Chapter 13

The debate on galaxy space distribution: an overview

Marco Montuori and Luciano Pietronero
Deptartment of Physics, University of Rome—'La Sapienza' and INFM, Rome, Italy

13.1 Introduction

A critical assumption of the hot big bang model of the universe is that matter is homogeneously distributed in space over a certain scale. It is usually assumed that under this condition the Friedmann–Robertson–Walker (FRW) metric correctly describes the dynamics of the universe. Investigating this assumption is then of fundamental importance in cosmology and much current research is devoted to this issue. In this chapter, we will review the current debate on the spatial properties of galaxy distribution.

13.2 The standard approach of clustering correlation

The usual way to investigate the properties of the spatial distribution of glaxies is to measure the two-point autocorrelation function $\xi(r)$ [1]. This is the spatial average of the fluctuations in the galaxy number density at distance r, with respect to a homogeneous distribution of the same number of galaxies. Let $n(\boldsymbol{r}_i)$ the density of galaxies in a small volume δV at the position $\boldsymbol{r}_i$. The relative fluctuation in δV is

$$\frac{\delta n(r_i)}{\langle n \rangle} = \frac{n(\boldsymbol{r}_i) - \langle n \rangle}{\langle n \rangle} \tag{13.1}$$

where $\langle n \rangle = N/V$ is the density of the sample.

It is clear that the fluctuations are defined with respect to the density of the sample $\langle n \rangle$. The two-point correlation function $\xi(r)$ at scale r is the spatial average of the product of the relative fluctuations in two volumes centred on data points at distance r:

$$\xi(r) = \left\langle \frac{\delta(\mathbf{r}_i + r)}{\langle n \rangle} \frac{\delta(\mathbf{r}_i)}{\langle n \rangle} \right\rangle_i = \frac{\langle n(\mathbf{r}_i) n(\mathbf{r}_i + \mathbf{r}) \rangle_i}{\langle n \rangle^2} - 1, \tag{13.2}$$

where the index i means that the average is performed over the all the galaxies in the samples. A set of points is correlated on scale r if $\xi(r) > 0$; it is uncorrelated if $\xi(r) = 0$. In the latter case the points are evenly distributed at scale r or, in another words, they have a homogeneous distribution at scale r. In the definition of $\xi(r)$, the use of the sample density $\langle n \rangle$ as a reference value for the fluctuations of galaxies is the conceptual assumption that the galaxy distribution is homogeneous at the scale of the sample.

In such a framework, a relevant scale r_0 for the correlation properties is usually defined by the condition $\xi(r_0) = 1$. The scale r_0 is called the *correlation length of the distribution.*

13.3 Criticisms of the standard approach

Let us summarize the conclusions of the previous section:

- The $\xi(r)$ analysis assumes homogeneity at the sample size; and
- a characteristic scale for the correlation is defined by the amplitude of $\xi(r)$, i.e. the scale at which $\xi(r)$ is equal to one [1].

These two points raise two main criticisms:

- As the $\xi(r)$ analysis assumes homogeneity, it is not reliable for *testing* homogeneity. In order to use $\xi(r)$ analysis, the density of galaxies in the sample must be a good estimation of the density of the whole distribution of the galaxies. This may either be true or not; in any case, it should be checked *before* $\xi(r)$ analysis is applied [2].
- The correlation length r_0 does not concern the *scale* of fluctuations. In this sense, it is not correct to refer to it as a measure of the characteristic size of correlations and call it the *correlation length.* According to the definition of $\xi(r)$, r_0 simply separates a regime of large fluctuations $\delta n/\langle n \rangle \gg 1$ from a regime of small fluctuations $\delta n/\langle n \rangle \ll 1$ [3,4].

Again the argument is valid if the average density $\langle n \rangle$ of the sample is the average density of the distribution or, in other words, if the distribution is homogeneous on the sample size. In statistical mechanics, the *correlation length* of the distribution is defined by how fast the correlations vanish as a function of the scale, i.e. by the functional form of $\xi(r)$ and not by its amplitude.

In this respect, the first step in a spatial correlation analysis of a data-set should be a study of the density behaviour versus the scale. This should be done without any *a priori* assumptions about the features of the underlying distribution [2].

