CORRELATIONS AND SCALING IN THE QDOT REDSHIFT SURVEY

VICENT J. MARTÍNEZ¹ AND PETER COLES

Astronomy Unit, School of Mathematical Sciences, Queen Mary and Westfield College, Mile End Road, London E1 4NS, United Kingdom Received 1994 February 24; accepted 1994 June 22

ABSTRACT

We have calculated the redshift space two-point correlation function, $\xi(r)$, for galaxies in the QDOT redshift survey. In the range 1-30 h^{-1} Mpc, the data are well described by a power law $\xi(r) \simeq (r/r_0)^{-\gamma}$ with $\gamma = 1.27 \pm 0.11$ and $r_0 = 4.82 \pm 0.81$ h^{-1} Mpc. Furthermore, we find that the sample exhibits clear multifractal scaling properties, but with dimensions different from those extracted from the CfA survey. For example, $D_0 = 2.90$, $D_2 = 2.77$ for QDOT on large scales (>10 h^{-1} Mpc), compared with 2.1 and 1.3 obtained from the CfA survey; on small scales (<10 h^{-1} Mpc), the correlation dimension for QDOT galaxies is $D_2 \simeq 2.25$. The differences in these values between the QDOT data on large and small scales, and between the QDOT and CfA data, provide further evidence that multifractal scaling properties are a useful diagnostic of the dynamics of galaxy clustering.

Subject headings: galaxies: clustering — galaxies: formation — large-scale structure of universe

1. INTRODUCTION

The increasing availability of well-controlled samples of galaxy redshifts allows us to study the clustering of galaxies on larger and larger scales and in greater and greater statistical detail. To match the increasing amount of data, more sophisticated statistical descriptors have been added to the traditional tools for studying galaxy clustering, the correlation functions (Peebles 1980). One particular galaxy sample—the QDOT redshift survey of *IRAS* galaxies—has, in the last few years, yielded extremely important information about the statistical properties of galaxy clustering on very large scales using a variety of descriptors (Efstathiou et al. 1990; Saunders et al. 1991; Moore et al. 1992; Saunders, Rowan-Robinson, & Lawrence 1992; Moore et al. 1994).

In this paper we perform an analysis of the clustering of QDOT galaxies on large scales, in which the more traditional correlation function approach is complemented by a qualitatively different method based on the multifractal formalism. This latter technique employs a "technology" drawn from nonlinear dynamics to investigate the scaling properties of the clustering pattern. The hope is ultimately to be able to relate the scaling behavior of galaxy clustering to its dynamical origin. Unfortunately, we do not at present know how to do this convincingly (Efstathiou, Fall, & Hogan 1979; Jones, Coles & Martínez 1992). Nevertheless, scaling is one way of describing a spatial pattern, and it is, in some sense, orthogonal to the study of correlations. Indeed, as we shall see, it reveals properties of the spatial distribution which are only hinted at in the correlation functions.

It is particularly interesting to apply the scaling description to a sparsely sampled catalog like QDOT to see if it functions effectively in the presence of sampling noise and selection effects. Furthermore, these descriptors have usually been used to describe clustering in the strongly nonlinear regime (Jones et al. 1988; Martínez et al. 1990; Martínez & Jones 1990). If it is true that the various scaling dimensions we can extract are a diagnostic of the dynamical origin of galaxy clustering, then, at the very least, they should demonstrate a clear difference

between nonlinear and linear clustering. The results we obtain for QDOT should, hopefully, be different from those obtained, for example, from the CfA survey.

The layout of this paper is as follows: in the next section we describe the main features of the QDOT catalog; in § 3 we explain how we extract estimates of the two-point correlation function and show the results. Section 4 is devoted to the scaling analysis, and in § 5 we draw the conclusions.

