THE ASTROPHYSICAL JOURNAL, 473:15–21, 1996 December 10

HIGHER ORDER STATISTICS FROM THE EDINBURGH/DURHAM SOUTHERN GALAXY CATALOGUE SURVEY. I. COUNTS IN CELLS

ISTVÁN SZAPUDI[1]
Fermi National Accelerator Laboratory, Theoretical Astrophysics Group, Batavia, IL 60510

AND

AVERY MEIKSIN[2] AND ROBERT C. NICHOL[3]
University of Chicago, Department of Astronomy and Astrophysics, 5640 South Ellis Avenue, Chicago, IL 60637
Received 1996 March 4; accepted 1996 July 2

ABSTRACT

Counts in cells are used to analyze the higher order properties of the statistics of the Edinburgh/Durham Southern Galaxy Catalogue Survey (EDSGC). The probability distribution is obtained from an equal area projection source catalog with infinite oversampling over the range 0°.015–2°. The factorial moments of the resulting distribution and the s_N's characterizing the non-Gaussian nature of the distribution are extracted. These results are compared to previous results from the Automatic Plate Measuring Facility (APM) survey and to theoretical results from perturbation theory. The deprojected three-dimensional values corresponding to the s_N's are also determined. We find that the three-dimensional values match the scaling relation for strongly nonlinear clustering found in N-body simulations remarkably well, which suggests that the galaxies are reliable tracers of the underlying mass distribution.

Subject headings: cosmology: theory — galaxies: clusters: general — large-scale structure of universe

1. INTRODUCTION

A leading hypothesis for the origin of the large-scale structure of the distribution of galaxies is that it is a consequence of gravitational instability in an initially homogeneous medium. The N-point correlation functions provide a set of statistics suited for quantifying the expected departure from homogeneity of the galaxy distribution under this hypothesis (Peebles 1980). The statistical analyses of recent galaxy catalogs has tended to provide support for this scenario. While the two-point correlation function has clearly demonstrated the non-Poisson character of the galaxy distribution, it is not a unique test of gravitational instability since it is reproduced by a variety of models for structure formation (Peebles 1993). If gravitational instability dominates the growth of structure, however, then it is possible to predict a relation between the higher order correlation functions and the two-point function. In particular, if the structure is hierarchical in nature, as expected in the strongly nonlinear limit, then the N-point functions are symmetrized products of $N-1$ two-point functions (Peebles 1980). In the limit of weakly nonlinear clustering, analytic forms for the amplitudes in analogous relations between spatial averages of the correlation functions have been derived (Juszkiewicz, Bouchet, & Colombi 1993; Colombi, Bouchet, & Schaeffer 1994; Bernardeau 1994a, 1994b).

Angular catalogs offer two advantages over their redshift analogs for measuring higher order correlations: their large size and their insensitivity to redshift distortions. A disadvantage is that because they are projections of the galaxy distribution, simplifying assumptions must be made concerning the clustering of galaxies to make the extraction of the higher order correlations practical. Thus the analyses of both types of catalogs are complementary. Measurements of the higher order correlation functions in angular catalogs have supported the form predicted by the hierarchical model. The amplitudes, however, have shown some variance, depending on the method of analysis and the catalog. Szapudi, Szalay, & Boschán (1992) confirmed and refined the estimate of Groth & Peebles (1977) for the three-point function of the Lick counts (Shane & Wirtanen 1967), although their estimate of the amplitude of the four-point function falls somewhat below that of Fry & Peebles (1978). Szapudi et al. (1992) provide estimates for higher order functions as well. Analyses of the *IRAS* catalogs have provided even stronger support for the hierarchical model, although the correlations of these infrared-selected galaxies tend to be somewhat weaker than those of their optical counterparts, perhaps reflecting a genuine morphology dependence in the nature of clustering (Meiksin, Szapudi, & Szalay 1992; Bouchet et al. 1993). Recently, the analysis of higher order functions has been extended to the Automatic Plate Measuring Facility (APM) catalog (Maddox et al. 1990a, 1990b, 1990c) by Gaztañaga (1994) and Szapudi et al. (1995, hereafter SDES), with generally good agreement with the Lick results of Szapudi et al. (1992), although there are some discrepancies. These may be due to differences in the scales over which these functions are averaged, but the differences between the catalogs or the measurement techniques cannot be precluded as the origin. Systematic variations in the measured magnitudes will induce artificial correlations, while different techniques will exhibit differing degrees of sensitivity to the sources of measurement error (Szapudi & Colombi 1996, hereafter SC96).

In this paper, we present an analysis of the higher order functions in the Edinburgh/Durham Southern Galaxy Catalogue (EDSGC), an angular catalog covering approximately 1000 deg^2 (Heydon-Dumbleton, Collins, & MacGillvray 1989; Collins, Nichol, & Lumsden 1992). We employ an efficient method based on factorial moments of cell counts. The infinite sampling of the catalog (Szapudi 1996) eliminates the measurement errors arising from the use of a finite number of sampling cells (Bouchet, Schaeffer, & Davis 1991; SC96).

[1] szapudi@astro1.fnal.gov.
[2] Edwin Hubble Research Scientist; meiksin@oddjob.uchicago.edu.
[3] nichol@huron.uchicago.edu.

In the next section, we describe the EDSGC catalog, which we follow with an account of the measurement technique in § 3. We present the results of the analysis in § 4 and discuss their relation to previous analyses of other catalogs and to theoretical expectations in § 5.

