

# Absence of self-averaging and of homogeneity in the large scale galaxy distribution

FRANCESCO SYLOS LABINI<sup>1,2</sup>, NIKOLAY L. VASILYEV<sup>3</sup>, LUCIANO PIETRONERO<sup>2,4</sup> and YURIJ V. BARYSHEV<sup>3</sup>

<sup>1</sup> *Museo Storico della Fisica e Centro Studi e Ricerche Enrico Fermi, - Piazzale del Viminale 1, 00184 Rome, Italy*

<sup>2</sup> *Istituto dei Sistemi Complessi CNR, - Via dei Taurini 19, 00185 Rome, Italy*

<sup>3</sup> *Institute of Astronomy, St.Petersburg State University - Staryj Peterhoff, 198504, St.Petersburg, Russia*

<sup>4</sup> *Dipartimento di Fisica, Università di Roma "Sapienza" - P.le Aldo Moro 2, 00185, Rome Italy*

PACS 98.80.-k – Cosmology

PACS 05.40.-a – Fluctuations phenomena in random processes

PACS 02.50.-r – 02.50.-r Probability theory, stochastic processes, and statistics

**Abstract.** - The properties of the galaxy distribution at large scales are usually studied using statistics which are assumed to be self-averaging inside a given sample. We present a new analysis able to quantitatively map galaxy large scale structures while testing for the stability of average statistical quantities in different sample regions. We find that the newest samples of the Sloan Digital Sky Survey provide unambiguous evidence that galaxy structures correspond to large amplitude density fluctuations at all scales limited only by sample sizes. The two-point correlations properties are self-averaging up to approximately 30 Mpc/h and are characterized by a fractal dimension  $D = 2.1 \pm 0.1$ . Then at all larger scales probed density fluctuations are too large in amplitude and too extended in space to be self-averaging inside the considered volumes. These inhomogeneities are compatible with a continuation of fractal correlations but incompatible with: (i) a homogeneity scale smaller than 100 Mpc/h, (ii) predictions of standard theoretical models, (iii) mock galaxy catalogs generated from cosmological Nbody simulations.

**Introduction.** – Understanding the large scale structure of the universe as mapped by galaxy distribution represents one of the cornerstones of modern cosmology. It provides the basic test for theories of structure formation in the universe. A primary question in the statistical analysis of three-dimensional galaxy catalogs (where, in addition to the angular coordinates, the redshift is measured and through Hubble’s law [1] the distance of each object) concerns the determination of a scale where the distribution becomes homogeneous. Such a scale  $\lambda_0$  can be defined to be the one beyond which counts of galaxies in three dimensional spherical volumes of radius  $r$  grow as  $r^3$  [2].

A decade ago, by measuring the conditional density, i.e., the local galaxy density seen by a galaxy in a spherical volume of radius  $r$  around itself [2,3], some of us found that galaxy correlations are power-law with an exponent  $\gamma \approx 1$  up to the sample sizes, i.e.,  $\sim 30$  Mpc/h<sup>1</sup>, corresponding to a fractal dimension  $D = 3 - \gamma \approx 2$  [4–6]. These results were in contrast with the analysis of the same samples by,

e.g., [7–9], who found  $\lambda_0 \approx 10$  Mpc/h and  $\gamma = 1.8$ . The reason for these differences lies in the a-priori assumption of homogeneity, inside a given sample, of the standard statistical analysis [2–4].

At larger scales, with weaker statistical significance, there was an evidence compatible with the fact that power-law correlations in the conditional density extend up to  $r \sim 100$  Mpc/h or more [4, 10]. These results generated a debate in the field [11–13] because even though galaxy structures were found in many different catalogs to extend to scales of the order of hundreds of megaparsecs, the characteristic length scale  $\lambda_0$  statistically describing their correlations was determined to be a few megaparsecs [7–9]. While for some this was a paradox [3–5], for others [7–9, 11, 12] the explanation was that large scale structures have small amplitude relative to the average density. However this interpretation is problematic as in the range of scales where the conditional density shows power-law correlations the sample density is not well defined while density fluctuations have large amplitude [2–4]. The determination of the crossover scale  $\lambda_0$ , where the conditional

<sup>1</sup>We use  $H_0 = 100h$  km/sec/Mpc, with  $0.4 \leq h \leq 0.7$ , for the Hubble’s constant

density from a power-law turns to a constant, allowing a meaningful determination of the average density, has been thus an important task of galaxy correlations studies in the last decade [14].