2. THE SAMPLE

The QDOT redshift survey is described in detail by Lawrence et al. (1994), so we give only an outline here. We use a sample selected from the QDOT galaxy redshift survey of one in six of IRAS point sources with 60 μ m flux, S_{60} , exceeding 0.6 Jy. The catalog contains 2387 galaxies or other sources. We use the revised version of the catalog; earlier versions suffered from an error in some of the redshifts. Excluding sources with $|b| < 10^{\circ}$ and those without a measured radial velocity leaves 2086 galaxies. Distances are calculated by means of the Mattig formula

$$R = \frac{c}{H_0 q_0^2 (1+z)} \left[q_0 z + (q_0 - 1)(\sqrt{2q_0 z + 1} - 1) \right], \quad (2.1)$$

with $H_0 = 100 \ h \ {\rm km \ s^{-1} \ Mpc^{-1}}$ and $q_0 = 0.5$. We further impose the condition that the distance to the galaxies in the sample, R, must be in the range $10 \ h^{-1} \ {\rm Mpc} < R < 200 \ h^{-1}$ Mpc, so that we are dominated neither by local structures nor Jy large-scale sampling problems. The survey covers 74% of the sky, after removal of the regions with low Galactic latitude and the masked regions; see Lawrence et al. (1994) for a detailed description of the survey and the mask.

We define the selection function, $\varphi(R)$, which gives the probability that a galaxy at a distance R is included in the sample. It is straightforward to calculate this function using the published luminosity function of IRAS galaxies (Saunders et al. 1990); in this paper, we have used the parametric luminosity function recommended by Lawrence et al. (1994). we must also, however, impose an absolute luminosity cutoff in order that $\varphi(R) = 1$ for R less than some critical distance R_l ; we select $R_l = 40 \ h^{-1}$ Mpc. The galaxies in the QDOT catalog all have $S_{60} \ge 0.6$ Jy; the logarithm of the limiting luminosity at $40 \ h^{-1}$

¹ Permanent Address: Departament de Matemàtica Aplicada i Astronomia, Universitat de València, Burjassot, 46100 València, Spain.

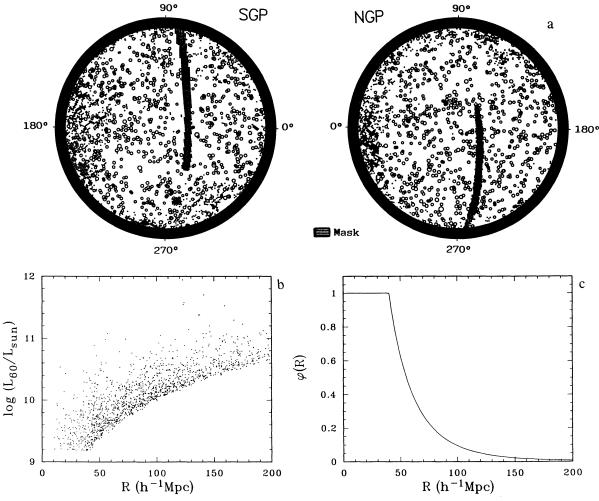


Fig. 1.—(a) Equal-area Lambert projection of the QDOT sample described in the text. The centers correspond to the south and the north Galactic poles. Galactic latitudes $|b| \le 10^\circ$ have not been considered. The circles representing galaxies are scaled proportionally to $1/\log(R)$, R being its distance. (b) The luminosity of all the galaxies in the sample as a function of the distance. (c) The selection function explained in the text.

Mpc is 9.19 in L_{\odot} units. The resulting sample has 1561 galaxies.

In Figure 1 we show the distribution of these galaxies on the sky, the distribution of galaxy luminosities with R, and the selection function of R.

