THE ASTROPHYSICAL JOURNAL, 492:439–451, 1998 January 10

NONLINEAR CONSTRAINED REALIZATIONS OF THE LARGE-SCALE STRUCTURE

V. BISTOLAS AND Y. HOFFMAN
Racah Institute of Physics, Hebrew University, Jerusalem, Israel
Received 1995 August 1; accepted 1997 August 18

ABSTRACT

The linear algorithm of the Wiener filter and constrained realizations (CRs) of Gaussian random fields is extended here to perform nonlinear CRs. The procedure consists of (1) using low-resolution data to constrain a high-resolution realization of the underlying field, as if the linear theory were valid; (2) taking the linear CR backward in time, by the linear theory, to set initial conditions for N-body simulations; (3) forwarding the field in time by an N-body code. An intermediate step is introduced to "linearize" the low-resolution data.

The nonlinear CR can be applied to any observational data set that is quasi-linearly related to the underlying field. Here it is applied to the *IRAS* 1.2 Jy catalog using 846 data points within a sphere of 6000 km s^{-1}, to reconstruct the full nonlinear large-scale structure of our "local" universe. The method is tested against mock *IRAS* surveys, taken from random nonlinear realizations. A detailed analysis of the reconstructed nonlinear structure is presented.

Subject headings: large-scale structure of universe — methods: statistical

1. INTRODUCTION

In the standard model of cosmology, galaxies and the large-scale structure (LSS) of the universe form out of a random perturbation field via gravitational instability. It is assumed that the primordial perturbation field constitutes a random homogeneous and isotropic Gaussian field and that on relevant scales its amplitude is small, hence its dynamics is described by the linear theory of gravitational instability (cf. Peebles 1980). The theoretical study of structure formation has been a major effort of modern cosmology (cf. Padmanabhan 1993). On the observational side, the large-scale structure has been studied mostly by means of redshift surveys (cf. Strauss & Willick 1995) and peculiar velocities (cf. Dekel 1994). A method for the reconstruction of the underlying dynamical (density and velocity) fields from a given observational database is presented here.

The problem of recovering the underlying field from given observations, which by their nature are incomplete and have finite accuracy and resolution, is one often encountered in many branches of physics and astronomy. It has been shown that for a random Gaussian field an optimal estimator of the underlying field is given by a minimum variance solution (Zaroubi et al. 1995, hereafter ZHFL), known also as the Wiener filter (WF) (Wiener 1949; Press et al. 1992). This approach is based on the assumed knowledge of the second moments of the random field. These moments, also known as covariance matrices, are to be deduced directly from the data or calculated from an assumed model, the so-called *prior*. Within the framework of Gaussian fields, the WF coincides with the Bayesian *posterior* and the maximum entropy estimations (ZHFL). Indeed, in the cosmological case where on large enough scales the linear theory applies and the (over)density and velocity fields are Gaussian, the WF is the optimal tool for the reconstruction of the large-scale structure. This is further complemented by the algorithm of constrained realizations (CRs) of Gaussian fields (Hoffman & Ribak 1991) to create Monte Carlo simulations of the residual from this optimal estimation. This combined WF/CR approach has been applied recently to a variety of cosmological databases in an effort to recover the large-scale structure. This includes the analysis of the *COBE*/DMR data (Bunn et al. 1994; Bunn, Hoffman, & Silk 1996), the analysis of the velocity potential (Ganon & Hoffman 1993), the reconstruction of the density field (Hoffman 1993, 1994; Lahav 1993, 1994; Lahav et al. 1994; Webster, Lahav, & Fisher 1996) and the peculiar-velocity field (Fisher et al. 1995b) from the *IRAS* redshift survey (Fisher et al. 1993). It has also been recently applied to the MARK III peculiar velocities (Willick et al. 1995) to reconstruct the underlying dynamical fields (Zaroubi, Hoffman, & Dekel 1997).

A major limitation of the WF/CR approach is that it applies only in the linear regime. Yet on small scales the perturbations are not small, and the full nonlinear gravitational instability theory has to be used. Here the WF/CR method is extended to the quasi-linear regime, and a new algorithm of nonlinear constrained realizations (NLCRs) is presented. The general method of WF and CRs and its modification to the case of quasi-linear data is presented in § 2. The method is tested against N-body simulations, and its application to the *IRAS* 1.2 Jy catalog (Fisher et al. 1995a) is given in § 3. The NLCRs of our "local" universe are presented in § 4 and a short discussion (§ 5) concludes the paper.

2. NONLINEAR CONSTRAINED REALIZATIONS

2.1. *Linear Theory*

The general WF/CR method has been fully described in ZHFL, and only a very short outline of it is presented here. Consider the case of a set of observations performed on an underlying random field (with N degrees of freedom) $\boldsymbol{s} = \{s_1, \ldots, s_N\}$ yielding M observables, $\boldsymbol{d} = \{d_1, \ldots, d_M\}$. Here, only measurements that can be modeled as a linear convolution or mapping on the field are considered. The act of observation is represented by

$$\boldsymbol{d} = \mathscr{R}\boldsymbol{s} + \boldsymbol{\epsilon}\,, \tag{1}$$

where $\mathscr{R}$ is a linear operator that represents a point-spread function, and $\boldsymbol{\epsilon} = \{\epsilon_i, \ldots, \epsilon_M\}$ represents the statistical

errors. Here the notion of a point-spread function is extended to include any linear operation that relates the measurements to the underlying field. The WF estimator is (ZHFL)

$$\boldsymbol{s}^{\mathrm{WF}} = \langle \boldsymbol{s}\boldsymbol{d}^{\dagger}\rangle \langle \boldsymbol{d}\boldsymbol{d}^{\dagger}\rangle^{-1}\boldsymbol{d} \,. \tag{2}$$

Here the angular brackets represent an ensemble average and $\langle \boldsymbol{s}\boldsymbol{d}^{\dagger}\rangle$ is the cross-correlation matrix of the data and the underlying field. The data autocorrelation matrix is

$$\langle \boldsymbol{d}\boldsymbol{d}^{\dagger}\rangle = \mathscr{R}\langle \boldsymbol{s}\boldsymbol{s}^{\dagger}\rangle \mathscr{R}^{\dagger} + \langle \boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\dagger}\rangle \,, \tag{3}$$

where the second term represents the statistical errors, i.e., shot noise.

In the case of a random Gaussian field, the WF estimator coincides with the conditional mean field given the data. A CR of the random residual from this mean field is obtained by creating an unconstrained realization of the underlying field ($\tilde{\boldsymbol{s}}$) and the errors ($\tilde{\boldsymbol{\epsilon}}$), and "observing" it in the same way the actual universe is observed. Namely, a mock database is created by

$$\tilde{\boldsymbol{d}} = \mathscr{R}\tilde{\boldsymbol{s}} + \tilde{\boldsymbol{\epsilon}} . \tag{4}$$

A CR is then simply obtained by (Hoffman & Ribak 1991)

$$\boldsymbol{s}^{\mathrm{CR}} = \tilde{\boldsymbol{s}} + \langle \boldsymbol{s}\boldsymbol{d}^{\dagger}\rangle \langle \boldsymbol{d}\boldsymbol{d}^{\dagger}\rangle^{-1}(\boldsymbol{d} - \tilde{\boldsymbol{d}}) \,. \tag{5}$$

2.2. *Covariance Matrices and Shot Noise*

The WF/CR and the NLCR presented here can be used with any database whose relation to the underlying field can be modeled by equation (1). Thus, for example, observations of the velocity field can be used to reconstruct the density field and vice versa. The concrete case studied here is the reconstruction of the continuous density field from a discrete galaxy catalog within a given volume subject to certain selection criteria. Here we assume that the galaxy distribution in configuration space is known, and redshift distortions are ignored. The formalism of WF/CR can be expressed in any functional representation, such as Fourier or spherical harmonics/Bessel functions. The choice of the particular representation is usually dictated by the geometry of the observations. In the case of a full sky coverage and radial selection function, the obvious choice is the spherical harmonics/Bessel basis (Fisher et al. 1995b; Webster et al. 1996). However, in the case of an incomplete sky survey the so-called zone of avoidance (ZOA) couples the spherical harmonics, and this results in a complicated covariance matrix. For the general case we choose here to use the configuration space representation, and the field is evaluated on a Cartesian grid. The estimation of the continuous field is done by smoothing, i.e., the convolution of the discrete galaxy distribution with a certain kernel. The discreteness of the galaxies introduces shot-noise errors, and the smoothing procedure causes these errors to be correlated with each other. The smoothing kernel depends on the nature of the data and is determined by a compromise between two conflicting considerations, namely, high resolution and low noise level. A high resolution is achieved by using a narrow smoothing kernel, and the noise level is reduced by widening the kernel. Here a Gaussian filter is used with two smoothing-length radii, R_L and R_S. The data are smoothed on a large scale, R_L, and high-resolution CRs are created on the scale R_S; thus $R_L > R_S$.

