Dartmouth College
Dartmouth Digital Commons

Open Dartmouth: Faculty Open Access Articles

8-1995

Determination of Malmquist Bias and Selection Effects from Monte Carlo Simulations

Wolfram Freudling
European Southern Observatory

Luiz N. da Costa
Institut d'astrophysique de Paris

Gary Wegner
Dartmouth College

Riccardo Giovanelli
Cornell University

Follow this and additional works at: https://digitalcommons.dartmouth.edu/facoa

Part of the External Galaxies Commons

Recommended Citation

DETERMINATION OF MALMQUIST BIAS AND SELECTION EFFECTS
FROM MONTE CARLO SIMULATIONS

WOLFRAM FREUDLING

Space Telescope—European Coordinating Facility, European Southern Observatory, Karl Schwarzschild Str. 2, D-85748 Garching b.
München, Germany
Electronic mail: wfreudli@eso.org

LUIZ N. DA COSTA

Institut d'Astrophysique, 98 bis Boulevard Arago, F75014, Paris, France
and Observatorio Nacional, Rua Gen. J. Cristino 77, São Cristovão, Rio de Janeiro, Brazil
Electronic mail: ldacosta@iap.fr

GARY WEGNER

Department of Physics and Astronomy, 6127 Wilder Laboratory, Dartmouth College, Hanover, New Hampshire 03755
and Astronomisches Institut, Ruhr-Universität Bochum, D-44780 Bochum, Germany
Electronic mail: wegner@kayz.dartmouth.edu

RICCARDO GIOVANELLI AND MARTHA P. HAYNES

Center for Radiophysics and Space Research and National Astronomy and Ionosphere Center,[1] Cornell University,
Ithaca, New York 14953
Electronic mail: riccardo@astrosun.tn.cornell.edu, haynes.astrosun.tn.cornell.edu

JOHN J. SALZER

Department of Astronomy, Wesleyan University, Middletown, Connecticut 06457
Electronic mail: slaz@parcha.astro.wesleyan.edu
Received 1994 December 28; revised 1995 February 13

ABSTRACT

Maps of the peculiar velocity field derived from distance relations are affected by Malmquist type bias and selection effects. Because of the large number of interdependent effects, they are in most cases difficult to treat analytically. Monte Carlo simulations are used to understand and evaluate these effects. In these simulations the "true" spatial distribution and relevant properties of galaxies as well as selection effects and observational uncertainties are realistically modeled. The results of the simulation can be directly applied to correct observed peculiar velocity maps. The simulation is used to investigate biases in samples of measured peculiar velocities by Lynden-Bell *et al.* (ApJ, 326, 19, 1988). Willick (ApJ, 351, L5, 1990) and the new sample of spiral galaxies by Haynes *et al.* (BAAS, 25, 1403, 1993). Based on the results obtained from the application of our method to toy models we find that the method is a useful tool to estimate the bias induced both by inhomogeneities and selection effects. This is a crucial step for the analysis of the Haynes *et al.* sample which was selected with a redshift dependent criterion. © *1995 American Astronomical Society.*

1. INTRODUCTION

Maps of the peculiar velocity field of galaxies are a powerful tool to test scenarios for formation of large scale structure. However, the interpretation of such maps must take into account biases introduced by the coupling of errors in the distance measurements and the spatial distribution of galaxies and/or selection effects in the samples considered. Such biases in the derived distances are present even if the underlying distance relation is exactly known (which we will as-

sume throughout this paper). A discussion of how a bias-free distance relation can be derived from data will be presented by Giovanelli *et al.* (1995). Biases originating from the spatial distribution are usually referred to as Malmquist bias and correcting for it requires a detailed knowledge of the underlying distribution of galaxies from which the sample is drawn. Biases introduced by the selection of the sample, which we will call selection biases throughout this paper, are caused by the correlation of the parameters used to select the sample with parameters used to estimate distances. In order to correct for such biases, the distribution function describing relevant galaxy properties is needed, as well as the detailed criteria adopted in selecting the sample being considered.

---

[1]The National Astronomy and Ionosphere Center is operated by Cornell University under a cooperative agreement with the National Science Foundation.

A number of correction schemes for Malmquist bias have been suggested by different workers (e.g., Feast 1987; Lynden-Bell *et al.* 1988; Landy & Szalay 1992; Willick 1991). A common feature of all these approaches is that they require *a priori* knowledge of the distribution of galaxies in space, which has to be either assumed (e.g., Lynden-Bell 1988) or, in the nontrivial case of an inhomogeneous distribution, derived from the data before the correction for the bias can be made. Different methods for estimating the required density distribution have been considered. Dekel (1993) built a treatment of the inhomogeneous Malmquist bias into their reconstruction of the density field, assuming that the distribution of galaxies traces the underlying matter distribution derived from peculiar velocities. Another approach utilizes the self-consistently reconstructed density field derived from redshift surveys (e.g., Dekel 1994; Hudson 1994).

Biases resulting from selection effects can in some situations be either comparable or larger than the Malmquist bias. To correct for selection biases, detailed knowledge of the selection process is necessary. However, even if this information is available it is difficult to derive simple expressions for situations where the parameter used to select the sample is not identical but correlates to one or several parameters used to derive distances. This is especially true when a complex set of different selection criteria is used to select the sample. For example, samples where the selection depends explicitly on the redshift, like the recently completed survey of peculiar velocities of Sc galaxies (Haynes *et al.* 1993, hereafter referred to as the Sc sample), are particularly difficult to correct for this kind of bias.

The biases mentioned above are strongly related to the scatter of the particular distance relation used. For example, the scatter in the Tully–Fisher (TF) relation depends on the linewidths of the galaxies in the sense that the scatter is large for galaxies of lower luminosity (Giovanelli *et al.* 1995). This fact and the unique selection criteria of the Sc sample have motivated us to address the question of biases, both Malmquist and selection biases, using a Monte Carlo (MC) technique.

MC simulations of Malmquist bias and/or selection effects have been used in the past by a number of workers (e.g., Landy & Szalay 1992; Teerikorpi 1993; Willick 1994), mainly for the purpose of testing analytic corrections. In contrast, here the MC technique is used to determine the bias at each point in space to correct measured peculiar velocity fields. The essence of our method is to simulate galaxy catalogues using the spatial and parametric distribution derived from observations, and select redshift-distance samples similar to those currently available. The strength of the method is that all relevant effects can be incorporated to predict the bias without the need for any simplifying assumptions.

In Sec. 2 we review analytical corrections for Malmquist bias which are used to test our MC results in simple cases. In Sec. 3 a detailed description of the technique is presented. In Sec. 4 we compare the bias derived from our MC simulations with analytical corrections and in Sec. 5 we derive the bias field for observational samples. A brief summary is presented in Sec. 6.

2. BIASES IN THE PECULIAR VELOCITY FIELD

Different analytical expressions to correct for the Malmquist and selection type bias can be found in the literature (see, e.g., Willick 1991; and references therein). These expressions depend critically on the properties of the scatter in the distance relation used. For example, to discuss the bias on distances estimated from a TF relation given by

$$I = a_{TF} \log w + b_{TF} \qquad (1)$$

one can distinguish between the "direct" and "inverse" TF relation. In Eq. (1) I is the absolute I-band magnitude and w is the corrected H I linewidth. The direct relation (called "forward" relation by Willick 1994) is derived from a least-squares fit of I as a function of $\log w$, while the inverse relation is obtained from a least-squares fit of $\log w$ as a function of I. Independent of the distribution of the scatter about the line defined by Eq. (1), a normal distribution in $\log w$ can be used as an approximation for the inverse relation, while a normal distribution in I will be a better approximation for the direct relation. Although in this paper we focus our attention on the TF relation, our conclusions apply equally well to the $D_n - \sigma$ relation.

A raw estimate of the distance d_{raw} from a measured magnitude m_I and linewidth w

$$\log d_{raw} = 0.2(m_I - a_{TF} \log w - b_{TF}) \qquad (2)$$

is a biased estimate of the distance. The most general expression to correct d_{raw} obtained with the direct TF relation under the assumption that the scatter is normal in magnitudes is

$$d_{cor} = \frac{\int_0^\infty r^3 n(r) \exp\left(-\dfrac{[\ln(r/d_{raw})]^2}{2\Delta^2}\right) dr}{\int_0^\infty r^2 n(r) \exp\left(-\dfrac{[\ln(r/d_{raw})]^2}{2\Delta^2}\right) dr}, \qquad (3)$$

where d_{cor} is the corrected estimated distance, Δ^2 is the variance in the logarithm of the distance, and $n(r)$ is proportional to the space distribution of galaxies along the line of sight.

Landy & Szalay (1992) have suggested a different expression given by

$$d_{cor} = d_{raw} \exp\left(\frac{\Delta^2}{2} \frac{h[\ln d_{raw} + \Delta^2]}{h[\ln d_{raw}]}\right), \qquad (4)$$

where $h(d)$ is the *observed* distribution of galaxies along the line of sight. As pointed out by Teerikorpi (1993), this expression is applicable when distances have been derived using the inverse TF relation.

