in redshift and real space, and its Fourier conjugate, the power spectrum (see e.g. [8], [9] for 2dFGRS-related results and [10], [11], [12] for some recent studies of SDSS). These methods as discussed in [13] can be affected by effect of a finite volume of a sample being analyzed, moreover the estimators commonly used to compute these functions indirectly suppose that the homogeneity scale is achieved within the sample (i.e. that the structures and correlations can be considered as small fluctuations). An alternative statistics useful to determine correlation properties in the regime of strong clustering (i.e. in case of irregular distributions) is the conditional average density, [14]. Recent analyses of the SDSS data using conditional density method show a simple power-law scaling with correlation exponent $\gamma \approx 1$ up to 20-30 Mpc/h ([15], [16]).

Estimations of the two-point correlation properties in a finite sample volume can be affected by two systematic factors: (i) systematical biases specific for a certain estimator and (ii) specifics of the sample itself. Latter can be, for example, a systematic error of the observations and/or data pre-processing (e.g. selec-

tion effects) or a presence of real super-large structures inside the volume which can not be properly averaged. The samples can be simply tested for a presence such distortions by e.g. considering the standard number counts and tracking average density changes between different parts of the sample. In case of a generally homogenous distribution these quantities are very well predicted, so expected and actual values can be easily compared

The text below mostly summarizes the results of our previous published papers [17] and [17]. A refreshment is given in case of SDSS as the results for the newest data release are given. Moreover the tests of the samples' systematics are added with a brief discussion of the current data limitations.

2. Building volume limited subsamples

Before proceeding with an analysis of the two-point correlations in three-dimensional galaxy distributions one has to extract some volume limited (VL) samples (i.e. including all objects in a given range of radial distances for a given range of luminosities, see [13] for more details) out of the given survey data which is typically magnitude limited (i.e. includes all galaxies for a certain range of apparent magnitudes). A comparison between VL samples with various distance and magnitude limits, in different sky regions, allows one to test the statistical stationarity of galaxy distributions in these samples and to estimate the luminosity dependence of galaxy clustering.

To construct VL subsamples one first has to compute metric distances as

$$r(z) = \frac{c}{H_0} \int_{\frac{1}{1+z}}^{1} \frac{dy}{y \cdot \left(\Omega_M / y + \Omega_\Lambda \cdot y^2\right)^{1/2}}, \tag{1}$$

where we use the standard model parameters $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$. Then absolute magnitudes can be calculated as

$$M = m - 5 \cdot \log_{10}\left[r(z) \cdot (1+z)\right] - K(z) - 25, \tag{2}$$

where m is an apparent magnitude and $K(z)$ is a K-correction which can estimated by a certain method depending on particular galaxy catalog considered.

2.1 Two degree Field Galaxy Redshift Survey

The 2dFGRS survey project was completed in 2002 and the complete catalog data were made public in 2003 (the Final Release, available at http://www.mso.anu.edu.au/2dFGRS/). A detailed description of the volume limited samples extraction from this catalog is given in [17]. We select two rectangular regions on the celestial sphere near the northern galactic pole (NGP, $60^\circ \times 6^\circ$) and the southern galactic pole (SGP, $84^\circ \times 9^\circ$). Having metric distances and absolute magnitudes computed as Eq.(1) and Eq.(2) respectively, we then select three distance intervals and corresponding magnitude limits to obtain $3 \times 2 = 6$ VL subsamples whose main parameters are presented in table 1.

VL sample	$r_{\min}$	$r_{\max}$	$M_{\min}$	$M_{\max}$	Ω	N_g
SGP250	50	250	-19.5	-17.8	0.20	14177
SGP400	100	400	-20.8	-19.0	0.20	29373
SGP550	150	550	-21.2	-19.8	0.20	26289
NGP250	50	250	-19.5	-17.8	0.11	12474
NGP400	100	400	-20.8	-19.0	0.11	23208
NGP550	150	550	-21.2	-19.8	0.11	18030

Table 1. Properties of the 2dFRGS VL samples: $r_{\min}$ and $r_{\max}$ are the chosen limits for the metric distance; $M_{\min}$ and $M_{\max}$ are the corresponding limits for the absolute magnitude; Ω is the solid angle of a region in steradians; N_g – the resulting number of galaxies in each subsample.

2.2 Sloan Digital Sky Survey

The SDSS is currently the largest spectroscopic survey of the extragalactic objects and one of the most ambitious observational programs ever undertaken in astronomy. The data considered here is taken from the latest

public version of the data: Data Release Six (DR6, [20]) which can be accessed at http://www.sdss.org/dr6/. Extraction of the VL samples from the catalog is done similarly to the procedure described in section 2.1 and analogously to the data preparation explained in [18], while main features and differences are highlighted below.