3. THE TWO-POINT CORRELATION FUNCTION

We have estimated the two-point correlation function, $\xi(r)$, of the QDOT sample using the following estimators:

$$1 + \xi_{1\varphi}(r) = \frac{1}{N} \sum_{i=1}^{N} \frac{D_i(r)}{\bar{w} V_i(r)}, \qquad (3.1)$$

and

$$1 + \xi_{\varphi\varphi}(r) = \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i \frac{D_i(r)}{\bar{w}V_i(r)}, \qquad (3.2)$$

where

$$D_i(r) = \sum_{r-dr/2}^{r+dr/2} w_j , \qquad (3.3)$$

and the weights $w_j = 1/\varphi(r_j)$, with ϕ the selection function defined in § 2. The quantity $D_i(r)$ thus represents a sum over the j neighbors of a given galaxy i in a shell of radius r and thickness dr, with each neighbor assigned a weight w_j . The mean

weighted density $\bar{w} = 1/V \sum_{i=1}^{N} w_i$. The total volume of the sample is V and $V_i(r)$ is the volume of that part of a shell of thickness dr that lies within the survey region. This is less than the total volume of such a shell because of the operation of the Galactic latitude selection and the mask.

The estimator (3.1) is known to be very stable for the kind of sample we are dealing with here, whereas estimator (3.2) tends to emphasise the poorly sampled and most distant galaxies (Rivolo 1986; de Lapparent Geller & Huchra 1988; Martinez et al. 1993). Estimates of the redshift space correlation function for QDOT galaxies $\xi(r)$ have also been obtained by Moore et al. (1994) using a different estimator:

$$1 + \xi_{DR}(r) = \frac{DD(r)}{DR(r)} \left(\frac{N_R}{N_D}\right), \tag{3.4}$$

where DD(r) is the number of pairs with separation r in the catalog of N_D objects and DR(r) is the number of pairs with separation r between the actual data and a random catalog of N_R objects. The weights used by Moore et al. (1994) are also different and are based on the correlation integral J_3 (Peebles 1980). Our estimators (3.1) and (3.2) are not optimal (Hamilton 1993), but we use them to get some idea of the robustness of different estimates of $\xi(r)$. The estimators (3.1) and (3.2) are also more closely related to the cell-count moment statistics we

552

In the range 1-30 h^{-1} Mpc, the correlation function is well fitted by a power law $(r/r_0)^{-\gamma}$ with $\gamma = 1.27 \pm 0.11$ and $r_0 = 4.82 \pm 0.81 \ h^{-1}$ Mpc. The open circles, corresponding to $\xi_{\varphi\varphi}$. are more scattered, but the power-law trend is clearly still obeyed. The slope and amplitude we have obtained for the QDOT galaxies is in good agreement with that obtained by Moore et al. (1994) for the same catalog and with the estimate of the redshift space correlation function of 1.2 Jy IRAS galaxies calculated by Fisher et al. (1994). Note, however, that the real space correlation function inferred by Fisher et al. (1994) differs significantly from the redshift space version because of the effect of peculiar velocities; they find $r_0 = 3.76 \ h^{-1} \ \mathrm{Mpc}$ and $\gamma = 1.66$ in real space. An indirect estimate of the real space correlation function by Saunders et al. (1992) using a cross-correlation technique produces a result consistent with this. The slope of $\xi(r)$ even after correcting for redshift-space distortions is shallower than that of the correlation function of optical galaxies (Davis & Peebles 1983) on small scales, consistent with claims of extra large-scale power. On the other hand, we find no evidence for a change in slope of the correlation function with distance for the QDOT galaxies. Indeed, if we try to fit a power law profile to $1 + \xi(r)$ rather than $\xi(r)$ itself, as

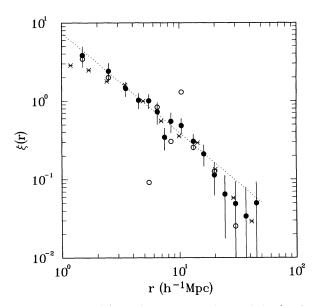


Fig. 2.—Estimates of the redshift-space two-point correlation function for the QDOT data. We show our results using two different estimators (filled circles for $\xi_{1\varphi}$ and open circles for $\xi_{\varphi\varphi}$, as well as the Moore et al. (1994) results [asterisks]). The dotted line is the best power-law fit to $\xi_{1\varphi}$ in the range 1–30 h^{-1} Mpc. Errors are calculated from the 100 bootstrap resamples.

