

THE SHANE WIRTANEN COUNTS: OBSERVABILITY OF THE GALAXY
CORRELATION FUNCTION

MICHAEL E. BROWN AND EDWARD J. GROTH

Joseph Henry Laboratories, Physics Department, Princeton University

Received 1988 February 8; accepted 1988 August 23

ABSTRACT

For an explicit test of the ability to recover the galaxy two-point correlation function from the Lick catalog of Shane and Wirtanen, we have applied the reduction and analysis methods of Seldner *et al.* and Groth and Peebles to model galaxy distributions that have known plate and field “errors” and that are high-fidelity simulations of the Lick sample. The model galaxy space distribution is constructed with the Soneira-Peebles prescription, which generates model distributions which have two-, three-, and four-point correlation functions in good agreement with the observed correlation functions. The space distribution is projected onto the sky with and without plate “errors.” The Seldner *et al.* analysis recovers the plate factors in the former case with an error of 6.3%, as originally estimated. The two-point correlation function estimated from the “corrected” model catalog reproduces the built-in correlation function including the break from the power law. This is also true if the angular scale of the break is increased or decreased by a factor of 1.76 from the observed value. We also compare a map of the corrected counts with a map of the counts projected without plate errors and find that the corrected map is a good visual representation of the galaxy distribution. Finally, we construct a simulation which includes systematic variations in plate sensitivity with observer and time—so called “plate shape gradients.” Once again, the correlation function of the model catalog reproduces the built-in correlation function.

Subject headings: cosmology — galaxies: clustering

I. INTRODUCTION

Groth and Peebles (1977, hereafter GP77) determined the galaxy two-point angular correlation function from the Lick counts of galaxies (Shane and Wirtanen 1967) as reduced by Seldner *et al.* (1977, hereafter SSGP). The principal result of the GP77 analysis is that galaxies are distributed as a scale-invariant power-law fractal on small scales, with

$$w^*(\theta) = A\theta^{1-\gamma}, \quad A = 0.0684, \quad \gamma = 1.741, \quad (1)$$

where $w^*(\theta)$ is the two-point angular correlation function and θ is measured in degrees. For $\theta \gtrsim 2.5$, $w^*(\theta)$ falls well below the extrapolation of the small-scale power law. This break may be important for theories of galaxy formation (e.g., Davis, Groth, and Peebles 1977).

Geller, de Lapparent, and Kurtz (1984, hereafter GdLK) and de Lapparent, Kurtz, and Geller (1986, hereafter dLKG) raised several concerns with the GP77 determination of the galaxy correlation function and suggested that the break might be significantly weaker (occur at larger θ) or stronger than deduced by GP77. These concerns were addressed by Groth and Peebles (1986*a, b*, hereafter GP86*a, b*) who showed that the original reduction and analysis are valid.

In this paper we examine one of the points raised by dLKG in more detail. Based on simulations of the Lick counts, dLKG concluded that the analysis method of GP77 was incapable of recovering the true correlation function of the galaxy distribution. GP86*b* argued that the dLKG simulations were not of sufficiently high fidelity to warrant such a conclusion. In addition, GP86*b* pointed out that dLKG did not properly apply the analysis method of GP77 and that if this method had been properly applied, it would have automatically accounted for the extra noise introduced by the lack of fidelity and would have recovered the true correlation function.

We create high-fidelity simulations starting with the distribution of galaxies in space. The space distributions are then projected onto a set of simulated Lick plates which are reduced to a uniform limiting magnitude with the method of SSGP. The simulated catalogs are then analyzed with the methods of GP77. We find that the resulting correlation functions accurately reproduce the correlation functions that were built in during the first step of the simulations.

In the next section, we provide a brief overview of the Lick counts, the SSGP reduction, and the GP77 analysis. Subsequent sections describe the construction of the simulated catalogs, the reduction of the catalogs, and the analysis of the catalogs. In § VI, we consider a simulation which includes “plate shape gradients.” Conclusions are presented in the final section.

II. OVERVIEW OF THE REDUCTION AND ANALYSIS OF THE
LICK COUNTS

Shane and Wirtanen counted galaxies to ~ 19 th magnitude in $10' \times 10'$ cells on $6^\circ \times 6^\circ$ plates. There are 1246 plates which cover the sky north of $\delta = -23^\circ$. They have 5° spacing in declination and $\sim 5^\circ$ spacing in right ascension.

Due to effects such as vignetting, emulsion variations, atmospheric conditions, personal factors, and so on, the limiting magnitude of the survey varies from cell to cell within plates and from plate to plate within the catalog. SSGP attempted to correct the counts to a uniform limiting magnitude by determining “correction factors” such that when an observed count is multiplied by the appropriate correction factor, the result is statistically equivalent (in a sense to be explained more fully below) to the count that would have been obtained from a survey with a uniform limiting magnitude. By stacking all the plates, SSGP determined 1296 field correction factors to

account for variations from cell to cell. By minimizing the difference of the counts in the plate overlap regions, 1246 plate correction factors were determined. With the field and plate correction factors, the counts are reduced to a uniform limiting magnitude but still contain the effects of atmospheric absorption and galactic obscuration. SSGP used the cosecant model to determine atmospheric and galactic factors which reduce the counts to a uniform limiting magnitude as seen “outside the Galaxy.” GP77 introduced additional “smoothing factors” to remove large-scale ($\gtrsim 40^\circ$) gradients in the counts. GP86b argued that the large-scale gradients might be due to an artifact of the SSGP edge-matching procedure, the inadequacy of the simple cosecant model for galactic absorption, or intrinsic large-scale gradients in the galaxy distribution. This point is discussed further in § IVc below.

GP77 used the counts (at $|b| > 40^\circ$) and the correction factors to determine the galaxy correlation function. In brief, the method computes $w(\theta)$ as the average of terms of the form

$$w_{ij} = (C_i^e n_i^r - n^*)(C_j^e n_j^r - n^*)/n^{*2}, \quad (2)$$

where n_i^r is the raw count for the i th cell, C_i^e is the estimate of the corresponding correction factor, n^* is the mean corrected count, and cells i and j are separated by angular distance θ . The fact that these terms contain the correction factors introduces several complications into the analysis of GP77. The two largest effects have to do with plate-to-plate variations in the depth of the catalog and spurious correlations introduced by the variances and covariances of the estimated correction factors. Although the correction factors normalize the counts to a uniform depth, they do not normalize the structure in the counts, measured by the correlation function, to a uniform depth. GP77 accounted for this by scaling each term of the form given above to the correct depth using the known scaling relations for the angular two-point correlation function (Peebles 1973). It should be noted that this correction is multiplicative and can neither mimic nor hide structure such as a break. The correction factors are estimated from the data and necessarily contain observational errors. This means that the average over expressions of the form given in equation (2) will contain spurious terms proportional to the variance and covariance of the correction factor error distribution. SSGP estimated these variances and the GP77 subtracted these terms from the estimates of the correlation function. If the correction for these terms were omitted and if they were sufficiently large, they could hide structure such as the break (this is what happened in the dLKG simulations). On the other hand, if the applied correction were much bigger than the needed correction, a spurious break could be introduced.

In fact, as first noticed by Fry and Seldner (1982), when the correlation function is computed by simply averaging terms of the form given in equation (2), without applying the corrections discussed above, the result is very close to the GP77 result. As noted by GP86a and GP86b, this results from the fortuitous circumstances that the width of the correction factor distribution is small, so the scaling correction is negligible, and the width of the correction factor error distribution is small, so the correction for spurious correlations is negligible.

III. SIMULATIONS OF THE LICK COUNTS

To construct the simulated counts we use the method of Soneira and Peebles (1978, hereafter SP) to generate the distribution of galaxies in space and assign absolute magnitudes.

The galaxies are then “observed” yielding a set of 1246 plates. These observations suffer from the same effects as do the actual observations, including cell to cell and plate to plate variations in limiting magnitude and atmospheric and galactic absorption. However, there is one important difference between the simulated and actual observations: the magnitudes of these effects are known for the simulated observations so we have a direct measure of how well the reduction and analysis procedures correct for these effects.

