THE ASTROPHYSICAL JOURNAL, 409:517–529, 1993 June 1

OBJECTIVE DETECTION OF VOIDS AND HIGH-DENSITY STRUCTURES IN THE FIRST CfA REDSHIFT SURVEY SLICE

E. SLEZAK,[1] V. DE LAPPARENT,[2] AND A. BIJAOUI[1]
Received 1992 August 3; accepted 1992 November 3

ABSTRACT

Voids of galaxies delineated by sharp one- or two-dimensional structures are the main features of the large-scale clustering of galaxies. In order to objectively identify the significant structures in the first slice of the Center for Astrophysics redshift survey, we perform a space-scale analysis of this galaxy catalog by means of the wavelet transform. The wavelet coefficients allow us to detect on all scales within reach both over- and underdense structures. The confidence level of the detection is computed under the assumption that galaxies are distributed according to a Poissonian process with a nonuniform law. Significant voids and groups of galaxies are identified at different scales, and small-scale clustering is taken into account to further assess the reliability of the detection. The morphological parameters of the structures are estimated using an elliptical Gaussian model of the local galaxy distribution. The hierarchical clustering of the high-density regions is clearly exhibited. The structures identified on small scale are in good agreement with the published group catalog of the slice obtained by dynamical criteria. Significant structures are also detected at the location of the Abell clusters. Because the analysis is isotropic, the large-scale structures are traced by the alignment of contiguous smaller structures. Voids are identified at a 99.5% confidence level and a selection criterion is applied to take into account the multiple detections inherent to the analysis. We obtain a catalog of the astrophysical voids in the slice.

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

1. INTRODUCTION

The complexity of the distribution of galaxies and of clusters of galaxies is now clearly established up to scales of 50 h^{-1} Mpc (see Oort 1983, Bahcall 1988, and Rood 1988 for reviews). Galaxy correlations on large scales are detected using two-dimensional catalogs from objective Schmidt plate processing which cover a significant fraction of the sky, like the APM catalog (Maddox, Efstathiou, & Loveday 1988). Valuable information on the three-dimensional clustering of galaxies on scales larger than 10 h^{-1} Mpc is provided by wide-angle redshift surveys, such as the Center for Astrophysics (hereafter CfA) redshift survey slices based on Zwicky's (1961–1968) and Nilson's (1973) magnitude-limited catalogs of galaxies (Huchra et al. 1983; de Lapparent, Geller, & Huchra 1986; Geller & Huchra 1989; Huchra et al. 1990), or the Southern Sky Redshift Survey based on the ESO (Lauberts 1982) diameter-limited catalog (da Costa et al. 1988; da Costa et al. 1991).

The main feature of the galaxy distribution is the departure from homogeneity at all scales within reach. The topology of the distribution is characterized by a complex network of sharp structures, one-dimensional filaments (Giovanelli et al. 1986) or two-dimensional sheets (de Lapparent et al. 1986), suggesting a cell-like geometry (Jôeveer & Einasto 1978). Statistical analysis of the CfA slices shows that these structures occupy only ~25% of the available volume (de Lapparent, Geller, & Huchra 1991). The high-density structures appear to connect clusters of galaxies and delineate large spherical regions which are devoid of bright galaxies (de Lapparent et al. 1986; Pellegrini, da Costa, & de Carvalho 1989). Voids, like that in Bootes (Kirshner et al. 1981), are thus frequent events of the distribution. The largest coherent structure detected so far in the galaxy distribution is the so-called Great Wall with a spatial extent of about 60 h^{-1} Mpc × 170 h^{-1} Mpc and ~5 h^{-1} Mpc thick (Geller & Huchra 1989): however, this structure appears to be made of portions of several surfaces which surround adjacent voids and are therefore geometrically connected. Hence the voids can be considered as the fundamental physical units of the large-scale galaxy distribution. The regular patterns revealed by pencil-beam redshift surveys (Broadhurst et al. 1990) are also consistent with the presence of voids alternating with sheets of galaxies out to $z = 0.5$ (see de Lapparent et al. 1991). Measures of the velocity field of the galaxy distribution provide dynamical evidence for the large-scale structure. Analysis of an *IRAS* sample of galaxies confirms the dipole anisotropy caused by the motion of the Local Group relative to the microwave background radiation (Yahil, Walker, & Rowan-Robinson 1986; Yahil 1988). On larger scale, peculiar motion of galaxies suggest a coherent infall into a "Great Attractor" (Dressler 1988) located behind the Hydra-Centaurus supercluster (Lynden-Bell et al. 1988; Faber & Burnstein 1989). A coherent motion on such large scales is nevertheless difficult to explain within the standard models for the formation of large-scale structure (Kaiser & Lahav 1989). The frequent occurrence of large voids in the galaxy distribution also challenges the theoretical models. The diameters of the largest detected voids (of order of 50 h^{-1} Mpc) put tight constraints both on the nature of the dark matter and on the spectrum of initial fluctuations (Melott 1987; Kofman 1989). These implications have thus stimulated systematic searches for low-density regions in galaxy and cluster catalogs (Burns et al. 1988; Willick, Brodie, & Bowyer 1990; Kauffmann & Fairall 1991). However, a fair sample of the galaxy distribution is not yet available because the size of the largest detected voids is close to the maximum size allowed by the corresponding surveys (de Lapparent, Geller, & Huchra 1988).

[1] Observatoire de la Côte d'Azur, BP 229, F-06304 Nice Cedex 4, France.
[2] Institut d'Astrophysique de Paris, 98 bis, Boulevard Arago, F-75014 Paris, France.

The main theoretical framework for describing the formation of large-scale structure is currently the standard cold dark matter model (hereafter CDM; see Blumenthal et al. 1984; Davis et al. 1985; White et al. 1987, and references therein). This model assumes an inflationary flat universe dominated by hypothetical weakly interactive massive particles, with $\Lambda = 0$ and $h = 0.5$, and a gravitational growth of initial Gaussian, scale-invariant, adiabatic fluctuations of density, and a galaxy formation biased toward high-density regions. This cosmological scenario is tightly constrained by the extreme homogeneity of the microwave background radiation (Hogan 1990): on the one hand, the large-scale perturbations must be sufficiently contrasted to enter the nonlinear regime with a Hubble time, but on the other hand they must remain small enough in amplitude for not introducing into the microwave background radiation fluctuations larger than the current observational limits (Smoot et al. 1991, Smoot et al. 1992). Consequently the growing evidence of structures and coherent motions on very large-scale is a serious challenge to the CDM model. The detected large-scale flows (Burstein, Faber, & Dressler 1990; Dressler & Faber 1990) and the positive angular galaxy correlations on large scales (Maddox et al. 1990; Efstathiou, Sutherland, & Maddox 1990; Picard 1991; Saunders et al. 1991) indicate at a high confidence level that the standard CDM model has to be at least revised. In order to reconcile the CDM model with the present observational limits on the microwave background anisotropy, one must either invoke a scale-dependent bias of the galaxy distribution relative to the mass distribution (Dekel & Rees 1987; Efstathiou et al. 1990) or a positive cosmological constant. Recent CDM simulations involving the Burgers equation approximation (Gurbatov, Saichev, & Shandarin 1989) to reproduce some nonlinear effects of gravity, as well as large N-body simulations of biased open CDM model using the Zel'dovich approximation of the linear theory, resemble the observed galaxy distribution (Weinberg & Gunn 1990; Park 1990). But a detailed quantitative comparison is still lacking and the consistency of the model with the limits on the microwave background has to be fully established. Several alternative models have been developed: hybrid inflationary models in which small-scale and large-scale parameters are formed through different processes (Lilje 1990); and non-Gaussian models involving explosions, topological defects, or textures to initiate the large-scale clustering (Ostriker & Cowie 1981; Vilenkin 1985; Turok 1989). These alternatives are also challenged by the recent observational results.

Various statistical methods have been used to detect local structures and to discriminate among theoretical models (see Pellegrini et al. 1990). The two-point correlation function introduced by Limber (1956) and developed by Peebles (1980) is the most widely used statistics to characterize galaxy clustering. However, its behavior at large scale where the growth of structures might still be in the linear regime is poorly determined because of the large uncertainties in the determination of the normalizing mean galaxy density (de Lapparent et al. 1988). Moreover, the sharp and strongly asymmetric structures as those in the CfA catalog are best constrained by the $N > 2$ order moments of the distribution. Direct access to these high-order statistics is provided by algorithms derived from the graph theory, like percolation analysis (Zel'dovich, Einasto, & Shandarin 1982; Tully 1987; West 1989), cluster analysis (Paturel 1979; Tago, Einasto, & Saar 1984; Barrow, Bhavsar, & Sonoda 1985) or minimal spanning tree techniques (Bhavsar & Ling 1988). But the results of these algorithms are often parameter dependent, and their ability to discriminate among models were questioned (Dekel & West 1985). Furthermore, these algorithms fail to completely describe the clustering hierarchy in which a galaxy can belong to several structures. The properties of low-density regions are studied using the void probability function, which can be written analytically in terms of correlation functions (White 1979; Maurogordato & Lachièze-Rey 1987; Fry et al. 1989; Vogeley, Geller, & Huchra 1991). Several quantitative characterizations of the overall geometry have been applied to galaxy catalogs: the multifractal analysis (Jones et al. 1988; Peebles 1989; Martinez et al. 1990); the variations of galaxy counts-in-cells (Efstathiou et al. 1990; de Lapparent et al. 1991); and the genus analyses, based on topological analyses of a smoothed version of the galaxy distribution (Gott et al. 1989; Ryden et al. 1989). The multifractal and counts-in-cells analyses suggest two-dimensional structures. The genus analyses suggest a spongelike topology, but the results are dependent on the smoothing length and on the density threshold used for defining the topological surfaces. All these indicators provide an objective way to compare observational data with numerical simulations, but they measure only a volume-averaged value of the parameters used to characterize the distribution of galaxies. A local description of the hierarchical properties of this distribution is by consequence out of their reach, and they cannot be used to locate structures.