suggested by Guzzo et al. (1992), we get a much poorer fit. The QDOT data on large scales therefore do not seem to display the same behavior as the Perseus-Pisces sample analyzed by Guzzo et al. (1992). To investigate this question in some more detail, we tried analyzing that part of the QDOT sample that lies only within the Perseus-Pisces region covered by the Guzzo et al. analysis: $22^h \le \alpha \le 4^h$, $0^\circ \le \delta \le 45^\circ$. Unfortunately, however, there are too few galaxies to make any reasonable fit to $1 + \xi(r)$ so that this question must remain open.

4. SCALING AND DIMENSIONALITY

4.1. Motivation

Consider the relationship between the moments of counts of neighbors and the two-point correlation function given by Peebles (1980, § 36). If a spherical cell of radius r is centered upon an object (labeled by i), then $n_i(r)$ is the count of objects in the cell excluding the central one. Averaging over all the N points in the sample we get the mean count

$$\bar{n} = \frac{1}{N} \sum_{i=1}^{N} n_i(r) . \tag{4.1}$$

The relationship among the mean number density ρ , the mean count of neighbors, \bar{n} , and the two-point correlation function $\xi(r)$ is

$$\bar{n} = \rho \int_0^r [1 + \xi(r)] 4\pi s^2 ds$$
 (4.2)

We will say that the first moment of the counts of neighbors scales if

$$\bar{n} \propto r^{D_2}$$
, (4.3)

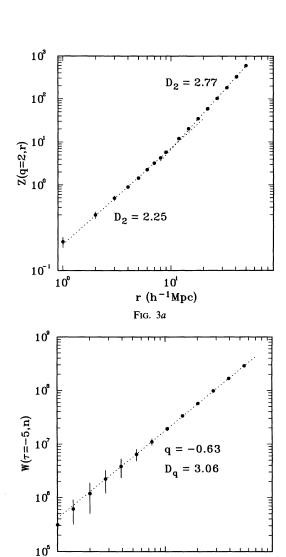
where D_2 is a constant. Most previous analyses of galaxy clustering probe the strongly nonlinear regime, where $\xi(r) \gg 1$. In this regime, $\xi(r) = (r/r_0)^{-\gamma}$ (Davis & Peebles 1983) and we can approximate $D_2 = 3 - \gamma$. In the regime where clustering is not strong—the case dealt with in this paper—it is not possible to extract a value of D_2 (Martínez & Jones 1990) directly from the slope of $\xi(r)$.

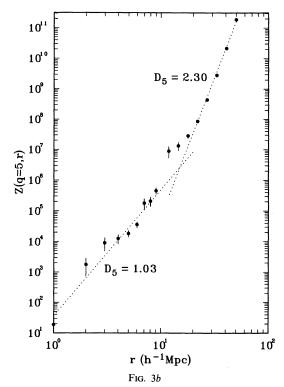
An obvious generalization of the scaling of the first moment of the counts of neighbors is to look at the scaling of the moments of any order:

$$Z(q, r) = \frac{1}{N} \sum_{i=1}^{N} n_i(r)^{q-1} \propto r^{\tau(q)}, \qquad (4.4)$$

where is principle one can have a different scaling exponent τ for each value of q. Our Z(q, r) is normalized differently than the "standard" definition, which is obtained by replacing $n_i(r)$ with $n_i(r)/N$ in equation (4.4). This makes no difference to the analysis of scaling indices. When q=2 we obviously recover equation (4.3). The Renyi dimensions (or generalized dimensions) are defined from the scaling exponents $\tau(q)$ by $D_q = \tau(q)/(q-1)$. If the relationship (4.4) holds for a wide range of r we will say that the point set we are analyzing is multifractal in nature.

