The Edinburgh/Durham Southern Galaxy Catalogue: An Investigation into the Large-Scale Structure of the Universe.

Neil H. Heydon-Dumbleton

Presented for the Degree of Doctor of Philosophy
at the University of Edinburgh.

January, 1989.

This thesis is my own composition, except where specifically indicated in the text.

January, 1989.

Dedicated to my parents,
with love and admiration.

Contents

List of Tables

List of Figures

Abstract

This thesis describes the construction and application of the Edinburgh/Durham Southern Galaxy Catalogue, which is a database of information on $\sim 1{\cdot}5$ million galaxies, covering ~ 2000 deg^2 of the South Galactic Cap. This catalogue is based on objective image detection and classification techniques, rather than the visual searches of photographic plates used in previous galaxy catalogues. The raw data for this project are digitised scans of 60 *ESO/SERC Atlas* plates using the COSMOS high-speed plate measuring machine. The quality controls employed during the production of the *ESO/SERC Atlas*, ensures that it is deeper and more uniform than set of plates used previously to construct a galaxy catalogue. The COSMOS machine objectively detects and parameterises $\sim 2 \times 10^5$ images on each photographic plate. Image deblending software has been introduced to ensure the accurate detection and parameterisation of images in the crowded regions of compact clusters.

Star-galaxy classification and photometric calibration techniques have been investigated and optimised to reduce and quantify any systematic variations that could introduce spurious structure. A classification algorithm has been used to automatically classify images over the whole range of magnitudes in the survey. Accurate intra-plate photometry is possible for galaxies, because a COSMOS magnitude can be defined which is linearly related to the object magnitude. Inter-plate calibration is carried out using CCD galaxy sequences for every second field in the catalogue. Unlike global calibration techniques used previously, this arrangement of CCD's prevents propagation of calibration errors. Statistics are given to show that the final catalogue of galaxies will be $\gtrsim 95\%$ complete for $b_j \leq 20{\cdot}0$ with $\lesssim 10\%$ contamination by stars and that the point-to-point variation in galaxy number density, due to the combined residual systematic errors in classification and calibration, is $\sim 8\%$.

To date a mosaic of 35 plates covering a contiguous region of 1000 deg^2 has been constructed. The large-scale galaxy distribution, seen in maps of this data, is dominated by two large supercluster complexes separated by $\sim 15° - 20°$. Several filamentary arc structures can also be seen, with clusters distributed along them. The number–magnitude counts derived from this database show significant deviation from a no-evolution model at $b_j > 18{\cdot}75$. The variation in the amplitude of the counts across the survey cannot be accounted for by systematic variations in limiting magnitude and so is probably due large-scale clustering of galaxies. The two-point correlation functions calculated for this 35-plate mosaic confirm a break from power-law, though at larger scales ($\sim 20h^{-1}$ Mpc) than previously estimated. In the context of current theories of galaxy formation, models involving standard cold dark matter with extra large-scale power would still seem to be excluded.

Acknowledgements

First I would like to thank Chris Collins for all his advice and encouragement over the past three years and for boldly assuming the role of supervisor. I wish him every success with the Catalogue in the future. Thanks are also due to Harvey MacGillivray for his work with COSMOS, to Lance Miller for numerous useful discussions and to Luigi Guzzo for his enthusiasm about this work.

Many people deserve a mention in these acknowledgements and I am bound to forget someone, so apologies to anyone who feels left out. Thanks go to the *Lads* at the Observatory, especially Paul and Stuart (for all things TEX-nical), to Patrick and Fiona (for making that awful cottage bearable), to Alison and Stewart (for the party in Montpelier Pk.), to Campbell (for great money making schemes) and to Susan (for putting up with our conversations about computing), to Mandy and Caroline (for letters from down-south), to Julie (for the help in getting a proper job), to Tim (for the many pints of Guiness), to Ian (for being a *Bloody-Arab*), to Boyd (for dealings with the DHSS), to Ted (for 80p), to all the helpers and members at the Scottish Society for the Mentally Handicapped (for Thursday Nights), and to Adrian who has been a remarkably good friend over the past few years.

I could not possibly have finished this thesis without the love and support of my family. Many thanks go to Mum, Dad, Nan, Grandad and to my sister, Claire, and her husband, David. Last, but not least, I thank Helen who is very special.

The SERC is acknowledged for financial support to carry out this project.

Chapter 1

The Large-Scale Distribution of Galaxies

1.1 Introductory Remarks

The Edinburgh/Durham Southern Galaxy Catalogue is a database of galaxies, and their properties, distributed across $\sim$ 2000 deg^2 of the South Galactic Cap. When completed this database will contain information on a complete sample of $\sim$ 500,000 galaxies, though information will be available on $\sim$ 1·5 million galaxies in total.

Previous galaxy catalogues and observations of their large-scale distribution in the Universe, are based on visual searches of photographic plates by several observers over a substantial period of time. Much work has been carried out since the 1950's to use these observations to constrain theories of galaxy formation. Recently however, doubts have been expressed about the validity of structures found in such *eyeballed* catalogues, as systematic variations in counting efficiency have been detected from plate to plate and over the duration of the survey's construction. Due to the subjective nature of the construction of theses catalogues, it is difficult to quantify the effects that these variations will have. Some of the constraints applied to theories of galaxy formation may not be entirely valid.

The Edinburgh/Durham Southern Galaxy Catalogue (hereafter EDSGC) differs from previous catalogues, because it is based on the objective detection of images from digitised scans of photographic plates. It is possible, with careful experimentation and data reduction techniques, to reduce systematic variations in image detection, classification and photometry and then to quantify any residual errors. This, coupled with the size of the catalogue, makes the EDSGC an ideal basis to investigate the

large-scale distribution of galaxies and will allow limits to be placed on the errors associated with the statistics of galaxy clustering. It should be possible, therefore, to confirm or refute some constraints that have previously been placed on theories of galaxy formation.

The aim of this thesis is to detail the work done to investigate the problems associated with the production of a catalogue of this kind, to present the solutions adopted and to estimate the level of any residual systematic variations. Work has also been carried out to construct a large part of the EDSGC, from which some preliminary results have been obtained and studies performed to test the feasibility of future applications.

The layout of this thesis is as follows. Chapter 1. discusses the context in which this work is carried out. The first section describes the construction of previous galaxy catalogues and some of the features observed in the distribution of galaxies. The second section is concerned with some theories of galaxy formation and the distribution of galaxies that is predicted by numerical modelling. A critical comparison between observation and theory is made in the third section and some limits are placed on the homogeneity required for the EDSGC.

The uniformity of the raw data, *ie.* the photographic plates and the digitised data, is vital for a project such as the EDSGC. The production and quality control of United Kingdom Schmidt Telescope plates is detailed in Chapter 2., together with a description of the COSMOS machine and the method by which digitised data is turned into lists of images and image parameters. Techniques are also described to overcome the incompleteness of COSMOS datasets previously found in crowded regions.

The major sources of inhomogeneity in galaxy catalogues are errors in image classification, as stars or galaxies, and in image photometry. The algorithms adopted, to perform accurate image classification in the EDSGC, are presented in Chapter 3., along with estimates of the residual variation in contamination (with stars and noise objects) and completeness of the resultant galaxy catalogue. Chapter 4. describes the methods used

to ensure consistent image photometry for individual galaxies and a uniform limiting magnitude across the whole survey region.

The final chapter (Chapter 5.) of this thesis details the construction of a contiguous mosaic of plates that forms the EDSGC database. To date a 35-plate mosaic has been constructed covering a region of ~ 1000 deg^2 centred on the South Galactic Pole. The distribution of galaxies in this mosaic is described and various maps of the large-scale structure are presented, both here and in the Appendix. Differential number–magnitude counts and two-point angular correlation functions are derived and some implications for theoretical models are drawn. The penultimate section in this chapter discusses some future applications of the EDSGC and the last section provides a brief summary of this and previous chapters.

This work has been carried out in collaboration with Dr C. Collins (University of Edinburgh/Royal Observatory Edinburgh) and with Dr H. MacGillivray (Royal Observatory Edinburgh). The early work on this project was carried out with useful contributions from Dr T. Shanks (University of Durham). The SERC grant awarded for this work is held jointly by the Universities of Edinburgh and Durham. The observations at Cerro Tololo Inter-American Observatory (Chapter 4) were carried out together with Chris Collins and Tom Shanks. Those at the South African Astronomical Observatory (Chapter 4) were carried out by Chris Collins. Observations at the European Southern Observatory, that will form the basis of future application of the EDSGC (Chapter 5.), were carried out in collaboration with Chris Collins and Luigi Guzzo (ESO/Osservatorio di Brera).