Both expressions (3) and (4) can be easily applied for the homogeneous case. They can also be easily applied to areas of the sky where the distribution of observed galaxies can be adequately described by one dimensional functions $h(d)$ or $n(r)$. However, if larger regions of the sky are considered,

variation with the direction in the sky must be taken into account. For example, estimates for $n(r)$ can be obtained from redshift surveys taking into account expected peculiar velocities (Hudson 1994).

Willick (1994) also presented a formal framework to correct for selection biases which is similar to the corrections for Malmquist bias given in Eq. (3). Unfortunately, his expressions cannot directly be applied to a sample selected as the Sc sample.

In the analysis of the Sc sample we expect significant contributions from both Malmquist bias and selection effects due to the selection criteria adopted. This has prompted us to develop the MC approach presented below. In Sec. 4 we compare the biases of the analytical expressions given in this section to results from our MC technique discussed below.

3. THE MONTE CARLO SIMULATION

3.1 Method

The ultimate goal of the method is to create simulated redshift-distance surveys with the same properties and thus the same biases as in existing observed samples. To achieve this, we have to simulate properties like the spatial distribution, magnitude-diameter function, color relations, distance relations, and their respective scatter properties. In addition, the selection criteria and completeness of the observed samples under consideration have to be taken into account.

For the spatial distribution of galaxies, we use the density field derived from redshift surveys corrected for peculiar velocities (see below). The derived density field depends on the assumed $\beta=\Omega/b^{0.6}$, where Ω is the density parameter and b is the linear biasing of galaxies relative to matter. In this paper, we assume that $\beta=1$. It should be emphasized that the density field and therefore the simulation is only slightly model dependent.

We use this density field to create a large number of artificial galaxies, to which we assign properties similar to those of the observed population of galaxies. We then simulate the selection of catalogues from which observational samples are drawn. Subsequently, the observational sample is drawn from the simulated catalogue, following the same selection criteria used for existing surveys, including any known incompleteness of the observational sample. Although below we assume that this incompleteness depends only on the magnitudes, this can be easily generalized to include incompleteness as a function of direction in the sky. Finally, estimated distances are derived from the assumed distance relation.

The final sample, which contains both the true and the estimated distances, can now be used to compute the bias field in estimated distance or redshift space. For each artificial galaxy, the bias is computed by subtracting from the estimated distance the known "true" distance of galaxies. This bias affects not only the position of the galaxies but also its estimated peculiar velocity which can also be corrected

using the same bias correction. Since the artificial catalogues of galaxies are large, the uncertainty in the determination of the biases can be reduced by binning them on a grid either in redshift or in estimated distance space. This average bias can then be applied to observed samples selected like the simulation. The raw estimated distance can be used to select the appropriate bias from the simulation and the estimated distance can be corrected for it. Peculiar velocities are subsequently computed from the bias corrected distances.

The items and parameters that represent the built-in assumptions for the estimate of biases are the parameters describing the luminosity function, the TF relation and the color–linewidth relation, the density distribution of galaxies, the peculiar velocities of galaxies, and the detailed selection and completeness of the observed sample. The main advantage of this approach is the flexibility it provides to evaluate a variety of effects and their impact on the estimated distances, including the scatter in the various relations used, the properties of the distance relations and the selection and incompleteness of observational samples.

3.2 Procedure

3.2.1 Creating the 3d galaxy distribution

To prepare the MC simulation, either a homogeneous distribution of galaxies or a distribution modeled after the local universe is used to derive a density field in real space. For the latter, the *Optical-IRAS* model without bias of Freudling *et al.* (1994) is used. This model uses the redshift distribution of optically selected galaxies in the galactic caps ($b<-30°$ or $b>40°$), the 1.9 Jy *IRAS* sample in regions $|b|>5°$, and interpolates the density field over the galactic plane ($|b|<5°$). The positions of the galaxies are computed iteratively from the redshifts, using the linear approximation to estimate the peculiar velocity of each galaxy in each iteration. The final output is a self-consistent density field, which is stored on a Cartesian grid of 500 km s^{-1} cell size. This final density field is shown by Willmer *et al.* (1995).

The first step in the MC simulation is to create a list of galaxy coordinates which distributes galaxies randomly but with density fluctuations identical to those in the *Optical-IRAS* model. This is accomplished by interpreting the local density of the model as the probability p to draw a galaxy at that location. The x, y, and z components of the position of potential galaxies are randomly drawn. The object is then rejected with a frequency proportional to $1-p$.

Finally, each galaxy is given a redshift computed from its distance, the velocity field of the *Optical-IRAS* model, and an assumed Gaussian random component. The distances are assumed to be proportional to the unperturbed redshift in the CMB rest frame and are expressed in km/s throughout this paper. For computational efficiency, the peculiar velocity vector is computed only once for each cell on the density mesh, and all galaxies within each cell are given the sample peculiar velocity field. Subsequently, a Gaussian random component to the redshift of $\sigma_f=150$ km s^{-1} is added, representing an estimate of the cosmic velocity dispersion.

3.2.2 Galaxy properties

Most observational maps of peculiar velocities (e.g., Willick 1991; Courteau *et al.* 1993, or the Sc sample) are originally drawn from diameter-limited catalogues such as the UGC or the ESO–Uppsala catalogues, and magnitude or other criteria are only imposed on the galaxies which already have been preselected by the criteria imposed in the parent catalogue. In order to incorporate such selection procedures into the MC simulation, galaxies distributed according to some density field must be given diameters, magnitudes, and other parameters relevant for the distance relation being used. Sodré & Lahav (1993) have investigated the joint diameter-magnitude distribution function, and have parametrized it in the following way:

$$\psi(D,B) \propto \exp\left[-\frac{D}{D^*} - \frac{(b-a\,\log(D)-B)^2}{2\sigma_M^2} \right], \quad (5)$$

where B is the blue absolute magnitude computed from B_T, D is the diameter of the galaxy, and the parameters a, b, D^*, and σ_M have been derived by Sodré & Lahav. We use such a distribution to assign both diameters and blue magnitudes to the artificial galaxies. For the simulation of the Sc sample, the parameters derived by Sodré & Lahav for their "$S+I$" sample with three different velocity models were averaged. The resulting parameters used are $D^*=2331$ km s^{-1} arcmin, $a=-4.87$, $b=-0.05$.

The properties of each galaxy are assigned as follows. First the diameters in km s^{-1} arcmin are drawn from an exponential distribution function $\propto\exp(D/D^*)$ [see Eq. (5)]. For a given diameter, we derive the most probable value for the blue absolute magnitudes B from the linear relation between diameters and magnitudes, $B=a\,\log(D)+b$, where a and b are the ones from the distribution functions. The final assigned magnitude will later be computed by adding to B a Gaussian random variable with a standard deviation of σ_M.

Since Eq. (5) is given in the blue band, it is necessary to relate the blue B_T magnitude to the I magnitude. For that we utilize a color–linewidth relation of the form $B_T-I=a_c\,\log w + b_c$. This together with the I-band TF relation as given in Eq. (1) are used to compute I and linewidths from B. Preliminary relations have been derived from the Sc data and are given in Table 1. As above, the scatter in these quantities will be added in the next section.

Note that our procedure differs from that of Willick (1991), who assumes that the galaxy properties originate from a Gaussian line width distribution rather than a luminosity function. Magnitudes are then computed from the widths via given TF relation, and a normally distributed scatter is added to the magnitudes. This results in a Gaussian luminosity function rather than the more Schechter-like function implicitly assumed in our approach.

3.2.3 Correlations and scatter

In the previous section, the parameters D, B, I, and $\log w$ were assigned to each galaxy using a distribution function for D and observed relations to compute the remaining pa-

TABLE 1. Model parameters for spiral galaxies.

Luminosity	function
D^* km s^{-1}	2231
a	-4.87
b	-0.05
σ_M	0.709
TF	relation
σ_{TF}	0.3
a_{TF}	-7.48
b_{TF}	-2.53
σ_m	0.3
a_c	1.78
b_c	-2.54
σ_c	0.59

rameters. Up to now, we have not considered the scatter in those relations. However, selection biases are produced because galaxy samples are selected by observables which are closely correlated with the quantity to be measured. In particular, blue magnitudes and diameters are used later to select samples. Therefore, their correlation with the parameters of the TF relation, I band magnitudes, and linewidth, must be modeled realistically in order to obtain reliable estimates for the biases. The scatter in the relation used to assign B from D Eq. (5) was assumed to be $\sigma_M=0.709$. For the TF relation, the scatter has been estimated to be $\sigma_{TF}=0.3$. It should be emphasized that the observed scatter in the TF relation and therefore the biases depend on the linewidth (e.g., Giovanelli *et al.* 1995). The value of 0.3 chosen here is the value measured from the Sc sample at a linewidth log w of about 2.4. The scatter in the color–linewidth relation, was derived from the Sc data to be $\sigma_c=0.59$. In Table 1 we list all the parameters used in the simulation, which are used later to determine the bias for samples of spiral galaxies. Finally, in the discussion below we also consider the uncertainty in the apparent blue magnitudes listed in the catalogues generally used to select the sample. Here we take this scatter to be $\sigma_m=0.3$.