An estimator of the fractional overdensity at the point $\boldsymbol{r}_{\alpha}$ is given by

$$\Delta_{\alpha} = \Delta(\boldsymbol{r}_{\alpha}) = \frac{1}{\bar{n}(2\pi R_L^2)^{3/2}} \sum_{\mathrm{gal}} \frac{1}{\phi(\boldsymbol{r}_{\mathrm{gal}})} \times \exp\left[-\frac{(\boldsymbol{r}_{\alpha} - \boldsymbol{r}_{\mathrm{gal}})^2}{2R_L^2} \right] - 1 \,, \tag{6}$$

where $\bar{n}$ is the mean number density of the galaxies and $\phi(r)$ is the data selection function. The data autocorrelation function is written as $\langle \Delta_{\alpha}\Delta_{\beta}\rangle = \xi_{\alpha\beta} + \sigma_{\alpha\beta}$. The first term is just the autocorrelation function of the smoothed field $[\xi^s(r)]$,

$$\xi_{\alpha\beta} = \xi^s(|\boldsymbol{r}_{\alpha} - \boldsymbol{r}_{\beta}|) = \frac{1}{(2\pi)^3} \int P(k) \exp\left[-(kR_L)^2\right] \times \exp\left[i\boldsymbol{k}\cdot(\boldsymbol{r}_{\alpha} - \boldsymbol{r}_{\beta})\right] d^3k \,, \tag{7}$$

and the shot-noise covariance matrix is

$$\sigma_{\alpha\beta} = \frac{1}{\bar{n}(2\pi R_L^2)^3} \int \frac{1}{\phi(x)} \exp\left[-\frac{(\boldsymbol{r}_{\alpha} - \boldsymbol{x})^2 + (\boldsymbol{r}_{\beta} - \boldsymbol{x})^2}{2R_L^2} \right] d^3x \,. \tag{8}$$

(The derivation of the error matrix follows Scherrer & Bertschinger 1991.) Note that the kernel introduces off-diagonal terms in the error covariance matrix (ZHFL). The cross-correlation of the high-resolution field and the low-resolution data is

$$\xi_{\alpha}(\boldsymbol{r}_i) = \frac{1}{(2\pi)^3} \int P(k) \exp\left[-\frac{(kR_S)^2 + (kR_L)^2}{2} \right] \times \exp\left[i\boldsymbol{k}\cdot(\boldsymbol{r}_i - \boldsymbol{r}_{\alpha})\right] d^3k \,. \tag{9}$$

Defining the WF operator $W_{i\beta}$,

$$W_{i\beta} = \xi_{\alpha}(\boldsymbol{r}_i)(\xi_{\alpha\beta} + \sigma_{\alpha\beta})^{-1} \,, \tag{10}$$

a linear high-resolution realization is now obtained by

$$\delta(\boldsymbol{r}_i) = \tilde{\delta}(\boldsymbol{r}_i) + W_{i\beta}(\Delta_{\beta} - \tilde{\Delta}_{\beta}) \,. \tag{11}$$

2.3. *Quasi-linear Approximation*

In the bottom-up model of gravitational clustering, small scales go nonlinear before large ones. Hence, our basic approach here is to smooth the data on a scale R_L which is large enough that the δ-field, smoothed on that scale, is approximately linear. Given an estimator of such a linear observable, the linear formalism of § 2.1 can be used to make high-resolution CRs, as if the linear theory is valid on these small scales. However, as will be shown below, numerical N-body simulations show that even for quite high R_L smoothing, the resulting field has undergone some nonlinear evolution. Thus, the linear procedure of the WF/CR has to be supplemented by an additional step of "linearizing" the input data, i.e., mapping the present epoch data back to the linear regime. Various algorithms have been proposed to trace back the nonlinear perturbation field to the linear regime (cf. Strauss & Willick 1995). All of these "time machines" recover the initial linear field in the

case where the quasi-linear field is known *exactly*, with no statistical uncertainty. The case of real observational data, where the shot noise increases with distance, poses a much more difficult problem. As the density field is sampled farther away, it becomes more dominated by the shot noise and in the mean its amplitude increases with distance. Thus, a procedure has to be developed that accounts for the statistical noise, separately from the nonlinear effects.

The analysis of the various nonlinear effects and the test of possible approximations to the mapping from the nonlinear to the linear regime is best done by an analysis of N-body simulations of nonlinear gravitational clustering. As our aim here is to apply the method to the *IRAS* 1.2 Jy catalog, N-body simulations with the *IRAS* power spectrum have been performed. Mock *IRAS* data sets are generated from these simulations. (A full description of the simulations is given in § 3.) The nonlinear evolution has been studied by smoothing the fully nonlinear field, $\delta^{\rm NL}$, and the linear field, δ^L, on the R_L scale. An examination of the ($\delta^{\rm NL}$, δ^L) relation (Fig. 5*a*) shows both a scatter and a systematic deviation from the desired $\delta^L = \delta^{\rm NL}$. The empirical fix to the "nonlinearity" of the smoothed data that are used here consists of two steps. First, to account for the scatter in the ($\delta^{\rm NL}$, δ^L) relation, a new term is introduced to the data autocovariance matrix, $\sigma^{\rm NL}$. Dealing with the scatter by statistical means is a manifestation of our inability to invert the exact nonlocal nonlinear mapping from the linear to the quasi-linear regime. Here we go to the extreme simplification and take $\sigma^{\rm NL}$ to be a scalar matrix. The value of this constant term is found to be of the order of 10^{-2}, and it is determined by the requirement that $\chi^2/{\rm dof} = 1$, where $\chi^2 = \boldsymbol{d}^\dagger \langle \boldsymbol{d}\boldsymbol{d}^\dagger \rangle^{-1} \boldsymbol{d}$ takes into account the theoretical variance, shot noise, and $\sigma^{\rm NL}$. A WF estimator of the R_L-smoothed field is obtained by applying a WF to the data, where R_S is replaced by R_L to obtain low resolution,

$$\delta^{\rm WF,QL}(\boldsymbol{r}_i) = (W_{i\alpha})_{R_S = R_L} \Delta_\alpha . \tag{12}$$

The estimation of the quasi-linear correction is given by $[\delta^{\rm WF,QL} - f(\delta^{\rm WF,QL})]$, where $f(\delta)$ is a polynomial fitting to the curve shown in Figure 5*a*. This correction is evaluated at grid points $\boldsymbol{r}_\alpha$ and is used to correct the data points:

$$\Delta^L_\alpha = \Delta_\alpha - [\delta^{\rm WF,QL} - f(\delta^{\rm WF,QL})] \ . \tag{13}$$

The modified ("linearized") Δ^L_α are now substituted in equation (11) to obtain a high-resolution CR of the underlying linear field, given the actual data.

The ad hoc linearization procedure presented here behaves correctly in the following limits. In the case of distant data points, where the data are dominated by shot noise, the WF attenuates the estimated field toward zero amplitude. The linearization operation (eq. [13]) would hardly change the amplitude of the WF estimator (eq. [12]). The WF/CR is therefore dominated by the random residual, and consequently the resulting realization lies in the linear regime. For nearby data points where the shot noise is negligible, the WF leaves the signal almost untouched, $\delta^{\rm WF,QL} \approx \Delta$. The procedure described here correctly recovers the Gaussian 1 point distribution function of the density field; however, it will not correctly reconstruct the 2 point function. The procedure suggested here works only for data close to the linear regime.

The mock database of equation (4) that is used to sample the residual from the mean, has to be subjected to the same nonlinear effects as the real data do. Thus, the mock data $\tilde{\Delta}_\alpha$ have to be sampled from a mock *IRAS* catalog drawn out of an N-body simulation, and then to be mapped to the linear regime in the same way the actual data are treated. The linear CRs thus generated are used to set the initial conditions for N-body simulations.

3. TESTING AGAINST N-BODY SIMULATIONS

Within the linear regime the WF/CR method provides a rigorous way of estimating the underlying field and making realizations that are consistent both with the data and with the prior model. However, beyond the linear regime, one should use various approximations to achieve such a goal. In particular, the mapping of the quasi-linear data to the linear regime has to be calibrated by testing it against N-body simulations. This empirical approach strongly depends on the nature of the data and the assumed model, and therefore it should be tested on mock data sets that are drawn from the full N-body systems in a manner that mimics the actual observations. It follows that in contrast to the linear regime reconstruction, where a general WF/CR formalism can be formulated, in the quasi-linear domain the approach should be fitted to the problem at hand.

Our aim is to apply the present method to the *IRAS* 1.2 Jy redshift survey. The prior model assumed here is a flat, $\Omega_0 = 1$, cold dark matter (CDM)–like model with a shape parameter of $\Gamma = 0.2$, normalized by $\sigma_8 = 0.7$ and with no biasing. Four random linear realizations of this model were generated and used as initial conditions to a 128^3 particle mesh (PM) N-body code (written by E. Bertschinger) with a comoving box size of 1.6×10^4 km s^{-1}, periodic boundary conditions and grid spacing of 125 km s^{-1}. The low resolution is defined by a Gaussian filter of smoothing length $R_L = 1000$ km s^{-1} (here distances are given in units of kilometers per second), and the high resolution is limited by the Nyquist wavelength. The smoothed field is sampled within a sphere of radius 6000 km s^{-1} on a Cartesian grid at a sampling rate of R_L, yielding 925 constraints. The smoothing is done by a fast Fourier transform convolution on a count in cells (CIC) of the particles done on the basic grid. In the numerical experiments no Galactic ZOA is assumed.

A nontrivial task here is the calculation of the nondiagonal error correlation matrix. The symmetry of the problem leaves only three independent degrees of freedom out of the six ($\boldsymbol{r}_1$, $\boldsymbol{r}_2$) variables. The resulting integral is taken over a product of a Gaussian that peaks at $\sim(\boldsymbol{r}_1 + \boldsymbol{r}_2)/2$ and the inverse of the selection function, which makes it a three-dimensional integral. The integral is calculated numerically on a finite cube of size $6R_L$ centered at $\sim(\boldsymbol{r}_1 + \boldsymbol{r}_2)/2$. Special care has been taken to control the numerical errors, as the inverse of the autocorrelation matrix is found to be sensitive to the numerical noise. The matrix inversion is done by a Cholesky decomposition algorithm which is fast and stable (Press et al. 1992).

The numerical simulations provide an ensemble of four random nonlinear realizations of the "prior" model. The first step in analyzing the N-body simulation is the construction of the "galaxy" distribution out of the "dark matter" particles. No biasing is assumed here, and a fraction of the particles are randomly tagged as galaxies, so as to reproduce the *IRAS* mean number of galaxies. Here a mean number density of $n_g = 5.48 \times 10^{-8}$ (km s^{-1})$^{-3}$ (Fisher et al. 1994) is used. Given a "volume-limited"

mock *IRAS* catalog, a selection process is applied to generate a magnitude-limited mock *IRAS* 1.2 Jy catalog, based on the selection function

$$\phi(r) = \left(\frac{r}{r_s}\right)^{-2\alpha}\left(\frac{r^2 + r_*^2}{r_s^2 + r_*^2}\right)^{-\beta}, \quad (14)$$

with $I_s = 500$ km s^{-1}, $\alpha = 0.492$, $\beta = 1.830$, $r_* = 5184$ km s^{-1} (Yahil et al. 1991; Fisher 1992).

