The APM Galaxy Survey: Some Data Reduction Techniques

S.J. Maddox, J. Loveday,
W.J. Sutherland and **G. Efstathiou**
Institute of Astronomy
Cambridge, England, UK

Abstract

We have used the Automatic Plate Measuring machine in Cambridge to scan 176 photographic survey plates from the UK Schmidt Telescope and have compiled a galaxy catalogue covering 4400 degrees^2 and containing over four million galaxies with $B_J < 21$. We give a brief description of the plate material, measuring procedure and data reduction techniques. The star–galaxy separation technique and tests of its reliability are described. CCD observations of groups of faint galaxies have been used to construct 66 photometric sequences of galaxies in the range $B_J = 17 - 21$.

1 Introduction

Wide-field surveys of galaxies and clusters are an indispensable tool for studying large scale structure in the universe. The Abell catalogue (Abell 1958), Zwicky catalogue (Zwicky *et al.* 1961-1968), and the Lick survey (Shane and Wirtanen 1967, Seldner *et al.* 1977) have provided many statistical results of key importance to our understanding of galaxy formation and clustering (see *e.g.* Peebles 1980). However, these surveys were constructed more than 20 years ago. Since then, there have been major technological developments in photographic emulsions, automatic scanning machines and computers. It is therefore possible to improve significantly on earlier surveys by generating deep galaxy catalogues with high photometric precision and uniformity over wide areas of sky. Over the last four years, we have taken advantage of these developments to construct a new survey of several million galaxies.

The photographic material that we have used is based on the $6° \times 6°$ SERC IIIa-J plates taken by the UK Schmidt Telescope Unit (UKSTU). The UKSTU plates cover the entire sky south of $\delta = -20°$. Our limiting magnitude corresponds to $B_J \approx 21$, more than 2 magnitudes deeper than the Lick survey. The European Southern Observatory (ESO) has supplied us with high quality glass copies of the original plates (West 1978) which we have scanned in Cambridge.

The SERC Automatic Plate Measuring (APM) machine in Cambridge is a high speed automated densitometer with on-line sky subtraction and image analysis (Kibblewhite *et al.* 1984). It can scan the central $5\overset{\circ}{.}8 \times 5\overset{\circ}{.}8$ of a UKSTU plate in about 14 hours giving accurate measurements of the position, magnitude and shape parameters for each image above a fixed detection threshold. We have used the APM machine to scan 176 of the 190 UKSTU fields within the area $b < -40°$ and $\delta < -20°$. The remaining

14 plates have yet to be supplied by ESO. Over the 4400 square degrees covered by our survey, we have detected about 3.6 million galaxies at a magnitude limit of $B_J = 21$. *Fig. 1* shows the field boundaries in an equal area projection centred on the south galactic pole. For statistical studies of the large-scale structure, it is important that the galaxy selection function is uniform over the whole survey. We have therefore used a combination of photometric CCD calibrations and comparisons between plate overlaps to ensure that the magnitude limit and star-galaxy classification are uniform. The positions of the calibration sequences are shown by the dots in *Fig. 1*.

2 The APM measurements

An outline of the technical details of the APM system is given by Kibblewhite *et al.* (1984). The microdensitometer uses a scanning laser to sample the transmission of the emulsion on a grid of positions with $\approx 8\,\mu$m separation. The image of a star or galaxy forms a group of pixels that are brighter than the local average sky and during the measurement the APM system locates all the images which contain more than a certain number of pixels brighter than a set threshold above sky.

To estimate the local sky brightness at each point, the APM carries out a preliminary scan of the plate. During the scan the pixel values are analysed in 64 × 64 pixel groups, corresponding to 0.5 mm squares on the plate. For each square a frequency histogram of the number of times each pixel value occurs is calculated. Then the sky brightness is estimated by fitting to the peak of the histogram. The sky values from all of the squares over the plate form a 640 × 640 pixel map of the background sky brightness. The map is filtered with a two dimensional non-linear filter and used to give an estimate of sky at the position of each pixel.

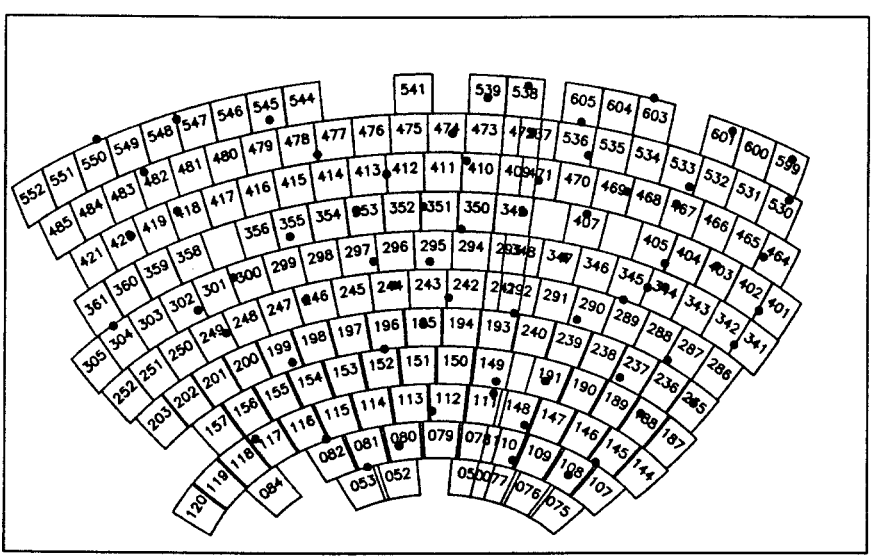

Fig. 1. Equal area projection of the APM survey fields. The UKSTU field numbers are indicated. The dots show the positions of our CCD sequences.

The pixel values input to the image processor after sky subtraction are truncated to 8 bits. For each detected image containing more than a certain minimum number of pixels above the threshold the image processor calculates parameters from the sky subtracted densities,

$$D = D_o - D_s \,,$$

where D_o is the density of an object pixel, and D_s is the local sky value.

