THE ASTROPHYSICAL JOURNAL, 332:L1–L5, 1988 September 1

MULTIFRACTAL DESCRIPTION OF THE LARGE-SCALE STRUCTURE OF THE UNIVERSE

BERNARD J. T. JONES, VICENT J. MARTÍNEZ,[1] ENN SAAR,[2] AND JAAN EINASTO[2]
NORDITA, Copenhagen, Denmark
Received 1987 June 4; accepted 1988 May 17

ABSTRACT

We show that the clustering structures in the observed universe and in numerical simulations are not well represented by a homogeneous measure on a fractal. We propose instead that a good description of these clustering structures is given by "multifractals"—fractals having more than one scaling index.

We evaluate the multifractal characteristics of a data sample and a numerical simulation of an axion-dominated universe. The clustering structures revealed by the multifractal analysis are quite similar, both showing a variation of dimensionality from 1 to 3 over similar ranges of scales. The multifractal description provides a natural way of scaling the numerical models to the data.

If a bias is applied to the simulation, the result is a distribution of points that has almost constant dimension, 1.5, on all scales. This reduction of the distribution of dimensionality by biasing may be a problem for theories that invoke naive biasing schemes.

Subject headings: cosmology — galaxies: clustering

I. INTRODUCTION

As catalogs of redshifts of galaxies (Tully and Fisher 1978; Kirshner *et al.* 1981; de Lapparent, Geller, and Huchra 1986) have become available, we have seen a complexity of structure previously only hinted at by the earlier surveys of galaxies projected on the celestial sphere (Soneira and Peebles 1978). We see in these surveys "voids," "filamentary," "bubble," and "sponge" structures, but these remain qualitative descriptors and so provide only subjective bases for comparison between theory and observation.

Characterizing the large-scale clustering structure of our universe is an important problem with a long history. Not only do we wish to understand the nature of the large-scale clustering, but we also wish to compare it with models purporting to explain that structure. The most common way of quantifying the clustering has been to use the two-point and higher order correlation functions. However, the interpretation of these low-order correlation functions is not without ambiguity.

The new view of the structure from the redshift surveys suggests that the fractal description (Efstathiou, Fall, and Hogan 1979; Peebles 1980; Mandelbrot 1977, 1982) may be only a first approximation: there appears to be a density-dependent quality to the clustering in the sense that the largest and most rarefied structures are the voids, whereas the denser systems show a filamentary character.

In this paper we shall show how to quantify this apparent density-dependent structuring, using a recently developed generalization of the fractal concept—the multifractal. The multifractal provides us with both a measure of the dimensionality on various length scales and a relative frequency of the structures of various dimensions that are identified on these scales.

II. THE DATA SAMPLES

In this paper we use two data samples: a version of the CfA catalog of Huchra's compilation of redshifts and a simple, axion-dominated N-body simulation of gravitational clustering.

From the CfA redshift survey we selected a cone centered on the north Galactic pole with $b \geq 40°$ with $v_{\max} = 800$ km s^{-1}, the limiting absolute magnitude (for $H = 50\,h$ km s^{-1} Mpc^{-1}) being -21.5 and the total number of galaxies 452. This sample is relatively complete and has been used before to study the behavior of the correlation function (Einasto, Klypin, and Saar 1986).

For comparison we used data from a numerical simulation of an axion-dominated cosmological model with nonzero Λ-term, which has been claimed to represent the observations rather well (Gramann 1987). We shall also look at a biased version of this sample created by selecting the galaxies lying in the peaks of the initial density distribution. The threshold is chosen so as to select out 10% of the initial sample.

III. IS COSMIC CLUSTERING A FRACTAL?

To calculate the box or fractal dimension of the different samples we are working with, we use a statistically corrected box-counting method, which provides good accuracy, even for small data sets. The box dimension of a fractal point set can be calculated by partitioning the sample volume into cells of size ϵ and counting for each value of ϵ the number of occupied cells $N(\epsilon)$. The plot of $\log N(\epsilon)$ versus $\log(1/\epsilon)$ yields a straight line for a fractal, and its slope is the fractal dimension.