Expansions of finite-energy functions onto orthogonal bases are widely used since the last century (e.g., Fourier analysis). These decompositions are based on a scalar product defined on the set R where the whole function is used instead of some peculiar values. This unlocalized description is particularly well suited to functions which are not continuous. It can then be applied with great benefit to describe natural structures, i.e., which are intrinsically irregular, like the distribution of galaxies. However, to have a good description of the data, one should also be able to locate independently structures which may exist at different scales. This is particularly crucial for characterizing the galaxy distribution because of the structures are hierarchical. The wavelet transform introduced by J. Morlet (see Goupillaud, Grossman, & Morlet 1984) allows us to scan structures on different scales without deteriorating the spatial information. Performed according to a scale parameter, this space-scale analysis does provide indeed a description which is as localized as possible given the smoothness and the oscillatory behavior of the function chosen as analyzing wavelet (see, for instance, Grossmann, Kronland-Martinet, & Morlet 1988 and references therein). The main difference between the wavelet transform and previous time-frequency methods (e.g., the Gabor transform) is its invariance with respect to dilations of the signal which insures a scale-invariant description of the signal. One of the major applications of the wavelet transform is therefore to detect and to measure locally fractal behaviors (Holschneider 1988; Argoul et al. 1989; Gill & Henriksen 1990).

The wavelet transform can also be used as a new cluster analysis method. Its ability to unfold data in a space-scale phase-space allows one to identify clusters of points according to their characteristic scale. To quantitatively characterize the large-scale structures in the distribution of galaxies, we developed an objective procedure based on the level of statistical significance of each wavelet coefficient (Slezak, Bijaoui, & Mars 1990). This tool was already used successfully to search

for subclustering in Abell clusters (Escalera & Mazure 1992) and to define asteroid families (Bendjoya, Slezak, & Froeschlé 1991). Because a lack of objects with respect to the local background can be analyzed in the same way as an excess of objects, the method also allows detection and characterization of the voids in the galaxy distribution. Here we apply the wavelet analysis to the first slice of the CfA redshift survey (de Lapparent et al. 1986; Huchra et al. 1990).

The paper is organized as follows. Section 2 presents first an overview of the basic definitions and properties of the wavelet transform. Its application to discrete data analysis is described in details in § 3; we discuss in particular the reliability of the detection with respect to a Poissonian noise, and the parameterization of the detected structures. We then report in § 4 on the application of the wavelet transform to the first CfA redshift survey slice and discuss in § 5 the results of our analysis. Finally, § 6 summarizes our conclusions.

2. THE WAVELET TRANSFORM

The wavelet transform of the one-dimensional function $f(x) \in L^2(R)$ is given by its decomposition onto a family of functions which are derived by translations and dilations of a unique bounded function $\psi(x)$. The function $\psi(x)$ is called the analyzing wavelet and can take complex values. Let us denote $\psi^*(x)$ the complex conjugate and $\hat{\psi}(v)$ the Fourier transform of $\psi(x)$. The wavelet coefficient $C(a, b)$ corresponding to the strictly positive scale a and the location b belongs to $H = R^+ \times R$ and it is given by

$$C(a, b) = N(a) \int_{-\infty}^{+\infty} f(x)\psi^*\left(\frac{x-b}{a}\right)dx , \quad (1)$$

where $N(a) = a^{-1/2}$ when the wavelets $N(a)\psi(x/a)$ are normalized in energy. The wavelet transform can also be viewed as a filtering process with a set of passband filters $\hat{\psi}(av)$. Let us denote $\hat{f}(v)$ the Fourier transform of $f(x)$, then the correlation integral in equation (1) can be written as

$$\hat{C}(a, v) = a^{1/2}\hat{f}(v) \times \hat{\psi}^*(av) . \quad (2)$$

In contrast to the standard Fourier analysis, the wavelet transform analyses the signal both in scale and position, especially if the localization and smoothness properties of $\psi(x)$ and $\hat{\psi}(v)$ prevent sinc (x) oscillations in both spaces. The wavelet transform provides a description of $f(x)$ around the location b and an analysis of $\hat{f}(v)$ around v_0/a in Fourier space, where v_0 is the center of the passing band of the filter. This scan of the Fourier space with filters derived from one another by an affinity implies an invariance of the transform under dilations and provide us with a multiscale analysis of the signal. Because the resolution of the wavelet transform increases in the spatial domain and decreases in the frequency domain when the scale decreases, thinner and thinner details of the signal are displayed when smaller and smaller scales are investigated. These variations enable the wavelet transform to zoom into the irregularities of the signal and to characterize them locally. Actually the wavelet transform of a given signal has the same scaling behavior as the signal itself (Holschneider 1988). One can therefore extract from the data more information than a mere global description of the scaling properties.

The main property which must be satisfied by a function in order to be a wavelet comes from the energy conservation between the spaces R and H. To satisfy this criterion matching the Parseval-Plancherel formula for the Fourier transform, the following integral must exist:

$$C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(v)|^2}{v} dv . \quad (3)$$

Therefore $\hat{\psi}(v)$ must be strictly equal to zero at the origin. For differentiable functions $\psi(x)$, it implies that the integral of $\psi(x)$ must be null. The best compromise between spatial and scale resolution is obtained for a Gaussian law. Analyzing wavelets have thus been defined from the even derivatives of this function because they satisfy the admissibility criterion of equation (3). A popular analyzing wavelet is the so-called Mexican hat

$$\psi\left(\frac{x}{a}\right) = \left(1 - \frac{x^2}{a^2}\right)e^{-(1/2)(x^2/a^2)} , \quad (4)$$

which exhibits fast convergence properties and has an even and symmetric Fourier transform.

Reconstruction of the function $f(x)$ from its wavelet coefficients $C(a, b)$ is given by (Grossmann & Morlet 1985):

$$f(x) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \int_0^{+\infty} \frac{1}{\sqrt{a}} C(a, b)\psi\left(\frac{x-b}{a}\right) \frac{da\, db}{a^2} . \quad (5)$$

This solution is however not unique due to the genuine overdetermination of the wavelet transform which associates a two-dimensional function to a one-dimensional function $f(x)$. Nevertheless equation (5) allows us to interpret the value of $f(x)$ at a given location x as the limit of the weighted sum of the contributions at various scales around x. In other words the whole set of structures at different scales around the location x determine the value $f(x)$. Equations (1)–(5) can be generalized to multidimensional signals with or without distinguishing any spatial orientation (Mallat 1989b; Antoine et al. 1990).

The redundancy of the wavelet transform allows us to define a complete and bounded representation of $f(x)$ by restricting the half-plane H to a suitable discrete lattice $\{a_m, b_{n;m}\}$. The goal is of course to extract a finite set of coefficients $C(a_m, b_{n;m})$ which enables one to restore entirely the information by use of an interpolation. The errors of reconstruction give thus a measure of the quality of the sampling. It can be easily shown from the invariance of the wavelet transform under dilations that the sampling step of the space axis must be proportional to the scale parameter. The fact that the filters $\hat{\psi}^*(av)$ have the same bandwidth $\Delta v/v$ on a logarithmic scale suggests to use for the frequencies a uniform sampling on a logarithmic scale (see also the binary decomposition of the information implied by the Littlewood-Paley's scheme 1937). Therefore the discrete wavelet transform is based on the following samplings:

$$a_m = a_0^m \quad b_{n;m} = nb_0 a_0^m .$$

Daubechies (1990) studied in details the completeness and stability of the discrete wavelet transform. In particular, she showed that in order to reconstruct the signal the norm of the restored function must be bounded by two constants A and B which depend on the sampling lattice and on the wavelet itself. Using the coefficients computed with the set of wavelets $\{\psi[(x - nb_0 a_0^m)/a_0^m]\}$ the error on the norm is then proportional to the value $\epsilon = BA^{-1} - 1$. For the Mexican hat, Daubechies showed that there exists always a small error in the reconstruction whatever the sampling steps are. For example, with $a_0 = 2^{1/2}$ and $b_0 = 0.5$, she derives numerically a negligible $\epsilon = 2\ 10^{-4}$, which ensures in practice an accurate enough restoration of the initial function $f(x)$. No error occurs when

the bounds A and B are equal. This is the case for an important class of discrete wavelets for which $A = B = 1$: these functions define orthonormal bases of $L^2(R)$ and are called orthogonal wavelets (see Meyer 1989, and references therein). A major breakthrough in the understanding of orthonormal wavelet bases came with the concept of multiresolution analysis developed by S. Mallat (1989a) similar to the Laplacian pyramid scheme of Burt & Adelson for image decomposition and reconstruction (Burt & Adelson 1983).

Apart from functional analysis, the wavelet transform, and especially the multiresolution description of S. Mallat, are well suited to applications in signal processing such as matching algorithms, texture discrimination, signal coding and image restoration. In the field of astronomy, the multiscale analysis provides a new way to analyze images: no parameters, like a definite smoothing length, the background scale or the detection threshold are needed. Intricate objects are decomposed into features with different scales detected at given resolution levels (Bijaoui et al. 1989). The wavelet transform can also be viewed as a new powerful method for clustering analyses (Slezak et al. 1990). The interest of this technique is that it describes the structures with no a priori assumption about the hierarchical nature of the clustering, as in single or complete linkage methods (Materne 1978), and with no a priori knowledge of an initial partition of the set of points, as in non-hierarchical methods (Paturel 1979; Moles, del Olma, & Perea 1985). Moreover, because of the strong mathematical support which exists for the wavelet transform, the confidence level of the detection with respect to a Poissonian process can be computed analytically as explained in the next section.