The relative brightness of each image is measured by the summed density above sky for pixels above the threshold, t,

$$I = \sum_{D_o > D_t} (D_o - D_s) = \sum_{D>t} D \,, \tag{1}$$

where $D_t = D_s + t$. The position of each image is calculated as the density weighted mean of the X and Y of the pixels contained in it, so $X_0 = \langle XD \rangle / I$, and $Y_0 = \langle YD \rangle / I$. The shape of each image is parameterised by $\langle x^2 D \rangle / I$, $\langle xyD \rangle / I$, $\langle y^2 D \rangle / I$, where $x = X - X_0$, and $y = Y - Y_0$. These parameters can be used to calculate the radius of gyration, the eccentricity and orientation of an equivalent bivariate Gaussian image. The surface brightness profile of each image is given by the peak density above threshold, $D_p - D_t$, and the areal profile. The areal profile is the area of the image at 8 increasing surface brightness levels. The levels that are used are $D \geq D_t + 1, D_t + 2, D_t + 4$, $D_t + 8, D_t + 16, D_t + 32, D_t + 64, D_t + 128$. This combination of parameters gives all the useful information contained in the faint images and simultaneously gives a good representation of the brighter images.

The threshold set for each scan was twice the *rms* noise in the measured sky value. For the UKSTU J copies the noise is typically between 5 and 6 density units. The thresholds correspond to between 8% and 11% of the sky surface brightness, or $\approx$ 24.5 to 24 B_J magnitudes arcsec^{-2}. Measures of the sky subtracted density for pixel values lower than this threshold are dominated by the grain noise of the emulsion and so would not give any significant improvement in the accuracy of the measured parameters. Also, if the threshold is set much lower than this, many useless small noise images are recorded. A scan of a typical plate records about 300 000 images using this threshold with a minimum area of 16 pixels and minimum integrated isophotal intensity of 300.

The (X, Y) positions measured from the plate are converted into right ascension and declination, (α, δ), using a standard six parameter transform, with a radial correction factor for the projection of the sky onto the Schmidt plate (*e.g.* Murray 1983). The positions of the 20 brightest stars that are on the plate and in the Perth 70 astrometric catalogue (Høg *et al.* 1970) were measured before each scan was started, and a least squares fit used to estimate the constants in the transform. The *rms* residuals between the measured positions of the stars and this fit are typically about $0''.5$. The absolute accuracy of the plate coordinates is set by the standard star positions, and is about $1''$. We also find small distortions over the $5°.8 \times 5°.8$ field caused by trailing, differential field rotation, misalignment of the linear transform and emulsion deformations. These give rise to systematic errors of $\approx 1''.5 - 2''$. For studies of galaxy clustering these positional errors are insignificant. The galaxy positions are well within the tolerances

required for multi-object fibre spectroscopy at large telescopes. For example, Colless and Hewett (1987) have used the APM survey to generate fibre masks at the Anglo-Australian Telescope for their study of the dynamics of rich clusters.

Figure 2 shows an image surface density map in the same projection as *Fig. 1* for 20 368 362 images before making any corrections for uniformity. The magnitude limits vary between plates, reflecting differences in plate quality and detection threshold. These variations can be seen as changes in the number density of images across the field boundaries. However, even with no corrections the survey is fairly uniform and interesting features can be seen. For example, the fall in the stellar density away from the galactic plane and away from the galactic centre is clearly visible. Also the dwarf galaxies in Sculptor and Fornax, and a few globular clusters are partly resolved and show up as small areas of high image density. Many other intermediate scale features from galaxy clustering can be seen.

3 Measurements of the APM density *vs.* flux relation

The APM parameters are calculated using density measurements rather than flux measurements. Therefore the relation between the measured intensities and standard photometric magnitudes needs to be investigated. The APM initially measures the transmission for each pixel on a 12 bit scale using a photomultiplier to measure the intensity of the stabilised laser beam after it has gone through the plate. The transmission values are converted to a 10 bit density using a look-up table. The look-up table was chosen so that on an average copy plate the density range from the

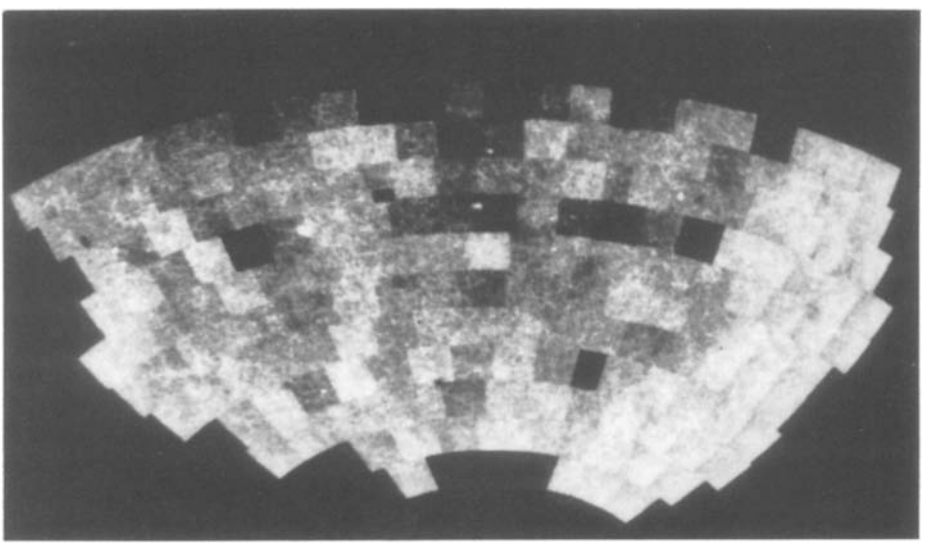

Fig. 2. Equal area projection showing the surface density of images in the unmatched APM survey scans at a magnitude limit of $M = 8$, corresponding to $B_J \approx 21.5$. There are 20.4×10^6 images in the $120° \times 60°$ area shown. The grey levels are set so that black = 0 image density and white = 7.3×10^3 images per square degree.

minimum sky value to the saturation level of the copies roughly fitted in the 10 bit density scale. After sky subtraction, only the most significant 8 bits are retained and so the effective range of the densities used in the calculation of the image parameters is 0 to 255. The optical density range measured on a copy plate is roughly 0 to 2.5. Therefore the APM density used to calculate the image parameters is related to the standard optical density, O,

$$D = \frac{256}{2.5} \log\left(\frac{4096}{T}\right) \approx 102.4\, O + const\,. \tag{2}$$

The density *vs.* flux relation can be measured using the step wedges that the UKSTU exposed on the edge of each plate during the main exposure. Unfortunately the emulsion is desensitised near the edges of the plates (Dawe and Metcalfe 1982), so the response of the emulsion for the central area of plate is different to that at the location of step wedges. The halo around the laser spot also alters the effective density *vs.* flux relation. The halo makes the density measurements at a point dependent on the surrounding density values. For a pixel in the wedge, the surrounding pixels have the same density, but for an image pixel most of the surroundings are close to the local sky. Therefore the density measured for a high flux pixel in an image is different to the density measured for the same flux in the step wedge.

