International Journal of Theoretical Physics, Vol. 28, No. 9, 1989

Statistics of the Galaxy Distribution

Marc Lachièze-Rey[1]

Received June 20, 1989

The universe appears from recent observational results to be a highly structured but also highly disordered medium. This accounts for the difficulties with a conventional statistical approach. Since the statistics of disordered media is an increasingly well-studied field in physics, it is tempting to try to adapt its methods for the study of the universe (the use of correlation functions also resulted from the adaptation of techniques from a very different field to astrophysics). This is already the case for the fractal analysis, which, mainly developed in microscopic statistics, is increasingly used in astrophysics. I suggest a new approach, also derived from the study of disordered media, both from the study of percolation clusters and from the dynamics of so-called "cluster aggregation" gelification models. This approach is briefly presented. Its main interest lies in two points. First, it suggests an analysis able to characterize features of unconventional statistics (those that seem to be present in the galaxy distribution and which conventional indicators are unable to take into account). It appears also *a priori* very convenient for a synthetic approach, since it can be related to the other indicators used up to now: the link with the void probability function is very straightforward. The connexion with fractals can be said to be contained in the method, since the objects defined during this analysis are themselves fractal: different kinds of fractal dimensions are very easy to extract from the analysis. The link with the percolation studies is also very natural since the method is adapted from the study of percolation clusters. It is also expected that the information concerning the topology is contained in this approach; this seems natural since the method is very sensitive to the topology of the distribution and posses some common characteristics with the topology analysis already developed by Gott *et al.* (1986). The quantitative relations remain however to be calculated. Additionally, this approach concerns the variation of clustering properties of galaxy groups and clusters with their richness. Although such studies have been made for various cases (like comparison of the correlation functions between galaxies and clusters, or between clusters of different richness classes), the analysis presented here deals with it in a more systematic and synthetic way.

1. INTRODUCTION

The exploration of the universe is a very active field of research, with the goal to understand the structure of our cosmos. Among cosmological

[1]Service d'Astrophysique, CEN-Saclay, 91191 Gif sur Yvette, France.

0020-7748/89/0900-1125$06.00/0

models, the class of the big-bang models has a large favor among astrophysicists and cosmologists, and they are used as a general frame for astrophysics. Since very few arguments have been advanced against them, it is very important to explore their weak points, to check if they resist or not a deep analysis, and if they can or cannot be refuted.

The study of the distribution of the galaxies, and more generally of the luminous matter in the universe, may provide crucial tests. It has been and still is a very active area of research, with the result that the distribution of the galaxies in the universe appears less simple than was suspected: at scales larger than that of clusters of galaxies, the matter seems to be organized in flattened, filamentary, lacunary structures. This picture, very different from a homogeneous distribution, has led cosmologists to ask if the universe is really homogeneous at a large scale, and if the cosmological principle (at the basis of the big-bang models) is justified. The answer is probably yes, but the big-bang models may suffer from difficulties, even without exploring these extremely large scales, since it appears very difficult to understand how galaxies and other structured systems did appear in a homogeneous universe. Many models of galaxy formation have been proposed, with various successes, but no one appears to be entirely satisfactory. There is an urgent need to check them, in order to finally know if galaxy formation is possible in the frame of the big-bang models. For this task, one possibility is to compare the predicted distribution of galaxies with observations. This will be done here.

In Section 2, I give a brief qualitative view of the matter distribution as it appears now. In Section 3, I present the basic statistical analysis used up to now. In Section 4, I show that high-order indicators allow one to get complementary information; and, in Section 5, I present some possible prospective work.

2. A QUALITATIVE VIEW OF THE GALAXY DISTRIBUTION

It was suspected for more than a century that the universe is made of separate structures, the galaxies. They are the basic objects of study for cosmology, but it is only from the beginning of the century that we have known that galaxies themselves are not randomly distributed in the universe, but are organized into larger structures.

This organization appears in two-dimentional views of the sky, in photographic plates or celestial maps: the angular distribution of galaxies designates groups, clusters, filaments, and even larger structures which seem to occupy a large fraction of the sky. These impressions are confirmed by the true spatial analyses, when the third dimension, i.e., the distance from the galaxies to us, becomes available.