3. DISCRETE DATA ANALYSIS

Let us consider a one-dimensional set of discrete data to be analyzed, for example a set of P points with coordinates $\{x_k\}$. This catalog of points can be represented in a continuous manner by means of Dirac functions.

$$f(x) = \sum_{k=1}^{P} \delta(x - x_k) \, . \tag{6}$$

This modeling is fully justified because the wavelet transform is linear. The discrete wavelet transform of this catalog is then the correlation product, at each location $b_{n;m}$ associated to the spatial resolution, of the wavelet function at the studied scale a_m by the sum of the unit Dirac functions which model the discrete data:

$$C(a_m, b_{n;m}) = N(a_m) \int_{-\infty}^{+\infty} \sum_{k=1}^{P} \delta(x - x_k) \psi^* \left(\frac{x - b_{n;m}}{a_m} \right) dx \, . \tag{7}$$

If the analyzing wavelet is not defined on a compact domain, one can choose a regular wavelet with fast convergence properties so that it can be usually considered as equal to zero outside a limited area. Therefore only the K points inside this limited area $V(a)$ are involved into the computation, and by consequence the wavelet coefficient C is the discrete sum of K values of the wavelet function. Equation (7) thus becomes the following equation which can be extended to multidimensional spaces:

$$C(a_m, b_{n;m}) = N(a_m) \sum_{k=1}^{K} \psi^* \left(\frac{x_k - b_{n;m}}{a_m} \right) . \tag{8}$$

The aim of a wavelet analysis applied to a discrete distribution is similar to that of cluster analysis: the description of how the data cloud is structured. The advantage of the wavelet analysis is to provide several classifications for the same point according to the investigated scale, and to allow examination of the relationships between the different levels of the hierarchy of structures, which is very useful for fractal or multifractal distributions. The wavelet transform also allows one to examine regions with a deficiency in points with respect to a uniform distribution. These underpopulated regions are a major feature of the large-scale galaxy distribution in the CfA redshift survey slices, but their individual detection is out reach of a classical technique, such as for instance, the void probability function which gives only a global result at the examined size.

3.1. *Significant Coefficients*

One characteristic property of the wavelet transform is to yield coefficients equal to zero for a constant signal (i.e., a uniform distribution in the case of a catalog of points). Consequently the existence of structures at a given scale is tied to the presence at this scale of wavelet coefficients with a large enough absolute value. A random distribution of points can exhibit coefficients different from zero due to the statistical fluctuations in the spatial repartition of these points. Hence, the study of structures is based on the analysis of the statistical significance of the coefficients, by comparison with the values obtained from a locally uniform Poissonian distribution. The level of significance is a function of the contrast of the structure with respect to its local environment.

The statistical significance of a coefficient C is determined by evaluating the probability $P(c > C)$ to obtain a higher value caused by a chance fluctuation of the underlying random process. We therefore need to compute the probability distribution function of the wavelet coefficient $F(c)$, and thus its probability density function $p(c)$, in the case of a Poisson distribution:

$$P(c > C) = 1 - F(C) = 1 - \int_{-\infty}^{C} p(c) dc \, . \tag{9}$$

Besides its dependence on the analyzing wavelet, the law $p(c)$ depends on the number K of events involved in the coefficient determination, and therefore on the local density $\rho(x)$ and on the volume $V(a)$ under investigation. Because the characteristic function (i.e., the Fourier transform of the probability density) of a sum of K independent random variables is the product of their characteristic functions, the probability density function of the coefficient is equal to $K - 1$ convolution products of the law associated to one event. This law is precisely the density of the values of the analyzing wavelet: the value of the wavelet coefficient for a single event is determined by the location of this event inside the domain where the wavelet is defined, and every position of the domain has the same probability to be occupied. A Taylor expansion of the probability density $p(c)$ with respect to $[\rho(x)V(a)]^{1/2}$ shows that it converges toward a Gaussian law for a large number of events all the more rapidly as the dimension of the space increases (Bijaoui 1989).

The value of the distribution function $F(c)$ for a given coefficient provides information on the amplitude of the chance fluctuations and allows a statistical interpretation. For a uniform distribution of events, $F(c) = 0.5$. A value close to zero $F(c) < \epsilon_d$, or to unity $F(c) > 1 - \epsilon_d$, implies a deficiency, respectively an excess of events, with an increasing significance level as ϵ_d decreases. Note, however, that K has to be greater

than 2 for the overdense regions and at least equal to 4 for the underdense regions in order to provide a reliable measurement. Otherwise a location b and a scale a can always be found such that the value of the wavelet coefficient seems to be significant.

3.2. *Detection of Structures*

We detect the structures by examining the extrema of the significance level map built from the set of wavelet coefficients. Because the most contrasted structures are those which are most likely to be real, the maxima and the minima of this map locate the overdense and underdense regions, respectively. The quality of the detection is insured by selecting the extrema which are greater or equal to a high confidence level threshold (we use 99.5%).

The independence of the wavelet coefficients is essential if their statistical significance is to be derived from the probability density $p(c)$. However a detection of the extrema with a sufficiently high spatial resolution implies use of a denser sampling than the one leading to disjointed integration areas $V(a)$ for the computation of each coefficient. The overlap of these areas produces a correlation between the values of the coefficients for neighboring regions. If the spatial resolution is fixed, the larger the studied scale, the greater is the correlation. Therefore, from a mathematical point of view, the ensemble average cannot be replaced by a spatial average, and the ergodicity of the stochastic process is lost. The use of the probability density of the wavelet coefficient thus yields only an estimation of their significance level called confidence level. By consequence the threshold chosen for the detection of structures only approximates a thresholding at the same value of significance.

Nevertheless, despite the spatial correlation of the coefficients, the significance level of the detected structures can be estimated. In practice, the distance between the extrema which locate structures is at least equal to that necessary to insure their mutual independence. The maximum number n_i of these independent coefficients is given by the ratio of the total number of wavelet coefficients by the area $V(a)$. So let us consider these independent coefficients computed from a uniform distribution. By definition, and due to statistical fluctuations, only $\epsilon_d n_i$ of them are detected with a confidence level $1 - \epsilon_d$. Hence, the quantity $\epsilon_d n_i$ is an estimation of the maximum number of extrema—and thus of structures—which are a false detection for the considered detection threshold and wavelet scale. The ratio between $\epsilon_d n_i$ and the total number of detected structures in the analyzed distribution gives therefore the probability that a given structure is observed by chance. All the structures have obviously this same significance level. Additional information about the structures themselves has to be taken into account in order to assess more precisely the physical reality of a given structure. It can be noted of course that detected structures with the highest confidence level have the lowest risks of being random fluctuations.

The confidence level of the wavelet coefficients is computed at each scale with respect to a Poisson distribution with a spatially varying density parameter. However, if a structure is detected at a given scale, the null hypothesis for larger scales at this location is no longer a uniform distribution. The Poisson noise depends on the signal itself and applying the previous scheme results in a biased detection for these large scales since the high-density clumps lead to a higher noise value. This problem occurs for instance in the multiple detections induced by the multiscale linear analysis of the wavelet transform. Thus, even the wavelet transform separates the different scales almost totally, the Poisson noise forces us to examine the influence of small scale structures on the statistical level of the detection of larger ones (see § 5). This question can be examined via numerical simulations of the data. Using a density map built from the real data in order to generate pseudorandom catalogs, statistics on the values of the wavelet coefficients can be computed at the locations where structures have been detected using the Poisson hypothesis. The standard deviations give a distance criterion with respect to a null coefficient and then yield a more accurate measure of the statistical significance of the real coefficients at these locations. These significance levels are obviously lower (higher) for the high- (resp. low-) density structures than those computed from the r.m.s. statistic obtained from uniform Poisson distributions.

3.3. *Descriptions of Structures*

The detection of the extrema allows us to locate the significant structures but lacks the fundamental information about their morphology because of the local status of this kind of detection. Position uncertainties can furthermore result from too large a discrepancy between the shape of the analyzing wavelet and that of the structure. The morphological properties can formally be derived from the shape of the contour outlining the coefficients which have a confidence level identical to the chosen detection level. Most of the time the domains defined in this way include one extremum. Nevertheless one must be able to take into account the correlation between the coefficients before extracting from this contour relevant measurements about the morphology. We solve this difficulty by analyzing the morphology of the initial point distribution inside the area $V(a)$ around the extremum.

Dealing with essentially bidimensional data, let us consider a catalog of positions on a plane. As indicated by equation (8), the discrete isotropic wavelet transform of a two-dimensional catalog yields a set of two-dimensional images. One way to characterize the structures detected at a given scale is to fit to each of these structures a two-dimensional Gaussian elliptical distribution. This choice is justified by the compromise between the number of shapes modeled by this fit and its simplicity. However one can already note that nonconvex structures will be poorly described. The algorithm that we have developed is applied directly to the data points. The wavelet analysis provides only the approximate coordinates of the structure (indicated by the position of the corresponding extremum) and an estimation of the size of the structure (the wavelet scale a). The characteristics of the elliptical model—the center, the position angle, and the ellipticity—are determined by identifying the statistical moments of the local distribution with those of an ellipse. The value of the major axis (a quantitative measure of the background-corrected extension of the structure) is derived by fitting the radial density profile around the position of the extremum by a Gaussian law and adopting 3 times the FWHM of the fitted Gaussian (the mathematical method is that of Slezak et al. 1990 with some minor technical improvements).

The application of the elliptical modeling to underpopulated areas sets a problem which is imperfectly solved. The difficulty lies in the computation of the moments of these regions: there is no way to attach to the lack of data points a function analogous to the sum of Dirac functions which would stand for their presence. Therefore, the straightforward moments calculated for the underpopulated regions using for each of them the

points within the $V(a)$ area centered on the minimum do not describe the characteristics of the void but those of the data points delineating the void. Actually the difference is large when the distribution of data does not isotropically surround the voids and when the latter have a high contrast. The highly deficient voids delineated by sharp sheetlike structures of the CfA survey illustrate this extreme case. A better approximation to the moments of the underdense regions is the difference between the classical moments describing the surroundings of the void and the moments computed from the same region uniformly populated. We therefore chose to fit ellipses to these moments in the same fashion as for the overdense regions.

The limits of this kind of description are those of the modeling which is performed. Structures with a curved shape and underdense regions are obviously not well described by a two-dimensional Gaussian elliptical modeling fitted on moments based on data points. A solution which overcomes the problem of the correlation between the coefficients lies in the reconstruction of structures from the significant coefficients at all scales and the analysis of the resulting contour at a given resolution. The detection, localization and parameterization of the structures will however remain unchanged. This issue will be discussed in a further paper.

3.4. *The Physical Objects*

The scales at which the wavelet transform can detect structures in a set of points range from the scale defined by the mean separation between points up to about one third of the largest separation between points. Knowledge about the physics of the objects or the processes under study allows us, however, to stop this hierarchy before these extreme limits. These are the free parameters common to all techniques for describing a cloud of data.