A more direct way to measure the density *vs.* flux relation for images is to compare a raster scan from the APM with a CCD frame of the same area (Cawson *et al.* 1987). This gives a direct measurement of the flux corresponding to the density measured for each image pixel. These comparisons include the adjacency effects of the spot halo and any field dependent variations in the response curve. We have many CCD frames which were taken with the 1.0 m telescope at the South African Astronomical Observatory to provide photometric calibration. The area of plate covered by the CCD frames for field 344 was raster scanned with the APM. The size of pixels in the raster scan is the same as that used for the image mode scans, but is different to the CCD frames. The CCD frames need to be accurately rescaled and positioned to match with the APM scan so that the image pixels match up exactly. If there is a mismatch between APM pixels and CCD pixels, the calibration curve is spread out into a series of loops corresponding to the difference between the image cross sections in the two pixel sets. We used an image detection algorithm similar to that in the APM to locate and measure the image positions in the data sets. Then we paired up each image and did a least squares fit to give 6 constants in a linear transform between the two coordinate systems. Each B and V frame was transformed so that the image positions matched the images in the corresponding raster scan.

The seeing on the survey plate and the two frames are different, so the image profiles are not exactly the same, even when correctly matched up. We smoothed both the CCD frames and the APM scans with a Gaussian blurring function to reduce the relative difference. The smoothing also reduces the random noise in the curves.

If a B or V band CCD frame is used alone, each image gives a slightly different curve. This is because the images have different colours, and so need different colour correction terms to convert from the B or V to the J band of the UKSTU plates. The

frames were added using standard colour equations (Blair and Gilmore 1982, Walker 1984) to give a composite J frame which was compared pixel by pixel to the raster scan. *Fig. 3* shows a point for each pixel pair. The lowest pixel values are those for the sky, which is well above the fog level for these frames.

These measurements show that the APM density D below the emulsion saturation is fairly well approximated by

$$D = \alpha f + \beta \,, \tag{3}$$

where α, β are constants, and f is the incident flux. This approximation is plotted as the solid line in *Fig. 3*. The more conventional way of relating optical density O to incident flux is to assume

$$O = \gamma \log_{10}(f) + c \,, \tag{4}$$

but for lower densities this is a poor approximation to the measured curve. A better approximation uses the Baker density (Baker 1925), defined as

$$B = \log_{10}(10^{O} - 10^{O_{fog}}) \,, \tag{5}$$

so that

$$B = \gamma \log_{10}(f) + c \,. \tag{6}$$

This fits both the low and high density parts of the curves very well below the emulsion saturation. The linear approximation is just as good as Baker density before the copy

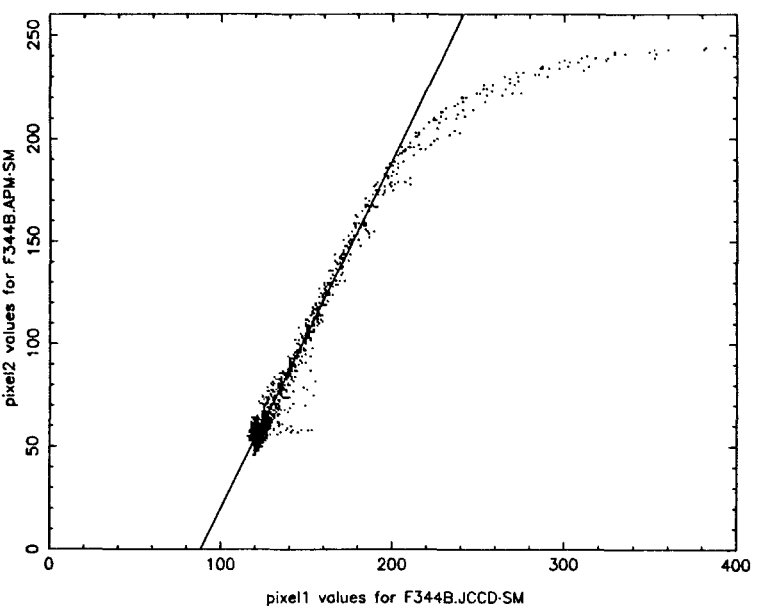

Fig. 3. Comparison of APM density measurements with CCD flux measurements in field 344. The density of each pixel from a $2' \times 3'$ raster scan of field 344 is plotted against the CCD flux measurement for the same piece of sky. The solid line is the $D = \alpha f + \beta$ approximation, with $\alpha = 1.7$ and $\beta = -150$.

emulsion saturates and is much simpler. Since the density is linearly related to the flux below saturation in this approximation, using densities throughout the image analysis is equivalent to using fluxes. Therefore the APM magnitude for galaxies is simply related to the total magnitude, $M = 2.5 \log_{10}(I) \approx z - B_J$, where z is the zero point for the plate. We have constructed 66 photometric sequences of faint galaxies from the CCD observations, and used them to measure zero points for the fields. They typically give $z \approx 29 - 30$.

Multiple scans of the same plate show that the measured magnitudes are repeatable to 5%, even for images close to the plate limit. However, comparisons between different plates of the same field show that variations in sensitivity and grain noise lead to large magnitude errors for images fainter than $B_J \approx 21$. Also, it becomes very difficult to distinguish galaxies from stars near the plate limit (see Sect. 6). For the final catalogue we have therefore aimed at uniformity for objects brighter than $B_J = 20.5$. At this limit we find about 17 000 galaxies per plate.

4 Plate overlaps and matching

The field centres are separated by 5° so there is a generous overlap area of about $6° \times 1°$ between each plate. The parameters of matched pairs of images in the overlaps provide our primary means of ensuring uniformity in the selection function over the survey area.