From the four different realizations, one is treated as the "real" data, and the other three are used to make the NLCRs. These are labeled *R* and 1, 2, and 3 in the various figures. The four different linear random realizations are shown in Figure 1. (In all figures the continuous δ-field is shown at the low resolution of $R_L = 1000$ km s^{-1} with contour spacing of 0.2 in δ). The *R* nonlinear realization has been sampled to produce a mock *IRAS* database, from which the CRs are generated. The three linear CRs, which serve as initial conditions to the PM code, are shown in Figure 2, where the *R* realization is given as well. The success of the whole procedure is actually determined at this stage, namely, by how close the various 1, 2, 3 fields are to the *R* field. The general "cosmography" is nicely recovered by the different CRs, with a little scatter. The smoothed density of the fully evolved NLCRs and the original *R* field are shown in Figure 3. Figure 4 shows the *R* volume-limited nonlinear "galaxy" distribution, which is to be recovered by the three NLCRs. The bottom panel shows the magnitude-limited "galaxy" distribution, using the *IRAS* selection function, from which the constraints are drawn. Various objects that dominate the "cosmography" of the simulations are tagged by capital letters A–F in Figures 2–4. The circle of radius 6000 km s^{-1} marks the region from which constraints are drawn.

The quality of the reconstruction provided by the NLCRs can be judged from Figures 3 and 4. The main objects that define the cosmography of the *R* simulation, within the sampling radius of $\approx$6000 km s^{-1}, are the A, B, D, and E clusters and the extended void that occupies most of the sampling volume (C). F denotes a filament running from A toward the object at (SGX, SGY) = (0, 8000), which is clearly seen in the particle distribution but is hardly noticed in the smoothed field. At the level of the smoothed field, the NLCRs (1–3) recovers the *R* realization to within one or two contour lines. No systematic discrepancy between the recovered fields and the original one is found

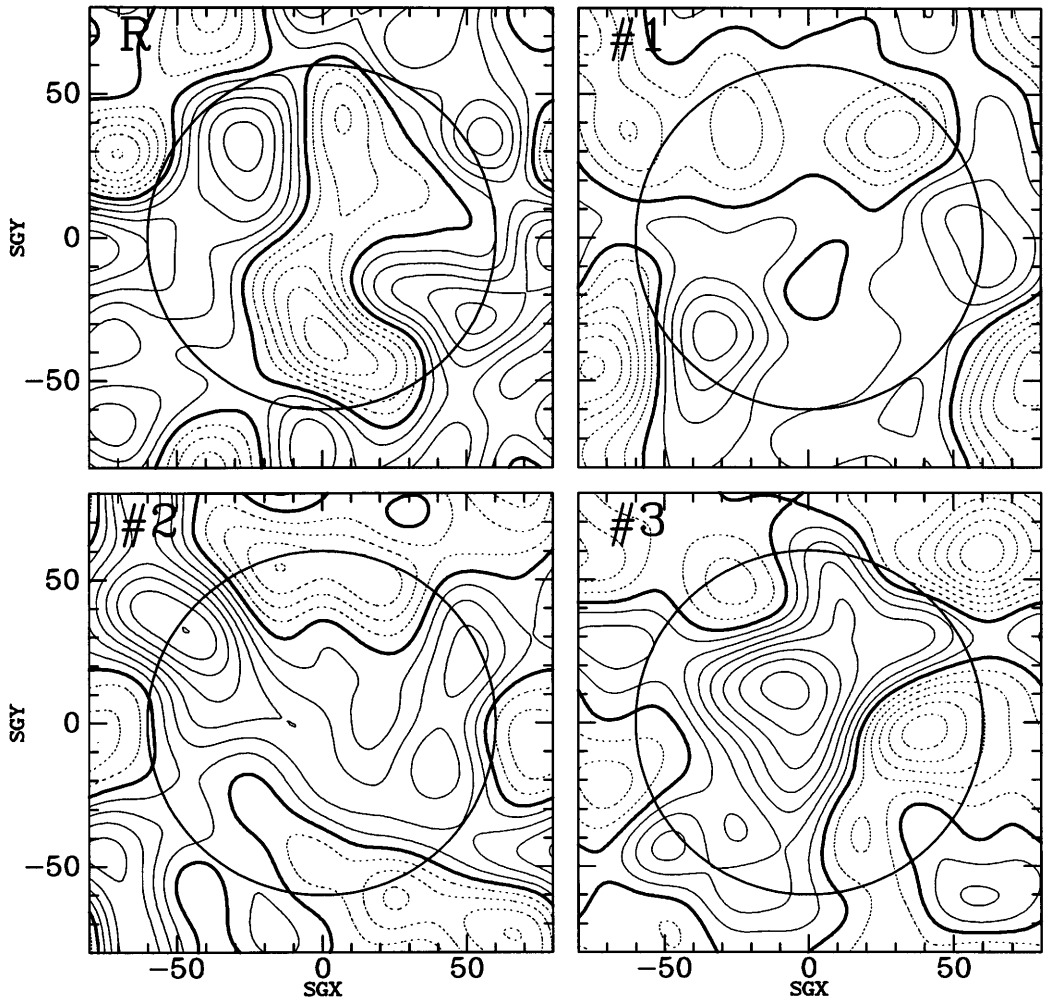

FIG. 1.—Random fields. The linear fluctuation density fields of the four random realizations given the prior as described in the text, smoothed with a 1000 km s^{-1} Gaussian window. The supergalactic plane is shown; the units are in 100 km s^{-1}, and the contour spacing is 0.2 in δ. The heavy contour belongs to $\delta = 0$, while solid and dot contours represent overdensity and underdensity regions, respectively. The circle of radius 6000 km s^{-1} marks the region from which constraints are taken.

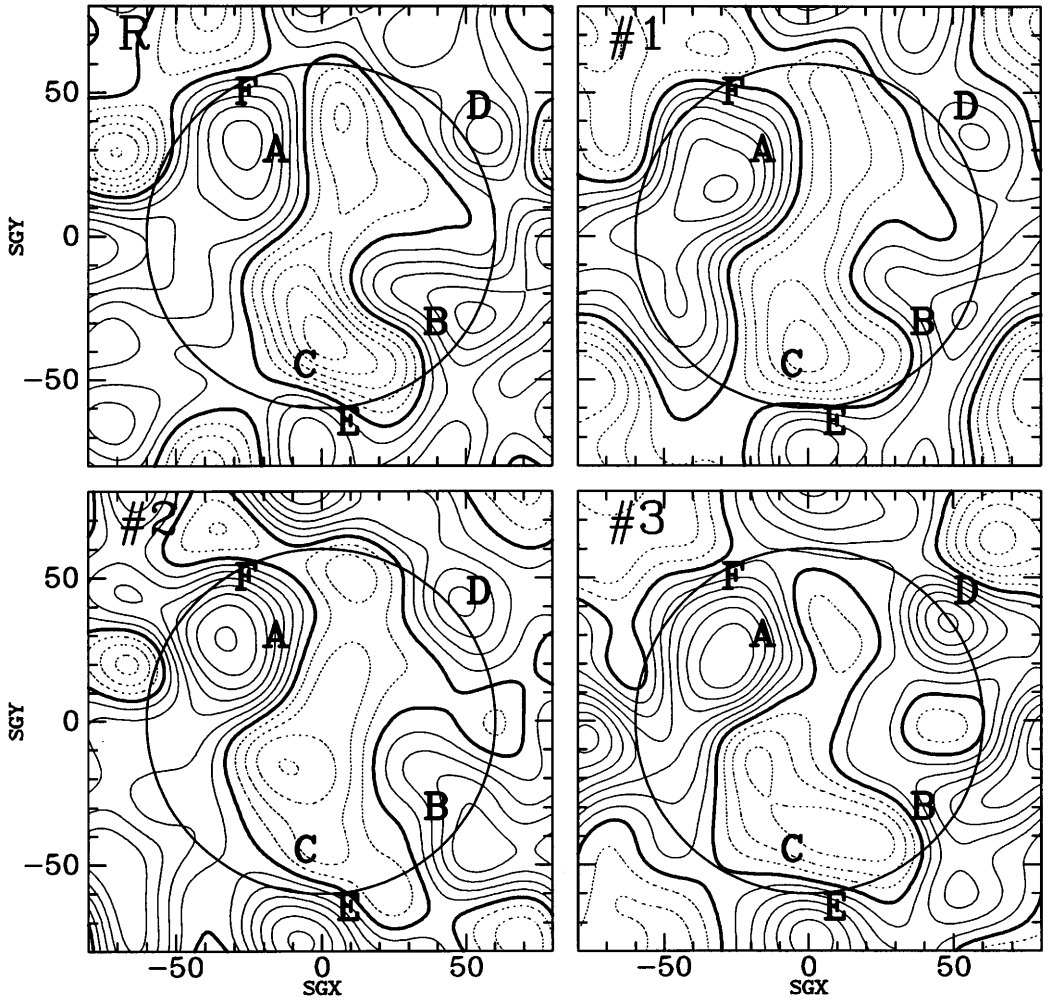

FIG. 2.—Linear CRs from nonlinear data and the linear R field. The linear CRs (Nos. 1–3) are constrained by data sampled from the fully evolved R field. These are supposed to recover the linear R field, which is presented here. The various objects that define the gross structure are marked by the letters A–F, and these are introduced for the sake of the comparison of the reconstructed fields with the original. The parameters of the plot are the same as in Fig. 1.

here. The comparison at the particle distribution level constitutes a much more challenging test of our method. The robust peak at A breaks into subclumps that are located at the intersection of a few filaments. The actual subclump positions and distribution vary in the NLCRs, but the overall statistical nature of the objects is recovered. As expected, the reconstruction of the filaments is less robust than the density peaks. The F filament that starts at one side of the A cluster and continues at its other side is reproduced by all the NLCRs. This is also the case with the filament that runs from cluster B toward the origin. However, there is quite a large scatter in the overall structure of the network of filaments. The conclusion that follows is that the NLCRs provide a faithful reconstruction of the nonlinear density field at the 1000 km s^{-1} Gaussian smoothing level. The highly nonlinear small-scale structure is formed within the correct "boundary conditions" provided by the "actual universe" (i.e. the R realization), but with some degree of random variability as dictated by the unconstrained random realizations used.

