1978AJ.....83

A COMPUTER MODEL UNIVERSE: SIMULATION OF THE NATURE OF THE GALAXY DISTRIBUTION IN THE LICK CATALOG[a)]

RAYMOND M. SONEIRA and P. J. E. PEEBLES
Joseph Henry Laboratories, Physics Department, Princeton University, Princeton, N. J. 08540
Received 23 February 1978

ABSTRACT

We describe the development of a nondynamical computer model universe designed to match the character of the galaxy distribution in the Lick survey. The model assigns "galaxy" positions in a three-dimensional clustering hierarchy, fixes absolute magnitudes, and projects angular positions of objects brighter than $m = 18.9$ onto the sky of an imaginary observer. This yields a galaxy map that can be compared to that of the Lick data. In the model there are 7.5×10^6 galaxies at mean space density 0.065 h^3 Mpc^{-3} ($H = 100$ h km s^{-1} Mpc^{-1}), and 386 000 galaxies are visible at $m \leq 18.9$ and $b \geq 40°$. By adjusting parameters in the model within the limits allowed by the correlation functions to fourth order we have arrived at a galaxy map with a visual appearance that seems a reasonable first approximation to that of the Lick data.

I. INTRODUCTION

a) Goal

The n-point correlation functions for the galaxy distribution (n = 2, 3, and 4; Groth and Peebles 1977, called GP; Fry and Peebles 1978, called FP; and earlier references therein) suggest rather a specific clustering hierarchy pattern (Peebles and Groth 1975). An interesting and important test of this interpretation is to reverse the process and use the hierarchical clustering as a prescription for laying down galaxy positions in a simulated "universe," and then use this distribution to construct galaxy maps whose appearance can be visually compared to the real data. We describe here the results of an experiment at matching the very rich sample provided by the Lick galaxy counts (Shane and Wirtanen, published in condensed form in 1967; reduced and mapped by Seldner *et al.* 1977 hereinafter called SSGP). This is a continuation of the studies of Neyman, Scott, and Shane (1953) and Scott, Shane and Swanson (1954) on the simulation of the Lick data, though we use rather a different model for the spatial galaxy distribution.

This experiment has a peculiar problem arising from the fact that we do not understand how the eye judges the texture and patterns in a map of the sort we are studying. We know the eye does tend to judge in a biased way—for example, one readily picks out "chains" of points in a uniform random distribution. If the statistics in the model map were a close enough approximation to the data such biases would cause no problem because they would apply equally to model and data. However, the statistical model adopted here is at best only a good approximation, now exact, for it does not describe relaxed objects like the Coma cluster and it does not describe the true spatial extent of the clustering of Abell (1958) clusters. Furthermore, the statistical measures we have do not uniquely fix the process by which the galaxies are distributed. The clustering hierarchy picture seems naturally indicated, but it is not the only possibility. And within this model there are many choices of parameters that equally well fit the statistics (cf. Sec. IV*c ii*). The goal is to judge whether a particular method of reproducing the statistics yield a reasonable approximation to the data map. The peculiar problem is to judge whether the eye has been unduly influenced by either the particulars of a distribution or by the inevitable minor deviations from the true statistical situation.

A related point is that since we do not understand how the eye judges texture, we do not have a prescription for how to adjust the model parameters to make the map more accurate. Instead we have resorted to trial and error, and, since each model costs a considerable amount of effort and computation, we have not been able to make a systematic study of all the options.

Despite these problems we consider the final result (Fig. 7 below) pleasing and useful. The visual impression of the map is a "measure," though hard to quantify and certainly not complete. The n-point correlation functions are directly quantified measures, but they also are not complete. The combination may be expected to yield a fuller picture than does either separately, and indeed as will be described in Sec. V this is the case.

b) The Organization of the Paper

In a previous paper (Soneira and Peebles 1977, called SP) we described a computer model to compare to the distribution of the Zwicky (1961–1968) galaxies at $m \leq 14$. The model was designed (and adjusted) to match

[a)] This research was supported in part by the National Science Foundation.

 0004-6256/78/8307-0845$00.90 © 1978 Am. Astron. Soc.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1978AJ.....83

the galaxy two-point correlation function and the frequency distribution of nearest neighbor distances, the latter providing some measure of the higher order correlation functions: the resulting galaxy map closely matches the appearance of the Zwicky map (SP Fig. 11). Our first attempt to match the Lick map simply was to extend the magnitude limit (and the depth of space within which the model galaxies are placed). With this richer sample it became quite clear that the model could be improved. In part this is a result of the Local Supercluster; the galaxies at $m \leq 14$ are not a fair sample of the Lick universe. Because it is interesting to see how well a model tuned to the data at $m \leq 14$ does when extrapolated to $m \leq 19$, we describe the results of this experiment in Sec. IV. *a* below. We briefly list in this section our attempts at improvement (Sec. IV*b*) and then describe the final model (Sec. IV*c*).

Though the process of improving the model involved considerable trial and error (~150 trials in all), we did come to realize that the appearance of the final map depends on two apparently minor aspects of the match to the correlation functions, the variation of the three-point function with elongation and the roll-off of the two-point function at large separation. In Sec. II we explain the general method of construction of the model and the role of these two aspects. In Sec. III we list a few details of the model—the luminosity function, the geometry of the space sample, and some shortcuts in the computation. Our main conclusions are listed in Sec. V.

II. NATURE OF THE MODEL

a) General Method of Construction

We review here the method of laying down the model galaxy distribution. For further details see SP and Peebles (1978).

The model commences with a clump of galaxies constructed as follows. Within a sphere of radius R ($R \gtrsim 10$ h^{-1} Mpc, $H = 100$ h km s^{-1} Mpc^{-1}) one places η spheres, each of radius R/λ; within each of these one places η spheres of radius R/λ^2; and so on through L levels to η^L positions in η^{L-1} spheres each of radius

$$r_0 = R/\lambda^{L-1}. \tag{1}$$

A galaxy is placed at each of these positions. These clumps are distributed uniformly at random in the model universe space. The galaxies are assigned absolute magnitudes drawn at random from a luminosity function, and then the angular positions of the galaxies brighter than m in the sky of an imaginary observer are used to make a map.

In the SP model, $m \leq 14$, the η subsphere centers are placed at random within the spheres at each level, and η is a random variable (Part d below). In the final model adopted here η is fixed at $\eta \equiv 2$, the number of levels L in the clump is a random variable (Part *e* below), and the computation is streamlined (and the map perhaps slightly improved) by placing the two subspheres at fixed separation, R/λ^n, in a random orientation, and centered on the sphere.

One way to picture the space distribution of galaxies in the model is to imagine viewing it with resolution r ($r_0 < r < R$, where r_0 is defined in Eq. (1), and also $r \lesssim$ typical distance between clumps), so that the structure scales $< r$ is washed out. One would see that galaxies are concentrated in clusters of size $\sim r$. These clusters correspond roughly to the lth level of the clump (counting from the smallest scale), where

$$r \sim r_0 \lambda^l, \tag{2}$$

so each contains

$$N_l \sim n^l, \tag{3}$$

galaxies. The mean density within the clusters is then

$$n_r \sim \eta^l r^{-3} \sim \eta^{\log(r/r_0)/\log\lambda} r^{-3},$$
$$\propto r^{-\lambda}, \qquad r_0 \ll r \ll R,$$
$$\lambda = \eta^{1/3-\gamma}. \tag{4}$$

That is, the density varies as a power of the cluster size, with index γ determined by λ. If η is a random variable, in the last of Eqs. (4) η is replaced with $\langle\eta\rangle$ (cf. Appendix A).

This distribution is a clustering hierarchy very similar to that discussed by de Vaucouleurs (1970). One case discussed by de Vaucouleurs is equivalent to a single clump infinite in extent ($R = \infty$) with a characteristic density run given by Eq. (4), so there does not exist a mean galaxy number density. To match the Lick sample we need many rather sharply truncated clumps (cf. f below) randomly distributed in the survey region.