1.2 The Observed Distribution of Galaxies

1.2.1 THE HISTORY OF GALAXY CATALOGUES

The first recorded catalogue of non-stellar objects was prepared by Charles Messier in 1781. This catalogue contains 103 objects with a nebulous appearance that Messier feared could be mistaken as comets. In 1781 William Herschel also began his systematic study and classification of the *nebulae*. By 1802 he had catalogued some 2500 nebulae. His son John Herschel continued his father's work, publishing the *General Catalogue of Nebulae* in 1864. This contained 5079 objects classified in 32 types, representing supposed evolution of the nebulae. By the beginning of the twentieth century around 15,000 nebulae had been catalogued, mainly by J.L.E. Dreyer in his *New General Catalogue* (1888) and its two index catalogues (1895 & 1908).

The extra-galactic nature of these nebulae was first postulated by Immanuel Kant in 1755. He considered that by analogy with our solar system, the Milky Way was a disk of stars and that the nebulae were similar systems. He wrote in *Universe Natural History and Theory of the Heavens* that:

> The form of the starry heavens has thus no other explanation than just such a systematic constitution on the whole that the planetary system has in miniature, in which all stars make up a system whose general plane relation is the Milky Way.
>
> By analogy with the system of stars in which we find ourselves...is in perfect agreement with the concept that these elliptical objects are just [island] universes — in other words, Milky Ways...

William Parsons was the first astronomer to record the spiral shape of many nebulae during the 1880s. The nature and distribution of these spiral nebulae was at the heart of the famous Shapley-Curtis debate, which took place at the National Academy of Science in Washington, America, on April 26th 1920. Harlow Shapley proposed that the spiral nebulae were scattered throughout our own galaxy, while Heber Curtis

supported the idea that these nebulae were rotating systems of stars similar to the Milky Way. Though both sides had observational data supporting their point of view, the core of the problem was the lack of an accurate distance indicator. The debate was effectively resolved on Oct 6th 1923 when Edwin Hubble took the photograph of Andromeda from which he was to discover a cepheid variable within the Great Nebula.

During the 1930s the first systematic surveys of galaxies and clusters were carried out. The Shapely-Ames' survey of bright galaxies was published in 1934, as was Hubble's survey of faint galaxies. Shapely had catalogued 25 clusters by 1933, while in 1938 Zwicky proposed a model for all galaxies being distributed in clusters. As pointed out by Oort (1983) even larger structures than clusters where also known at that time. The excess of galaxies due to the local supercluster is visible in both the Messier catalogue and the *General Catalogue* of the Herschels. The flattened nature of the local supercluster is very prominent in the Shapley-Ames catalogue.

The largest catalogues of galaxies and clusters to date were compiled during the 1950's and 1960's. The *Lick Catalogue* of galaxies and Abell's catalogue of clusters have formed the basis of the investigation of the large-scale structure of the Universe and the statistical properties of galaxy clustering. Though these catalogues were produced by visual scans or counts of objects on photographic plates, the problems encountered and methods used to ensure homogeneity provide a useful introduction to a discussion of the production of Edinburgh/Durham Southern Galaxy Catalogue.

1.2.2 THE LICK CATALOGUE

From 1947 to 1954 C.D. Shane and C.A. Wirtanen, of the Lick Observatory, counted 801,000 separate galaxies visible on 1246 photographic plates taken with the Carnegie 20 inch Astrograph. With each plate subtending $6° \times 6°$ and $1°$ overlap between plates, the survey covered $\sim \frac{2}{3}^{\text{rd}}$ of the sky, ranging from a declination $-20°$ to the North Galactic Pole. The limiting magnitude of the *Lick Catalogue* is estimated at $m_{pg} \simeq 18{\cdot}6$ (Groth

& Peebles 1977). The plates were originally taken as part of the Lick Observatory's proper motion survey and so each image has a companion image due to the objective grating used. Shane and Wirtanen claim these extra images did not affect the galaxy counts.

Galaxies were counted on each plate in a grid of 1296 $10' \times 10'$ cells. Each plate was counted in a consistent manner using a travelling microscope moving from east to west along each strip of cells. The definition of which objects constituted galaxies was somewhat subjective, as explained in the final published version of the catalogue (1967):

> To make the counts meaningful it is necessary to set some criteria for identifying galaxies. The most practical criteria...seems to be the observers' confidence that the images represent galaxies...it was agreed that any image judged with 90% probability to be a galaxy should be counted.

Some comparison of the galaxy classification was made with deeper exposure plates. One image in 70 cells was found to be a contaminant, that is a non-galaxy image, but no statistics were given on the survey completeness, or the variation of contamination with galactic latitude.

There were many corrections applied to the raw data to reduce the counts to a uniform system and counter variations in counting efficiency and survey limiting magnitude, that took place during the construction of the catalogue. Variations in the survey limiting magnitude (m_{lim}) will introduce spurious number density gradients because of the number–magnitude relation:

$$\frac{\Delta N}{N} = \ln(10).\gamma.\Delta m_{lim} \tag{1.1}$$

where N = number density of galaxies to m_{lim}. Assuming a uniform distribution of galaxies and flat space, the slope, γ, is 0·6.

The aim of the correction procedure was to ensure that the same number of galaxies would be counted regardless of plate sensitivity, epoch, or observer. Shane and Wirta-

nen's approach was to derive a correction factor for each variation in efficiency that they could identify. The details of these correction factors are listed in Table 1.1. As part of the correction procedure the raw data counted in $10' \times 10'$ cells were combined into $1^\circ \times 1^\circ$ cells.

Shane and Wirtanen estimated their counting error using cells in plate overlaps. They conclude that the experimental standard deviation of a count N per square degree is $1{\cdot}13N^{\frac{1}{2}}$ ($N^{\frac{1}{2}}$ is expected from counting statistics). They also re-photographed, counted and corrected 27 fields and compared the results with the original data. They comment that:

> Comparison [with the original data] appears to be reasonably satisfactory except for the area covered by the plate at $14^h\ 40^m$, $+15^\circ$. Here the number of galaxies on the repeat plate is greater that on the original in the ratio 1·6 to 1...the only reasonable conclusion seems to be that since the determination of nearly all the correction factors is subject to errors of one kind or another, several conspired to produce the large discrepancy in the results from these two plates.

The raw data from the *Lick Catalogue* were re-analysed in 1977 by Seldner, Siebers, Groth and Peebles (hereafter SSGP). Rather than apply a set of correction factors they used the $10' \times 10'$ cells in the overlaps of adjacent plates to obtain a single plate correction factor for each plate. Similar correction factors to those of Shane and Wirtanen had previously been applied to correct for errors within plates (*e.g.* vignetting). The SSGP plate correction factors were obtained using a complicated global fitting technique to compare each plate with all the adjacent plates and so "make the counts in overlapping regions as consistent as possible" (SSGP 1977). The SSGP correction factors are therefore equivalent to the product of the date, personal, exposure, emulsion and smoothing factors (Table 1.1) used by Shane and Wirtanen. Further correction factors were applied by SSGP to correct for atmospheric extinction, a north–south discrepancy and to normalise the total number of galaxies back to the original total number counted. The SSGP correction factors are listed in Table 1.2 and the maps of

Shane & Wirtanen Correction Factors

Corrections applied to cells according to their position on the plate.

Field Corrections. The average number of galaxies per cell ($1° \times 1°$) is forced to be the same for all cells. One correction for vignetting, one for an east–west bias.

Atmospheric Extinction. Plates were taken symmetrically about the meridian, so a correction was applied to each cell calculated from the declination of the plate centre using $\Delta \log N = 0{\cdot}11[\sec(37{\cdot}3 - \delta) - 1]$.

Corrections applied to each plate as a whole.

Date Factor. Applied to correct for the varying counting standard of each observer over the counting period (1947 – 1954). Correction factors obtained by randomly re-counting plates.

Personal Factor. Correction applied for the different counting standard of each observer. Corrections obtained from the overlaps between plates.

Exposure Factor. Corrections applied to plates with less than 2^{hrs} exposure due to bad weather, using $\Delta \log N = 0{\cdot}99 \log(\frac{t_{exp}}{120})$

Batch Factor. Correction depending on the sensitivity of each batch of plates using the plate overlaps (weighted by $N^{\frac{1}{2}}$).

Smoothing Factor. Correction for unseen errors obtained by comparing the number ($N^{\frac{1}{2}}$) of galaxies on overlapping plates after application of the above corrections.

Corrections applied to large regions of the survey

Normalising Factor. Correction applied to the data as a whole to correct the total number of images back to the original number counted. The value (0.768) is the reciprocal of the product of average values of each of the correction factors listed above.

Table 1.1: The correction factors applied by Shane and Wirtanen to ensure homogeneity over the *Lick Catalogue*.

Seldner, Siebers Groth & Peebles Correction Factors

Corrections applied to cells according to their position on the plate.

Field Corrections. The average number of galaxies per cell ($10' \times 10'$) is forced to be the same for all cells. The correction factor (f_i) for the i^{th} cell is obtained from: $\langle n_i \rangle$ = the mean count at the i^{th} cell, $\langle n \rangle$ the average count per cell, with $f_i = \langle n \rangle / \langle n_i \rangle$

Corrections applied to each plate as a whole.