There are two independent parts of the galaxy survey in the SDSS: the Main Galaxy (MG) sample and the Luminous Red Galaxy (LRG) sample: here both are analyzed. The former consists of the objects similar to the ones targeted in the 2dFGRS survey, being magnitude limited in r filter as $r < 17.77$. At the same time LRG sample galaxies are selected on a basis of color and magnitude to yield a sample of luminous intrinsically red galaxies that extends fainter and farther than the SDSS main galaxy sample (see [21] for details). Instead of typical magnitude limited scenario these galaxies are selected using a variant of the photometric redshift technique and form approximately volume-limited sample of objects in the range of redshifts $0.16 < z < 0.38$ with additional galaxies extending up to $z \approx 0.6$.

To select objects from the MG sample we first of all constrain the flags indicating the type of object so that we select only the objects from required sample. We then consider galaxies in the redshift interval $10^{-4} < z < 0.3$ and with the redshift confidence parameter $z_{conf} \geq 0.35$ having no redshift determination errors. Finally we apply $r < 17.77$ magnitude limit as it is required and thus select 479417 galaxies. For the luminous red galaxies we follow a similar procedure selecting objects of the appropriate type with the same redshift quality constraints in $0.16 < z < 0.6$ range, without magnitude limits; this results in 99799 matched galaxies.

The angular coverage of the survey is not uniform but observations have been done in different sky regions. We have considered three rectangular fields (named R1, R2 and R3) in the SDSS internal angular coordinates (η, λ) not taking into account completeness and fiber collision corrections (assuming completeness in the selected regions enough for our goals). The parameters of the chosen fields are reported in table 2.

Region	$\eta_{\min}$	$\eta_{\max}$	$\lambda_{\min}$	$\lambda_{\max}$	Ω
R1	-48.0	32.5	-6.0	36.0	0.94
R2	-54.0	-17.0	-33.5	-16.5	0.15
R3	-14.0	43.0	-36.0	-26.5	0.15

Table 2. Properties of the angular regions considered for the SDSS: the limits (in degrees) are chosen using the intrinsic coordinates of the survey (η, λ); Ω is the solid angle of each region in steradians.

Analogously to the 2dFGRS data processing metric distances and absolute magnitudes are computed as Eq.(1) and Eq.(2) respectively. We use Petrosian apparent magnitudes corrected for galactic absorption in r filter for MG sample and in g filter for LRG sample, the latter chosen in order to comply with a procedure described in [22]. The K-corrections for both analyzed samples are extracted from the latest release of the Value Added Galaxy Catalog (VACG, [23]). A simple data extraction procedure results in a successful retrieval of the K-corrections from VAGC for more than 98% of the selected galaxies, while for the rest a simple polynomial approximation of $K(z)$ based on extracted VAGC data is taken. We have considered 4 different VL cuts (defined by limits in absolute magnitude and metric distance) for main galaxies (named MG1, MG2, MG3 and MG4) and 1 VL cut for luminous red galaxies (named LRG). The limits defined analogously to 2.1 are reported in table 3.

VL cut	$r_{\min}$	$r_{\max}$	$M_{\min}$	$M_{\max}$
MG1	50	200	-21.0	-19.5
MG2	100	300	-22.0	-20.0
MG3	150	500	-22.5	-21.0
MG4	200	600	-23.0	-21.5
LRG	460	990	-22.4	-20.5

Table 3. Selected VL cuts for SDSS galaxies: $r_{\min}$ and $r_{\max}$ are the chosen limits for the metric distance; $M_{\min}$ and $M_{\max}$ are the corresponding limits for the absolute magnitude (r-band for MG1 – MG4 and g-band for LRG).

While MG1 and MG2 cuts actually contain relatively faint galaxies in the local universe, MG3 and MG 4 cuts cover a wide range of distances for main galaxies, and LRG cut consists of the bright red galaxies

at distances up to 1 Gpc/h. Considering three different rectangular areas, as a result we consider $5\times 3=15$ VL subsamples, the number of galaxies in each is reported in table 4.

Region / VL cut	R1	R2	R3
MG1	33259	5435	3891
MG2	43394	8765	9256
MG3	51717	9153	9324
MG4	30532	5061	5020
LRG	27150	4440	3957

Table 4. Number of galaxies in each VL subsample extracted from SDSS.

3. Two-point correlations

A number of statistical methods can be used to study galaxy distribution; the main ones involve the determination of two-point properties although the study of the higher order correlations, has also been applied to SDSS data by some researchers. The methods discussed below are the conditional density, which is a suitable method for highly correlated samples and the two-point reduced correlation function which is a widely used method for studying galaxy correlations. For a discussion of the small scale correlations characterization using the nearest neighbor distance probability distribution one may refer [17] and [17] for 2dFGRS and SDSS data analysis respectively.