We use three simulations—a fourth simulation is described in § VI. All have a built-in correlation function which matches the amplitude and exponent of the observed power law at small angles. The built in correlation function of the first simulation has a break at $\theta = 2.5$ in order to test the reduction and analysis methods when applied to a catalog that matches the actual data as closely as possible. The other simulations have breaks at smaller (nominally, $\theta = 1.4$) and larger (nominally, $\theta = 4.4$) angles in order to verify that the reduction and analysis methods do not have some obscure systematic problem which always generates a break at $\theta = 2.5$.

a) Simulation of the Space Distribution of Galaxies

The SP method of constructing the spatial distribution of galaxies builds clusters containing a hierarchy of subclustering and then distributes the cluster centers at random throughout the volume considered. The SP method, when properly tuned, reproduces the observed two, three, and four-point correlation functions of the Lick counts. Maps generated from the SP model have a visual appearance similar to the Lick map. To generate simulation 1, we use the parameters suggested by SP as these have already been tuned by SP to simulate the Lick counts. For simulations 2 and 3, (with breaks at $\theta = 1.4$ and $\theta = 4.4$) we retune the parameters to give the desired two-point function. There are no observations of the three- and four-point correlation functions corresponding to these modified two-point functions, so matching the three- and four-point functions is not a constraint in this retuning. Instead, we make the simplest possible changes to the parameters which gives the desired two-point function. Aside from this retuning, we follow the method of SP except in the following areas. First, for distant clusters, SP calculate the fraction of galaxies which should be visible and choose galaxies at random until the correct fraction of visible galaxies is obtained. Since our observations have a nonuniform limiting magnitude, we do not determine the visibility of a galaxy until its apparent magnitude and the limiting magnitude of its Lick cell have been determined. In fact, in regions of plate overlap, a given galaxy may be visible on some plates but not on others. Second, we generate enough clusters to fill a sphere around the observer rather than just a cone. This is because we want to simulate the entire Lick map, not just that portion with $b > +40^\circ$. Third, once the spatial distribution of galaxies in a cluster is generated, that same spatial distribution is used 100 times, although new absolute magnitudes are chosen each time the distribution is used. This scheme saves computation time and was used by SP except for the two richest levels of clustering. We are unable to think of any reason why this small difference in construction techniques should cause any overall difference in the quality of the simulations.

The parameters which vary from simulation to simulation are listed in Table 1. Hubble’s constant scales out of the simulations and for convenience we have taken $H = 100 \text{ km s}^{-1}$

TABLE 1
SIMULATION PARAMETERS

PARAMETER	SIMULATION		
	1 and 4	2	3
Break location	2°5	1°4	4°4
d_1 (Mpc)	11.35	6.45	19.98
P_6	0	0.683	0
P_7	0.683	0.234	0
P_8	0.234	0.061	0.683
P_9	0.061	0.015	0.234
P_{10}	0.015	0.0059	0.061
P_{11}	0.0059	0.0011	0.015
P_{12}	0.0011	0	0.0059
P_{13}	0	0	0.0011
Cluster density (Mpc ⁻³) ...	0.000310	0.000620	0.000155

Mpc⁻¹ throughout this paper. The break location in the first line of Table 1 characterizes the simulation while d_1 is the diameter of the top level in the clustering hierarchy. Recall that the SP method constructs clusters by choosing two points with separation d_1 and a random orientation. With each of these points as centers, two more points with random orientations and separations d_2 are chosen. With each of these four points as centers, two more points with random orientations and separations d_3 are chosen. This continues for n steps and galaxies are placed at the locations of the 2^n points chosen in the final step. At the top level, $d_1/d_2 = 1.10$, while for all other levels, $d_i/d_{i+1} = 1.76$. The ratio 1.76 is chosen to generate a power-law exponent of 1.77 in the spatial function, while the smaller ratio at the top level enhances the sharpness of the break to agree with what is observed. In Table 1, P_n is the fraction of n level clusters in the model. The last line of Table 1 gives the space number density of clusters in the model.

As can be seen from Table 1, in order to increase the location of the break by a factor of 1.76, we increase d_1 by this factor and this increases the spatial and angular scales over which galaxies are correlated by the same factor. To maintain the same number of correlated galaxies per galaxy at small scales, the number of levels in the clustering hierarchy is increased by one. This doubles the average number of galaxies per cluster, so the space density of clusters is halved. The net effect of this retuning of the parameters is to increase the location of the break by a factor of 1.76 while leaving unchanged the mean density of galaxies and the amplitude of the two-point correlation function at small scales. The opposite retuning is used to place the break at 1°4.

The simulations with breaks at 2°5 and 1°4 turned out as expected on the first realization. However, simulation 3, with a break at 4°4, required nine realizations before the break in the built-in correlation function occurred at a substantially larger angle than that in simulation 1. Simulation 3 is noisier than simulation 1 since it contains half as many clusters; it may be that fluctuations in the galaxy distribution are the dominant source of noise when the amplitude of the correlation function is somewhat smaller than that of the observed break. We plan to investigate the fluctuations to be expected from different samples of the universe in a future paper. For the present work, the goal is to make a simulated universe with a given correlation function and to determine if the reduction and analysis methods recover that correlation function.

b) Simulation of the Absolute Magnitudes

Each galaxy in the simulation is assigned an absolute magnitude chosen at random from an Abell (1962) luminosity function truncated at both faint and bright magnitudes,

$$\begin{aligned}
 dN/dM &= 0, & M^* + 3 < M; \\
 &\propto \beta \text{ dex } [\beta(M - M^*)], & M^* < M < M^* + 3; \\
 &\propto \alpha \text{ dex } [\alpha(M - M^*)], & M^* - 2 < M < M^*; \\
 &= 0, & M < M^* - 2; \\
 \alpha &= 0.75, \quad \beta = 0.25, & M^* = -18.6.
 \end{aligned} \quad (3)$$

This luminosity function is the same as that used by SP.

c) Simulation of the Apparent Magnitudes

The apparent magnitudes of the galaxies are calculated according to

$$m = 5 \log D + 25 + M + (3 + 5/\ln 10)HD/c, \quad (4)$$

where D is the distance to the galaxy, and the last term models the effects of expansion and the K -correction.

d) Simulation of the Observations

Once the apparent magnitudes have been assigned, the galaxies are "observed." At the position of each Lick cell, on each Lick plate, all visible galaxies are counted. A galaxy is visible if its apparent magnitude is brighter than the limiting magnitude of the Lick cell. The limiting magnitude is based on the product of the simulated field and plate correction factors and atmospheric and galactic absorption factors. Each cell is assigned the field correction factor determined by SSGP. Each plate is randomly assigned a plate correction factor drawn from the distribution of plate correction factors determined by SSGP for plates with centers that satisfy $|b| > 40^\circ$. We do not use the distribution of all the plate correction factors since the errors in the SSGP determination of the plate correction factors are greater at low galactic latitudes. The width of this distribution is the width of the distribution of true plate correction factors broadened by the distribution of errors in the plate correction factors. Therefore we expect that the simulated catalogs have plate variations in limiting magnitude which are slightly greater than those of the Lick Catalog. The consequence of this slight infidelity in the simulations is that the reduction and analysis methods are subject to a slightly more stringent test when applied to the simulated catalogs than when applied to the Lick counts.

The atmospheric and galactic factors are given by

$$C_a = \exp [0.454(\sec \delta - 37.3) - 1], \quad (5)$$

and

$$C_g = \exp [0.587 \csc |b| - 0.717], \quad (6)$$

as determined by GP77.

If C is the overall correction factor assigned to a cell, then the limiting magnitude of the cell is

$$m_L = m_0 - 2.079 \log C, \quad (7)$$

where $m_0 = 18.9$ is the nominal limiting magnitude of the catalog. The factor 2.079 in this equation results from the fact that the correction factors are applied to counts, rather than magnitudes. With the luminosity function given in equation

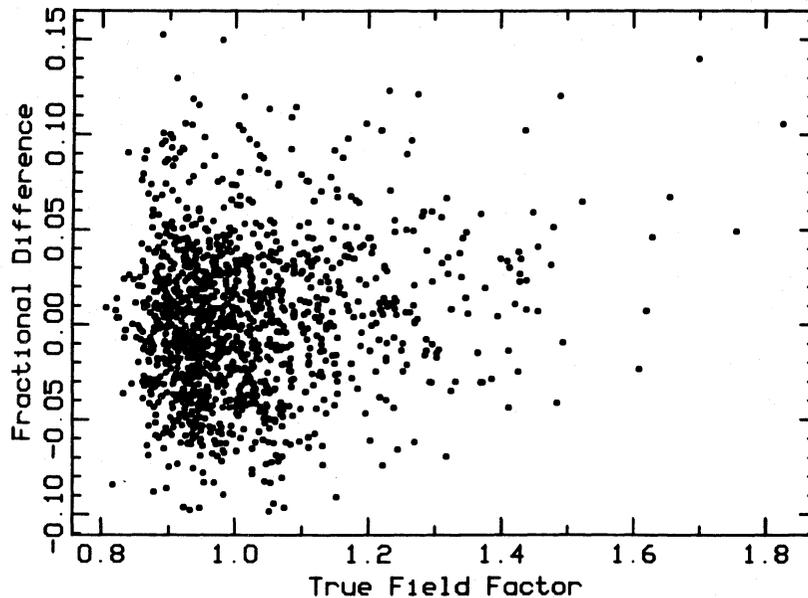


FIG. 1.—Fractional error in the field correction factor in the sense estimated factor minus true factor divided by true factor vs. the true field correction factor. These data are from simulation 1.