Atmospheric Factor. Correction applied according to the declination (δ_c) of the plate centre from $\exp\{0.454[sec(\delta_c - 37{\cdot}3) - 1]\}$.

Plate Factor. Corrections applied to make the counts in overlapping regions of adjacent plates as consistent as possible. Equivalent to the, Date, Personal, Exposure, Batch and Smoothing Factors of Shane and Wirtanen. Correction factors C_p are calculated by minimising the quantity:

$$F = \frac{1}{2} \sum_p \sum_{q \subset p} (n_{pq} + n_{qp}) (B_{pq} + \theta_p - \theta_q)^2 ,$$

using: n_{pq} = the sum of counts in cells on p^{th} plate which overlaps the q^{th} plate, A = the number of overlapping cells, $B_{pq} = \ln(n_{pq}/A_{pq}) - \ln(n_{qp}/A_{qp})$, $\theta_p = \ln C_p$, and $q \subset p$ indicating plates q overlapping p.

Corrections applied to large regions of the survey

N–S Factor. The Plate Factors systematically increase counts in the northern hemisphere by 10% relative to the southern hemisphere. Plate factors in the south are multiplied by 1·10.

Normalising Factor. Correction applied to the data as a whole to correct the total number of images back to the original number counted. The value 0.85504 is applied after the application of the above corrections.

Table 1.2: Correction factors applied during the SSGP reduction of the *Lick Catalogue.*

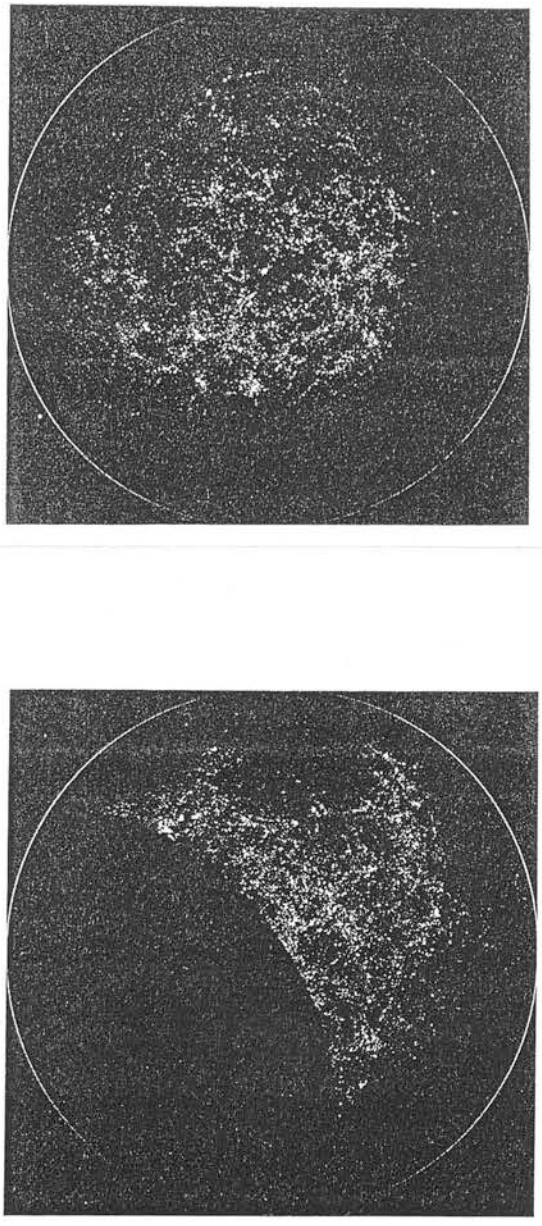


Figure 1.1: Greyscale maps of the galaxy distribution produced in the Seldner *et al.* (1977) reduction of the *Lick Catalogue*.(a) Northern hemisphere. (b) Southern Hemisphere. Reproduced from Seitter (1988).

the galaxy distribution that they obtained are shown in Fig. 1.1.

The distribution of galaxies as seen in the *Lick Catalogue* is described by Shane and Wirtanen (1967):

> The distribution of galaxies in space is characterized by a hierarchy of structural features ranging upward from double to multiple galaxies, through groups, clusters and clouds. The latter may be several tens of mega-parsecs in diameter. The possibility of even larger irregularities measured in hundreds of mega-parsecs is suggested. These considerations raise the question as to whether there is any observable upper limit to the size of structural features in galaxian distribution. If there is no such limit, surveys of the distribution in depth will reflect these features rather than more general properties of the Universe.

1.2.3 THE ABELL CLUSTER CATALOGUE

George Abell's *Catalogue of Rich Clusters of Galaxies* (1958, hereafter the Abell Catalogue) was prepared by visual inspection of Palomar Observatory Sky Survey (POSS) plates and contains 2712 clusters covering $\frac{2}{3}^{\text{rd}}$ of the sky above galactic latitude $-27°$.

Each of the 879 plates used was initially inspected for interesting objects and if accepted for the survey later "carefully inspected" for clusters. The centroid of each "collection of galaxy images was determined solely by judgement" and the magnitudes were obtained by comparison with galaxies of known magnitude on film copies. 2712 candidate clusters were identified. From this candidate list the *Statistical Sample* was compiled. Candidates were only included in the sample if they satisfied each of 4 criterion based on cluster richness, compactness, estimated distance and galactic latitude. The motivation for the choice of these criteria, given in Table 1.3, is described by Abell:

> There is a general field of galaxies which varies from point to point in the sky...superimposed upon that general field there are occasionally very rich clusters of galaxies which stand out conspicuously and which we shall assume to be physical associations. There will generally be a few galaxies

> belonging to the general field...However, their number will be relatively small if we consider only the very richest aggregates...criteria have been set up to exclude those associations which have a non-negligible chance of being optical only or which are insufficiently rich to insure identification.

Several possible sources of error within the survey were identified, using duplicate plates and by comparison with the work of others (Humason *et al,* 1956). The standard deviation in image magnitudes due to varying observing conditions and image quality was estimated at $0^{m}.19$. The standard error on the redshift (from the $\log(cz)$ verses m_{10} relation) is 26% and the standard error in counting was estimated at 17%. The Abell classification scheme, using richness and distance classes (Table 1.4), has the width of each class set to be 3·5 times the respective standard error in counting or magnitude.

Using the statistical sample Abell examined the distribution of clusters on the sky and as a function of redshift. He concluded that "the data allow no significant conclusion that the spatial density of cluster centres varies with distance." Across the sky however he noted a "highly significant non random surface distribution" that "suggest the existence of second order clustering or clustering of clusters."

Recently the Abell Catalogue has been repeated in the southern sky (Abell *et al.* 1988). The method of production of the southern catalogue is similar to that of the original catalogue, except that SERC (J) survey plates (Section 2.2.1) were used rather than POSS red plates and the $\log(cz)$ verses m_{10} relation has been refined. Some details of this new survey are discussed in detail in Section 1.4.

1.2.4 SUMMARY OF OTHER WORK

The *Catalogue of Galaxies and Cluster of Galaxies* (CGCG) produced in six volumes by Zwicky *et al.* (1961–1968) contains a list of galaxies complete to $m_{pg} = 15{\cdot}5$ and a list of clusters complete to the limit of the POSS. This work, often referred to as the Zwicky Catalogue, has a brighter limiting magnitude than the *Lick Catalogue*, but

Abell Selection Criteria

Richness Criterion. A cluster must contain at least 50 members within $2^{m}.0$ fainter than the third brightest member.

Compactness Criteria. A cluster must contain 50 or more members within a given angular radius of its centre. The angular radius calculated for each cluster corresponds to a fixed distance in space ($r_{abell} = 1.5h^{-1}$ Mpc). The cluster redshift (z) is estimated from the magnitude (m_{10}) of the 10^{th} brightest member and is converted to a distance on the plate using, $r_{plate} = 4{\cdot}6 \times 10^5/cz$ mm (see Abell 1958 Sect.Ie for $\log cz$ *vs* m_{10} relation).

Distance Criterion. The lower distance limit ($cz \sim 6600$ km/sec) is set to ensure clusters do not cover more than one plate. The upper limit ($cz \sim 60000$ km/sec) is set by the requirement that the third brightest member is $2^{m}.0$ above the plate magnitude limit.

Galactic-Latitude Criterion. Areas of sky near the Milky Way are excluded because of the high density of stellar images. The latitudes excluded vary as a function of longitude.

Table 1.3: The criteria for inclusion in the Abell *Statistical Sample* of galaxy clusters.

Richness Group Intervals

Richness Group	Count	Richness Group	Count
0	30–49	3	130–199
1	50–79	4	200–299
2	80–129	5	300+

Distance Group Intervals

Distance Group	Magnitude Range	Distance Group	Magnitude range
1	13·3–14·0	4	15·7–16·4
2	14·1–14·8	5	16·5–17·2
3	14·9–15·6	6	17·3–18·0
		7	Over 18.0

Table 1.4: Abell cluster classifications: (a) Richness classification in terms of the number of galaxies between m_3 (third brightest) and $m_3 + 2^{\mathrm{m}}.0$. (b) Distance classification in terms of the brightness of the tenth brightest member (m_{10}).

lists the positions, magnitudes, morphological types and redshifts, where available, for individual galaxies. The criteria used to define clusters (Table 1.5) are less strict than those used by Abell and the resultant catalogue contains systems that are less rich.