2. THE EDINBURGH/DURHAM SOUTHERN GALAXY CATALOG (EDSGC)

The Edinburgh/Durham Southern Galaxy Catalogue (EDSGC) is a catalog of 1.5 million galaxies covering $\simeq 1000$ deg^2 centered on the south Galactic pole (SGP). The database was constructed from COSMOS scans (a microdensitometer) of 60 adjacent UK IIIa-J Schmidt photographic plates and reaches a limiting magnitude of $b_j = 20.5$. The entire catalog has $<10\%$ stellar contamination and is $\gtrsim 95\%$ complete for galaxies brighter than $b_j = 19.5$ (Heydon-Dumbleton et al. 1989). The two-point galaxy angular correlation function measured from the EDSGC has been presented by Collins et al. (1992) and Nichol & Collins (1994).

A rectangular area of the catalog between $\alpha = 22^h$, passing through 0^h–3^h, and declination $-42 \leq \delta \leq -23$, was suitable for our purposes. The original coordinates were converted to physical ones using an equal area projection: $x = (\alpha - \alpha_{\min}) \cos \delta$, $y = \delta - \delta_{\min}$. This simple formula is suitable for the small angular scales considered in this paper. The projection did not affect the declination range, but to obtain a rectangular area, the physical coordinates corresponding to right ascension were restricted to values less than 55°. This resulted in a sample of 2.9×10^5 galaxies and a total effective survey area of 1045 deg^2, or $\simeq 997$ deg^2 after accounting for the cutout regions.

Magnitude cuts were determined by practical considerations. The catalog is complete to about mag 20.3. We adopt a limit half a magnitude brighter for our analysis to be conservative. To permit a direct comparison with results from the APM survey (Gaztañaga 1994), we used the magnitude cut $16.98 \leq m_{\rm EDS} \leq 19.8$. There is an offset in the magnitude scales of the two catalogs (Nichol 1992). Based on matching the surface densities listed in SDES, the magnitude range we have adopted corresponds approximately to the APM magnitude range $17 \leq m_{\rm APM} \leq 20$.

3. THE METHOD OF ANALYSIS

The calculation of the higher order correlation functions consists of a sequence of three consecutive steps: estimation of the probability distribution, calculation of the factorial moments, and extraction of the normalized, averaged amplitudes of the N-point correlation functions. We present the relevant definitions and theory below.

Let P_N denote the probability that a cell contains N galaxies, with implicit dependence on the cell size ℓ. The best estimator for P_N from the catalog is the probability that a randomly thrown cell *in the catalog* contains N galaxies (excluding edge effects, which are negligible for the scales in the present study, except perhaps on the largest scales as a result of the holes cut out around bright stars). This may be either calculated from the configuration of the points using a computer algorithm (see Szapudi 1996) or estimated by actually throwing cells at random,

$$\tilde{P}_N = \sum_{i=1}^{C} \delta(N_i = N) , \tag{1}$$

where C is the number of cells thrown and N_i is the number of galaxies in cell i. It is desirable to use as many cells as possible, since for large C, the errors behave as (SC96)

$$E^{C,V} = \left(1 - \frac{1}{C}\right) E^{\infty,V} + E^{C,\infty} , \tag{2}$$

where the $E^{C,V}$ is the total theoretical error (not including the systematic errors of the catalog), $E^{\infty,V}$ is the "cosmic" error associated with the finiteness of the catalog, and $E^{C,\infty}$ is the error associated with the finite number of cells used for the estimator. Since $E^{C,\infty} \propto C^{-1}$ (SC96), the lowest possible error is obtained for $C \to \infty$. We employed such a code on scales up to 2°.

The factorial moments (see, e.g., Szapudi & Szalay 1993) may be obtained from the probability distribution using

$$F_k = \sum P_N (N)_k , \tag{3}$$

where $(N)_k = N(N - 1) \ldots (N - k + 1)$ is the kth falling factorial of N. The F_k's directly estimate the moments of the underlying continuum random field, which is Poisson sampled by the galaxies. This is equivalent to the ordinary moments after shot noise subtraction as can be seen from the relation with ordinary moments

$$\langle N^m \rangle = \sum_{k=0}^{m} S(m, k) F_k , \tag{4}$$

where $S(m, k)$ are the Stirling numbers of the second kind. The use of factorial moments simplifies all the expressions, since sums weighted by the Stirling numbers (shot noise) are eliminated. For instance, the factorial moments of a Poisson distribution are $F_k = \langle N \rangle^k$. These could have been obtained from a constant probability density $\delta(\epsilon - \langle N \rangle)$, which is the underlying continuum process. The ordinary moments of the Poisson distribution, however, will be more complicated, containing "Poisson noise" from the previous equation. It is worthwhile to note that we implicitly assume infinitesimal Poisson sampling throughout this paper. It is the most widely accepted assumption, although it cannot account for certain distributions, such as ones derived from collisions of hard spheres.