The above construction reproduces the features of the n-point correlation functions. The two-point function is defined by the probability of finding a galaxy centered in the volume element dV at distance r from a randomly chosen galaxy,

$$dP = \langle n\rangle\,[1 + \xi(r)]dV. \tag{5}$$

The term $\langle n\rangle\,dV$ is matched by the chance of finding a neighbor from another randomly placed clump, $\langle n\rangle$ being the mean galaxy density. Thus ξ is determined by the density within the clump [Eq. (4)],

$$\langle n\rangle\,\xi(r) \approx n_r \propto r^{-\gamma}, \qquad r_0 \ll r \ll R. \tag{6}$$

This reproduces the observed power law behavior of ξ.

The three-point correlation function $\zeta(r_1,r_2,r_3)$ is defined by the chance of finding two galaxies separated by r_3 and at distances r_1 and r_2 from a randomly chosen galaxy, all in the same clump (Peebles and Groth 1975) so if $r_1 \ll r_2$,

$$\langle n\rangle^2\zeta \sim n_{r_1} n_{r_2},$$
$$\zeta \sim \xi_1\xi_2. \tag{7}$$

(For further details see Peebles 1978.) The model thus

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1978AJ.....83

at least roughly matches the observation that ζ is given to good accuracy by the expression (GP),

$$\zeta(r_1,r_2,r_3) = Q\,[\xi(r_1)\xi(r_2) + \xi(r_2)\,\xi(r_3) + \xi(r_3)\xi(r_1)], \qquad Q \sim 1.25. \quad (8)$$

(The value of Q depends on the assumed luminosity function and cosmological model. We adjust the model map to fit the observed angular three-point function. With the luminosity function and cosmological model adopted here this makes the spatial Q larger, ~1.6.)

b) *Adjustment to $\mathcal{N}$ and $w(\theta)$*

Having chosen η one can adjust λ to fit the wanted power law index γ of the two-point correlation function [Eqs. (4) and (6)]. There remain as free parameters the mean number of clumps per unit volume and the mean and rms number of galaxies per clump, which can be adjusted to fit the wanted mean density of galaxies $\langle n\rangle$ and the amplitude of ξ. In practice, $\langle n\rangle$ is fixed at the density of galaxies brighter than some absolute magnitude cutoff M_c chosen to exclude from the computation the very faint galaxies that are abundant in space but not in a catalog selected by apparent magnitude. If M_c were then adjusted to double the mean number of galaxies per clump. Provided r_o [eq. (1)] is smaller than the resolution in the map, this has negligible effect on the results (cf. See. IV*b* below).

The parameter M_c provides a convenient way to make fine adjustments in the amplitude of the two-point correlation function. If, in a given model, M_c is changed, it changes the number of visible galaxies from $\mathcal{N}$ to $f\mathcal{N}$, say, but it leaves ξ unchanged, and (when M_c is faint enough that few of the observed galaxies are at $M \approx M_c$) it leaves w unchanged. To bring the density of visible galaxies back to its original value, one must change the number of randomly placed clumps by the factor f^{-1}, and this changes ξ and w to $f\xi$ and fw.

c) *The Parameter Q*

In the simplest model all clumps are constructed in the same way, using the same parameters and different random numbers, and using a fixed number $\eta = 2$ subclusters in each cluster. Here Q [Eq. (8)] is very nearly constant at the value 0.5, which is too small by a factor of about 3. To increase Q, one must arrange to increase the dispersion in the values of the mean density seen within the clusters at resolution r, for that increases the triplets counts. The need for this also is apparent as a lack of highlights in a map constructed from this simplest model (Groth *et al.* 1977, called GPSS). Indeed, it is clear that a considerable dispersion is wanted, for one knows there is considerable spread in mean densities within groups and clusters of size ~ 1 h^{-1} Mpc, say. We discuss next two ways to increase this dispersion in the model.

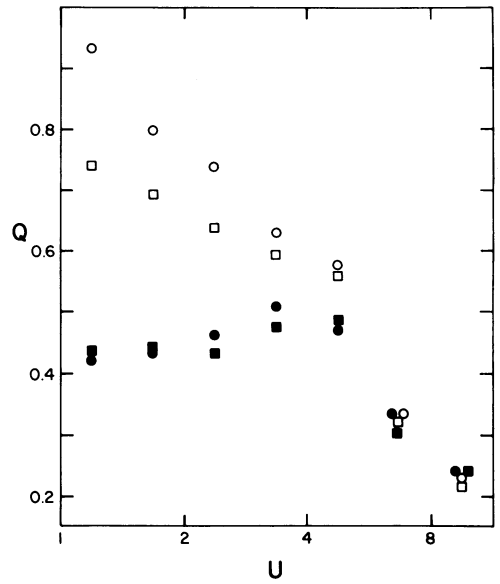

FIG. 1. Estimates of Q from ensembles of clumps constructed in two different ways. The closed symbols correspond to a simple model with $\eta \equiv 2$, the open symbols to a model where η is a random variable drawn from a strongly skewed distribution P_η. In the first case Q is nearly constant, in the second the variation is well outside the bounds of the observations. The circles correspond to $v < 0.5$, the squares to $v > 0.5$.

d) *Dispersion in the Number of Subclusters Per Cluster*

SP increased the dispersion in cluster density by taking η, the number of subspheres placed in each sphere, to be a random variable drawn from a universal distribution P_η. As is described in the Appendix, Q [defined by Eq. (8)] at $r_1 \sim r_2 \sim r_3$ varies roughly as $\langle\eta^3\rangle/\langle\eta^2\rangle^2$ when $\langle\eta\rangle$ is fixed. Thus, if P_η has strong skewness, Q is increased. However, as is discussed in Sec. IV*a*, this does not yield a very pleasing map because with a strongly skew P_η there are occasional large values of η and these show up as prominent isolated spots. Only after observing this did we recognize the connection to a relatively minor but significant discrepancy with the observed three-point function ζ. The occasional large values of η produce large triplets counts at $r_1 \sim r_2 \sim r_3$, hence increasing Q, but they do not so strongly increase the count or the effective value of Q at $r_2 \gg r_1$. The model thus conflicts with the observation that Q is very nearly constant. The effect is illustrated in Fig. 1. We have arranged the arguments of ζ so

$$r_1 < r_2 < r_3, \quad (9)$$

and then defined the new variables

$$r = r_1, \qquad u = r_2/r_1, \qquad v = (r_3 - r_2)/r_1. \quad (10)$$

Q is estimated from ensembles of clumps (100 clumps for the case $\eta \equiv 2$, 400 clumps for the model with variable η where the scatter from clump to clump is larger), each

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1978AJ.....83

with 10 levels, by the method described by Peebles (1978). In the model the minimum cluster radius is $r_0 \sim 0.005$, the maximum is $R \sim 1$, and the Q estimates are for $0.014 < r < 0.020$. The circles correspond to $v < 0.5$, the squares to $v > 0.5$. The closed symbols refer to the simple model $\eta \equiv 2$. The open symbols refer to a model where $\eta = 1$ with probability 0.13, $\eta = 2$ with $P_2 = 0.86$, and $\eta = 15$ with $P = 0.01$. In the first case Q is nearly constant, and the drop at $u \gtrsim 6$ is at least partly because the triangle approaches the boundaries of the clump. In the second case the variation with u and v is well outside the bounds of the observations (GP). Thus the SP method of increasing Q is not satisfactory.