Cosmologists try to recognize the exact nature of these structures, up to the largest scales, as well to understand how they were formed. Galaxies have typical sizes of a few tens of kpc (1 kpc = 1000 pc; 1 pc = 1 parsec = 3×10^8 cm); their average separation is of the order 1 Mpc ($= 10^6$ pc). Unfortunately, if their positions over the sky (i.e., their celestial coordinates) are known without ambiguity, this is not the case for their distances to us, and this imperfect knowledge strongly limits our ability to explore the structure of the universe. Most of the galaxies have their distance estimated only through their redshift z. In the conventional cosmological interpretation, z is related to the distance via the Hubble law: $cz = H_0 D$ (for the galaxies not too far away from us), where c the velocity of light, and H_0 is defined as the Hubble constant. This law expresses the expansion of the universe; it allows us to estimate the distance D of a galaxy after having measured its redshift. Unfortunately, the Hubble constant is not known with a good precision and there is a strong controversy—the distant scale problem—concerning its value, between limits of 50 and 100 km sec^{-1} Mpc^{-1}. In the following, *I* will assume for convenience a value of 100 km sec^{-1} Mpc^{-1} for expressing the cosmic distances, but it must be kept in mind that all distance determinations from redshifts suffer from the relative uncertainty on H_0.

Although the Hubble law is probably well verified statistically, it leads to large uncertainties when applied to individual objects. The reason is that galaxies have proper velocities in addition to their Hubble (expansion) velocities, due, for instance, to the gravitational attraction of nearby mass concentrations, like clusters or superclusters. Since the (unknown) proper velocity contributes to the observed redshift, a wrong estimation of the distance is obtained. Although I will not discuss further this problem (which, for the local galaxies, is known as the problem of Virgo infall corrections), it should be kept in mind that the distances of galaxies, when obtained from redshifts, are largely uncertain. There are fortunately other methods for determining the distances. They are, however, very time consuming, and only a few galaxies have their distances measured independently of redshifts. Although the situation is quickly improving, the majority of statistical studies up to now involve distances estimated from the redshifts.

Mapping the three-dimensional sky has been a long and difficult task; we have begun to have a knowledge of the spatial distribution of galaxies, at least not too far from us. Locally, our own galaxy (the Milky Way) is a member of the so-called Local Group. This local concentration gathers about 20 galaxies in mutual gravitational interaction, so that it is not a pure geographical coincidence, but a real physical association. Among other members, the best known is the Andromeda galaxy, also called Messier 31. The size of the Local Group is about 1 Mpc.

Most of the galaxies lie much farther than the Local Group, but they are also members of groups, or clusters. The nearest cluster is the Virgo cluster, at about 10 Mpc from us. It is much larger than the Local Group, but still poorer and smaller than a typical cluster of galaxies. It is dominated by a huge galaxy in its core, called Messier 87. Much further is the Coma cluster, at about 100 Mpc from us. Larger and richer than Virgo, it gathers thousands of galaxies. More than 2000 other clusters have been found over the sky, with a large diversity in sizes, richnesses, and characteristics. Their typical size is about 5 Mpc. But clusters of galaxies are not the ultimate structures in the universe. Around us, the distribution of matter in the nearest tens of Mpc makes up what has been called the Local Supercluster, or the Virgo Supercluster. This very large structure was first recognized by Gérard de Vaucouleurs around 1958. It contains many clusters, groups, and isolated galaxies, and appears as a flattened structure, shaped like a disk, about 1 Mpc wide and 20 Mpc long.

Superclusters seem to be characteristic of the organization of matter in the universe. Their existence was proven beyond any doubt by the systematic measurements of redshifts in the last decades. All appear elongated, or flattened, or both, with characteristic lengths between 20 and 100 Mpc. Complementary, large voids also exist between the galaxies. A void is defined as a region of the universe inside which no galaxy is present. Recent observations have proved that voids are common between the galaxies, at all scales. Clusters and superclusters may be analyzed as positive density fluctuations (see below), voids as negative ones. There is, however, no consensus on whether the voids are devoid of any material content, or only of galaxies and luminous matter.

