Persistent fluctuations in the distribution of galaxies from the Two-degree Field Galaxy Redshift Survey

F. Sylos Labini, N. L. Vasilyev and Y. V. Baryshev

1 Museo Storico della Fisica e Centro Studi e Ricerche Enrico Fermi - Piazzale del Viminale 1, 00184 Rome, Italy, EU
2 Istituto dei Sistemi Complessi CNR - Via dei Taurini 19, 00185 Rome, Italy, EU
3 Institute of Astronomy, St. Petersburg State University - Staryj Peterhoff, 198504, St. Petersburg, Russia

received 17 November 2008; accepted in final form 17 December 2008
published online 2 February 2009

PACS 98.80.-k – Cosmology
PACS 05.40.-a – Fluctuation phenomena, random processes, noise, and Brownian motion
PACS 02.50.-r – Probability theory, stochastic processes, and statistics

Abstract – We apply the scale-length method to several three-dimensional samples of the Two-degree Field Galaxy Redshift Survey. This method allows us to map in a quantitative and powerful way large scale structures in the distribution of galaxies controlling systematic effects. By determining the probability density function of conditional fluctuations we show that large-scale structures are quite typical and correspond to large fluctuations in the galaxy density field. We do not find a convergence to homogeneity up to the samples sizes, i.e. $\approx 75 \text{Mpc}/h$. We then measure, at scales $r \lesssim 40 \text{Mpc}/h$, a well-defined and statistically stable power-law behavior of the average number of galaxies in spheres, with fractal dimension $D = 2.2 \pm 0.2$. We point out that standard models of structure formation are unable to explain the existence of the large fluctuations in the galaxy density field detected in these samples. This conclusion is reached in two ways: by considering the scale, determined by the linear perturbation analysis of a self-gravitating fluid, below which large fluctuations are expected in standard models and through the determination of statistical properties of mock galaxy catalogs generated from cosmological $N$-body simulations of the Millenium consortium.

Copyright © EPLA, 2009

Introduction. – In the past twenty years observations have provided growing evidences that galaxy distribution is organized in a complex network of structures and voids [1–4]. Despite the fact that large-scale galaxy structures, of size of the order of several hundreds of Mpc/$h^3$, have been observed to be the typical feature of the distribution of visible matter in the local universe, the statistical analysis measuring their properties has identified a characteristic scale which has only slightly changed since its discovery forty years ago in angular catalogs. This scale, $r_0$, was measured to be the one at which fluctuations in the galaxy density field are about twice the value of the sample density and it was indeed determined to be $r_0 \approx 5 \text{Mpc}/h$ in the Shane and Wirtanen angular catalog [5]. Subsequent measurements of this scale — see, e.g., [6–13] — found a similar value, although in several samples larger values of $r_0$ have been found (i.e. $r_0 \approx 6–12 \text{Mpc}/h$). This variation was then ascribed to a luminosity-dependent effect — see, e.g., [7–9,12].

However, recently in a CCD survey of bright galaxies within the Northern and Southern strips of the 2dF Galaxy Redshift Survey (2dFGRS) [3] conclusive evidences were found that there are fluctuations of the order $\sim 30\%$ in galaxy counts as a function of apparent magnitude [14] (see also [15,16] for similar observations in other galaxy samples). Further, since in the angular region toward the Southern Galactic Cap (SGC) a deficiency, with respect to the Northern Galactic Cap (NGC), in the counts below magnitude $\sim 17$ (in the $B$ filter) was found, persisting over the full area of the APM and APMBGC catalogs, this would be an evidence that there is a large void of radius of about $150 \text{Mpc}/h$ implying that there are spatial correlations extending to scales larger than the scale detected by the 2dFGRS correlation function [10,11]. Indeed, by considering the two-point correlation function,
and thus by normalizing the amplitude of fluctuations to the estimation of the sample density, the length scale \( r_0 \approx 6-8 \text{ Mpc}/h \) was derived \[10,11\].

Structures and fluctuations at scales of the order of 100 Mpc/h or more are at odds with the prediction of the concordance model of galaxy formation \[14-16\], while the small value of the correlation length is indeed compatible. In what follows we try to clarify this puzzling situation, i.e. the coexistence of the small typical length scales measured by the two-point correlation function analysis with the large fluctuations in the galaxy density field on large scales as measured by the simple galaxy counts. Because of the difference in the counts amplitude, and thus in the sample density between the NGC and the SGC samples, the estimation of the sample density is not stable and thus one must critically consider the significance of the normalization of fluctuations amplitude to the estimation of the sample density as used in the correlation analysis employed to measure the length scale \( r_0 \).