Other visual searches for galaxies include the *Jagellonian Catalogue* (Rudnicki *et al* 1972), a deep survey in Zwicky field number 185, the *Uppsala General Catalogue of Galaxies* (UGC) (Nilson 1973), the *ESO/Uppsala Survey of the ESO (B) Atlas* (Lauberts 1982) and the *Southern Galaxy Catalogue* (SGC) (Corwin *et al.* 1985). The *Uppsala* surveys are diameter limited surveys down to 1′, while the SGC consists of 5481 bright galaxies visually inspected on the SERC J Survey. The *Reference Catalogue of Bright Galaxies* (RGC) (de Vaucouleurs *et al.* 1964) and the *Second Reference Catalogue of Bright Galaxies* (RGC2) (de Vaucouleurs *et al.* 1975) are compilations of all data on bright galaxies available at the time of publication. Tully (1987) has recently produced a *Catalog of Nearby Galaxies* (NBG) which includes all galaxies with known redshift less than 0·1.

Other lists of clusters, detected in 3D space, or outside the limits of Abell catalogue include the work of Humason *et al.* (1956), Klemola (1969), Snow (1970), Sersic (1974), Gunn & Oke (1975), Rose (1976), Lugger (1976), Sandage *et al.* (1976) and Spinrad (1977). Shectman (1984) used the *Lick Catalogue* to find clusters. These comprise sets of connected pixels (cells) lying above a given threshold. Catalogues of smaller groups of galaxies have been produced by Holmberg (1937), Sandage and Tammann (1975), de Vaucouleurs (1976), Turner & Gott (1976), Huchra & Geller (1982), and Stevenson (1985).

Supercluster structures have been catalogued by Bachall and Soneira (1983) and the properties of individual structures is discussed by Oort (1983). The *Atlas of Nearby Galaxies* (Tully & Fisher 1987) uses NBG to divide the nearby galaxy distribution into groups, associations and clouds.

Redshift surveys have the advantage that it is possible to detect structures that are

CGCG Selection Criteria

Richness Criterion. A cluster must contain at least 50 members within $3^{m}.0$ fainter than the brightest member.

Compactness Criteria. These galaxies must lie within the *cluster contour*, defined as the isopleth at ~twice the local field density.

Distance Criterion. There are no distance limits, but clusters covering very large areas of the sky (eg. Virgo) are not included.

Galactic-Latitude Criterion. Clusters must lie north of declination $-3°$ and within the areas specified in the introduction to Volume VI of the catalogue.

Table 1.5: The criteria for the inclusion of clusters in the CGCG (Zwicky *et al.* 1961–1968).

washed out by projection effects in 2D surveys. An all sky redshift survey of bright galaxies ($m_{pg} \lesssim 14.5$) in the CGCG was carried out by Davis & Huchra (1982) and is at present being extended to fainter magnitudes (de Lapparent *et al.* 1986). Deeper surveys over more limited areas of the sky have been produced by Gregory & Thompson (1978), Tarenghi *et al.* (1979,1980), Gregory *et al.* (1981), Kirshner *et al.* (1981,1987), Tago *et al.* (1984), Haynes & Giovanelli (1986), Peterson *et al.* (1986) Bicay & Giovanelli (1987), Shanks *et al.* (1983,1988).

The large-scale distribution of galaxies deduced from these surveys is varied and complex. Galaxies are seen not only to lie in the groups, clusters and clouds that are visible in the *Lick Catalogue*, but also in filaments or bridges between clusters (Chincarini 1981, Batuski & Burns 1985: $300h^{-1}$ Mpc filament) and in sheets and ridges (Haynes & Giovanelli 1986). In the "Slice of the Universe" map, produced from the extended CFA redshift survey (de Lapparent *et al.* 1986), all galaxies surveyed seem to lie in either a cluster or a filament. There is no evidence in this data for an underlying, uniformly distributed, *Field* population. Tully (1986) suggests there is evidence to connect structures on scales up to $z = 0{\cdot}1$. Differences in the clustering of galaxies of different morphological types or as a function of surface brightness has been found by Giovanelli *et al.* (1986), Davis & Djorgovski (1985), and Parker *et al.* (1988). Variation in clustering as function of luminosity has been suggested by Hamilton (1988). Recently the emphasis has changed to that of mapping the structure of voids, large volumes where the density of galaxies is extremely low (Davis 1982, Bachall & Soneira 1982, Kirshner *et al.* 1983 & 1987, Burns *et al.* 1988, de Lapparent *et al.* 1986). Some workers (Gott *et al.* 1986) are now attempting to quantify the topology of the large-scale structure, through the connectivity of high and low density regions.

1.2.5 STATISTICAL MEASURES OF CLUSTERING

A widely used statistic in galaxy clustering is the two-point correlation function or covariance function (Hauser & Peebles 1973). The angular two-point function ($\omega(\theta)$) measures the excess number of galaxies above random, at a given angular separation (θ), seen in projection on the sky. The spatial two-point function ($\xi(r)$) is similarly defined for 3-dimensional distributions. A detailed description of the two-point and other n-point correlation functions can be found in *The Large-Scale Structure of the Universe* (Peebles 1980).

For any distribution of objects in 2-dimensions, $\omega(\theta)$ is defined such that:

$$\delta P = \mathcal{N}^2[1 + \omega(\theta_{12})]\delta\Omega_1\delta\Omega_2, \tag{1.2}$$

where δP is the joint probability of finding objects in both elements of solid angle $\delta\Omega_1$ and $\delta\Omega_2$, separated by θ_{12}. $\mathcal{N}$ is the mean number density of objects.

There are various methods of estimating $\omega(\theta)$ (Peebles 1980 chapter III), but the most commonly used involves comparison with a catalogue of randomly distributed objects, as:

$$1 + \omega(\theta) = \frac{\mathrm{N}_{obj}(\theta)}{\mathrm{N}_{ran}(\theta)}, \tag{1.3}$$

where N_{obj} and N_{ran} are respectively, the number of object pairs and random pairs with separation $\theta \pm \delta\theta$. An advantage of this method is that known systematic variations or selection effects in the object distribution (eg. variations due to galactic extinction or boundary effects) can be included when generating the random distribution and hence removed from the estimation of $\omega(\theta)$. For counts of objects in bins or cells (eg. *The Lick Catalogue*) N_{obj} and N_{ran} can be replaced by the number of objects per bin, providing that $\omega(\theta)$ is only estimated on scales larger than the bin size.

Comparison of estimates of $\omega(\theta)$ for different datasets is facilitated by the *Scaling Law*. If $\omega(\theta)$ is estimated from two catalogues complete to different depths D_1 and D_2, then

providing the selection functions for the two catalogues are the same:

$$\omega_2\left(\theta\frac{D_1}{D_2}\right) = \frac{D_1}{D_2}\omega_1(\theta). \tag{1.4}$$

The form of this relation (see Peebles 1980 p 193) arises because the angular separation equivalent to a given spatial separation drops as D^{-1}, while the number of background galaxies rises as D. Hence $\omega \propto D^{-1}$ at fixed θD.

The agreement between estimates of $\omega(\theta)$ from independent surveys scaled to the same depth provides a strong constraint that the features seen are intrinsic to the galaxy distribution. It is unlikely that structure in $\omega(\theta)$ due to unrecognised selection effects or systematic errors would scale correctly.

One limitation on estimates of $\omega(\theta)$ from Eqn. 1.3 comes from using the total number of objects N to estimate the average number density $\mathcal{N}$. From Eqn: 1.2, the number of objects in $\delta\Omega$ at θ from a randomly chosen object is:

$$\delta n = \mathcal{N}[1 + \omega(\theta)]\delta\Omega. \tag{1.5}$$

The expected number of neighbours within separation θ is therefore:

$$\langle N\rangle_{neigh} = \mathcal{N}\int_0^\theta [1 + \omega(\theta)]d\Omega. \tag{1.6}$$

If we consider the total solid angle of the survey region Ω_R, then $\langle N\rangle_{neigh} = N - 1$ and as $\mathcal{N}$ is estimated as $\frac{N}{\Omega_R}$:

$$N - 1 = \frac{N}{\Omega_R}\int_{\Omega_R} [1 + \omega(\theta)]d\Omega, \tag{1.7}$$

which implies:

$$\int_{\Omega_R} \omega(\theta)d\Omega = -\frac{\Omega_R}{N}. \tag{1.8}$$

Hence any estimates of the true correlation function made in this way will be biased low. If the intrinsic correlation function is, for example, a power-law ($\omega(\theta) = A\theta^{-\delta}$, see below), then the estimated function will be $\omega(\theta) = A\theta^{-\delta} - B$. The correction B is

termed the *integral constraint.* Given the number of objects in a sample of area $\pi\theta_R^2$ is large, then:

$$B = \left(\frac{2A}{2-\delta}\right)\theta_R^{-\delta}. \tag{1.9}$$

It should also be noted that $\omega(\theta)$ will be underestimated on angles of the order of the scale size of the region used and therefore small samples may not represent a fair sample of the Universe. In this case estimates of $\omega(\theta)$ may not be a true representation of the correlation function over larger regions of the sky (Collins *et al.* 1989)

Groth and Peebles (GP) (1977) have estimated the galaxy two-point correlation function ($\omega_{gg}(\theta)$) using the SSGP reduction of the *Lick Catalogue.* The main result from their study was a power-law form for $\omega_{gg}(\theta)$ on scales $\lesssim 2{\cdot}0°$ (at a limit depth of $m_{pg} = 18{\cdot}6$). They estimate:

$\omega_{gg}(\theta) = A\theta^{-\delta}$ with $A = 0{\cdot}0684 \pm 0{\cdot}0057$ and $\delta = 0{\cdot}734 \pm 0{\cdot}035$.