We do not have a clear vision of the distribution of the galaxies at scales larger than those of the superclusters, and the organization of the cosmic matter at scales beyond 25 Mpc remains uncertain. We know that, at these large scales, the matter is organized into flattened, elongated structures, with numerous voids between them. Galaxies seem to occupy a kind of cellular network between the voids, as suggested, for instance, by Einasto and co-workers. Although a large number of astronomers would probably agree with the term, it remains to make quantitatively precise what could be the properties of such a structure. Among other answers, future observations will probably tell us if superclusters are mutually connected, and if they are organized into larger systems. In any case, our explorations have not reached the scale at which the universe can be considered as homogeneous. Other observations suggest, however, that the universe can be seen as homogeneous beyond a scale of about 100 Mpc, justifying therefore the cosmological principle.

One of the goals of the study of the galaxy distribution is to put constraints on the hypothetical processes by which these structures could have formed, i.e., the models for galaxy formation. For this approach, it is convenient to distinguish two ranges of scales. The so-called linear range concerns the largest scales, beyond a limit which, although not precisely known, probably lies below 10 Mpc. The structures at these large scales (for instance, the shapes and distribution of superclusters, the presence of large voids, the cellular network) are thought not to have been subject to a strong dynamical evolution. Gravitational interactions have remained weak enough so that a linear treatment can be applied. The result is that, as astronomers believe, the distribution of matter at these linear scales results from a simple evolution of the initial conditions. This is not the case for the nonlinear range, at scales below about 10 Mpc. The structures are expected at these scales to result from a strong evolution which a conventional dynamical analysis cannot follow.

Unfortunately, observations concerning the linear scales are incomplete and inhomogeneous (although the situation is improving rapidly). It is therefore difficult to quantify the information concerning these large scales. This task—necessary to confront observations with the predictions of the models—has been initiated by various groups thanks to the introduction of new statistical tools, such as percolation, topological analysis, areas of isodensity contours, etc. I will, however, concentrate on the nonlinear range, where conventional statistical analyses have been applied successfully, with the use of well-suited linear indicators.

3. CORRELATIONS

In order to compare observations with the predications of the models, suitable tools have been defined. A first piece of information comes from the analysis of the density of galaxies in space, $n(x)$. A density fluctuation is defined as $\delta n = n - \langle n \rangle$, where $\langle n \rangle$ is the average density of galaxies in space. We can also define a relative density fluctuation $\delta(x) = \delta n / n$, varying from place to place. An analysis in terms of δ makes sense in the range where δ is not too far from 1. Nevertheless, it is very important to realize that this density n fluctuates today, at all the scales where the universe has been explored; thus, for any three-dimensional catalog (up to size larger than, say, 50 Mpc), the measured value of n cannot be considered as the average value for the universe. In other words, these catalogs are not fair samples of the universe.

Among the different possible indicators possible, the two-point correlation function (hereafter CF) has been the most popular. Its introduction results from the adaptation of the BBGKY formalism to galaxy statistics (see, for instance, Peebles, 1980). It can be defined in a probabilistic way: for a distribution of galaxies of average density n, the probability for finding one galaxy in the elementary volume dV around x is $P = n(x)\, dV$. For a homogeneous distribution (stationary in the language of statistics), n would be constant in space. The probability of finding simultaneously a galaxy in dV_1 around x_1 and a second one in dV_2 around x_2 is written as $P_{12} = n^2[1 + \xi(r_{12})]$, which defines the two-point correlation function ξ. Homogeneity and isotropy imply that ξ is only a function of $r = r_{12} = r_1 - r_2$. It is convenient to remark that this definition is in fact a conditional one: P_{12} is the probability to find two galaxies, given that the probability to find one is P. This is the reason why the calculation of ξ is subject to normalization problems, as I indicate below.

In the case where galaxies are randomly distributed, the presence of a galaxy in x_1 does not depend on the presence of another one in x_2. The joined probability is then the product of the two elementary probabilities, and ξ has a zero value. This indicates how ξ measures the excess probability of clustering, with respect to a random distribution taken as a reference. Practically, this probabilistic definition has been widely used since the pioneering work of Peebles. An intuitive insight might also be gained from the fact that $1 + \xi$ is a measure of the average number of neighbors at a distance r from a galaxy, in excess of a pure random distribution. This helps us to understand in which sense ξ measures the clustering properties of galaxies. It can also be remarked that ξ is the autocorrelation function for the density of galaxies: $\langle n(x+r)n(x)\rangle = \langle n(x)\rangle^2[1 + \xi(r)]$, where the averages are taken over space.