When q > 2 the meaning of the moments of the neighbors is clear: as q increases the denser parts of the set dominate the sum in equation (4.4). The scaling exponents $\tau(q)$ (and therefore the D_q) give us information predominantly about the high-density regions of the set. Likewise, if $q \to -\infty$ the sum in equation (4.4) will be dominated by low-density regions. However, because the cell has been centered on a particular





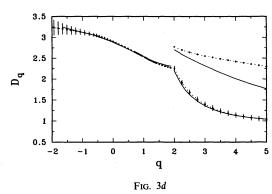


Fig. 3.—Scaling properties of the QDOT data. The panels show (a) the count-moment sum Z(q, r) (eq. [4.4]) for q = 2; (b) the count-moment sum Z(q, r) (eq. [4.4]) for q = 5; (c) the inverse sum $W[\tau(q), n]$ (eq. [4.5]) for $\tau = -5$; and (d) the generalized dimensions D_q [for $q \ge 2$, we show (upper lines) the dimensions for the range 10-50 h^{-1} Mpc and (lower lines) those for the range 1-10 h^{-1} Mpc; for $q \le 2$, we have used all values of n in the range 10-50]. In all these panels, the dotted lines correspond to the best fits taking errors into account and the solid lines are unweighted fits; errors are based on 100 bootstrap resamples of the data.

 10^3

object (galaxy) and the galaxies tend to be clustered, the scaling does not always hold for $q \le 2$. This is a common numerical problem in the estimation of the generalized dimensions for a point set, even one which is known a priori to be multifractal. To resolve this problem, we can consider instead an inverse function of $n_i(r)$. Let us denote by $r_i(n)$ the radius of the smallest sphere centered at a point i and enclosing n neighbors. In an unbounded point set $r_i(n)$ is just the distance to the nth nearest neighbor. The inverse relationship we need to investigate scaling is

102

n Fig. 3c

10¹

$$W[\tau(q), n] = \frac{1}{N} \sum_{i=1}^{N} r_i(n)^{-\tau} \propto n^{1-q} . \tag{4.5}$$

It is particularly interesting to evaluate equation (4.5) at q = 0:

$$W[\tau(0), n] = \frac{1}{N} r_i(n)^{-\tau(0)} \propto n.$$
 (4.6)

The value of $D_0 = -\tau_0$ is an estimator of the Hausdorff dimension (or, more correctly, the capacity dimension: Grassberger, Badii, & Politi 1988). This quantity is more directly related to the geometry of the set than the correlation dimension D_2 . In a multifractal set $D_2 < D_0$, while in a simple fractal $D_2 = D_0$.

The dimension D_0 is also related to the scaling properties of the void probability function (VPF). Let $P_0(r)$ be the probability that a randomly selected cubic cell of side r contains

554

no galaxies. If r is expressed in units of $V^{1/3}$, where V is the sample volume, then the number of occupied cells $N_{\rm oc}(r) = r^{-3}[1-P_0(r)]$. The quantity $N_{\rm oc}(r) \propto r^{-D_0}$ (Jones et al. 1992). This turns out not to be a particularly effective way to estimate D_0 because of problems with discreteness, but it does furnish a simple way to interpret the geometrical information carried by D_0 . We shall return to the relationship between D_0 and the VPF in future work.

4.2. Analysis

It is well established that equation (4.4) converges better for $q \ge 2$ (dominated by high-density regions) and equation (4.5) converges better for q < 2 (dominated by low-density regions), because n_i increases in the presence of neighbors while $r_i(n)$ increases with a lack of neighbors. Our procedure is therefore to use equation (4.4) to determine the dimensions with $q \ge 2$ and equation (4.5) for q < 2. The details of the procedure for dealing with edge corrections and the effect of the selection function can be found in Borgani et al. (1994). Applying this procedure to the QDOT sample yields results displayed in Figure 3 for various values of q.