More generally the problem of the statistical characterization of these structures in a finite sample, of volume \( V \) containing \( M \) galaxies, can be rephrased as the problem of measuring volume-averaged statistical quantities. The basic issue concerns whether these are meaningful descriptors, i.e. whether they give or not stable statistical estimations of ensemble-averaged quantities \[17\]. In general it is assumed that galaxy distribution is an ergodic stationary stochastic process \[17\], which means that it is statistically translationally and rotationally invariant, thus avoiding special points or directions. Stationary stochastic distributions satisfy these conditions also when they have zero average density in the infinite volume limit \[17\]. The assumption of ergodicity implies that in a single realization of the microscopic number density field \( n(\vec{r}) \) the average density \( n_0 \) in the infinite volume is well defined and equal to the ensemble average density \[17\]. The constant \( n_0 \) is strictly positive for homogeneous distributions and it is zero for infinite inhomogeneous ones \[17\]. The infinite volume limit must be considered in the definition of probabilistic properties, but in physical systems one is concerned only with finite volumes and statistical determinations. For inhomogeneous distributions, in a finite sample, the estimation of the average mass density gives a large relative error with respect to the ensemble value and it is thus systematically biased \[17\]. This situation occurs as long as the sample size is smaller than the scale \( \lambda_0 \) at which the distribution turns to homogeneity, i.e. beyond which density fluctuations are small \[17\]. In the finite sample analysis it is then necessary to study the conditional scaling properties of statistical quantities, by an analysis of fluctuations and correlations which explicitly considers whether a distribution can be or not homogeneous. Before turning to the description of the methods employed to study galaxy distributions and mock galaxy samples, we discuss the properties of the samples considered.

**The data.** – The Two-degree Field Galaxy Redshift Survey (2dFGRS)\(^2\) \[3\] measured redshifts for more than 220,000 galaxies in two strips in the Southern Galactic Cap (SGC) and in Northern Galactic Cap (NGC). The median redshift is \( z \approx 0.1 \) and the apparent magnitude corrected for galactic extinction in the \( b_j \) filter is limited to 14.0 < \( b_j < 19.45 \). The selection of the samples used in the analysis discussed below is described in \[18\]. To avoid the effect of the irregular edges of the survey, we selected two rectangular regions whose limits are: for SGC: 84° × 9° (−33° < \( \delta < −24° \), −32° < \( \alpha < 52° \)) and for NGC: 60° × 6° (−4° < \( \delta < 2° \), 150° < \( \alpha < 210° \)). We construct Volume Limited (VL) samples, which are unbiased for the observational selection effects due to the limit in apparent magnitude \[12\]. To this aim, we computed the metric distance \( R(z) \) with parameters \( \Omega_M = 0.3 \) and \( \Omega_\Lambda = 0.7 \) (i.e. the concordance model) and determined the absolute magnitudes \( M \) using K-corrections from \[19\]. Two couples of VL samples, in each galactic cap, are identified by i) 100 Mpc/h < \( R < 400 \text{ Mpc}/h \) and −19.0 < \( M < −20.8 \) (SGC400 and NGC400) and ii) 150 Mpc/h < \( R < 550 \text{ Mpc}/h \) and −19.8 < \( M < −21.2 \) (SGC550 and NGC550). Each sample contains about \( N \approx 2-3 \times 10^4 \) galaxies \[18\].

**Statistical methods.** – The Scale Length (SL) analysis \[20\] consists in the determination of the number \( N(r; R_i) \) of galaxies in spheres of radius \( r \), centered on the \( i \)-th galaxy at the radial distance \( R_i \) from the observer. When this is averaged over the whole sample it gives an estimate of the average conditional number of galaxies in spheres of radius \( r \) \[17,20\]:

\[
\overline{N(r)} = \frac{1}{M(r)} \sum_{i=1}^{M(r)} N(r; R_i),
\]

where the sum is extended to the \( M(r) \) galaxies whose distance from the boundaries of the sample is smaller than or equal to \( r \). In this way when \( r \) growths \( M(r) \) decreases with \( r \) because only those galaxies for which the sphere is fully included in the sample volume are considered as centers \[17\]. In addition when \( r \) is large enough only a part of the sample is explored by the volume average \[17,20\]. Thus for large sphere radii \( M(r) \) decreases and the location of the galaxies contributing to the average in eq.(1) is mostly at radial distance \( \sim [R_{\text{min}} + r, R_{\text{max}} - r] \) from the boundary of the sample at \( [R_{\text{min}}, R_{\text{max}}] \). By using these boundary conditions eq.(1) gives the so-called full-shell estimator \[17,21\]. This has the advantage to make the weakest \textit{a priori} assumptions about the properties of the distribution outside the sample volume. Indeed one may use incomplete spheres, by counting the galaxies inside a portion of a sphere and by weighting this for the corresponding volume \[21\]. However, this method implicitly assumes that what is inside the