The linearity of the wavelet transform implies that even low-order moments structures are detected at several scales. Therefore, one must keep in mind the fundamental difference between the physical object and the structures detected at the same location by the wavelet transform. These structures depict only the object at each different scale whether it is isolated or embedded into a larger structure. An object is defined only by the full hierarchy of these structures.

4. PROCESSING OF THE CfA REDSHIFT SURVEY SLICE

In order to obtain preliminary results about the large-scale distribution of galaxies in the Universe, the ongoing CfA Redshift Survey is divided into contiguous strips across the sky. Each strip is 135° wide in right ascension and 6° thick in declination. The projected catalog was derived from a merge version of Zwicky's (Zwicky et al. 1961–1968) and Nilson's (Nilson 1973) catalogs of galaxies. Radial velocities measurements for all galaxies brighter than $B = 15.5$ have been already completed for several slices (Geller & Huchra 1989). Here we analyze the available first CfA slice bounded by $8^h \leq \alpha \leq 17^h$ and $26^\circ.5 \leq \delta \leq 32^\circ.5$ (Huchra et al. 1990). Because the large-scale structures are poorly defined at velocities larger than 15,000 km s^{-1} (the separation between galaxies becomes comparable to the size of the voids). we include in our analyses only galaxies below this velocity.

Because the spatial extent of the slice is much smaller in declination than in right ascension, it is reasonable to ignore the declination coordinate and project the sample onto a constant declination surface. We thus consider the slice as a two-dimensional distribution where the polar coordinates of each object are its right ascension and radial velocity. The discrete wavelet analysis of this catalog using an isotropic wavelet becomes a set of two-dimensional images which can be handled easily and contain most of the relevant information about the large-scale clustering in the slice.

The first CfA slice contains the Coma Cluster, with its prominent finger-of-God superposed on the underlying large-scale clustering. The ability of the wavelet transform to detect small departures from homogeneity makes the detection of this preeminent feature doubtless. However, this structure can prevent the detection of nearby small clumps due to the linearity of the wavelet transform. We thus remove the 175 galaxies which are members of the Coma Cluster as defined in the group and cluster catalog-extracted from the slice by Ramella et al. (1989).

The velocities of the 880 remaining objects have then been scaled so that a homogeneous Poissonian process with the same limiting apparent magnitude and luminosity function as the data (see de Lapparent et al. 1989a) leads to a uniform distribution of constant density. This transformation is called an anamorphosis. The anamorphosed velocities $V_{\rm ana}$ are derived from the true velocities V by

$$V_{\rm ana}(V) = K \int_0^V N_{\rm exp}(v)dv + K_0 \,, \tag{10}$$

where

$$N_{\rm exp}(v) \propto V^2 \Gamma\{\alpha + 1, \text{dex}\,[0.4(M^* - m_{\rm lim} - 5\log v - 15]\} \tag{11}$$

is the expected number of galaxies per velocity interval unit at velocity v (given in km s^{-1}) in a survey with a limiting magnitude $m_{\rm lim}$ and a Schechter luminosity function with characteristic parameters M^* and α. The function $N_{\rm exp}$ resembles a skewed bell distribution with a peak at $V_{\rm peak} = 5400$ km s^{-1} for the luminosity function derived for the slice ($M^* = -19.2$ and $\alpha = 1.1$; de Lapparent et al. 1989a). The constants K and K_0 are determined by setting

$$V_{\rm ana}(V_{\rm peak}) = V_{\rm peak} \tag{12}$$

$$V_{\rm ana}(V_{\rm max}) = V_{\rm max} \tag{13}$$

with $V_{\rm max} = 15{,}000$ km s^{-1}.

With the anamorphosis, the pie-shaped slice becomes a square with right ascension and velocity as Cartesian coordinates: we arbitrarily scaled the velocities so that the 8^h–17^h range in right ascension is equal to the 15,000 km s^{-1} range in velocity. The distance between two objects are therefore given as velocities. Because the large-scale structures in the CfA slice are likely to be unbound, and because the peculiar velocities of these structures are small compared to the size of the voids (Bothun et al. 1989), this confusion of the velocity and spatial coordinate for structures which are not aligned or perpendicular to the line of sight is reasonable. On the contrary, a separate processing of the two coordinates would have been necessary for analyzing relaxed structures such as clusters of galaxies. Note that the anamorphosis distorts the structures. Voids at $V \ll V_{\rm peak}$ or $V \gg V_{\rm peak}$ are systematically smaller (resp. larger) than those at $V_{\rm peak}$ and their shape is also modified. However, the anamorphosis ensures that a given structure would be detected with nearly the same detection level at any velocity inside the survey. Subsequent morphological analyses can then be performed on the true velocity distribution in the pie diagram configuration.

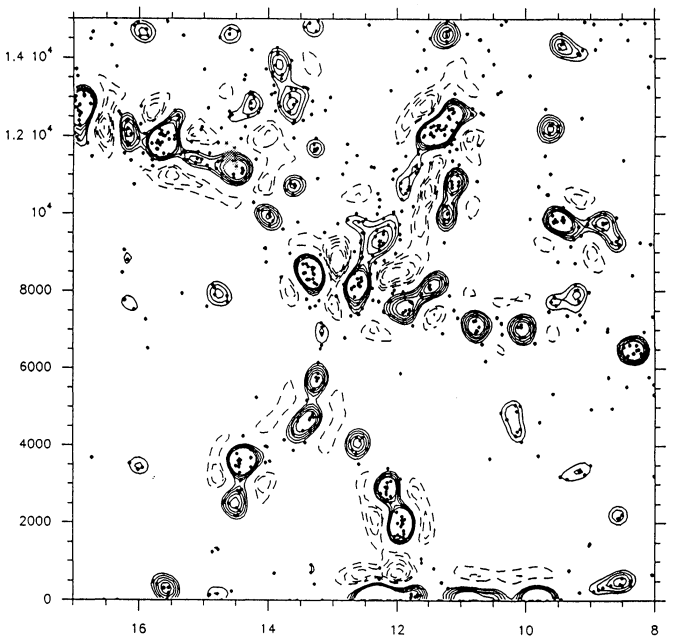

FIG. 1*a*

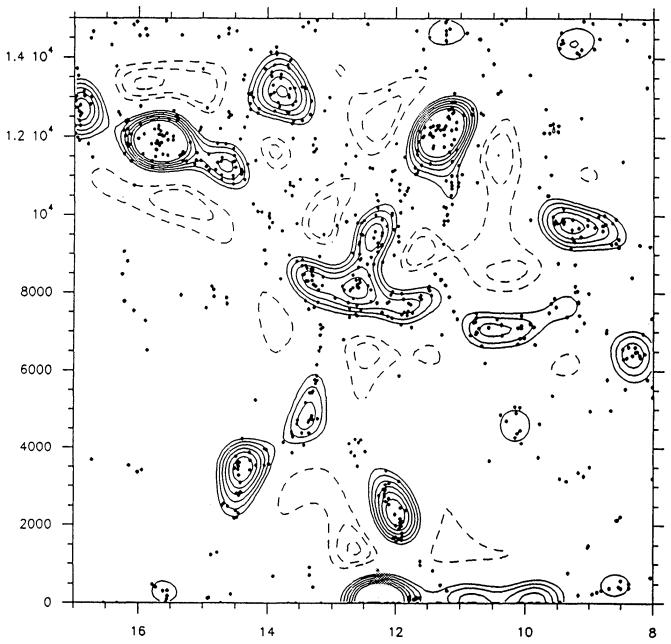

FIG. 1*b*

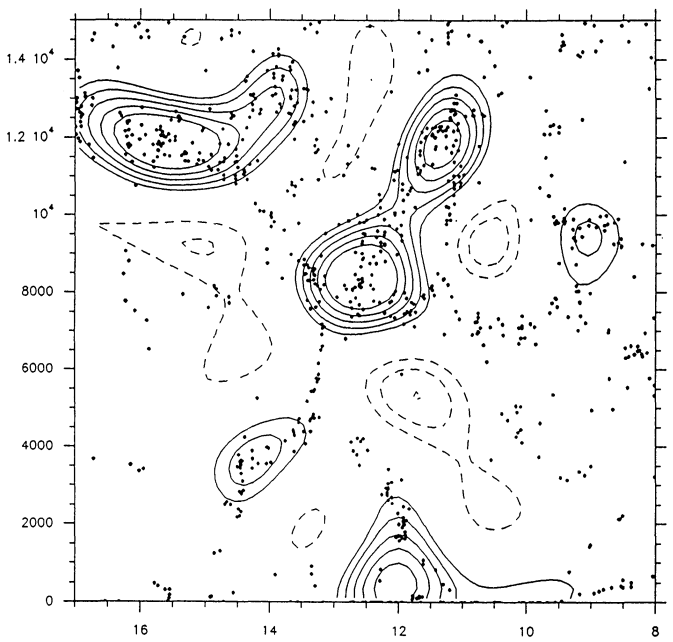

FIG. 1*c*

FIG. 1.—Isopleth maps of the wavelet coefficients superposed on the first CfA redshift survey slice with the Coma Cluster removed (each galaxy is indicated by a dot), for scale values (Fig. 1*a*) $a = 350$ km s^{-1}, (Fig. 1*b*) $a = 700$ km s^{-1}, and (Fig. 1*c*) $a = 1400$ km s^{-1}. Solid (dashed) contours stand for isopleths which are between 1 r.m.s. and 3 r.m.s. above (below) zero with increment of 0.4 r.m.s. Overdense regions are detected for values of a corresponding to their characteristic scale.

We use for the analyzing wavelet the radial Mexican hat given in equation (4). This isotropic wavelet is the simplest compromise for analyzing the CfA slice where large-scale structures seem to occur with any orientation with respect to the line of sight. Moreover, this second derivative of the Gaussian function can be considered from a practical point of view as compactly supported, which results in fast computations. Another advantage of the Mexican hat function is that its two first moments are equal to zero, and therefore the resulting wavelet analysis is insensitive to additive backgrounds or to gradients. Finally, it does not differ very much from the other wavelets actually used in image processing (e.g., spline wavelets). The dependence of our results on the analyzing wavelet will be all the more weak that the probability density function $p(c)$ takes into account the detailed shape of the selected wavelet.