The simulations are used to calibrate the nonlinear correction of equation (13). A scatter plot of the (δ^L, δ^{NL}) is shown in Figure 5*a* (*left panel*), where both the bias and the scatter are clearly seen. The nonlinear mapping is applied to δ^{NL} to produce a corrected linear δ^{CL}. Indeed, the scatter plot of (δ^L, δ^{CL}) (Fig. 5*a*, *right panel*) shows that the bias is removed without increasing the scatter (see also Nusser et al. 1991). The mapping recovers the normal 1 point distribution function (Fig. 5*b*) and the theoretical power spectrum (Fig. 5*c*). Here only the dynamical aspects are studied, and no "observational" uncertainties are considered.

The mock realization has been used to simulate the effect of the improvement of the sampling on the quality of the reconstruction. The R simulation has been resampled by the PSCZ selection function, which corresponds to an *IRAS* survey sampled down to a 0.6 Jy flux at 60 μm. (The PSCZ survey is an extension of the 0.6Jy QDOT 1-in-6 survey [Saunders et al. 1990], which is now being completed to all galaxies brighter than the above limit.) The improvement in the errors is shown in Figure 6 as they are normalized to the rms value of the underlying field. Here, only one NLCR has been reconstructed, and realization 1 is used. Figure 7 shows the smoothed R realizations, to be reconstructed; the linear CR that serves as the initial conditions; and the smoothed density field and particle distribution of the NLCR. A comparison of Figure 7 with Figures 3 and 4 shows the improvement gained by the reduction of the shot noise. This is clearly seen in the smoothed nonlinear field, where

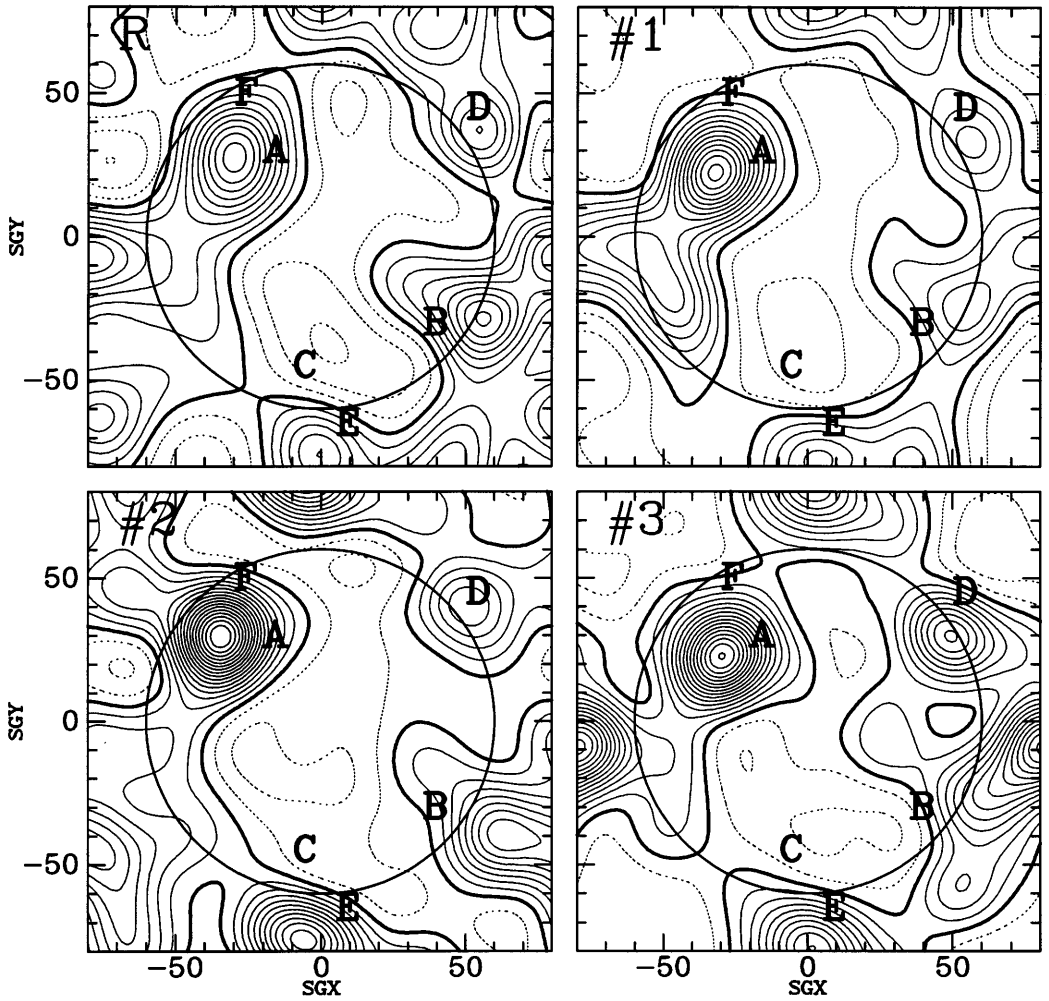

FIG. 3.—NLCRs and the nonlinear R field. The smoothed nonlinear evolved density fields of the *real* universe, R, and the ensemble of NLCRs (Nos. 1–3). The NLCRs reconstruct the R field, within the limitations of the method.

the PSCZ-based reconstruction recovers the original field much better than the 1.2 Jy–based field.

4. VOLUME-LIMITED *IRAS* CATALOG

The *IRAS* 1.2 Jy catalog includes 5321 galaxies. These are used to evaluate the smoothed density field on a Cartesian grid of 1000 km s^{-1} spacing within a 6000 km s^{-1}, excluding the ZOA, yielding 846 constraints. The NLCRs are created on a finer 128^3 grid of 125 km s^{-1} spacing, assuming periodic boundary conditions. However, for a CDM-like power spectrum the structure within the 6000 km s^{-1} is hardly affected by the periodicity on the ± 8000 km s^{-1} box. A word of caution should precede the analysis of the NLCRs. The neglect of the redshift distortions affects the reconstructions in two ways. One is the displacement of the objects by a few times 100 km s^{-1}, and the other is the amplification of the (over)density amplitude, due to the gravitational focusing (cf. Kaiser 1986).

The actual *IRAS* survey is affected by Galactic ZOA of $b \leq 5°$. Our choice of a Cartesian three-dimensional representation is especially suited for handling zones of missing data, where the only correction to be applied is for the truncated Gaussian smoothing near the ZOA. Otherwise, no assumption is made on the completeness of the survey sky coverage. This is to be contrasted with the case of orthogonal representations such as Fourier or spherical harmonics/Bessel functions, where the treatment of regions of missing data usually results in a very complicated mode-mode coupling (Fisher et al. 1995b) The Wiener filter extrapolates the density field within the ZOA, using the galaxy distribution at the two sides of the ZOA.

An "ensemble" (namely, three) of NLCRs has been generated, from which one can estimate the variance of the different reconstructions and thereby assess the quality of the reconstruction. The "observed" smoothed *IRAS* 1.2 Jy density field is presented in Figure 8 (panel labeled R), and the three linear CRs are shown as well, all presented on the supergalactic plane. Note that in the linear CRs the amplitude of the positive (over)density field never exceeds the observed one. The voids, on the other hand, become more empty in the CRs because of the nonlinear correction. The three NLCRs of the actual universe are shown in Figure 9. The random realizations used here are the same ones used for the reconstruction of the mock data (Fig. 1). The fundamental features of the supergalactic plane are the Great Attractor (GA), the Perseus part of the Perseus-Pisces (PP) supercluster, and the Local Void (LV). The overdensity containing the Coma Cluster at (SGX, SGY) $\approx$ (0, 8000) km s^{-1} is clearly seen. However, it lies at the edge of the simulation and is affected by the periodic boundary conditions,