Finally, we are in the position of assigning values to all of the observable quantities relevant to our simulated samples such as $B_{\rm scatter}$, $I_{\rm scatter}$, and $w_{\rm scatter}$. These are computed by adding normal deviates to the corresponding quantities B, I, and w computed before. For the inverse TF relation, four independent Gaussian random variables ξ_i with zero mean and a $\sigma=1$ are used in the following manner. First four random variables σ_1 to σ_4 are defined:

$$\sigma_1 = \sigma_{\rm TF}\xi_1, \quad (6)$$

$$\sigma_2 = \sigma_m\xi_2, \quad (7)$$

$$\sigma_3 = \sqrt{\sigma_M^2 - \sigma_m^2}\,\xi_3, \quad (8)$$

$$\sigma_4 = \sqrt{\sigma_c^2 - \sigma_m^2}\,\xi_4. \quad (9)$$

These random variables are added to the observable quantities as computed in the previous section in the following manner. First, a Gaussian with a scatter σ_M has to be added to B so that the distribution of B and $\log D$ follows Eq. (5). Using the random variables defined above, we define

$$B_{scatter} = B + \sigma_2 + \sigma_3 \tag{10}$$

in such a way that the distribution of $B_{scatter}$ satisfies Eq. (5). The scatter in the color relation is also reproduced if we define $I_{scatter}$ to be given by

$$I_{scatter} = I + \sigma_3 + \sigma_4. \tag{11}$$

In order to reproduce the scatter in the TF relation, we define

$$w_{scatter} = w \, 10^{(\sigma_1 + \sigma_3 + \sigma_4)/a_{TF}}. \tag{12}$$

It can be verified that this leads to a Gaussian scatter term for $B - I$ with a dispersion of σ_c by substituting σ_1, σ_2, and σ_3 in the relations. This is also approximately the scatter of the color–linewidth relation, because the additional terms added to w are small compared to the scatter in the color.

Similarly, for the direct TF relation implemented by scatter in magnitudes, Eqs. (8)–(12) change to

$$\sigma_3 = \sqrt{\sigma_M^2 - \sigma_m^2 - \sigma_{TF}^2}\,\xi_3, \tag{13}$$

$$\sigma_4 = \sqrt{\sigma_c^2 - \sigma_m^2}\,\xi_4, \tag{14}$$

$$B_{scatter} = B + \sigma_1 + \sigma_2 + \sigma_3, \tag{15}$$

$$I_{scatter} = I + \sigma_1 + \sigma_3 + \sigma_4, \tag{16}$$

$$w_{scatter} = w \, 10^{(\sigma_3 + \sigma_4)/a_{TF}}. \tag{17}$$

3.2.4 Creating simulated surveys

The next step in the MC simulation is to create galaxy catalogues similar to existing ones which we assume to be complete within the specified criteria. This step is necessary to simulate the exact selection procedure of the observer. For example, a "magnitude-limited" sample drawn from the UGC catalogue is not identical to a truly magnitude limited sample because the inclusion of a galaxy in the UGC already depends on a diameter selection threshold. Since all the observables of the artificial galaxies are known, this is an easy step. For example, for the UGC we select galaxies with diameters larger than 1 arcmin or brighter than 14.5 mag.

Next, "observed samples" of galaxies are drawn from the simulated catalogues. It is here that we can take full advantage of the MC approach since besides the known selection effects (e.g., on magnitudes and sizes of galaxies), even complicated selection criteria, such as imposed redshift cutoffs, variations with direction on the sky, and combination of different samples can be incorporated into our simulations. In addition, observational samples are characterized by an incompleteness function that must also be taken into account in building our simulated samples. In general, we assume that the samples are incomplete as a function of apparent blue magnitude only. The incompleteness of each sample was taken into account in generating our simulated samples.
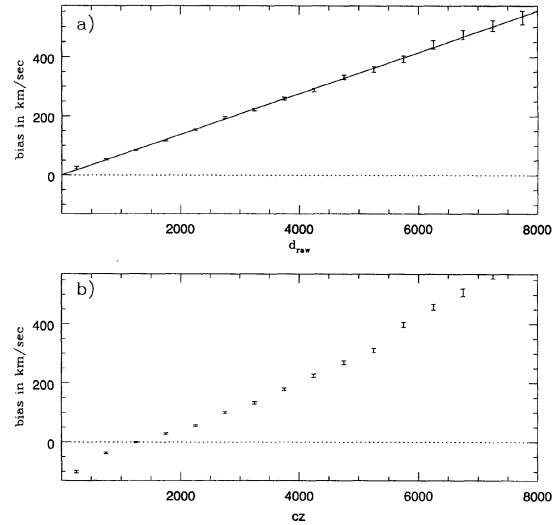

FIG. 1. Biases of a homogeneous distribution of galaxies observed with 14.5 mag limit drawn from the UGC catalogue. The scatter of the TF relation of this sample is distributed in I, corresponding to the direct TF relation. The velocities were binned in 500 km s^{-1} bins in estimated distance d_{raw} in panel (a) and redshift cz in panel (b). The units for distances, redshifts, and biases are km/s. The solid line in panel (a) is the homogeneous Malmquist bias computed from Eq. (3).

Finally, peculiar velocities are computed from the estimated distances obtained from the TF relation and observed magnitudes, widths, and redshifts.

3.2.5 Computation of bias

The peculiar velocity of a galaxy is defined as $v_{pec} = v_r - d_{true}$, where v_r includes the recessional velocity and the streaming motion at its location, and d_{true} is the true velocity of the galaxy. This quantity is estimated from measurements of the redshift cz and the distance d_{raw} as

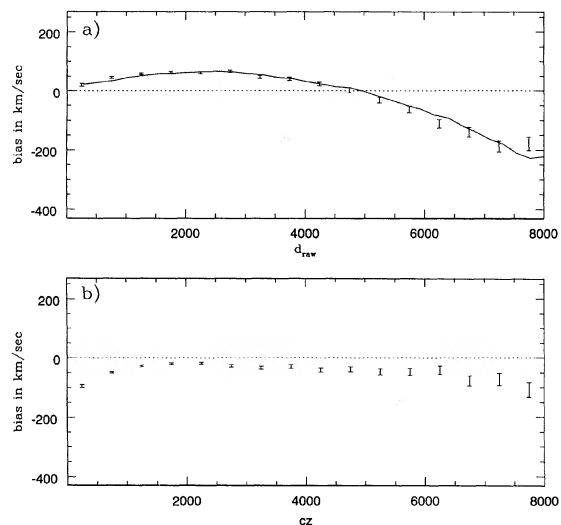FIG. 2. Same figure as Fig. 1, but with the scatter in the TF relation distributed in $\log w$, corresponding to an inverse TF relation. The solid curve in panel (a) is the expected bias for the observed distribution of galaxies computed from Eq. (4).
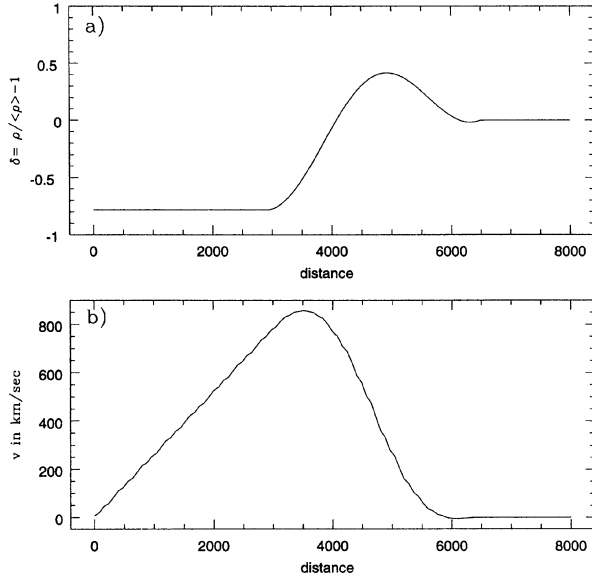
FIG. 3. Density profile and peculiar velocities of the toy model described in the text.

$v_{pec,m} = cz - d_{raw}$, where the measured redshift includes the cosmic random component σ_f and the estimated distance a large measurement uncertainty. These random components lead to a bias of the measured peculiar velocity relative to v_{pec}.

The final step in our MC simulation is to estimate these biases. This is done by computing raw estimates of the distances, corrected for homogeneous Malmquist bias, d_{HMB} and measured redshifts cz from the observed parameters, exactly in the same manner as in the observations. The error in the estimated distance δ_d is computed for each galaxy as

$$- \delta_d = d_{HMB} - d_{true}. \tag{18}$$

Similarly, the error in the measured redshift relative to the underlying streaming motion and recessional velocity is computed as

$$\delta_z = cz - v_r. \tag{19}$$

The signs in these definitions were chosen so that the error in the measured peculiar velocity is $\delta_v = \delta_d + \delta_z$.