The errors are estimated by evaluating $\omega_{gg}(\theta)$ in 4 separate zones of the *Lick Catalogue.* On scales $\gtrsim 2{\cdot}5°$ the correlation function breaks sharply below the power-law as can be seen in Fig. 1.2, which shows $\omega_{gg}(\theta)$ for each of the 4 zones and in Fig. 1.3, which shows $\omega_{gg}(\theta)$ as estimated from the *Jagellonian Catalogue* (Peebles 1975), the CGCG (Peebles 1974) and the GP estimate, all scaled to an effective depth (the Zwicky depth) for the CGCG ($m_{pg} = 14{\cdot}9$). At this depth the amplitude and slope for the power-law form are:

$A = 0{\cdot}073$ and $\delta = 0{\cdot}77 \pm 0{\cdot}04$.

The agreement between these scaled estimates provides good evidence that the observed correlation function is intrinsic to the galaxy distribution rather than due to systematic errors or selection effects.

The spatial correlation function ($\xi(r)$) has been estimated from various redshift surveys. Given the power-law form of the angular function, the spatial function is expected to

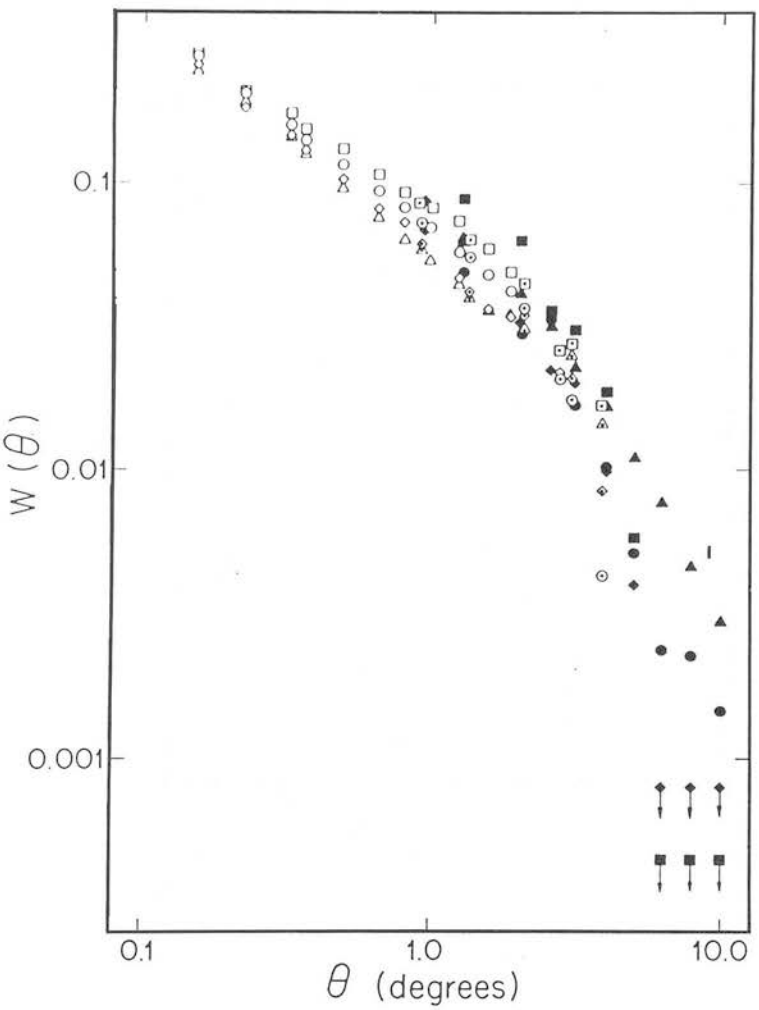


Figure 1.2: Angular 2-point correlation functions for 4 zones in the SSGP reduction of the *Lick Catalogue*. Reproduced from Groth & Peebles 1977.

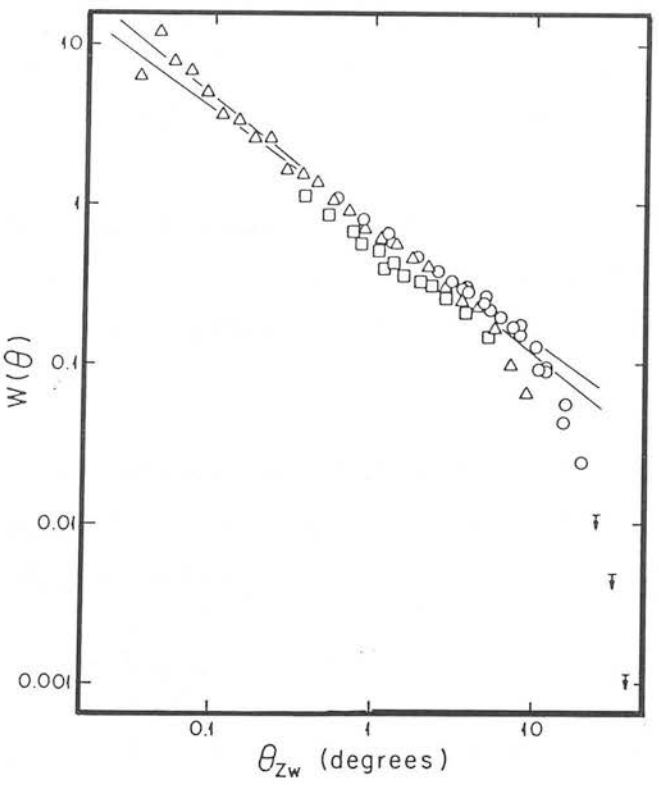


Figure 1.3: Angular 2-point correlation functions for data from the CGCG, SSGP and *Jagellonian Catalogue* scaled to the Zwicky depth. Reproduced from Groth & Peebles 1977.

be a power-law of exponent $-\gamma = -(1+\delta)$, such that:

$$\xi_{gg}(r) = Ar^{-\gamma} \equiv \left(\frac{r}{r_0}\right)^{-\gamma}, \tag{1.10}$$

where r_0 is the scale length of clustering. Using $\gamma = 1{\cdot}77$, Davis & Peebles (1983) estimated that $r_0 = 5{\cdot}3 \pm 0{\cdot}3h^{-1}$ Mpc using the CFA survey (Huchra & Davis 1982), while de Lapparent *et al.* (1988) estimate that $5 \leq r_0 \leq 12h^{-1}$ Mpc and that $1{\cdot}3 \leq \gamma \leq 1{\cdot}9$ from the extended CFA survey. Both these results are very uncertain on scales $\gtrsim 10h^{-1}$ Mpc (*ie.* the break scale from ω_{gg}) because of the errors in determination of the background density and the luminosity function. The second of these results is based on a survey that may not be a fair sample of the Universe, because there is structure visible on the scale size of the sample and because it is dominated by the Coma cluster.