The two-point correlation functions have been measured from different positions in the universe. First measurement came from two-dimensional catalogs only, giving the angular positions of galaxies. For their analysis, a two-dimensional correlation function $w(\theta)$ is defined between galaxies separated in the sky by an angle θ. Its definition is analogous to that of the three-dimensional correlation function, except that the three-dimensional distance r is replaced by the angular distance θ: w measures the probability of finding two galaxies in the sky separated by θ. The knowledge of $w(\theta)$ allows one to estimate $\xi(r)$ by a procedure of deprojection, provided that we know the average depth (i.e., the average distance of the galaxies to us) of the catalog. When three-dimensional catalogs became available, ξ was estimated directly.

All reported results agree now that, in first approximation, the correlation function of galaxies may be fitted by a universal power law $\xi = (r/r_0)^{-\gamma}$, with $\gamma = 1.8$ and r_0, the correlation radius, around 5 Mpc. There is, however, no agreement concerning the value of ξ beyond 10 Mpc, so that these estimations concern the nonlinear range only. Unfortunately, it is not possible to calculate analytically the dynamical development of ξ during the nonlinear phase, so that the models of galaxy formation cannot be checked directly with this function. The development of ξ can be followed, however, during the nonlinear regime with numerical simulations. As a matter of fact, the matching with the observed value (level and shape) is used as a kind of normalization for these simulations: astrophysicists have chosen a criterion using this function to decide when a simulation must be stopped, i.e., to choose which instant in the time coordinate of the simulation is to represent the present state of the universe. It is important to realize that this requirement—that the simulation provides both the shape and level of ξ—has already led to the rejection of some models for galaxy formation, namely the hot dark matter models, and has required the introduction of the concept of biased galaxy formation. It is clear, however, that this test is very crude, and not sufficient to definitely qualify a model. To improve the situation, other statistical tools have to be developed.

3.1. Normalization

Some difficulties have additionally been pointed out with the use of correlation functions. The reason is that the probabilistic definition requires the knowledge of the density. Measurements of ξ are performed by counting pairs of galaxies in a given catalog, and then dividing this measured quantity by a corresponding number for an uniform Poisson distribution (this number is introduced as the best way to estimate the volume sampled). Without getting into details, this requires assigning some value n to the spatial point density of the catalog under study (i.e., the number of galaxies per unit volume). This can be seen as a normalization for calculating ξ and it is easy to show that any change in the normalization also leads to changes of the estimated value of ξ (Lachièze-Rey, 1989).

Important density fluctuations (for instance, large voids or superclusters) are present in all the catalogs of galaxies, up to the sizes of these catalogs themselves, even for the largest available. We have therefore no absolute estimation of the density of galaxies (or of their luminosity function) in the universe and this implies that no absolute value of ξ can be measured: any measured value of ξ is relative to the given sample. The largest three-dimensional sample available now is the CfA catalog, so its measured value of ξ is the nearest to a universal one. In such a situation, the recent publication of a three-dimensional catalog of southern galaxies—the SSRS catalog (da Costa *et al.*, 1988)—appears as a major step, since it provides another sample, statistically independent from the CfA. Recently, Davis *et al.* (1988) have shown that the northern and southern galaxy distributions have the same value of ξ. This seems to indicate that this value may be in fact universal.

3.2. Fractals

On the other hand, it has been reported by Einasto *et al.* (1986) that, at scales smaller than 100 Mpc, the level of ξ apparently depends on the size of the sample in which it is estimated. This tendency has been interpreted in different manners. Maurogordato and Lachièze-Rey (1987; hereafter ML1) suggested that it could be due to an effect of the luminosity segregation, but they later showed that this is not the case (Maurogordato and Lachièze-Rey, 1989; hereafter ML2). On the other hand, Davis *et al.* (1988) and Lachièze-Rey *et al.* (1989) have shown a similar effect in the southern hemisphere. They all conclude that this is not a luminosity effect.

Calzetti *et al.* (1988) have suggested that this could be the sign of a fractal distribution of the galaxies. This is probably an overinterpretation of the data. Although some fractal distribution would lead to the reported effect, the opposite is not necessarily true. More simply, the effect may be due to the strong variations of the density of galaxies at the relevant scales, ξ being very sensitive to normalization effects, as reported above. Davis *et al.* (1989), for instance, explain it by the density gradients inside the catalogs. In passing, it should be remarked that, even if a fractal description is correct, it would not be valid beyond the range where the universe may be seen as homogeneous, probably no more than 100 Mpc.