A good demonstration of our photometric accuracy and the extent to which we can remove systematic field effects is provided by plate pairs near $\alpha = 0$. Unlike the majority of plates in the survey, some of the plates at $\alpha = 0$ have large overlaps of up to $3° \times 6°$. Thus we can compare measurements made at one plate centre with those made at another plate edge. Any systematic errors in the magnitudes on each plate will show up as a positional variation in the mean difference between the two sets of measurements. *Fig. 4* shows a contour map of the average difference between the magnitudes of images measured from one plate at $\alpha = 0$ and its neighbour. The differences over most of the area are less than $0^m.04$ peak-to-peak.

We use an algorithm similar to that applied to the Lick counts by Seldner *et al.* (1977) to determine corrections to the magnitudes from the plate overlaps. A polynomial fit to the magnitude-magnitude plot for each overlap is used to give a conversion between the magnitudes from each plate to those from its neighbours. We then apply an iterative algorithm to find the set of field corrections which is the most consistent with the measured overlap conversions. Matching the magnitudes in this way gives a residual scatter of $0^m.02$ in the zero point of each plate.

The edge matching procedure is sensitive to residual field effects which introduce small systematic drifts in the plate zero points. Without any additional checks, these would cause large-scale gradients in the survey in the same way as that described for the Lick survey by Groth and Peebles (1986). To prevent this evidently undesirable effect, we use the CCD sequences as tie points in the plate matching algorithm.

5 Star - galaxy separation

The APM measures parameters for all images on each field, including stars, galaxies and dust, emulsion defects and other noise. Many techniques to distinguish stars from galaxies using parameterised shape data have been developed by several groups including Godwin *et al.* (1983), Jarvis and Tyson (1981), Sebok (1979) and MacGillivray *et al.* (1976). All of the techniques rely on the fact that the profile and shape of a stellar image is determined only by the instrumental response and seeing conditions during the observation, whereas a galaxy image has a different intrinsic profile. Therefore the image parameters of stars lie in a well defined region of the parameter space and can be excluded to leave a sample of galaxies.

The first attempt to define a galaxy sample from the APM survey used the standard APM automated classification program STATS. Pairs of the measured parameters are plotted against each other, and the stars fall along a well defined line. The program locates the peak position and measures the *rms* scatter about the stellar locus in four parameter plots. The plots are shown for a typical field in *Fig. 5*. The images which are more than 2σ away from the stellar locus are flagged as non-stellar. In the plot of peak density *vs.* total intensity an attempt is made to get a reasonable measure of the unsaturated value of the peak by extrapolating the areal profile to the centre of the image. Due to measuring noise and emulsion defects some images have parameters that are impossible for real images and these are excluded from the plots and flagged as noise images.

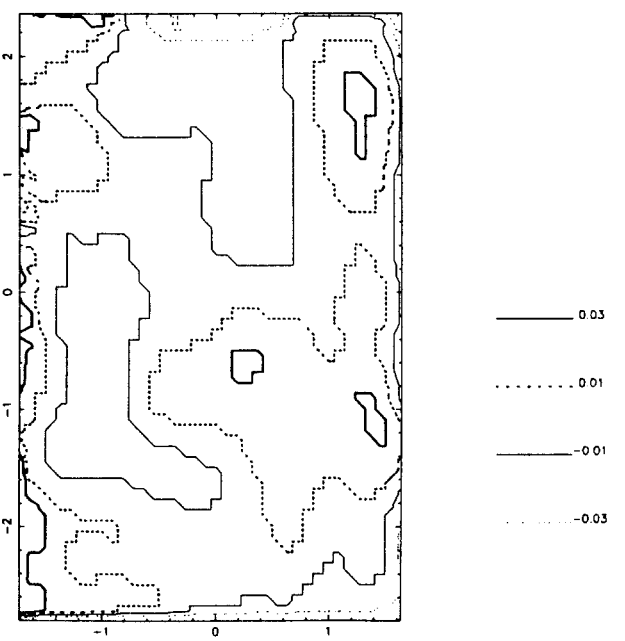

Fig. 4. Contour map illustrating magnitude differences between two plates with a large overlap (fields 292 and 241, see *Fig. 1*). The x and y scales are marked in degrees. The contours are spaced at $0^{\mathrm{m}}.02$ intervals according to the key shown.

The positions of each image on the four plots are combined by adding the distances of each image from the stellar lines in quadrature. Assuming that the differences from the stellar line are caused by Gaussian errors, this combination of distances is effectively a maximum likelihood estimate of whether the image is a star. Finally, each image is flagged as a star, eccentric star, galaxy, merged image or noise image, depending on the combined estimate and also the individual estimates. The eccentricity of each image is also calculated, and if an otherwise stellar image has a very different eccentricity than the average for stars, it is flagged as an eccentric star and grouped with the