13.4 Mass–length relation and conditional density

The *mass–length* relation links the average number of points at distance r from any other point of the structure to the scale r. Starting from an ith point occupied by an object of the distribution, we count how many objects $N(< r)_i$ ('mass') are present within a volume of linear size r ('length') [5]. The average over all the points of the structure is:

$$\langle N(< r)_i \rangle = B \cdot r^D. \tag{13.3}$$

The exponent D is called *the fractal dimension* and characterizes in a quantitative way how the system fills the space, while the prefactor B depends on the lower cut-off point of the distribution.

The conditional density $\Gamma(r)$ is the average number of points in a shell of width dr at distance r from any point of the distribution.

According to equation (13.3), $\Gamma(r)$ is:

$$\Gamma(r) = \frac{1}{4\pi r^2 \mathrm{d}r} \frac{\mathrm{d}\langle N(< r)_i \rangle}{\mathrm{d}r} = \frac{BD}{4\pi} \cdot r^{D-3} \tag{13.4}$$

(see [2, 6] for details of the derivation).

13.5 Homogeneous and fractal structure

If the distribution crosses over to a homogeneity distribution at scale r, $\Gamma(r)$ shows a flattening toward a constant value at such a scale. In this case, the fractal dimension in equations (13.3) and (13.4) has the same value as the dimension of the embedding space d, $D = d$ (in three-dimensional space $D = 3$) [2, 5, 6].

If this does not happen, the density of the sample will not correspond to the density of the distribution and it will show correlations up to the sample size. The simplest distribution with such properties is a fractal structure [5]. A fractal consists of a system in which more and more structures appear at smaller and smaller scales and the structures at small scales are similar to those at large scales. The distribution is then self-similar. It has a value of D that is smaller than d, $D < d$. In three-dimensional space $d = 3$, a fractal has $D < 3$ and $\Gamma(r)$ is a *power law*. The value of $N(< r)_i$ largely fluctuates by changing both the starting ith point and the scale r. This is due to the scale-invariant feature of a fractal structure, which does not have a characteristic length [5, 7].

13.6 $\xi(r)$ for a fractal structure

Equation (13.4) shows that $\Gamma(r)$ is a well-defined statistical tool for the generic distribution of points, since it depends only on the intrinsic quantities (B and D). The same is not true for $\xi(r)$ statistics.

Assuming for simplicity a spherical sample volume with radius R_s ($V(R_s) = (4/3)\pi R_s^3$), containing $N(R_s)$ galaxies. The average density of the sample will be

$$\langle n \rangle = \frac{N(R_s)}{V(R_s)} = \frac{3}{4\pi} B R_s^{-(3-D)}. \tag{13.5}$$

For a fractal, $D < 3$ and its average density is a decreasing function of the sample size: $\langle n \rangle \to 0$ for $R_s \to \infty$. Then the average density depends explicitly on the sample size R_s and it is not a meaningful quantity.

From equation (13.2), the expression for $\xi(r)$ for a fractal distribution is [2]:

$$\xi(r) = ((3-\gamma)/3)(r/R_s)^{-\gamma} - 1. \tag{13.6}$$

From equation (13.6) it follows that, for the *fractal sample* the so-called correlation length r_0 (defined as $\xi(r_0) = 1$) is a linear function of the sample size R_s:

$$r_0 = ((3-\gamma)/6)^{1/\gamma} R_s. \tag{13.7}$$

It is then a quantity *without* any statistical *significance*, one simply related to the sample size [2].

Neither is $\xi(r)$ *a power law*. For $r \le r_0$,

$$((3-\gamma)/3)(r/R_s)^{-\gamma} \gg 1 \tag{13.8}$$

and $\xi(r)$ is well approximated by a *power law* [6].

For larger distances there is clear deviation from power-law behaviour due to the definition of $\xi(r)$. This deviation, however, is just due to the size of the observational sample and does not correspond to any real change in the correlation properties. It is clear that if one estimates the exponent of $\xi(r)$ at distances $r \approx r_0$, one systematically obtains a higher value of the correlation exponent due to the break of $\xi(r)$ in the log–log plot. Only if the sample set has a crossover to homogeneity inside the sample side, is $\xi(r)$ correct. However, this information is given only by the $\Gamma(r)$ analysis which, for this reason, should always come before the $\xi(r)$ investigation.

13.7 Galaxy surveys

Galaxy catalogues are *angular catalogues (three-dimensional)*, which can be computed in *real* or in *redshift space*. The latter defines the galaxy positions by the *redshift distance* $\mathbf{s}$, which is derived by the *galaxy redshift* z, according to *Hubble's law*. $\mathbf{s}$ is not the *real* distance, but contains an additional term called the *redshift distortion*, which is small on scales $s > 5h^{-1}$ Mpc [8].