e) Dispersion in the Number of Levels

The adopted method of increasing Q is to let the number of levels L vary from clump to clump while holding fixed the maximum radius R and fixing the number of subclusters per cluster at $\eta \equiv 2$. The effect on the two-, three- and four-point correlation functions is computed as follows. Let ξ_0, ζ_0 and μ_0 be the functions in a model with fixed η, λ, R and L, with $N_0 \gg 1$ galaxies per clump. Now suppose L is changed to a random variable but η, λ, and R are unchanged. The abundance of galaxy pairs, both in the same clump, at separation r, varies as $\langle N_L{}^2 \rangle$, where $N_L \gg 1$ is the number of galaxies in a clump, Therefore the new two-point function is

$$\xi(r) = \xi_0(r)\, \langle N_L{}^2 \rangle / \langle N_L \rangle^2, \qquad r \gg r_0. \tag{11}$$

The shape of ξ is unchanged ($r \gg r_0$) because the number of correlated pairs at each separation changes by the same factor. Since ζ is proportional to the probability of finding two neighbors of a randomly chosen galaxy, all from the same clump (cf. Peebles 1978, Appendix), one finds

$$\zeta = \zeta_0 \langle N_L{}^3 \rangle / \langle N_L \rangle^3,$$

whence by Eqs. (8) and (11)

$$Q = Q_0 \langle N_L{}^3 \rangle \langle N_L \rangle / \langle N_L{}^2 \rangle^2. \tag{12}$$

As before this applies independent of the arguments r, u, and v, if $r \gg r_0$, so Q agrees with the observations in being very nearly independent of u and v. Finally, the four-point function μ varies as

$$\mu / \xi^3 = (\mu_0 / \xi_0{}^3) \langle N_L{}^4 \rangle \langle N_L \rangle^2 / \langle N_L{}^2 \rangle^3. \tag{13}$$

In the Lick data the correlation functions are known so these relations fix the moments $\langle N_L{}^2 \rangle$, $\langle N_L{}^3 \rangle$, and $\langle N_L{}^4 \rangle$ in terms of the other parameters in the model.

f) Introduction of a Sharp Break in the Two-Point Function

In the model with fixed λ, $\xi(r)$ rather gently falls away from the power law at $r \sim R$, and the angular two-point correlation function $w(\theta)$ falls away from the power law even more slowly because at fixed apparent magnitude m one sees clumps at a considerable range of distances. To make w approximate a power law at $\theta \lesssim 2°$ as well as it does in the Lick data we at first simply chose a large clump radius, $R \sim 30\ h^{-1}$ Mpc, so that the break appears at rather a large angle, $\theta \sim 10°$, and at a very small value of w. We at first thought the presence or absence of this extended tail at very small w would have negligible effect on the appearance of the map. That proved wrong, apparently because the tail makes the integral of w large, meaning there are many galaxies per clump, hence relatively few independent clumps. The result is that the map has too many low density "uninteresting" patches. The adopted remedy is to reduce the clump radius to $r \sim 5\ h^{-1}$ Mpc, then eliminate the slow roll-off in $w(\theta)$ by reducing λ in the first sub-clustering level, thereby increasing the number of pairs at separations $\sim 10\ h^{-1}$ Mpc.

The interpretation of the two-point correlation function in the Lick data is somewhat uncertain. In the direct estimates $w(\theta)$ has a kink at $\theta \sim 2.^\circ 5$, and a long positive tail at larger θ. It has been interpreted as a rather sharp dimunition in clustering at $\theta \sim 2.^\circ 5$, $\theta D \sim 10\ h^{-1}$ Mpc, together with a component that varies on scales $\sim 40°$ (GP). This agrees with what seems to be needed for a pleasing map. The large-scale component, whether due to obscuration or to true large-scale clustering, would not affect the appearance of the map because the eye is quite insensitive to slow gradients in density.

III. SOME DETAILS OF THE MODEL

a) Luminosity Function

The adopted differential luminosity function is (SP Eq. 7)

$$\begin{aligned} dN/dM &= 0, && M > M_c; \\ &= \beta C\, dex\, \beta M, && M_c > M > M^*; \\ &= \alpha C\, dex\, \alpha M, && M^* - 2 < M < M^*: \\ &= 0, && M < M^* - 2; \\ \alpha &= 0.75, \qquad \beta = 0.25, && M^* = -18.6 + 5 \log h. \end{aligned} \tag{14}$$

The Hubble parameter h scales out of the model.

The faint end cutoff is $M_c \approx M^* + 3$. This causes all galaxies closer than $D_c \sim 75\ h^{-1}$ MPC to be visible, which is comfortably small compared to the typical distance, $\sim 230\ h^{-1}$ Mpc, in the model. Then with the chosen geometry one galaxy is visible for every ~ 188 placed in space. If M_c were increased to ∞, with fixed slope β and fixed amplitude C, the number of visible galaxies would increase by 3%.

The adopted expression for the distance modulus is

$$m - M = 5 \log D + 25 + (3 + 5/\ln 10) HD/c. \tag{15}$$

D is the distance in megaparsecs, computed in the flat

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1978AJ.....83

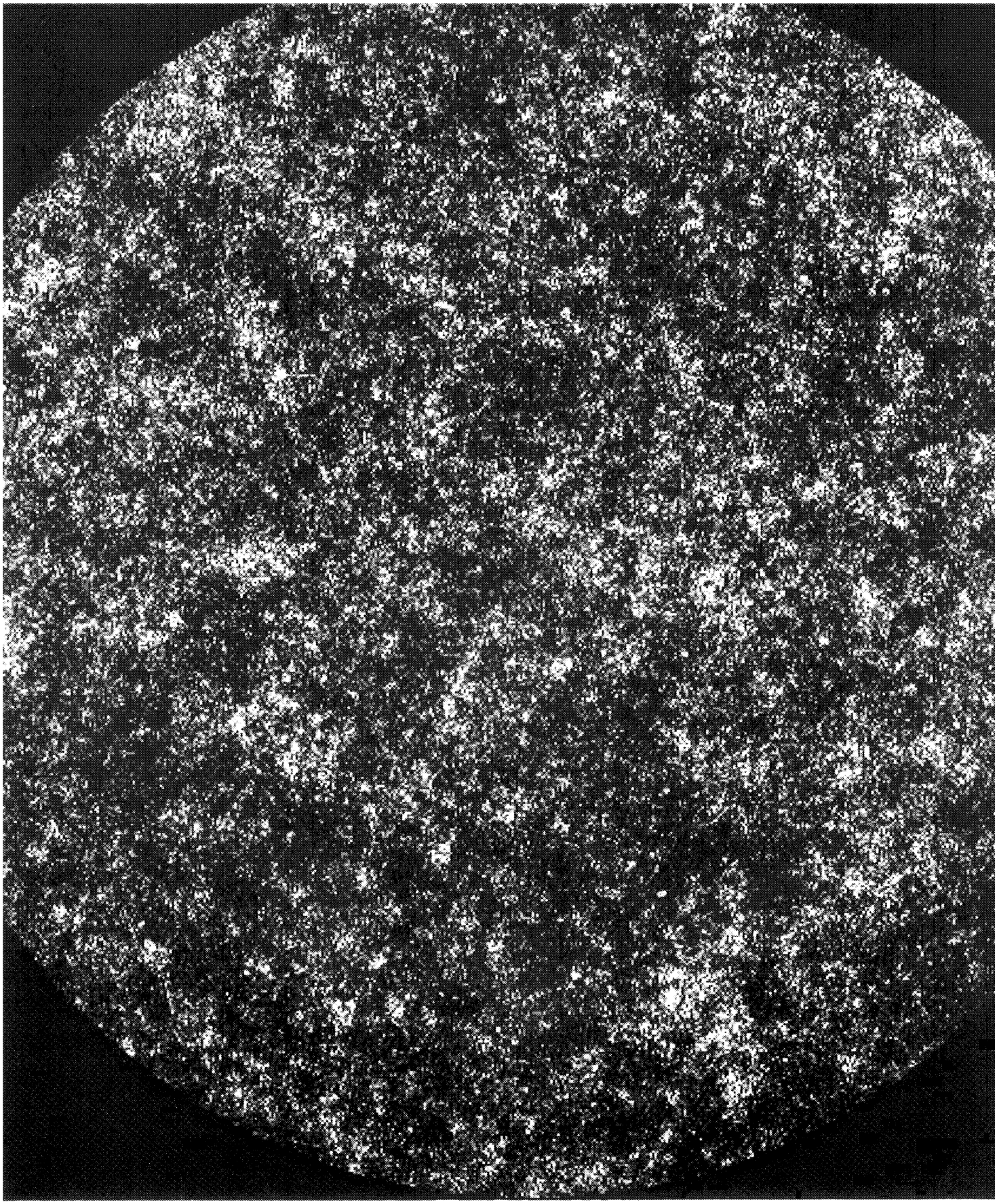

FIG. 2. Galaxy map for the extrapolated $m \leq 14$ model as in Fig. 3.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1978AJ.....83