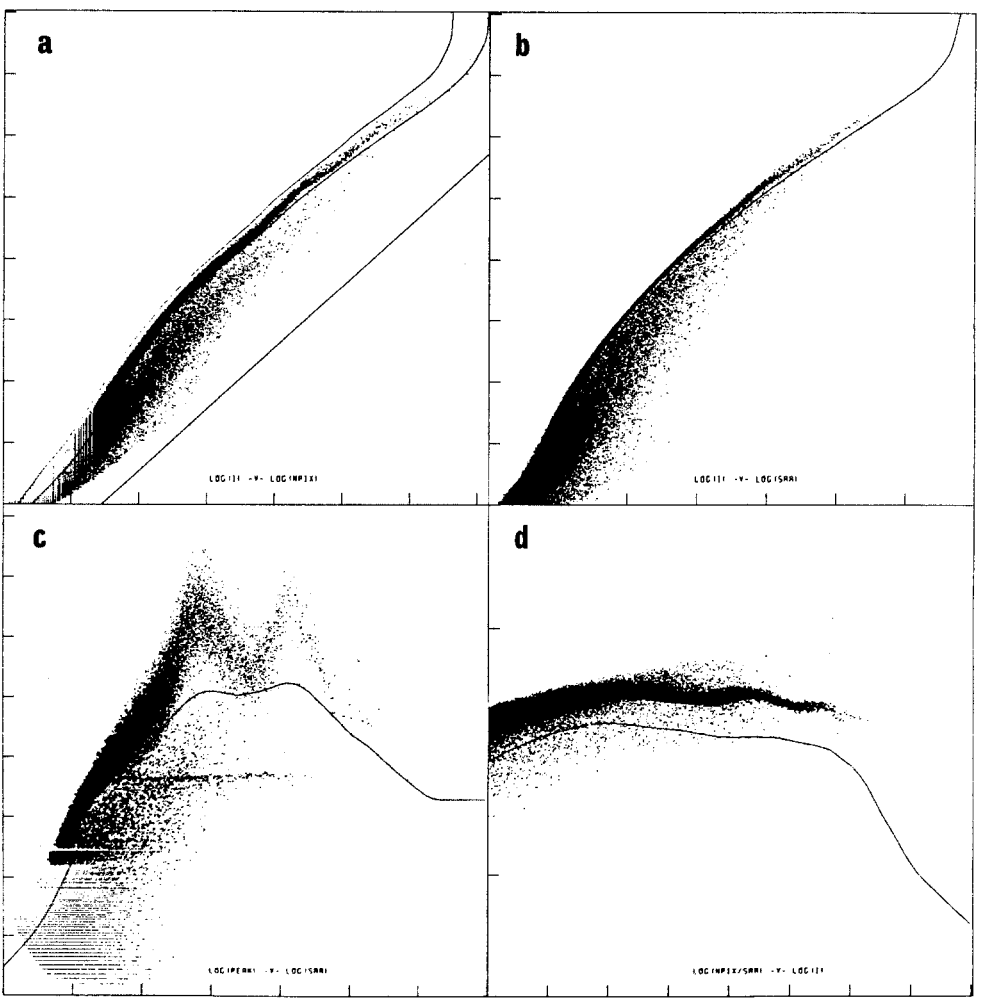

Fig. 5. Parameters used in separating stars from galaxies. (a) shows area *vs.* magnitude, (b) mean size *vs.* magnitude, (c) size *vs.* peak surface brightness, (d) magnitude *vs.* area/size. Points between the curved lines in (a) and above the curved lines in (b) and (c) are classified as stars. Points below the line in (d) are classified as merged images. The straight line in (a) shows the minimum possible area for a given magnitude.

merged images. The eccentricity of stars need not be zero because small errors in the tracking during the exposure can lead to all stellar images having a small eccentricity.

Using this classification algorithm, the main residual source of error in our matched galaxy maps is caused by variations in star-galaxy separation. Nevertheless some correlation analyses can be carried out as described in Maddox *et al.* (1988). Over the survey area the stellar density varies by a factor of ~ 3 from the galactic pole to low galactic latitudes. Most of the residual variation is caused by an increase in stellar contamination in the directions of the galactic centre and anticentre. Therefore we have developed a more sophisticated classification algorithm which uses all of the APM parameters.

6 Profile classification

The areal profile gives 8 measurements of the surface brightness profile for each image. Some typical stellar profiles are shown in *Fig. 6* together with some galaxy profiles. For a large range of magnitudes, the stellar profiles are very similar to each other, and the galaxy profiles are more varied with much lower surface brightnesses. Stars with $M < 10$ are faint enough to peak below the emulsion saturation and the profiles are linear in the $\log_{10}(D)$ *vs.* N_{pix} plot. Therefore their surface brightness profile is given by $f = Pe^{-(r^2/2\sigma^2)}$. The galaxy profiles are not very different to those of the stars for $M < 9$, so at faint magnitudes the classifications are not very reliable. The profiles of stars $13 > M > 10$ are affected by saturation, but the stellar profiles are still easily distinguishable from galaxies. Stars with $M > 13$ are bright enough that the halo affects the profile, and so the stellar profile becomes indistinguishable from the galaxy profiles.

A simple way of using all of the APM parameters is to treat each of the levels and the peak density as different measurements of the surface brightness profile, and use nine separate pairwise plots against magnitude. The radius of gyration is included as a tenth plot against magnitude. The distance of each image from the peak of the median stellar line is measured from each plot. Then for each image the residuals from the plots are summed in quadrature with a set of weighting factors, to give the final profile residual, ψ, which is equivalent to the integrated difference between the particular image profile and the stellar profile at the same magnitude,

$$\psi = \sum_{levels} \frac{(P(M) - P_{locus}(M))^2}{\sigma^2_{locus}(M)} \,. \tag{7}$$

Some fields show significant variation in the position of the stellar locus on the plots as a function of position over the plate. Therefore the position of the peak of the stellar locus for each paramater is calculated separately for each cell in a 8×8 array over the field. Then for every image the residual in each parameter is calculated from the peak position interpolated from the neighbouring cell centres.

The residuals are weighted by $\frac{1}{\sigma^2_{locus}}$ and so, assuming the errors are approximately Gaussian, the residual sum for stars has a χ^2 distribution. The σ_{locus} is calculated