The average of the N-point angular correlation functions on a scale ℓ is defined by

$$\bar{\omega}_N(\ell) = A(\ell)^{-N} \int d^2 r_1 \ldots d^2 r_N \, \omega_N(r_1, \ldots, r_N) , \tag{5}$$

where ω_N is the N-point correlation function in the two-dimensional survey, and $A(\ell)$ is the area of a square cell of size ℓ. We define s_N in the usual way,

$$s_N = \frac{\bar{\omega}_N}{\bar{\omega}_2^{N-1}} . \tag{6}$$

This definition is motivated by the assumed scale invariance of the N-point correlation functions in the strongly nonlinear limit (Balian & Schaeffer 1989),

$$\omega_N(\lambda r_1, \ldots, \lambda r_N) = \lambda^{-(\gamma-1)(N-1)} \omega_N(r_1, \ldots, r_N) , \tag{7}$$

where γ is the slope of the three-dimensional two-point function. The coefficients also quantify the deviation from Gaussian statistics, like skewness ($N = 3$) and kurtosis ($N = 4$). Derivations of the coefficients from perturbation theory have recently been performed in the weakly nonlinear limit for three dimensions by Juszkiewicz et al. (1993)

and Bernardeau (1994a, 1994b), and for two dimensions by Bernardeau (1995).

The factorial moments have an especially simple relation to the s_N's through the recursion relation (Szapudi & Szalay 1993),

$$s_k = \frac{F_k \bar{\omega}_2}{N_c^k} - \frac{1}{k} \sum_{q=1}^{k-1} \frac{(k-q) s_{k-q} F_q^{(k)}}{N_c^q} , \quad (8)$$

where $N_c = \langle N \rangle \bar{\omega}_2$. Note that although the notation indicates a projected catalog, there are corresponding expressions for three dimensions.

The deprojection of the s_N's to their three-dimensional counterparts has an intrinsic limitation due to the finite sizes of the cells. While the deprojection of any individual tree structure is well defined, care must be taken in interpreting the deprojected values of the cell-count–determined s_N's, since these implicitly contain an averaging over trees within each cell (see SDES for a discussion). On small scales, where clustering is strongly nonlinear, the coefficients deproject to the three-dimensional coefficients S_N defined by $S_N = \bar{\xi}_N / \bar{\xi}_2^{N-1}$, where the hierarchical assumption may be presumed valid. In this case,

$$s_N = R_N S_N , \quad (9)$$

where S_N is the corresponding three-dimensional value for the spherically averaged $\bar{\xi}_N$'s. The projection coefficients R_N's are fairly insensitive to slight variations of the magnitude cut (see Table 2 in SDES), and the shape dependence is neglected according to the findings of Boschán, Szapudi, & Szalay (1994). We adopt the R_N's of SDES with a Hubble constant of $H_0 = 100$ km s^{-1} Mpc^{-1}. In the intermediate range of weakly nonlinear clustering, hierarchy-breaking terms become significant, and the differences between the conical averaging of the projected correlation functions and the spherical averaging of the three-dimensional functions become large (Bernardeau 1995). In this limit, the s_N's deproject according to

$$s_N = R_N \Sigma_N , \quad (10)$$

where the Σ_N's involve averages only over the orthogonal parts of the wavevectors. (The expressions for R_N are identical in eqs. [9] and [10] for power-law power spectra.) Expressions for Σ_N for pure power-law power spectra have been worked out by Bernardeau (1995). For the depth of the EDSGC, the weakly nonlinear region corresponds to separations of $\theta \gtrsim 1°$ (see § 4).

4. RESULTS

We measured counts in cells by calculating the results corresponding to an infinite number of square cells, placed according to the algorithm of Szapudi (1996), with sizes in the range 0°.015125–2° (corresponding to 0.1–13 h^{-1} Mpc with $D \simeq 370$ h^{-1} Mpc, the approximate depth of the catalog). The largest scale is limited by the geometry induced by the cutout holes: the number of available cells would be severely limited for a measurement on significantly larger scales, since cells intersecting with the cutout holes were rejected. The smallest scale approaches that of galaxy halos for the typical depth of the catalog. Note that even at the smallest scale, where the average count is only 0.0645 per cell, the s_N's are measured to high accuracy because of the infinite oversampling and the efficient Poisson subtraction through the use of factorial moments. By comparison, the practice common in the literature is to stop at scales 4 times that at which the Poisson noise starts to dominate, i.e., where the average count approaches unity. Note that this method extracts almost all the information available through cell counts, except that we did not sample different orientations of the cells, which in principle could have a slight effect. However, studying different orientations properly would most likely involve cutting off more of the existing catalog to prevent potential weighting problems. In practice this could even diminish the available information by enhancing cosmic errors; the thorough examination of this effect is left for future work.

The results of both infinite and low sampling measurements for P_N are displayed in Figure 1. The low sampling corresponds to covering the area with cells once only, i.e., the number of sampling cells is $C_V = V/v$, where V is the volume of the survey and v is the volume of the sampling cell at the given scale. As proved in SC96, the "number of statistically independent cells," C^*, depends strongly both on scales and on the aims of the measurement, but for higher order statistics, it is generally much larger than C_V. Therefore, to minimize the errors as much as possible, we used infinite oversampling for all measurements in this paper. A comparison of the two curves shows the substantial improvement in accuracy achieved through oversampling. Note that covering the area fully with C_V number of cells is not fully equivalent to throwing the same number of random cells, because these might overlap, thus more effectively sampling clusters and in principle decreasing the bias toward low values visible in Figure 2.

