The EFAR Project: Monte Carlo Testing of the Fundamental Plane Distance Estimator*

R.P. Saglia[1], Matthew Colless[2], G. Baggley[3], Edmund Bertschinger[4], David Burstein[5], Roger L. Davies[3], Robert K. McMahan, Jr.[6] and Gary Wegner[7]

[1] Institut für Astronomie und Astrophysik, Scheinerstraße 1, D-81679 Munich, Germany
[2] Mount Stromlo and Siding Spring Observatories, The Australian National University, Weston Creek, ACT 2611, Australia
[3] Dept of Physics, University of Durham, South Road, Durham, DH1 3LE, UK
[4] Dept of Physics and Astronomy, Arizona State University, Tempe, AZ 85287-1504
[5] Dept of Physics, MIT, Cambridge, MA 02139
[6] Dept of Physics and Astronomy, University of North Carolina, CB#3255 Phillips Hall, Chapel Hill, NC 27599-3255
[7] Dept of Physics and Astronomy, Dartmouth College, Wilder Lab., Hanover, NH 03755

Abstract. The EFAR project is a long-term study of 736 candidate early-type galaxies in 84 clusters lying in two directions towards Hercules-Corona Borealis and Perseus-Pisces-Cetus at distances 6000-15000 km/s. In the following we discuss the goals of the project and the properties of the photometric and spectroscopic database. We present a new method, based on maximum likelihood, to determine the parameters of the Fundamental Plane and describe the distribution of galaxies in it. The algorithm takes into account the effects of errors and selection and is superior to least-squares procedures.

1 Introduction

The EFAR (Ellipticals FAR away) project started in 1986 to construct an accurate and homogeneous photometric and spectroscopic database for a large sample of early-type galaxies with redshifts 6000-15000 km/s, distributed in 84 clusters in the Hercules-Corona Borealis and Perseus-Pisces-Cetus regions. Our goals were to study the structural properties of elliptical galaxies and determine their peculiar motions in two independent regions outside the local supercluster using the Fundamental Plane distance estimator. Both scientific goals are important, since they provide powerful tests of theories of galaxy evolution and of the formation of large scale structure in the Universe.

We have completed the observing and data reduction phases of the project and are in the process of publishing the database and analyzing it. The following papers describe the results obtained so far. Colless et al. (1993) present the photoelectric photometry database, Saglia et al. (1993) discuss the effects of seeing

* To appear in the Proceedings of the 3^{rd} ESO-VLT Workshop "Galaxy Scaling Relations: Origins, Evolution and Applications", Garching (18-20 November 1996)

on the photometric properties of early-type galaxies, Wegner et al. (1996, Paper I) introduce the EFAR cluster and galaxy sample and its selection properties, Wegner et al. (1997, Paper II) present the spectroscopic database, Saglia et al. (1997a, Paper III) the photometry database, and Saglia et al. (1997b, Paper IV) the photometric fitting technique. In the following we summarize the properties of the photometric and spectroscopic database (Section 2) and describe the maximum likelihood method we have developed to determine the EFAR Fundamental Plane (Section 3). Section 4 gives our conclusions.

2 The photometric and spectroscopic database

We invested 185 nights of photoelectric and CCD observations with 1m class telescopes and more than 100 nights of spectroscopic observations with 2m to 4m class telescopes to build up the EFAR database.

We collected 2846 R band CCD images for 776 galaxies with 87% of the galaxies having more than one profile. A total of 1278 photometric calibrations were obtained using photoelectric photometry (in both the R and the B bands) and CCD R band photometry. Repeated calibrations show a zero-point precision of 0.03 mag per observation (see Paper III). We computed the circularly averaged surface brightness profiles and the photometric parameters of the EFAR galaxies using a seeing-convolved, $R^{1/4}$ plus exponential fitting algorithm which optimally combines multiple profiles and corrects for sky subtraction errors (see Paper IV).