Biases are produced by the fact that the expectation value of these errors at a given estimated distance or measured redshift are not zero for an observed galaxy sample. In the MC simulation, the biases in estimated distances b_d or in redshift space b_z are computed by averaging the errors for individual galaxies in the sample within spatial cubes of 500 km s^{-1} size, i.e.,

$$b_d = \langle \delta_d \rangle_{cell}, \tag{20}$$

$$b_z = \langle \delta_z \rangle_{cell}. \tag{21}$$

Depending on whether the direct or inverse TF relation was used, the averaging in cells $\langle \rangle_{cell}$ was performed differently. In case of the direct TF relation, galaxies were assigned to spatial cells based on the estimated distances corrected for homogeneous Malmquist bias. This results in a

bias field which contains the nontrivial biases not removed by the homogeneous Malmquist correction. In case of the inverse TF relation, this assignment to cells was made in redshift space.

The average and root mean square σ_b^2 of the errors for all N galaxies in each cell were estimated, homogeneous Malmquist bias corrected distances were computed and recorded. The resulting field of biases was then smoothed with a Gaussian, where the σ in grid size was chosen to be $(\sigma_b^2/\sqrt{N}/150 \text{ km s}^{-1})^{2/3}$. The minimum smoothing applied was a Gaussian with a σ of 1 and the maximum σ used was 3. This results in heavy smoothing of regions with few galaxies and large discrepancies of the biases, and less smoothing in high density regions and regions where the biases of the galaxies agree well. A similar smoothing should be applied to the real data in order to make the derived bias field fully applicable.

These bias fields are later applied to real observations in the following manner. First, the raw distance estimate of real observations is corrected for homogeneous Malmquist bias. Next, this corrected distance of the galaxy and its direction on the sky is used to find the spatial cube in which the galaxy is located and the corresponding biases are determined. The distance bias b_d is then subtracted from d_{HMB} to find a final, bias corrected estimate of the distance. Similarly, the redshift bias b_z is added to the measured redshift. The corrected distance and redshift are subsequently used to compute the peculiar velocity of the galaxy.

4. SIMULATIONS

In this section, we present a series of simulations of "observed" peculiar velocities distributions and compare them to the true velocities. The raw distances are computed from the I-band TF relation as discussed above. The simulations differ in the assumed spatial distribution of galaxies. We first present a homogeneous universe and a toy model density distribution for the purpose of demonstrating the general behavior of the bias and testing our procedure. Subsequently, we use a realistic density distribution, modeled after the density distribution derived from redshift surveys. Similar models will later be used to determine biases for existing observational samples.

4.1 Tests of the MC Method

We begin our analysis by applying our MC method to spatial distributions for which the expected biases can be computed analytically. This allows us to test the bias field determined from the MC method. As our first model we consider a magnitude-limited sample ($m_B = 14.5$) drawn from a UGC-like catalogue of a homogeneous distribution of galaxies with no streaming motion. In this experiment, the peculiar velocities of the galaxies are exclusively due to the cosmic dispersion σ_f. Figures 1 and 2 show the biases in the observed peculiar velocities. In this simulation, the biases in distance b_d are computed without the HMB correction to the estimated distances, i.e., $b_d = \langle d_{raw} - d_{true} \rangle_{cell}$.
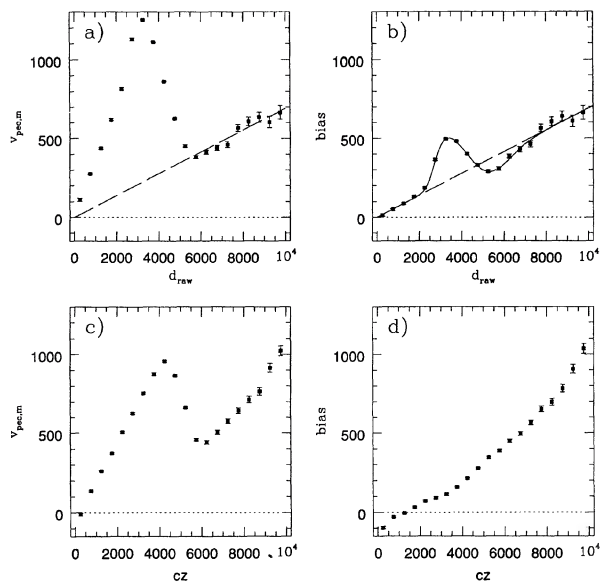
FIG. 4. Observed peculiar velocities [panels (a) and (c)] and biases [panels (b) and (d)] from the toy model for the direct TF relation. Panels (a) and (b) show the velocities and biases binned in estimated distances d_{raw}, whereas panels (c) and (d) show them binned in redshift. The dashed lines in panels (a) and (b) are the expected homogeneous Malmquist bias. The solid curve in panel (b) is the bias as computed from Eq. (3), using the known density profile as shown in Fig. 3. Note the strong bias for the redshift binning [panel (d)].
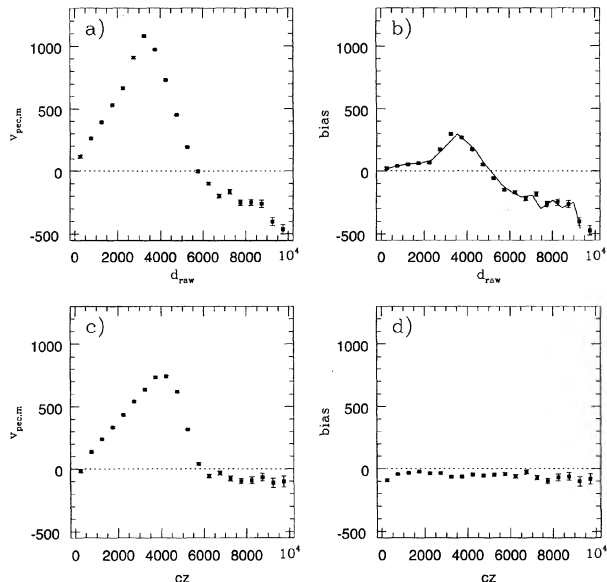

FIG. 5. Observed peculiar velocities and biases from the toy model for the inverse TF relation. The presentation is the same in Fig. 4. The solid curve in panel (b) is the bias as computed from Eq. (4), using the observed distribution of galaxies derived from the data. Note that there is some bias even for the redshift binning [panel (d)].

Figure 1 is the case corresponding to the use of the direct TF relation. Panel (a) shows the biases binned in estimated distances. Superimposed is the bias expected from a homogeneous distribution computed from Eq. (3). It can be seen that the bias behaves as predicted, indicating that additional selection biases are negligible. In redshift space [panel (b)], Malmquist-type effects result from the cosmic random component σ_f and produce a bias for small redshifts. The magnitude of this bias is directly determined by the value of 150 km s^{-1} assumed for σ_f. The cosmic component plays the role of the error in distance estimate. This effect contributes to a negative bias because on the average the redshift will be underestimated. At larger redshift, this contribution becomes small and selection biases dominate.

In Fig. 2 we show the bias when the inverse TF relation is used. As before, the solid line in panel (a) represents the expected bias as estimated using now Eq. (4). Note that although relatively small, biases are present even for the inverse TF relation binned in redshift (panel b). At small redshifts, the cause is the same as in the case of the direct relation discussed above. At large redshifts, selection biases are present because of the correlation between the linewidth and the blue magnitudes used to select the sample due to the indirect selection on line widths introduced by the color-line width relation. This is different from the assumption of Schechter (1980) that there are no selection effects on line-widths.

In our second model, we assume spherical symmetry and a radial density profile which simulates a void centered on

the observer which is surrounded by a region of high density at intermediate distances. The density falls off to the mean density at large distances. The density profile is shown in Fig. 3. From this profile, the expected streaming motion was computed in linear approximation using $\beta=1$.

In Figs. 4 and 5, the measured peculiar velocities are plotted for the direct and for the inverse TF relation, respectively. In Figs. 4 and 5, the measured peculiar velocities and their biases are plotted as a function of estimated distances and redshifts. Again, the biases do not include the HMB correction and the analytical solutions computed from Eqs. (3) and (4) are superimposed on the MC results.

Comparing Figs. 4 and 5 one can clearly see the advantage of studying the peculiar velocities in redshift space. This has previously been recognized by Willick (1994). Figure 4 shows that although selection effects are important, they are almost independent of density enhancements. This is an important property, because in this case biases do not produce any apparent infall. More importantly, Fig. 5 shows that in the case of the inverse TF binned in redshift space, the biases are smaller than in all other combinations.

The good agreement between the analytical curves and both the homogeneous and inhomogeneous bias recovered utilizing our methodology demonstrates that our MC technique is reliable.