If the number of points in the data set is not too large, the estimate of $N(\epsilon)$ will depend sensitively on the detailed position of the grid. To avoid this problem, we could replace $N(\epsilon)$ by a mean taken over several realizations of the random process, $\bar{N}(\epsilon)$. However, while we could in principle generate many realizations of a numerical simulation, we in fact have only one realization of the observed universe. So we generate different samples by simply shifting the counting grid from one position to another with a random vector $\boldsymbol{a}$ having components $|a_i| < \epsilon$. We have verified the effectiveness of this technique in a variety of situations. Finding $\bar{N}(\epsilon)$ for different box sizes ϵ, the dimension D associated with a given value of ϵ will be

$$D = \frac{d[\log \bar{N}(\epsilon)]}{d[\log(1/\epsilon)]}. \qquad (1)$$

If this value is constant for a particular range of scales ϵ, the set has fractal behavior in that range.

[1] Departament de Matemàtica Aplicada i Astronomia, Universitat de València, Spain.
[2] Tartu Astrophysical Observatory, Toravere 202444, Estonia, USSR.

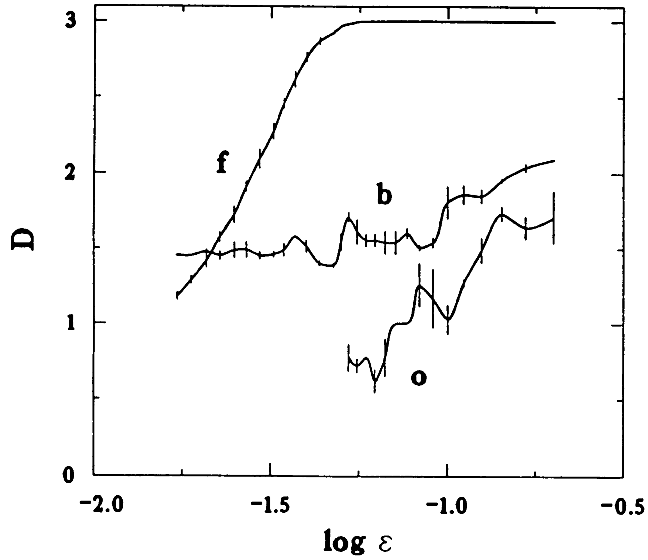

FIG. 1.—Scaling behavior of the fractal dimension (capacity) D with the size ϵ for observations (curve o), the full axion model (curve f) and the biased version (curve b). Plotting D vs. log ϵ, one would observe constant behavior for simple fractals. Instead, for the observations and the full model there is significant evidence for an increase of dimensionality with scale size. Only the biased data display scale-independent dimensionality (though it is not a simple fractal: see Fig. 2).

The results for our samples are given in Figure 1, where the observations are denoted by o, the full model (all particles used) by f, and a biased version (bias level 0.1) by b. The scale ϵ is expressed in the units of the side of a cube enclosing the sample.

As we see, the box dimension (capacity) of the observed sample, as well as that of the unbiased simulation data, keeps growing with the scale size, ruling out homogeneous fractal models. It is surprising, however, that the biased version of the simulation does have a plateau on the D versus log ϵ plot. However, one should beware of jumping to the conclusion that this distribution is a homogeneous fractal simply because it displays a scale-independent dimensionality. The condition that the fractal dimension be constant is necessary but not sufficient for the distribution to be a homogeneous fractal. We shall show that the biased numerical model (curve b) is not a homogeneous fractal despite its scale-independent dimensionality.

The observed data set is relatively small (452 points), and we should consider the possibility that this method of obtaining the fractal dimension is not convergent. As a check, we can calculate directly the correlation dimension D_2, which is always a lower bound on the fractal dimension D_0 (Grassberger and Procaccia 1983). D_2 has been calculated as the exponent of a power-law fit for the correlation integral at small distances:

$$C(\epsilon) = \int_0^{\epsilon} 4\pi r^2[1 + \xi(r)]dr \propto \epsilon^{D_2} . \qquad (2)$$

D_2 for the observed data sample has a constant value of 1.3. [The two-point correlation function $\xi(r)$ for this sample is a power law of slope $\gamma = 1.70 \pm 0.04$.]