Figure 2 shows the scale dependence of the s_N's determined from the counts in cells. The solid line corresponds to the measurements of the entire survey area with high oversampling. The dotted line is the same measurement with undersampling. For the error determination, we divided the survey into four equal parts, similar to the approach of

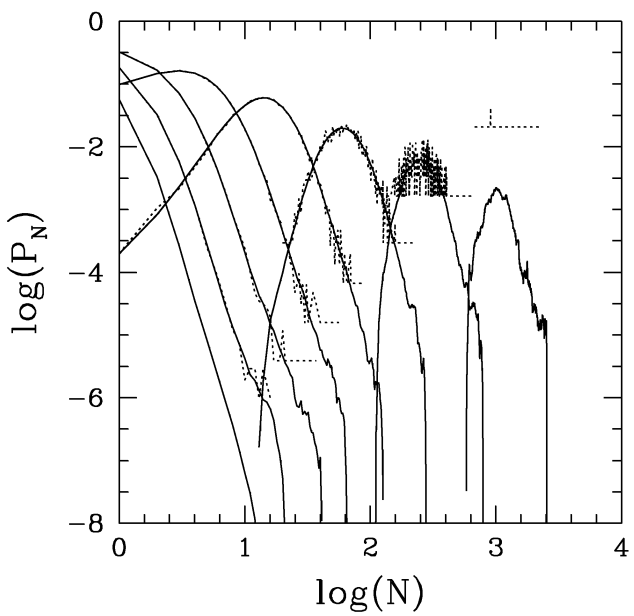

FIG. 1.—Distributions P_N of counts in cells measured in the EDSGC catalog. The solid line corresponds to infinite sampling, while the dotted line corresponds to severe undersampling. The curves from left to right correspond to cell sizes from 0°.015125 doubling up to 2°. Exhaustive sampling is essential on all scales.

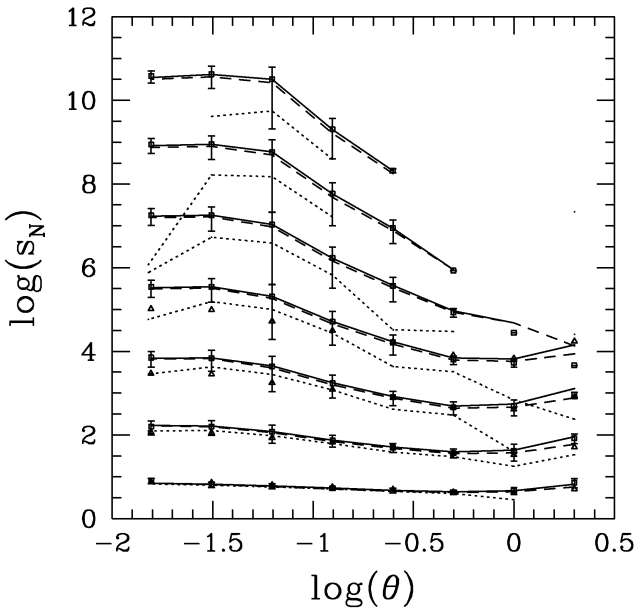

FIG. 2.—Solid line is the measurement of the values of s_N over the entire survey area with infinite sampling; the dotted line is the same with low sampling. Undersampling results in a systematic underestimate of the coefficients. The squares show the mean of the measurements (see text for details) in four equal parts of the survey, and the errors are calculated from the dispersion. The misalignment of the squares and the solid line at the largest scales may be a result of edge effects. The triangles show the values of s_N corresponding to the best-fitting formal $n_{\rm eff}$ (see text).

Gaztañaga (1994); this procedure can overestimate the cosmic error, because subcatalogs have smaller area, but it could also lead to underestimation, because the subcatalogs are not independent volumes (SC96). The squares show the mean of the measurements taken in the following way: to avoid the bias introduced by the fact that the s_N's are nonlinear functions of the factorial moments, the mean of the moments was taken first, and then the cumulants were calculated. The error bars, estimated by a determination of the dispersion of the (possibly biased) s_N's calculated from the factorial moments from each zone, are shown only for those points for which there was sufficient physically valid data permitting a determination. Note that, as mentioned above, the results of this crude estimate must be taken with caution, because at certain scales it can either under- or overestimate the true error bars, and the error distribution is non-Gaussian (SC96). On large scales, the squares deviate from the solid line: this is presumably a result of edge effects. For s_3 and s_4, the errors range over 8%–36% and 19%–56%, respectively. These may be compared with theoretical estimates for the errors. We base the estimates on the errors of the correlator moments over the entire catalog, according to SC96. For the first four moments, respectively, the errors are, ranging from large scales to small, 3%–2%, 7%–12%, 13%–45%, and 23%–63%. Although there is no simple formula relating the errors of the moments to the errors of the s_N's, it is likely that the errors at each order are dominated by the largest error; i.e., the highest moment. Thus, unless some cancellation effects are present, the last two values should well represent the errors on s_3 and s_4. These compare well with the empirical errors from the dispersion above. Possible systematics were also checked for; perturbing the magnitudes of the galaxies with the measured magnitude zero-point distribution yields virtually identical results.

Figure 2 exhibits two plateaus, one at small scales ($<0.^\circ03$) and a second at large scales ($>0.^\circ5$). The large-scale plateau is approaching the width of the survey and so may merely reflect edge effects. The plateau at small separations, however, may indicate that the strongly nonlinear clustering limit has been reached, in which case the hierarchical form for the angular correlations should apply, for which the coefficients appear to converge. The values of s_N are provided in Table 1, and the ratios s_N/R_N in Table 2.

In order to probe more deeply into the weakly nonlinear regime, we performed a separate analysis extending to 4°.