(3), the number of galaxies as function of limiting magnitude varies as

$$d \log N/dm = 0.481, \quad (8)$$

in the neighborhood of 19th magnitude.

Each cell of the simulated observations has a sharp limiting magnitude while the Lick catalog most likely has a small spread in limiting magnitude within each cell since Shane and Wirtanen could not determine apparent magnitudes perfectly. We do not model this spread in limiting magnitude since it has negligible effect on the simulations. It is equivalent to a slight broadening of the already broad galaxy luminosity function together with selection to a well defined limiting magnitude. SSGP and GP77 account for this effect with a counting error term which modifies the estimated correlation function only at scales comparable to the cell size. Such scales are not important for our simulations. The counting error is an additional source of error in the determination of plate factors by the method of matching the counts in regions of plate overlap. However, the dominant source of error is the nonuniform distribution of galaxies. According to the SSGP analysis, the contribution of the counting error term to the mean square error in a typical plate overlap is less than 4% of the total mean square error (SSGP eqs. [A15] and [A16]).

IV. REDUCTION OF THE SIMULATED COUNTS

a) Field Correction Factors

The SSGP method is used to calculate the field correction factors. All the plates are stacked and the field correction factor for a cell is proportional to the inverse of the stacked count in that cell. Figure 1 shows the fractional difference between the estimated and true field correction factors plotted against the true field correction factors. The data shown are from simulation 1; plots for the other simulations are similar. We note that there are no systematic trends evident in this figure. Further-

more, the mean difference in correction factors is 0.0022 and the sample standard deviation of the difference, i.e., the typical error in an individual field correction factor, is 0.040, which is close to the SSGP estimate of 0.045 to 0.047.

b) Plate Correction Factors

The plate correction factors are estimated with the SSGP method which minimizes, in the least squares sense, the number weighted differences of galactic density in plate overlap regions. The plate correction factors can be multiplied by an overall constant without affecting the solution. Therefore, in the following comparisons of the estimated and true plate correction factors, we have normalized the estimated factors such that the average of the estimated factors agrees with the average of the true factors for plates at $|b| \geq 40^\circ$. Figure 2 shows the fractional difference between the estimated and true plate correction factors plotted against the true plate correction factors. The data are from simulation 1 and include only those plates with $|b| > 40^\circ$. Figure 3 shows the data for plates with $|b| < 40^\circ$. Plots for the other simulations are similar. Once again, there are no systematic trends; this indicates that the SSGP procedure generates unbiased estimates of the plate correction factors. For $|b| > 40^\circ$, the mean difference is -0.0035 with sample standard deviation 0.063, while for $|b| < 40^\circ$ these quantities are -0.0287 and 0.181.

The SSGP estimates of the standard deviations are 0.063 and 0.11 for high and low latitudes, respectively. As can be seen, the result from the simulation and the SSGP estimate are in extremely good agreement for plates at high latitudes. GP86b pointed out that the large scale gradients in the counts might be due in part to an artifact of the SSGP edge matching procedure and that the SSGP estimate for the error in the plate correction factors might need to be increased to $\sim 8\%$. This point is discussed in greater detail in the next section.

At low latitudes, the standard deviation in plate correction factors from the simulation is substantially larger than the

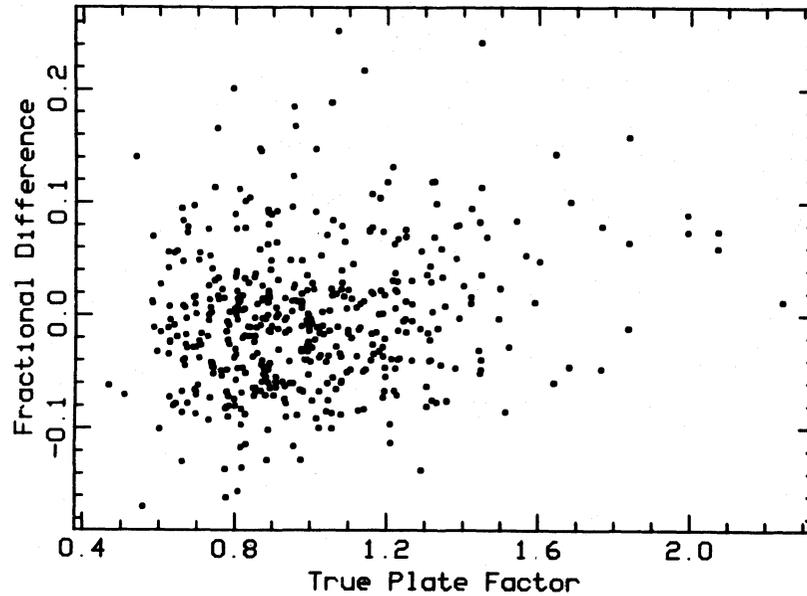


FIG. 2.—Fractional error in the plate correction factor in the sense estimated factor minus true factor divided by the true factor vs. the true plate correction factor. These data are for plates with $|b| > 40^\circ$ from simulation 1.

SSGP estimate. In the simulation, the cosecant model of galactic absorption (eq. [6]) is applied all the way down to $b = 0^\circ$, so the simulation contains many fewer galaxies at low latitudes than the Lick catalog. In fact, simulated plates on the galactic plane contain no galaxies (something which did not happen in the Lick catalog). The correction factors for these plates are arbitrarily set to 1 and are responsible for the hyperbolic arc which appears in Figure 3. Since our simulation does not reproduce the apparent distribution of galaxies at low latitudes, we do not consider the discrepancy between the result of the simulation and the SSGP estimate of the errors in the plate correction factors at low latitudes to be a serious problem.

c) Large-Scale Gradients

SSGP calculated correction factors to account for atmospheric and galactic absorption. GP77 applied these correction factors to the data and also applied smoothing factors which were obtained by fitting the counts to smooth functions and then correcting the counts by these smooth functions. Since atmospheric and galactic absorption produce large-scale gradients in the counts (at least those components modeled by SSGP), it is not necessary to calculate separately absorption factors and smoothing factors. Since we are not interested in checking how well SSGP estimated the absorption factors, we

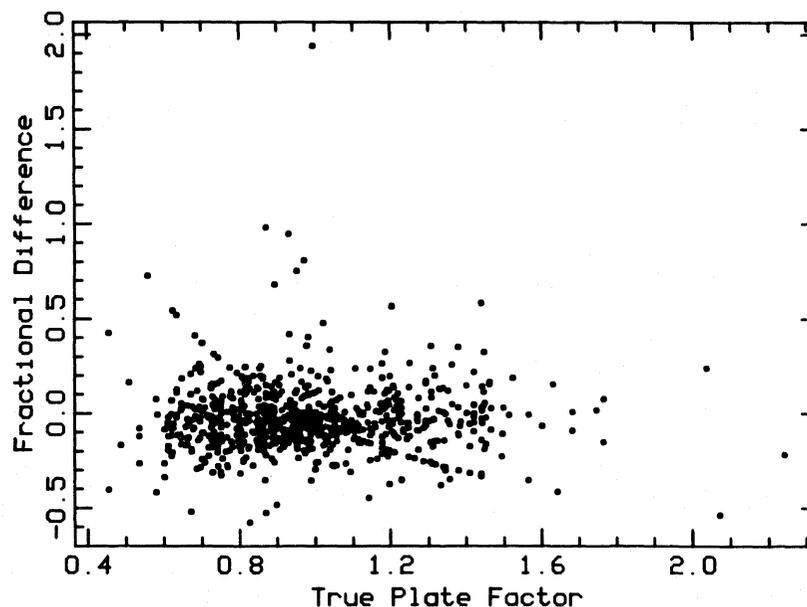


FIG. 3.—Same as Fig. 2, but for plates with $|b| < 40^\circ$ from simulation 1

lump all the factors together and treat them as smoothing factors.

The simulated counts, corrected by the field and plate correction factors, are projected onto a polar map like those shown in SSGP. That portion of the map with $b > +40^\circ$ is fitted to a two-dimensional, second-order polynomial which we call the smoothing polynomial. The counts are divided by the inverse of the smoothing polynomial in order to remove large-scale gradients from whatever source.