In order to show the previous effect, and, more generally, to compare the statistical properties of different samples, it is necessary to take into account the strong density variations between these samples. The conventional prescriptions for the calculation of ξ may not be most convenient. This is why ML1 and ML2 proposed to use an intrinsic normalization for the estimations of ξ. This prescribes using for normalization, not the universal value of the density (which, being unknown, has to be guessed), but the one derived from the catalog itself, i.e., the number of galaxies divided by the total volume. Although not free from effects due to the gradients of the density inside the sample, this indicator seems to express the desired properties; it has been used, for instance, to exhibit the effects reported above. It has also been used to test for possible segregations in the galaxy distribution (see below). Alimi and Blanchard (1988) and Pietronero (1987) also proposed normalization-free indicators in the same spirit. As a conclusion, I will, however, point out that great care must be taken in drawing conclusions from the use of the correlation functions, because of the strong density fluctuations always present in the galaxy distribution.

3.3. Segregations

Normalization problems occur not only for absolute estimations, but every time the statistical properties of different samples of galaxies are to be compared. For instance, some models predict that samples with different morphologies (spiral versus elliptical galaxies), luminosities (bright versus faint), or surface brightness characteristics, with different sizes, localizations in space, etc., may have different statistics. Usually, such samples also have different densities, so that strong normalization problems are present. ML1 and ML2, for instance, have considered different (complete) subsamples of the CfA, with different shapes, sizes, and locations. Although the "conventional" normalization would have required using a universal density and a universal luminosity function, such quantities are not available. the "intrinsic" normalization was therefore used, which requires no hypothesis about a universal density or luminosity function of galaxies, but can be entirely performed from the available data (see also Alimi *et al.*, 1988).

3.4. High-Order Indicators

It is clear that the two-point correlation functions are not able to express the whole informational content of the galaxy distribution. ξ is in fact only the lowest order indicator in the hierarchy of correlation functions, and n-point correlation functions can be defined in a similar way. For instance, the three-point correlation function is defined from the excess probability that three galaxies lie at three positions in space. This probability is written $P(1,2,3)=n^3(1+\xi_{12}+\xi_{23}+\xi_{13}+\zeta_{123})$, which defines the (reduced) three-point function ζ_{123}. Although the determination is quite noisy, Groth and Peebles (1977) showed that this function can be fitted by an expression of the particular form $\zeta_{123}=Q(\xi_{12}\xi_{23}+\xi_{12}\xi_{13}+\xi_{13}\xi_{23})$. The determination of four-point correlation functions is hardly possible, but it has been shown that a similar (generalized) form is compatible with the observations (Fry and Peebles, 1978). Measurements involving more than four points appear impossible for the moment. It is interesting to note that the form present above for the three-point function may be generalized to higher orders, implying that the reduced N-point function $\xi^{(N)}$ is a sum of products of the form $(\xi^{(2)})^{N-1}$. This is the basic assumption of the so-called hierarchical models (Fry, 1986; Schaeffer, 1985). This implies that the n-point correlation function obeys a scaling relation:

$$\xi^{(N)}(\lambda r_1, \lambda r_2, \ldots, \lambda r_N) = \lambda^{\alpha} \xi^{(N)}(r_1, r_2, \ldots, r_N)$$

We have also shown that the void probability function (hereafter VPF) is an efficient statistical indicator, complementary to the CF, and giving information about the large-order correlations (two-point correlations are of order two). We will also see that it does not suffer normalization problems.

3.5. The Void Probability Function

The VPF $P_0(V)$ is defined as the probability that a sphere of volume V ($V=4\pi R^3/3$) at a random place in the catalog contains no galaxy at all. White (1979) has shown that it is linked to the hierarchy of correlation functions of all orders (in this hierarchy, the point density and the two point CF are, respectively, first- and second-order quantities). It appears, in this sense, complementary to the CF. Concerning normalization properties, the VPF does not seem at first sight very attractive since it depends on both the volume and the density, although the two-point CF, conversely, depends only on the volume, not on the density (at least in principle). But the situation is in fact much more favorable if we do not use $P_0(V)$ as a function of the volume, but rather the quantity $\chi(V) = \log P_0(V)/nV$.