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1978AJ.....83

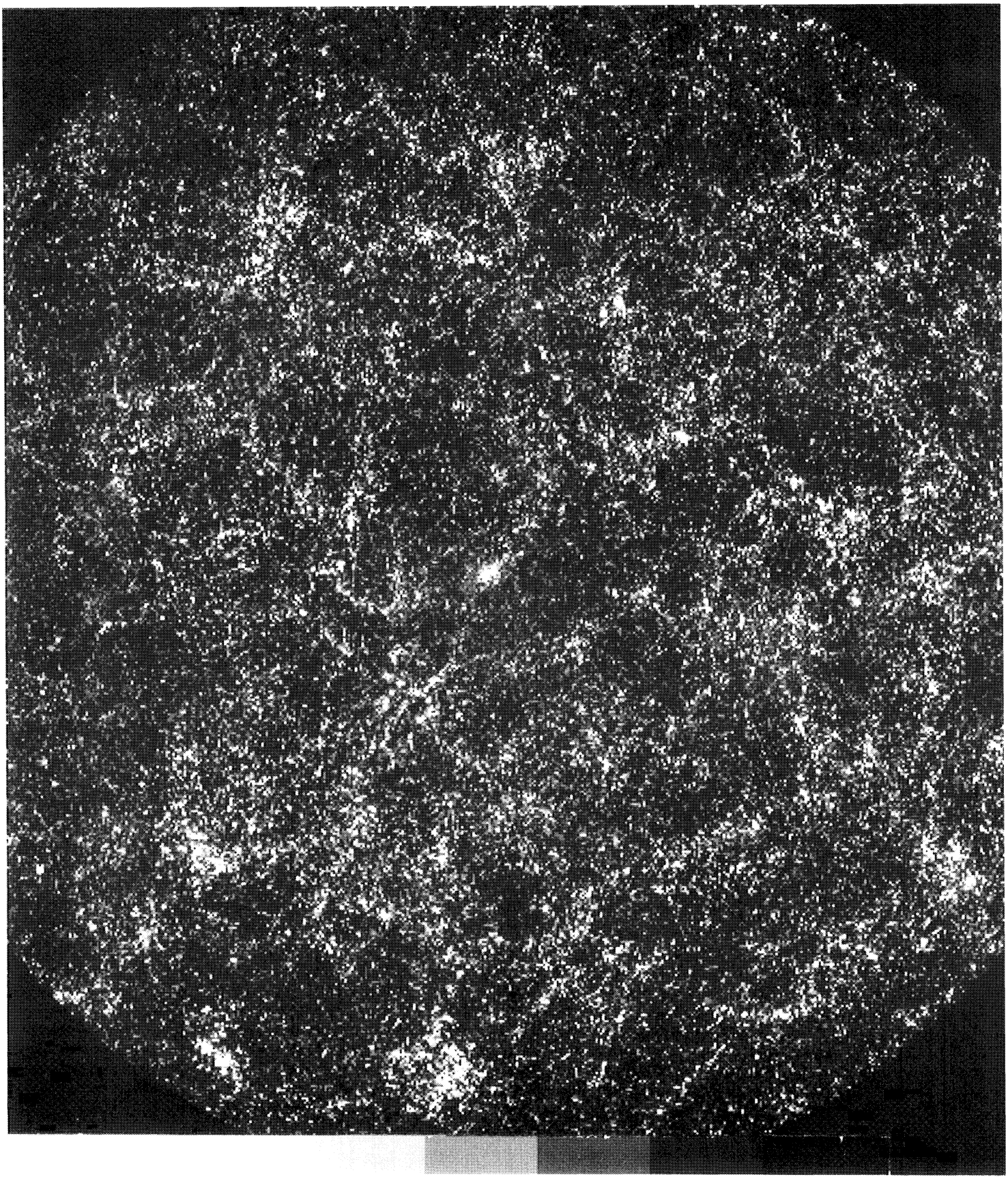

FIG. 3. Galaxy map for the Lick data with $b^{II} \geq 40°$. The map is an array of picture elements, each about the size of one 10′ by 10′ Shane–Wirtanen cell, in which the grey tone represents the galaxy count in the nearest cell. The grey scale shows the steps for successive increases of one galaxy commencing with black for zero.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1978AJ.....83

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1978AJ.....83

static space of the model. The K-correction, followed Pence (1976), is $3Z$ for all galaxies. The last term is an approximate correction for the effect of expansion.

In the model absolute magnitudes are assigned at random from Eq. (14), and the galaxies in the map are those with $m \leq 18.9$. The maximum distance of a visible galaxy is then $D = 524\ h^{-1}$ Mpc.

b) *Spatial Boundaries*

The model catalog is designed to be complete for the area of the sky $b \geq 40°$. We place clump centers at random in a cone of opening angle 100° truncated by a sphere with center at the apex. The point of observation is on the axis within the cone. The distance from this point to the apex, and the sphere radius, are chosen large enough that the boundary has negligible effect on the catalog—the perpendicular distance between the 100° cone of the observer and this full cone is almost as large as the maximum distance of a galaxy from the center of its clump.

c) *Shortcuts in the Computation*

We introduce some shortcuts to speed up the computation. We take the distance moduli for all galaxies in a clump to be equal to the distance modulus at the clump center. This is a good approximation because the typical distance of a clump is $\sim 230\ h^{-1}$ Mpc, the radius of a clump $\lesssim 10\ h^{-1}$ Mpc. (This approximation is not used in computing the angular position of each visible galaxy.) The distance modulus with Eq. (14) fixes the mean number of galaxies visible in a clump, and for distances greater than about 350 Mpc we simply select this number of galaxies at random from the clump. This ignores the fluctuations that would be present if the M were randomly drawn from the distribution. Also, the same galaxy occasionally is drawn more than once, but this causes no problems because the probability for duplication never exceeds 3% (of the small number visible at those large distances) and the positions are measured only to within the 10′ by 10′ cells. The mean number of duplicates expected is 0.18%, 700 galaxies in a total of 386 000.

Where the wanted number of clumps with the same parameters (but different random numbers) is greater than about 400, we build a small number of patterns and use each more than once, with separate random drawings of the wanted numbers of visible galaxies. For example, in the final model there are 24 000 clumps with seven levels, and we build 240 patterns.

IV. MODEL RESULTS

a) *Extrapolation of the $m \leq 14$ Model*

This model was tuned to match in some detail the statistics of the galaxy distribution at $m \leq 14$ (with $b \geq 40°$, $\delta \geq 0°$).

The distribution of η, the number of subclusters in a cluster, is

$$P_1 = 9.12\,A; \qquad P_\eta = A\exp -\eta/4, \qquad 2 \leq \eta \leq 20. \tag{16}$$

A is the normalizing constant, and $\langle \eta \rangle = 2$. The sphere radius at the largest level is $r_1 = 22\ h^{-1}$ Mpc, and the ratio of sphere radii at successive levels is λ [Eq. (4)] with $\gamma = 1.77$. The η subsphere centers are distributed uniformly at random in the sphere.