The precision of the derived D_n diameters, half-luminosity radii R_e, total magnitudes m_T and average effective surface brightness $\langle SB_e \rangle$ is estimated with the help of extensive Monte Carlo simulations to test for systematic effects. We find that the systematic errors on m_T and R_e are less than 0.15 mag and 25% respectively for 90% of our sample, while the systematic errors on the combined quantity $\log R_e - 0.3\langle SB_e \rangle$ which enters the Fundamental Plane equation always remain smaller than 0.03 dex, and remain below 0.015 dex for 90% of the galaxies. The morphological classification based on CCD images shows that 32% of the sample objects, visually selected from photographic images as early-type, are in fact spirals or barred S0 galaxies. The remaining 68% classified as early-type can be subdivided into cD (8%), E (11% with a simple $R^{1/4}$ law best fit) and E/S0 (49% with a disk plus bulge best fit). Therefore most of the profiles deviate significantly (at the 3σ level and in 90% of the cases at more than the 5σ level, see Paper III) from a de Vaucouleurs law, needing a non-negligible exponential component, with a typical "disk-to-bulge" ratio $D/B \approx 0.5$. In these cases, photometric parameters derived using a simple $R^{1/4}$ law (as commonly done in the literature) are affected by large systematic errors, easily larger then 0.3 mag on m_T or 0.2 dex on $\log R_e$.

A total of 1327 spectra (see Paper II) have been collected for 659 EFAR galaxies (1250 spectra) plus 48 galaxy standards (77 spectra). We used both long-slit and fiber-fed spectrographs, spanning MgI b, the Fe lines at 5207 Å and 5269 Å, and (usually) Hβ for cz= 6000 − 15000 km/s, with some spectra going from CaI H+K (3933 Å) to NaI D (5892 Å). The typical resolution was

$\Delta\lambda = 1-2$ Å/pixel, or $\sigma_{instr} = 80-150$ km/s, and we reached a typical signal-to-noise ratio S/N$\approx$ 26 per Å. We derive galaxy redshift and central velocity dispersions σ using the cross-correlation algorithm and we estimate the precision of the results by means of Monte Carlo simulations. The median error on the combined redshift measurements for each galaxy is 18 km/s; 90% of the redshifts have errors less than 36 km/s. Likewise, the median error on the combined σ measurements is 9% per galaxy, and 90% have errors less than 18%. We can determine reliably the Mgb index, while we find that the Mg_2 index is subject to significant systematic effects arizing from continuum shape variations. We apply aperture corrections to both σ and Mgb values which take into account the slit width, the galaxy distance and the galaxy size. The Mgb–σ relation for the EFAR galaxies does not deviate from previous determinations. The histogram of the zero-points of the relation for the sample of the EFAR clusters has an rms of 0.02 dex. This suggests that environmental effects, if present, are small.

3 A maximum likelihood Fundamental Plane algorithm

Figure 1 shows the distribution of EFAR galaxies in the $\log R_e, \log\sigma, \langle SB_e\rangle$ space which was constructed using redshift distances. Early-type galaxies do not fill this three-dimensional space, rather they occupy a narrow region around a Fundamental Plane defined by the equation $\log R_e = a\log\sigma + b\langle SB_e\rangle + c$ (Dressler et al. 1987, Djorgovski & Davis 1987). The canonical methods used to determine the coefficients a and b are of the least-squares type: one minimizes the orthogonal (absolute or squared) residuals from the plane (Jørgensen et al. 1996, Pahre 1996, Scodeggio et al. 1996) or the (square) residuals from one of the variables (Smith et al. 1996).

We find that this approach can lead to biased coefficients, if selection effects, sizeable random errors and well-defined exclusions (e.g., due to resolution we cannot observe galaxies with $\sigma \leq \sigma_{cut}$ km/s) are present, as in the case of the EFAR sample. Fig. 2, top, shows the histograms of the coefficients recovered using the regressions on one of the variables, for 99 Monte Carlo realizations of the EFAR data sample, which take into account the actual photometric and spectroscopic errors. We assumed that only $\sigma > \sigma_{cut} = 120$ km/s could be measured. The three regressions (and any sort of mean of them) perform very poorly. The mean values of the reconstructed parameters are $a = 0.90, 1.47, 0.93$ and $b = 0.33, 0.34, 0.37$ for the regressions against $\log R_e$, $\log\sigma$, and $\langle SB_e\rangle$ respectively, while the true values are $a = 1.3$, $b = 0.35$. Fig. 2, bottom, shows the histograms of the coefficients recovered using an improved orthogonal minimization algorithm. The algorithm excludes points falling below an orthogonal cut determined by the largest $\sigma < \sigma_{cut}$, with an optional selection weighting, where every point is inversely weighted by the sample selection function, based on galaxy diameter (see Eq. 6). While the improved orthogonal minimization with selection weighting gives the best result, still the mean of the recovered coefficients is biased.