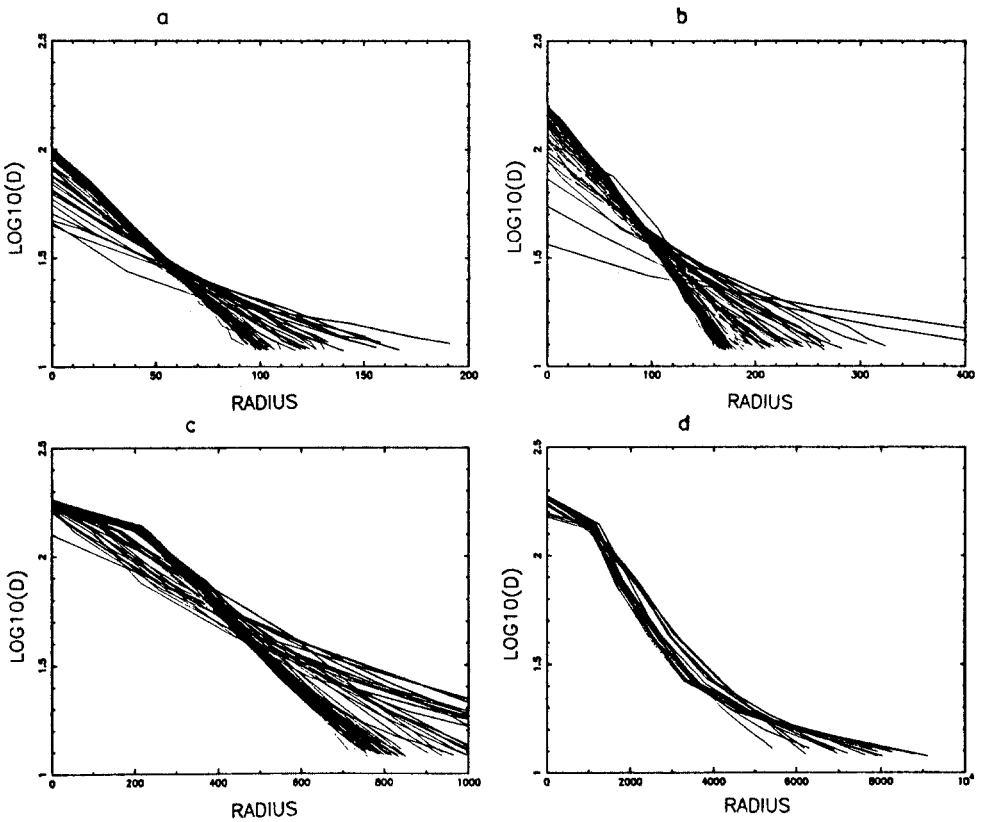

Fig. 6. Areal profiles of images at different magnitude slices. The density of each level is $D_n = D_t + 2^{n-1}$, and $\log(D_n)$ is plotted against N_n, the number of image pixels with $D > D_n$. For the fainter two slices, (a), $M = 9.0 - 9.1$, and (b), $M = 10.0 - 10.1$, the stellar profiles are all straight lines with the same slope. The galaxies are lower surface brightness, and more varied. For (c), $M = 12.0 - 12.1$, the stellar profiles are saturated near their peak, but are still easily distinguishable from galaxies. Brighter than $M = 14.0$, (d), the stars are saturated except for the halo, and so their profiles are indistinguishable from galaxies.

as a function of magnitude in each of the plots, so a fixed limit of ψ gives a sample of images which have a fixed minimum confidence level that they are significanctly non-stellar.

The σ_{locus} can be measured internally from the widths of the stellar locus in each of the plots, as in the STATS algorithm. Unfortunately, the variation in σ_{locus} for different plates makes the confidence levels different for each field and so it would be neccessary to adjust the classification boundary for each field to obtain a uniform sample of galaxies. In order to keep the scale of ψ the same for each field and also maintain uniformity as a function of magnitude, we assume that the σ_{locus} has the same magnitude dependence in each field. We have used the overlap between two fields to estimate the *rms* error in the measurements of each parameter and hence the σ_{locus} as a function of magnitude. The comparison of the measurements of each

of the profile areas from two neighbouring fields show the *rms* errors are fairly well approximated by $\sqrt{N_{pix}}$, equivalent to Poisson noise in the count of pixels above each density level.

The distribution of $\log \psi$ with these weightings, is shown for images from a typical field in *Fig. 7*. The stellar images fainter than $M = 12.5$ have a distribution which is well approximated by a constant χ^2 for all magnitudes. The galaxy images are well separated from the stars for $12.5 > M > 9.5$. For images fainter than this, the seeing and increased noise make it impossible to distinguish a stellar image from a galaxy, and the confidence level that can be given to any image being non-stellar is smaller. For images with $M < 9.0$, no reliable distinction can be made.

In order to measure the reliability of this classification algorithm we have inspected 6 areas on each of 16 plates in a test region and compared with the parameter classification. Each area is about $1 \times 1\,\mathrm{cm}^2$ and contains about 100 images brighter than $B_J \approx 21$, which we classified as stars, galaxies and merged objects by eye. We agreed with each other for about 95 % of the objects, and the other 5 % of the images were too faint to be sure about. Histograms of the number of images as a function of the $\log \psi$ classification parameter are plotted for the visually checked samples of stars and galaxies in *Fig. 8*. Each plot shows images in a different magnitude range.

For images in the faintest range the star and galaxy distributions have a large overlap.

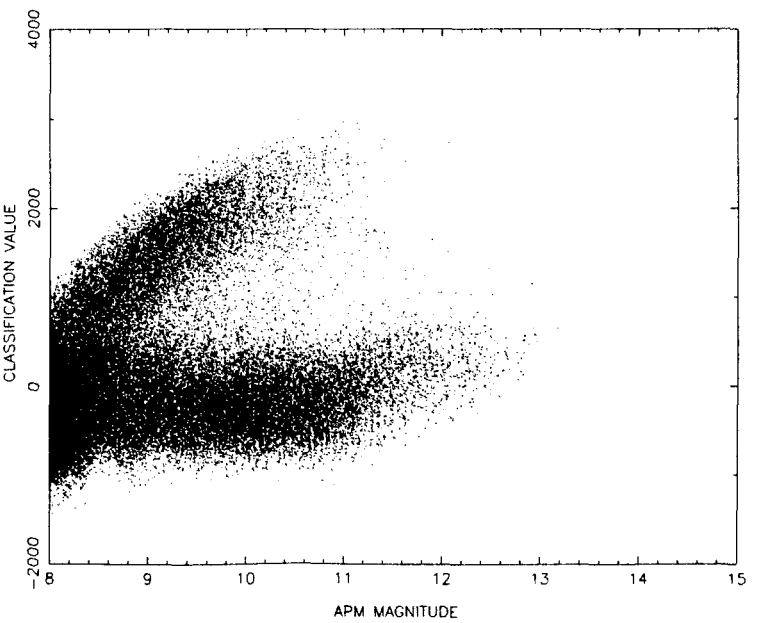

Fig. 7. The distribution of the ψ classification value for field 78. The stars fall along a well defined ridge with $\psi \approx 0$. The galaxies with $9 > M > 13$ fall in a well separated ridge with $\psi > 1000$. Near the plate limit, the atmospheric seeing blurs any image structure so the two ridges merge together and no distinction can be made for images fainter than $M \approx 8.5\,(B_J \approx 21)$.

These images are close to the plate limit, and their shapes are determined mainly by the the seeing profile. This means that it is very hard to tell the difference between stars and galaxies by eye, or from the measured parameters, and so it is not possible to select a complete and uncontaminated sample of galaxies. For the brighter ranges where the intrinsic galaxy profiles are visible, most stars fall in a narrow peak near zero as expected. The galaxies form a wider distribution extending to higher values because the intrinsic profiles are very different from the median stellar profile. Some high surface brightness galaxies have profiles which are very close to stellar, and so the galaxy distribution extends into the stellar distribution.

For images $M > 13$ the parametrisation of the shape does not contain all of the information available from the plate. The haloes around bright stars, and saturation of both stars and galaxies make the parameters for all objects very similar, and also increase the noise in the measurements. Therefore classifications based on the image parameters become unreliable and it is neccessary to visually check each bright image. We have visually classified galaxies brighter than $B_J = 16.3$ to give a sample of 11 963 galaxies with morphological types.