In these plots we have used 100 bootstrap resamples to determine the errors on equation (4.4) and 40 bootstrap resamples for equation (4.5), and results are shown using the errors in the fit and also, for comparison, the fits obtained without errors. Note that the scaling exponents $\tau(q) = D_q/(q-1)$ so the errors have been translated to D_q for plotting the relevant figure. It is clear that the effect of "lacunarity" (Borgani et al. 1994) is much stronger in the case where q=5 than when q=2. One can see clearly that when one considers different scale ranges $10-50~h^{-1}$ Mpc and $1-10~h^{-1}$ Mpc for $q \ge 2$, a different value of D_q is obtained. This is not surprising since we expect these two regimes to be dominated by linear and nonlinear clustering respectively. It is encouraging that one obtains different D_q in these two regimes, even when there is no clear change in the behavior of $\xi(r)$.

It is not possible to separate different scales clearly using the function $W(\tau, q)$, so we have considered all values of $10 \le n \le 550$. We do not use all n down to unity, because the estimator is known to suffer from a systematic error when $n \le N$ (Grassberger 1988). The figure shows W(-5, n), from which the slope obtained is 1 - q = 1.63, so q = -0.63 and $D_q = 3.06$. Again, we have shown the corresponding D_q plot using error-weighted and unweighted data.

Notice that, because Z and W are both integral quantities the errors get progressively smaller as the argument increases, i.e., Z(q, r) is more accurate for large r and $W(\tau, n)$ is more accurate for large n. More generous and probably more realistic error bars for the estimates for $q \ge 2$ are given by the difference between weighted and unweighted estimates, rather than by these formal bootstrap errors.

We do obtain quite good scaling for the QDOT data, despite the sparseness of the sample and the fact that it probes mainly the weakly nonlinear regime. It is interesting to focus upon D_2 and D_0 . From the CfA survey (Martinez et al. 1990), one obtains $D_0 = 2.1$ and $D_2 = 1.3$. We obtain an estimate of $D_0 = 2.90 \pm 0.02$ for QDOT, which we can assume is valid for all scales. The values for D_2 vary from 2.25 on small scales to 2.77 on large scales.

Jones et al. (1992) defined an intermittency exponent by $I=(D_0-D_2)/2$; for the CfA data this exponent is I=0.4. The intermittency exponent for the QDOT galaxies varies with scale: I=0.325 ($<10\,h^{-1}$ Mpc), i.e., close, but not quite equal, to the CfA number; I=0.065 ($>10\,h^{-1}$ Mpc). This indicates that the large-scale spatial distribution is less intermittent, i.e., less heterotopic (Jones et al. 1992), which, in turn, agrees with the smaller amount of nonlinear dynamical evolution we expect to see on scales probed by the QDOT sample compared with CfA.

5. CONCLUSIONS

We have calculated the two-point correlation function, $\xi(r)$, for *IRAS* galaxies in redshift space and found that it is well modeled by a single power law over the range of scales considered. Our results agree with those of Moore et al. (1994) and Fisher et al. (1994) for the redshift space correlation function; these analyses also suggest that the *real space* correlation function has a somewhat steeper slope than the redshift space version.

We find no evidence for the significant shoulder in the correlation function seen by Guzzo et al. (1992). This latter sample may be dominated by a single large-scale feature.

The scaling analysis we have performed using the multifractal formalism demonstrates the usefulness of this technique as a diagnostic of galaxy clustering. Although there is not a strong signal in the two-point correlation function between small scales (nonclustering) and large scales (linear clustering), the scaling dimensions are clearly different in these two regimes. This shows that our descriptors can distinguish between these two physically distinct domains. Moreover, even on small scales ($\leq 10~h^{-1}$ Mpc), the scaling properties of the QDOT data are different from those of, for example, the CfA survey. Since the latter contains a much larger fraction of bright ellipticals than the QDOT data (which is dominated by spiral galaxies), this is an interesting indication that our method can elucidate differences in the clustering of these two types of galaxy which are otherwise difficult to detect.