Better results can be obtained by modeling the distribution of galaxies within

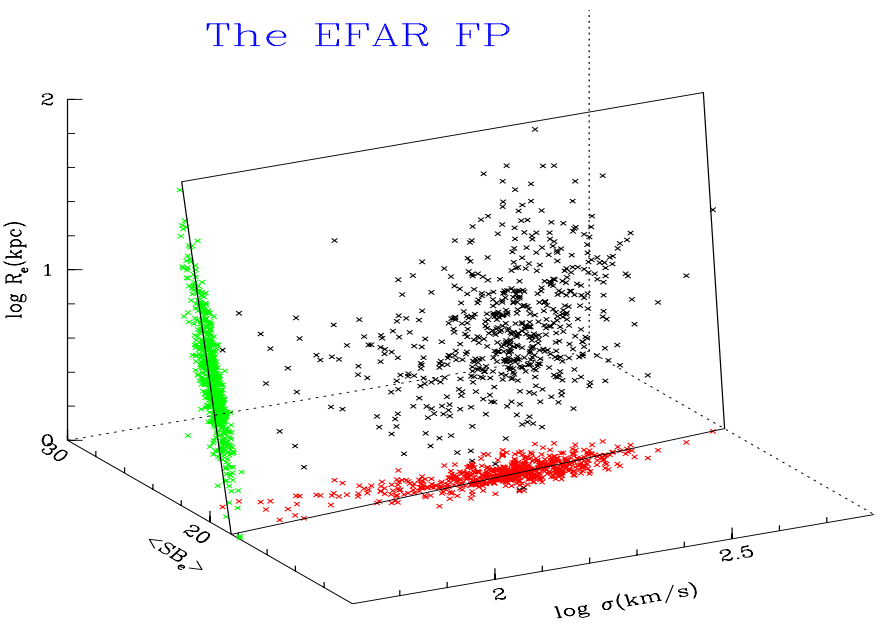

Fig. 1. The distribution of EFAR galaxies in the $\log R_e, \log \sigma, \langle SB_e \rangle$ space, using redshift distances.

the FP. Let us define the coordinates relative to the cosmic averages as $r = \log R_e - \overline{\log R_e} + \log(1+z_j)$, $s = \log \sigma - \overline{\log \sigma}$, $u = \langle SB_e \rangle - \overline{\langle SB_e \rangle}$, where z_j are the cluster peculiar redshifts, and the set of rotated coordinates v_1, v_2, v_3:

$$
\begin{aligned}
v_1 &= \frac{r - as - bu}{\sqrt{1 + a^2 + b^2}}, \\
v_2 &= \frac{br + u}{\sqrt{1 + b^2}}, \\
v_3 &= \frac{-ar - \sqrt{1 + b^2}s + abu}{\sqrt{1 + b^2}\sqrt{1 + a^2 + b^2}},
\end{aligned}
\tag{1}
$$

where v_1 measures the displacement from the FP, and v_2 and v_3 are two orthogonal vectors in the FP. A look at the distribution of EFAR galaxies (Figure 3) shows that with this choice of coordinates the (normalized) distribution P of galaxies in the FP can be factorized as $P(|v >= (v_1, v_2, v_3)) = f_1(v_1) f_2(v_2) f_3(v_3)$. Moreover, it is not unreasonable (for example, see Willick