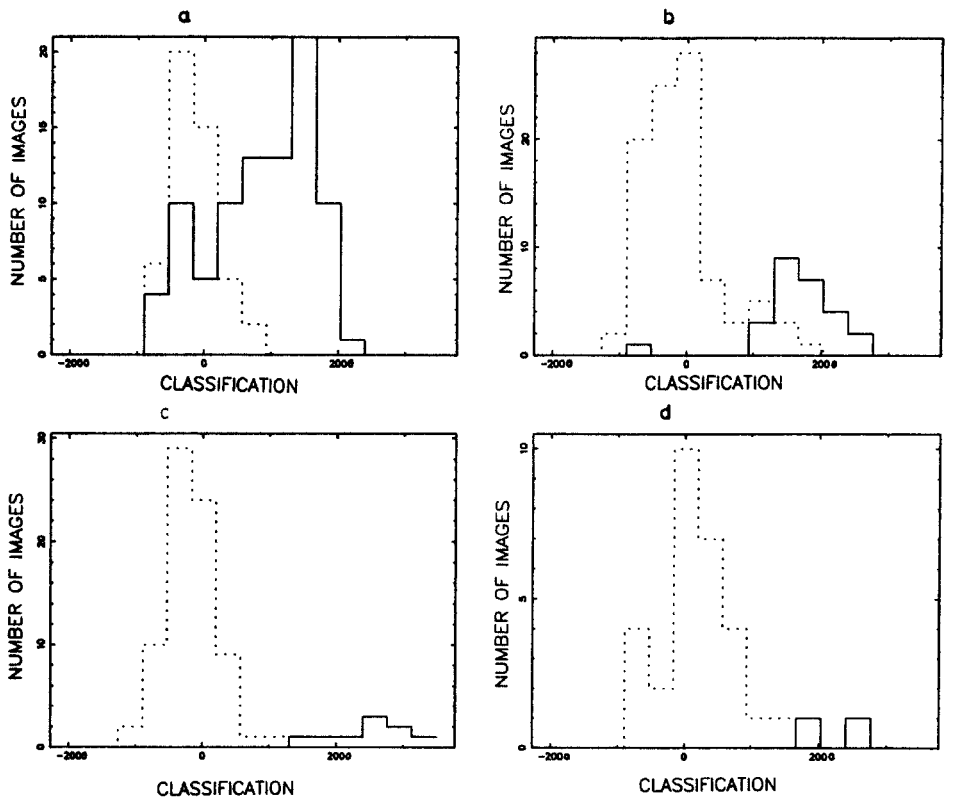

Fig. 8. Comparison of visual classification checks and ψ for 4 magnitude ranges as in *Fig. 6*. In each plot the dotted histogram shows the number of objects visually classified as stars as a function of ψ, and the solid histogram the number of objects visually classified as galaxies.

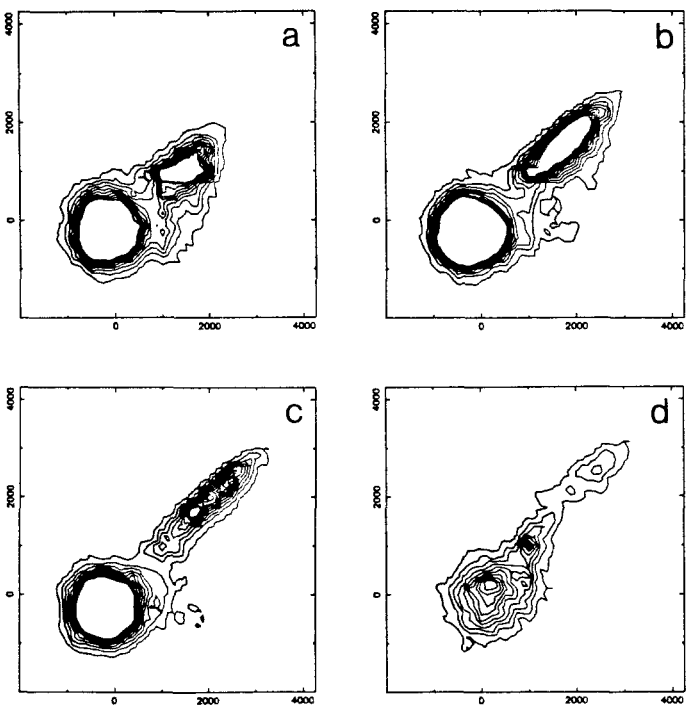

Fig. 9. Comparison of the ψ classification value parameter for images in the overlap between field 286 and field 287 for 4 magnitude ranges similar to those in *Fig. 6*. The axes of each plot are 1000 $\log_{10}(\psi_{286})$ and 1000 $\log_{10}(\psi_{287})$. The contours show the number of images in the cells of a 50×50 grid, smoothed by a Gaussian with $\sigma = 1$ cell. The lowest contour is for 4 images per cell and the interval between the levels is 2 images per cell. If the normalisation of the noise were exactly correct, the stars would be centred on (0,0) and the slope of the galaxy ridges would be 1.

The combined classification parameter ψ, measured from a field, can be compared with rescans of the same field and neighbouring fields to measure the repeatability of the classification. Comparison of the classifications of objects from repeat scans of the same plate shows that the ψ parameter is very repeatable. Similar comparisons of the classification parameter of objects in the overlaps show that the parameter is also repeatable for different plates and gives a reliable separation between stars and galaxies. A typical overlap is shown in *Fig. 9*, which is a series of contour maps of image number density in the ψ_1^2 *vs.* ψ_2^2 plane. The maps are for the same magnitude ranges as in *Fig. 8*. The stars form the large circular concentration of images at $(0, 0)$ and the galaxies fall in the elongation towards $(2000, 2000)$. Over the magnitude range $9 < M < 13$ there is a clear separation between stars and galaxies that is the same for each plate.