4.2 Realistic Model

In order to deal with real samples, it is necessary to apply our simulations to a density field which adequately represents the true distribution of galaxies. As discussed above, we utilize the *Optical-IRAS* density field for this purpose. In
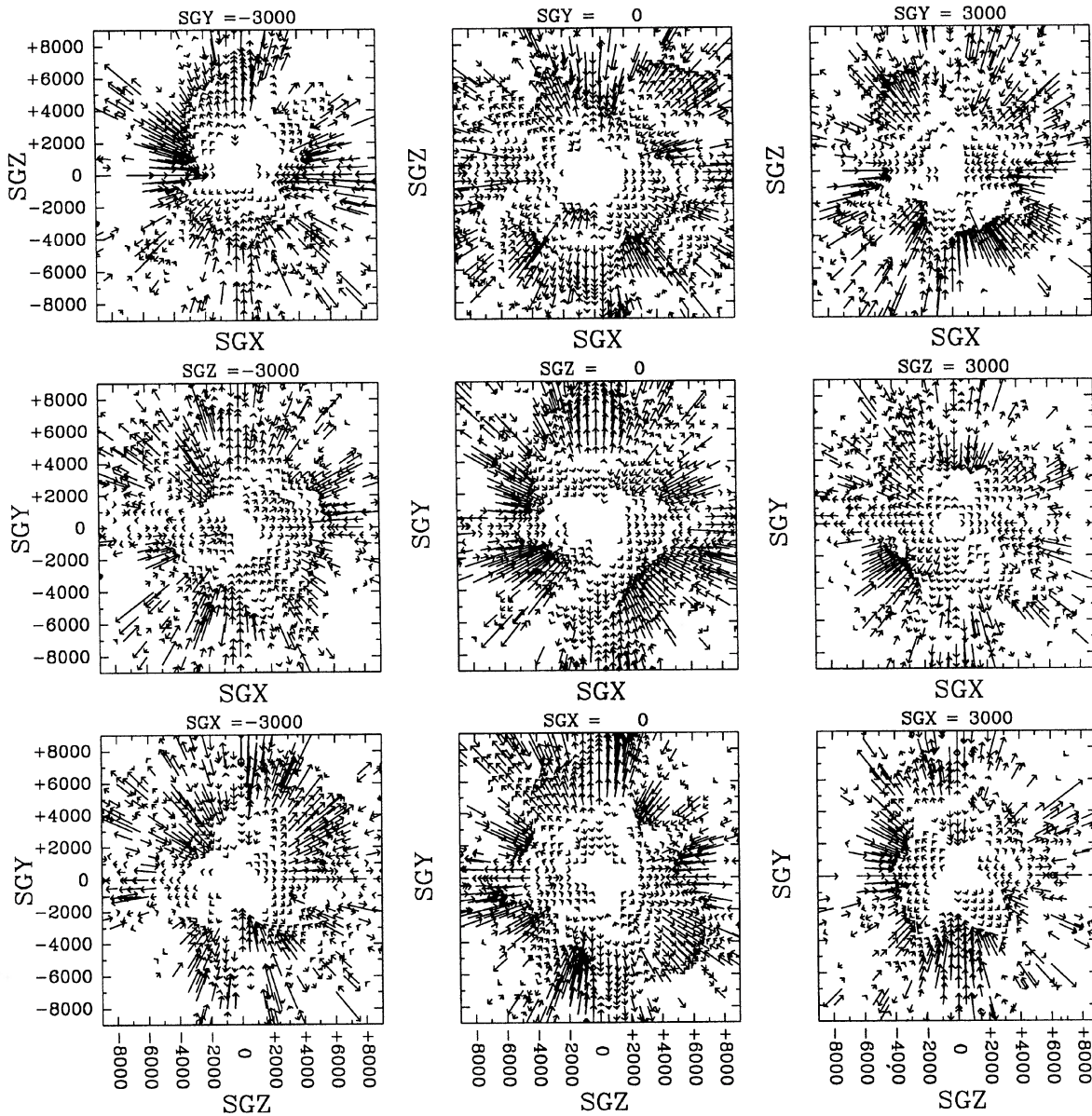
FIG. 6. Biases of a 14.5 mag limited sample shown as a function of HMB corrected positions. The biases were computed using the direct TF relation and are shown as vectors at their positions in Supergalactic Cartesian coordinates. This plot corresponds to a peculiar velocity field an observer would derive *if there were no true peculiar velocities* and the observer computed peculiar velocities from estimated distances corrected for HMB. The units for distances are km/s. The scale of the velocity vectors is five times the scale of distances.

contrast to the simple cases considered earlier, the bias b_d in this section is computed from the HMB corrected distance estimates d_{HMB}.

The resulting velocity biases are shown in Fig. 6 as vectors along the line of sight at their d_{HMB} positions for the case of a magnitude limited sample ($m=14.5$) and TF relation with scatter distributed in magnitudes. For clarity, the scale of the bias vectors is five times the scale of the distances. This plot corresponds to a peculiar velocity field an observer would derive if there were no true peculiar velocities and the observer computed peculiar velocities from estimated distances corrected for HMB. It can be seen that the effect of the bias is to produce artificial infall into high-density regions. For instance, this effect can be seen in the central panel near the locations of the Great Attractor (SGX≈ −5000, SGY≈+2000), and the Pisces–Perseus complex (SGX≈+3000, SGY≈−3000).
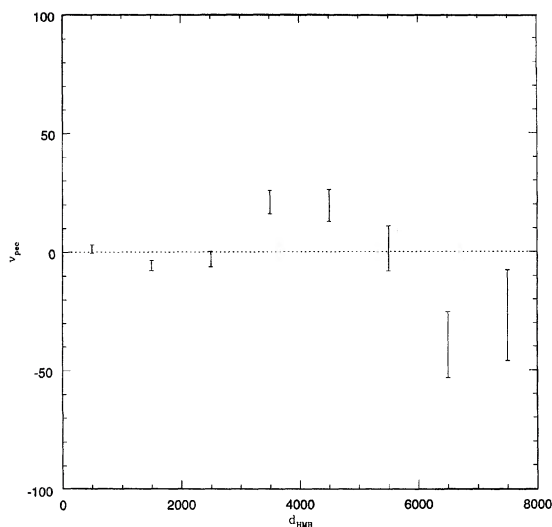
The similarity of this map with maps of the predicted

FIG. 7. The biases shown in Fig. 6 binned in HMB corrected estimated distances bins. Note the very small amplitude of the bias if averaged over the whole sky.
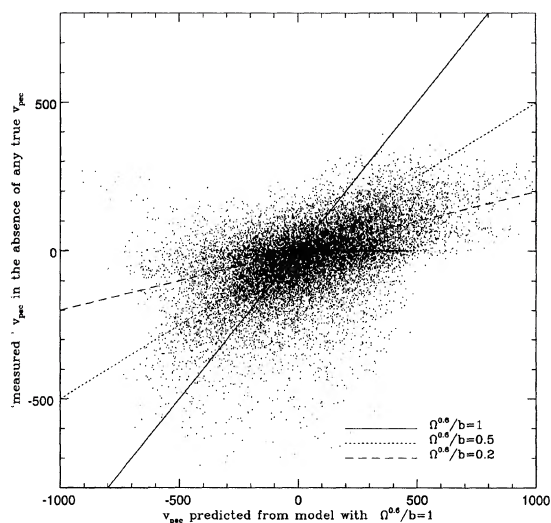

FIG. 8. Estimated HMB corrected peculiar velocities for a model without true peculiar velocities plotted vs the peculiar velocity as predicted by the *Optical-IRAS* model for $\beta=1$. If no biases were present, the two quantities should not correlate. The observed correlation between estimated and predicted velocities could be mistaken as evidence for gravitationally induced streaming motion if biases are not taken into account. The lines show expected correlations for several different values of β. The observed correlation is similar to the one expected for values of β between 0.2 and 0.5.

peculiar velocities from redshift surveys such as the one presented by Freudling *et al.* (1994) is evident. Note that these strong local biases average out if binned into homogeneous Malmquist bias corrected estimated distance bins. This is shown in Fig. 7. This shows that the HMB correction works well globally, but fails locally because of the large density variations.

Above, we noted that the bias field in estimated distance space looks similar to peculiar velocity maps predicted from the *Optical-IRAS* model reconstruction from redshift surveys. To quantify this impression, we present in Fig. 8 the apparent velocities shown in Fig. 6 and generated by the bias versus the velocities predicted from the *Optical-IRAS* model. It is such a plot which in principle can be used to derive a value for β. The comparison of the distribution of points with the lines for different β suggests that from the data one would derive a β in the interval 0.2–0.5 even when the true peculiar velocities are zero, i.e., the true $\beta\approx0$. This demonstrates that β cannot be derived from data which have been corrected only for HMB.

In the previous section, it was argued that distances estimated with the inverse TF relation are not dominated by Malmquist bias if they are plotted as a function of redshift. Instead, such distances are biased by selection effects (See Fig. 2). In order to demonstrate this difference, we plot in Fig. 9 the biases for a model identical to the one shown in Fig. 6 but with the scatter in the TF relation taken to be in the linewidth, and the biases are shown in redshift space. It can be seen that the biases show a completely different behavior than those in estimated distance space shown in Fig. 6. They are mainly a function of distance, and do not concentrate around high density regions. For the chosen magnitude limited sample of 14.5, the biases are significantly smaller than the one seen in Fig. 6.