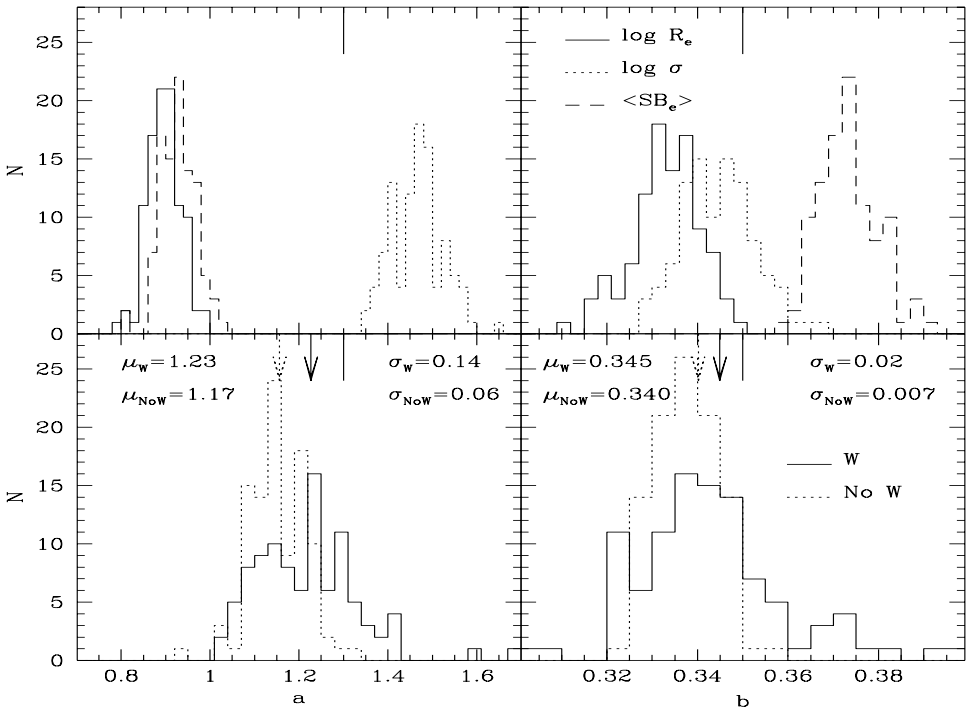

Fig. 2. The histograms of the FP coefficients a and b recovered with different least-square procedures for 99 Monte Carlo realizations of the EFAR data sample, without peculiar velocities. The top panels show the results of the regressions against $\log R_e$ (full line), $\log \sigma$ (dotted line), and $\langle SB_e \rangle$ (dashed line). The bottom panels show the results of the orthogonal residual minimization, with selection function weighting (full line) and without (dotted line). The bars at the top show the true values of a (1.3) and b (0.35), the arrows the mean values recovered. The mean μ and dispersion σ of the distribution of the recovered values are also given for the case with selection weighting (label W) and without (label NoW).

1994) to assume that f_1, f_2, f_3 can be well approximated by gaussians:

$$P(|v>) = \frac{1}{(2\pi)^{3/2}\sigma_1\sigma_2\sigma_3} \exp\left[-\left(\frac{v_1^2}{2\sigma_1^2} + \frac{v_2^2}{2\sigma_2^2} + \frac{v_3^2}{2\sigma_3^2}\right)\right], \qquad (2)$$

where σ_1, σ_2, σ_3 are the relative sigmas of the distributions, with $\sigma_1 << \sigma_2, \sigma_3$. When written as a function of the original coordinates $|x>= (r, s, u)$ the distribution P becomes:

$$P(|x>) = \frac{|F|^{1/2}}{(2\pi)^{3/2}} \exp(- <x|F|x> /2), \qquad (3)$$

where we have adopted the quantum mechanical formalism for the scalar product, so that $|v>= R|x>$ and R is the rotation matrix defined in Eq. 1. In addition one has $F = R^T DR$, $D_{ij} = 1/\sigma_i^2 \delta_{i,j}$, and $|M|$ denotes the determinant of the matrix M. F is called the covariance matrix and is symmetric and positively defined; its inverse $V = F^{-1}$ is the variance matrix.