We will report the statistical properties of *redshift surveys*, which contain the large majority of avalaible three-dimensional data.

13.7.1 Angular samples

$\xi(r)$ can be obtained from two-dimensional data, by means of the angular two-point function $w(\theta)$. $\xi(r)$ is reconstructed using the luminosity function, which is derived assuming homogeneneity in the sample [1]. No independent check is usually performed on this assumption. The procedure is currently considered one of the best estimates of three-dimensional clustering properties of galaxies, at least on a small scale ($\leq 20h^{-1}$ Mpc) [9, 10]. Such a claim is considered to be justified by the great quantity of available data in angular catalogues with respect to three-dimensional surveys and by the absence of redshift distortions in the two-dimensional data. The main conclusion obtained by this approach is that the galaxy correlation (more precisely for optical selected galaxies) $\xi_{gg}(r)$ is *quite close to a power law in the range* $10h^{-1}$ kpc–(10–$20h^{-1}$) Mpc and more precisely [9, 10]:

$$\xi_{gg}(r) = \left(\frac{r_0}{r}\right)^{-1.77} \tag{13.9}$$

with a *correlation length* $r_0 \approx 4.5 \pm 0.5h^{-1}$ Mpc.

This is considered to give the 'canonical shape and parameter values' of $\xi(r)$ and is a well-established result in cosmology [1, 10–14].

13.7.2 Redshift samples

13.7.2.1 ML samples

An ML sample is simply the whole *redshift catalogue*. By construction, any ML sample is incomplete in the distribution of galaxies. At larger distances, it contains fewer and fewer galaxies, as more and more galaxies fall beyond the threshold of detectability. To account for such an effect, the galaxies in the sample are weighted, according the luminosity function [1].

The value of s_0 in different ML catalogues is found to span from 4.5–$8h^{-1}$ Mpc [10, 13].

$\xi(s)$ does not appear to be *a power law*. According to Guzzo [15], the shape of $\xi(s)$ at very small scales ($<3h^{-1}$ Mpc) is well fitted by a power law with exponent $\gamma = -1$.

13.7.2.2 VL samples

It is possible to extract subsamples from the ML catalogues, which are unaffected by the aforementioned incompleteness. Such samples are called *VL samples* [16]. The main result of $\xi(s)$ analysis is that different VL samples have different values for the correlation length s_0. The general trend is that deeper and brighter samples show larger s_0 (figure 13.1) [6, 15, 19–23]. Again, $\xi(s)$ is *not a power law* in the whole observed range of scale ($\approx$1–$50h^{-1}$ Mpc). This has been recognized by several authors, who have performed the fit with the power law in a limited range of scales. The value of the exponent γ (see equation (13.9)) is in the range