TABLE 1

VALUES OF s_N[a]

θ (deg)	s_3	s_4	s_5	s_6	s_7	s_8	s_9
2	6.60	90.0	1.29(3)	1.42(4)	...	...	...
1	4.65	43.7	546	6.60(3)	4.77(4)	...	...
0.5	4.37	39.1	492	6.93(3)	9.29(4)	8.97(5)	...
0.25	4.76	51.2	832	1.67(4)	3.76(5)	8.83(6)	2.06(8)
0.125.....	5.37	75.6	1.76(3)	5.18(4)	1.70(6)	5.80(7)	2.01(9)
0.0625	6.04	121	4.44(3)	2.08(5)	1.08(7)	5.84(8)	3.21(10)
0.03125	6.58	161	6.99(3)	3.50(5)	1.79(7)	8.89(8)	4.18(10)
0.015125.....	7.05	168	6.81(3)	3.32(5)	1.70(7)	8.24(8)	3.53(10)

[a] Powers of 10 are denoted by numbers in parentheses.

TABLE 2

VALUES OF s_N/R_N[a]

r (Mpc)	s_3/R_3	s_4/R_4	s_5/R_5	s_6/R_6	s_7/R_7	s_8/R_8	s_9/R_9
13.0.....	5.68	62.7	707	5.98(3)	...	...	...
6.5	4.01	30.4	298	2.76(3)	1.51(4)	...	...
3.2	3.76	27.3	269	2.90(3)	2.94(4)	2.12(5)	...
1.6	4.10	35.7	455	7.02(3)	1.19(5)	2.09(6)	3.62(7)
0.81.....	4.63	52.7	962	2.17(4)	5.37(5)	1.37(7)	3.53(8)
0.41.....	5.20	84.5	2.43(3)	8.70(4)	3.41(6)	1.38(8)	5.64(9)
0.20.....	5.67	113	3.82(3)	1.47(5)	5.67(6)	2.11(8)	7.36(9)
0.10.....	6.07	117	3.72(3)	1.39(5)	5.37(6)	1.95(8)	6.21(9)

[a] Powers of 10 are denoted by numbers in parentheses.

On these scales, the majority of cells overlaps with some of the cutout regions; therefore, the analysis had to be done without the elimination of such cells; otherwise, edge effects and cosmic errors from the resulting small area would have severely affected the measurement. After reanalyzing all scales without eliminating cells containing the cutout holes, we found the effect of the holes is to bias the measurements to slightly low values (Fig. 2), but by an amount that is well within the statistical errors. We nonetheless find good agreement with the smaller scale analysis for $\theta \leq 1°$. We obtain in the larger scale analysis $s_3 = 5.75$ and $s_4 = 60$ at $\theta = 2°$, and $s_3 = 7.9$ and $s_4 = 71$ at $\theta = 4°$. These correspond to $\Sigma_3 = 4.95$ and $\Sigma_4 = 42$ at 13 h^{-1} Mpc separation, and $\Sigma_3 = 6.8$ and $\Sigma_4 = 50$ at 26 h^{-1} Mpc separation (although these angular scales are outside the range of strict applicability of the theory for the Σ_N's; Bernardeau 1995). The errors on these measurements could be as much as 30% and 50%, respectively.

5. DISCUSSION

5.1. *Comparison with the APM Catalog*

Figure 3 compares our results for s_3–s_6 with estimates from the APM catalog kindly provided by E. Gaztañaga. The heavy solid line extending to the smallest scales is our measurement for the EDSGC catalog, the dotted lines are the measurements for the full APM catalog, and the light solid line is the measurement of a subregion of the APM that corresponds to the EDSGC. Between scales of about 0°.2–2°, the agreement is good between the full EDSGC and the same region of the APM. It seems that the increase of the S_N's at the largest scales is due to edge effects: a similar phenomenon appears in the full APM at larger scales. This figure again shows that the error bars (which are obtained from the APM using a similar zoning procedure) do not necessarily reflect the true errors: the EDSGC zone of the APM lies about 2 σ outside of the full APM measurements at the same scales. As was mentioned above, the reason for this could be both that the estimation of the dispersion by dividing into four subcatalogs is not accurate and that the errors are probably distributed in a non-Gaussian fashion. At scales smaller than 0°.2, the APM measurements seem to be systematically low. The reason for this could be insufficient sampling in the APM estimates: since the APM measurement was derived from a density map at the lowest scale shown at the figure (E. Gaztañaga 1996, private communication), only minimal sampling and 4 times oversampling was used for the two leftmost points because this is the most possible with shifting the grid at these scales. As shown in SC96, it is most important to have high oversampling at small scales; therefore, low sampling could have introduced measurement errors. Insufficient sampling in principle does not cause a low bias, i.e., the mean is always recovered as an average over many low-sampling measurements. However, if the error distribution is skewed, a realization is likely to undershoot the mean, which is compensated by a few larger overshoots in the ensemble average sense. Physically, this corresponds to the fact that with low sampling, it is not likely to have cells that fully cover small, dense clusters. Since these clusters dominate the higher order statistics, insufficient sampling usually results in an underestimate and only rarely in an overestimate. This is a possible explanation both for the effect shown in Figure 2 (*dotted line*) and for the systematic deviation between the APM and the EDSGC in Figure 3.