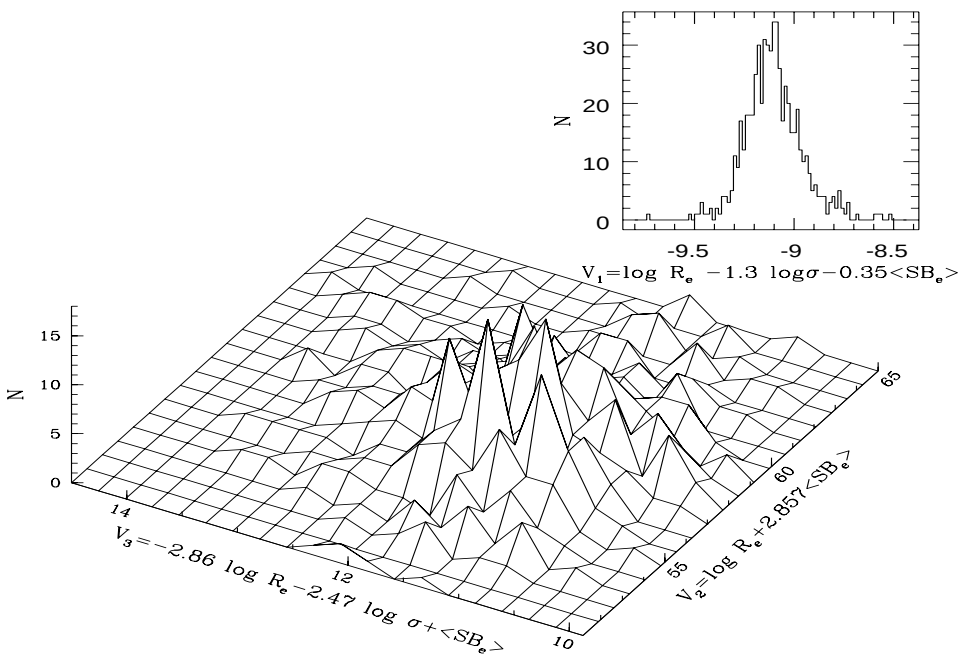

Fig. 3. The distribution of EFAR galaxies within the FP of coefficients $a = 1.3$, $b = 0.35$. The inserted plot to the upper right shows the distribution orthogonally to the plane.

The observed galaxy distribution deviates systematically from the true one (which we model by Eq. 2), since measurement errors and selection effects are present. Assuming gaussian error distributions, measurement errors act as a convolution of Eq. 3 with a gaussian function of variance matrix E:

$$E = \begin{pmatrix} (1+\alpha^2)\delta r^2 & 0 & \frac{(1+\alpha^2)\delta r^2 - \delta FP^2}{\alpha} \\ 0 & \sigma_s^2 & 0 \\ \frac{(1+\alpha^2)\delta r^2 - \delta FP^2}{\alpha} & 0 & \frac{(\alpha^4-1)\delta FP^2 + (1+\alpha^2)^2 \delta r^2)}{\alpha^2(1+\alpha^2)} + \delta ZP^2 \end{pmatrix}, \qquad (4)$$

where δr is the error on $\log R_e$, δFP is the error on the combined quantity $FP = \log R_e - \alpha\langle SB_e\rangle$, with $\alpha \approx 0.3$, δZP is the photometric zero-point error (see Paper III and IV), and σ_s the error on $\log\sigma$ (see Paper II). The resulting normalized distribution reads

$$P^c(|x>) = \frac{|C|^{1/2}}{(2\pi)^{3/2}} \exp(-<x|C|x>/2), \tag{5}$$

where $C = (E+V)^{-1}$.

Selection effects operate on all variables (r, s, u). Only velocity dispersions larger than the instrumental resolution σ_{cut} (see Paper II) can be measured, therefore $P^c(|x>) = 0$ if $s \leq s_{cut} = \log\sigma_{cut} - \overline{\log\sigma}$. The EFAR galaxy selection S is described in Paper I and is a function of the D_W diameter measured in arcsec:

$$S_j(\log D_W) = 0.5\left\{1 + \mathrm{erf}\left[\frac{\log D_W - \log D^0_{Wj}}{\delta_{Wj}}\right]\right\}, \tag{6}$$

where $\log D^0_{Wj}$ is the midpoint, δ_{Wj} the width of the cutoff in the selection function, and j is the cluster index. The diameter $\log D_W$ correlates with $FP = \log R_e - 0.3\langle SB_e\rangle$ with 0.09 dex scatter in FP (see Paper III). Taking into account this correlation, one can write the observed probability distribution as:

$$P^o(|x>) = \exp(-<x|C|x>/2)\theta(s - s_{cut}) \left[1 + \mathrm{erf}\frac{1.28FP + d_0 - \log D^0_W}{\sqrt{\delta^2_W + 2\sigma^2_W}}\right]/N, \tag{7}$$

where d_0 is a constant depending on the distance D, σ_W the dispersion in the relation between FP and $\log D_W$, $s_{cut} = \log\sigma_{cut} - \overline{\log\sigma}$, $\theta(s - s_{cut}) = 0$ if $s \leq s_{cut}$, $\theta = 1$ otherwise, and N is the normalization constant so that $\int P^o d^3x = 1$.