To investigate whether the SSGP edge matching procedure generates spurious large-scale gradients, we multiply the smoothing polynomial by the inverse of the absorption factors (eqs. [5] and [6]) which were built into the simulation. The resulting map is then an estimate of the large gradients in the simulated counts. These gradients may result from either the SSGP edge-matching procedure or intrinsic gradients in the distribution of galaxies. No intrinsic gradients were built in to the simulations, so any intrinsic gradients must result from statistical fluctuations and are believed to be small.

The root mean square value of the large-scale gradients in the map just described is only 4.2%. This is less than half that estimated by GP86*b* from a model which allowed for the propagation, throughout the grid of plates, of errors due to fluctuations in the counts in plate overlap regions. This small result may be due to a statistical fluctuation although the results for the other simulations are also small: 7.6% for simulation 2 and 6.3% for simulation 3.

The GP86*b* model assumed a rectangular grid of plates approximating the Lick map for $b > +40^\circ$. The prediction of this model is that the mean square error in plate correction factors (for $b > +40^\circ$) is 0.0127. This is the mean square error due to all sources including plate to plate errors and whatever large-scale gradients are induced by the SSGP edge matching procedure. The value of the mean square error deduced from the simulation (see the previous section) is 0.004, so the discrepancy is actually a factor of 3.

What is the source of this discrepancy? It may be that GP86*b* were too conservative in their modeling. For example, the fluctuations in the counts in the overlap region depend on the galaxy correlation function. GP86*b*, following SSGP, took a conservative approach in which the correlation function (for the purposes of estimating the fluctuations in the overlap region) was assumed to have no break. If the break is included, the estimated fluctuations drop by $\sim 50\%$. The GP86*b* model also assumes that each plate has a $1^\circ \times 6^\circ$ overlap with its four nearest neighbors. In fact, there is $\sim 30\%$ more overlap area due to the curvature in the plate grid. These two effects can account for about a factor of 2. In addition, there are several other effects whose magnitude is hard to estimate analytically. Since the typical overlap is bigger than 6 square degrees, the correlation function averaged over an overlap is smaller than assumed by GP86*b*. Also, the GP86*b* model ignores "corner" overlaps with diagonally adjacent plates. For these reasons, we believe the simulation gives a better estimate of the errors in the plate correction factors and the large-scale gradients induced by the SSGP edge-matching procedure. If this is the case, the error in the plate correction factors for high-latitude plates is $\sim 6.3\%$, the amplitude of the large-scale gradients induced by the edge-matching procedure is $\sim 4\%$, and the observed large-scale gradients (amplitude $\sim 20\%$) in the Lick counts must result from intrinsic gradients in the galaxy distribution or an inadequacy in the simple cosecant model of galactic absorption.

DLKG attributed the observed large-scale gradients to the results of "plate shape gradient" effects. On the basis of several arguments, GP86*b* concluded that the plate shape gradient effects were too small to introduce systematic errors in the SSGP and GP77 results. One of the arguments of GP86*b* is that the large-scale gradients could be produced by the known errors propagating through the grid of plates. If this were the case there would be no need to introduce plate shape gradients to explain the large-scale gradients. In view of the present simulations, this is not necessarily a valid argument. However, the other arguments of GP86*b* concerning plate shape gradients remain valid. These include their lack of statistical significance and the absence of a "contrast stretch effect." In § VI, we show explicitly that plate shape gradients do not generate large-scale gradients. One might well expect that a significant fraction of the observed large-scale gradients are due to large-scale gradients in the intrinsic galaxy distribution, especially in view of the results from redshift surveys concerning voids (Kirshner *et al.* 1981), bubbles (de Lapparent, Geller, and Huchra 1986), and filaments (Haynes and Giovanelli 1986).

d) Visual Examples of the Correction Procedure

In Figures 4, 5, and 6 we show maps of the counts from simulation 1. These maps are polar projections such that the north galactic pole is at the center, the direction to the Galactic center is at the bottom, and the radial coordinate is proportional to $+90^\circ - b$. This is the same projection used in the SSGP maps and the map described in the previous section. At the centers of the maps, pixels are $10' \times 10'$. The maps contain 180×180 pixels or $30^\circ \times 30^\circ$, so they contain $\sim 6 \times 6$ plates, although the plate grid runs approximately diagonally through the maps.

Figure 4 shows the simulated space distribution projected onto the sky with a uniform limiting magnitude. That is, Figure 4 represents what might be seen by an observer with an ideal galaxy counting apparatus located above the Earth's atmosphere and outside the Galaxy. In Figure 5 are shown the simulated raw counts, and in Figure 6 are shown the simulated corrected counts. With some difficulty, the plate grid can be picked out in Figure 5, but it is almost impossible to find the grid in Figure 6. (A caution to the reader: before you conclude that you have found a plate boundary in Fig. 6, check Fig. 4 for the same feature you've identified as a plate boundary, and remember that Fig. 4 contains *no* plate artifacts.) Figure 6 is quite similar to Figure 4, and large scale features can be followed across plate boundaries in Figure 6. Thus, we have a visual demonstration that, as discussed by GP86*a*, the map of the corrected counts is indeed a useful representation of the galaxy distribution.

V. TWO-POINT CORRELATION FUNCTION FOR THE SIMULATED COUNTS

As already noted in § II, the width of the plate correction factor distribution and the width of the plate correction factor error distribution are sufficiently small that the correction terms in the GP77 method of estimating the correlation function are negligibly small and simpler methods give the same result for the correlation function. Since we have constructed high-fidelity simulations, the widths of these distributions in the simulations are also negligibly small, and we need not reconstruct the full GP77 apparatus in order to estimate the

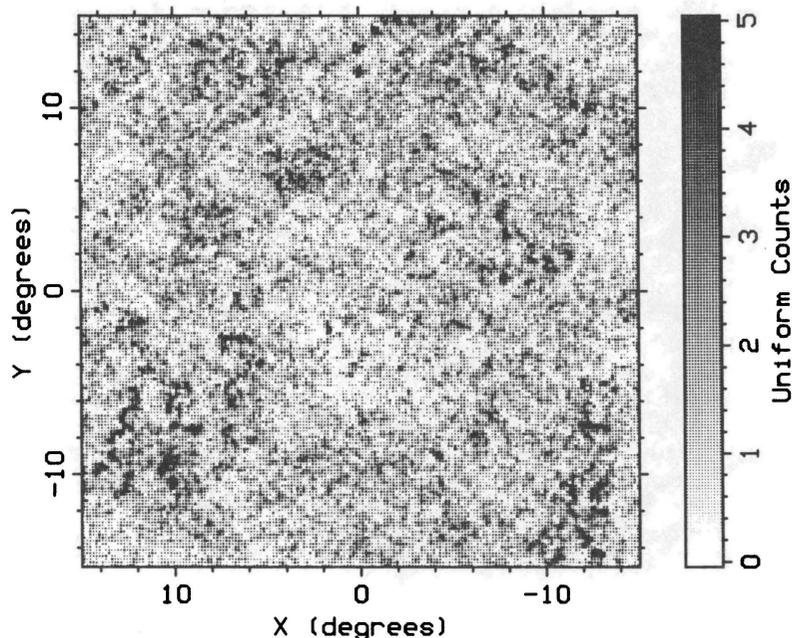


FIG. 4.—Simulated ideal portion of the Lick map. The space distribution of galaxies is projected onto the sky with a uniform limiting magnitude. That is, there are no plate artifacts in this map. Shown is a $30^\circ \times 30^\circ$ portion of the map with $10' \times 10'$ pixels. These data are from simulation 1.

correlation function of the simulated data. Instead, we estimate the correlation function using polar maps of the simulated counts and Fourier transform techniques. This method of estimating the correlation function is described in GP86*a* and GP86*b*. For each simulation, we estimate the correlation function twice, once for counts projected onto the sky with a uniform limiting magnitude, as in Figure 4, and once for the corrected counts, as in Figure 6. The former we call the “true” correlation function and the latter is the “test” function. In all

cases, we use only those portions of the maps with $b > +40^\circ$.

Figures 7, 8, and 9 show the true and test correlation functions for simulations 1, 2, and 3. There are no error bars in these figures, since in each case we have only one estimate of the correlation function. However, the question we are addressing is not the dispersion in correlation functions one might expect from different samples of the universe, but how well the test functions reproduce the true functions. As can be seen, the test functions track the true functions extremely well.