3.1 Conditional density

The conditional density in spheres $\Gamma^*(r)$ is defined for an ensemble of realizations of a given point process as

$$\Gamma^*(r)=\frac{\langle N(r)\rangle_P}{\|C(r)\|}. \tag{3}$$

This quantity measures the average number of points $\langle N(r)\rangle_P$ contained in a sphere of volume $\|C(r)\|=\frac{4}{3}\pi r^3$ with the condition that the center of the sphere lies on an occupied point of the distribution (and $\langle ...\rangle_P$ denotes the conditional ensemble average). Such quantity can be estimated in a finite sample by a volume average (supposing stationarity of the point distribution)

$$\Gamma_E{}^*(r)=\frac{\overline{N(r)}_P}{\|C(r)\|}=\frac{1}{N_c(r)}\sum_{i=1}^{N_c(r)}\frac{N_i(r)}{\|C(r)\|}, \tag{4}$$

where $N_c(r)$ – the number of points (centers) with spheres fully contained in the sample volume, $\overline{(...)}_P$ means averaging by the sample points. Given a sample of an arbitrary geometry and a scale r at which correlations are measured, only a subsample of the points contained in it will satisfy the following requirement: when chosen as center of a sphere of radius r, the sphere is fully contained in the sample volume. When the average in Eq.(4) is made over such a subsample one considers the full-shell (FS) estimator of the conditional density.

The behavior of the conditional density in spheres is reported on Fig.1 – Fig.3. Volume limited samples with the same luminosity and distance cuts show approximately the same behavior for different angular regions (Fig.1, Fig.3). The slopes of $\Gamma^*(r)$ for 2dFGRS (Fig.1) show a power-law behavior for most of the scales without a crossover towards a constant value (the scales are limited by at most 40 Mpc/h due to a narrowness of the survey areas). The deviations from a power-law at smallest and largest scales can be explained by considering the limitations of the estimators which is done in section 0. Note that all results here are provided for the redshift space while a comparison of redshift space and real space approaches is given in [17] and [17].

The range of scales analyzed with the SDSS data at R1 region extends up to 200 Mpc/h (R1LRG sample). Conditional density estimation for SDSS (Fig.2) clearly shows a crossover towards a constant mode

which starts at 20-40 Mpc/h depending on VL subsample. A scale where $\Gamma^*(r)$ becomes a constant value (which can be interpreted as the evidence of homogeneity) ranges from 30 Mpc/h to 80 Mpc/h increasing for the deeper subsamples.

If one fits the behavior of the estimated $\Gamma^*(r)$, with a power-law function of type $B \cdot r^{\gamma}$ (in the range of scales where it conforms to a power-law) one finds that $\gamma = 0.8 \pm 0.2$ and $\gamma = 0.9 \pm 0.1$ for 2dFGRS and SDSS subsamples respectively. This corresponds to the values $D = 2.2 \pm 0.2$ and $D = 2.1 \pm 0.1$ of the metric dimension $D = 3 - \gamma$ which can be interpreted as fractal dimension if one considers a fractal model for the irregular structures (see [13] for an extensive discussion).

The conditional density in spherical shells $\Gamma(r)$ is defined as

$$\Gamma(r) = \frac{\langle N(r, dr) \rangle_P}{\|C(r, dr)\|}, \tag{5}$$

where $\langle N(r, dr) \rangle_P$ represents the ensemble average number of points in a sphere of radius r and thickness dr, of volume $\|C(r, dr)\| = 4\pi r^2 dr$, around a point of distribution (and thus this is a conditional ensemble average $\langle ... \rangle_P$ as in case of $\Gamma^*(r)$). Note that one can also write Eq.(6) as

$$\Gamma(r) = \frac{\langle n(r) n(0) \rangle}{\langle n(0) \rangle}, \tag{6}$$

where $\langle ... \rangle$ represents the (unconditional) ensemble average and $n(r)$ is the microscopic number density at distance r from a given point. The conditional density in shells can be estimated in a finite sample by the following volume average

$$\Gamma_E(r) = \frac{\overline{N(r, \Delta r)}_P}{\|C(r, \Delta r)\|} = \frac{1}{N_c(r + \Delta r)} \sum_{i=1}^{N_c(r+\Delta r)} \frac{N_i(r, \Delta r)}{\|C(r, \Delta r)\|}, \tag{7}$$

where we consider again only the full-shells, i.e. $N_c(r + \Delta r)$ represents the number of points (centers) contained in the spherical shells fully contained in the sample volume.