Given the sample of EFAR datapoints ($|x_i>$), the maximum likelihood choice of parameters $a, b, \overline{\log R_e}, \overline{\log\sigma}, \overline{\langle SB_e\rangle}, \sigma_1, \sigma_2, \sigma_3$, and the vector of cluster peculiar redshifts z_j can be determined by maximising the likelihood function $L = \prod P^o_i$. We adopt a simplified version of this approach, which yields an unbiased estimate of the parameters (Eadie et al. 1971), by maximizing the following quantity:

$$\begin{aligned} \ln Prob &= \sum_{\sigma>\sigma_{cut}} \frac{1}{S_j(\log D^i_W)} \ln\frac{P^c(|x_i>)}{f_i} \\ &= -\sum_{\sigma>\sigma_{cut}} \frac{1}{S_j(\log D^i_W)} \left[0.5 <x_i|C_i|x_i> + \ln f_i + 1.5\ln(2\pi) - 0.5\ln|C_i|\right], \end{aligned} \tag{8}$$

where $f_i = \int P^c_i\theta(s - s_{cut})d^3x = 0.5(1 - \mathrm{erf}(y_i))$ with $y_i = s_{cut}/\sqrt{2|C_i|/|C^{(2,2)}_i|}$ is the fraction of galaxies with $\sigma > \sigma_{cut}$ and $C^{(2,2)}_i$ is the matrix obtained by dropping the second column and row of the matrix C_i.

Fig. 4 shows the histograms (full lines) of the FP distribution parameters recovered maximizing Eq. 8 for the same 99 Monte Carlo realizations of the EFAR data sample of Fig. 2. The plots at the top show the recovered coefficients of the Fundamental Plane a and b, the ones in the middle the values of the cosmic averages $\overline{\log R_e}, \overline{\log \sigma}, \overline{\langle SB_e \rangle}$ and of the normalizing coefficient f_i for one of the simulated galaxies. The plots at the bottom show the reconstructed values of σ_1, σ_2 and σ_3 (see Eq. 2) and the values of Eq. 8. On the mean, the parameters are recovered without bias. As a second step, we add a random "peculiar redshift field" $\log(1+z_j)$ of rms=0.05 dex, which corresponds to peculiar velocities of the order of 12% of the distance, or to the (rather large for testing purpose) peculiar velocity of 1200 km/s at 10000 km/s (a typical redshifts of the EFAR clusters). The same result as before holds (Fig. 4, dotted lines) if the parameters are reconstructed ignoring first the peculiar redshifts. The rms of the mean values are however larger. In addition, the reconstructed σ_1 are systematically larger than the input value.

The Monte Carlo simulations can also be used to estimate how well we can determine the peculiar redshifts of the galaxy groups (z_j) if we fix the FP slope to the determined values. Fig. 5(a) shows the values of $\log(1+z_j)$ we obtain from the simulations by taking the average over the 99 Monte Carlo realizations, versus the input values. The relative rms are also given. In order to fix the FP zero-point, we assume that one cluster (our best-observed "reference cluster") has zero peculiar redshift, although in our simulation we have $\log(1+z_{ref}) = 0.15$ dex to provide a severe test for our distance predictions. Fig. 5(a) shows that the method determines the peculiar redshifts accurately, even when the best-observed cluster has the second largest peculiar redshift. The mean systematic residuals are of the order of 0.01 dex (Fig. 5(b)), or $\approx 2\%$ of the distance (200 km/s at 10000 km/s redshift). The clusters with the largest residuals (≈ 0.03) have ≤ 3 objects only; smaller systematic residuals are achieved for clusters with larger numbers of galaxies (Fig. 5(c)). The random errors are of the order of 0.005 dex for clusters with four or more galaxies (Fig. 5(d)).