Two new galaxy catalogs, the Sloan Digital Sky Survey (SDSS — [15]) and the Two degree Field Galaxy Redshift Survey [16], have recently provided great advances in the mapping of the local universe both for the number of objects measured in continuously growing volumes and for the determination of several parameters for each of them. Several studies [17–19,22] of different samples of these surveys confirmed the small scale correlations measured by [4,6]. In addition it has been claimed that a slow crossover toward homogeneity occurs [17] with the average conditional density in spheres at  $\sim 20$  Mpc/h having twice the amplitude of the asymptotic density reached at  $r > 70$  Mpc/h [18]. It was however noticed that galaxy structures could bias the determination of correlations in these samples introducing uncontrolled systematic effects [19,22,23]. Recently in the Two degree Field Galaxy Redshift Survey it has been found that [20,21] galaxy distribution characterized by large amplitude fluctuations with a large spatial extension, whose size is only limited by the sample’s boundaries. In addition at scales  $r < 40$  Mpc/h, it has been observed a well defined and statistically stable power-law behavior of the average number of galaxies in spheres in agreement with previous determinations.

A different and complementary method to characterize structures is provided by galaxy counts as a function of the radial distance from us or of the apparent luminosity [24]. These show large fluctuations around the average behavior both in redshift [25] and angular surveys [26,27]. There have been controversies as to whether these are due to real clustering or to incompleteness of the catalogs [11,24]. Recent results support the conclusion that the local galaxy distribution is characterized by large scale structures with significant correlations on scales  $r > 50$  Mpc/h [28,29].

In this paper we use a new method which is able to establish, in a given sample, the link between the small scales  $r < 30$  Mpc/h correlations and the large scales  $r > 30$  Mpc/h fluctuations in galaxy counts and which clarifies how the latter influence the determination of the former. Using it we can test whether sample means, variances and correlations are well defined, i.e., whether they are statistically stable in different sub-volumes of the given sample. By applying this method to the data of the SDSS project [30] we detect large density fluctuations of spatial extension limited by the samples’ sizes. We show that these introduce systematic biases in the determination of large scale correlations.

**The Data.** — The SDSS [15] is currently the largest spectroscopic survey of extragalactic. Here we consider the main galaxy (MG) sample (DR6) [30] containing redshifts for about 800,000 galaxies which covers an area of 7425 square degrees on the celestial sphere. To query the DR6 database we constrain the flags indicating the type

of object so that we select only the galaxies from the MG sample. We then consider galaxies in the redshift range  $10^{-4} \leq z \leq 0.3$ . The redshift confidence parameter is constrained to be  $z_{conf} \geq 0.35$  with flags indicating no significant redshift determination errors. In addition we apply the filtering condition  $m_r < 17.77$ , using Petrosian apparent magnitudes in the  $r$  filter which are corrected for galactic absorption, and thus taking into account the target magnitude limit for the MG sample in the DR6 [31]. In this way we have selected 479,417 objects. We considered also more stringent limits in apparent magnitude, to test whether a possible incompleteness of the survey at bright and/or faint apparent magnitudes could generate a fake signal. To this aim we used we have  $14.5 \leq m_r \leq 17.5$  and we selected 370,893 objects, i.e., about 25% less than with less conservative constraints. We have considered a rectangular angular fields, with uniform coverage, in the SDSS internal angular coordinates  $(\eta, \lambda)$  limited by  $-6^\circ \leq \eta \leq 36^\circ$  and  $-48^\circ \leq \lambda \leq 32.5^\circ$ .

To construct volume limited (VL) samples that are unbiased for the selection effect related to the cuts in the apparent magnitude, we have applied a standard procedure [33]. Firstly we compute metric distances  $R(z)$  using the standard cosmological parameters  $\Omega_M = 0.3$  and  $\Omega_\Lambda = 0.7$ . Secondly the galaxy absolute magnitude is determined to be  $M_r = m_r - 5 \cdot \log_{10} [R(z) \cdot (1+z)] - K_r(z) - 25$ , where  $K_r(z)$  is the K-correction. We determine the  $K_r(z)$  term from the NYU VACG data [32]. Finally, we have considered two different VL samples (named VL1 and VL2) defined by two chosen limits in absolute magnitude and metric distance: for VL1  $R \in [100, 300]$  Mpc/h,  $M \in [-22, -20]$  and for VL2  $R \in [200, 600]$  Mpc/h,  $M \in [-23, -21.5]$ . The number of galaxies is about  $4 \cdot 10^4$  in VL1 and  $3 \cdot 10^4$  in VL2. Different cuts in absolute magnitude do not introduce substantial differences in the results presented in this paper. When more conservative and stringent limits in apparent magnitude are applied we find that in the samples with the same limits in distance, there are up to three times less galaxies. However the main results presented in this paper are affected only in the fact that statistics is less robust.