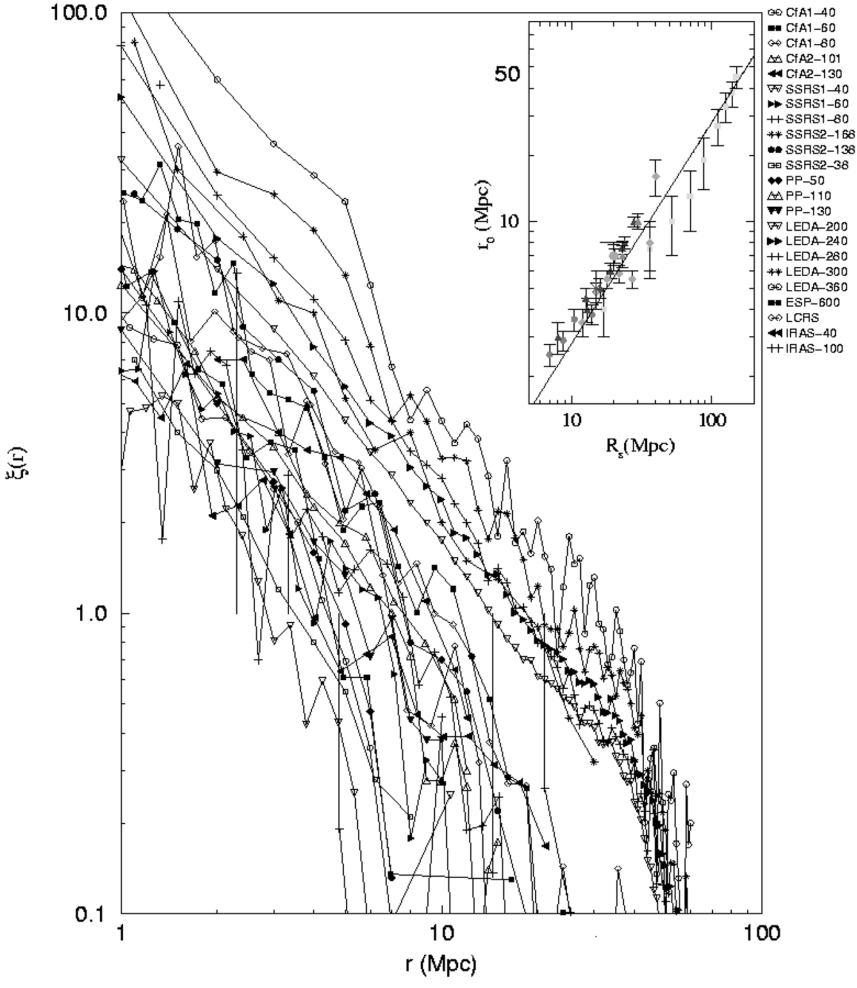

Figure 13.1. $\xi(r)$ measure in various VL galaxy samples. The general trend is an increase of the $\xi(r)$ amplitude for brighter and deeper samples. In the *insert panel* we show the dependence of *correlation length* r_0 on *sample size* R_s for all samples. The linear behaviour is a consequence of the fractal nature of galaxy distribution in these samples.

1.17–2.1, which demonstrates the deviation of $\xi(s)$ from the *canonical power-law* shape [6, 15, 21, 22].

13.8 $\Gamma(r)$ analysis

According to our criticism of the standard analysis, we have measured the galaxy conditional average density $\Gamma(r)$ in all the three-dimensional catalogues

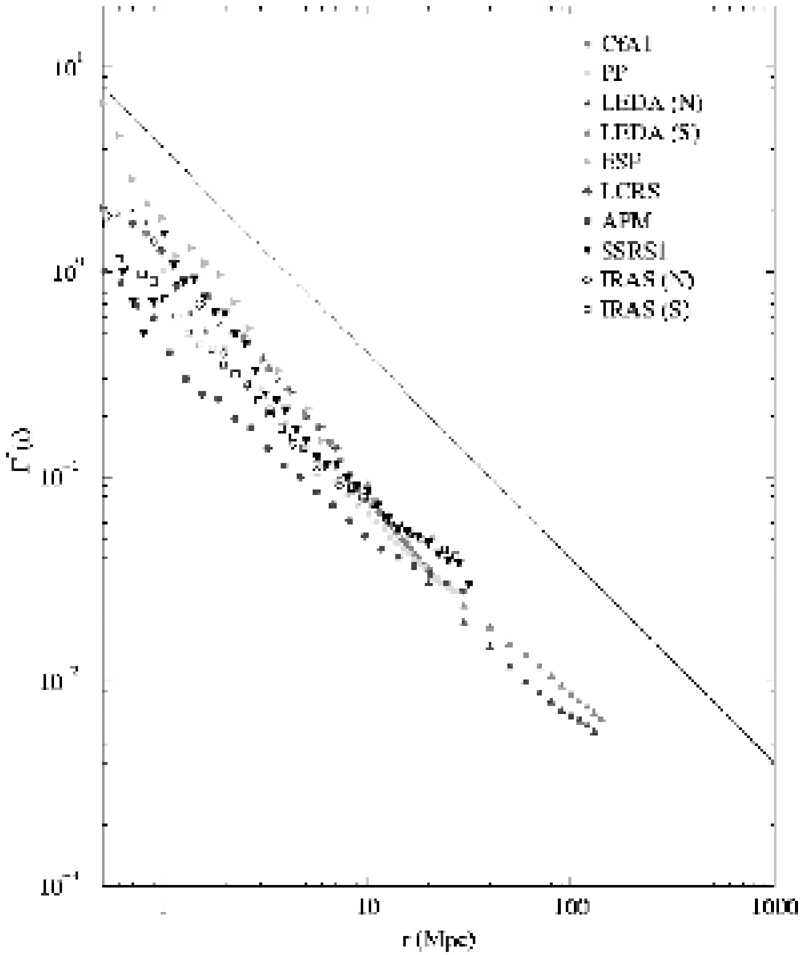

Figure 13.2. Full correlation for the same data of figure 13.1 in the range of distances 0.5–100h^{-1} Mpc. A reference line with a slope -1 is also shown (i.e. fractal dimension $D = 2$.

avalaible. Our analysis was carried out for VL samples; the results are collected in figure 13.2 [6].