We compute the discrete wavelet transform of the CfA slice for eight scales a increasing with a factor $\simeq 2^{1/2}$ from 175 km s^{-1} to 2000 km s^{-1}: 175 km s^{-1}, 250 km s^{-1}, 350 km s^{-1}, 500 km s^{-1}, 700 km s^{-1}, 1000 km s^{-1}, 1400 km s^{-1}, and 2000 km s^{-1}. The lower limit is close to the typical velocity dispersion of galaxies in groups while the upper limit is constrained by the extent of the survey. For each scale a, the grid of locations b yields a two-dimensional square map. The resolution of the maps varies from 256 × 256 pixels for the smallest value of a to 64 × 64 pixels for the three largest values.

Figures 1*a*–1*c* show the contour maps of the wavelet coefficients derived from the CfA slice for resp. $a = 350$ km s^{-1}, $a = 700$ km s^{-1}, and $a = 1400$ km s^{-1}. The isopleths correspond to significance levels of 1–3 σ with increment of 0.4 σ. The significant coefficients correspond to high-contrast structures with respect to their local neighborhood: for example a 3 σ level corresponds to a density contrast $\rho/\bar{\rho}$ of ~ 10 (~ 5) for scale $a = 350$ km s^{-1} (resp. $a = 700$ km s^{-1}). The contours in Figures 1*a*–1*c* show that visible structures and voids are efficiently identified by the wavelet transform. The wavelet transform provides in addition the location of the structures within the survey, an important piece of information which is lacking in classical cluster analysis methods and in global statistical analyses. Note that structures are exhibited only at a wavelet scale a corresponding to their characteristic size. The different scales are thus clearly decoupled and can be analyzed separately. The hierarchy of structures corresponding to a same object can, however, be derived by comparing the coefficient maps for different scales.

One may wonder how the maps of Figures 1*a*–1*c* would compare with maps of density inferred from the distance between a given gridpoint and its *n*th nearest data point, using varying values of *n*. Even if greater prominence would be given to filaments because spherical symmetry is not enforced as in our isotropic wavelet analysis, the main drawback of this other approach is that different length scales are mixed. A characteristic length scale can be associated to each value of *n* by considering the peak of the distribution of neighbor separations. But the frequency distribution of these separations is not Gaussian for small *n*'s due to small-scale clustering and the tail at large distances prevents a narrow-banded scale resolution.

5. RESULTS AND DISCUSSION

For objective detection of the overdense and underdense structures, we use the maps of confidence levels of the wavelet coefficients with a threshold value of $\epsilon_d = 5 \times 10^{-3}$ (see § 3.1). Figure 2 shows the corresponding isopleths for the overdense structures detected at wavelet scales $a = 250$ km s^{-1}, $a = 350$ km s^{-1}, $a = 500$ km s^{-1}, and $a = 700$ km s^{-1}. The complex organization of the high-density regions is clearly visible as well as the multiple belongings of most of the objects. However, the limits of our method appear clearly. The isotropic wavelet used in this analysis is not suited for detecting the highly anisotropic sheetlike structures delineating the voids. Therefore, as far as the overdense regions are concerned, the radial Mexican hat analyzing wavelet appears essentially best designed for detecting structures on scales comparable to or less than the thickness of the sheets, i,e, groups and clusters of galaxies.

In order to describe the significant structures detected in the CfA slice further, we fitted ellipses to the distribution of objects around each maximum from the confidence level map according to the method described in § 3.3. The local background density is estimated for each structure from its radial density profile, allowing to compute the excess of galaxies in each ellipse. Figure 3 shows the fitted ellipses, displaying only those associated to an excess of objects greater or equal to 3.

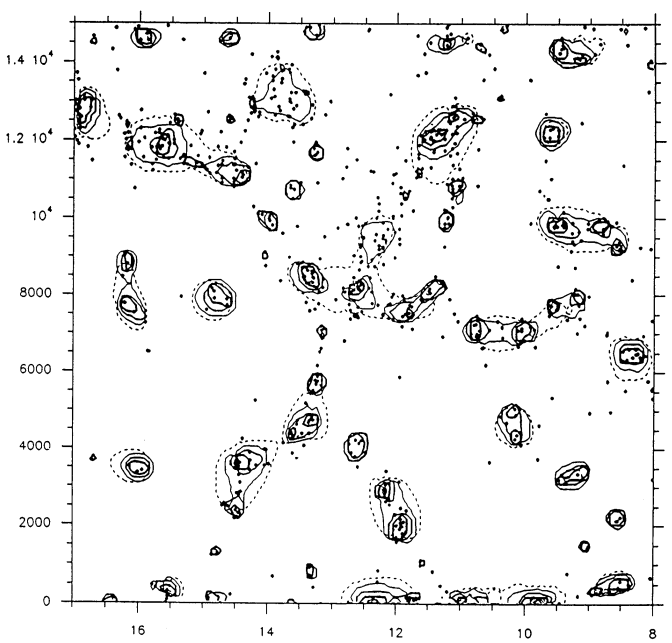

FIG. 2.—Map of the significant overdense structures. The isopleths define areas where there are less than 5 odds in 1000 of measuring in a random distribution a higher wavelet coefficient than the observed positive value. The four plotted contours correspond to wavelet scales $a = 250$ km s^{-1}, $a = 500$ km s^{-1}, and $a = 700$ km s^{-1}. For a given structure, the contours enclosing the larger areas correspond to the larger wavelet scales.

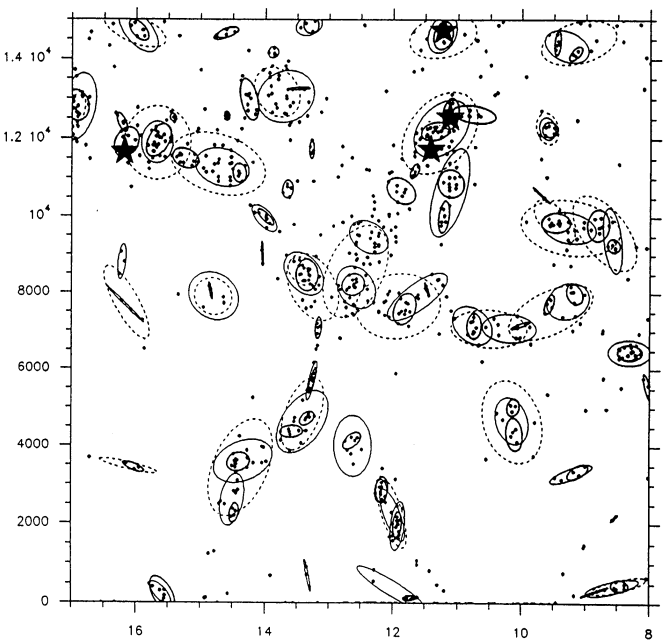

FIG. 3.—Map of the ellipses fitted to the significant structures. The dashed line ellipses correspond to scale $a = 700$ km s^{-1}, and the solid line ellipses (with increasing thickness) correspond to scales $a = 500$ km s^{-1} and $a = 250$ km s^{-1}. The three Abell clusters included in the slice in addition to the Coma Cluster are indicated by black stars.

The Poisson noise depends on the signal. As a consequence, we have examined the dependence of the detection of large-scale structures on the small-scale clustering by using 100 pseudorandom catalogs sampled from an underlying density field. This density map is obtained from the first CfA slice catalog using Dressler's definition (1980) of the local density with the third nearest neighbor. In practice, the third nearest neighbor appears to be a good compromise between keeping highly contrasted structures and minimizing the small-scale noise. In each simulation the wavelet coefficients are computed for scales $a = 500$ km s^{-1}, $a = 700$ km s^{-1}, and $a = 1000$ km s^{-1} at the location of the structures previously detected, and the mean values and r.m.s. fluctuations of these coefficients are calculated at each location. The mean coefficients are obviously always smaller in absolute value than those computed from the observational catalogue due to the smoothing. A measure of the expected fluctuations caused by the small-scale clustering of the data is given by the r.m.s. values. Each r.m.s. value can be used for deriving a new estimate of the statistical significance of the wavelet coefficient computed at the corresponding location from the CfA data.

Most of the detected structures are found to be above the 3 σ level ($\simeq 70\%$ for $a = 1000$ km s^{-1}). Thus, one can deduce that these large-scale features are still highly significant when the noise caused by the small-scale clustering is taken into account. The least significant structures are associated to a 2 σ level. They correspond mainly to galaxies which are located inside the two largest voids of the first CfA slice (e.g., galaxies

around 16^h, 7500 km s^{-1}). These structures are sparsely populated and the small number of objects ($\simeq 5$) yields a large fluctuation with respect to the value of the wavelet coefficient when the scale is greater than 500 km s^{-1}. The other marginal detections ($\simeq 25\%$ detections between a 2.5 and 3 σ significance level) are obtained for sparse groups containing a few objects and for the large structures including two well-defined components (e.g., structure at scale $a = 700$ km s^{-1} around 9^h30, 7500 km s^{-1}). This example is a good illustration of the importance of a detailed analysis of the significance level. In this particular case, each of the subclumps is individually detected at a small scale with a high level of confidence. But the detection level of the larger structure which includes these clumps is lowered due to the higher Poisson noise caused by the subclumps. Ideally a real multiscale approach would be required to fully address this issue. The use of the previous Poisson hypothesis for detecting structures corresponds to a scale-by-scale examination of the data where no information on other scales but the linearity of the wavelet transform is considered. Obviously this may not be sufficient in the case of highly contrasted hierarchical structures with a Poisson noise.