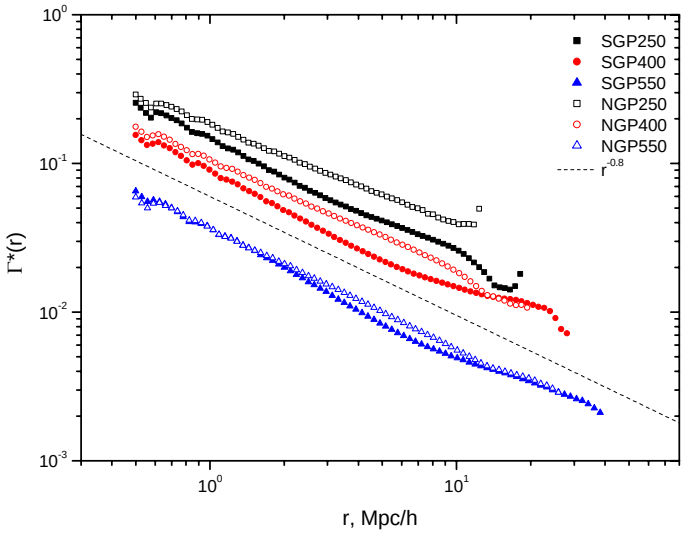

Fig.1: Conditional density in spheres for different VL subsamples of 2dFRGS (see labels). The reference line has a power-law behavior with slope $\gamma = 0.8$.

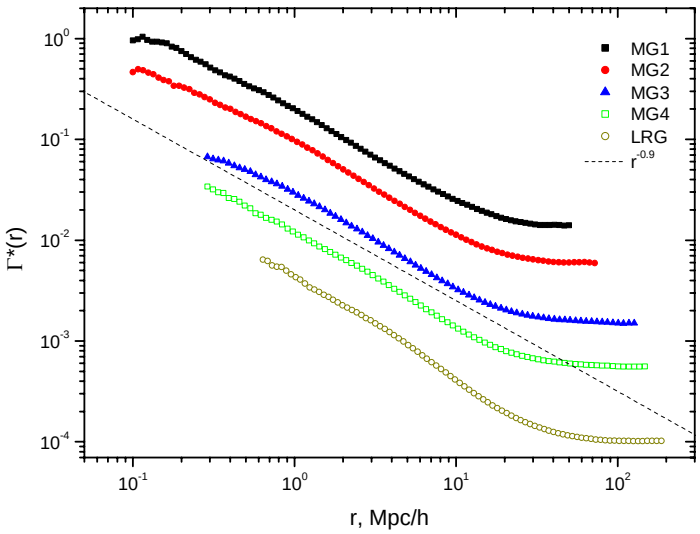

Fig.2: Conditional density in spheres for different SDSS VL subsamples in R1 angular region. The reference line has a power-law behavior with slope $\gamma = 0.9$.

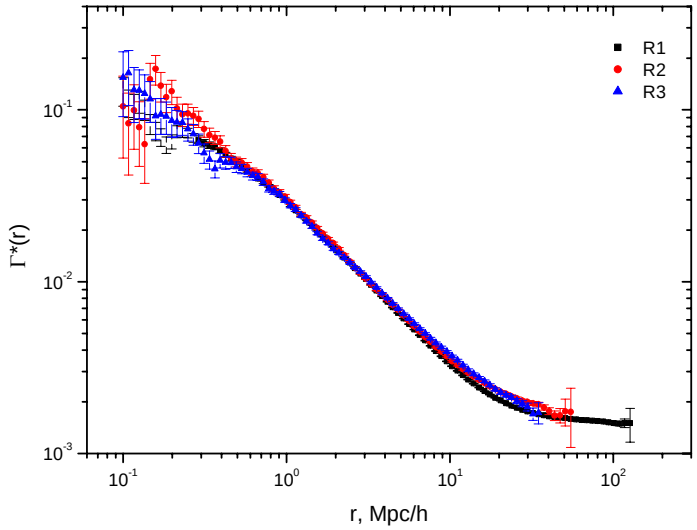

Fig.3: As for Fig.2 but MG3 distance-magnitude cut is considered and formal statistical errors are shown.

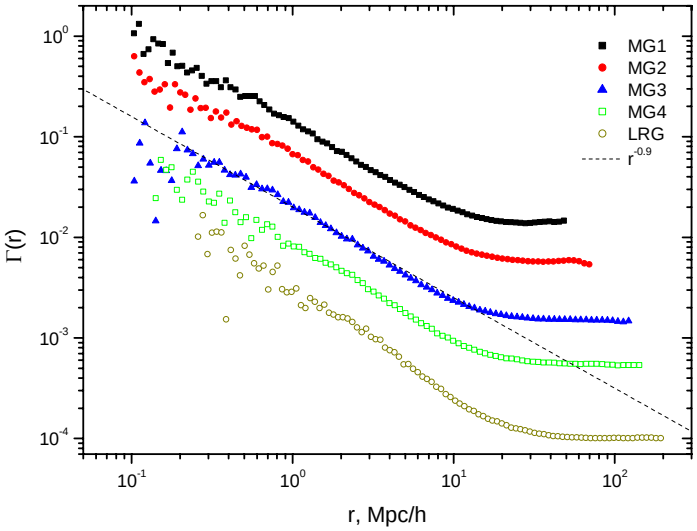

Fig.4: Conditional density in shells for different SDSS VL subsamples in R1 angular region. The reference line has a power-law behavior with slope $\gamma = 0.9$.