We can derive the following conclusions:

- $\Gamma(s)$ measured in different catalogues is a *power law* as a function of the scale s, extending from approximately 1 to 40–50h^{-1} Mpc, without any tendency towards homogenization (flattening) [6]. This implies that all the optical catalogues show well-defined fractal correlations up to their limits, with the fractal dimension $D \simeq 2$ [6].

- Only in a single case, the LEDA database [17, 18], is it possible to reach larger scales of $\sim 100h^{-1}$ Mpc. This data sample has been largely criticized but, to our knowledge, never in a quantitative way. The statistical tests we performed show clearly that up to $50h^{-1}$ Mpc the results are completely consistent with all other data [6]. This agreement also appears to extend to the range 50–$100h^{-1}$ Mpc, with the same overall statistical properties found at smaller scales [6].
- We do not detect any difference between the various optical catalogues, as expected if they are simply different parts of the same distribution.
- Such results imply that the $\xi(s)$ analysis is inappropriate as it describes correlations as deviations from an assumed underlaying homogeneity. According to the $\Gamma(s)$ results, the value of s_0 (derived from the $\xi(s)$ approach) has to scale with the sample size R_s. The behaviour observed corresponds to a fractal structure with dimension $D \simeq 2$.

13.9 Interpretation of standard results

Here we attempt a comparison between the different interpretations.

In the standard interpretation, the rough constancy of s_0 for the different ML samples ($s_0 \simeq 4.5$–$8h^{-1}$ Mpc) and within the angular data is considered evidence for the validity of this approach. Moreover, since the samples have different volumes, these results should discount a *fractal* interpretation, which predicts an increase in s_0 with sample volume [13, 22].

In contrast, in the fractal approach, in our opinion, the analysis of the angular and ML samples is heavily biased by the use of the luminosity function and the corresponding homogeneity assumption. To measure the correlation function of such samples, one has to estimate the number of missing galaxies and their positions in the space. This is done by assuming the existence of a homogeneity scale. As an aside, we stress that the three-dimensional correlation in a fractal structure cannot be reconstructed in such a way from its angular features [6].

Regarding the *shape* of the $\xi(s)$, the difference from a power law is attributed:

(1) in the standard model, to the presence of *redshift distortions* [15]; and
(2) in the fractal model, to the fact that $\xi(s)$ is not in itself *a power law* [6]. If $\Gamma(s)$ is a power law, $\xi(s)$ is expressed by equation (13.6). In particular, it should be close to a power law only on very small scales and with an exponent $\gamma \sim 1$, as in the data reported by Guzzo [15].

With regards to the *VL results*, the increase in s_0 could be due to either of the following cases.

(1) *Luminosity segregation* (standard model). The increase in $\xi(s)$ corresponds to a real change in clustering properties for galaxy distribution, called *luminosity segregation*. This is considered just one aspect of the general

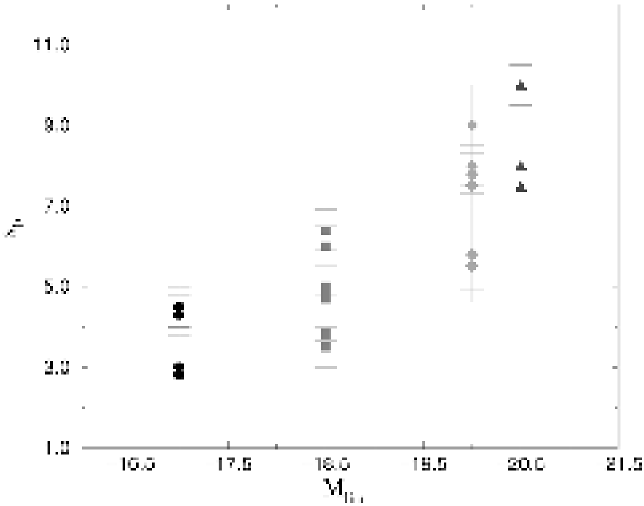

Figure 13.3. *Correlation length* s_0 versus sample luminosity $M_{\rm lim}$, for several VL samples. VL samples, with the same luminosity $M_{\rm lim}$, have different *volumes* and very different s_0. This is in contrast to *luminosity segregation* and in agreement with a fractal distribution inside the sample *Volume*.

expected dependence of the clustering features on the internal properties of galaxies, such morphology, colour, surface brightness and internal dynamics [15, 19–23].