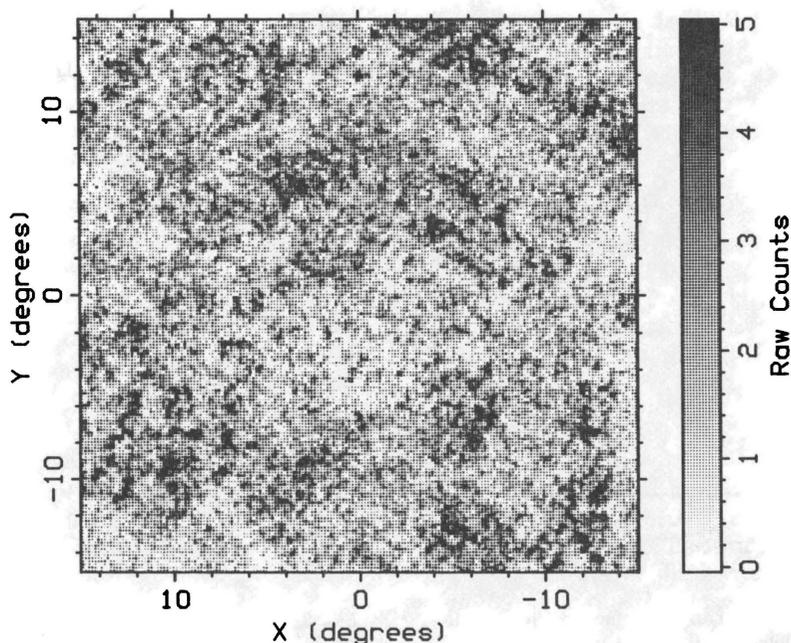


FIG. 5.—Simulated raw counts. These are the same data as shown in Fig. 4 except the simulated variation in limiting magnitude has been applied.

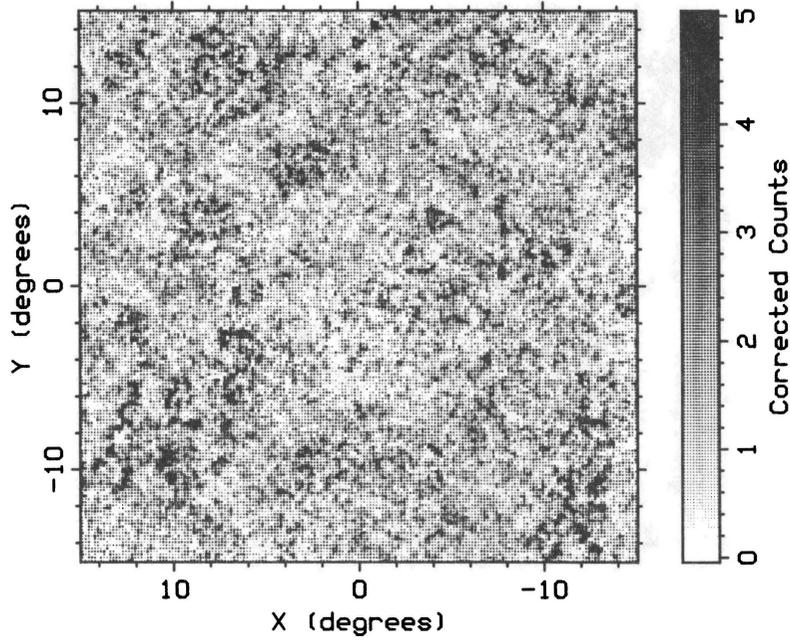


FIG. 6.—Simulated corrected counts. These are the same data as shown in Fig. 5 except the counts have been corrected by the estimated field, plate, and smoothing factors. Compare with Fig. 4.

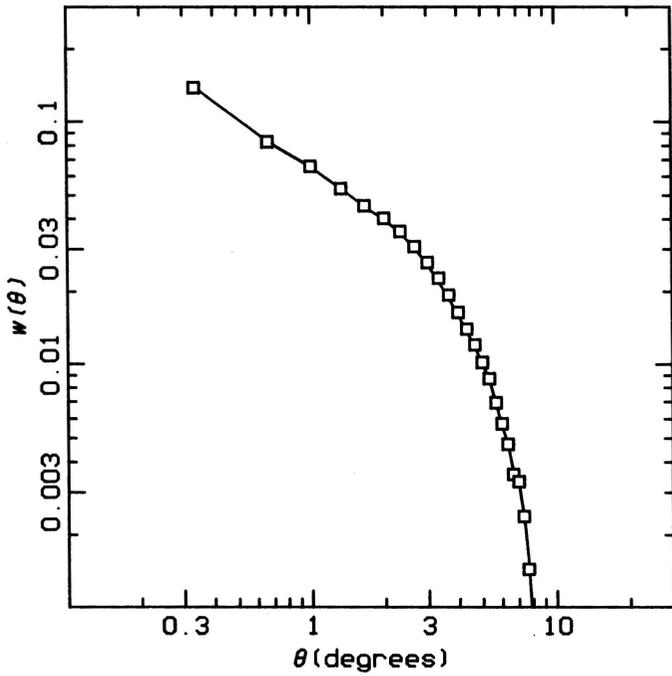


FIG. 7.—Comparison of the true and test correlation functions from simulation 1. The solid line connects estimates of the true correlation function, while the squares are the estimates of the test function.

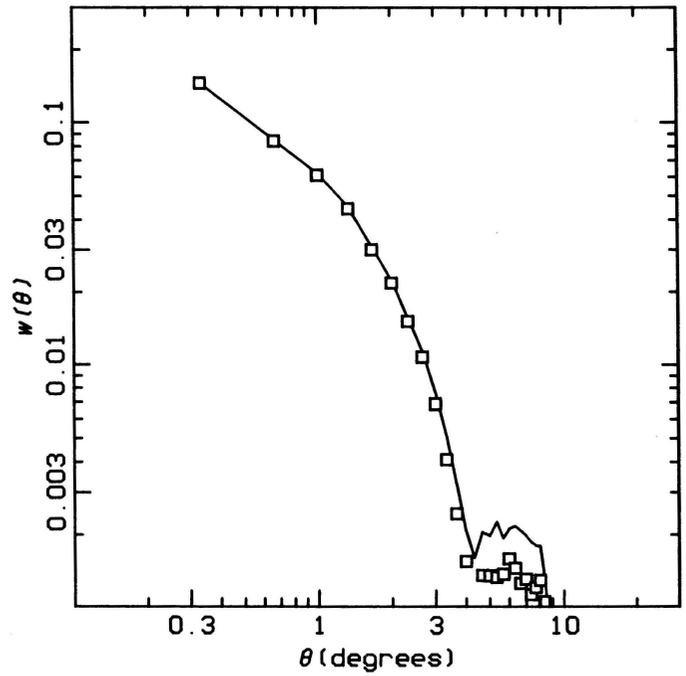


FIG. 8.—Comparison of the true and test correlation functions from simulation 2. The solid line connects estimates of the true correlation function, while the squares are the estimates of the test function.

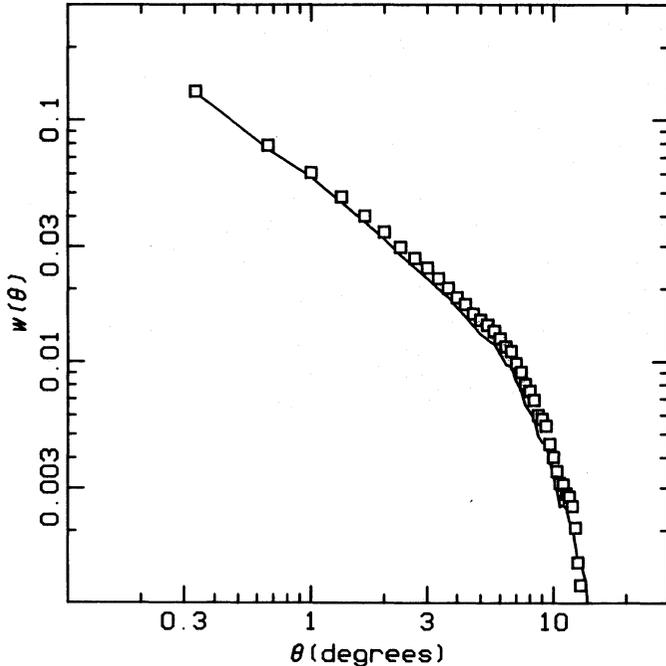


FIG. 9.—Comparison of the true and test correlation functions from simulation 3. The solid line connects estimates of the true correlation function, while the squares are the estimates of the test function.

VI. PLATE SHAPE GRADIENTS

The field factors are used to correct for systematic variations in limiting magnitudes over a plate. Presumably these variations are due to effects such as vignetting, distortions, and scattering. However, dLKG proposed that some of the systematic variation is due to observer (Shane or Wirtanen) differences and that these observer variations are a systematic function of time. DLKG termed these systematic observer and time-dependent variations “plate shape gradients.” They further proposed that these systematic effects would invalidate the GP77 estimate of the galaxy correlation function.