Because we are using an isotropic wavelet, the sheetlike structures are described by contiguous ellipses of size comparable to the thickness of the sheets and corresponding to small-scale clustering within the sheets. In particular, the three Abell clusters included in the slice (Coma A1656 being removed) are included in one of the significant ellipses. Hence this method appears promising for the objective detection of clusters of galaxies in redshift maps. In these maps, the elongation along the line of sight and the projected size of the bounded and/or relaxed structures are two physically different parameters. Use of an isotropic wavelet prevents us from making the distinction between these two quantities. A physically meaningful detection would thus have to be based on an anisotropic wavelet elongated along the line of sight. The velocity dispersion and the projected size of each detected cluster could be derived by varying the two characteristic dimensions of the wavelet and choosing the values yielding the best detection level. As any other method for detecting clusters in redshift maps, this anisotropic wavelet analysis might still confuse sheets aligned along the line of sight with dynamically elongated structures.

To test the ability of the wavelet analysis for detecting small-scale clustering, we compared the maps of ellipses fitted to the significant structures detected at scales $a = 250$ km s^{-1} and $a = 500$ km s^{-1} with the catalog of groups extracted from the CfA slice by Ramella, Geller, & Huchra (1989). In this catalog, the groups are detected by treating separately the velocity width and projected size of the structures. Figure 4 shows the ellipses fitted to the significant structures detected at scales $a = 250$ km s^{-1} and $a = 500$ km s^{-1} superposed on the groups by Ramella et al. after restoration of the wedge diagram configuration (the ellipses are slightly distorted due to the cone geometry). The Great Wall (Geller & Huchra 1989) appears clearly in Figure 4 as a continuous alignment of fitted ellipses. Moreover, there is good agreement between the map of ellipses and the map of groups: most of the dynamically elongated groups are included at least in one of the fitted ellipses. In order to quantify the agreement between the ellipses derived by the wavelet analysis and the groups detected by Ramella et al., we calculate the fraction of groups with their center included inside the ellipses fitted to the structures detected on scales of $a = 250$ km s^{-1} and $a = 700$ km s^{-1}, of order of the typical velocity dispersion of groups and clusters, respectively (see Zabludoff, Huchra, & Geller 1990), The fractions are 53% (resp. 68%) for groups with $N = 2$ members, 43% (resp. 71%) for groups with $N = 3$–4 members, and 45% (resp. 90%) for groups with $N \geq 5$ members. In addition the fractions of groups included at least in one of the ellipses on scales ranging from 175 km s^{-1} to 500 km s^{-1} are 88% for groups with $N = 2$ members, 89% for groups with $N = 3$–4 members, and 83% for groups with $N \geq 5$ members. The isotropic wavelet analysis detects efficiently small-scale clustering because the projected size of these small groups of galaxies is close to their velocity dispersion.

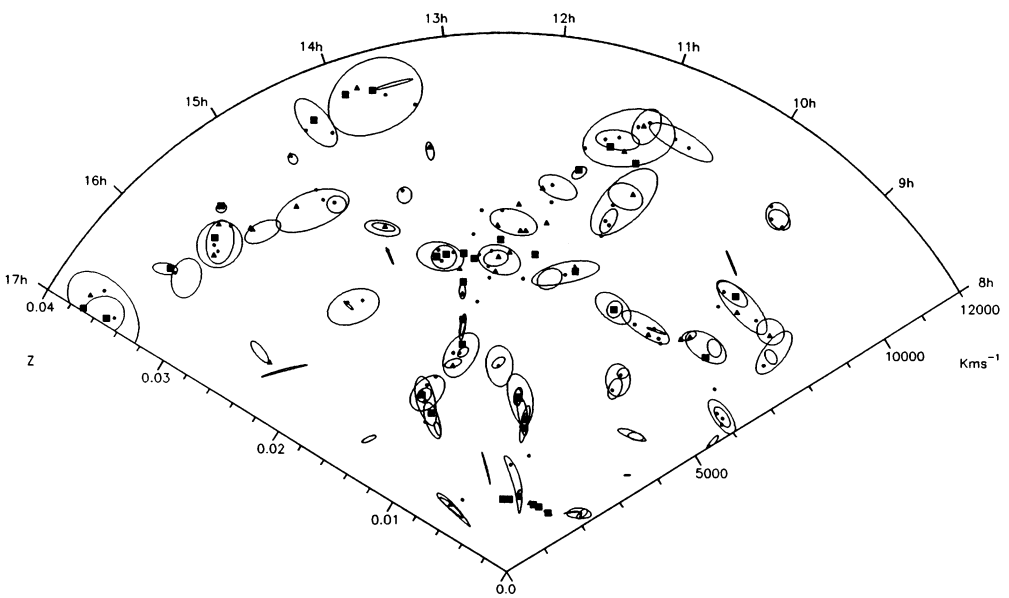

FIG. 4.—Comparison of the ellipses fitted to the significant overdense structures detected for scales $a = 250$ km s^{-1} and $a = 500$ km s^{-1} with the groups identified by Ramella et al. (1989). Dots, triangles, and squares correspond to groups with $n = 2$, $n = 3$–4, and $n \geq 5$ members, respectively. The ellipses appear distorted due to the cone geometry. Note that only structures beyond $V_r = 1635$ km s^{-1} can be detected due to the anamorphosis (see text).

To avoid the difficulty mentioned above in detecting elongated structures by means of an isotropic wavelet, one can also consider the sheetlike structures in the CfA slice as the counterparts of the voids. The isotropic analyzing wavelet is best suited for detecting the nearly spherical voids of the slice. Moreover, in our analysis of this sample, the underdense regions are the most significant structures for wavelet scales a greater than 1000 km s^{-1}, which is twice the typical FWHM thickness of the sheets (cf. de Lapparent et al. 1991). The coherence of sheetlike structures like the Great Wall is indeed likely to be geometric and not dynamical (de Lapparent et al. 1988). We can thus consider the voids as the physical units of the distribution, and the network of walls is then uniquely defined by a map of the voids indicating their sizes and relative positions.

Figure 5 shows the map of confidence levels for the negative wavelet coefficients obtained at scales $a = 350$ km s^{-1} and $a = 700$ km s^{-1}. The contours correspond to a threshold value of $\epsilon_d = 5 \times 10^{-3}$ (see § 3.1). Figure 5 shows that the wavelet method is able to detect, locate, and decouple voids existing at different scales. The need for such a decoupling is illustrated by the fact that a structure can be significant at a given scale but not at higher and/or lower scales. To derive relevant physical parameters for the voids, we fit to each detected void an ellipse according to the procedure described in § 3.3 and we compute its deficiency in galaxies with respect to the local background density from its radial density profile. The influence of the small-scale clustering on the detection of the underdense structures has been tested for scales a between 700 km s^{-1} and 2000 km s^{-1} in a similar fashion as for the overdense structures. All the detected voids have a confidence level above the 3 σ limit with the two largest voids at scales $a = 2000$ km s^{-1} reaching values close to 6 and 9 σ.

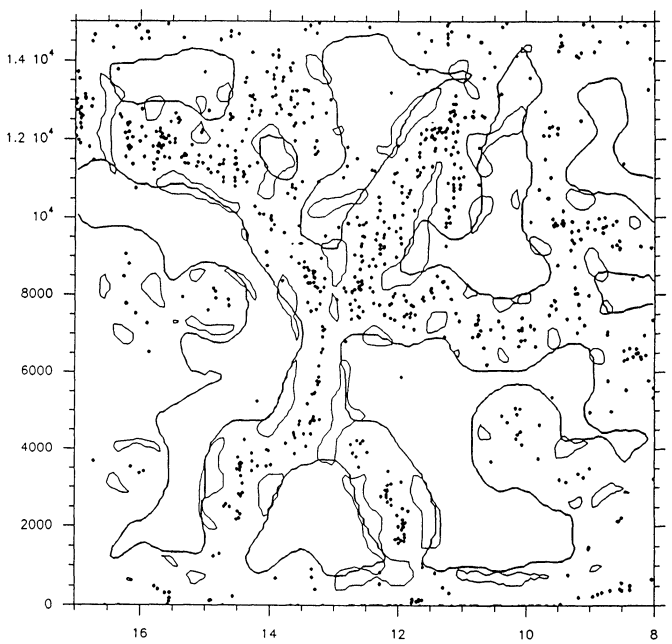

FIG. 5.—Map of the significant underdense structures. The isopleths define areas where there are less than 5 odds in 1000 of measuring in a random distribution a lower wavelet coefficient in absolute value than the observed (negative) value. The two plotted contours correspond to wavelet scales $a = 350$ km s^{-1} (*light contour*) and $a = 700$ km s^{-1} (*heavy contour*). The wavelet method can locate voids, whatever their typical size.

Our goal is to produce a catalog of the significant underdense regions in the CfA slice which could be used for follow-up investigation of the properties of these structures. To this purpose we must further process the list of fitted ellipses because some of the corresponding voids are artifact and others are redundant with larger voids. First, one can notice on Figure 5 that isolated clouds of galaxies and sharp sheets generate on small scales significant underdense regions which cannot be identified with genuine voids—they are part of larger voids and most of them are delineated by overdense structures only in a sector $< \pi$. This occurs because the wavelet analysis detects all density variations. To remove these artifacts we use a density threshold criteria designed to detect strong anisotropies in the *local* density contrast of the regions surrounding the detected voids. We compute the density angular profile of each structure using a 45° resolution and including only objects which lie in the shell defined by the fitted ellipse and the ellipse derived from the former by a similarity with a factor 1.5. Then the density of the 135° sector centered on the position corresponding to the maximum value of the angular profile is divided by the mean value over the other sectors. The corresponding structure is subsequently withdrawn from the final catalog if the density ratio exceeds a value of 5, thereby indicating a large anisotropy—unless part of the ellipse contour crosses the boundaries of the map.

The goal of the second processing applied to the list of remaining ellipses is to eliminate redundant voids. Although in the sheetlike topology of the CfA slice most voids have a well-defined edge, a void of average radius a_0 might induce detections at wavelet scales $a \leq a_0$ (see § 3.4). We thus apply the following inclusion criterion to the ellipses: starting with the largest voids, smaller ones are included in the final catalog if the position of their center is not inside a void which already belongs to the catalog. Figure 6 displays all the significant underdense regions selected as real voids by the above treatment. Note that the two small voids around $a \simeq 13^h$ and $V \simeq 7{,}500$ km s^{-1} correspond to the region left artificially empty by the removal of the Coma Cluster from the sample. The ellipses plotted in the wedge diagram of Figure 6 are evidently distorted because they were fitted in the anamorphosed space. A direct fit in the wedge diagram configuration would, however, be biased by the varying density of objects with velocity. The center of all these ellipses corresponds to a wavelet coefficient with a confidence level greater than 99.5%. Table 1 gives the "elliptical" parameters and the density ratio inferred from the angular profile for all ellipses except those two which are artificially created by the removal of Coma, ordered by increasing velocity of the center of the void after restoration of the real geometry of the data. The shape parameters are the distances in the pie diagram map between the extrema of the major (minor) axis obtained in the anamorphosed space. These distances take into account the scaling of the right ascension by the cosine of the average declination of the slice.