4 Conclusions

We described the first results of the EFAR project, which measured the structure parameters of 434 early-type galaxies at redshifts 6000-15000 km/s in 84 clusters in the Hercules-Corona Borealis and Pisces-Perseus-Cetus regions. We discussed the properties of the photometric and spectroscopic data, together with new accurate methods used to determine these quantities. We examined the usual least-squares based algorithms to determine the parameters of the Fundamental Plane. Simulations show that these methods can provide biased answers due to the combined effects of selection, random errors and exclusions. These potential problems can be solved by modeling the distribution of galaxies in the Fundamental Plane in a maximum likelihood sense. We tested our algorithm on simulated EFAR catalogues and found that it provides unbiased estimates of the searched parameters and of the underlying peculiar velocity field.

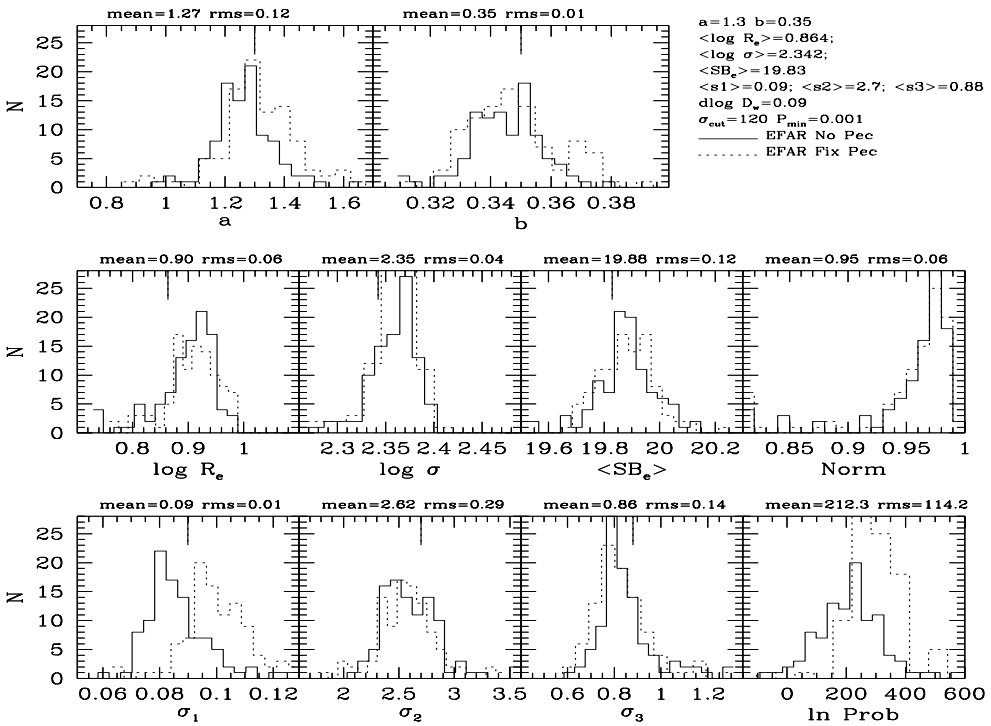

Fig. 4. Full lines: the histograms of the FP distribution parameters recovered with the maximum likelihood algorithm for the 99 Monte Carlo realizations of the EFAR data sample of Fig. 2. Mean and rms values are given at the top. Dotted lines: the same quantities reconstructed from simulations with a random "peculiar redshift field" $\log(1+z_j)$ of rms=0.05 dex.

RPS acknowledges the financial support by the Deutsche Forschungsgemeinschaft under SFB 375. Support was also received from the Durham Visitors grant PPARC GR/K44992 and Durham University.