The Millennium project [38] has performed several cosmological simulations of standard theoretical models. Amount of dark matter and cosmological parameters are given in agreement with standard models. The dark matter simulations have about  $10^{10}$  particles. From these galaxies are identified according to semi-analytics models of galaxy formation [39]. We have cut a sample with exactly the same geometry as the SDSS VL1 sample and a sample close to the geometrical parameters of the SDSS VL2 applying the same absolute magnitude limits in  $r$ -filter as for the SDSS data. In the SDSS we use a redshift space analysis while in mock catalogs a real space one. The difference between the real and redshift space analysis is relevant for very small scales, i.e.,  $r < 5$  Mpc/h and thus does not influence results on scales of the order of 100 Mpc/h [14,20,21].

**Statistical methods.** – Statistical properties are determined by making averages over the whole sample volume [2]. In doing so one implicitly assumes that a certain quantity measured in different regions of the sample is statistically stable, i.e., that fluctuations in different sub-regions are described by the same probability density function (PDF). However it may happen that measurements in different sub-regions show systematic differences, which depend, for instance, on the spatial position of the specific sub-regions. In this case the considered statistic is not statistically stationary in space, the PDF systematically differs in different sub-regions and its whole-sample average value is not a meaningful descriptor [2].

In general such systematic differences may be related to two different possibilities: (i) that the underlying distribution is not translationally and/or rotationally invariant; (ii) that the volumes considered are not large enough for fluctuations to be self-averaging [40]. On very general grounds, we expect the galaxy distribution to satisfy the condition of *statistical* stationarity in space to avoid special points or directions [1, 2]. Hence the question we face in a finite volume analysis concerns whether it is large enough to obtain statistically stable results. Note that stationary stochastic distributions satisfy the condition of spatial statistical isotropy and homogeneity also when they have zero average density in the infinite volume limit [2]. This condition is called the Conditional Cosmological Principle [2] to differentiate it from the stronger Cosmological Principle which requires exact homogeneity and deterministic rotational and translation invariance [1, 2].

For the case of galaxy surveys there is an intrinsic preferred direction which is set by the radial position from the observer, i.e., the Earth. It is thus necessary to show that statistical quantities do not depend on the radial distance from us. To this aim, in a given sample, a simple approach is to determine the number  $N(r; R)$  of galaxies in spheres of radius  $r$ , centered on a galaxy whose distance from the origin is  $R$ : we call it the scale-length (SL) analysis. As we discuss below, this is found to be very efficient in mapping large scale structures which appear as large fluctuations of  $N(r; R)$ . For instance by studying it in various angular slices of the SDSS samples we identify a giant filament covering, in the largest contiguous angular area of the survey, more than 400 Mpc/h at  $R \sim 500$  Mpc/h. In different sky directions the SL analysis reveals a variety of structures, showing that large density fluctuations are quite typical.

Averaged over the whole sample the quantity  $N(r; R)$  gives an estimate of the average conditional number of galaxies in spheres of radius  $r$ . An estimator making the weakest a-priori-assumptions about the properties of the distribution outside the sample volume is [2–4]

$$\overline{N(r)} = \frac{1}{M(r)} \sum_{i=1}^{M(r)} N_i(r), \quad (1)$$

where  $N_i(r)$  is the number of galaxies seen by the  $i^{\text{th}}$

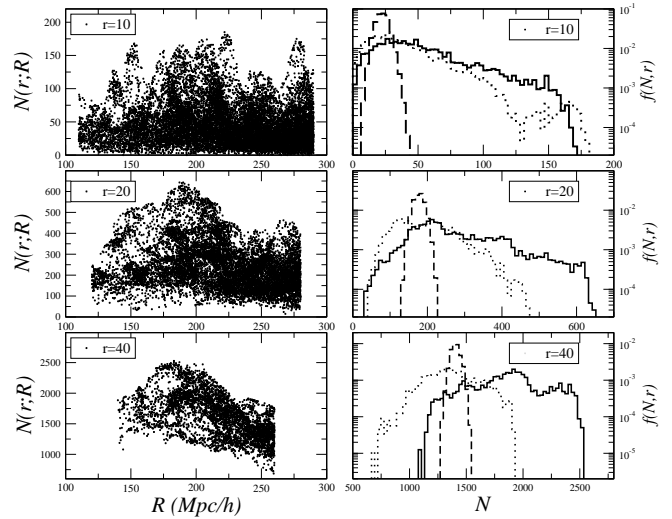


Fig. 1: *Left panels:* From top to bottom the SL analysis for the SDSS sample VL1, with  $r = 10, 20, 40$  Mpc/h. *Right panels:* Probability density function of  $N(r; R)$  in two non-overlapping sub-samples with equal volume (each half of the sample volume) at small and large  $R$ . While for  $r = 10, 20$  Mpc/h the PDF (nearby sub-sample solid line and faraway sub-sample dotted line) is reasonably statistically stable, for  $r = 40$  Mpc/h there is a clear difference. The dashed line corresponds to the Poisson distribution: a Gaussian function gives very good fits for all  $r$ .