Given the poor sampling of the structures for velocites larger than 10,000 km s^{-1}, the edge effects, and the limits of our analysis (isotropic detection, second-order moments description of structures, lack of local reconstruction) this catalog of objective void provides a fairly complete list—from visual inspection—of the voids with diameter larger than about 1000 km s^{-1} in the slice. Similar analyses applied to larger catalogs than the CfA slice should allow to derive a spectrum of void sizes. The requirement for reliable derivation of this spectrum

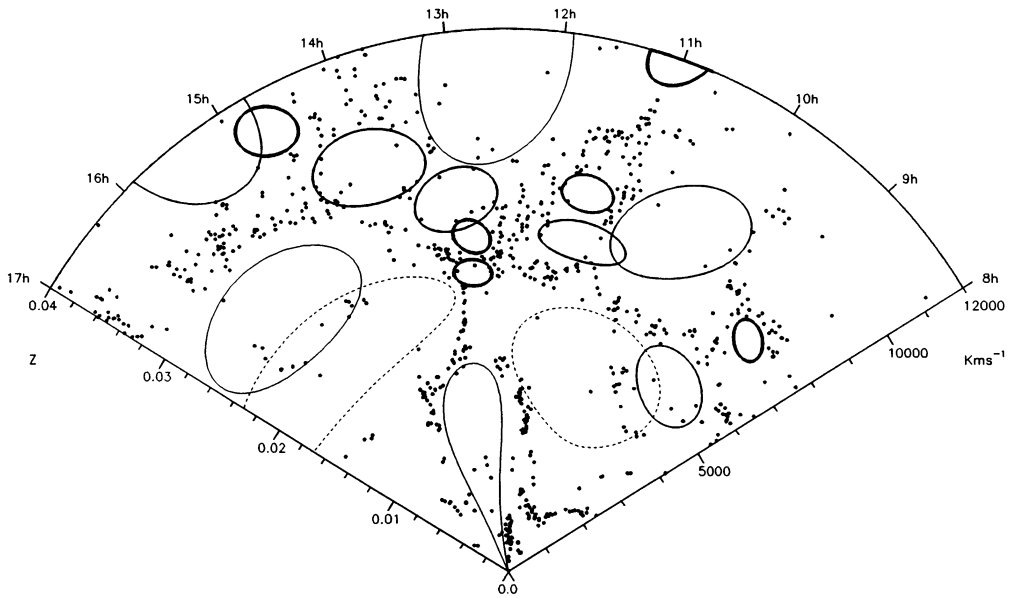

FIG. 6.—Map of the significant voids detected by means of the wavelet technique in the pie diagram of the slice. The ellipses fitted to the significant underdense structures after removal of the artifacts and redundant structures is superposed on the galaxy distribution.

is that the size of the largest voids in the map be significantly smaller than the extent of the map itself. The spectrum of void sizes is of cosmological interest because it puts tight constraints on N-body models for the formation of large-scale structures. In the standard gravitational instability picture, the diameter of the largest existing structures is limited by the upper limits on the cosmic microwave background anisotropy via the Sachs-Wolfe effect (see, for instance, Blumenthal et al. 1991). Given the observational limits (cf. Smoot et al. 1991), the existing cold dark matter models have difficulties generating voids larger than $\sim 55\ h^{-1}$ Mpc in diameter (White et al. 1987). Voids with diameters of $\sim 50\ h^{-1}$ Mpc *are* detected in the CfA slice and detection of larger voids will require well-designed larger surveys (de Lapparent et al. 1991).

6. CONCLUSION

The extension of the Center for Astrophysical Redshift survey redshift has suggested that at large scale galaxies gather along sharp sheetlike structures and delineate vast regions devoid of luminous objects. Here we propose an algorithm

TABLE 1

CATALOG OF ASTROPHYSICAL VOIDS

Number	Scale[a] (km s^{-1})	R.A.[b]	V_r[c] (km s^{-1})	A[d] (km s^{-1})	B[e] (km s^{-1})	Ellipticity[f]	PA[g]	Lack[h]	Ratio[i]
1.....	1400	13^{h}32^m	3035	4670	1165	4.0	170°	−21	*6.17
2.....	2000	10 45	4695	3370	2740	1.2	75	−31	1.45
3.....	1000	9 17	5460	1765	1440	1.2	55	−7	3.61
4.....	2000	15 49	6060	6355	2105	3.0	95	−10	*8.34
5.....	350	8 55	7440	870	710	1.2	35	−4	5.00
6.....	700	11 28	7495	1955	865	2.3	95	−4	2.29
7.....	1400	15 43	7540	3965	2445	1.6	85	−62	1.45
8.....	700	13 06	8290	1660	1585	1.0	145	−6	1.30
9.....	1000	10 25	8340	3020	2145	1.4	145	−16	2.22
10.....	500	11 33	8550	1125	835	1.3	100	−6	0.77
11.....	700	13 57	9405	2355	1820	1.3	60	−12	1.38
12.....	1400	12 37	10960	6220	3205	1.9	170	−8	3.70
13.....	350	14 43	11090	1370	1100	1.2	35	−3	1.67
14.....	1000	15 20	12195	6175	3055	2.0	5	−10	2.12
15.....	350	11 01	12200	2100	1285	1.6	160	−3	5.00

[a] Wavelet scale used for the detection of the voids.
[b] Right ascension of the center of the fitted ellipse.
[c] Recession velocity of the center of the fitted ellipse.
[d] Major axis of the fitted ellipse.
[e] Minor axis of the fitted ellipse.
[f] Ellipticity of the fitted ellipse.
[g] Position angle of the major axis of the fitted ellipse increasing from the line of sight to the east.
[h] Deficiency in galaxies with respect to the local background density.
[i] Density ratio inferred from the angular profile. A star indicates that part of the ellipse contour crosses the boundaries of the galaxy map.

based on the wavelet transform to objectively detect and characterize the voids and the high-density regions in the large-scale galaxy distribution. The wavelet transform provides a space-scale analysis in which over- and underdense structures are detected according to their typical size. Applied to bidimensional discrete data like galaxy catalogs, the wavelet transform yields for each scale a wavelet coefficient image. We then compute for each value of the wavelet coefficient the probability to observe a higher value with respect to a Poissonian process with uniform density using the associated distribution function. We define the significant structures as the connected areas where this probability is greater than the specified detection threshold. High-density regions are identified by a value of the distribution function close to unity while voids correspond to values close to zero. We finally calculate morphological parameters for the detected structures by fitting an elliptical Gaussian model to the local galaxy distribution.

We then apply the wavelet transform to the first slice of the CfA redshift survey projected onto a constant declination surface. Many significant over- and underdense structures are detected using a 99.5% confidence level threshold. The algorithm does detect significant clustering at the location of the Abell clusters other than the Coma Cluster (removed from the sample) included in the slice. We also find good agreement between the significant structures detected at small scale by the wavelet analysis and the groups of galaxies indentified by Ramella et al. (1989). Because we are using an isotropic analyzing wavelet, the large-scale sheetlike structures are detected as alignments of contiguous smaller structures. We also detect the various voids present in the CfA slice as strongly significant structures. We perform a selection based on an inclusion criterion in order to restrict the void hierarchy to the astrophysically interesting structures, and we calculate the elliptical parameters for each void in the resulting catalog. The wavelet analysis allows for the first time to objectively detect and locate the voids in the CfA slice and to replace the often used subjective criteria with quantitative parameters.

This preliminary wavelet analysis of the CfA redshift survey data suggests several directions for further statistical description of the large-scale clustering. First, the isotropic analyzing wavelet used here is not well suited to the detection of structures dynamically elongated along the line of sight like groups and clusters of galaxies, and to the detection of highly asymmetric structures like filaments and walls. Specific processing appears necessary for these objects. We are hence currently investigating the use of an anisotropic wavelet transform, and it appears promising for the objective detection of dynamical clustering from redshift maps. We are also examining the possibilities offered by local reconstruction of the distributions from significant data—using only the significant wavelet coefficients. Both the anisotropic wavelet and the local reconstruction will be important tools for comparing observational data with numerical simulations. The hierarchical clustering of the overdense regions is also clearly exhibited due to the ability of the wavelet transform to decouple structures. One should thus be able to calculate the local fractal dimensions directly (cf. Martinez, Paredes, & Saar 1991), as well as to investigate the relationships between high-order and low-order structures.

Because the wavelet method can disentangle complex hierarchical features, detect significant clustering and voids as well as line-of-sight dynamical elongation, this statistic might provide a complete description of the observed galaxy distribution. In particular, the range of voids sizes can put tight constraint on the theoretical models. Because of the small number of voids contained in the CfA slice (15), it is at the moment premature to attempt a derivation of the spectrum of void sizes from this catalog. However, larger redshift surveys are currently under completion and should be available within the next years: for example the other slices of the CfA survey (see Geller & Huchra 1989) and deep pencil-beam surveys (de Lapparent et al. 1989b; Broadhurst et al. 1990). In those of these surveys for which data is already available, structures qualitatively in agreement with the patterns detected in this first CfA slice are observed. Because they intersect a large number of voids, application of the wavelet analysis to the deep pencil-beam surveys should provide better quantitative constraints on the typical and maximum size for the voids.