5. APPLICATIONS

While the purpose of the previous discussion was to understand the properties of Malmquist bias in combination with selection effects, the purpose of this section is to predict the magnitude of these effects for real surveys. Therefore we create simulated redshift-distance surveys which correspond to samples actually observed. A summary of the surveys considered and their selection criteria is listed in Table 2. The samples considered are those of Lynden-Bell *et al.* (1988, hereafter referred to as 7S), Willick (1990, hereafter referred to as W90), and the new Sc sample.

5.1 7S Sample

The parameters for the simulation of the 7S sample of elliptical galaxies are given in Table 3 and the incompleteness as a function of magnitude was taken from Faber *et al.* (1989). Since the TF relation is substituted by a $D_n-\sigma$ relation (see Table 2), the color–linewidth relation must also be replaced by the equivalent relation. We could use a relation between $D-D_n$ and σ, derived from the data, which plays the same role as the color–linewidth relation and allows for possible correlation between the parameters used to select the sample with both of the quantities of the distance relation. No significant correlation between $D-D_n$ and σ was found, therefore D_n was computed from D by adding the constant and Gaussian scatter given in Table 3.
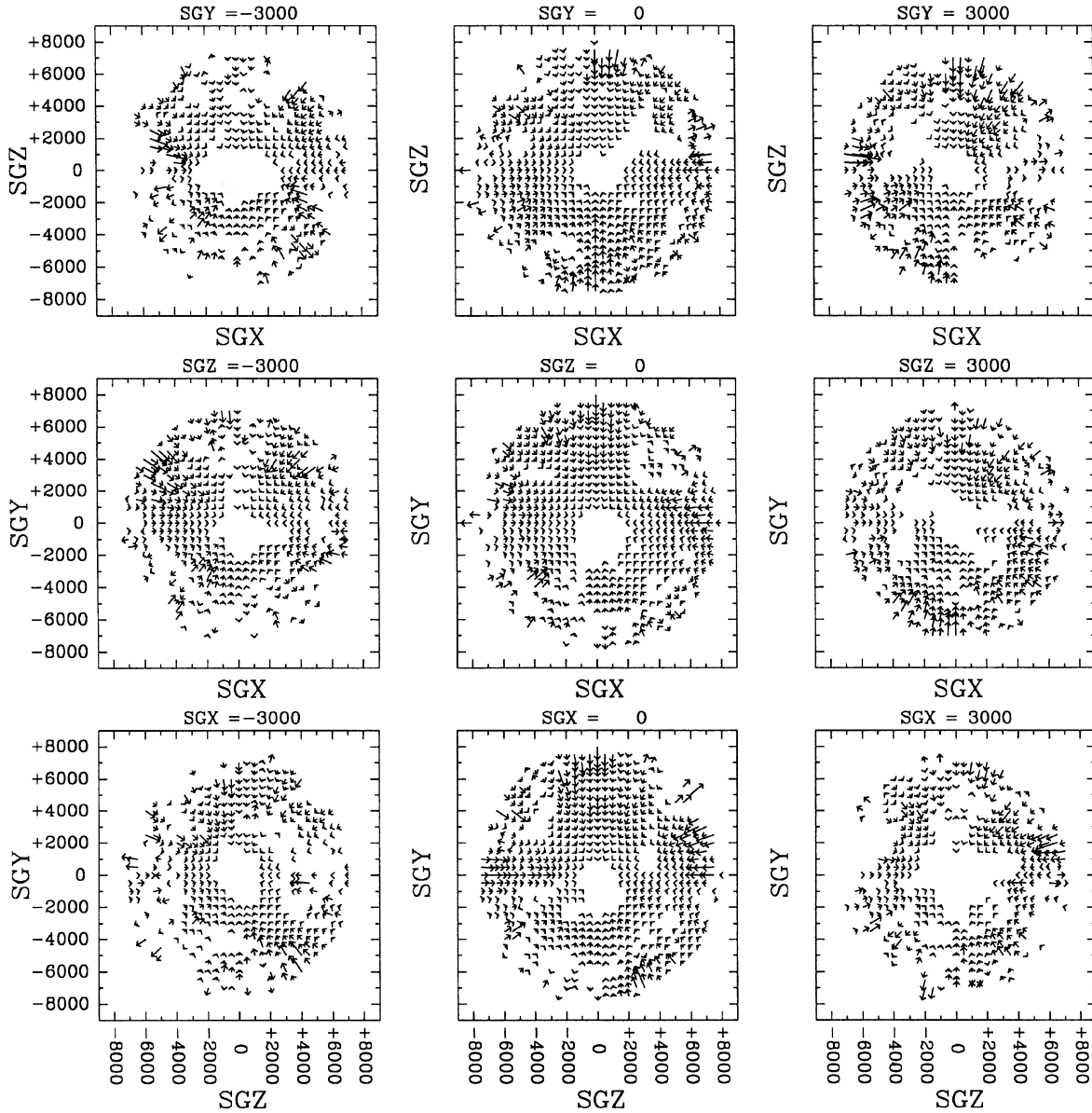
FIG. 9. Biases of distances derived from the TF relation for a 14.5 mag limited sample shown in redshift space. The bias map shown here is again plotted as a function of Supergalactic coordinates. The velocity vectors are five times the scale of distances. This is to be compared to Fig. 6, which shows a similar plot shown in corrected estimated distance space, which has a completely different behavior of the biases.

TABLE 2. Selection criteria of samples.

Sample	catalogs	type	area	mag.	sizes	redshifts in km/sec
Sc	UGC	Sc	$|b| > 10°$		$5' > d > 2.5'$	< 3000
	ESO				$5' > d > 1.5'$	3000 to 5000
	MGC				$5' > d > 1.3'$	5000 to 7500
7S	RC2	E	$|b| > 2.5°$	$B_T^o < 13.0$	$d > 2'$	< 8000
	RSA					
	UGC					
	ESO					
Willick	UGC	Spirals	$70° < l < 160°$	< 16.5	$d > 0.6'$	
	CGCG		$b < 0°$			

TABLE 3. Model parameters for E+S0.

Luminosity	function
D^* km s^{-1}	2489
a	-4.62
b	-1.66
σ_M	0.396
$D_n - \sigma$	relation
$\sigma_{Dn-\sigma}$	0.6
$a_{Dn-\sigma}$	1.2
σ_m	0.3
$D - D_n$	0.05
σ_{D-D_n}	0.5

In Fig. 10, the derived bias field for the 7S sample is shown. The biases are again shown in estimated HMB corrected distances space. The biases produce some apparent infall into the Great Attractor (at SGX≈ -4200 km s^{-1}, SGY $\approx +800$ km s^{-1}, SGZ≈ -700 km s^{-1}), but the effect is not much stronger than in other parts of the sky. In order to address the question whether such biases could produce the velocity field as suggested by 7S, in Fig. 11 we bin the biases as a function of the cosine of the angular distance from the GA in several distance shells. The binning and presentation corresponds exactly to the one used by 7S in their Fig. 5. It can be seen that the biases binned in such a manner are independent of the angular distance from the GA, unlike the behavior the 7S find in their data. We conclude that residual
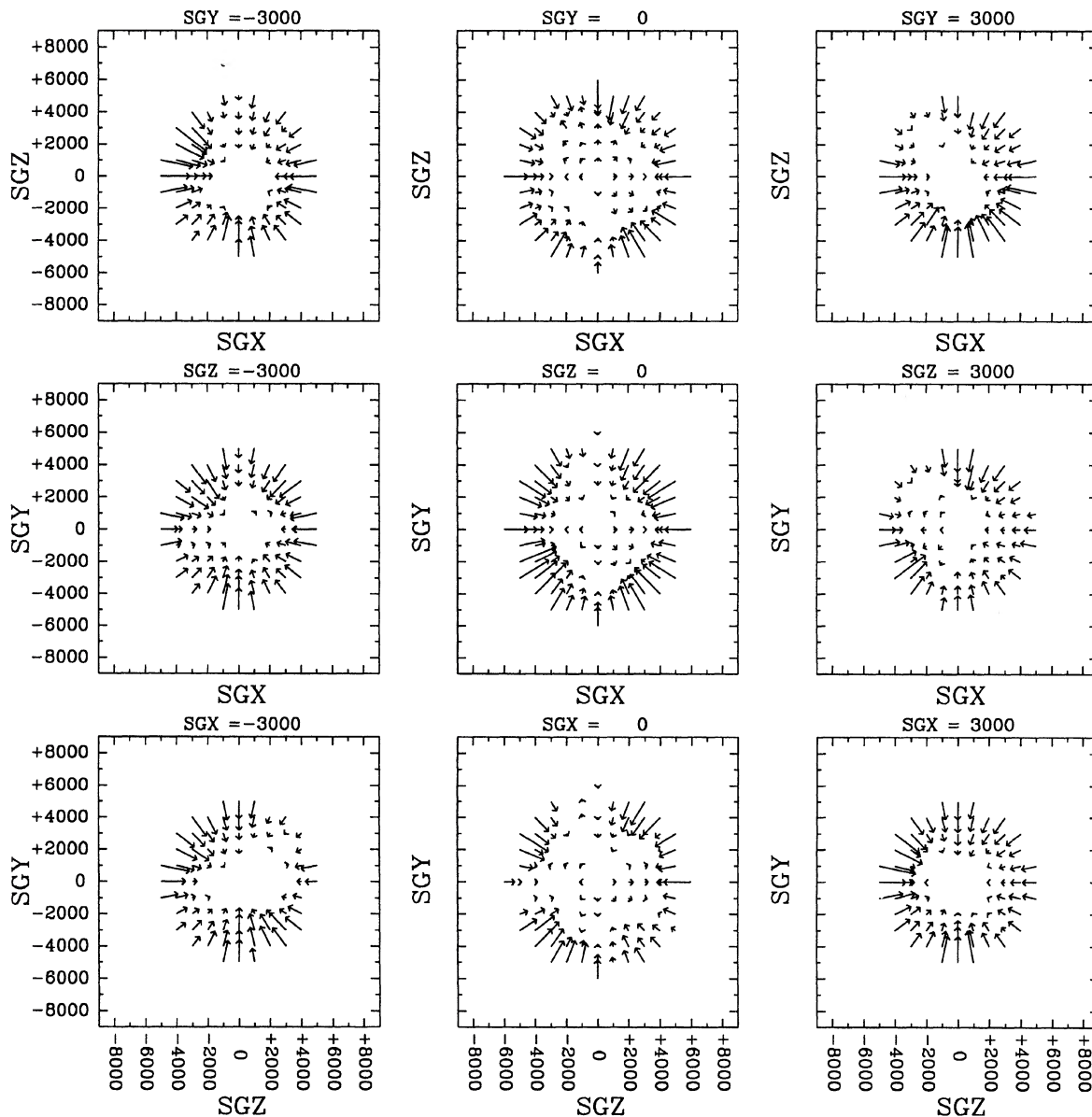