center-point and the number of centers  $M(r)$  varies with  $r$  because only those galaxies for which the sphere is fully included in the sample volume are considered as centers [2]. Even in this case, there is an intrinsic selection effect related to the geometry of the samples, which are portions of spheres: when  $r$  is large only a part of the sample is explored by the volume average. Hence 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  $\sim [R_{\min} + r, R_{\max} - r]$  from the radial boundaries of the sample at  $[R_{\min}, R_{\max}]$ .

When Eq.1 scales as  $\overline{N(r)} \sim r^D$  and  $D = 3$  the distribution is homogeneous, while for  $D < 3$  it is fractal [2, 24]. Furthermore fluctuations  $\delta^2(r) = [\overline{N(r)^2} - \overline{N(r)}^2] / \overline{N(r)}^2$  are small for a homogeneous distribution with any kind of small-amplitude correlations ( $\delta^2(r) \ll 1$ ) and large for a fractal one ( $\delta^2(r) \sim 1$ ) [2, 24]. To study fluctuations we determine the PDF of  $N_i(r)$ , which is expected to converge to a Gaussian when  $r \gg \lambda_0$  [2].

**Results.** – Let us now consider the VL1 sample. Here the SL analysis (Fig.1) detects large density fluctuations without a clear radial-distance dependent trend. Correspondingly the PDF has a regular shape characterized by a peak with a long  $N$  tail and it is sufficiently statistically stable in different non-overlapping sub-samples of equal volume. This occurs except for the largest sphere radii, i.e., for  $r > 30$  Mpc/h, for which the number of independent centers becomes too small.

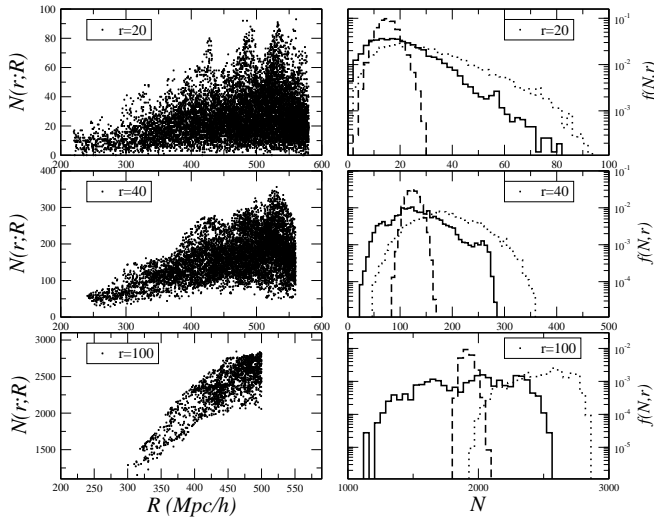


Fig. 2: The same of Fig.1 for the VL2 sample with  $r = 20, 40, 100$  Mpc/h. The fact that the signal for SDSS data becomes smoother when  $r$  increases, but still with a systematic radial distance-trend, is due to the fact that many spheres overlap when  $r$  grows. While for  $r = 20$  Mpc/h the PDF is reasonably statistically stable, for  $r = 40, 100$  Mpc/h there is a clear difference.

In the deeper VL2 sample we find instead a systematic trend of  $N(r;R)$  as a function of  $R$  (Fig.2). Particularly  $N(r;R)$ , for  $R > 300$  Mpc/h, grows without any clear saturation, for sphere radii up to  $r \sim 100$  Mpc/h. The PDF in two non-overlapping sub-samples of equal volume is found to differ systematically for  $r > 30$  Mpc/h, and its average value moves as a function of  $R$ . This shows that fluctuations are not self-averaging at those scales. This behavior is due to the large scale structures at scales  $R > 300$  Mpc/h. However at smaller scales, i.e.,  $r < 30$  Mpc/h, the PDF in different sub-samples is reasonably statistically stable and similar to the one found in VL1. This shows that, at those scales, fluctuations are self-averaging because the volume average can explore different regions of the sample. Thus these results show that, at the largest scales probed, there are large density fluctuations which are not self-averaging because of the limited sample volume. These determine relative variations larger than unity in the estimation of the average density in spheres of radius  $r = 100$  Mpc/h. We thus conclude that the homogeneity scale must be  $\lambda_0 > 100$  Mpc/h, the largest sphere radius we considered.