\(^2\text{http://www2.aao.gov.au/2dFGRS/}.\)
Persistent fluctuations in the distribution of galaxies

Incomplete sphere is a statistically meaningful estimate of the whole spherical volume. This is incorrect when a distribution presents large fluctuations. For example in the part of a spherical volume which lies outside the sample boundaries there can be an empty region or a large scale structure: in this situation the weighted estimate is biased [17]. In a finite sample, together with the average given by eq. (1), one may determine amplitudes of density fluctuations by measuring their variance \( \delta(r)^2 \equiv [N(r)^2 - \bar{N}(r)^2] / \bar{N}(r)^2 \sim 1 \), where the last equality holds for inhomogeneous distributions (e.g., for a fractal \( N(r) \sim r^D \) and \( D < 3 \)) and it means that fluctuations are persistent [17]. In such a situation, because of the strong correlations, the Central-Limit Theorem does not hold and the Probability Density Function (PDF) of fluctuations does not generally converge to a Gaussian function as for homogeneous ones [17], where \( \delta(r)^2 < 1 \) [17].

Thus, galaxy distribution in these samples are dominated by several large-scale structures which cross their volumes. These structures are typical, i.e. they are detected at different radial distances and in two different sky areas of the 2dFGRS. They correspond to large density fluctuations, i.e. large variations of \( N(r; R_i) \). For the largest sphere radius we considered is \( r \approx 40\, \text{Mpc}/\text{h} \), we find fluctuations of order four in \( N(r; R_i) \). This implies that \( \lambda_0 > 40\, \text{Mpc}/\text{h} \).

Convergence to homogeneity? – We can now use the data obtained by the SL analysis to investigate whether there is a convergence to homogeneity at some large scales \( r > 40\, \text{Mpc}/\text{h} \). This is achieved by dividing the whole range of radial distances in bins of thickness

Fig. 1: Left panels: from top to bottom the SL analysis for the different 2dFGRS samples and \( r = 5, 10\, \text{Mpc}/\text{h} \) (NGC400) and \( r = 20, 30\, \text{Mpc}/\text{h} \) (SGC550). Right panels: probability density function \( f(N, r) \) of \( N(r; R_i) \) in the whole sample (thick solid line) and in two non-overlapping sub-samples with equal volume (each half of the sample volume) at small (think solid line) and large (dashed line) \( R_i \).
The geometry of the two-point correlation function can be written as [17]

$$\xi(r) + 1 = \frac{dN(r)}{4\pi r^2 dr} \frac{1}{n_S},$$

(2)

where $n_S$ is the sample density. If $N(r) \propto r^D$, then $\xi(r)$ has the following features: i) its amplitude is proportional to the sample size and ii) it shows a break from a power law at a scale of the order of the sample size. The amplitude of $\xi(r)$ is then a ratio between a local and a global quantity ($n_S$). The former one can be estimated for instance as $n_S = N/V$. When $V$ is spherical of radius $R_e$, we get that $\xi(r_0) = 1$ for $r_0 = (D/6)^{1/(3-D)}R_e$. The geometry of the 2dFGRS samples is a spherical portion for which the radius of the maximum sphere fully enclosed is about $R_e \approx 40$. Given that $D \approx 2$, we get $r_0 \approx 10$ Mpc/h, which is approximately the value obtained by [10,11]. The normalized mass variance is equal to unity at approximately the same scale [17]. A more detailed discussion of the determination of the two-point correlation function can be found in [18,24].

Comparisons with standard models of galaxy formation. – In cosmological-structure formation cold-dark-matter models [25] gravitational collapse firstly forms non-linear structures (i.e. large fluctuations) at small scales and then larger and larger scales become non-linear. The theoretical homogeneity scale $\lambda_0^m$ identifying the range of distances where large inhomogeneities are formed, can be defined to be such that the unconditional relative mass variance in spheres is $\sigma^2(\lambda_0^m) = 1$ [26]. Thus from the time dependence of the power spectrum in the linear perturbation analysis of the self-gravitating fluid equations it is possible to derive the time-dependence of $\lambda_0^m(t)$ [26]. By normalizing the initial amplitude of density fluctuations to the Cosmic Microwave Background Radiation (CMBR) anisotropies it is found that at the present time, in the concordance model, $\lambda_0^m \approx 10$ Mpc/h [27–29].