The uniformity of galaxy selection from these parameters is very important when measuring the clustering properties of the distribution. If the selection varies from plate to plate, the fraction of galaxies that are included, and the fraction of contami-

nating stars will be different for each field within the survey. Such variations introduce spurious clustering in a similar way to changes in the magnitude limit. Therefore in our final catalogue we will match the classification parameters between plates in the same way as the magnitudes. The ψ limit used to define the galaxy sample can then be adjusted for each field so that the completeness and stellar contamination are kept constant over the survey. Even without the matching, the selection function is quite accurately uniform over large areas. *Fig. 10* shows the galaxies in the whole survey area selected using a fixed cut of the ψ classifier. The confidence level chosen is fairly high so that the variations in plate quality are not very significant and the boundaries between most fields are not visible (*cf. Fig. 2*). With this cut in ψ the completeness of the galaxy sample is about 70% at a limit of $B_J = 20.5$.

7 Summary

A preliminary reduction of the APM galaxy survey including relative and absolute magnitude calibration, and star galaxy separation has been completed. A more sophisticated and reliable star galaxy separation technique has been developed and applied to the survey. The technique allows full matching of the galaxy selection function on different fields. Also, a visually classified sample of 11 963 galaxies with $B_J < 16.3$ has been completed. This sample provides reliable classifications for the brighter galaxies which are difficult to classify from the image parameters, and so complements the deeper sample. Detailed correlation and cluster analyses of the galaxy distribution will be described elsewhere.

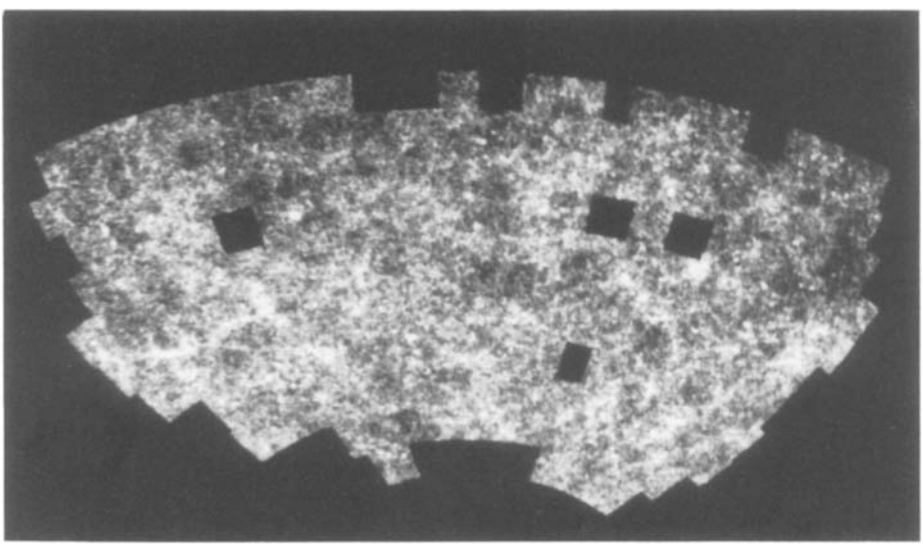

Fig. 10. Equal area projection of the surface density of 2.6×10^6 galaxies in the APM survey at a limit of $B_J = 20.5$ before matching the classification selection. A few field boundaries can be picked out, but most boundaries are not visible even though no classification adjustments have been made. The grey levels are set so that black = 0 image density, and white = 1.8×10^3 images per square degree.

Acknowledgements

We are indebted to the APM group, E.J. Kibblewhite, M.J. Irwin, P. Bunclark and M. Bridgeland, for their constant advice during this project. This work has been supported by the SERC and by the Nuffield Foundation. SJM, JL and WJS acknowledge the award of SERC research studentships.

References

Abell, G.O., 1958. *Astrophys. J. Suppl.*, **3**, 211.
Baker, A.E., 1925. *Proc. R. Soc. Edinburgh*, **45**, 166.
Blair, M., Gilmore, G., 1982. *Publ. astr. Soc. Pacific*, **94**, 742.
Godwin, J.G, Metcalfe, N., Peach, J.V., 1983. *Mon. Not. R. astr. Soc.*, **202**, 113.
Cawson, M.G.M., Kibblewhite, E.J., Disney, M.J., Phillipps, S., 1987. *Mon. Not. R. astr. Soc.*, **224**, 557.
Colless, M., Hewett, P., 1987. *Mon. Not. R. astr. Soc.*, **224**, 453.
Dawe, J.A., Metcalfe, N., 1982. *Proc. astr. Soc. Australia*, **4**, 466.
Groth, E.J., Peebles, P.J.E., 1986. *Astrophys. J.*, **310**, 507.
Høg, E., von der Heide, J., von Fischer-Trenenfeld, F., Holst, G., Loibl, B., Ziegler, U., Nikoloff, I., Helmer, L., Behr, A., 1970. *Abhandl. Sternw. Hamburg-Bergedorf*, Band 9.
Jarvis, J.F., Tyson, J.A., 1981. *Astr. J.*, **86**, 476.
Kibblewhite, E.J., Bridgeland, M.T., Bunclark, P., Irwin, M.J., 1984. In *Astronomical Microdensitometry Conference*, ed. Klinglesmith, D.A., NASA CP-2317, p. 277.
MacGillivray, H.T., Martin, R., Pratt, N.M., Reddish, V.C., Seddon, H., Alexander, L.W.G., Walker, G.S., Williams, P.R., 1976. *Mon. Not. R. astr. Soc.*, **176**, 265.
Maddox, S.J., Efstathiou, G., Loveday, J., 1988. In *IAU Symp. 130, The Structure of the Universe*, eds. Audouze, J., Szalay, A., Kluwer, Dordrecht (in press).
Murray, C.A., 1983. *Vectorial Astrometry*, Hilger, Bristol.
Peebles, P.J.E., 1980. *The Large-Scale Structure of the Universe*, Princeton Univ. Press, Princeton.
Sebok, W.L., 1979. *Astr. J.*, **84**, 1526.
Seldner, M., Siebers, B., Groth, E.J., Peebles, P.J.E., 1977. *Astr. J.*, **82**, 249.
Shane, C.D., Wirtanen, C.A., 1967. *Publ. Lick Obs.*, **22**, Part 1.
Walker, A.R., 1984. *Mon. Not. R. astr. Soc.*, **209**, 83.
West, R.M., 1978. In *Modern Techniques in Astronomical Photography*, eds. West, R.M., Heudier, J.L., ESO Publ., Geneva, p. 193.
Zwicky, F., Herzog, E., Wild, P., Karpowicz, M., Kowal, C.T., 1961-1968. *Catalogue of Galaxies and Clusters of Galaxies*, **I - VI**, California Institute of Technology, Pasadena.