In the extrapolation from $m \leq 14$ to 18.9 we simply adjust the m limit and increase the depth of space in the model. There are $L = 6$ levels (two less than in SP to reduce computation time). The result has 1.5 times the galaxy density in the Lick sample. (The density at $m \leq 14$ apparently is high because of the Local Supercluster; the discrepancy may be due in part also to the uncertainty in the relative limiting magnitudes in the two catalogs.) To reduce the density, we accept each galaxy at $m \leq 18.9$ into the map with probability $P \approx 2/3$, independently decided for each galaxy. This leaves all the correlation functions unchanged.

The resulting map of the 386 000 visible galaxies at $b \geq 40°$ is shown in Fig. 2. It can be compared to the Lick data in Fig. 3. Each map is an array of 556 by 603 picture elements, representing counts binned in 10′ by 10′ cells, the grey tone in each cell representing the galaxy count. The Lick map differs slightly from that of SSGP because the number of picture elements here approximates the number of cells in the counts of Shane and Wirtanen. We have not introduced a correction for absorption in the galaxy.

Though this model produces maps that look very much like the real data at $m \leq 14$ (SP), the map at $m \leq 18.9$ certainly is disappointing. The Lick map has a crisp filamentary appearance, while the model map can only be described as "mushy."

Figure 4 shows that the two-point angular correlation functions of the model and the Lick data agree fairly well. At $\theta \lesssim 0.3°$ the model w tips slightly down. This is because the smallest sphere radius in the model is $1.3\ h^{-1}$ Mpc (about comparable to the cell size at the typical clump distance) so there are too few close pairs. At intermediate angles the model w falls with logarithmic slope 0.77 but has a slightly greater amplitude than the data. The amplitude has not been adjusted from its initial calibration at 14th mag. For $\theta \gtrsim 1.5°$ the model w begins a gentle roll-off due to the finite clump radius (Sec. II*f*) but at $\theta \gtrsim 5°$ it is much greater than the "corrected" w for the Lick sample.

The frequency distributions of galaxy counts in the 10′ by 10′ cells are compared in Fig. 5. The SP model and the Lick frequencies are quite similar at $n \lesssim 5$ galaxies per cell, but the model fails to match the extended tail at $n \gtrsim 10$. The discrepancy in a sense is quite small, for it involves only the $\sim$ 3% of the cells at $n \gtrsim 6$, but it is

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1978AJ.....83

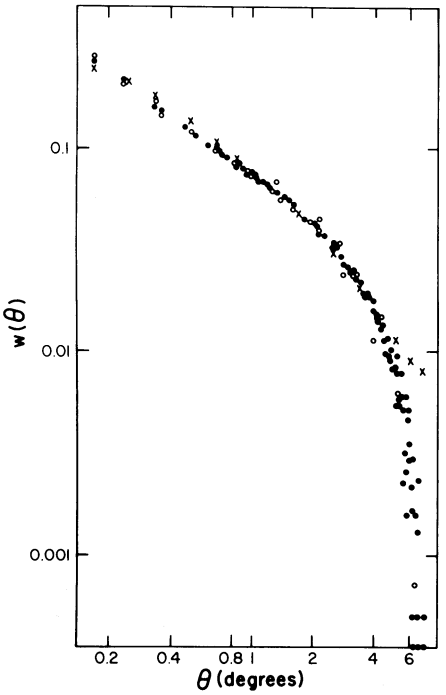

FIG. 4. The two-point angular correlation function, $w(\theta)$, for the Lick data (open circles), the extrapolated $m \leq 14$ model (crosses), and the final model (closed circles).

very important. To see this, we observe that both distributions at small n closely match a Poisson distribution with the same $\langle n \rangle$ (dotted line in Fig. 5). For example, the frequencies f_i for the Lick data and the Poisson distribution differ by <30% at $i = 0$ through 4 galaxies per cell. Despite this close numerical agreement the map of a Poisson distribution of galaxies looks quite different from the other maps in GPSS (pp. 96–97).

Because the tail of the frequency distribution determines the higher order correlation functions integrated over the 10′ cell, another way to state the discrepancy in Fig. 5 is that the higher order functions in the model are too small at $\theta \sim 10'$. Thus it is clear that part of the problem with the SP model is that it does not produce enough extreme concentrations. The eye takes particular notice of these occasional extremes of density, the unusual knots of galaxies. These will typically have many cells with large counts grouped closely together. Such configurations almost completely determine the high nth order correlation functions because the contributions to them go as the product of the cell counts taken n at a time. Hence the eye is very sensitive to the higher order correlation functions and it is at least partly the lack of strong high order correlations that accounts for the "mushy" appearance of the model.

b) Steps Toward the Final Model

Having realized that the extrapolated model is unsatisfactory at least in part because Q is too small, we entered on a long series of trials of remedies. (Many of the ~150 models we tried produced the distribution at $b \geq 70°$. This restricted sample is much less expensive to produce, but we found that it gives a much less useful visual impression than does the $b \geq 40°$ sample.)

The first approach was to adjust P_η [Eq. (16)]. By making this distribution strongly skew we can improve the frequency distribution of galaxy counts per cell. However, the result is that there are occasionally very large η, so that many or all clumps are rich at some level; Sec. II*d* contains a simple example of such a distribution. These large values of η show up as isolated spots or blotches on the map. We have found this to be true for a broad variety of forms for P_η.

We have found the same problems if $\langle n \rangle$ is reduced to a value well below $\langle n \rangle = 2$, so that the clustering develops more slowly, over many more levels.

The minimum cluster size in the original model is ~1 h^{-1} Mpc. Therefore we match the correlation between counts in neighboring cells but not the rms fluctuations in counts per cell, for that is determined by an integral over w at $\theta < 10'$, $\theta D \lesssim 0.7\ h^{-1}$ Mpc. One can remedy this by increasing the number of levels. This increases somewhat the dispersion in counts per cell, but does not significantly improve the picture.

To suppress the tendency of the model to produce isolated dense spots, we tried placing the clumps in groups, and we tried varying by a moderate amount the

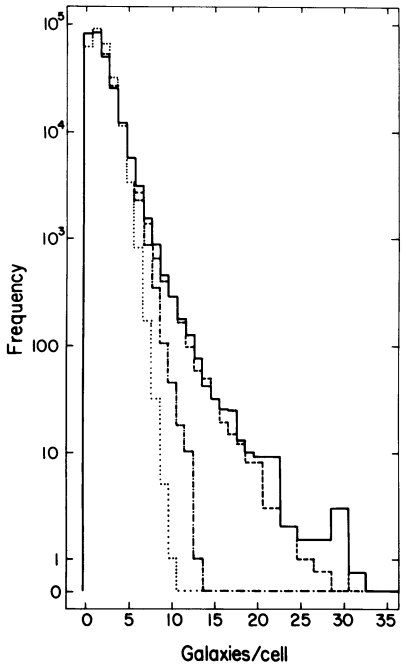

FIG. 5. The frequency distributions of galaxy counts in 10′ by 10′ cells for the Lick data (solid lines), a Poisson distribution (dots), the extrapolated $m \leq 14$ model (mixed dots and dashes), and the final model (dashes). All have 386 000 galaxies in 265 000 cells. A zero count is plotted the same distance below one that one is below two.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1978AJ.....83

number of levels in each clump in the group (with η still a random variable). However, the construction seemed awkward and the results not very pleasing.

The method finally adopted increases Q by introducing a large dispersion in L rather than η (Sec. II*e*). Only after we had seen that this yields more pleasing maps did we realize that it also gives a better fit to the shape of the three-point function.