GP86b showed that plate shape gradients could not seriously effect the GP77 analysis for the following reasons:

1. Contrary to the analysis of dLKG, the plate shape gradient effects are not statistically significant. In order to estimate the plate shape (that is, a set of field factors), the galaxy counts on the plates counted by a particular observer during a particular time interval are averaged. Galaxy counts are not normally distributed, so it is inappropriate to use Gaussian statistics (as was done by dLKG) to estimate the significance of the resulting plate shapes. When proper statistics are used (as was done by GP86b) the statistical significance of the effect disappears.

2. Although plate shape gradients are not statistically significant, they might still exist in the data at a level that is below detectability. Since the SSGP procedure for estimating plate factors relies on the edges of the plates, the existence of plate shape gradients in the data will cause the corrected plates to sometimes have high centers and sometimes low centers. That is, the plate correction factor will correct the edges of a plate to agree with its neighbors but will under- or overcorrect the center of the plate since the field factors have generated plates in which the centers are too high or too low relative to the

edges. DLKG call this the “contrast stretch effect.” The details of the contrast stretch effect depend on the pattern in which plates were counted; one might expect the contrast stretch effect to be most severe when plates with high and low centers occur in a checkerboard pattern. GP86b were able to treat this case analytically and found that for the maximum amplitude plate shape gradient effect permitted by the data, there is no significant distortion of the galaxy correlation function.

3. GP86b divided the plates into the six observer and time interval samples used by dLKG and estimated the galaxy correlation function for each of the six samples. Of course, these correlation functions were much noisier than that for the whole sample. Nevertheless, they each agree well with the correlation function estimated by GP77 and their average tracked the GP77 estimate extremely well.

DLKG suggested that plate shape gradients—interacting with the SSGP edge matching procedure for estimating plate correction factors—are responsible for the large-scale gradients observed in the corrected counts. In effect, the existence of large-scale gradients is taken as evidence for the existence of plate shape gradients. However, dLKG provided no calculations to show how plate shape gradients generate large-scale gradients—their suggestion is simply speculation. GP86b showed that the SSGP edge-matching procedure naturally generates large-scale gradients due to the propagation of fluctuations in the galaxy distribution into the plate correction factors. However, as discussed in the § IV, GP86b may have overestimated the amplitude of the large-scale gradients resulting from a given amplitude of fluctuations in the galaxy distribution. Finally, one must always keep in mind the possibilities that the large-scale gradients in the counts are due to inadequacies in the simple models of atmospheric and galactic absorption or may represent true large-scale gradients in the galaxy distribution. In any event, the large-scale gradients were removed by GP77 before estimating the galaxy correlation function, so they are of no direct consequence for the estimates of the correlation function.

The main thrust of the present paper is to show that when the galaxy distribution, the plate counting, the SSGP reduction procedure, and the GP77 analysis procedure, are all simulated with sufficiently high fidelity, then these procedures recover the galaxy correlation function. This is in contrast to dLKG whose simulations did not recover the correlation function built into the simulated galaxy distribution. GP86b pointed out the main areas where the dLKG simulations were lacking in fidelity and showed in one case that if the dLKG simulations were corrected for the extra noise induced by the lack of fidelity, the correlation function would be recovered.

Since our goal is to validate the reduction and analysis methods using simulations, and since the plate shape gradients have been shown not to exist or at least to be too small to affect the correlation function, we have not included plate shape gradients in the simulations discussed in the previous sections. (We note that dLKG did not include plate shape gradients in their simulations.) Nevertheless, we have essentially all the tools necessary to simulate the effects of plate shape gradients on the GP77 estimate of the correlation function.

Of course, since the plate shape gradients are not detectable at a statistically significant level, one has to deal with the question of simulating a noneffect. We have adopted the following approach. We use the same six observer and time interval subdivisions as dLKG. For each subdivision, we stack all the raw Shane-Wirtanen plates in order to generate a set of field

TABLE 2
OBSERVER AND TIME INTERVAL SAMPLES

Sample Name	Observer	Time Interval	Number of Plates	Figure Number
S1	Shane	1947–1951	382	10a
S2	Shane	1952–1955	249	10b
S3	Shane	1956–1959	181	10c
W1	Wirtanen	1947–1951	98	10d
W2	Wirtanen	1952–1955	148	10e
W3	Wirtanen	1956–1959	188	10f

factors for that sample. We then construct a fourth simulation which is identical to simulation 1 (including the same random numbers) except that the limiting magnitude in a cell is determined by the plate factor (which is the same as in simulation 1) and one of the six sets of field factors. The set corresponds to the actual observer and time interval in which the plate at the position was counted. Thus the pattern of counting in the simulated catalog is the same as that in the actual Lick catalog.

The observer and time intervals are listed in Table 2. The two plates counted by Mayall are included with Shane's plates. Figure 10 shows the field factors for the various samples. Figures 10a–10c show those corresponding to Shane's plates; Figures 10d–10f show those corresponding to Wirtanen's plates. Note that these are all based on the actual Lick catalog and are input to simulation 4. Figure 10g shows the field factors determined from simulation 4. Although six sets of field factors are input, SSGP used only one set of factors, so the simulation determines only one set of factors and uses them to correct all the data. For comparison, Figure 10h shows the field factors determined from simulation 1, and Figure 10i shows those used in the original SSGP reduction.

There are substantial variations among the six sets of field factors for the observer and time interval samples. In particular, sample W1 (Fig. 10d) appears substantially different from the others. This sample contains only 98 plates. The variations seen in these samples are due to intrinsic fluctuations in the galaxy distribution and possibly due to systematic variations in counting sensitivity. According to GP86b (who used only high-latitude plates, as did dLKG), the variations due to fluctuations in the galaxy distribution mask any possible variations due to counting sensitivity. We have not repeated the GP86b analysis of the significance of the variations for these samples. This is because it is difficult to characterize the fluctuations in the galaxy distribution when the sample contains all latitudes and hence a wide range of effective depths.

Another point to be made about the six field factor samples is that they are substantially noisier than the field factors for the entire catalog. When we construct a simulation based on these field factors we are adding considerable noise to the simulation. The fidelity of the simulation is lowered and effects which are driven by the input noise are increased. An exception can be seen in Figure 10. The field factors estimated from simulation 4 are almost identical to those estimated from simulation 1. Of course, this is because when the field factors for the