Angular and spatial correlation functions have also been estimated for samples of clusters (ω_{cc} and ξ_{cc}). Clusters are observed to be considerably more clustered than galaxies. Ba all & Soneira (1983) estimate that:

$$\xi_{cc} = 360(r)^{-1.8}, \quad r \leq 150h^{-1} \text{ Mpc}$$

for Abell clusters of richness class > 1. Postmann *et al.* have used a volume limited version of this sample and several sub-samples to re-estimate the correlation function. They conclude that:

$$\xi_{cc} = 300(r)^{-1.8}, \quad r \leq 60h^{-1} \text{ Mpc}$$

and that original data was affected by the Corona Borealis Supercluster. The cluster–cluster correlation function is also known to be a function of Cluster richness, with richness class ≥ 2 clusters having a correlation function amplitude 3 times that of richness class $= 1$ clusters (Bahcall 1984, Postmann *et al.* 1986). Bahcall and Burgett (1985) suggest there is a relation between the amplitude of clustering and the mean

separation (d_i) of the objects concerned, such that:

$$\xi_i \left(\frac{r_i}{d_i}\right) = 0{\cdot}3 \left(\frac{r_i}{d_i}\right)^{-1.8} \tag{1.11}$$

1.3 The Theoretical Distribution of Galaxies

1.3.1 STANDARD BIG-BANG SCENARIO

In the Standard Big-Bang theory the Universe is considered to be an isotropic and homogeneous expanding fluid of matter and radiation. In this case General Relativity implies that the expansion is described by:

$$\frac{\ddot{\mathrm{R}}}{\mathrm{R}} = -\frac{4\pi \mathrm{G}}{3}\left(\rho + \frac{3\mathrm{P}}{\mathrm{c}^2}\right), \tag{1.12}$$

$$\dot{\rho} = -\frac{3\dot{\mathrm{R}}}{\mathrm{R}}\left(\rho + \frac{\mathrm{P}}{\mathrm{c}^2}\right), \tag{1.13}$$

where R(t) is the scale factor and P is the pressure. In the Newtonian approximation Eqn. 1.12 is the equation of motion and Eqn. 1.13 is the equation of mass conservation for an expanding fluid. If the present day Universe is considered to be dust filled (*ie.* pressure free) then integration gives:

$$\frac{\mathrm{d}}{\mathrm{dt}}(\rho \mathrm{R}^3) = 0 \implies \rho \mathrm{R}^3 = \mathrm{M} = \text{const}, \tag{1.14}$$

and the Friedmann equation:

$$\left(\frac{\dot{\mathrm{R}}}{\mathrm{R}}\right)^2 = \frac{8\pi \mathrm{G}\rho}{3} - \frac{\mathrm{kc}^2}{\mathrm{R}^2}, \tag{1.15}$$

with k = to the constant of integration (≡ the curvature constant). Solutions of this equation require:

$$\mathrm{kc}^2 = \mathrm{R}_0^2 \mathrm{H}_0^2 (\Omega_0 - 1), \tag{1.16}$$

where $\mathrm{H}_0 = (\dot{\mathrm{R}}/\mathrm{R})_0$ is the (present day) Hubble constant and the density parameter $\Omega_0 = 8\pi \mathrm{G}\rho_0 / 3\mathrm{H}_0^2$. The Friedmann equation can thus be rewritten:

$$\dot{\mathrm{R}}^2 - \frac{8\pi \mathrm{GM}}{3\mathrm{R}} = \mathrm{R}_0^2 \mathrm{H}_0^2 (\Omega_0 - 1), \tag{1.17}$$

with the behaviour of the expansion of the Universe dependent on whether Ω_0 is $< 1, > 1$ or $= 1$

As the Universe expands it cools. Initially the energy density of the Universe is dominated by radiation and the temperature is inversely proportional to the scale factor. All particles (p) and anti-particles ($\bar{\mathrm{p}}$) are in equilibrium via:

$$\mathrm{p} + \bar{\mathrm{p}} \rightleftharpoons \gamma + \gamma, \tag{1.18}$$

as long as $kT/c^2 \gtrsim$ the mass of the particle. In *baryonic* universes, the only particles of interest are electrons, positrons, massless neutrinos and anti-neutrinos, photons, protons and neutrons. For no other reason that, 'that is the way it is', a slight excess of matter over antimatter is built in at the start.

The ratio of neutron and proton number densities is set at $\exp(-2{\cdot}5 m_e c^2/T)$ by weak interactions, providing $T \gtrsim$ a few MeV. Below this temperature the reaction $e^- + e^+ \rightleftharpoons \nu + \bar{\nu}$ is slow. The neutrinos decouple and the ratio of neutrons to protons is effectively fixed at that implied by the decoupling temperature. As T drops below $\sim$ 0·1MeV deuterium begins to form because the reverse reaction to $\mathrm{p} + \mathrm{n} \longrightarrow \mathrm{D} + \gamma$ becomes slow. Further nucleosynthesis takes place once there is sufficient deuterium, turning all the neutrons and an equal number of protons into ^{4}He and trace amounts of ^{3}He, ^{7}Be and ^{7}Li. Detailed computation of the network of these nuclear reactions, coupled with observations of the abundances of light elements (*e.g.* in globular clusters or HII regions), require the density parameter for baryonic matter $\Omega_B \lesssim 0{\cdot}035 h^{-2}$ (Yang *et al.* 1984), *ie.* $\lesssim 0{\cdot}14$ for $h = 0{\cdot}5$. Thus if $\Omega = 1$, the Universe cannot be composed of purely baryonic matter.

As the Universe expands, the energy density of radiation drops as R^{-4}, while that of matter drops as R^{-3}. Consequently at $T \sim$ 2eV the Universe becomes matter dominated and at *recombination* ($T \sim$ 0·3eV), it becomes a pressureless fluid as the formation of atomic hydrogen implies it is transparent to photons. For this epoch onwards it evolves by Eqn. 1.15 (as P is effectively zero) according to the value of the

density parameter.

1.3.2 GRAVITATIONAL INSTABILITY

Formation of structure in the Universe is based around the concept of *gravitational instability*, *ie.*, that when gravity dominates pressure, small overdensities in an otherwise uniform mass distribution will evolve (collapse) into high density peaks. This concept was noted by Newton:

> It seems to me, that if the matter of our sun and planets, and all the matter of the Universe, were evenly scattered though all the Heavens, and every particle had an innate gravity towards all the rest, and the whole space throughout which this matter was scattered, was finite, the matter on the outside on this space would by gravity tend towards all matter on the inside, and by consequence fall down into the middle of the whole space, and there compose one great spherical mass. But if the matter were evenly disposed, throughout an infinite space, it could never convene into one mass; but some of it would convene into one mass and some into another, so as to make an infinite number of great masses, scattered at great distances from one another throughout all that infinite space.

For any density fluctuation in a fluid, there is a mass at which gravitational collapse takes over from the supportive effects of pressure. The effects of introducing perturbations into a fluid are found by linearising the the equation of continuity, Euler's and Poisson's equations, to yield the Jeans' wavelength:

$$\lambda_j = 2\pi \Big/ \left(\frac{4\pi G\rho}{c_{ad}^2}\right)^{\frac{1}{2}}, \tag{1.19}$$

with ρ equal to the density and c_{ad} equal to the adiabatic sound speed. If the wavelength of the perturbation is less than λ_j than the perturbation just oscillates. If it exceeds λ_j, *ie.* if the mass in an overdensity exceeds the Jeans' mass:

$$M_j = \left(\frac{4\pi}{3}\right)\rho\left(\frac{\lambda_j}{2}\right)^3, \tag{1.20}$$

then the perturbation collapses.

Consider *adiabatic* perturbations in an expanding baryonic universe, *ie.* the perturbations are such that the relative number densities of baryons and photons is constant. The evolution of a perturbation in the Universe depends on the variation of Jeans' mass as the Universe expands. Fig. 1.4, reproduced in part from Primack (1984), shows the variation of the mass enclosed within the horizon (M_H) and the Jeans mass as a function of temperature, epoch *etc.*

During the radiation dominated era M_j grows approximately as M_H. This is because the Universe is a relativistic fluid, so c_{ad} is $\sim c/\sqrt{3}$, and because the free fall timescale, $(G\rho)^{-\frac{1}{2}}$, must be $\sim t$ (the age of the Universe) in order for perturbations to collapse. M_j is therefore $\sim \frac{4\pi}{3}\rho(ct)^3$, *ie.* $\sim M_H$.

Consider a cluster sized perturbation ($\sim 10^{14}\,M_\odot$). This enters the horizon during the radiation dominated era, with $M_{pert} < M_j$ and is therefore pressure supported and so just oscillates. These perturbations cannot grow until after recombination when radiation pressure is no longer important and M_j has dropped to the value ($\sim 10^6\,M_\odot$) determined by standard gas pressure ($P = nkT$). Larger perturbations may begin to grow after matter domination, while smaller perturbations with $M_{pert} < 10^{12}\,M_\odot$ are washed out by Silk Damping. This is the process whereby Thompson scattering washes out the perturbations in the radiation-baryon fluid on the scale length of the photon random-walk. Galaxy size objects can thus only form by fragmentation during the aspherical collapse (into *pancakes*) of larger fluctuations.