FIG. 10. Simulated biases for the 7S sample, binned in estimated distance space. The scale of the arrows is 5 times the scale of distances.
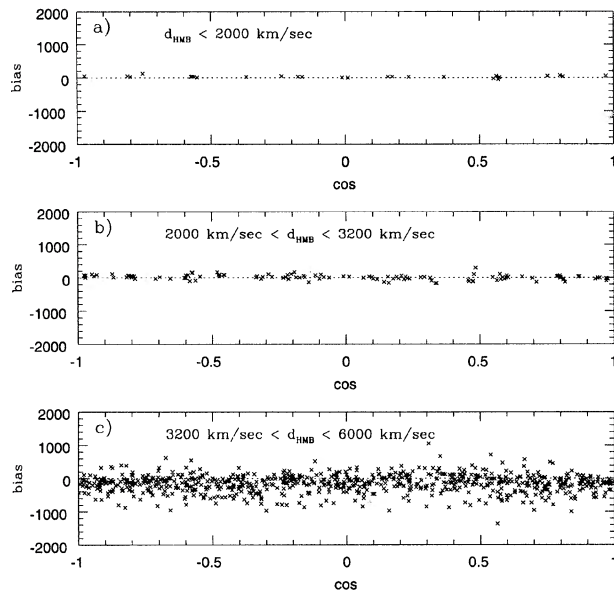
FIG. 11. The unsmoothed biases from Fig. 10 plotted as a function of the cosine of the angular distance from the Great Attractor in 3 estimated distance ranges, similar to Fig. 5 in 7S. It can be seen that conclusions by 7S based on this plot are not subject to biases.

biases do not significantly affect the conclusion of 7S regarding the existence of a large infall towards the GA within the volume considered.

5.2 W90 Sample

In order to estimate the incompleteness of the W90 sample, we have binned the data in 0.5 mag intervals and
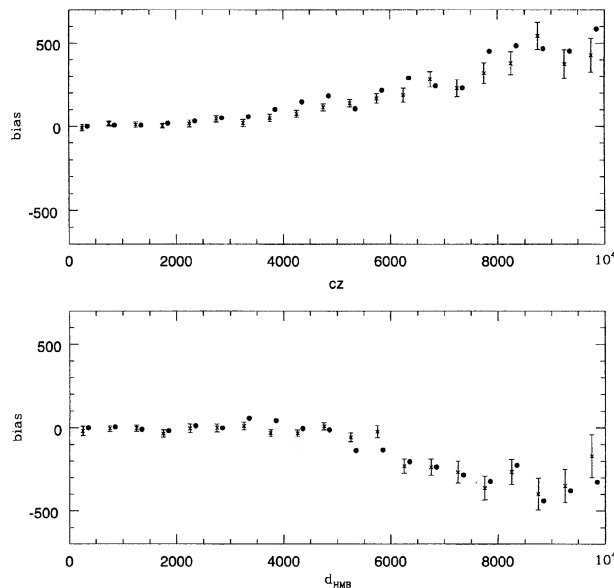

FIG. 12. Biases computed for the simulated sample of Willick (1990). The panel (a) shows the biases as a function of redshift. Panel (b) shows the biases as a function of HMB corrected distances. The error bars show the biases computed for a model with a velocity field as described in the text. The solid points show the biases for a model without any true velocities. The solid points are offset horizontally from their true position by 100 km s^{-1} in order to make them easier to distinguish from the error bars.
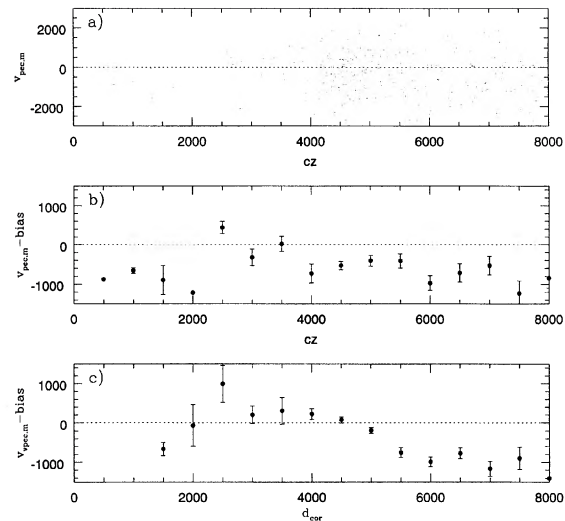

FIG. 13. The W90 sample corrected for biases and binned in bins of 500 km s^{-1}. Panel (a) shows the raw data as a function of redshift in the CMB rest frame. Panel (b) shows the corrected peculiar velocities binned in redshift bins of 500 km s^{-1} size. Panel (c) shows the corrected data binned in equivalent estimated distance bins.

compared the resulting counts to those obtained in the CGCG (Zwicky *et al.* 1961–68) catalogue. This incompleteness was imposed on the simulated sample. Subsequently, the biases were computed using the direct TF relation as discussed in Sec. *3.3*. It should be noted that the TF relation we use differs from that assumed by Willick (1990). In principle we could use the same quadratic relation, but we believe that the major source of uncertainty stems instead from
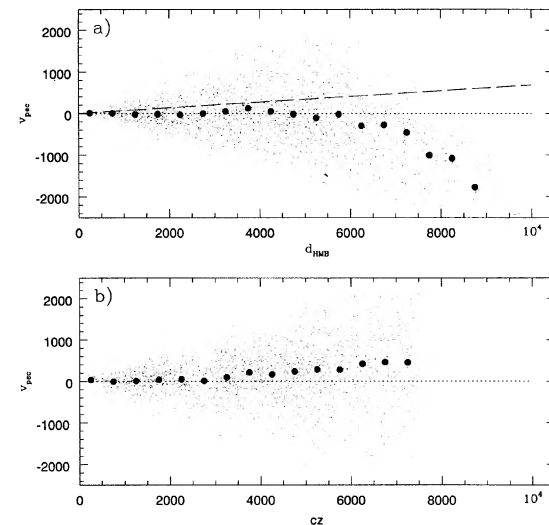