Previous analyses of smaller galaxy catalogs, e.g., [4–6, 17–19, 22, 33, 34], considered sample averaged statistics without quantitatively testing whether a significant bias could affect the results (but see [19, 22]). For instance the estimator of the most commonly used statistics, the two-point correlation function [35], can be written as [2]

$$\xi(r) + 1 = \frac{\overline{N(r, \Delta r)}}{V(r, \Delta r)} \cdot \frac{V}{N}. \quad (2)$$

The first ratio in the r.h.s. of Eq.2 is the average con-

ditional density, i.e., the number of galaxies in shells of thickness  $\Delta r$  averaged over the whole sample, divided by the volume  $V(r, \Delta r)$  of the shell. The second ratio in the r.h.s. of Eq.2 is the average density estimated in a sample containing  $N$  galaxies and with volume  $V$ . When measuring this function we implicitly assume, in a given sample, that: (i) fluctuations are self-averaging in different sub-volumes [2] (ii) the linear dimension of the sample volume is  $V^{1/3} \gg \lambda_0$  [2, 3], i.e., the distribution has reached homogeneity inside the sample volume. When the latter condition is not verified the  $\xi(r)$  analysis is biased by systematic finite size effects even if fluctuations are self-averaging [2, 3]. To show how non self-averaging fluctuations inside a given sample bias the  $\xi(r)$  analysis, we consider the estimator

$$\xi(r; R, \Delta R) + 1 = \frac{\overline{N(r, \Delta r)}}{V(r, \Delta r)} \cdot \frac{V(r^*)}{\overline{N(r^*; R, \Delta R)}}, \quad (3)$$

where the second ratio on the r.h.s. is the density of points in spheres of radius  $r^*$  averaged over the centers lying in a shell of thickness  $\Delta R$  around the radial distance  $R$ . If the distribution is homogeneous, i.e.,  $r^* > \lambda_0$ , and statistically stationary, Eq.3 should be statistically independent on the range of radial distances  $(R, \Delta R)$  considered. For instance we consider, in the VL2 sample,  $\Delta R = 40$  Mpc/h and  $R = 240$  Mpc/h or  $R = 520$  Mpc/h, with  $r^* > 50$  Mpc/h. We thus find large variations in the amplitude of  $\xi(r)$  (Fig.3). This is simply an artifact generated by the large density fluctuations on scales of the order of the sample sizes. The results that the estimator Eq.2 (or others based on pair-counting [2, 36]) has nearly the same amplitude in different samples, e.g., [7–9, 33, 34], despite the large fluctuations of  $N(r;R)$ , are simply explained by the fact that  $\xi(r)$  is a ratio between the local conditional density and the sample average density: both vary in the same way when the radial distance is changed and thus the amplitude is nearly constant.

On the other hand Eq.1, *averaged over the volumes where the PDF has a statistically stable shape*, shows in both considered samples a power-law behavior for  $r < 30$  Mpc/h corresponding to a fractal dimension  $D = 2.1 \pm 0.1$  in agreement with [4, 6, 17, 19, 22] (Fig.3). Due to the non self-averaging nature of fluctuations at larger scales, i.e., due to limited volumes, we are not able to determine correlations for  $r > 30$  Mpc/h.

**Discussion.** – According to standard models of cosmological structure formation, gravitational clustering gives rise to non-linear perturbations from homogeneous initial conditions in the early universe [1]. If the initial amplitude of fluctuations is normalized to the anisotropies of the Cosmic Microwave Background Radiation (CMBR) [37], then the homogeneity scale is about  $\lambda_0^m = 10$  Mpc/h [23], i.e., twice the value at which  $\xi(r) = 1$  [1, 2].