The problems of galaxy formation is further impaired because fluctuations can only grow linearly in an expanding universe, ie $\delta\rho/\rho \propto R \propto (1+z)^{-1}$ and while $\Omega_0 z \gtrsim 1$. Galaxy structures can be expected to form when the perturbations go non linear, *ie.* $\delta\rho/\rho > 1$, and so at recombination perturbations should be $\sim 10^{-3}$ (z= 1500). Because the perturbations are adiabatic, the temperature fluctuation of the background radiation should be of the same order ($\frac{\delta T}{T} \sim \frac{1}{3}\frac{\Delta\rho}{\rho}$). Such fluctuations ($\Omega_B = 1$) are only marginally consistent with the measured inhomogeneity on arcmin scales of the

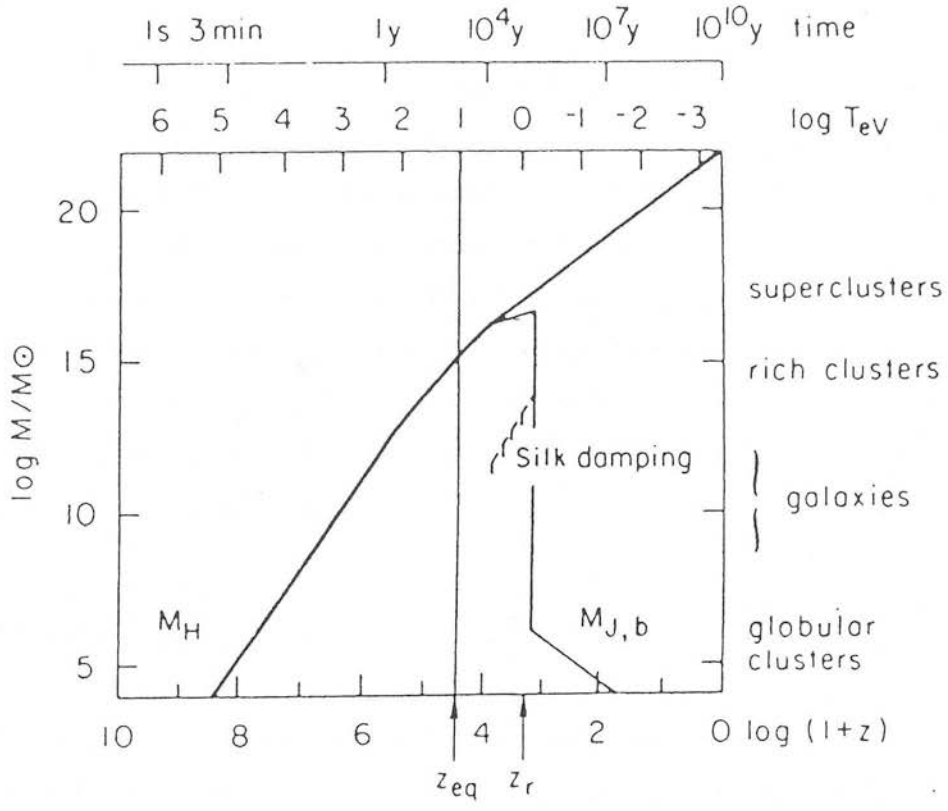


Figure 1.4: Variation of the baryonic Jeans mass (M j, b) and the horizon mass (M_H) as a function of temperature and redshift in a baryon universe. Reproduced in part from Primack 1984

micro-wave background radiation (Bond 1988). $\Omega_B = 1$ can however be ruled out by nucleosynthesis and observations of the light element abundances (Yang *et al.* 1984). $\Omega_B < 1$ implies larger density fluctuations at recombination and consequently larger temperature fluctuations in the MWBR. These have already been ruled out by observations (Bond 1988).

In universes where the mass density is dominated by baryonic matter, it would seem that there are significant observational constraints on the formation of galaxies. Solutions to overcome these constraints involving non-baryonic matter are described below, though Peebles (1987) has suggested an alternative where the perturbations are not adiabatic. In *isocurvature* models the perturbations in the baryons and the photons have opposite signs and amplitudes that result in no net fluctuation in the gravitational field. Such fluctuations predict temperature fluctuations in the MWBR that are consistent with the observations ($\Omega_B \leq 1$), but suffer because "there are no realistic proposals for generating them" (Bond 1988).

1.3.3 DARK MATTER SCENARIOS

Recent theories of galaxy formation invoke a form of dark matter that can only be detected by its gravitational influence. The value of Ω in universes composed of dark matter is not constrained by nucleosynthesis ($\Omega_{\rm DM}$ can be ~ 1). Evidence for the existence of such matter comes from the dynamics of galaxies and clusters. The measured rotation curves of spirals are constant out to several times the effective radius of the luminous mass, instead of falling off as $(\text{radius})^{-\frac{1}{2}}$. The mass of a cluster can be inferred from the dynamics of its constituent galaxies via the virial theorem. Zwicky (1933) was the first to realise that there was more mass present than could be accounted for in the luminous mass (M_{lum}). Present estimates suggest that $M/M_{lum} \sim 10$ for a wide range of masses, from dwarf spheroidals to galaxy clusters (Primack 1984).

Formation of structure in a universe dominated by dark matter is aided by the fact that

the candidate particles cease to interact with the radiation field (except gravitationally) well before recombination. Density fluctuations in the dark matter can effectively start growing as soon as the universe is matter dominated, *ie.* when the expansion timescale is comparable to the collapse timescale. Galaxy formation occurs after recombination when the baryonic matter falls into the gravitational wells of collapsed dark matter structures.

The structures expected in dark matter universes are dependent on the primordial fluctuation spectrum and the properties of the candidate particles. With $\delta_k(t)$ equal to the Fourier Transform of the density fluctuations ($\frac{\rho(r)}{\rho} - 1$), and $\delta_k(p)$ equal to the primordial *power-spectrum*:

$$|\delta_k(t)| = \mathrm{T}(k,t)\,|\delta_k(p)|\,. \tag{1.21}$$

$\mathrm{T}(k,t)$ is the transfer function that models the properties of the dark matter and its interaction with the radiation field. $|\delta_k(p)|^2$ is usually taken to be a power-law, $\propto k^n$, to ensure there is no preferred scale. $n = 1$ (predicted by inflation) is also often adopted as this ensures the amplitude of the density fluctuation is constant for all masses as they enter the horizon and so there is no preferred mass scale for which $\delta\rho/\rho$ would be unity as it enters the horizon.

In hot dark matter (HDM) scenarios the candidates are massive neutrinos which are relativistic at decoupling (T∼few Mev, Section 1.3.1). *Free streaming* of the particles therefore washes out fluctuations on scales smaller than the horizon size at which they become non-relativistic (R_ν), *ie.* on mass scales smaller than $\mathrm{R}_\nu^3 m_\nu n_\nu$ ($\sim 10^{15}\,\mathrm{M}_\odot$, with $m_\nu = 30\mathrm{eV}$). Fig. 1.5 shows the post-recombination fluctuation spectrum for neutrinos as a function of mass. Only large scale fluctuations have survived and so the first structures to form are supercluster size pancakes, inside which galaxies must form by some fragmentation process.

In cold dark matter (CDM) the candidates invoked have small thermal velocities, either because they are extremely massive (*e.g.* black holes or photinos) or because they are

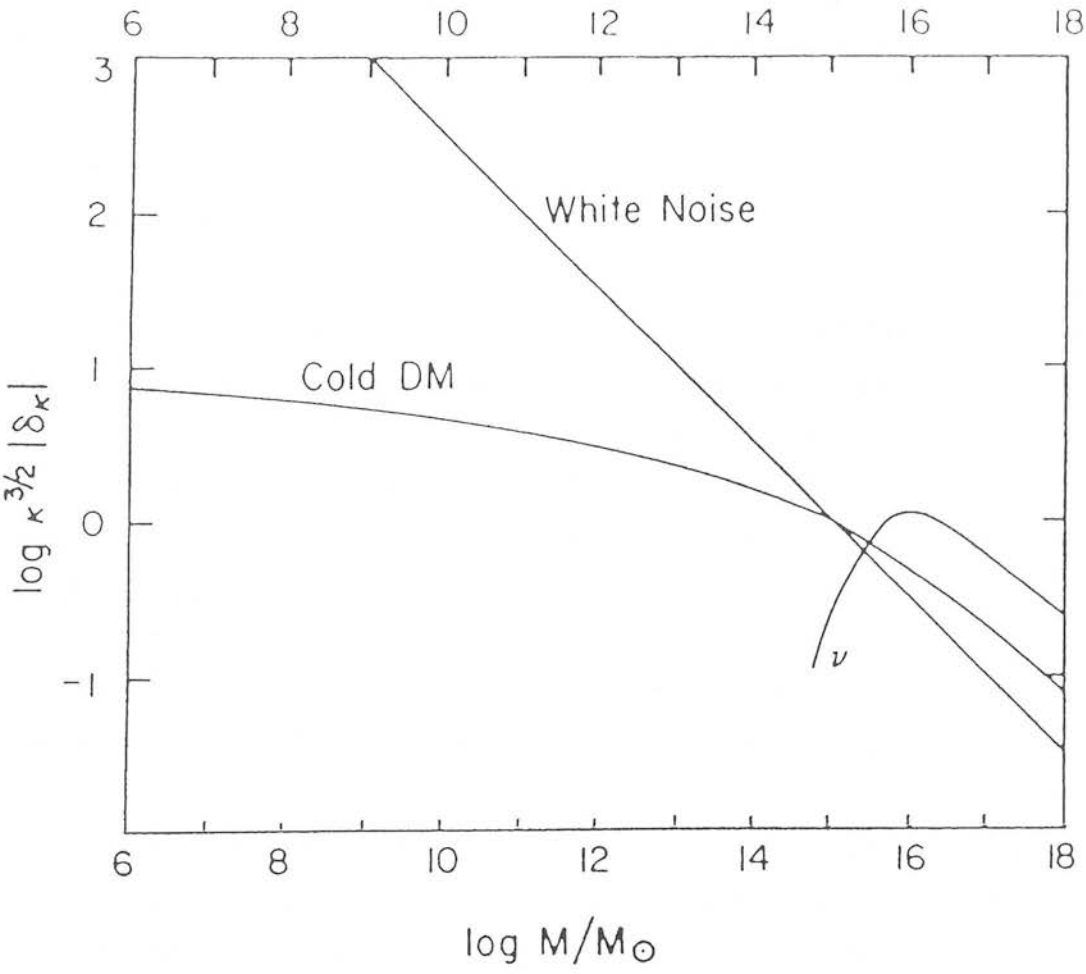


Figure 1.5: Spatial density fluctuations, post recombination for cold dark matter and neutrino dominated universes ($\Omega = 1$). Reproduced in part from Primack & Blumenthal 1984.