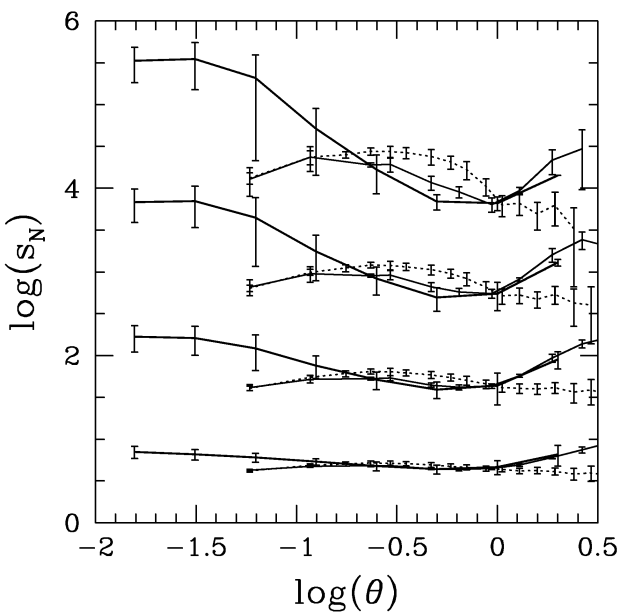

FIG. 3.—Comparison with the results from the APM survey for $s_3 \ldots s_6$. The heavy solid line is the measurement in the full EDSGC survey, as in Fig. 2. The dotted line is the full APM measurement by Gaztañaga (1994), while the light solid line is the measurement in the APM catalog by E. Gaztañaga (1996, private communication) for the region on the sky corresponding to the EDSGC. There is excellent agreement except for the smallest scales, possibly caused by insufficient sampling in the APM measurements.

5.2. *Comparison with Theory*

At small (nonlinear) scales, the hierarchical tree model (described below) is believed to be a good approximation to the clustering. At larger (weakly nonlinear) scales, perturbation theory of gravitating matter starting from Gaussian initial conditions provides a basis for comparison. In a projected catalog, the transition between scales is somewhat uncertain, since a length scale is assigned to angular scales using the depth of the catalog. This procedure is physically correct although somewhat arbitrary, and there could be effects associated with mixing of different scales in the selection function–weighted cone corresponding to a cell in an angular catalog. While no existing measurement has clearly demonstrated the validity of either of the above models, the results based on moment correlators seem to support the hierarchical model, at least on small scales (Szapudi et al. 1992 ; Meiksin et al. 1992; SDES), as does the present work. In what follows, a direct comparison is made without taking into account the possibility of biasing: the data are consistent with the galaxies' acting as faithful tracers of the underlying mass distribution.

A plateau at small separations may be expected when the clustering becomes strongly nonlinear. The effect is found, for example, in the N-body experiments for scale-free clustering by Colombi, Bouchet, & Hernquist (1995). If the clustering we measure is strongly nonlinear on the smallest scales, then we are permitted to identify $S_N = s_N/R_N$ in Table 2 at small separations. We may then in turn derive the three-dimensional clustering amplitudes Q_N. These are defined within the hierarchical model

$$\xi_N(r_1, \ldots, r_N) = \sum_{k=1}^{K(N)} Q_{Nk} \sum^{B_{Nk}} \prod^{N-1} \xi(r_{ij}) , \qquad (11)$$