This estimation is in agreement with results of cosmological $N$-body simulations which are used to study the non-linear regime for $r < \lambda_0^m$. Here we considered the cosmological simulations performed by the Millennium project which are the largest ones performed until now [29]. The amount of dark matter and cosmological

radii. This is due to fact that, for large sphere radii, the volume average cannot explore properly the full sample because of the geometrical selection effect present in the determination of $N(r; R_c)$ and it is dominated by only few structures. The determination of the whole-sample average statistics, i.e. eq. (1), provides a meaningful statistical quantity as the PDF in all samples is reasonably statistically stable. We find $N(r) \propto r^D$ with $D = 2.2 \pm 0.2$ up to $r \sim 40$ Mpc/h (fig. 2) in agreement with previous determinations by [18].

Estimation of the standard two-point correlation function. – The estimator of the two-point correlation function can be written as [17]
parameters are given in agreement with the standard concordance models. Dark-matter simulations have about $10^{10}$ particles and galaxies are identified according to semi-analytics models of galaxy formation [30]. We used a mock galaxy catalog with about 9 millions objects, where absolute magnitudes of mock galaxies can be transformed in the same filter $b_J$ of the 2dFGRS [30]. We have cut samples with almost same geometry, number of objects and limits in magnitude and distance as the 2dFGRS samples. Note that we computed $N(r; R_i)$, where $R$ is the real-space position of each object with respect to the observer: in the redshift space the dimension for $r < 10$ Mpc/h is slightly different [24]. We find (fig. 2) that $\bar{N}(r) \sim r^{1.2}$ for $r < 10$ Mpc/h and $\bar{N}(r) \sim r^3$ for $r > 10$ Mpc/h. The PDF of conditional fluctuations rapidly converges to a Gaussian for $r > 10$ Mpc/h and it is statistically stable (fig. 3). Correspondingly $N(r; R_i)$ shows different and more quiet fluctuations than the real data. Note that at scales $r > 10$ Mpc/h density fluctuations in the dark-matter field are in the linear regime and thus the understanding of biasing in that case is simple. Indeed, according to the simple threshold sampling a Gaussian field [31] biasing is linear when fluctuations are small and Gaussian [27]. In addition only non-local biasing mechanisms, which at the moment have not been explored in the literature, could possibly produce large-scale density fluctuations of the kind observed in the galaxy distribution.

**Discussion.** – In summary, by applying the SL method to the 2dFGRS samples we detect large density fluctuations of considerable spatial extension. At scales $r \lesssim 40$ Mpc/h we find statistically stable power law correlations with fractal dimension $D = 2.2 \pm 0.2$ in agreement with previous determinations [17,18,20,32–34]. For $r > 40$ Mpc/h we find that the galaxy distribution is strongly inhomogeneous and fluctuations are large up to the samples sizes, in agreement with a similar analysis of the SDSS data [20]. Persistent large-scale density fluctuations are compatible [17] with fractal power law correlations extending to scales $r > 40$ Mpc/h, but incompatible with homogeneity at $\lambda_0 \lesssim 75$ Mpc/h. On the other hand, standard models of galaxy formation, normalized to CMBR anisotropies, predict $\lambda_0^m \approx 10$ Mpc/h [26,29], i.e. smaller than our lower limit $\lambda_0 > 75$ Mpc/h. This prediction is in agreement with the results we found in mock galaxy catalogs where we measured that, fluctuations are smoother than in the 2dFGRS samples, and their PDF rapidly converges to a Gaussian function for $r > 10$ Mpc/h.

Our results are in contrast with the standard determinations that the characteristic length scale of galaxy distribution, marking the transition to the regime of small fluctuations, is of the order of $10$ Mpc/h [11–13]. This is because this length scale is derived by measuring the amplitude of two-point correlation function $\xi(r)$. When considering this quantity, which is normalized to the estimation of the sample density, it is implicitly assumed that the distribution is homogeneous (i.e. with small amplitude fluctuations) well inside the sample volume, i.e. $\lambda_0 \ll V^{1/3}$ [17]. When fluctuations are large, as in the case of the 2dFGRS samples, this descriptor is systematically biased by finite size effects [17,18,20] and so is the characteristic length scale derived from its amplitude. On the other hand, our results fairly agree with studies of galaxy counts as a function of the apparent magnitude $\mathcal{N}(m)$, which indirectly probe radial distance fluctuations. These show large fluctuations around the average behavior: particularly $\mathcal{N}(m)$ in the SGC are down by 30% relative to the NGC counts [14]. These behaviors can be now directly related to large scale galaxy structures and particularly to the fact that in the NGC samples there are more structures, and thus an higher amplitude of $\bar{N}(r; R, \Delta r)$ (fig. 2), than in SGC samples.