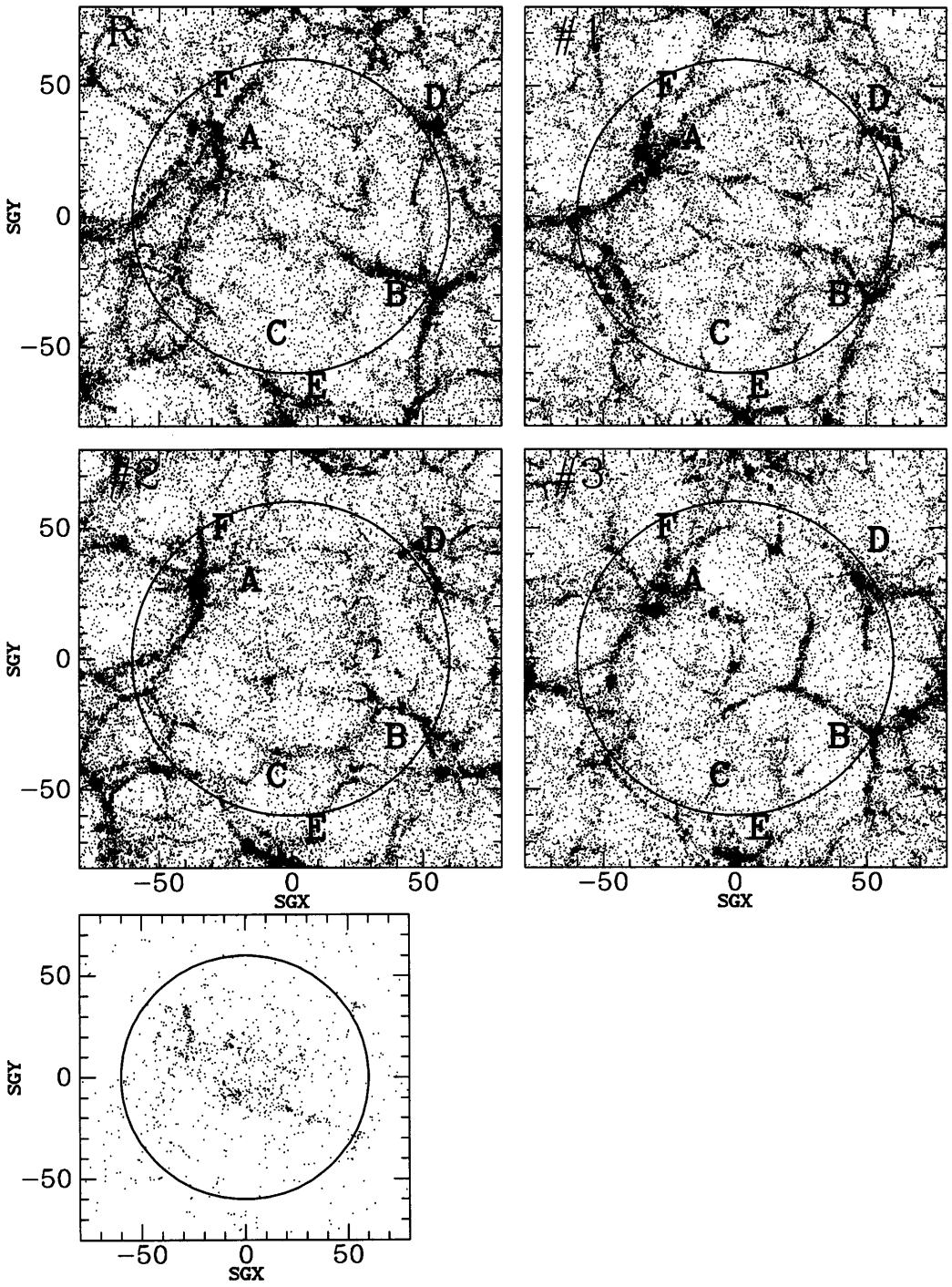

FIG. 4.—Volume-limited mock "galaxy" catalogs. Slices of thickness 2000 km s^{-1} of the volume-limited "galaxy" distribution normalized to the mean galaxy density of the *IRAS* 1.2 Jy survey. The top left panel shows the "galaxy" distribution of the *R* field. Applying the *IRAS* selection function produces the magnitude-limited sample that is shown in the bottom left panel. The mock survey is used to set constraints on the random realizations of the underlying field. The "galaxy" distributions of the three NLCRs are shown here, and these are to be compared with the *R* field.

and therefore its physical characteristics cannot be studied here. The "volume-limited" galaxy distribution of the NLCRs as well as the *IRAS* 1.2 Jy survey itself are shown in Figure 10. The structure of peaks and the network of filaments at the GA region on one side of the Local Group, and at the PP region on the other side, is a robust feature of all the NLCRs. The Local Void is recovered in all the NLCRs, but the details of the filamentary network within the voids vary from one realization to the other. It is in this sense that the term "nonlinear reconstruction" of the local LSS can be used; that is, the gross features of the LSS are imprinted onto the otherwise random nonlinear realizations.

To further study the cosmography recovered by the NLCRs, Aitoff projections of the "galaxy" distribution are plotted. Here only one NLCR (No. 1) is analyzed in that way. Aitoff projections of the "galaxy distribution" within shells of thickness 1000 km s^{-1} at distances of 2000, 3000, and 4000 km s^{-1} (Fig. 11*a*) and 5000, 6000, and 7000 km s^{-1} (Fig. 11*b*) are plotted. Contours of the smoothed field are superimposed on the "galaxy" distribution. Variable

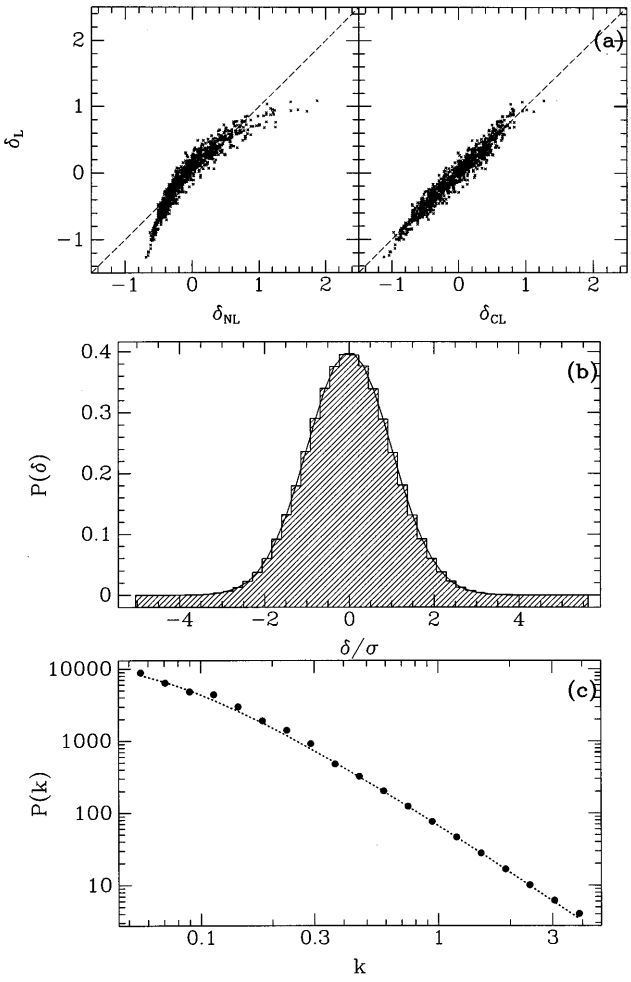

FIG. 5.—(*a*) *Left:* Scatter plot of linear vs. nonlinear evolved density field smoothed with a 1000 km s^{-1} Gaussian filter. *Right:* The same plot after the application of a polynomial correction on the nonlinear field. (*b*) The 1 point probability distribution function of δ (histogram) as calculated from the linear CR given corrected linear constraints. The solid line represents the theoretical Gaussian PDF. (*c*) The power spectrum of the linear CR (*points*). The dotted curve is the CDM-like power spectrum that is used as a prior for the reconstructions.

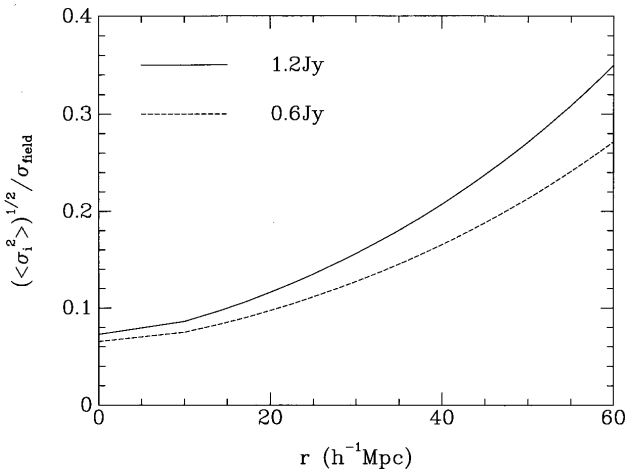

FIG. 6.—Diagonal elements of the error correlation matrix, normalized to the rms value of the 1000 km s^{-1} smoothed field vs. the distance. The two curves show the variation of the errors in the case of the *IRAS* 1.2 Jy and PSCZ selection functions.

smoothing is used here to compensate for the variation of the angular particle density with depth. The smoothing kernel radii used here are 500, 640, 780, 920, 1060, and 1200 km s^{-1}, corresponding to $\delta_{rms} = 0.80, 0.67, 0.56, 0.46, 0.41, 0.37$. In all projections a contour spacing of 0.2 is used. The identification of the various objects grossly follows the cosmographical analysis of Webster et al. (1996), who used a linear Wiener filter in the spherical harmonics/Bessel representation to analyze the same database used here.

The nearby structure, at 2000 km s^{-1}, is dominated by the foregrounds of Centaurus at $(l, b) \approx (305°, 30°)$. The 2000 km s^{-1} plot shows also the Ursa Major $(l, b) \approx (170°, 25°)$ and the outskirts of the Fornax-Doradus-Eridanus complex $(l, b) \approx (180°, -60°)$. Underdense regions are identified at the two sides of the supergalactic plane. These include the Local Void $(330° < l < 120°, -70° < b < 60°)$, a void at $(120° < l < 200°, -40° < b < 0°)$, and a void at $(210° < l < 330°, -85° < b < 0°)$, which extends over all the slices out to 7000 km s^{-1}.

The structure at 3000 km s^{-1} is dominated by the GA complex, which lies close to the supergalactic plane and is almost aligned along the line of sight toward $(l, b) \approx (300°, 25°)$. The Pavo-Indus-Telescopium (PIT) complex $[(l, b) \approx (330°, -15°)]$ forms an extension of the GA. On the other side of the supergalactic plane the structure is dominated by two ridges, one in the direction of $(l, b) \approx (150°, 25°)$ connecting Ursa Major (at 2000 km s^{-1}) with Camelopardalis (at 4000 km s^{-1}) and the other that extends all the way toward the PP at $(l, b) \approx (165°, -15°)$, which peaks at a distance of $\approx$6000 km s^{-1}.

At 4000 km s^{-1} Centaurus/GA $[(l, b) \approx (315°, 20°)]$ and PIT $[(l, b) \approx (325°, -15°)]$ are the dominant overdensities, and these are connected by a filamentary structure running close to the supergalactic plane at $l \sim 320°$, $-10° < b < 10°$. Other features in the plot are a concentration at $(l, b) \approx (180°, -25°)$ and the Camelopardalis cluster $(l, b) \approx (150°, 15°)$.