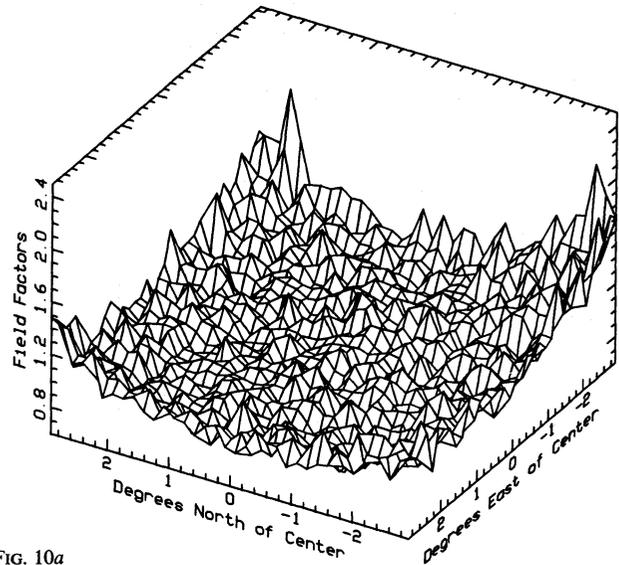


FIG. 10a

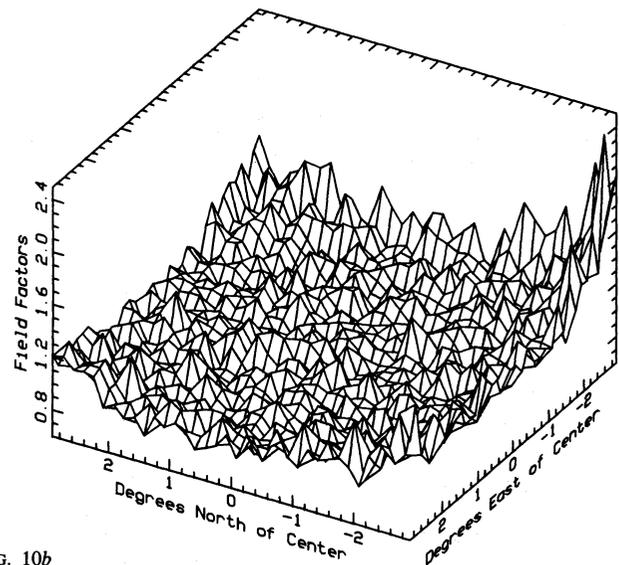


FIG. 10b

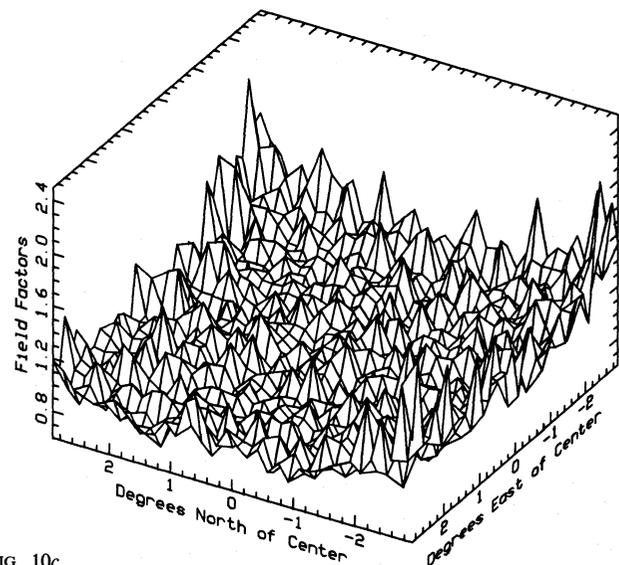


FIG. 10c

FIG. 10.—Field factors (*plates shapes*) determined from various samples of actual or simulated data. (a) Plates counted by Shane during 1947–51. (b) Plates counted by Shane during 1952–55. (c) Plates counted by Shane during 1956–59. (d) Plates counted by Wirtanen during 1947–51. (e) Plates counted by Wirtanen during 1952–55. (f) Plates counted by Wirtanen during 1956–59. (g) Simulation 4. (h) Simulation 1. (i) The entire Lick catalog.