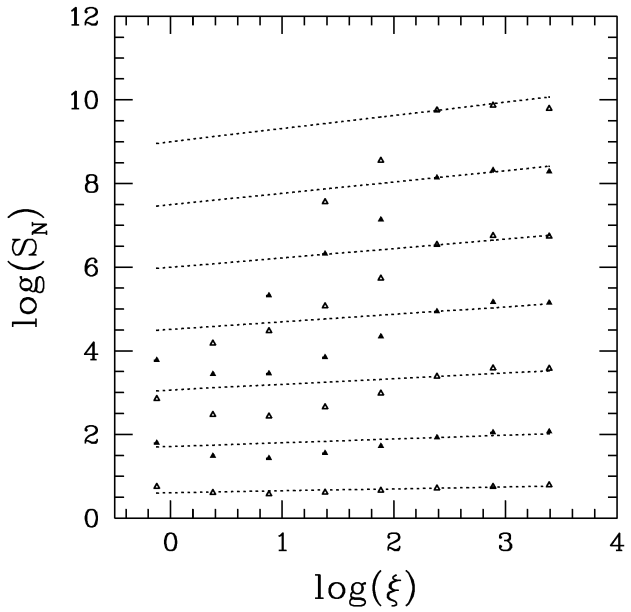

FIG. 4a

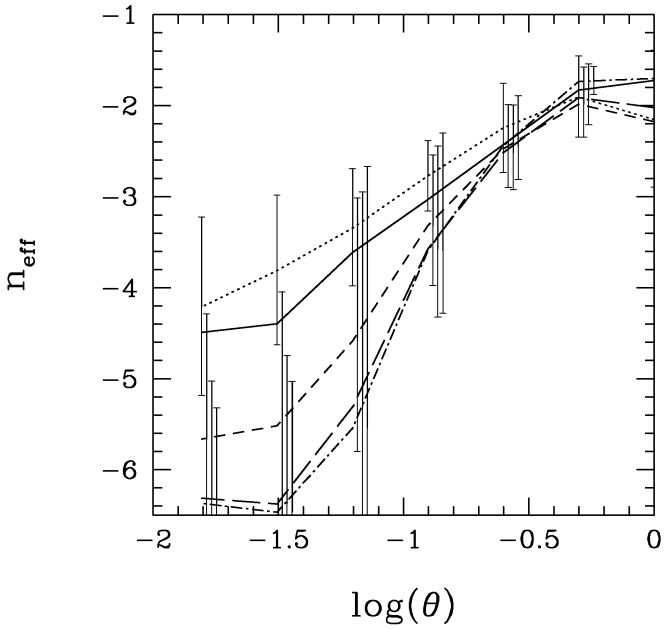

FIG. 4b

FIG. 4.—(*a*) Clustering amplitudes s_N/R_N as a function of the average two-point function $\bar{\xi}_2$. Also shown is the scaling relation of Colombi et al. (1995) found in the strongly nonlinear limit in N-body experiments with scale-free initial conditions. (*b*) Best formal fits for $n_{\rm eff}$ (*solid*), using up to sixth-order quantities. Also shown are the values determined from each N separately, including an indication of the errors based on the upper and lower quartile values for each S_N. Within the errors, the clustering may be described by a single value of $n_{\rm eff}$. Shown are the values of $n_{\rm eff}$ for $N = 3$ (*dotted line*), $N = 4$ (*short-dashed line*), $N = 5$ (*long-dashed line*), and $N = 6$ (*dot-dashed line*).

where $\xi(r) \equiv \xi_2(r) = (r/r_0)^{\gamma}$, as the average of the Q_{Nk}

$$Q_N = \frac{\sum_{k=1}^{K(N)} Q_{Nk} B_{Nk} F_{Nk}}{N^{(N-2)}}, \tag{12}$$

where F_{Nk} are the form factors associated with the shape of cell of size unity (see Boschán et al. 1994 for details)

$$F_{Nk} = \int_1 d^3q_1 \ldots d^3q_N \prod^{N-1} \times \left(|q_i - q_j|^{\gamma} \int_1 d^3p_1\, d^3p_2\, |p_1 - p_2|^{-\gamma} \right)^{-1}. \tag{13}$$

The product above runs over the $N - 1$ edges of a tree. The summation in equation (11) is over all possible N^{N-2} trees with N vertices. In the sum, every $\xi(r_{ij})$ corresponds to an edge $r_{ij} = |r_i - r_j|$ in a tree spanned by $r_1, \ldots, r_N$. For each tree, there is a product of $N - 1$ two-point functions and a summation over all the B_{Nk} labelings of all the $K(N)$ distinct trees. Note that this definition implies $S_N = Q_N N^{N-2}$.

Using the values for $r = 0.1$ Mpc in Table 2, we find for $N = 3$–9, $Q_N = (2.02, 7.3, 30, 108, 320, 745, 1298)$. The values for Q_3 and Q_4 somewhat exceed those found for the Lick-Zwicky sample (Groth & Peebles 1977; Fry & Peebles 1978; Szapudi et al. 1992) and greatly exceed the values found for the CfA1 and SSRS surveys (Gaztañaga 1992). The discrepancy between the larger angular samples and the smaller samples used for the redshift surveys has been previously noted by Fry & Gaztañaga (1994). Our results suggest the discrepancy at small scales may be even larger. The reason for the difference in the values is unknown but may be a result of cosmic variance (Colombi et al. 1994). It appears not to be a result of the added redshift information, since Gaztañaga (1994) found that suppressing the redshifts in the CfA1 and SSRS surveys and treating the samples as angular catalogs had little effect on the values.

In the limit of weakly nonlinear clustering, it is possible to compare the clustering coefficients with theoretical predictions for a given power spectrum (Juszkiewicz et al. 1993; Bernardeau 1994a, 1994b, 1995). If $n_{\rm eff}$ is the local slope of a hypothetical power spectrum that would yield the measured moments in the weakly nonlinear regime, we find from the values of Σ_3 and Σ_4 at separations of 6.5, 13, and 26 h^{-1} Mpc the values $n_{\rm eff} = (-1.2, -1.9, -3.1)$ for $N = 3$, and $n_{\rm eff} = (-1.3, -1.7, -1.9)$ for $N = 4$, using the results of the larger 4° analysis for $\theta > 1°$ from the previous section and the expressions relating Σ_N to n in Bernardeau (1995) in the small angle approximation, which is accurate up to scales 1°. Since 1° corresponds to roughly 6.5 Mpc, clustering is just entering the weakly nonlinear regime, for which theory and measurement may be best compared. The trend of decreasing $n_{\rm eff}$ with increasing scale is suspect. For a power spectrum like CDM, $n_{\rm eff}$ increases with increasing scale. The inverse trend may indicate that edge effects are significant on these scales and are compromising the determination of s_N on scales exceeding 1° or that the theory for Σ_N indeed starts to break down.

While the weakly nonlinear limit should break down on scales smaller than 1°, it is informative to explore the inferred dependence of $n_{\rm eff}$ on scale to smaller values. Colombi et al. (1995) find from N-body experiments for scale-free initial conditions that the values for S_N vary only slightly with scale, increasing for small separations where the clustering becomes strongly nonlinear. They find, independent of n,

$$S_N \simeq [D(\bar{\xi}_2)]^{N-2} \tilde{S}_N, \tag{14}$$

for $N = 3$, 4, and 5, where $D(\bar{\xi}_2) = (\bar{\xi}_2/100)^{0.045}$ and $\tilde{S}_N$ is the value of S_N for $\bar{\xi}_2 = 100$. The relation implies a weak departure from the hierarchical clustering behavior, since the S_N's depend on scale. The dependence is so weak,

however, that the departure is slight. We compare the clustering amplitudes found in the EDSGC with this relation in Figure 4*a*. The agreement in the strong clustering limit is remarkably good, particularly for $N = 3$ and 4. Because we have only angular information, it is not possible to determine whether the deviation from the scaling relation for log $\bar{\xi}_2 < 2$ is a real effect or a consequence of the inherent limitations of extracting three-dimensional information from a projected catalog. It should be noted that the agreement is particularly remarkable since hierarchical clustering is assumed for the underlying distribution in order to convert from the projected correlations to the three-dimensional, while the relation of equation (14) violates this assumption. This suggests that the hierarchical model is a good, though perhaps not perfect, description of the clustering.