(2) In the fractal model, the increase in $\xi(s)$ is just a geometrical effect and it is not related to any variation of the clustering of the corresponding data-set, as shown in equation (13.7). The effect is simply a byproduct of the inappropriate use of a statistical tool for the distribution under analysis [2,6].

The two interpretations seem to be equivalent; this is the reason why the same data-set is considered to confirm both fractality (and no luminosity segregation) and homogeneity (with luminosity segregation).

In our opinion there is a difference between the two interpretations: for the fractal case we have a quantitative prediction of an increase in s_0 within the sample size, while the theoretical expectation for luminosity segregation does not have a general consensus [24].

In principle it is possible to disentangle the two effects. A possible test is presented in figure 13.3. Here, we have reported the value of s_0 for the collection of VL samples versus the luminosity $M_{\rm lim}$ of the corresponding samples. Samples with the same $M_{\rm lim}$ have *different volumes*.

For each value of luminosity $M_{\rm lim}$ there is a range of values of s_0. These appear *in contradiction* to the *luminosity segregation effect*, according to which we should find only a single value for s_0 for samples with the same luminosity

$M_{\rm lim}$. Experimental uncertainities in the determination of s_0 do not explain such a spread. Conversely, the behaviour seems to be in agreement with a *fractal* distribution of galaxies within the sample size. The spread in the s_0 values for a single $M_{\rm lim}$ is due to the difference in the *volume* between different samples (in agreement with fractal prediction).

Acknowledgments

We warmly thank F Sylos Labini with whom we carried out the large part of this work. We are also grateful to A Gabrielli and M Joyce for many fruitful discussions. We would like to thank the organizers of the Graduate School in Contemporary Relativity and Gravitational Relativistic Cosmology for the invitation to the School. This work has been supported by INFM Forum Project: 'Clustering'.

References

[1] Peebles P J E 1993 *Principles of Physical Cosmology* (Princeton, NJ: Princeton University Press)
[2] Coleman P H and Pietronero L 1992 *Phys. Rep.* **213** 311
[3] Gaite J, Domnguez A and Pérez-Mercader J 1999 *Astrophys. J. Lett.* **522** L5
[4] Gabrielli A, Sylos Labini F and Durrer R 2000 *Astrophys. J. Lett.* **531** L1
[5] Mandelbrot B 1983 *The Fractal Geometry of Nature* (San Francisco, CA: Freeman)
[6] Sylos Labini F, Montuori M and Pietronero L 1998 *Phys. Rep.* **293** 66
[7] Montuori *et al* 1997 *Europhys. Lett.* **39** 103
[8] Hamilton A J S 1998 *The Evolving Universe* ed D Hamilton (Dordrecht: Kluwer Academic) p 185
[9] Peebles J P E 1998 *Les Rencontres de Physique de la Vallee d'Aosta* ed M Greco (Gif-sur-Yvette: Frontières)
[10] Strauss M A and Willick J A 1995 *Phys. Rep.* **261** 271
[11] Kolb E W and Turner M S 1989 *The Early Universe* (Frontiers in Physics) (Boston, MA: Addison-Wesley)
[12] Davis M 1997 *Critical Dialogues in Cosmology* ed N Turok (Singapore: World Scientific) p 13
[13] Guzzo L 1997 *New Astronomy* **2** 517
[14] Wu K K S, Lahav O and Rees M 1999 *Nature* **397** 225
[15] Guzzo L 1998 *Abstracts of the 19th Texas Symposium on Relativistic Astrophysics and Cosmology* ed J Paul and T Montmerle (Paris)
[16] Davis M and Peebles P J E 1983 *Astrophys. J.* **267** 465
[17] Paturel G, Bottinelli L and Gouguenheim L 1994 *Astron. Astrophys.* **286** 768
[18] Di Nella H *et al* 1996 *Astron. Astrophys. Lett.* **308** L33
[19] Davis M *et al* 1988 *Astrophys. J. Lett.* **333** L9
[20] Park C, Vogeley M S, Geller M and Huchra J 1994 *Astrophys. J.* **431** 569

[21] Benoist C *et al* 1999 *Astrophys. J.* **514** 563
[22] Willmer C N A, da Costa L N and Pellegrini P S 1998 *Astrophys. J.* **115** 869
[23] Cappi A *et al* 1998 *Astron. Astrophys.* **335** 779
[24] Colbert J M *et al* 2000 *Mon. Not. R. Astron. Soc.* **319** 209