Since the simple fractal description manifestly fails to describe the distributions, we shall try to find whether the structure can be represented as a mixture of different fractal sets using the recently developed multifractal ideas and techniques (Hasley *et al.* 1986). (See Mandelbrot 1986 for a comment on the history of multifractals and fractals.)

The fact that the observed data sample (o) and the unbiased numerical simulation sample (f) have the same slope is significant. It indicates that the clustering structures are similar from the point of view of their dimensionality. The lateral displacement of the curves relative to one another comes from the different number of points in each sample and is not relevant in terms of fractal dimensionality. This lateral displacement does, however, indicate how to scale the numerical model to the observed data.

The dimensionality of the sample must tend to 3 as the sample cell size increases. This is because beyond some particular scale all the cells must be occupied. The point at which the dimensionality attains this plateau gives a characteristic scale to the point distribution which is presumably related to other clustering scales, such as the scale on which the correlation function falls to unit amplitude or the percolation length scale.

IV. THE MULTIFRACTAL FORMALISM

To be specific to cosmology, let us consider a point set embedded in a three-dimensional Euclidean space, each point representing the position of a galaxy. The first step in developing the multifractal formalism is to assign a probability measure to the set (Mandelbrot 1982). A natural way to do this is to partition the sample volume into cells (cubes) of size ϵ (ϵ small) and then to count the number of points within each cell, $n_i(\epsilon)$. The probability that a random point falls in a given cell is then

$$p_i(\epsilon) = \frac{n_i(\epsilon)}{N} , \qquad (3)$$

where the subscript i labels the cells and N is the number of points in the sample.

The moments of this measure define a new function,

$$\tau(q) = \lim_{\epsilon \to 0} \frac{\log \sum_{i=1}^{N(\epsilon)} [p_i(\epsilon)]^q}{\log \epsilon} . \qquad (4)$$

Homogeneous fractals display a linear dependence of τ on q, whereas any departure of linearity implies multifractality. Therefore, $\tau(q)$ provides a generalization of the concept of dimension (eq. [1]) to sets that are not homogeneous fractals. It is related to the generalized dimension

$$D_q = (q - 1)^{-1}\tau(q) \qquad (5)$$

(Hentschel and Procaccia 1983). D_0 is the Hausdorff dimension, while D_1 is the so-called information dimension and D_2 is the so-called correlation dimension. The D_q form an infinite set of relevant dimensions characterizing chaotic systems. It is easy to see from equation (5) that D_q is always decreasing, i.e., if $q \geq q'$ then $D_q \leq D_{q'}$.

It is useful to note that large values of q select out the terms where p is largest, in other words the denser regions of the distribution. Likewise, the large negative values of q refer to the least dense regions.

Obviously, $\sum_{i=1}^{N} p_i^q$ is directly related to the familiar moment generating function of the underlying random point set distribution (Peebles 1980, § 38). Thus the function D_q provides a compact and intuitively accessible way of describing what has hitherto been buried in the higher statistical moments of the galaxy distribution.

There is an alternative description of such a distribution which some may find more appealing than D_q. Following a standard procedure dating back to Hölder (Mandelbrot 1982; Jensen *et al.* 1985), we can make the physically reasonable assumption that the cell occupancy probabilities $p_i(\epsilon)$ are related to the choice of ϵ by a power-law expression

$$p_i = \epsilon^{\alpha_i} . \tag{6}$$

The α_i are the scaling indices of the set for that choice of ϵ. In a uniform measure on a fractal there will be only one scaling index, and that will be related to the Hausdorff dimension of the fractal. Studying how the values of the scaling indices are distributed provides a useful tool to characterize the geometry of the point set.