The estimation of the conditional density in spheres for SDSS VL samples in R1 region is shown on Fig.4. Basically the behavior of $\Gamma(r)$ is similar to $\Gamma^*(r)$ but clearly more affected by a statistical noise. Note that formal statistical errors (standard errors of the mean values) for the estimations of $\Gamma^*(r)$ and $\Gamma(r)$ are only plotted on Fig.3. Actually these errors appear to be quite small in comparison with systematic variations due to the fact that the volume average cannot be performed at large scales and sample to sample variance.

3.2 Reduced correlation function

The reduced two-point correlation function $\xi(r)$ for a stochastic point process is defined (see e.g. [24]) as

$$\xi(r) = \frac{\langle n(r)n(0)\rangle}{\langle n(0)\rangle^2} - 1 = \frac{\Gamma(r)}{\langle n\rangle} - 1, \tag{8}$$

where $\langle ...\rangle$ indicates the ensemble average and $\langle n\rangle$ is the ensemble average number density. The last equality follows from the definition of the conditional density (see Eq.(6)).

There are several estimators of $\xi(r)$ (see [25] for a detailed discussion). Analogously to the full-shell estimator of the conditional density, one may define the full-shell estimator $\xi_{FS}(r)$ which gives a very conservative estimation of the correlation function, its application to the 2dFGRS data is discussed in [17]. An estimator considered below is the Landy & Szalay (LS) estimator that is the most widespread in modern studies of correlation function for large scale structures as it has the minimal variance estimator for a Poisson distribution. This can be written as:

$$\xi_{LS}(r) = \frac{N_R(N_R-1)}{N_D(N_D-1)}\frac{DD(r)}{RR(r)} - 2\frac{N_R-1}{N_D}\frac{DR(r)}{RR(r)} + 1, \tag{9}$$

where N_D is the number of data (sample) points; N_R – the number of random points homogeneously distributed in the sample geometry; $DD(r)$ is the number data-data pairs, $DR(r)$ – data-random pairs and $RR(r)$ – random-random pairs. We evaluate $\xi(r)$ with LS estimator using artificial random catalogs with a number of points N_R about three times more than N_D.

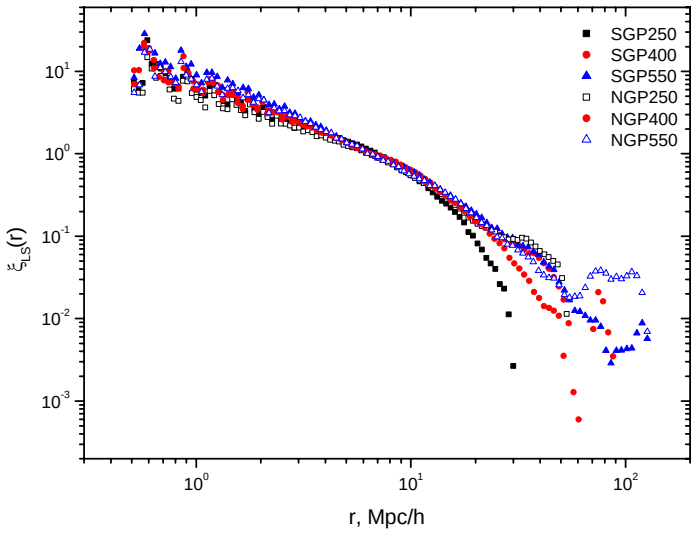

Fig.5: Two-point correlation function for different VL subsamples of 2dFRGS using Landy & Szalay estimator.

The behavior of the reduced two-point correlation function evaluated with LS estimator for the 2dFGRS VL subsamples is presented on Fig.5. The amplitude of $\xi(r)$ appears to be similar for all samples in contrast to the full-shell estimator as shown in [17]. While at small scales $\xi(r) >> 1$ and the behavior similar to $\Gamma(r)$ i.e. power-law is observed (which follows from Eq.(8)), at larger scales $\xi(r)$ decays rapidly

and crosses zero. The values of characteristic scales r_0 and r_{zc} such that $\xi(r_0)=1$ and $\xi(r_{zc})=0$ depend on the sample analyzed; this can be explained as a systematic effect of the finite volume in each sample. If one fits $\xi(r)$ with a power law at scales where $\xi(r)>>1$ one should obtain a value of $\gamma\approx 0.8\div 0.9$. However at the scales around r_0 an exponent value $\gamma\approx 1.7\div 1.8$ is typically obtained, see [17] for more details.