Colombi et al. (1995) find that the clustering for $N = 3$, 4, and 5 may be described by a single effective spectral index $n_{\rm eff}$, found from the expressions for weakly nonlinear clustering (Juszkiewicz et al. 1993; Bernardeau 1994a, 1994b). Although the relations between the S_N and n from weakly nonlinear theory do not apply for strong clustering, and even less so in an angular catalog, we may adopt them to obtain a formal value for $n_{\rm eff}$ as done by Colombi et al. (1995, 1996). We do so by fitting s_N/R_N to the expressions for S_N in the limit of weakly nonlinear clustering for $N = 3 \ldots$ using least squares, for $\theta \leq 1°$. The results are shown in Figure 4*b*, including the values derived for each N individually. Within the error estimates, a single value of $n_{\rm eff}$ appears to provide an adequate description of the clustering amplitudes, although the errors are large for small separations. A comparison with N-body results for scale-dependent clustering models, like a CDM-dominated cosmology, could be very illuminating.

I. S. thanks S. Colombi, J. Frieman, and A. Szalay for stimulating discussions. S. Colombi provided the theoretical error estimates in § 4. We are indebted to C. Collins, S. Lumsden, N. Heydon-Dumbleton, and H. MacGillivray for full use of the EDSGC. The authors would like to thank the referee, E. Gaztañaga, for providing his estimates from the APM catalog for comparison, and for useful suggestions, and E. Wright for suggested improvements. Additionally, we would like to acknowledge useful discussions with G. Dalton, J. Loveday, and S. Maddox about properties of the APM survey. I. S. was supported by DOE and NASA through grant NAG 5-2788 at Fermilab. A. M. is grateful to the W. Gaertner Fund at the University of Chicago for support.

REFERENCES

Balian, R., & Schaeffer, R. 1989, A&A, 220, 1
Bernardeau, F. 1994a, ApJ, 433, 1
———. 1994b, A&A, 291, 697
———. 1995, A&A, 301, 309
Boschán, P., Szapudi, I., & Szalay, A. 1994, ApJS, 93, 65
Bouchet, F. R., Schaeffer, R., & Davis, M. 1991, ApJ, 383, 19
Bouchet, F. R., Strauss, M. A., Davis, M., Fisher, K. B., Yahil, A., & Huchra, J. P. 1993, ApJ, 417, 36
Collins, C. A., Nichol, R. C., & Lumsden, S. L. 1992, MNRAS, 254, 295
Colombi, S., Bernardeau, F., Bouchet, F. R., & Hernquist, L. 1996, in preparation
Colombi, S., Bouchet, F. R., & Hernquist, L. 1995, A&A, 281, 301
Colombi, S., Bouchet, F. R., & Schaeffer, R. 1994, A&A, 281, 301
Fry, J. N., & Gaztañaga, E. 1994, ApJ, 425, 1
Fry, J., & Peebles, P. J. E. 1978, ApJ, 221, 19
Gaztañaga, E. 1992, ApJ, 319, L17
———. 1994, MNRAS, 268, 913
Groth, E. J., & Peebles, P. J. E. 1977, ApJ, 217, 385
Heydon-Dumbleton, N. H., Collins, C. A., & MacGillivray, H. T. 1989, MNRAS, 238, 379
Hoaglin, D. C., Mosteller, F., & Tukey, J. W., ed. 1983, Understanding Robust and Exploratory Data Analysis (New York: John Wiley & Sons)
Juszkiewicz, R., Bouchet, F. R., & Colombi, S. 1993, ApJ, 412, L9
Maddox, S. J., Efstathiou, G., Sutherland, W. J., & Loveday, L. 1990a, MNRAS, 242, 43P
Maddox, S. J., Sutherland, W. J., Efstathiou, G., & Loveday, L. 1990b, MNRAS, 243, 692
———. 1990c, MNRAS, 246, 433
Meiksin, A., Szapudi, I., & Szalay, A. 1992, ApJ, 394, 87
Nichol, R. C. 1992, Ph.D thesis, Univ. of Edinburgh
Nichol, R. C., & Collins, C. A. 1994, MNRAS, 265, 867
Peebles, P. J. E. 1980, The Large-Scale Structure of the Universe (Princeton: Princeton Univ. Press)
———. 1993, Principles of Physical Cosmology (Princeton: Princeton Univ. Press)
Shane, C. D., & Wirtanen, C. A. 1967, Pub. Lick. Obs., 22, part 1
Szapudi, I. 1996, in preparation
Szapudi, I., & Colombi, S. 1996, ApJ, 470, 000 (SC96)
Szapudi, I., Dalton, G., Efstathiou, G. P., & Szalay, A. 1995, ApJ, 444, 520 (SDES)
Szapudi, I., & Szalay, A. 1993, ApJ, 408, 43
Szapudi, I., Szalay, A., & Boschán, P. 1992, ApJ, 390, 350