Indeed in mock galaxy catalogs generated from N-body simulations of standard cosmological models [38, 39],  $N(r;R)$  does not show, for  $r > \lambda_0^m$ , large fluctuations



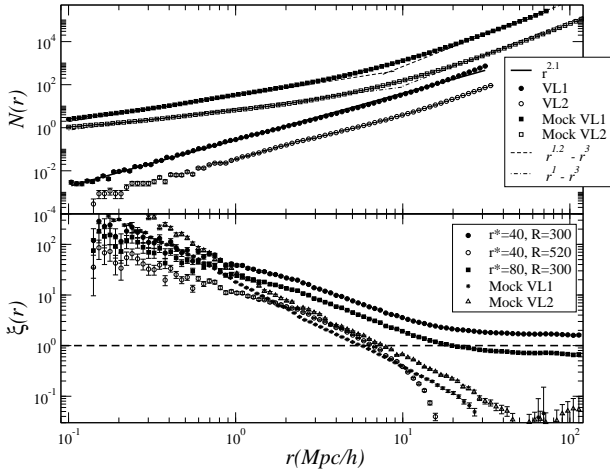


Fig. 3: *Upper Panel:* The sample average conditional number of galaxies (Eq.1) for the SDSS VL1 and VL2 samples up to  $r < 30$  Mpc/h. The best fit slope (solid line) gives  $D = 2.1 \pm 0.1$ . The difference in amplitude between the two samples is simply ascribed to a luminosity selection effect [2], as VL1 contains fainter galaxies than VL2. The amplitude of the mock VL1 and VL2 samples has been rescaled by the same arbitrary factor for seek of clarity. *Bottom Panel:* Standard two-point correlation function in the VL2 sample estimated by Eq.3: the sample average density is computed in spheres of radius  $r^*$  and considering all center-points lying in a bin of thickness  $\Delta R = 40$  Mpc/h centered at different radial distance  $R$ . The case  $r^* = 80$  Mpc/h and  $R = 520$  Mpc/h gives the average over the whole sample, i.e., Eq.2, and it coincides with the estimation of [33] in a similar sample. The flat tail of  $\xi(r)$ , for  $r^* = 40, 80$  Mpc/h and  $R = 300$  Mpc/h, reflects the inhomogeneities with strong correlations at large scales. The stars and triangles correspond to the behavior of the mock VL1 and VL2 samples: in this case the amplitude is statistically stable and thus meaningful. The small amplitude difference in this case is ascribed to the different selection in luminosity [38].

or systematic trends as a function of  $R$  (Fig.4). Because in these artificial catalogs fluctuations are small and self-averaging, whole-sample averaged statistics are meaningful at all scales. From the  $\overline{N(r)}$  analysis we find, differently from the real galaxy data, that  $\overline{N(r)} \sim r^D$  with  $D = 1.1 \pm 0.1$  for  $r < \lambda_0^m$  and  $D = 3$  for  $r > \lambda_0^m$  (Fig.3). Correspondingly the PDF rapidly converges to a Gaussian for  $r > \lambda_0^m$ . The  $\xi(r)$  function, estimated by Eq.2 or Eq.3, has a statistically stable amplitude. The residual small amplitude difference between the mock VL1 and VL2 samples is ascribed to a different selection in luminosity [38]. The amplitude of  $\xi(r)$  is equal to unity at  $r \approx 6$  Mpc/h in agreement with [38, 39]. In addition its shape reasonably agrees with the standard estimation of  $\xi(r)$  from galaxy catalogs [7–9, 33, 34]. However, as discussed above, the latter estimation is biased by systematic effects making the agreement fortuitous.

Standard theoretical models predict that, for  $r > \lambda_0^m$ , the small fluctuations in the early universe are linearly amplified by gravitational clustering. Therefore for  $r > \lambda_0^m$

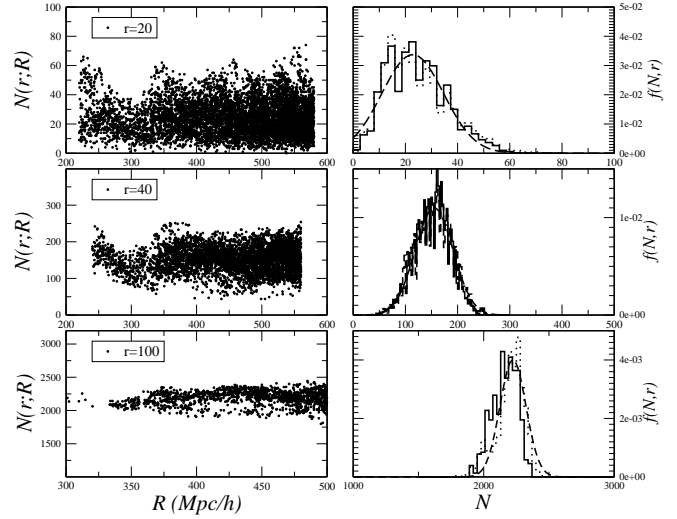


Fig. 4: The same of fig.1 for mock sample VL2, with  $r = 20, 40, 100$  Mpc/h. Although for  $r = 20$  Mpc/h fluctuations are still important, they rapidly become small for larger radii. The PDF is statistically stable for all  $r$  and  $R$ . The solid line corresponds to a Gaussian fit.