3.3 Differences and limitations of methods

The main difference between two approaches of studying two-point correlations described above is that estimation of the correlation function actually requires a reliable estimation of the ensemble average density $\langle n\rangle$ as follows from Eq.(8). This can be easily achieved in case of weakly correlated distributions, i.e. homogenous with weak fluctuations. For the intrinsically irregular distributions (such as fractals) a sample average value $\bar{n}$ can not be a reliable estimator of $\langle n\rangle$ and strongly depends on the volume being analyzed. Moreover it was shown in [13] that some of the estimators of $\xi(r)$ such as $\xi_{LS}(r)$ can display a false tendency of a crossover towards homogeneity in case of the model irregular samples with a non-spherical shape. Therefore conditional density functions $\Gamma^*(r)$ and $\Gamma(r)$ estimated using the full-shell estimator are in general more suitable for the characterization of galaxy distribution in the regime of strong clustering (one may refer to [13] for a comprehensive discussion).

An important consequence following from Eq.(8) is that the functions $\xi(r)$ and $\Gamma(r)$ can not have a power-law behavior for the same sample. When performing a power-law fitting at different scales we find that for most of the VL subsamples extracted from 2dFGRS and SDSS it is the conditional density (in spheres and spherical shells) that shows a power-law behavior with almost constant exponent at the scales of strong clustering. The same fitting applied to the correlation function (locally) shows a systematic increase of the exponent when the scale r increases. For example Hawkins et al. [8] studied 2dFGRS data in redshift space with $\xi(r)$ and found that in the full magnitude limited value of the correlation exponent is $\gamma=0.75$ in the range $[0.1, 4]$ Mpc/h and then $\gamma=1.75$ in the range [4,10] Mpc/h (see their Fig.6).

The maximum scale available for the FS estimator (for $\Gamma^*(r)$, $\Gamma(r)$ and $\xi(r)$) is limited by the maximum radius of the galaxy-centered sphere r_s^m fully contained in the sample volume. The minimal scale up to which correlations can be reliably measured by all considered estimators is given by the average distance between neighbor galaxies r_{sep}: clearly for smaller scales discrete shot-noise dominates estimations of any statistical quantity. The values of r_{sep} for the VL subsamples considered above are about $1\div 10$ Mpc/h while r_s^m ranges as $10\div 200$ Mpc/h.

Note that for the full-shell estimators of $\Gamma^*(r)$ and $\Gamma(r)$ the number of centers $N_c(r)$ is a function of the scale r at which correlations are estimated. At scales approaching to r_s^m only the points lying close to the outer boundary of the sample will be considered as centers (due to the specifics of the shape of analyzed samples). One can define various characteristics in addition to $N_c(r)$ to check statistical relevancy of $\Gamma^*(r)$ and $\Gamma(r)$ determination at certain scale. It is shown in [17] that one can rely on results at scales 2-3 times smaller than r_s^m. Therefore finite size effects such as presence of the super-large structures become especially important for the large scales (e.g. a large cluster or void near the sample's outer boundary). A couple of very simple tests for a presence of such structures in the considered samples is briefly discussed below.

4. Tests for structures and sample effects

To verify whether a considered VL sample is statistically uniform as a whole (i.e. that the statistical properties are more or less the same in each part of the sample) one may use various methods. In case of a homogenous sample (i.e. a sample where all fluctuations are well averaged at large scales) one knows what exactly to expect for various tests and can simply compare expected values with actual results for a real sample. Here we apply two simple tests to the studied samples: the differential number counts by radial distance

and the distribution of the average density along radial distance. These methods could show over-density or under-density regions at scales comparable to a sample's size.

4.1 Differential number counts

The behavior of the differential number counts $dN(r)/dr$ as a function of distance for different VL samples of the 2dFGRS catalog is presented on Fig.6. As an example the best fit for the SGP400 volume limited sample is reported, which shows an exponent corresponding to a metric dimension $D = 3.7$ larger than the space dimension ($D = 3.0$). This can be interpreted a purely finite-size effect corresponding to the large fluctuations still visible at scales of the order of 100 Mpc/h. Similar fit for the R1MG3 sample of SDSS also gives a metric dimension larger than the space dimension ($D = 3.3$). The counts up to certain apparent magnitude $N(m)$ applied to 2dFGRS and SDSS magnitude limited catalogs also yield to the same conclusion.

An important question is whether the observed systematics are due to the real structures or it is produced by some selection effects or errors in the original data of the catalogs. This question is yet to be answered and is beyond the of the scope of the current text.