Going further to 5000 km s^{-1}, the plot shows Centaurus/GA at $(l, b) \approx (320°, 10°)$, a continuation of the $(l, b) \approx (180°, -25°)$ structure; the PP at $(l, b) \approx (150°, -15°)$; a prominent overdensity at $(l, b) \approx (275°, 10°)$ that is quite separated from Centaurus; PIT $[(l, b) \approx (345°, -20°)]$ continuing from the previous slice; and an overdensity at $(l, b) \approx (130°, -65°)$ that extends out to 7000 km s^{-1}. It is not clear whether the remarkable concentration at $(l, b) \approx (40°, -15°)$, which seems to extend from 4000 to 7000 km s^{-1} is a real object or a fluke induced by the shot noise.

The *IRAS* density field is sampled within a distance of 6000 km s^{-1}, and beyond it the random nature of the NLCR becomes more dominant. To study the reconstructed structure at these distances, a more detailed comparison of the members of the "ensemble" of NLCRs is needed, so that the robust features of the realizations can be separated from the random ones. Such an effort is beyond the scope of the present work. Here only the main objects of the 6000 km s^{-1} slice are noted. The dominant feature at this shell is the PP at $(l, b) \approx (165°, -15°)$. Also shown in this plot are the "tail" of Centaurus/GA $[(l, b) \approx (320°, 10°)]$, Leo $[(l, b) \approx (270°, 50°)]$, Cancer $[(l, b) \approx (190°, 35°)]$, and PIT $[(l, b) \approx (350°, -15°)]$. The strong overdensity at $(l, b) \approx (40°, -20°)$ extends from the 5000 km s^{-1} slice to the 7000 km s^{-1} one.

The almost complete sky coverage of the *IRAS* survey, of a ZOA of $|b| \leq 5°$, and the method used here enable a good

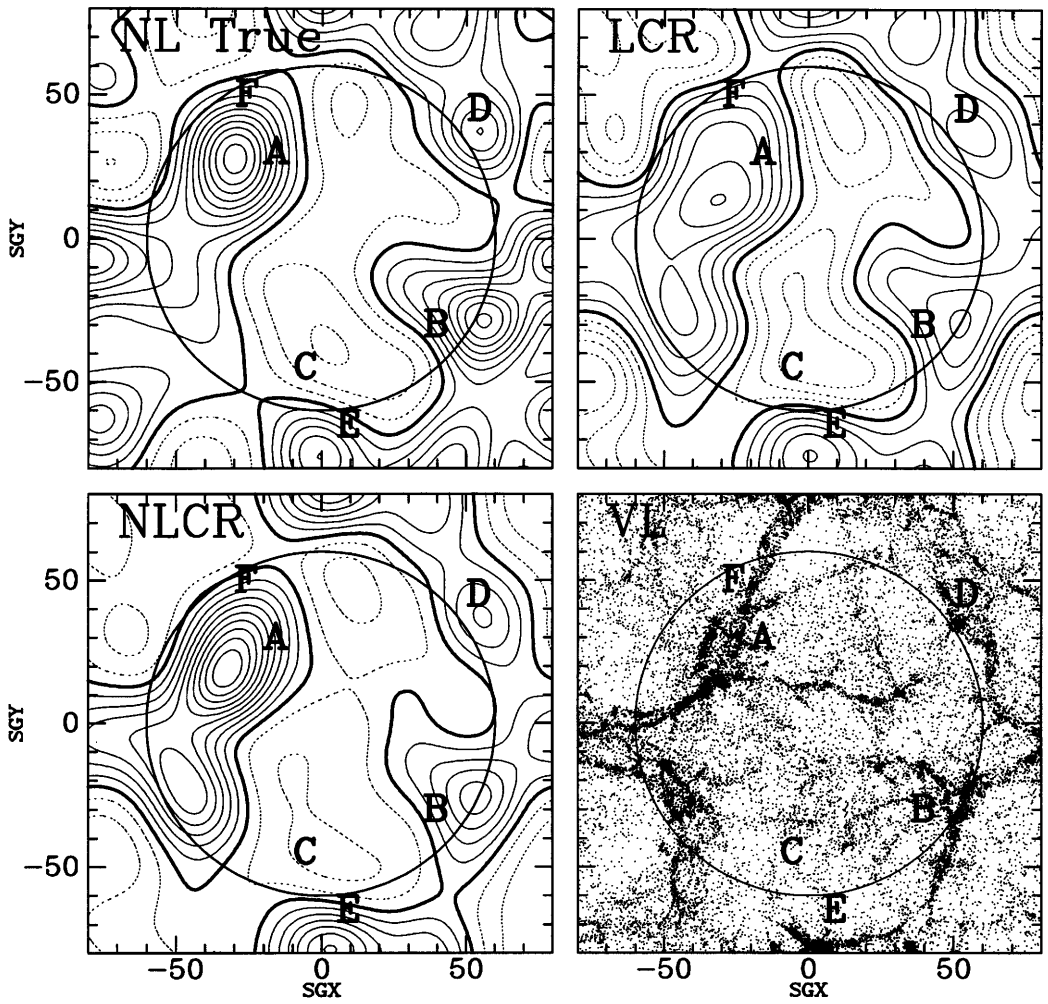

FIG. 7.—PSCZ reconstruction of the mock catalog. *Upper left:* The nonlinearly evolved R field. *Upper right:* Linear CR from a PSCZ-like survey of the R field. The constraints are applied to random realization 1 of the ensemble of realizations. *Lower left:* The smoothed NLCR field for the PSCZ case. *Lower right:* Slice of thickness 2000 km s^{-1} centered at the supergalactic plane of the reconstructed volume-limited PSCZ-like survey.

restoration of the obscured structure. Indeed, the density field found here agrees with optical studies of the ZOA, in particular the optical survey of the southern ZOA [$(l, b) \approx (265°–340°, -10°–10°)$] of Kraan-Korteweg and her colleagues (e.g., Kraan-Korteweg, Woudt, & Henning 1997).

5. DISCUSSION

The NLCR algorithm presented here enables one to perform controlled Monte Carlo N-body simulations of the formation of our "local" universe. These are designed to recover the actual observational LSS within the statistical uncertainties of the data. The new ingredient introduced here is the reconstruction of the nonlinear regime, i.e., the extrapolation in Fourier space from small to large wavenumbers that are deep in the nonlinear regime.

The NLCR introduced here can serve as a tool for studying and analyzing the large-scale structure of the universe. Some of the obvious areas in which NLCRs are expected to be very useful are (1) the reconstruction of the velocity field from redshift catalogs; (2) mapping the zone of avoidance and extrapolating the dynamical fields into unobserved regions; (3) studying the dynamics of observed rich clusters with the actual initial and boundary conditions; (4) analysis of filaments and pancakes as probes of the initial conditions and the cosmological model; and (5) as probes of the biasing mechanism. The main virtue here lies in the fact that different data sets, which in principle can represent different biasing of the underlying dynamical field, can be used to simultaneously set constraints on the realizations. Given all these, and the technical simplicity of the algorithm, we expect it to be a standard tool of N-body and gasdynamical simulations.

The algorithm used here relies on using a Cartesian spatial representation of the density field. This should be considered as an attractive alternative to the spherical harmonics/Bessel function representation (Fisher et al. 1995b). The latter provides a representation that is well suited to handle full-sky redshift surveys, including an elegant treatment of redshift distortions. However, this method has two main shortcomings. One is the complicated mode-mode coupling that is introduced by an incomplete sky coverage. The other is the effective variable smoothing of the reconstruction, which is not suitable for setting initial conditions for N-body simulations. These two problems are naturally solved by using the Cartesian representation, where no a priori geometry of the survey is assumed, and an arbitrary sky coverage of various data sets can be handled. The representation used here provides CRs that have a constant spatial resolution, which makes it the optimal tool for

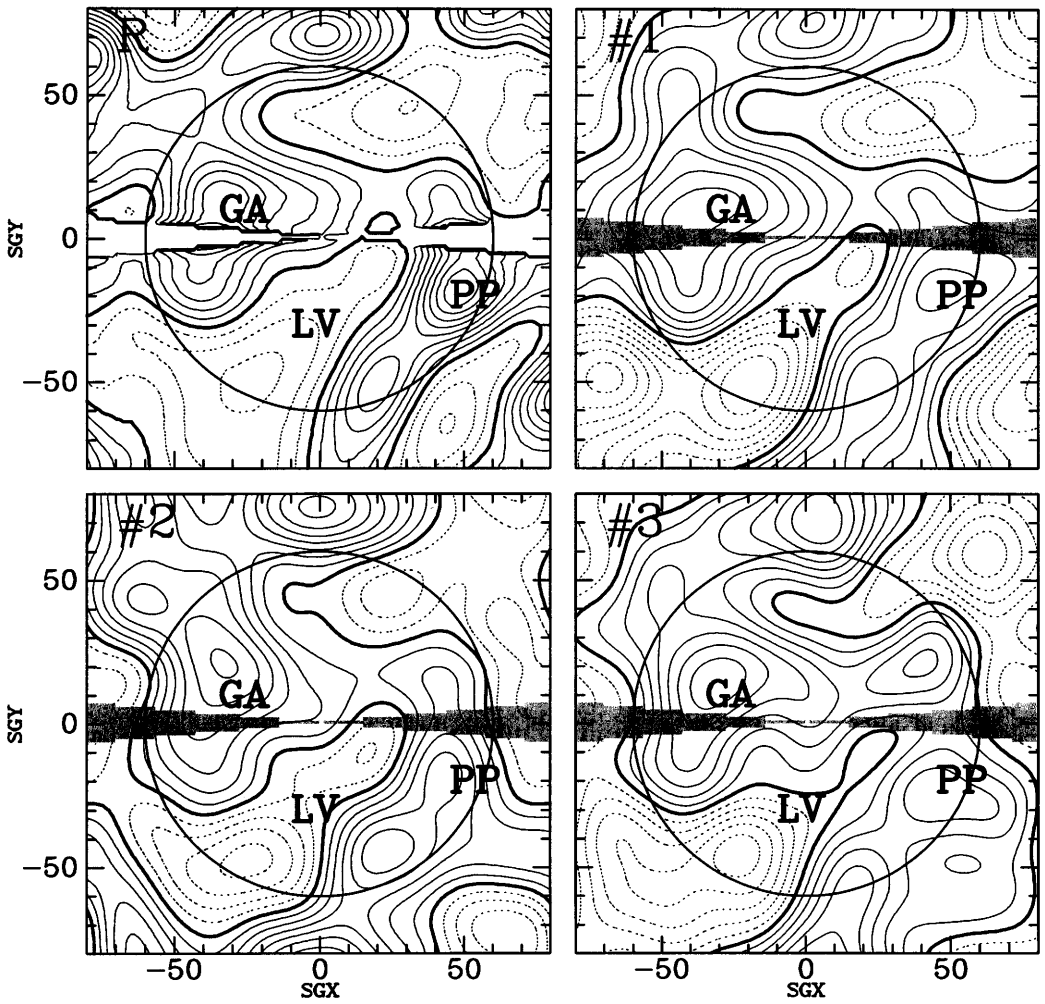

FIG. 8.—Linear CRs of the *IRAS* 1.2 Jy survey. The *R* plot shows the supergalactic plane of the 1000 km s^{-1} smoothed raw *IRAS* 1.2 Jy survey. The remaining three plots show the linear CRs, which are reconstructed by imposing constraints from the smoothed actual *IRAS* survey. The three linear CRs are generated from the three random realizations (Nos. 1–3) used for the reconstruction of the mock catalog.