the shape of the theoretical  $\xi^m(r)$  must be the same as the initial one [1]. This is characterized by a length-scale  $r_c$ , where  $\xi^m(r_c) = 0$ , which is fixed by the physics of the early universe and estimated from CMBR anisotropies to be  $r_c \approx 100$  Mpc/h [36, 37]. For  $r > r_c$ ,  $\xi^m(r)$  becomes negative, corresponding to super-homogeneous correlations characterized by the most rapid possible decay of fluctuations [41, 42]. This theoretical framework applies to the whole mass distribution, where dark matter is supposed to provide with the main contribution. Galaxies would form on the largest peaks of the density field. Standard models of galaxy formation describe this physical phenomenon as a selection mechanism [44]. This leaves unperturbed the scale  $r_c$  and slightly changes  $\lambda_0^m$  [36, 43]. We find  $\lambda_0 > 100 \text{ Mpc/h} \geq r_c \gg \lambda_0^m$ . This raises a fundamental inconsistency for the relation between galaxy structures and CMBR anisotropies as no physical mechanism is known, which by sampling a super-homogeneous density field transforms it into a strongly inhomogeneous one [2, 36, 43].

**Conclusion.** – In summary, by applying the SL analysis to the newest SDSS galaxy samples, we measure large density fluctuations of spatial extension limited by sample sizes. At scales  $r < 30$  Mpc/h we detect statistically stable fractal correlations with  $D = 2.1 \pm 0.1$ . On larger scales,  $r > 30$  Mpc/h, we find that the galaxy distribution is strongly inhomogeneous and fluctuations are not self-averaging in the samples considered. This situation is compatible with fractal power-law correlations extending to such length-scales but incompatible with homogeneity at  $\lambda_0 \leq 100$  Mpc/h. Indeed, in a portion of a fractal, large structures are expected to be present at any scale, fluctuations being self-averaging only if the sam-

ple volume is large enough [2]. These results have important consequences on the theoretical interpretation of the large scale universe, where models, normalized to CMBR anisotropies, predict there is not enough time to form structures with relative density fluctuations larger than unity on scales larger than  $\lambda_0^m \approx 10\text{Mpc/h}$  [1, 38]. This length scale is more than ten times smaller than our lower limit to  $\lambda_0$ . Indeed the latter is of the order of the scale  $r_c$  where theoretical model predict matter distribution to have negative correlations, a situation which is in contrast with the results from the data analyzed here. Thus the large scale inhomogeneities detected in the SDSS samples are incompatible with the predictions of standard theoretical models relating the early universe physics, with CMBR normalization, to structures in the present universe. Moreover we found that for  $r < \lambda_0^m$ , mock galaxy catalogs have different correlations from real galaxy data, i.e.,  $D = 1.1 \pm 0.1$  instead of  $D = 2.1 \pm 0.1$ . Thus structures generated by N-body simulations are intrinsically different from observed ones.

Recently three dimensional maps of dark matter distribution from weak lensing observations have been published [45]. Dark matter is observed to trace the same structures as galaxies. Thus the whole matter distribution would be inhomogeneous on scales larger than 100 Mpc/h. This has a great impact on the whole theoretical framework for the physical understanding of the large scale universe. For instance it may imply a new type of evolution scenario within an open Friedmann model [46] or new types of spatial averaging of the Einstein equations [47, 48] which relate the observed inhomogeneities to the apparent acceleration measured from supernovae observations [49].

The determination of statistical properties of those very large structures, which we detected but could not analyze in detail, should be possible when sample volumes become large enough so that the corresponding fluctuations will be self-averaging on scales larger than the ones studied here. The application of the SL analysis to the forthcoming galaxy redshift surveys, like the complete SDSS [15], thus represents an important task.

\*\*\*

We warmly thank M. Joyce for his careful reading of the manuscript and the very detailed comments and suggestions. We thank A. Gabrielli, R. Durrer, M. Lopez Corredoira for useful remarks and discussions. We acknowledge the use of the Sloan Digital Sky Survey data [30] and of the Millennium run semi-analytic galaxy catalog [39]