It has been established that different kinds of assumptions lead to what has been called a scaling relation for the VPF. This latter is defined by the prescription that the quantity χ, which *a priori* depends on the two variables n (the density) and V, reduces to a function of one variable only, which is called the scaling variable. Schaeffer (1987) has shown that the hypothesis of the hierarchical models implies this relation. Schaeffer and Balian (1988) also showed that it is implied by the assumption that N-point correlation functions are themselves scale invariant, in the sense defined above. In these two cases, the scaling variable q takes the form $q = nV\langle\xi\rangle$, where $\langle\xi\rangle$ is defined as the integral of $\xi(x_1, x_2)$ over the two spatial variables x_1 and x_2, in two volumes V_1 and V_2 equal to V. It is interesting to note that these assumptions allowed Balian and Schaeffer (1988) to make many predictions concerning the distribution of the galaxies.

It appears therefore very important to check if this property is verified in the true galaxy distribution. Bouchet and Lachièze-Rey (1986), ML1, and ML2 have shown, from various samples of galaxies, that the galaxy distribution obeys this scaling invariance. (ML2) have also shown that, if $\chi(q)$ shares with the CF the property of being density independent, it has the advantage of being almost normalization independent. This allows us to compare the high-order properties of different samples, without normalization problems. It is, for instance, a striking fact that the apparent existence of a segregation in clustering properties, as measured by the correlation functions, does not imply a corresponding segregation as measured by the VPF: we have shown that samples containing different galaxy morphological types, or of different sizes, have different values of two-point CF (although the sign and the level of the differences depend on the adopted normalization), but that there are no differences in the VPF. Conversely, the VPF makes a luminosity segregation apparent, although the two-point correlation function did not. This clearly indicates (if necessary) that the CF and VPF do not measure the same kind of clustering properties.

Many reasons make the VPF an efficient tool to determine the statistical distribution of galaxies. An increasing number of papers are devoted to its calculation. Its main interest has probably been to allow the property of scale invariance to be established. This made it possible to compare sample with different densities, with different luminosity functions, at various places, of different sizes, etc., without normalization problems. Moreover, since the VPF has been predicted by dynamical or statistical models, there is a reference value to compare with observations. By the way, I suggest to those involved in numerical simulations to use it as a test, as it has been done with the CF. Finally, in contrast with the two- (or three-) point CF, the VPF gives information at the highest orders, and it becomes increasingly clear that the most specific features of the galaxy distribution can be expressed through such high-order indicators only. The present data are unfortunately too scarce for the VPF to be applied to the study of the large, linear strucures, but this indicator will probably appear very useful in the near future. In addition, it may be that voids, filaments, sheets, etc., are also present at the nonlinear scales. This would explain the apparent organization of the galaxies in the catalogs in the form of a network of bubbles, or cells, or of a spongelike structure, etc. If this impression is true, it will soon appear necessary to give a quantitative account of this network, by measuring the dimensionality, the connectivity properties, the sharpness of the edges, etc.

On the other hand, it is presently very difficult to predict statistical properties from dynamical studies since, at nonlinear scales, we are not able to follow the dynamics. This would certainly require the use of different kinds of statistics.

Different suggestions have been made recently to these ends, with more or less success. The VPF is only a first step, including information at all orders. It is encouraging that scale invariance has been shown from its use. Among other methods, we could retain percolation (Shandarin and Zeldovich, 1986), analysis in terms of fractals or multifractals (Jones *et al.*, 1987; Balian and Schaeffer, 1988), topology analysis (Gott *et al.*, 1986; Melott *et al.*, 1988), tesselation (Icke and van de Weygaert, 1987), and probably many others of which the author is unaware. Unfortunately, these different indicators give different kinds of information, not easy to compare among themselves or with the predictions of dynamical models. There is no general agreement concerning the optimal analysis, which ideally should be related to some dynamical processes. I will give the broad lines of a possible approach derived from the so-called coagulation formalism.

4. STATISTICS OF DISORDERED MEDIA

The study of the galaxy distribution in recent years has led to a picture very different to what was proposed before: the distribution of galaxies appears highly irregular and disordered, with the presence of peculiar features like filaments or sheets around voids, i.e., large regions devoid of luminous matter. The visual appearance of the catalogs is the first clue that this is so, and one only needs to look at the first slices of the new CfA catalog (De Lapparent *et al.*, 1986) or the SSRS catalog (da Costa *et al.*, 1988) to be convinced. More quantitatively, this impression is confirmed by many facts: the presence of density fluctuations at all scales, up to those of the catalog themselves, and maybe beyond; the reported tendency of the two-point correlation function to vary with the depth of the catalog; the presence of voids at all scales, and especially large ones; the presence of filaments; etc.