Finally it is worth noticing that our results agree with the conclusion of [23] who found that large-scale structures (e.g. super-clusters) are more frequent in observed samples than in the simulations.

***

We thank M. Joyce, A. Gabrielli, L. Pietronero for discussions and M. Blanton and M. Lopez-Corredoira for comments. YVB thanks the partial financial support by the Russian Federation grants “Scientific school” and “Russian Education”. We acknowledge the use of the 2dFGRS data [3] and of the Millennium simulations [30].
REFERENCES

[1] de Vaucouleurs G., Science, 167 (1970) 1203.
[2] de Lapparent V., Geller M. J. and Huchra J. P., Astrophys. J., 302 (1986) 1.
[3] Colless M. et al., Mon. Not. R. Acad. Soc., 328 (2001) 1039.
[4] York D. et al., Astron. J., 120 (2000) 1579.
[5] Totsuj H. and Kihara T., Publ. Astron. Soc. Jpn., 21 (1969) 221.
[6] Davis M. and Peebles P.J.E., Astrophys. J., 267 (1983) 465.
[7] Davis M. et al., Astrophys. J., 333 (1988) L9.
[8] Park C., Vogeley M. S., Geller M. J. and Huchra J. P., Astrophys. J., 431 (1994) 569.
[9] Benoist C., Maurogordato S., da Costa L. N., Cappi A. and Schaeffer R., Astrophys. J., 472 (1996) 452.
[10] Norberg E. et al., Mon. Not. R. Acad. Soc., 328 (2001) 64.
[11] Norberg E. et al., Mon. Not. R. Acad. Soc., 332 (2002) 827.
[12] Zehavi I. et al., Astrophys. J., 571 (2002) 172.
[13] Zehavi I. et al., Astrophys. J., 608 (2004) 16.
[14] Busswell G.S., Shanks T., Frith W. J., Metcalfe N. and Fong R., Mon. Not. R. Acad. Soc., 354 (2004) 991.
[15] Frith W. J., Busswell G. S., Fong R., Metcalfe N. and Shanks T., Mon. Not. R. Acad. Soc., 345 (2003) 1049.
[16] Frith W. J., Metcalfe N. and Shanks T., Mon. Not. R. Acad. Soc., 371 (2006) 160.
[17] Gabrielli A., Sylos Labini F., Joyce M. and Pietronero L., Statistical Physics for Cosmic Structures (Springer Verlag, Berlin) 2004.
[18] Vasilyev N.L., Baryshev Yu. V. and Sylos Labini F., Astron. Astrophys., 447 (2006) 431.
[19] Madgwick D.S. et al., Mon. Not. R. Acad. Soc., 333 (2002) 133.
[20] Sylos Labini F., Vasilyev N. L., Pietronero L. and Baryshev Yu. V., http://arxiv.org/abs/0805.1132v1 (2008).
[21] Kerscher M., Astron. Astrophys., 343 (1999) 333.
[22] Eke V. et al., Mon. Not. R. Acad. Soc., 348 (2004) 866.
[23] Einasto J. et al., Astron. Astrophys., 459 (2006) L1.
[24] Sylos Labini F., Vasilyev N. L. and Baryshev Yu. V., to be published in Astron. Astrophys.
[25] Peacock J. A., Cosmological Physics (Cambridge University Press, Cambridge) 1999.
[26] Peebles P. J. E., The Large-Scale Structure of the Universe (Princeton University Press, Princeton) 1980.
[27] Sylos Labini F. and Vasilyev N. L., Astron. Astrophys., 477 (2008) 381.
[28] Szapudi I. et al., Astrophys. J., 170 (2007) 377.
[29] Springel V. et al., Nature, 435 (2005) 629.
[30] Croton D. J. et al., Mon. Not. R. Acad. Soc., 365 (2006) 11.
[31] Kaiser N., Astrophys. J., 284 (1984) L9.
[32] Sylos Labini F., Montuori M. and Pietronero L., Phys. Rep., 293 (1998) 61.
[33] Hogg D. W. et al., Astrophys. J., 624 (2005) 54.
[34] Sylos Labini F., Vasilyev N. L. and Baryshev Yu. V., Astron. Astrophys., 465 (2007) 23.

29002-p6