## REFERENCES

- [1] Peebles P. J. E., *The Large-Scale Structure of the Universe* (Princeton University Press, 1980)
- [2] Gabrielli A., Sylos Labini F., Joyce M., Pietronero L., *Statistical Physics for Cosmic Structures* (Springer, 2004)
- [3] Pietronero L., *Physica A*, **144**, 257 (1987)
- [4] Sylos Labini, F., Montuori, M., Pietronero, L., *Phys. Rep.*, **293**, 61 (1998)
- [5] Coleman P., Pietronero L., *Phys.Rep.* , **213**, 311 (1992)
- [6] Joyce, M., Montuori, M., Sylos Labini, F., *Ap.J.*, **514**, L5 (1999)
- [7] Davis, M., Peebles, P.J.E., *Ap.J.*, **267**, 465 (1983)
- [8] Park, C., et al., *Ap.J.*, **431**, 569 (1994)
- [9] Benoist, C., et al., *Ap.J.*, **472**, 452 (1996)
- [10] Joyce, M., et al., *A&A*, **344**, 387 (1999)
- [11] Wu, K.K., Lahav, O., Rees, M., *Nature*, **397**, 225 (1999)
- [12] Davis, M., Proc. of the conference *Critical Dialogues in Cosmology* ed. N. Turok, p.12 (World Scientific, 1997)
- [13] Pietronero L., Montuori M., Sylos Labini F., *ibidem* pg.24
- [14] Baryshev, Yu., Teerikorpi P., *Bull.Spec.Astrophys.Obs.*, **59**, 92-154 (2006)
- [15] York, D., et al., *Astronom.J.*, **120**, 1579 (2000)
- [16] Colless M., et al., *MNRAS*, **328**, 1039 (2001)
- [17] Hogg, D.W., et al., *Ap.J.*, **624**, 54 (2005)
- [18] Tikhonov, A.V., *Astron.Lett.*, **32**, 72 (2006)
- [19] Vasilyev, N.L., Baryshev, Yu. V., Sylos Labini, F., *A&A*, **447**, 431 (2006)
- [20] Sylos Labini, F., Vasilyev, N.L., Baryshev, Yu. V., *Europhys.Lett*, **85**, 29002-p1 (2009)
- [21] Sylos Labini, F., Vasilyev, N.L., Baryshev, Yu. V., *Astron.Astrophys.*, **496**, 7 (2009)
- [22] Sylos Labini, F., Vasilyev, N.L. Baryshev, Yu.V., *A&A* **465**, 23 (2007)
- [23] Joyce, M., et al., *A&A*, **443**, 11 (2005)
- [24] Gabrielli, A., Sylos Labini, F., *Ep.L.* , **54**, 286 (2001)
- [25] Kerscher, M., et al., *A&A*, **333**, 1 (1998)
- [26] Picard A., *A.J.*, **102**, 445-453 (1991)
- [27] Bertin E. and Dennefeld M., *A.&A.*, **317**, 43 (1997)
- [28] Buswell, G.S. , et al., *MNRAS*, **354**, 991 (2004)
- [29] Frith, W.J. et al., *MNRAS*, **345**, 1049 (2003)
- [30] Adelman-McCarthy, J. K. , et al., *Ap.J.Suppl.*, **75**, 297, (2008)
- [31] Strauss, M.A., et al., *A.J.* , **124**, 1810 (2002)
- [32] Blanton, M.R., Roweis, S., *Astron.J.*, **133**, 734 (2007)
- [33] Zehavi, I., et al., *Ap.J.*, **571**, 172 (2002)
- [34] Norberg E., et al., *MNRAS*, **332**, 827, (2002)
- [35] Totsuji, H., Kihara, T., *PASJ*, **21**, 221 (1969)
- [36] Sylos Labini, F., Vasilyev, N.L., *A&A*, **477** 381-395 (2008)
- [37] Spergel, D.N., et al., *Ap.J.Suppl.*, **170**, 377 (2007)
- [38] Springel, V., et al., *Nature*, **435**, 62 (2005)
- [39] Croton, D.J., et al., *MNRAS*, **365**, 11 (2006)
- [40] Aharony, A., Harris, B., *Phys. Rev. Lett*, **77**, 3700 (1996)
- [41] Gabrielli, A., Joyce, M., Sylos Labini, F., *Phys.Rev.*, **D65**, 083523 (2002)
- [42] Torquato, S., Stillinger, F.H., *Phys.Rev.*, **E68**, 041113 (2003)
- [43] Durrer, R., et al., *Ap.J.*, **585**, L1 (2003)
- [44] Kaiser, N., *Ap.J.* **284**, L9 (1984)
- [45] Massey, R., et al., *Nature*, **445**, 286 (2007)
- [46] Joyce, M., et al. *EpL*, **50**, 416 (2000)
- [47] Buchert, T., *Gen.Rel.Grav.*, **32**, 105 (2000)
- [48] Wiltshire, D.L., *Phys.Rev.Lett.*, **99**, 251101 (2007)
- [49] Perlmutter, S., et al., *Ap.J.*, **517**, 565 (1999)