The two-point CF are not sufficient to give an account of this disordered-structure, and the VPF approach is only a first step. It is presently not clear how efficient are the different indicators—percolation, topology, multifractals, etc.—nor how they compare.

The disordered aspect of the universe is the cause of the difficulties encountered with the conventional statistical approach. This suggests turning toward the statistics of disordered media, which is an increasingly well-studied field in physics, and trying to adapt its methods for the study of the universe (remember that the use of CF also resulted from the adaptation of techniques from a very different field than astrophysics). This is already the case for fractal or multifractal analysis, which, although mainly developed in microscopic statistics, is increasingly used in astrophysics.

Another approach, also derived from the study of disordered media, could be of interest: it is derived both from the study of percolation clusters and from the dynamics of the so-called cluster "aggregation," "coagulation," and "gelification" models.

4.1. The Coagulation Method

I will not present the details of the method, but only the general principle. Starting from the distribution of points to be studied, we create a geometrical object—hereafter the "aggregate"—by the following rule: given a length L (the "scale variable"), we "fill up" the space around each point (galaxy) of the distribution, up to distance L. In other words, we consider a distribution of spheres of radius L, each having as center one galaxy of the distribution. Some spheres are isolated and some merge to form what I will call a cluster. The geometrical object of interest—the aggregate—is the ensemble of all these clusters and of the isolated spheres (which can be considered as clusters of multiplicity 1). Each cluster is characterized by the volume V that it occupies in space. Of course, the volume of a cluster of multiplicity 1 is $4\pi L^3/3$, but the volume of a cluster of multiplicity k is between this value and k times this value. The interesting quantity is the distribution of clusters, as a function of the volumes V, for a given value of the scale variable L: $N(L, V)$. The "coagulation" method is based on the use of $N(L, V)$ as a statistical indicator. The "aggregate" so defined is an interesting geometrical object. It reflects in some sense the smoothed distribution, with a smoothing length L.

The calculation of N is without any special problem and we have already developed a code for such a purpose. This indicator characterizes some features of the distribution that the conventional ones are unable to take into account. Additionally, it may easily be related to other indicators, which is very convenient for a synthetic approach. The link with the VPF is, for instance, very straightforward and has already led to a very efficient method of evaluating the VPF in the galaxy distribution.

The connection with fractals is contained in the method, since the different objects defined during this analysis—the aggregate, the clusters—are themselves fractal. There is also a link with the percolation, since the "aggregate" is of the same nature as a percolation cluster. This method is, however, richer, since it considers information not only about the percolation cluster (i.e., the largest cluster in the sample), but also about clusters of all sizes (Stanley 1986). However, the tools used for percolation studies—i.e., the whole series of different fractal dimensions (Stanley 1986)—are also useful in this case. Also, the method presents some analogies with the "topological" analysis of Gott *et al.* (1986). It is also interesting to remark that this indicator is able to express the dimensionality (or dimensionalities) of the distribution, so that it might be used to check the presence of sheets, filaments, etc. This approach seems therefore a very efficient way to express results obtained with one method in the language adapted to another and, in our sense, in a more synthetic way.

Many benefits of this method come from the work already done with this formalism in a very different context. A similar quantity has been introduced for the study of the cluster aggregation models (Kolb *et al.*, 1986), where the dynamical growth of the clusters is considered. There is substantial hope that this may be applied to the dynamics of galaxy formation and this could make it possible to follow some statistical characteristics of the matter distribution.

REFERENCES

Alimi, J. M., and Blanchard, A. (1988). Preprint.
Alimi, J. M., Valls-Gabaud, D., and Blanchard, A. (1988). *Astronomy and Astrophysics*, **302**, L37.
Balian, R., and Schaeffer, R. (1988). *Astrophysical Journal*, **335**, L43.
Blanchard, A., and Alimi, J. M. (1988). *Astronomy and Astrophysics*, **203**, L1.
Bouchet, F. R. (1988). In *Proceedings of the 8th Moriond Astrophysics Meeting "Dark Matter"*, J. Audouze and J. Tran van Tranh, eds., Frontieres, Gif-sur-Yvette, France.