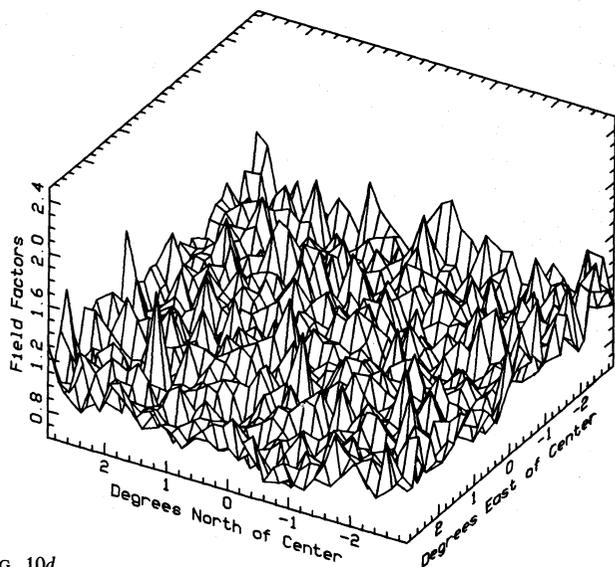


FIG. 10d

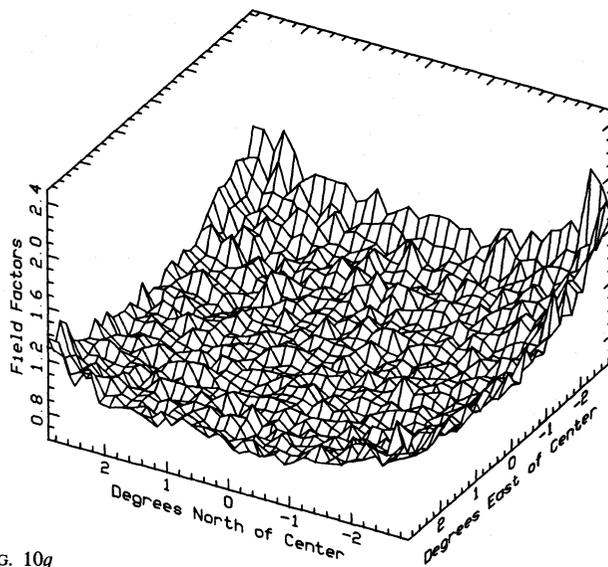


FIG. 10g

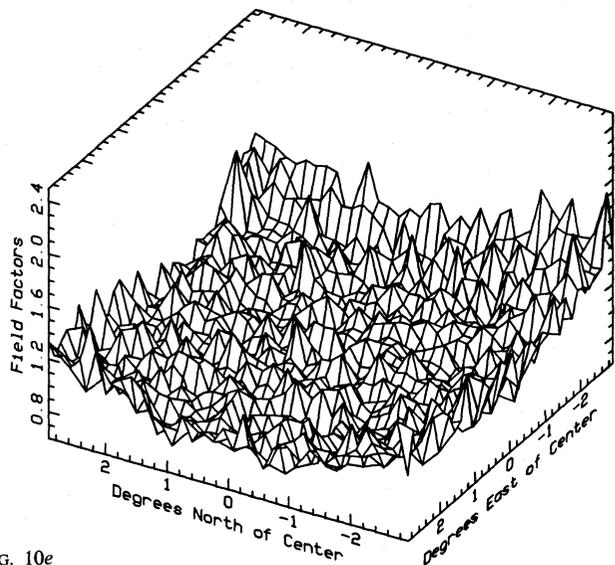


FIG. 10e

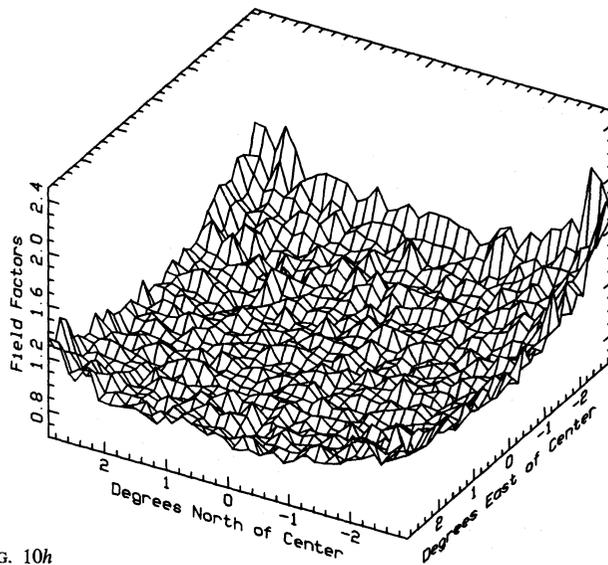


FIG. 10h

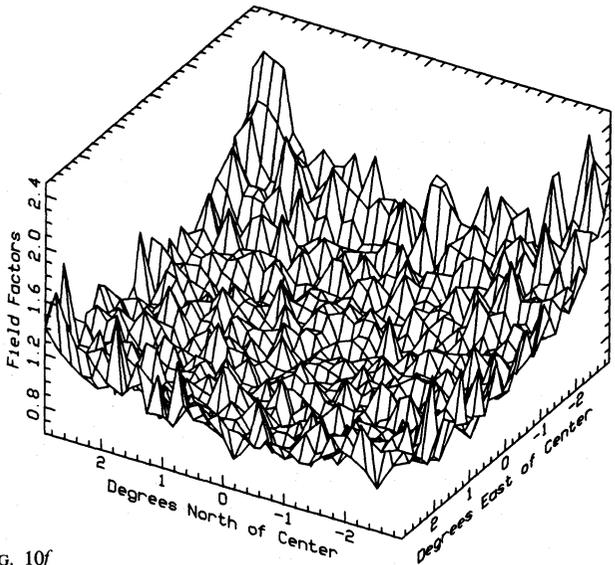


FIG. 10f

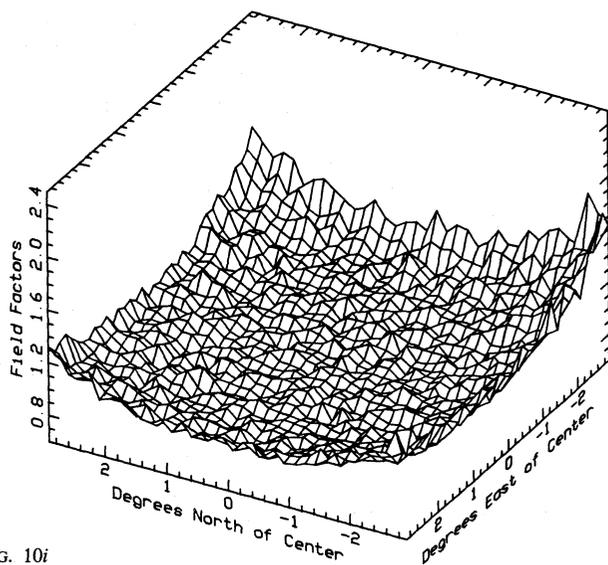


FIG. 10i

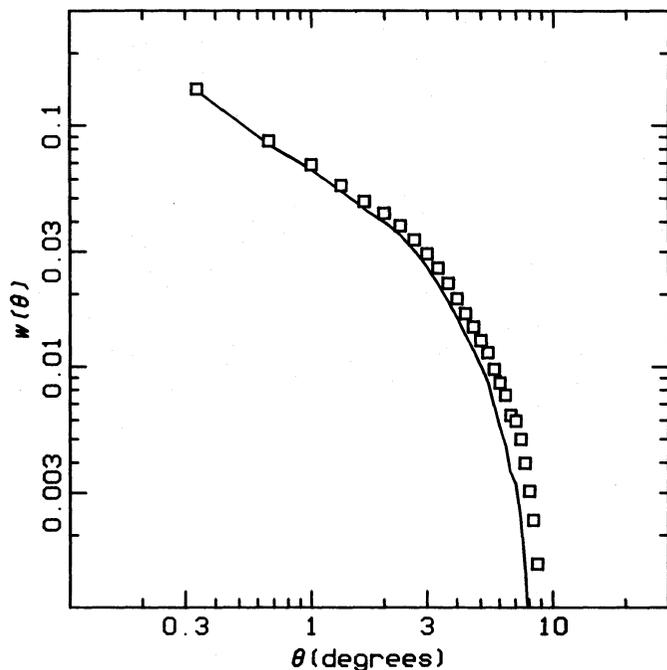


FIG. 11.—Comparison of the true and test correlation functions from simulation 4. The solid line connects estimates of the true correlation function, while the squares are the estimates of the test function.

six subsamples are averaged (weighted by the number of galaxies in each subsample) they yield the field factors for the full sample (Fig. 10i).

We now consider the results of simulation 4. The plate factors estimated from simulation 4 are similar to those from simulation 1, but the root mean square fractional difference between the estimated and true plate factors, for $|b| > 40^\circ$, increases from 6.3% to 8.0%. This is an example of the added noise in the simulation propagating into the errors in quantities estimated from the data in the simulation. Maps constructed from simulation 4, similar to the maps shown for simulation 1 in Figures 5 and 6 show no obvious differences from the simulation 1 maps.

The large-scale gradients, determined by the method described in § IV, actually decrease from a root mean square amplitude of 4.2% in simulation 1 to 3.2% in simulation 4. This would seem to indicate that the large-scale gradients in the Lick catalog are not the result of systematic variations in plate shapes—contrary to the speculation of dLKG.

Figure 11 shows the true and test correlation functions for simulation 4. It can be seen that the test correlation function lies above the true correlation function by ~ 0.003 . The extra noise from the six sets of simulated field factors is responsible for this increase. Nevertheless, the difference between the true and test correlation function is very slight—much less than the zone-to-zone differences in the original GP77 estimates of the correlation function and much less than the error bars shown in GP77. We conclude, again, that plate shape gradients have had no significant effect on the GP77 estimates of the galaxy correlation function.

VII. DISCUSSION

We have shown that, contrary to the assertions of dLKG, the GP77 method of estimating the galaxy two-point correla-

tion function from the Lick counts as reduced by SSGP is in fact capable of recovering the true correlation function. The simulations we generate and analyze are of sufficiently high fidelity that all the noise sources which effect the determination of the correlation function are present in the simulations at about the same level as they are present in the Lick counts. Furthermore, the method does not suffer from any strange anomaly which always produces a break in the correlation function at 2.5 , since it recovers breaks in the correlation function at the location of the observed break and at both larger and smaller angles.

Our simulations follow the entire counting, reduction, and analysis procedure with only small infidelities. The width of the simulated plate correction factor distribution is slightly wider than the actual width as described in § III*d*. This just means that the reduction and analysis methods are subject to a slightly more rigorous test when applied to the simulated counts than when applied to the actual counts. The second area where the simulations depart from the actual method used by SSGP and GP77 concerns the treatment of the atmospheric absorption, galactic absorption, and smoothing factors. Rather than compute three separate factors for each plate, we compute a single factor for each plate which is designed to remove the large-scale gradients in the counts. Since GP77 applied the smoothing factors after applying the absorption factors, and since the absorption factors themselves are smooth, essentially the same corrected counts are obtained whether or not the absorption factors are applied as an intermediate step. The other area in which the simulations differ from the actual GP77 method is that the simpler Fourier transform method, rather than the elaborate GP77 method, is used to calculate the actual correlation function of the simulated counts. Again, this is of no consequence as it has already been shown by Fry and Seldner (1982), GP86*a*, and GP86*b* that the two methods give negligibly different results when the widths of the plate correction factor and error distributions are as small as they are. We also note that the original GP77 software (and SSGP software) was created in the mid 1970s on punched cards. These cards have long since turned to mush, so it is not possible to rerun the original software on the simulated data.

We have concentrated on whether the SSGP and GP77 methods can recover the correlation function. GdLK and dLKG raised other questions concerning the reliability of the GP77 correlation function. Some of these have to do with systematic errors induced by the reduction and analysis methods, and some have to do with systematic errors in the original counting of the plates. GP86*a* and GP86*b* addressed all the concerns and showed that the suggested systematic errors are either not present or too small to have affected the results. Our simulation 4 shows that systematic errors in the original counting are not an important source of error. This simulation includes whatever plate shape gradients are present in the Lick catalog as well as extra noise due to the need to estimate six different plate shapes, yet it too shows that the estimated correlation function reproduces the built in correlation function extremely well. Our simulations reinforce the GP86*a* and GP86*b* conclusions concerning systematic errors induced by the reduction and analysis methods. For example, we have shown that the magnitudes of the errors in the field and plate correction factors are as originally estimated (§ IV*a, b*). We have shown that the SSGP edge matching procedure can induce large-scale gradients in the counts but the amplitude of these gradients is about half that estimated by

GP86b (§ IVc). We also have a visual demonstration of the utility of the Lick map (§ IVd).

Of course, the best way to verify an observational result is through independent observations. We note that the first results from the APM Galaxy Survey of the Southern galactic hemisphere are becoming available (Maddox, Efstathiou, and Loveday 1988). The galaxy two-point angular correlation function estimated from this survey, when scaled to the depth of the Lick survey, is indistinguishable from the GP77 result. The amplitude and power-law index at angles smaller than the

break are the same; the location of the break is the same; and even the slope of the correlation function longward of the break is the same.

This work is based partly on the Princeton University Senior Thesis of M. E. B. We thank Jim Peebles for a careful reading of this manuscript and several helpful suggestions. This research was supported in part by NASA (contract NAS5-29142) and the NSF (grant PHY 8419908).

REFERENCES

- Abell, G. O. 1962, in *Problems in Extragalactic Research*, ed. G. C. McVittie (New York: Macmillan), p. 213.
- Davis, M., Groth, E. J., and Peebles, P. J. E. 1977, *Ap. J. (Letters)*, **212**, L107.
- de Lapparent, V., Geller, M. J., and Huchra, J. R. 1986, *Ap. J. (Letters)*, **302**, L1.
- de Lapparent, V., Kurtz, M. J., and Geller, M. J. 1986, *Ap. J.*, **304**, 585 (dLKG).
- Fry, J. N., and Seldner, M. 1982, *Ap. J.*, **259**, 474.
- Geller, M. J., de Lapparent, V., and Kurtz, M. J. 1984, *Ap. J. (Letters)*, **287**, L55 (GdLK).
- Groth, E. J., and Peebles, P. J. E. 1977, *Ap. J.*, **217**, 385 (GP77).
- . 1986a, *Ap. J.*, **310**, 499 (GP86a).
- . 1986b, *Ap. J.*, **310**, 507 (GP86b).
- Haynes, M. P., and Giovanelli, R. 1986, *Ap. J. (Letters)*, **306**, L55.
- Kirshner, R. P., Oemler, A., Schechter, P. L., and Shectman, S. A. 1981, *Ap. J. (Letters)*, **248**, L57.
- Maddox, S. J., Efstathiou, G., and Loveday, J. 1988, in *Evolution of Large-Scale Structures in the Universe*, in press.
- Peebles, P. J. E. 1973, *Ap. J.*, **185**, 413.
- Seldner, M., Siebers, B., Groth, E. J., and Peebles, P. J. E. 1977, *A.J.*, **82**, 249 (SSGP).
- Shane, C. D., and Wirtanen, C. A. 1967, *Pub. Lick Obs.*, Vol. **22**, Part 1.
- Soneira, R. M., and Peebles, P. J. E. 1978, *A.J.*, **83**, 845 (SP).

MICHAEL E. BROWN and EDWARD J. GROTH: Joseph Henry Laboratories, Physics Department, Princeton University, P.O. Box 708, Princeton, NJ 08544