The number of times the scaling index α takes a value in the interval $[\alpha', \alpha' + d\alpha']$ can be expressed as

$$n(\alpha')d\alpha' = \epsilon^{-f(\alpha')}\, d\alpha' , \tag{7}$$

where $f(\alpha)$ is the fractal dimension of the points in the set which have the same value for the scaling index α, and will in general be a continuous function. The function $f(\alpha)$ is the distribution of dimensionalities that are present in the set. It plays a central role in the theory of multifractals, and it is often referred to as the "$f(\alpha)$ spectrum."

The dimensionality structure of the set is equally well characterized by $\tau(q)$, D_q, or $f(\alpha)$. It is a matter of taste as to which gives a better feeling for the character of the distribution. We can easily transform from one description to another. It can be shown (Hasley *et al.* 1986) that the variables (q, τ) are related to (α, f) through the Legendre transformation

$$\tau(q) = q\alpha - f(\alpha) , \tag{8}$$

$$\alpha(q) = \frac{d\tau}{dq} . \tag{9}$$

Thus it is easy to obtain $f(\alpha)$ by calculating the envelope of the straight lines: $y = qx - \tau(q)$ as q varies from $-\infty$ to $+\infty$. Mandelbrot (1988) has introduced the $f(\alpha)$ curve by rephrasing the Gibbs formalism of thermodynamics.

The curve $f(\alpha)$ has a unique maximum, and the value of the spectrum in that point is the fractal dimension, $f(\alpha_{\rm H})$, of the set. Directly from equation (8) we obtain $f'(\alpha) = q$, and hence the maximum of $f(\alpha)$ is reached when $q = 0$, corresponding to $f(\alpha_{\rm H}) = -\tau(0) = D_0$.

We can see that α takes values in a finite range $[\alpha_{\min}, \alpha_{\max}]$, where

$$\alpha_{\min} = \lim_{q \to \infty} D_q , \quad \alpha_{\max} = \lim_{q \to -\infty} D_q ;$$

$\alpha_{\min}$ is the scaling exponent of the region of the set where the concentration of points is maximum, and $\alpha_{\max}$ is the scaling exponent of the most rarefied parts. (This interpretation follows from the observation made earlier that large positive values of q select out the densest parts of the set, while large negative values of q select out the least dense parts.)

It should be noted that in practice we have a finite sample of discrete points, and so we cannot in fact take the limit $N \to \infty$ or $\epsilon \to 0$. We can merely estimate what would happen if this limit were attainable through a larger data sample. [See Jones and Martínez 1988 for a discussion of this and other technical points involved in calculating the $f(\alpha)$ spectrum.]

V. RESULTS AND CONCLUSIONS

The results for the data sets described above are shown in Figure 2. We have plotted the generalized dimensions D_q, as well as the $f(\alpha)$ spectrum. [Conceptually, we find $f(\alpha)$ to be the easier function to appreciate intuitively, although, as we said before, that is merely a matter of taste.] Note that we have elected to plot the curves for particular choices of the cell size ϵ, and this means that the curves are not directly comparable. To draw comparable curves would necessitate properly scaling the simulations. Also, because of the discreteness of the samples, we cannot let the cell size become arbitrarily small as demanded by the definition (4). We do not consider this restriction to be a problem for our description of these point distributions as multifractals.

The analysis reveals a whole range of scaling indices α in all three data samples. So none of the samples is a simple fractal, not even the biased sample that displayed the interesting scale-independent dimensionality. For each sample, the maximum of the $f(\alpha)$ curve yields the fractal dimension for the chosen value of ϵ, and that is what is plotted in Figure 1.

Our conclusions are summarized as follows:

1. The distribution of galaxies in the universe is not well represented by a simple fractal. The failure of the uniform measure on a fractal is shown in Figure 1, where we see that the box dimension of the CfA sample (curve *o* in the figure) changes from 0.5 to 1.5 over one decade of scale lengths. However, as we have mentioned before, fractal behavior may still be expected with dimension $D_0 \geq D_2 \approx 1.3$. In fact, the sample is well represented by a multifractal, and D_0 should be the maximum of the $f(\alpha)$ curve (see Fig. 2).