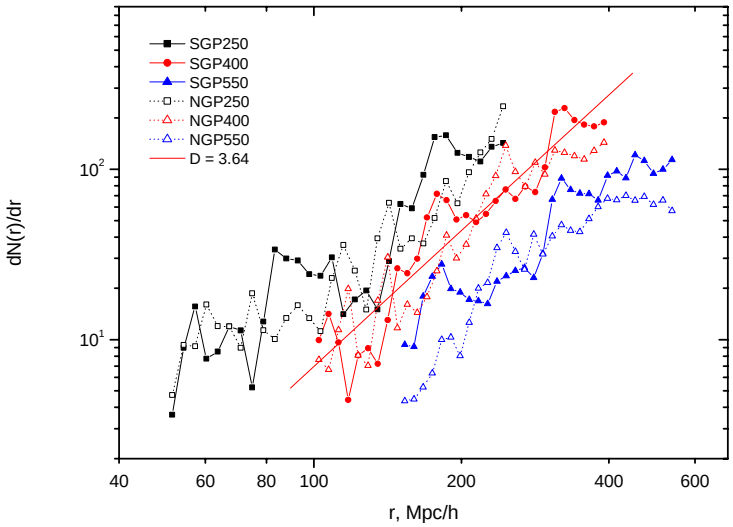

Fig.6: Differential number counts for VL subsamples of 2dFRGS. Best fit for SGP400 sample is shown.

4.2 Radial density distribution

A simple method clearly showing local properties of the various regions in a sample is a comparison between the values of the average density computed in each region. Here we consider the following simple test: a volume limited sample is divided into several slices along the line of sight and an average density $\bar{n}_{rad}(r)$ is computed in each slice as

$$\bar{n}_{rad}(r,\Delta r) = \frac{\Delta N(r,\Delta r)}{\Delta V(r,\Delta r)} = \frac{\Delta N(r,\Delta r)}{\frac{\Omega}{3}\cdot\left((r+\Delta r)^3 - r^3\right)}, \qquad (10)$$

where $\Delta N(r,\Delta r)$ is the number of galaxies in the slice delimited by radial distances r and $r+\Delta r$, $\Delta V(r,\Delta r)$ – the volume of this slice and Ω – the solid angle of the sample. While the volume of a slice increases along with the radial distance, it's obvious that in case of uniform distribution one should not see the systematical changes of density at large scales.

Examples of the radial density distributions are presented on Fig.7 and Fig.8 for some of the VL samples of 2dFGRS and SDSS respectively. One can clearly see that in 100-400 Mpc/h distance interval (VL400 cut) for 2dFGRS the value of density differs 2-3 times with a scale of fluctuations about 100 Mpc/h. Moreover all three independent angular regions of SDSS show the same systematical 3 times growth of the density towards outer boundary of the MG4 subsamples (2dFGRS regions are located at different poles of the spherical coordinate system thus showing no correlation with each other). These results lead us to a conclusion that the samples are not uniform at large scales comparable to their sizes.

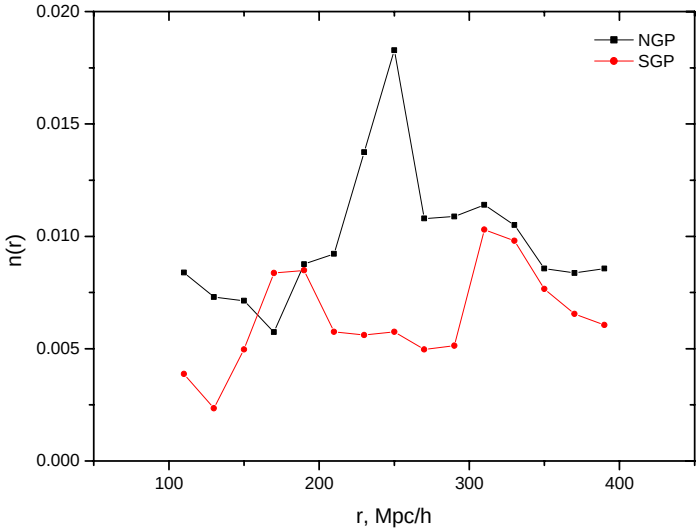

Fig.7: Radial density profile for various angular regions of 2dFRGS reported in 100-400 Mpc/h interval.

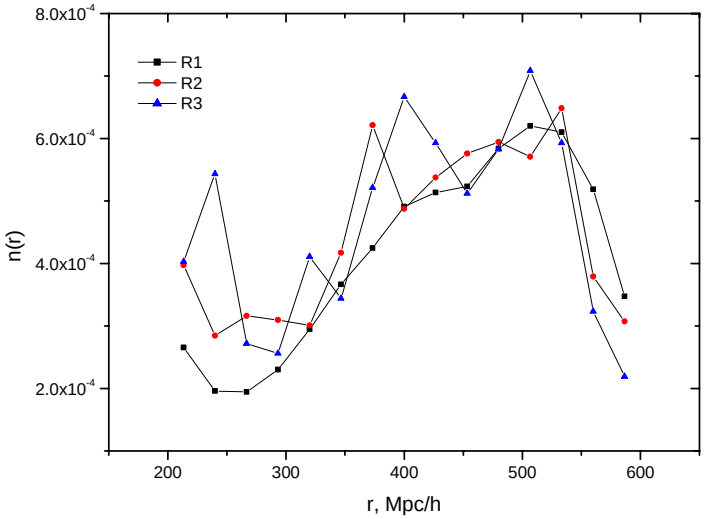

Fig.8: Radial density profile for various angular regions of SDSS reported in 200-600 Mpc/h interval.