The second important point in the development of the model was the realization that the appearance of the map is improved if the clump size is reduced so there are more clumps and the sizes of the random holes in the map are reduced. This was first demonstrated by cutting a model map into $\sim 3°$ by $3°$ squares and gluing the squares back together in random order.

c) The Final Model

i) Dumbbell Method of Construction

For the adopted model there are $\eta = 2$ subspheres in each sphere. The sub-sphere centers are placed at fixed separation d_l, in a randomly chosen orientation in three dimensions, with the pair centered on the sphere center. The separation at the largest level is

$$d_1 = 11.35\, h^{-1}\ Mpc. \tag{17}$$

The separation of the centers on the next level is d_2, with

$$d_1/d_2 = 1.10, \tag{18}$$

Thereafter the ratio of separations on successive levels is

$$d_n/d_{n+1} = \lambda, \qquad \lambda = 1.76. \tag{19}$$

The reduced ratio (18) increases the number of pairs at separations $\sim 10\ h^{-1}$ Mpc, thereby sharpening the roll-off of $w(\theta)$. One advantage of the dumbbell method of placing the subclusters is that clustering scales are less mixed than in the original method, so that it is easier to adjust radii on successive levels to get the wanted shape of $w(\theta)$.

ii) Distribution in Number of Levels

We match four parameters, the mean density $\mathcal{N}$ of visible galaxies, the amplitude A of the angular two-point correlation function, $w = A\,\theta^{1-\gamma}$, and the amplitudes of the angular three-point z and four-point μ functions (GP, FP),

$$\begin{aligned} K &= z/(w_1w_2 + w_2w_3 + w_3w_1) = 1.56, \\ R &= u/w^3 \approx 60, \end{aligned} \tag{20}$$

where the four-point function u is evaluated only for square configuration with $w(\theta)$ evaluated at θ equal to the side of the square. In the model these four parameters are adjusted by varying the mean space number density of clumps, M_c [Eq. (14)], and the first four moments of N_L, the number of galaxies in a clump [Eqs. (11)–(13)].

We take the distribution P_L in the number of levels in a clump to be different from zero in the range $L_1 \le L \le L_2$, and then seek by trial and error a P_L that matches the wanted three moment relations. The solution for P_L is not unique because not all moments are known, and even the first four moments are not uniquely known because there is an extra free parameter. The minimum number of levels L_1 is determined from the visual appearance of the map: a large number of galaxies from too small L makes the map appear too flat and Poisson-like. The upper value L_2 is strongly limited by required size of the third and fourth moments as N_L enters to a high power in eqs. (12) and (13). Using the condition that P_L ought to decrease in a reasonably smooth way with increasing L, we rather quickly settled on the distribution listed in column 4 of Table I. Column 2 lists the number of galaxies in each clump with L levels, the column 3 lists the separation between galaxy pairs at the final (smallest) level in the clump.

It is an amusing coincidence that the frequency distribution of the N_L turns out to be well approximated by the power law

$$P_L \propto N_L{}^{-1.835}, \tag{21}$$

close to the form of ξ, though P_L was not chosen with this in mind.

The final complication is the statistical fluctuations in the model. There are only 40 clumps at $L = 12$, and it is the nearest few of these that make the dominant contributions to z and u. The result is that there can be substantial fluctuations in these statistics from model to model. We do not know whether there would be similar fluctuations in different samples of the real universe—for example, if these massive systems were required to be well spaced, as might be reasonable because of the large mass needed to make them, it would reduce the fluctuations. In any case, to make the model statistics agree with the data we required that the total number of visible galaxies from all the clumps of given L be within 10% of the model ensemble average value. This number of galaxies seen is almost completely determined by the distribution of positions assigned to the centers of the clumps.

After 5 random trial placings of the 40 clumps with $L = 12$ we found one in which the number of galaxies visible from these clumps is within 10% of the ensemble average value. This is the adopted placing. The contributions from each of the other levels matched the ensemble averages within 10% on the first trial. To remove the remaining minor discrepancies in the four observed parameters we made slight adjustments to the numbers of clumps at $L < 12$. The final numbers in the full cone (Sec. III*b*) are listed in column (5) of Table I. The numbers of galaxies at $m \le 18.9$ contributed by the clumps of each sort are listed in column (6).

As a last step we have repeated the procedure with

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1978AJ.....83

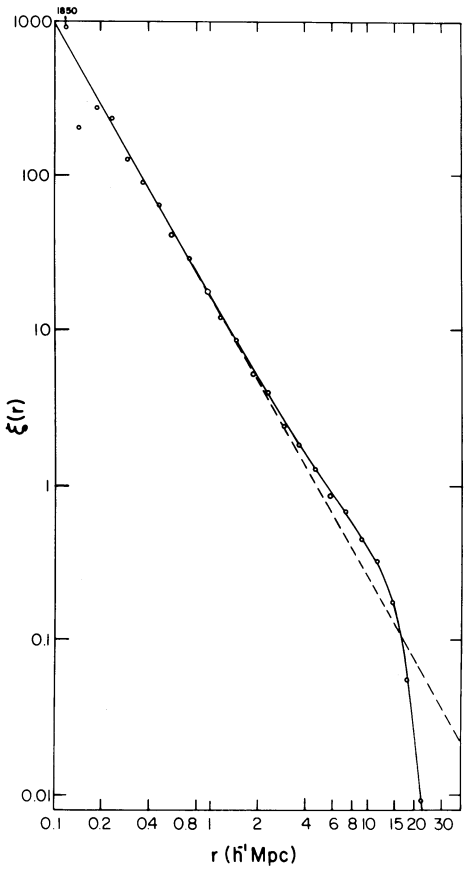

FIG. 6. The two-point spatial correlation function $\xi(r)$ for the model, determined by direct estimates in the known three-dimensional distribution. A smooth curve has been drawn through the points. The dashed line is a power law of slope $\gamma = 1.77$ and describes the asymptotic behavior of $\xi(r)$. To match the sharp roll-off of $w(\theta)$ we arranged that $\xi(r)$ gently rise above the power law before sharply falling beyond 15 h^{-1} Mpc.

different random numbers. Similar adjustments were required, and the final results were quite similar.

iii) Statistical Measures of the Final Model

The angular two-point correlation function w for the model is compared to the GP estimate in Fig. 4. The spatial two-point function ξ found by direct estimates from the known three-dimensional distribution is shown as the dots in Fig. 6. The solid line is a smooth curve connecting the dots. To match the sharp roll-off of $w(\theta)$ we had to arrange that ξ rises above the power law (dashed line in the figure) at $r \sim 10\ h^{-1}$ Mpc, then falls rather sharply below the power law (Davis, Groth, and Peebles 1977; Fall and Tremaine 1977).

The amplitude K of the three-point angular correlation function agrees with the GP estimate [Eq. (20)] to 1% accuracy. The amplitude of the four-point function is $R = 40 \pm 5$, 0.7 times the observed value [Eq. (20)], and within the uncertainty of the estimate. If there had been no dispersion in L, K, and R would have been smaller than the observed values by the factors 3.2 and 18.5.

Columns 8–10 in Table I give the fractional contributions to the two-, three- and four-point correlation functions $\propto P_L N_L{}^n$, $n = 2$, 3, and 4 [Eqs. (11)–(13)]. The two-point function is produced about equally by clumps of all richness, while the four-point function is almost entirely due to the 2% of the galaxies in the richest clumps.

In the model there is a very broad spread in clump richness N_L. The very rich clumps are needed to fit the higher order correlation functions. One should not conclude, however, that the clustering in the very poor clumps is negligible. One measure of this is that if all galaxies were in clumps with 7 levels, all other parameters in the model being unchanged, ξ at $r \ll R$ would have the power law form $B(\mathrm{hr})^{-\gamma}$ with $B = 4$ compared to $B = 15$ (r in Mpc) for the full distribution in L. Column (11) lists, for each L, the value of B that would have been obtained if all clumps had had that richness.

The frequency distribution of galaxy counts in 10′ by 10′ cells is shown as the dashed line in Fig. 5. the distribution closely matches the Lick data over a range of more than four orders of magnitude in the frequency. This might be expected since we have adjusted the model to match fairly well the first four moments. (The adjustment is not exact because (a) the angular correlation functions we match exclude the large-scale gradients in the Lick counts, and (b) the poorest clumps have minimum clustering scale, column (3) in Table 1, $\gtrsim$ the counting cell size.)