setting initial conditions for numerical simulations. Indeed, the amplitude of the mean (WF) field decreases as the shot-noise errors increase with distance, but this is compensated by the random component of the CRs. The resulting realizations have a constant resolution with a constant power. The method outlined here can be extended to handle very large data sets by using data compression methods, in particular the signal-to-noise eigenmode expansion (Zaroubi

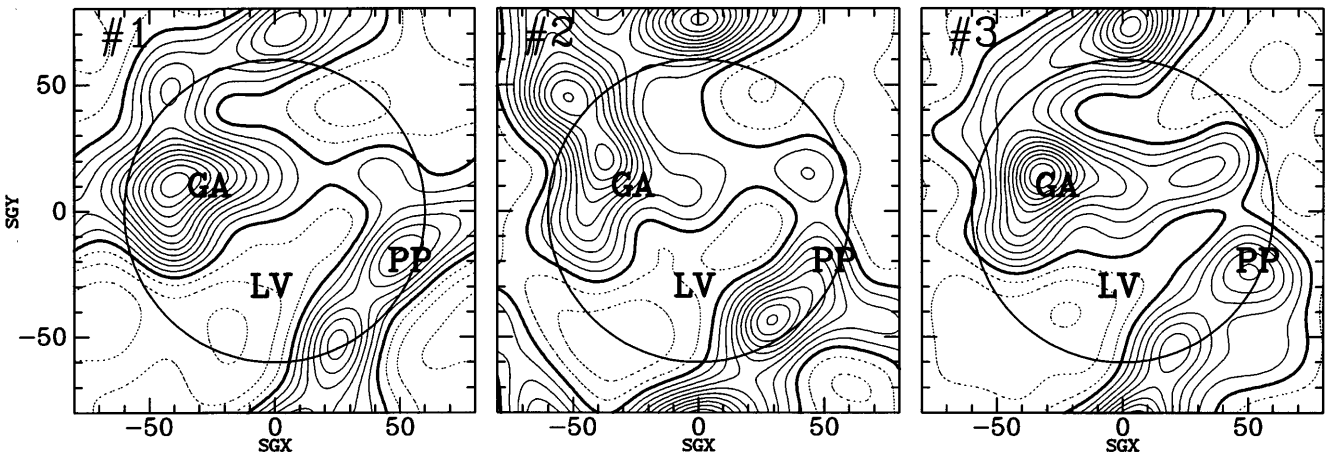

FIG. 9.—NLCRs of the *IRAS* 1.2 Jy survey. The ensemble of the three NLCRs based on the *IRAS* 1.2 Jy survey smoothed by a 1000 km s^{-1} Gaussian window.

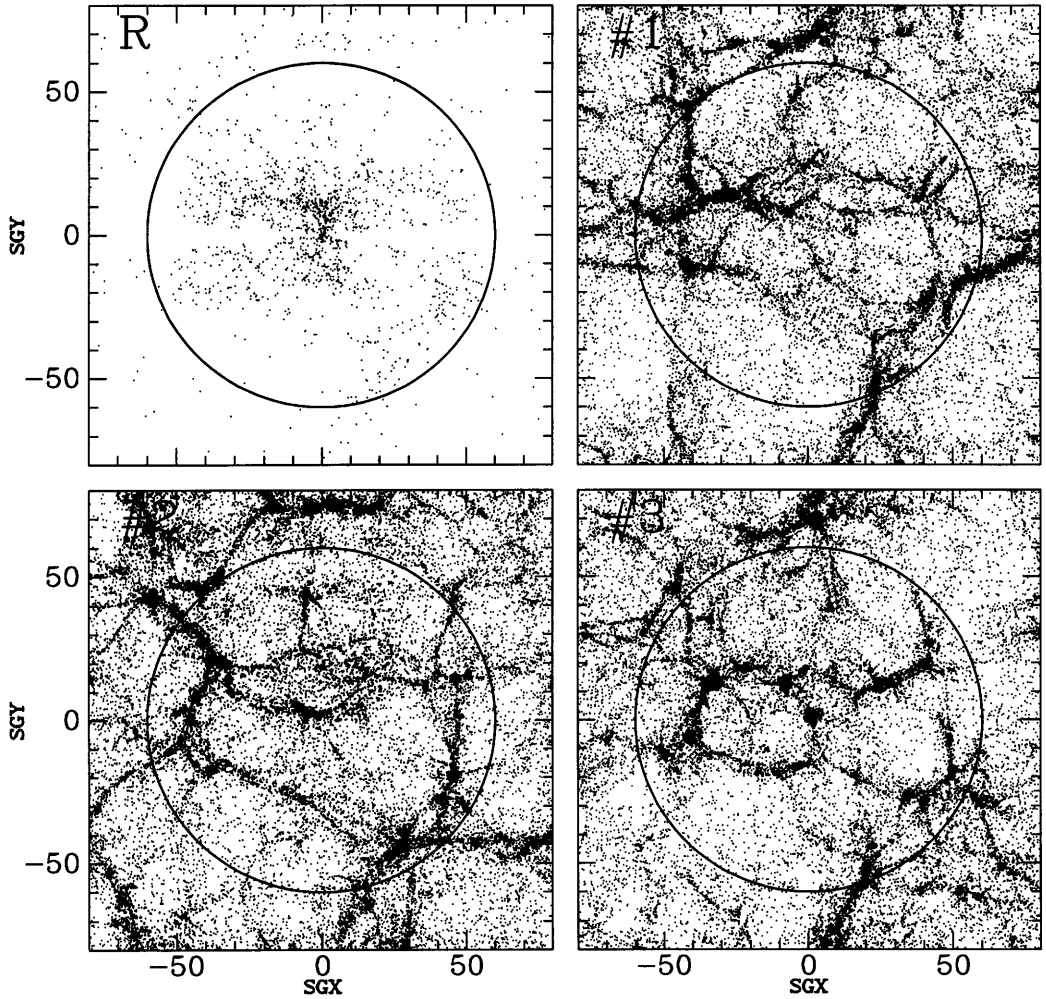

FIG. 10.—Volume-limited *IRAS* 1.2 Jy reconstructions. The *R* plot shows a 2000 km s^{-1} thick slice of the *IRAS* 1.2 Jy redshift survey centered at SGZ = 0. The other plots show the ensemble of the reconstructed volume-limited *IRAS* 1.2 Jy surveys.

1995; Vogeley & Szalay 1996; Tegmark, Taylor, & Heavens 1997). Using this representation, the dimensionality of the data space can be significantly reduced, and the WF/CR algorithm can be easily formulated in the reduced data space. A shortcoming of the present work is the neglect of redshift distortions. However, these can be handled by using correlation matrices of equation (2), properly expressed in terms of redshift-space dynamical variables. The general formalism of evaluating the redshift-space variables, within the linear theory, was outlined by Zaroubi & Hoffman (1996).

At the time this paper was originally submitted, Kolatt et al. (1996) reported on a similar project of NLCR of the *IRAS* 1.2 Jy catalog. Their procedure differs from the present one mainly in not distinguishing between low resolution (data) and high resolution (realizations). The input data are smoothed on the 500 km s^{-1} scale and are heavily dominated by the noise, which is partially "removed" by a power-preserving filter (PPF: a modified WF). The PPF is designed to preserve the power regardless of the noise level. The resulting estimator is therefore more dominated by the noise and less by the prior model, compared to our method. Next, the resulting 500 km s^{-1} power-preserving filtered field is taken from the quasi-linear to the linear regime by the Nusser & Dekel (1992) "time machine," where a further Gaussianization is applied to the linear field. Small-scale structure is then added to the linearized and Gaussianized 500 km s^{-1} smoothed field by constrained realizations. Compared to the work of Kolatt et al. (1996), the present algorithm is more rigorous and consistent with the theoretical framework. Our procedure involves only one ad hoc step, namely, the linearization procedure, where Kolatt et al. used the PPF, linearization, and Gaussianization to obtain the 500 km s^{-1} smoothed field. Yet Kolatt et al. used a better and more consistent linearization procedure, which can replace the one used here. In spite of the very different approaches, it seems that the two methods yield similar results and are equally efficient. Detailed comparisons against *N*-body simulations and mock catalogs are needed to judge the merits of each method.

The members of the *IRAS* collaboration are gratefully acknowledged for their help with the *IRAS* database. We have benefited from many stimulating discussions with L. da Costa, A. Dekel, O. Lahav, and S. Zaroubi. This work is supported in part by US-Israel Binational Science Foundation grant 94-00185 and by Israel Science Foundation grant 590/94.