2. The numerical models are capable of producing structures having the same range of dimensionality as the data (curve *f* in Fig. 1). The similarity of the curves *f* and *o* in Figure 1 (up to a scaling factor) indicates that the clustering structures are similar, at least as far as the distribution of fractal dimension is concerned, and that with an obvious rescaling they could be made to match.

3. We would be encouraged to reject the biased model *b* on the grounds that it had a clustering structure differing manifestly from that of the real universe. Biasing the model tends to reduce the spread in dimensionality because selecting only objects above a given high threshold removes points from the more spherical void regions of the simulation. Whether or not this is an argument against simple threshold biasing in general remains to be seen.

4. As an answer to the question "If not fractals, then what?" we have suggested that multifractals provide an important descriptor of the large-scale clustering of the universe and of numerical simulations (as was pointed out by Mandelbrot 1982). In Figure 2 we see the distribution of dimensionality in each sample when viewed on some length scale. The maximum of the $f(\alpha)$ curves corresponds to the fractal dimension, and, as we saw from Figure 1, this changes with the sampling length scale. A complete explanation of clustering in the universe should reproduce these curves in detail.

5. The final point is a caveat. We have departed from the usual description of multifractals (Hasley *et al.* 1986) in not taking the limits as the sample box size tends to zero (cf. eq. [4]). This is a consequence of the fact that our samples are discrete point sets and so have dimension zero in the limit of infinitesimally small sample cells. We shall discuss this approximation in detail elsewhere.