References

Colless M., Burstein D., Wegner G., Saglia R.P., McMahan R., Davies, R.L., Bertschinger E., & Baggley G., 1993, MNRAS, 262, 475

Djorgovski, S., Davis, M., 1987, ApJ, 313, 59

Dressler, A., Faber, S.M., Burstein, D., Davies, R.L., Lynden-Bell, D., Terlevich, R.J., Wegner, G., 1987, ApJ 313, L37

Eadie, W.T., Dryard, D., James, F.E., Roos, M., Sadoulet, B., 1971, Statistical Methods in Experimental Physics, North-Holland Publishing Company, Amsterdam

Jørgensen, I., Franx, M., Kjærgaard, P., 1996, MNRAS, 280, 167

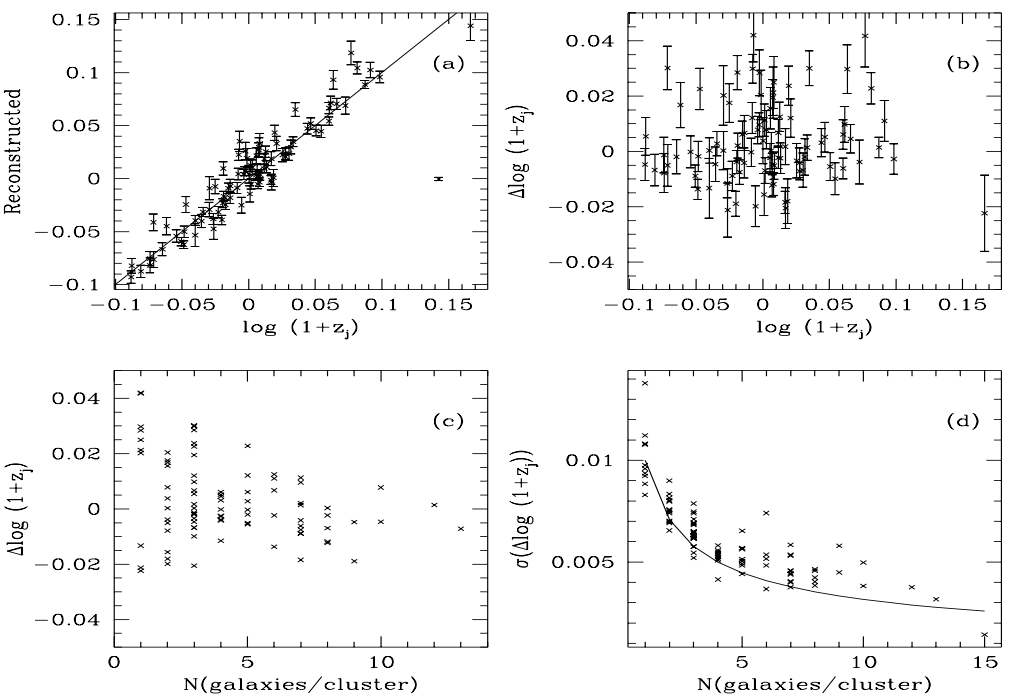

Fig. 5. (a) Peculiar redshifts $\log(1+z_j)$ as reconstructed with the maximum likelihood method, averaged over 99 Monte Carlo realizations, as a function of the input peculiar velocities. The peculiar redshift of the reference cluster Coma has been set to zero. The bars show the random errors.(b) Residual mean $\log(1+z_j)$ with their random errors. (c) Residual mean $\log(1+z_j)$ as a function of the number of galaxies per cluster. (d) Random errors as a function of the number of galaxies per cluster. The line shows the expected poissonian decline.

Pahre, M.A. 1996, this book

Saglia R.P., Bertschinger E., Baggley G., Burstein D., Colless M., Davies R.L., McMahan R., & Wegner G., 1993 MNRAS, 264, 961

Saglia R.P., Bertschinger E., Baggley G., Burstein D., Colless M., Davies R.L., McMahan R., & Wegner G., 1997a, MNRAS, in press (Paper III)

Saglia R.P., Bertschinger E., Baggley G., Burstein D., Colless M., Davies R.L., McMahan R., & Wegner G., 1997b, ApJS, in press (Paper IV, astro-ph/9609089)

Scodeggio M. et al. 1996, this book

Smith R. et al. 1996, this book

Wegner G., Colless M., Baggley G., Davies, R.L., Bertschinger E., Burstein D., McMahan R., & Saglia R.P., 1996, ApJS, 106, 1 (Paper I)

Wegner G., Davies, R.L., Baggley G., Saglia R.P., McMahan R., Colless M., Burstein D., & Bertschinger E., 1997 in preparation (Paper II)

Willick, J.A., 1994, ApJS, 92, 1