A final interesting statistic is the degree to which the clumps fill space. On the average, 75% of the galaxy pairs in a clump are at separations $\lesssim 15\ h^{-1}$ Mpc, 90% are at distances $\lesssim 18\ h^{-1}$ Mpc. In the model the total number

TABLE I. Distribution of cluster richness.

L (1)	N_L (2)	Minimum clustering diameter (3)	P_L (4)	No. of clumps (5)	No. of visible galaxies (6)	Fractional Contributions $\langle n \rangle$ (7)	$\bar{\xi}$ (8)	$\bar{\zeta}$ (9)	μ (10)	B (11)
7	128	612 h^{-1} kpc	0.68	24002	158030	0.41	0.11	0.01	0.0004	4
8	256	348	0.23	8214	106647	0.28	0.15	0.025	0.0023	8
9	512	198	0.061	2160	459459	0.15	0.16	0.052	0.0095	16
10	1024	113	0.015	534	30175	0.074	0.16	0.10	0.037	32
11	2048	64	0.0059	207	21597	0.057	0.24	0.32	0.23	63
12	4096	37	0.0011	40	9620	0.022	0.19	0.49	0.72	127

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1978AJ.....83

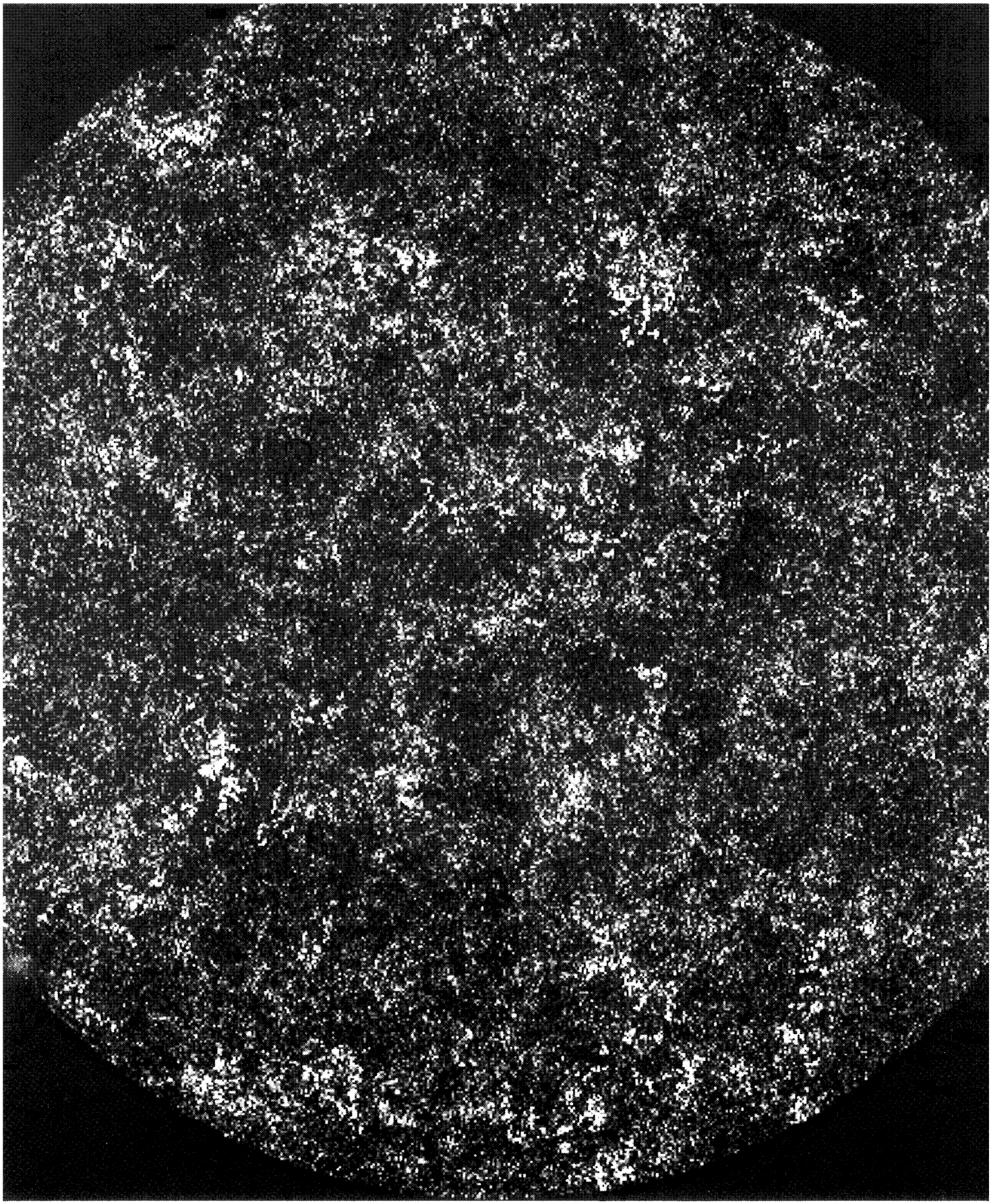

FIG. 7. Galaxy map for the final model as in Fig. 3.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1978AJ.....83

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1978AJ.....83

density of clumps is $n_{cl} = 3.1 \times 10^{-4}$ h^3 Mpc^{-3}, so the mean distance between clumps is $n_{cl}^{-1/3} = 15$ h^{-1} Mpc. Thus the clumps just about fill space with rather little overlap.

iv) The Model Map

The map of the model catalog is shown in Fig. 7. Though one can distinguish it from the Lick map we feel that is is a reasonable and pleasing approximation. The texture of the highlights is well reproduced. One sees holes between the highlights, and the abundance of the holes, their typical size, and the texture of the galaxy distribution within them all are reasonably well reproduced. Some interesting chains and filaments in the highlights can be made out in the model though it may be that such features are more common in the data. One cannot find objects like the Coma cluster (at the center of the Lick map), for the model was designed not to produce such apparently relaxed systems.

As was mentioned in Part ii) above we repeated the final prescription with a different set of random numbers, and obtained a similar-looking map.

v) Remarks on the Method of Construction

While the method for obtaining the model certainly has physical significance it also involves a number of artificial constructional devices—and after the building is completed such scaffolding should not be apparent. In the model the hierarchy is discrete, with $\eta = 2$ sub-clusters in each cluster on each level. The accidental overlaps from the random orientations have removed this discreteness from statistics such as $\xi(r)$ and from the appearance of clusters in the map. Another artificial device is the assignment of separate and independent clumps each with a definite number of levels. However, with a few exceptions, one cannot make out in the map the boundaries of any particular clump. (With other prescriptions, involving P_η, the boundaries were more apparent, and gave the map a mottled appearance.

One could think of many ways to place the sub-clusters in a cluster. We adopted a fixed separation because it is simple and it simplifies the adjustment of the roll-off of $w(\theta)$. We subsequently found that if the separation is a random variable, which would seem less artificial, the map is not quite as pleasing. Conceivably further experiments with the P_L would improve this, or perhaps the fixed separation really makes better looking filaments. Another choice presents itself in the method of producing the rather sharp roll-off in $w(\theta)$, which seems to be important to the appearance of the map (Sec. II.*f*). An alternative to the adopted approach of shrinking the cluster radius on the first level is to increase the mean number of sub-clusters on this level only. Maps produced by the latter method seem inferior. Yet another way is to make the clump positions anticorrelated. We have made only limited trials of this approach. The higher order correlation functions could provide a way to choose among methods, but we do not have useful estimates of these functions at large enough separations.

V. SUMMARY

We consider the following to be the main important points that have emerged from this study.