FIG. 14. HMB corrected estimates of the peculiar velocities of a sample of galaxies selected like the Sc sample, binned in estimated distances [panel (a)] and redshifts [panel (b)]. The small points indicate individual galaxies, whereas the large points are averages in bins of 500 km s^{-1}. The dashed line is the computed homogeneous Malmquist bias. The most prominent feature when plotted as a function of the estimated distance is the redshift cutoff at 7500 km s^{-1}, which has the effect that galaxies with a redshift>7500 km s^{-1} are only included if they have negative peculiar velocities.
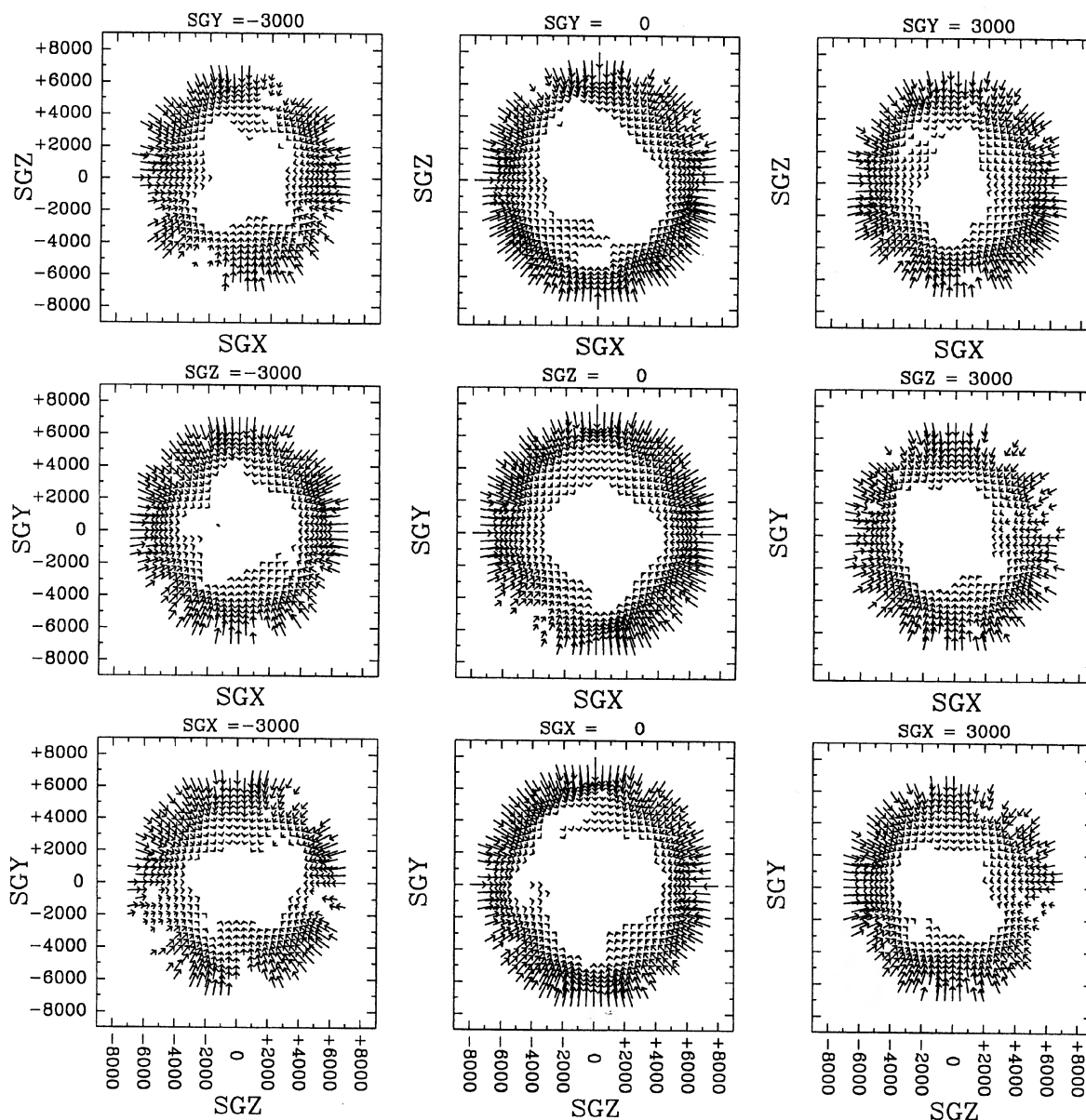
FIG. 15. Bias map for the simulated sample of Haynes *et al.* (1993). The unit of the velocity vectors is identical to the unit of distances. The vectors are plotted at the position of the estimated distances corrected for homogeneous Malmquist bias.

the lack of detailed information regarding the properties of the scatter in the TF relation. In particular, the sample contains galaxies as faint as magnitude 16.5, but the number of galaxies in each magnitude interval is small. This leaves the possibility of substantial unrecognized systematic effects in the incompleteness of the sample.

In Fig. 12, the biases for the simulated W90 sample is shown. The figure shows the biases both as a function of redshift and as a function of d_{HMB}. The biases are shown separately for models with no true peculiar velocities and with a velocity field given by the *Optical-IRAS* model. The biases are essentially independent of the assumed velocity

model. A comparison of the two panels in Fig. 12 shows that for a sample selected like W90, the biases out to 5000 km s^{-1} are smaller if the peculiar velocities are plotted as a function of d_{HMB} rather than redshifts. Beyond that distance, the bias is similar in amplitude but with opposite sign.

In Fig. 13, the W90 data are corrected for the bias shown in Fig. 12 and binned in bins of 500 km s^{-1}. Panel (a) corresponds to Fig. 1 shown by W90, whereas panels (b) and (c) show the bias corrected data as a function of redshift and distance, respectively. A significant infall into the Pisces–Perseus supercluster complex can be seen both in panels (b) and (c). While the corrections (as shown in Fig. 12) are sub-

stantial, the basic conclusion of W90 that there is a large bulk motion of the whole region in the direction of the Local Group cannot be ascribed to the biases discussed in this paper.

5.3 Sc Sample

Samples selected using a redshift criterion such as the Sc sample are a completely different case. The Sc sample also differs from the previously discussed sample by the fact that a large fraction of all galaxies which satisfy the selection criteria have actually been observed. This lowers the probability for large biases due to unrecognized systematic effects in the selection of the sample. In Fig. 14 the "measured" peculiar velocities of a sample of galaxies selected with the Sc sample criteria is shown as a function of estimated distance and redshift. As before, the estimated distances in Fig. 14 have been HMB corrected. At small distances (up to about 6000 km/s) the biases in addition to the HMB are small. This reflects the changing diameter limit as a function of redshift adopted in selecting the sample. On the other hand, at large distance the peculiar velocities are strongly affected by the redshift cutoff which leads to the depopulation of galaxies with positive peculiar velocities. The full bias map which we will actually use in our analysis of the Sc sample is shown in Fig. 15, again plotted against HMB corrected distances.

6. SUMMARY

We have presented a new approach to correct observed peculiar velocity fields. The Monte Carlo approach we have adopted allows us to simultaneously include all known sources of biases in a realistic way. The resulting simulated biases can be applied to existing observed samples with complex but well-defined selection criteria, such as the new Sc sample.

We have tested the MC simulation in simple cases, for which bias correction schemes are known, and recovered the expected biases with high accuracy. We have also applied our technique to real samples and computed the expected biases. We found that the results presented by Lynden-Bell *et al.* are virtually free of biases. On the other hand, the data as presented by Willick (1990) are significantly biased, but the biases do not affect his main conclusion. Finally, we derived the expected bias for the new Sc sample of Haynes *et al.*(1993) that will be used in future analysis of this sample.

REFERENCES

Bothun, G. D., & Cornell, M. E. 1990, AJ, 99, 1004
Courteau, S., Faber, S. M., Dressler, A., & Willick, J. A. 1993, ApJ, 412, 51
Dekel, A. 1993, in Observational Cosmology, ASP Conf. Ser. Vol. 51, edited by G. Chincarini, A. Iovino, T. Maccacaro, and D. Maccagni, p. 194
Dekel, A. 1994, ARA&A, 32, 371
Faber, S. M., Wegner, G., Burstein, D., Davies, R. L., & Dressler, A. 1989, ApJS, 69, 763
Feast, M. W. 1987, Observatory, 107, 185
Freudling, W., da Costa, L. N., & Pellegrini, P. S. 1994, MNRAS, 268, 943
Giovanelli, R., *et al.* 1995, in preparation
Haynes, M. P., Giovanelli, R., Salzer, J. J., Wegner, G., da Costa, L. N., Freudling, W., & Chamaraux, P. 1993, BAAS, 25, 1403
Hudson, M. J. 1994, MNRAS, 266, 468
Landy, S. D., & Szalay, A. S. 1992, ApJ, 394, L25

Lynden-Bell, D., Faber, S. M., Burstein, D., Davies, R. L., Dressler, A., Terlevich, R. J., & Wegner, G. 1988, ApJ, 326, 19
Nilson, P. 1973, Uppsala General Catalog of Galaxies, Uppsala Astron. Obs. Ann. 6 (UGC)
Schechter, P. L. 1980, ApJ, 85, 801
Sodré, Jr. L., & Lahav, O. 1993, MNRAS, 260, 285
Teerikorpi, P. 1993, A&A, 280, 443
Willick, J. A. 1991, Ph.D. thesis, University of California at Berkeley
Willick, J. A. 1990, ApJ, 351, L5
Willick, J. A. 1994, ApJS, 92, 1
Willmer, C. N. A., da Costa, L. N., Pellegrini, P. S., Fairall, A. P., Latham, D. W., & Freudling, W. 1995, AJ, 109, 61
Zwicky, F., Herzog, E., Will, P., Karpowicz, M., & Kowal, C. 1961–1968, Catalog of Galaxies and Clusters of Galaxies in six volumes (California Institute of Technology, Pasadena) (CGCG)