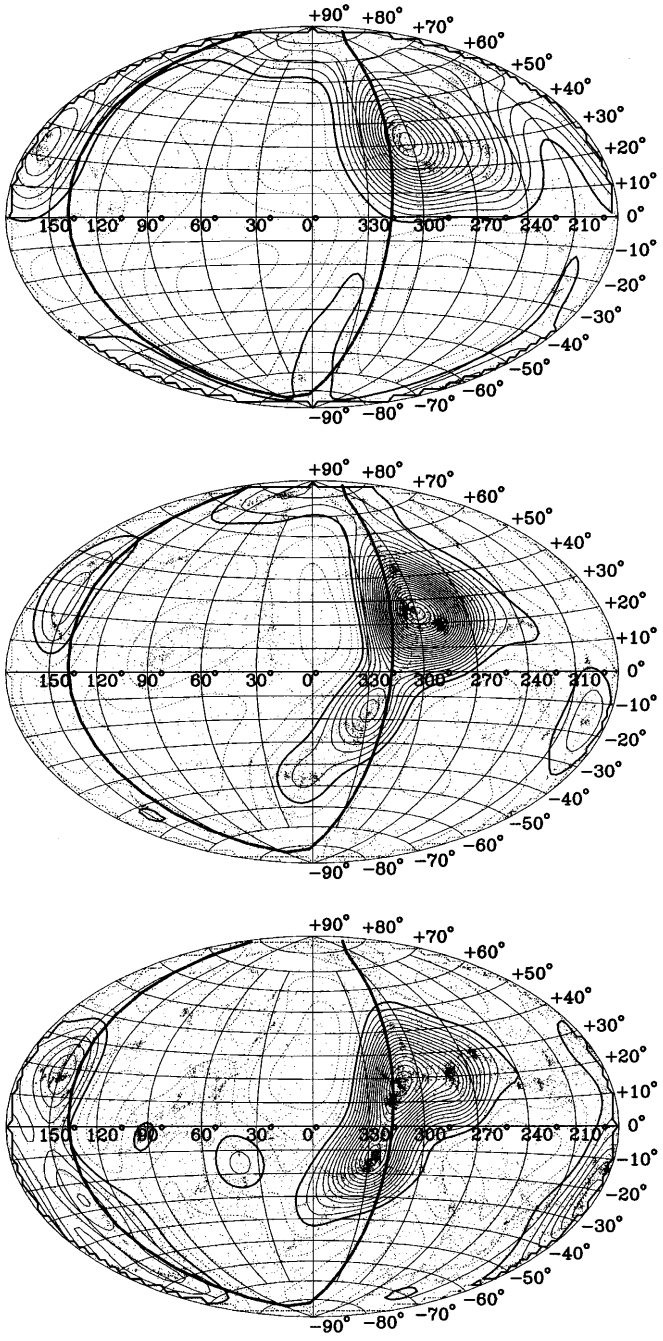

FIG. 11*a*

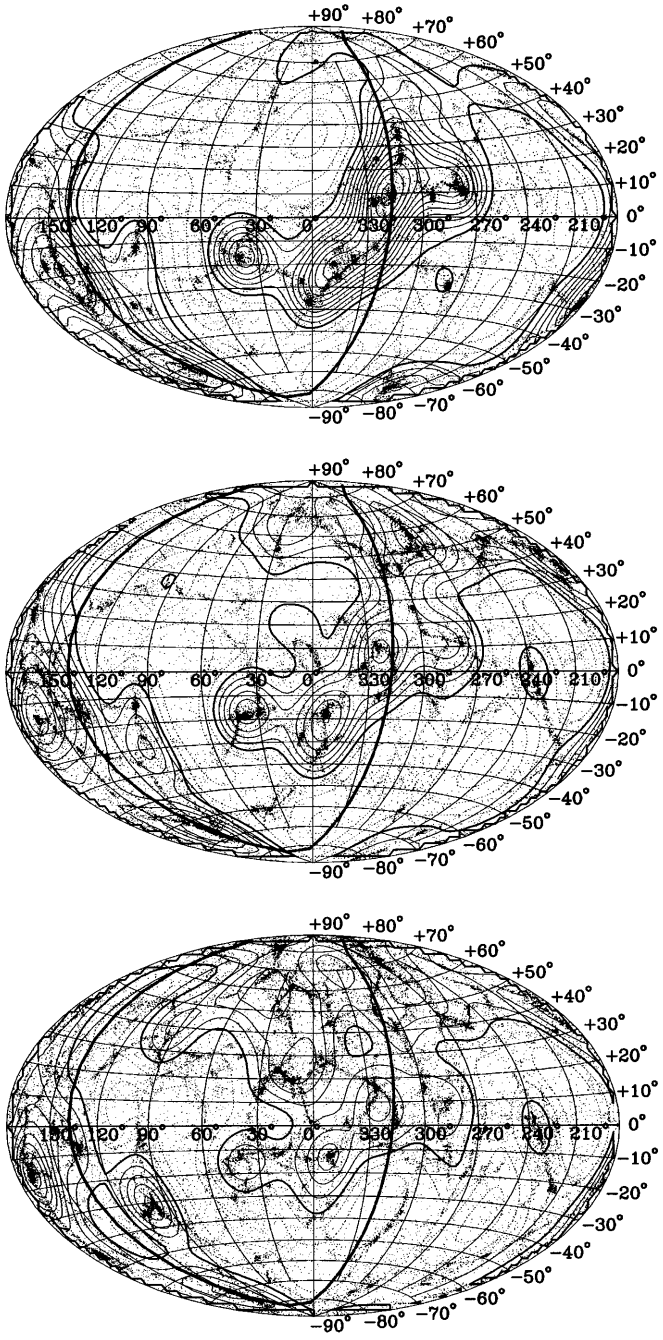

FIG. 11*b*

FIG. 11.—Aitoff projections in Galactic coordinates of the NLCR (No. 1) reconstruction of the volume-limited *IRAS* 1.2 Jy. (*a*) (*from top to bottom*), the 2000, 3000, and 4000 km s^{-1} shells; (*b*) the 5000, 6000, and 7000 km s^{-1} shells. The thick curve marks the supergalactic plane, and the particles are within a 1000 km s^{-1} thick slice centered on these distances.

REFERENCES

Bunn, E., Fisher, K. B., Hoffman, Y., Lahav, O., Silk, J., & Zaroubi, S. 1994, ApJ, 432, L75

Bunn, E., Hoffman, Y., & Silk, J. 1996, ApJ, 464, 1

Dekel, A. 1994, ARA&A, 32, 371

Fisher, K. B., Davis, M., Strauss, M. A., Yahil, A., & Huchra, J. P. 1993, ApJ, 402, 42

———. 1994, MNRAS, 267, 927

Fisher, K. B., Huchra, J. P., Davis, M., Strauss, M. A., Yahil, A., & Schlegel, D. 1995a, ApJS, 100, 69

Fisher, K. B., Lahav, O., Hoffman, Y., Lynden-Bell, D., & Zaroubi, S. 1995b, MNRAS, 272, 885

Ganon, G., & Hoffman, Y. 1993, ApJ, 415, L5

Hoffman, Y. 1993, in Proc. Ninth IAP Conf. on Cosmic Velocity Fields, ed. F. Bouchet & M. Lachiéze-Rey (Gif-sur-Yvette: Editions Frontières), 357

———. 1994, in ASP Conf. Ser. 67, Unveiling Large Scale Structures behind the Milky Way, ed. C. Balkowski & R. C. Kraan-Korteweg (San Francisco: ASP), 185

Hoffman, Y., & Ribak, E. 1991, ApJ, 380, L5
Kaiser, N. 1986, MNRAS, 227, 1
Kolatt, T., Dekel, A., Ganon, G., & Willick, J. A. 1996, ApJ, 458, 419
Kraan-Korteweg, R. C., Woudt, P. A., & Henning, P. A. 1997, Proc. Astron. Soc. Australia, 14, 15
Lahav, O. 1993, in Proc. Ninth IAP Conf. on Cosmic Velocity Fields, ed. F. Bouchet & M. Lachiéze-Rey (Gif-sur-Yvette: Editions Frontières), 205
———. 1994, in ASP Conf. Ser. 67, Unveiling Large Scale Structures behind the Milky Way, ed. C. Balkowski & R. C. Kraan-Korteweg (San Francisco: ASP)
Lahav, O., Fisher, K. B., Hoffman, Y., Scharf, C. A., & Zaroubi, S. 1994, ApJ, 423, L93
Nusser, A., & Dekel, A. 1992, ApJ, 391, 443
Nusser, A., Dekel, A., Bertschinger, E., & Blumenthal, G. R. 1991, ApJ, 379, 6
Padmanabhan, T. 1993, Structure Formation in the Universe (Cambridge: Cambridge Univ. Press)
Peebles, P. J. E. 1980, The Large-Scale Structure of the Universe (Princeton: Princeton Univ. Press)
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes (2d ed.; Cambridge: Cambridge Univ. Press)
Saunders, W., Rowan-Robinson, M., Lawrence, A., Efstathiou, G., Kaiser, N., Ellis, R., & Frenk, C. S. 1990, MNRAS, 242, 318
Scherrer, R. J., & Bertschinger, E. 1991, ApJ, 381, 349
Strauss, M. A., & Willick, J. A. 1995, Phys. Rep., 261, 271
Tegmark, M., Taylor, A. N., & Heavens, A. F. 1997, ApJ, 480, 22
Vogeley, M. S., & Szalay, A. S. 1996, ApJ, 465, 34
Webster, A. M., Lahav, O., & Fisher, K. B. 1996, MNRAS, in press
Wiener, N. 1949, Extrapolation and Smoothing of Stationary Time Series (New York: Wiley)
Willick, J. A., Courteau, S., Faber, S. M., Burstein, D., & Dekel. A. 1995, ApJ, 446, 12
Yahil, A., Strauss, M. A., Davis, M., & Huchra, J. P. 1991, ApJ, 372, 380
Zaroubi, S. 1995, in Proc. 30th Moriond Meeting, Clustering in the Universe, in press
Zaroubi, S., & Hoffman, Y. 1996, ApJ, 462, 25
Zaroubi, S., Hoffman, Y., & Dekel, A. 1997, preprint
Zaroubi, S., Hoffman, Y., Fisher, K. B., & S. Lahav, O. 1995, ApJ, 449, 446 (ZHFL)