The ability of this method to show up differences between linear and nonlinear regimes, and between different types of galaxy, shows that it is indeed a potentially useful statistical descriptor, though more theoretical work is needed to understand the physical implications of these results.

Vicent Martínez received a grant from the Conselleria de Educació i Ciència de la Generalitat Valenciana. Peter Coles is an SERC Advanced Fellow. We acknowledge the use of the QMW starlink and IFIC facilities during this work. We thank Andy Lawrence for letting us use some of his subroutines for the QDOT analysis, and we thank Ben Moore for providing us with results from Moore et al. (1994) prior to publication. We thank the (anonymous) referee for comments which helped us to clarify the presentation of this work. This work has been partially supported by the DGICYT project PB 90-0416, Spain, and by the European Community Human Capital and Mobility network project "Large-Scale Structure in the Universe: Evolution and Statistics" (CHRX-CT93-0129).

REFERENCES

Borgani, S., Martínez, V. J., Pérez, M. A., & Valdarnini, R. 1994, ApJ, in press Davis, M., & Peebles P. J. E. 1983, ApJ, 267, 465
de Lapparent, V., Geller, M. J., & Huchra, J. P. 1988, ApJ, 332, 44
Efstathiou, G., Ellis, R. S., Frenk, C. S., Kaiser, N., Lawrence, A., Rowan-Robinson, M., & Saunders, W. 1990, MNRAS, 247, 10P
Efstathiou, G., Fall, S. M., & Hogan, C. J. 1979, MNRAS, 189, 203
Fisher, K. B., Davis, M., Strauss, M. A., Yahil, A., & Huchra, J. P. 1994, MNRAS, 266, 50
Grassberger, P. 1988, Phys. Lett. A., 128, 369
Grassberger, P., Badii, R., & Politi, A. 1988, J. Stat. Phys., 51, 135
Guzzo, L., Iovino, A., Chincarini, G., Giovanelli, R., & Haynes, M. P. 1992, ApJ, 382, L5
Hamilton, A. J. S. 1993, ApJ, 417, 19
Jones, B. J. T., Coles, P., & Martínez, V. J. 1992, MNRAS, 259, 146
Jones, B. J. T., Martínez, V. J., Saar, E., & Einasto, J. 1988, ApJ, 332, L1
Lawrence, A., et al. 1994, MNRAS, in press

Ling, E. N., Frenk, C. S., & Barrow, J. D. 1986, MNRAS, 223, 21P
Martínez, V. J., & Jones, B. J. T., 1990, MNRAS, 242, 517
Martínez, V. J., Jones, B. J. T., Dominguez-Tenreiro, R., & van de Weygaert, R.
1990, ApJ, 357, 50
Martínez, V. J., Portilla, M., Jones, B. J. T., & Paredes, S. 1993, A&A, 280, 5
Mo, H. J., Jing, Y. P., & Borner, G. 1992, ApJ, 392, 452
Moore, B., Frenk, C. S., Efstathiou, G., & Saunders, W. 1994, MNRAS, 269,
742
Moore, B., et al. 1992, MNRAS, 256, 477
Peebles, P. J. E. 1980, The Large-Scale Structure of the Universe (Princeton: Princeton Univ. Press)
Rivolo, A. R. 1986, ApJ, 301, 70
Saunders, W., et al. 1991, Nature, 349, 32
Saunders, W., Rowan-Robinson, M., & Lawrence, A. 1992, MNRAS, 258, 134
Saunders, W., Rowan-Robinson, M., Lawrence, A., Efstathiou, G., Kaiser, N.,
Ellis, R. S., & Frenk, C. S. 1990, MNRAS, 242, 318