5. Discussion and conclusions

We have studied redshift space correlation properties of several volume limited samples extracted from the final release of 2dFGRS survey and the sixth data release of SDSS survey. We characterize correlations in volume limited subsamples with a conservative two-point statistics – the conditional density which appears to be well fitted with a power-law in the range of scales approximately from 0.5 Mpc/h up to 30 Mpc/h (exact range varies from sample to sample). The power-law exponent γ value ranges from 0.7 to 1.1 having average values of $\gamma = 0.8 \pm 0.2$ for 2dFGRS and $\gamma = 0.9 \pm 0.1$ in case of SDSS. Extending two-point correlation analysis to the scales of 100-200 Mpc/h with the latest SDSS data we observe almost constant behavior of the conditional density. This can be interpreted as an evidence of the detection of homogeneity scale, still one has to be sure that the analyzed samples are uniform at large scales to make such a conclusion. The two-point correlation function, which is commonly used as a two-point statistics, does not show a power-law behavior with a constant exponent γ.

While it is claimed in [15] and [16] that galaxy distribution is homogenous starting from at most 70 Mpc/h we put several arguments to show insufficient motivation of such strong statement. First of all it is noted here as previously in [17] and [17] that current surveys still do not provide continuous large samples to make proper averages for the scales around 100 Mpc/h without including some a priori artificial assumptions into the analysis procedure. Then by performing the differential number counts $dN(r)/dr$ and computing

the radial density $\overline{n}_{rad}(r)$ we find strong evidences of the existence of super-large scale structures which appear as 2-4 times density contrast and are comparable in size with the samples they belong to (more than 100 Mpc/h). This density contrast simply invalidates the statement of homogeneity detection for these scales.

Both 2dFGRS and SDSS data demonstrate significant variations of density at very large scales which can be interpreted as catalogs' errors or real structures (the latter being more probable). These effects obviously can not be properly averaged in the estimations of the conditional density and the two-point correlation function. Analysis of the possible biases which can be induced by these structures will be a subject of a forthcoming paper. Furthermore it's becoming more and more clear that we still need deeper samples with a larger solid angles to give a trustworthy answer to the question of scale of true homogeneity for the visible matter distribution.

References

[1] J.A. Peakock, Cosmological Physics, Cambridge Astrophysics, (2001)

[2] J.R. Gott III et al., Astrophys.J., **624**, 463, (2005)

[3] R. Massey et al., Astrophys.J. Supp., **172**, 239, (2007)

[4] L. Rudnick et al., astro-ph/0704.0908v2, (2007)

[5] D.M. Coward et al., astro-ph/0711.0242, (2008)

[6] M. Colless et al. (The 2dFGRS team), astro-ph/0306581, (2003)

[7] D.G. York et al. (The SDSS Collaboration), Astron.J., **120**, 1579, (2000)

[8] E. Hawkins et al. (The 2dFGRS team), MNRAS, **346**, 78, (2003)

[9] D.S. Madgwick et al. (The 2dFGRS team), MNRAS, **344**, 847, (2003)

[10] T. Okumura et al., Astrophys.J., **676**, 889, (2008)

[11] N.P. Ross et al., MNRAS, **381**, 573, (2007)

[12] W.J. Percival et al., Astrophys.J., **657**, 645, (2007)

[13] A. Gabrielli, F. Sylos Labini, M. Joyce, L. Pietronero, Statistical physics for cosmic structures, Springer, (2005)

[14] L. Pietronero, Physica A, **144**, 257, (1987)

[15] D.W. Hogg et al., Astrophys.J., **624**, 54, (2005)

[16] A.V. Tikhonov, Astron.Lett., **32**, 721, (2006)

[17] N.L. Vasilyev, F. Sylos Labini, Yu.V. Baryshev, Astron.Astrophys., **447**, 431, (2006)

[18] F. Sylos Labini, N.L. Vasilyev, Yu.V. Baryshev, Astron.Astrophys., **465**, 23, (2007)

[19] D.S. Madgwick, O. Lahav et al., MNRAS, **333**, 133, (2002)

[20] J.K. Adelman-McCarthy et al., Astrophys.J. Supp., **175**, 297, (2008)

[21] D.J. Eisenstein et al., Astron.J., **122**, 2267, (2001)

[22] I. Zehavi et al., Astrophys.J., **608**, 16 (2004)

[23] M.R. Blanton et al., Astron.J., **129**, 2562, (2005)

[24] P.J.E. Peebles, The Large-Scale Structure of the Universe, Princeton Univ. Press, (1980)

[25] M. Kerscher, I. Szapudi, A.S. Szalay, Astrophys.J., **535**, L13, (2000)