1. The known n-point correlation functions (1–4) do not give so detailed a description of the galaxy distribution that they alone can be used as the guide to how to construct a visually appealing model map at $m \leq 19$. This experiment has improved our understanding of the nature of the galaxy distribution. The final model is a good statistical and visual match to the Lick data. This seems to indicate that the hierarchical clustering picture must be a reasonably good approximation to reality.

2. The first model we tried had been adjusted to fit several statistical measures of the galaxy distribution at $m \leq 14$. When this model is extrapolated to $m \leq 19$ it is not very satisfactory. This is due in part to the limited statistics in such a small sample of galaxies, and it must also be a result of the Local Supercluster not being a fair sample of the Lick survey. We plan to report on 14th mag versions of the adopted model elsewhere.

3. In the Lick map the highlights have a striking filamentary appearance. In the model map one certainly can pick out equally striking filaments, even though the orientation of each subcluster pair is random so we are not building in linear structure. It may well be that the filaments in the Lick map are more common and finer, perhaps indicating a true filamentary character in the space distribution, though it has been difficult to find objective evidence of this. In any case the model map shows that many of these filaments will prove to be accidents in a clumpy distribution that are picked out because the eye is so adept at finding linear structures.

4. To fit the higher order correlation functions in the hierarchical clustering model, we need a very broad distribution in cluster richness—and this is pleasing because it is known from observation that there is a broad range in richness. We have tried two methods of introducing this spread, increasing the richness within randomly chosen clusters in randomly chosen levels, and increasing the richness within whole clumps. Though presumably both effects occur in nature, the indication, from the shape of the three-point correlation function (Sec. II*d*) and from the appearance of the maps, is that the latter dominates. That is, the neighbors of a rich cluster tend to be rich.

5. Along with the indicated spread in clump richness is the fact that the amplitudes of the angular three- and four-point functions are quite sensitive to how far away are the nearest several of the richest clumps. The shapes of the functions (manner of the variation with the arguments) are much less sensitive because each clump tends to produce the same functional form. It is not clear whether fluctuations in samples of the real universe at

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1978AJ.....83

$m \leq 19$ would be as large as they are in the model—possibly the fluctuations are suppressed by strong anticorrelation of the richest clumps.

6. Direct estimates of $w(\theta)$ for the Lick sample give tentative evidence of a dimunition in clustering at $\theta \sim 2.^{\circ}5$, $r \sim 10\ h^{-1}$ Mpc (GP). In the model a clump size of about this amount seems to be wanted—if the clump is much larger too many galaxies are tied up in each clump and so the map shows excessive gaps between the highlights. This is evidence in support of the GP interpretation of the break in ξ.

On much larger scales there is insufficient superclustering in the model. While one of the $\sim 15\ h^{-1}$ Mpc clumps can contain many $1.5\ h^{-1}$ Mpc Abell clusters grouped within it, by construction ξ is identically zero at $r \geq 35\ h^{-1}$ Mpc. It is known that there are small correlations on scales greater than or at least comparable to this (Abell 1958; Bogart and Wagoner 1973; Hauser and Peebles 1973; Seldner and Peebles 1977). Such large-scale effects are not very well explored. It could be, for example that although ξ is close to zero at $r \sim 30\ h^{-1}$ Mpc there is strong clustering among rich clumps, balanced, in ξ, by some degree of anticorrelation among lesser clumps.

7. In the model the clump size is just comparable to the mean distance between clumps. This says there is a fairly continuous sea of clustering, with no substantial gaps between clumps. On the other hand, since the subclusters occupy only a small fraction of the clump volume, there are holes in the galaxy distribution of the sort found by Chincarini and Rood (1976) and Tift and Gregory (1976) (Soneira and Peebles 1978).

We thank Ed Groth for help with the model catalog, Mike Seldner for help with the Lick catalog, and both for helpful discussions and opinions. The maps were made on the Princeton film scanner (Zucchino and Lowrance 1971; Heiles and Jenkins 1976). We are grateful to Paul Zucchino, John Opperman, and Eden Steiger for their help and cooperation in producing the maps and to the Princeton Astrophysical Sciences Department for the use of the film scanner.

APPENDIX A

We discuss here the behavior of the correlation functions when the number η of subclusters placed in a cluster is a random variable.

Let N_l be the number of galaxies in a particular cluster at the lth level of the hierarchy (l counted from the smallest scale). We can write this as a sum over the number of galaxies in each of the η subclusters on the next smaller level,

$$N_l = \sum_{i=1,\eta} N_{l-1}^{(i)}. \tag{A1}$$

Since η is drawn at random from P_η, the mean value of this equation is

$$\langle N_l \rangle = \langle \eta \rangle \langle N_{l-1} \rangle, \tag{A2}$$

whence

$$\langle N_l \rangle = \langle \eta \rangle^l. \tag{A3}$$

The mean square count is

$$\langle N_l^2 \rangle = \langle \eta \rangle \langle N_{l-1}^2 \rangle + (\langle \eta^2 \rangle - \langle \eta \rangle)\langle \eta \rangle^{2l-2}, \tag{A4}$$

where the second term has been reduced using eq. (A3). For $l \gg 1$ the solution to this equation is

$$\langle N_l^2 \rangle = \frac{\langle \eta^2 \rangle - \langle \eta \rangle}{\langle \eta \rangle^2 - \langle \eta \rangle} \langle \eta \rangle^{2l}. \tag{A5}$$

Similarly one finds that the third moment is

$$\langle N_l^3 \rangle = A \langle \eta \rangle^{3l}, \qquad A = \frac{\langle \eta^3 \rangle + b}{\langle \eta \rangle^3 - \langle \eta \rangle},$$
$$b = -3\langle \eta^2 \rangle + 2\langle \eta \rangle + \frac{3(\langle \eta^2 \rangle - \langle \eta \rangle)^2}{\langle \eta \rangle^2 - \langle \eta \rangle}. \tag{A6}$$

The two-point correlation function ξ at $r \sim R_l = r_0 \lambda^l$ is proportional to the ratio of the mean density within a cluster at level l to the overall mean density $\langle n \rangle$. The former density must be weighted by the change of choosing a galaxy in a particular cluster $\propto N_l$ so

$$\xi \approx \frac{\langle N_l^2 \rangle}{\langle n \rangle \langle N_l \rangle R_l^3} \propto \langle \eta \rangle^l R_l^{-3}. \tag{A7}$$

The last expression follows from Eqs. (A3) and (A5). It agrees with the first of Eq. (4) if η is replaced with $\langle \eta \rangle$.

The three-point function ζ at $r_1 \sim r_2 \sim r_3 \sim R_l$ is proportional to the square of the density, weighted by N_l,

$$\zeta \approx \langle N_l^3 \rangle / (\langle n \rangle^2 \langle N_l \rangle R_l^6), \tag{A8}$$

so Q varies with the moments of P_η as

$$Q \propto \frac{\langle N_l^3 \rangle \langle N_l \rangle}{\langle N_l^2 \rangle^2}$$
$$= \frac{\langle \eta^3 \rangle + b}{\langle \eta \rangle^3 - \langle \eta \rangle} \left[\frac{\langle \eta \rangle^2 - \langle \eta \rangle}{\langle \eta^2 \rangle - \langle \eta \rangle} \right]^2. \tag{A9}$$

If the dispersion and skewness of P_η are large this reduces to

$$Q \propto \langle \eta^3 \rangle / \langle \eta^2 \rangle^2, \tag{A10}$$

at fixed $\langle \eta \rangle$.

REFERENCES

Abell, G. O. (1958). Astrophys. J. Suppl. **3,** 211.
Bogart, R. S., and Wagoner, R. V. (1973). Astrophys. J. **181,** 609.
Chincarini, G., and Rood, H. J. (1976). Astrophys. J. **206,** 30.
Davis, M., Groth, E. J., and Peebles, P. J. E. (1977). Astrophys. J. Lett. **212,** L107.
de Vaucouleurs, G. (1970). Science **167,** 1203.

© American Astronomical Society • Provided by the NASA Astrophysics Data System