Bouchet, F. R., and Lachièze-Rey, M. (1986). *Astrophysical Journal Letters*, **302**, L37.
Calzetti, D., Giavalisco, M., and Ruffini, R., (1988). *Astronomy and Astrophysics*, **198**, 1.
Da Costa, L. N., Pellegrini, P., Sargent, W., Tonry, J., Davis, M., Meixksin, A., Latham, D., Menzies, J., and Coulson, I. (1988). *Astrophysical Journal*, **327**, 544.
Davis, M., and Efstathiou, G. (1987). In *Proceedings of the Vatican Study Week on* Large Scale Motions in the Universe.
Davis, M., and Peebles, P. J. E. (1983). *Astrophysical Journal* **267**, 465.
Davis, M., Meiksin A., Srauss M. A., da Costa L. N., and Yahil, A. (1988). *Astrophysical Journal*, **333**, L9.
De Lapparent, V., Geller, M. J., and Huchra, J. (1986). *Astrophysical Journal Letters*, **302**, L1.
Einasto, J., Klypin, A. A., and Saar, E. (1986). *Monthly Notices of the Royal Astronomical Society*, **219**, 457.
Fry, J. N. (1986). *Astrophysical Journal*, **306**, 358.
Fry, J. N., and Peebles, P. J. E. (1978). *Astrophysical Journal*, **221**, 19.
Gott, J. R., Melott, A. L., and Dickinson, M. (1986). *Astrophysical Journal*, **306**, 341.
Groth, E. J., and Peebles, P. J. E. (1977). *Astrophysical Journal*, **217**, 385.
Hamilton, A. J. S., Saslaw, W. C., and Thuan, T. X. (1985). *Astrophysical Journal*, **297**, 37.
Huchra, J., Davis, M., Latham, D., and Tonry, J. (1983). *Astrophysical Journal* (*Supplement*), **52**, 89.
Icke, V., and van de Weygaert, R. (1987). *Astronomy and Astrophysics*, **184**, 16.
Jones, B. J. T., Martinez, V. J., Saar E., and Einasto, J. (1988). *Astrophysical Journal Letters*, **332**, L1.
Kolb, M., Botet, R., Jullien, R., and Herrmann, J. (1986). In *Growth and Form, Fractal and Nonfractal Patterns in Physics*, H. E. Stanley and N. Ostrowsky, eds., p. 222, Martinus Nijhoff, Dordrecht.
Lachièze-Rey, M. (1989). In *Proceedings of the 8th Moriond Astrophysics Meeting "Dark Matter"*, J. Audouze and J. Tran Van Tranh, eds., p. 187, Frontieres, Gif-sur-Yvette, France.
Maurogordato, S., and Lachièze-Rey, M. (1987). *Astrophysical Journal*, **320**, 13.
Maurogordato, S., and Lachièze-Rey, M. (1989). In preparation.
Melott, J. A., Weinberg, D. H., and Gott, J. R. III (1988). *Astrophysical Journal*, **328**, 50.
Peebles, P. J. E. (1980). *The Large Scale Structure of the Universe*, Princeton University Press, Princeton, New Jersey.
Pietronero, L. (1987). *Physica* **144A**, 257.
Schaeffer, R. (1984). *Astronomy and Astrophysics Letters*, **134**, L15.
Schaeffer, R. (1985). *Astronomy and Astrophysics Letters*, **144**, L1.
Schaeffer, R. (1987). *Astronomy and Astrophysics Letters*, **181**, L23.
Schaefer, R., and Balian, R. (1988), preprint.
Shandarin, S. F., and Zeldovich, Ya. B. (1986). In *Cosmology, Astrophysics, and Fundamental Physics*, G. Setti and van Hove, eds.
Slezak, E., Mars, G., Bijaoui, A., Balkowski, C., and Fontanelli, P. (to appear). *Astronomy and Astrophysics* (Supplement Series), to appear.
Stanley, H. E. (1986). In *Growth and Form, Fractal and Non Fractal Patterns in Physics*, H. E. Stanley and N. Ostrowsky, eds., p. 21, Martinus Nijhoff, Dordrecht.
Tyson, J. A. (1988). *Astronomical Journal*, **96**, 1.
White, S. D. M. (1979). *Monthly Notices of the Royal Astromical Society*, **186**, 145.