REFERENCES

Antoine, J. P., Duval-Destin, M., Murenzi, R., & Piette, B. 1990, in Wavelets and Some of Their Applications, ed. Y. Meyer (Berlin: Springer-Verlag), 181
Argoul, F., Arnéodo, A., Elezgaray, J., Grasseau, G., & Murenzi, R. 1989, Phys. Lett. A, 135, 327
Bahcall, N. 1988, ARA&A, 26, 631
Barrow, J. D., Bhavsar, S. P., & Sonoda, D. H. 1985, MNRAS, 216, 17
Bendjoya, Ph., Slezak, E., & Froeschlé, Cl. 1991, A&A, 251, 312
Bhavsar, S. P., & Ling, E. N. 1988, ApJ, 331, L63
Bijaoui, A. 1989, Lois de Distribution des Coefficients en Ondelettes (Observatoire de Nice)
Bijaoui, A., Slezak, E., Mars, G., & Giudicelli, M. 1989, in 12ème colloque du GRETSI sur le Traitement du Signal et des Images, ed. GRETSI (Sophia-Antipolis) p. 209
Blumenthal, G. R., Da Costa, L. N., Goldwirth, D. S., Lecar, M., & Piran, T. 1992, ApJ, 388, 234
Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, Nature, 311, 517
Bothun, G. D., Geller, M. J., Huchra, J. P., & Schild, R. 1989, BAAS, 21, 1139
Broadhurst, T. J., Ellis, R. S., Koo, D. C., & Szalay, A. S. 1990, Nature, 343, 726
Burns, J. O., Moody, J. W., Brodie, J. P., & Batuski, D. J. 1988, ApJ, 335, 542
Burstein, D., Faber, S. M., & Dressler, A. 1990, ApJ, 354, 18
Burt, P., & Adelson, E. 1983, IEEE Trans., Comm., 31, 532
da Costa, L. N., Pellegrini. P. S., Davis, M., Meiksin, A., Sargent, W. L. W., & Tonry, J. 1991, ApJS, 75, 935
da Costa, L.N., et al. 1988, ApJ, 327, 544
Daubechies, I. 1990, IEEE Trans., Information Theory, 36, 961
Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371
de Lapparent, V., Geller, M. J., & Huchra, J. P. 1986, ApJ, 302, L1
———. 1988, ApJ, 332, 44
de Lapparent, V., Geller, M. J., & Huchra, J. P. 1989a, ApJ, 343, 1
———. 1991, ApJ, 369, 273
de Lapparent, V., Mazure, A., Mathez, G., & Mellier, Y. 1989b, Messenger, 55, 5
Dekel, A., & Rees, M. J. 1987, Nature, 326, 455
Dekel, A., & West, M. J. 1985, ApJ, 288, 411
Dressler, A. 1980, ApJ, 236, 351
———. 1988, ApJ, 329, 519
Dressler, A., & Faber, S. M. 1990, ApJ, 354, 13
Efstathiou, G., Sutherland, W. J., & Maddox, S. J. 1990, Nature, 348, 705
Escalera, E., & Mazure, A. 1992, ApJ, 388, 23
Faber, S. M., & Burnstein, D. 1989, in Large-Scale Motions in the Universe, ed. V. Rubin & G. V. Coyne (Princeton: Princeton Univ. Press), 115
Fry, J. N., Giovanelli, R., Haynes, M. P., Melott, A. L., & Scherrer, R. J. 1989, ApJ, 340, 11
Geller, M. J., & Huchra, J. P. 1989, Science, 246, 897
Gill, A. G., & Henriksen, R. N. 1990, ApJ, 365, L27
Giovanelli, R., Haynes, M., Myers, S. T., & Roth, J. 1986, AJ, 92, 250
Gott, J. R., et al. 1989, ApJ, 340, 625
Goupillaud, P., Grossmann, A., & Morlet, J. 1984, Géoexploration, 23, 85
Grossmann, A., Kronland-Martinet, R., & Morlet, J. 1988, in Wavelets, Time-Frequency Methods and Phase Space, ed. J. M. Combes, A. Grossmann, & Ph. Tchamitchian (Berlin: Springer-Verlag), 1
Grossmann, A., & Morlet, J. 1985, in Mathematics + Physics, Lectures on Recent Results, ed. L. Streit (Singapore: World Scientific), 1
Gurbatov, S. N., Saichev, A. I., & Shandarin, S. F. 1989, MNRAS, 236, 385
Hogan, C. J. 1990, Nature, 344, 107
Holschneider, M. 1988, J. Stat. Phys., 50, 963
Huchra, J. P., Davis, M., Latham, D., & Tonry, J. 1983, ApJS, 52, 89

Huchra, J. P., Geller, M. J., de Lapparent, V., & Corwin, H. G. 1990, ApJS, 72, 433
Jôeveer, M., & Einasto, J. 1978, in IAU Symp. 79, The Large-Scale Structure of the Universe, ed. M. S. Longair & J. Einasto (Dordrecht: Reidel), 241
Jones, B., Martinez, V. J., Saar, E., & Einasto, J. 1988, ApJ, 322, L1
Kaiser, N., & Lahav, O. 1989, MNRAS, 237, 129
Kauffmann, G., & Fairall, A. P. 1991, MNRAS, 248, 313
Kirshner, R. P., Oemler, A., Schechter, P. L., & Schectman, S. A. 1981, ApJ, 248, L57
Kofman, L. 1989, in Lecture Notes in Physics, 332, Morphological Cosmology, ed. P. Flin & H. W. Duerbeck (Berlin: Springer-Verlag), 354
Lauberts, A. 1982, The ESO-Uppsala Survey of the ESO(b) Atlas (Munich: ESO)
Lilje, P. B. 1990, ApJ, 351, 1
Limber, D. N. 1956, ApJ, 119, 9
Littlewood, J., & Paley, R. 1937, J. London Math. Soc., 42(2), 52
Lynden-Bell, D., Faber, S. M., Burnstein, D., Davies, R. L., Dressler, A., Terlevitch, R. J., & Wegner, A. 1988, ApJ, 326, 19
Maddox, S. J., Efstathiou, G., & Loveday, J. 1988, in IAU Symp. 130, Large-Scale Structure of the Universe, ed. J. Audouze, M. C. Pelletan, & A. Szalay (Dordrecht: Reidel), 151
Maddox, S. J., Efstathiou, G., Sutherland, W. J., & Loveday, J. 1990, MNRAS, 242, 438
Mallat, S. 1989a, IEEE Trans. Pattern Analysis and Machine Intelligence, 11, 674
———. 1989b, IEEE Trans. Acoustic Speech and Signal Processing, 37, 2091
Martinez, V. J., Paredes, S., & Saar, E. 1991, in The Distribution of Matter in the Universe, ed. D. Gerbal & G. Mamon, 2nd DAEC meeting (Observatorie de Paris), 311
Martinez, V. J., Jones, B., Dominguez-Teinero, R., & van de Weygaert, R. 1990, ApJ, 357, 50
Materne, J. 1978, A&A, 63, 401
Maurogordato, S., & Lachièze-Rey, M. 1987, ApJ, 320, 13
Melott, A. L. 1987, MNRAS, 228, 1001
Meyer, Y. 1989, in Wavelets, Time-Frequency Methods and Phase Space, ed. J. M. Combes, A. Grossmann, & Ph. Tchamitchian (Berlin: Springer), 21
Moles, M., del Olma, A., & Perea, J. 1985, A&A, 213, 365
Nilson, P. 1973, Uppsala General Catalogue of Galaxies, Uppsala Astron. Obs. Annaler, 6
Oort, J. H. 1983, ARA&A, 21, 373
Ostriker, J. P., & Cowie, L. L. 1981, ApJ, 243, L127
Park, C. 1990, MNRAS, 242, 59P
Paturel, G. 1979, A&A, 71, 106
Peebles, P. J. E. 1980, The Large-Scale Structure of the Universe (Princeton: Princeton Univ. Press)
———. 1989, Physica D, 38, 273
Pellegrini, P. S., da Costa, L. N., & de Carvalho, R. R. 1989, ApJ, 339, 595
Pellegrini, P. S., Wilmer, C. N. A., da Costa, L. N., & Santiago, B. X. 1990, ApJ, 350, 95
Picard, A. 1991, ApJ, 368, L7
Ramella, M., Geller, M. J., & Huchra, J. P. 1989, ApJ, 344, 57
Rood, H. J. 1988, ARA&A, 26, 245
Ryden, B., Melott, A., Craig, D., Gott, R., Weinberg, R., Scherrer, R., Bhavsar, S., & Miller, J. 1989, ApJ, 340, 647
Saunders, W., et al. 1991, Nature, 349, 32
Slezak, E., Bijaoui, A., & Mars, G. 1990, A&A, 227, 301
Smoot, G. F., et al. 1991, ApJ, 371, L1
Smoot, G. F., et al. 1992, ApJ, 396, L1
Tago, E., Einasto, J., & Saar, E. 1984, MNRAS, 206, 559
Tully, B. 1987, ApJ, 323, 1
Turok, N. 1989, Phys. Rev. Lett., 63, 2625
Vilenkin, A. 1985, Phys. Rep., 121, 263
Vogeley, M. S., Geller, M. J., & Huchra, J. P. 1991, ApJ, 382, 44
Weinberg, D. H., & Gunn, J. E. 1990, MNRAS, 247, 260
West, M. J. 1989, ApJ, 347, 610
White, S. D. M. 1979, MNRAS, 186, 145
White, S. D. M., Frenk, C. S., Davis, M., & Efstathiou, G. 1987, ApJ, 313, 505
Willick, J. A., Brodie, J. P., & Bowyer, S. 1990, ApJ, 355, 393
Yahill, A. 1988, in Large-Scale Motions in the Universe, ed V. C. Rubin & G. V. Coyne (Princeton: Princeton Univ. Press), 219
Yahil, A., Walker, D., & Rowan-Robinson, M. 1986, ApJ, 301, L1
Zabludoff, A. I., Huchra, J. P., & Geller, M. J. 1990, ApJS, 74, 1
Zel'dovich, Y. B., Einasto, J., & Shandarin, S. F. 1982, Nature, 300, 407
Zwicky, F., Wild, P., Herzog, E., Karpowicz, M., & Kowal, C. T. 1961–1968, Catalogue of Galaxies and Clusters of Galaxies (Pasadena: California Institute of Technology), vols. 1–6