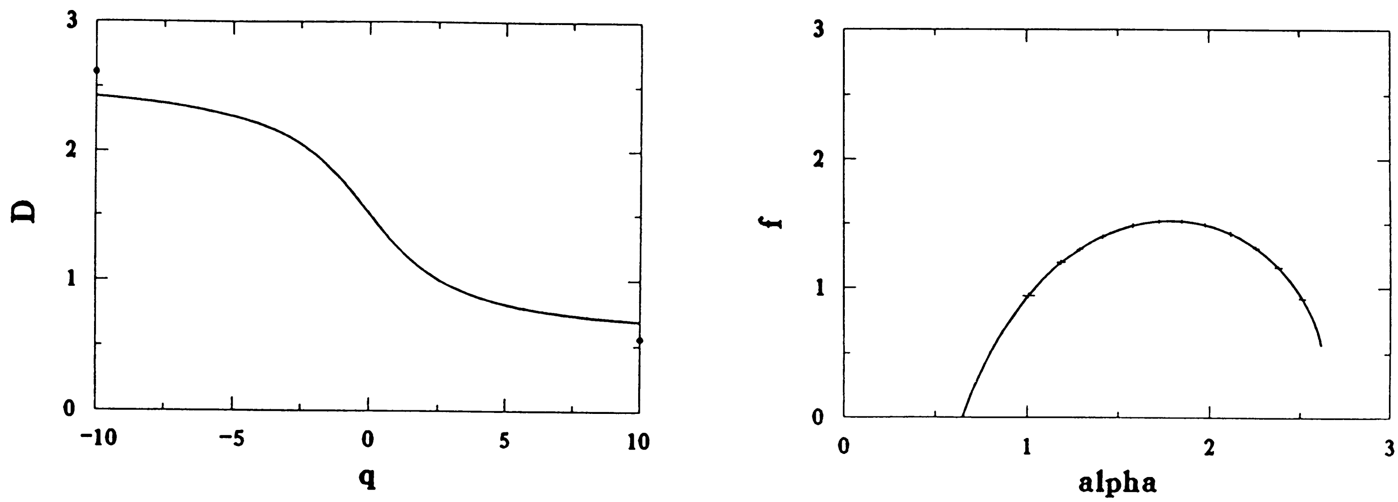

FIG. 2*a*

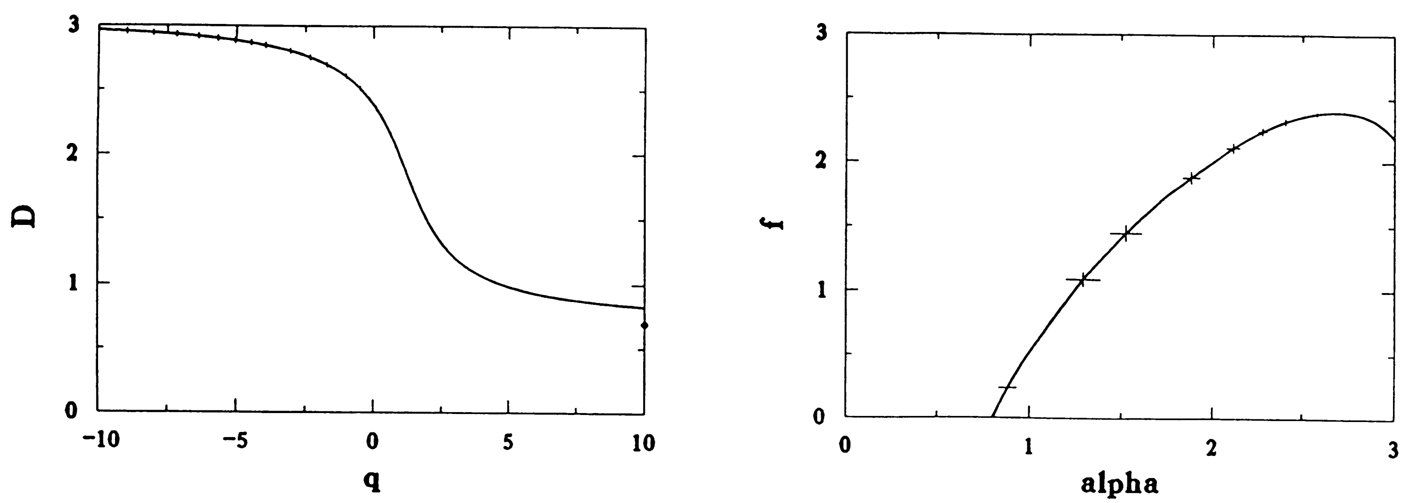

FIG. 2*b*

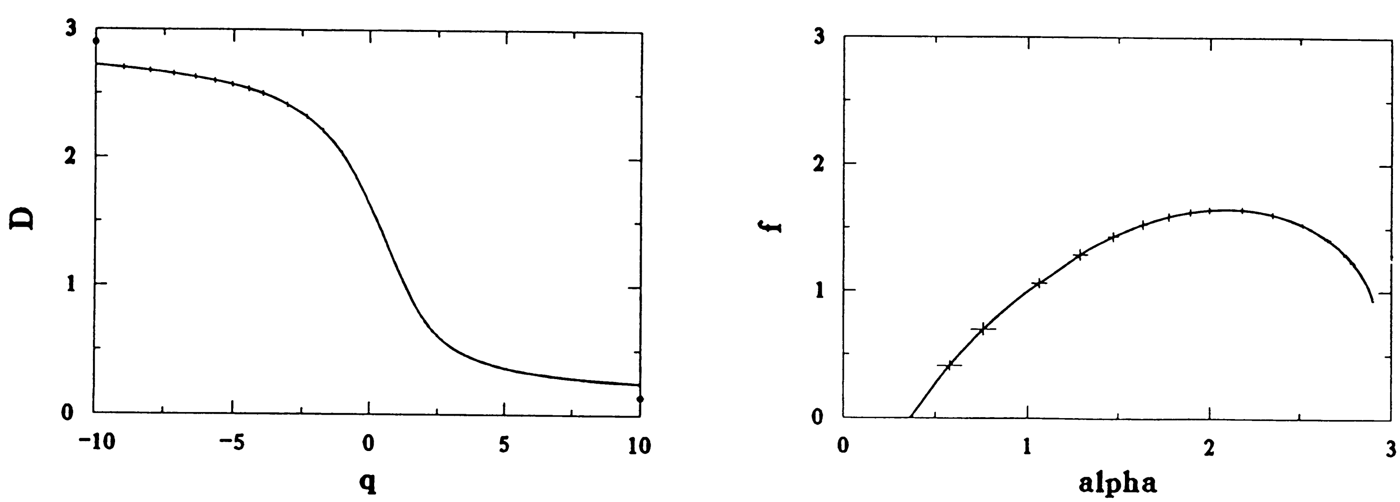

FIG. 2*c*

FIG. 2.—Generalized dimensions D_q and the $f(\alpha)$ spectrum for (*a*) observations, (*b*) the full axion model, and (*c*) the biased version. As we can see, the three cases are clearly not simple fractals, since they show a substantial spread in the scaling index α. These plots are for specific values of the cell size ϵ, but are nevertheless representative of the general shape of the $f(\alpha)$ spectrum.

We have made no attempt here to relate this measure of clustering to other measures such as the correlation functions, power spectra, or multiplicity functions, except to point out that the D_q curve is a compact and intuitively appealing way of presenting the moment generating function for the galaxy distribution. As a statistical descriptor of the clustering there can be little doubt that D_q is of considerable value. However, it would be inappropriate at this stage to suggest that there is any deep physics in the fact that the universe and models of the universe can be described in this way.

We would like to thank Møgens H. Jensen for stimulating discussions on the multifractal formalism. We enjoyed discussions with Benoit Mandelbrot, and we are grateful to John Huchra for permission to use the redshift catalog. This work was done while J. Einasto and E. Saar were guests of NORDITA, and they wish to express their gratitude for the Copenhagen hospitality.

REFERENCES

de Lapparent, V., Geller, M. J., and Huchra, J. P. 1986, *Ap. J.* (*Letters*), **302**, L1.
Efstathiou, G., Fall, S. M., and Hogan, C. 1979, *M.N.R.A.S.*, **189**, 203.
Einasto, J., Klypin, A. A., and Saar, E. 1986, *M.N.R.A.S.*, **219**, 457.
Gramann, M. 1987, *M.N.R.A.S.*, submitted.
Grassberger, P., and Procaccia, I. 1983, *Phys. Rev. Letters*, **50**, 346.
Hasley, Y. C., Jensen, M. H., Kadanoff, L. P., Procaccia, I., and Shraiman, B. I. 1986, *Phys. Rev. A*, **33**, 1141.
Hentschel, H. G. E., and Procaccia, I. 1983, *Physica*, **8D**, 435.
Jensen, M. H., Kadanoff, L. P., Libchaber, A., Procaccia, I., and Stavans, J. 1985, *Phys. Rev. Letters*, **55**, 2798.
Jones, B. J. T., and Martínez, V. J. 1988, in preparation.
Kirshner, R. P., Oemler, A., Schechter, P. L., and Shectman, S. A. 1981, *Ap. J.* (*Letters*), **248**, L57.
Mandelbrot, B. B. 1977, *Fractals* (San Francisco: Freeman).
———. 1982, *The Fractal Geometry of Nature* (San Francisco: Freeman).
———. 1986, *Physics Today*, **39**, 11.
———. 1988, preprint.
Peebles, P. J. E. 1980, *The Large Scale Structure of the Universe* (Princeton: Princeton University Press).
Soneira, R. M., and Peebles, P. J. E. 1978, *A.J.*, **83**, 845.
Tully, R. B., and Fisher, J. R. 1978, *IAU Symposium 79, The Large Scale Structure of the Universe*, ed. M. S. Longair and J. Einasto (Dordrecht: Reidel), p. 214.

Jaan Einasto and Enn Saar: Tartu Astrophysical Observatory, Toravere 202444, Estonia, USSR

Bernard J. T. Jones: NORDITA, Blegdamsvej 17, Copenhagen, Denmark DK2100

Vicent J. Martínez: Departament de Matemàtica Aplicada i Astronomia, Universitat de València, Spain