assumed to form that way (*e.g.* axions). The free streaming scale is therefore small and of no consequence. Fig. 1.5 shows the density fluctuation spectrum for CDM. Small scale power is moderated by the fact that fluctuations enter the horizon during radiation domination and hence cannot grow. In a CDM scenario, because there is power on all scales, a hierarchy of structures from globular to galaxy clusters forms rapidly, with the small scale limit being set by the baryonic Jeans mass just after recombination ($\sim 10^6 M_\odot$).

1.3.4 NUMERICAL MODELS

Much of the work on modelling the growth of structures in dark matter universes is done using N-body codes, where structures are allowed to grow by gravitation alone. In these models (White 1986, Frenk 1986 and references therein), the relevant power spectrum is imposed on a particle distribution which is then allowed to evolve in an expanding universe. For details on how the gravitational forces acting are calculated see White (1986).

In HDM simulations the resulting particle distribution consists of filamentary structure joining rich clusters, while at very late times the distribution is dominated by a few large clusters. One problem with all numerical simulations is that the galaxy distribution has to be determined from the particle distribution in some (simplified) way. Since in HDM models, galaxies must be formed by fragmentation of pancakes, galaxies are associated with particle cells in the simulation that have collapsed. This prescription allows the galaxy distribution and correlation function to be predicted from the particle distribution. Fig. 1.6 shows particle and 'galaxy' correlation functions for an $\Omega = 1$ neutrino dominated universe (from White 1986). The function is a power law (without a break) with a slope approaching the observed value at $z = 1$. The major constraint on such scenarios is, that for galaxies to have formed by the present day, the resultant distribution is far too clustered. The predicted galaxy correlation function amplitude is well above

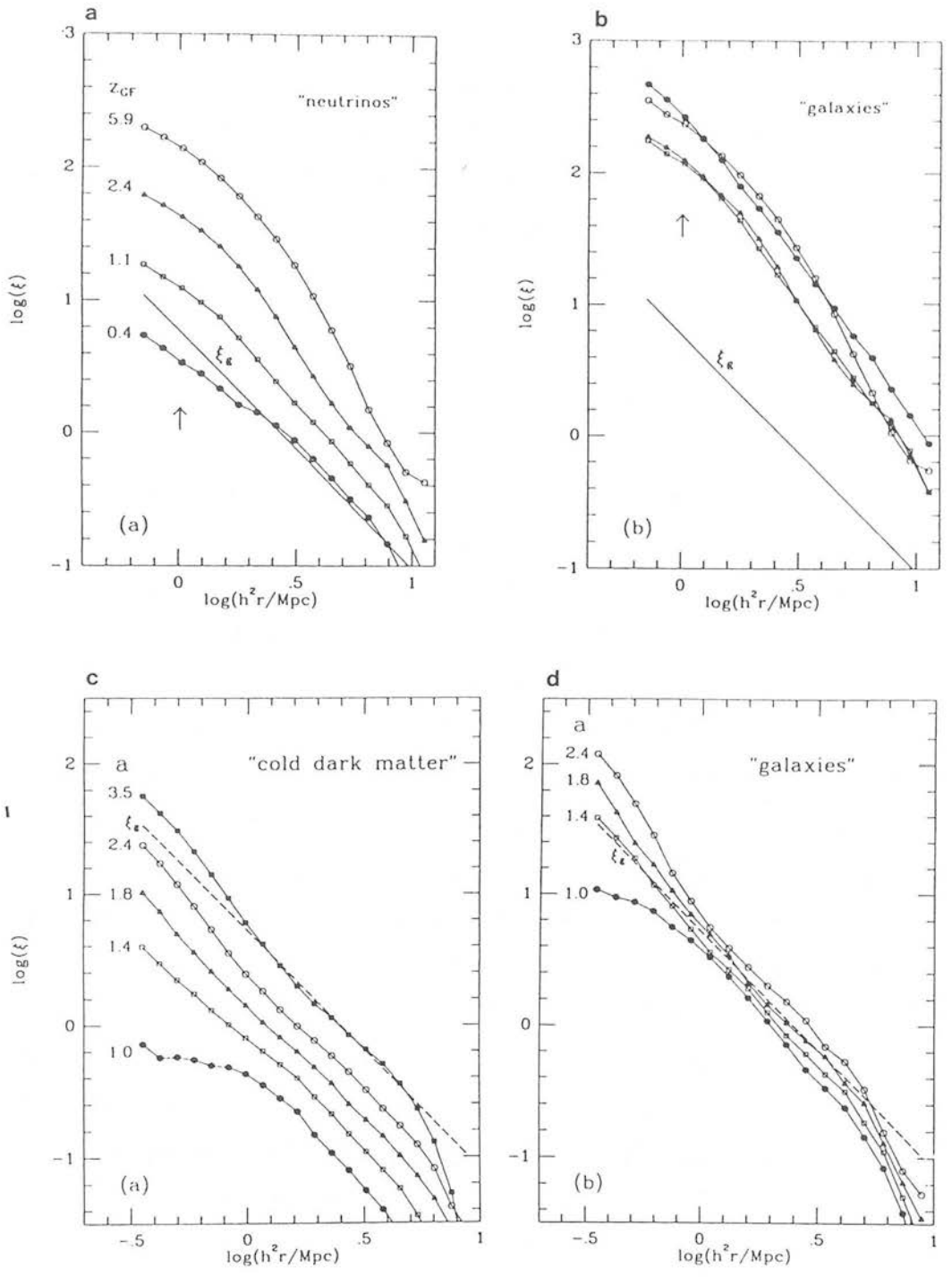


Figure 1.6: Correlation functions for numerical simulations of dark matter universes. Reproduced from White 1986. (a) & (b) particle and 'galaxy' autocorrelation functions for HDM models. (c) & (d) particle and 'galaxy' autocorrelation functions for CDM models. Functions are labelled by $(1 + z)$, the expansion factor, since the epoch when the first galaxies are formed. The solid line shows the observed function.

the observed (note $\Omega < 1$ increases clustering). One possible solution would involve a mechanism whereby galaxies could be made to avoid the most dense regions of the underlying mass distribution. More (1988) investigated the explosions as a possible mechanism for this, but concluded that far too much X-ray radiation would be produced in evacuating large enough bubbles.

In CDM simulations the power spectrum implies that fluctuations on many different scales go non-linear at almost the same time. This model can successfully produce filaments and superclusters of mass clumps (for $h = 0{\cdot}5$), but only at one time does the slope of the correlation function equal that of the observed and at that time the predicted amplitude is very much less than that observed (Fig. 1.6). Lower values of Ω produce more clustering, but the usual method to solve this (while keeping $\Omega = 1$) invokes *biased* galaxy formation. Here galaxies are only allowed to form at the those peaks of the matter distribution lying above some threshold and are thus intrinsically more clustered than the mass (Kaiser 1984). Biased CDM models have been extremely successful (Davis & Efstathiou 1987), for example they can predict the observed flat rotation curves of galaxy halos, but do have some serious shortcomings. The lack of power on large scales results in bulk motions on scales $\sim 50h^{-1}$ Mpc being several hundred km/s less than observed (Collins *et al.* 1986, James *et al.* 1988, White *et al.* 1988), and also the predicted cluster–cluster correlation function ($220r^{-1{\cdot}8}$, for $r < 150h^{-1}$ Mpc , White *et al.* 1987) is inconsistent with that observed for Abell clusters ($360r^{-1{\cdot}8}$ Bachall & Soneira 1983). Bond and Couchman (1988) added a spike into the CDM power spectrum at large scales in an attempt to remedy this situation. As shown in Fig. 1.7, the observed break in the $\omega_{gg}(\theta)$, which lends some support to the CDM model, is inconsistent with CDM+X.

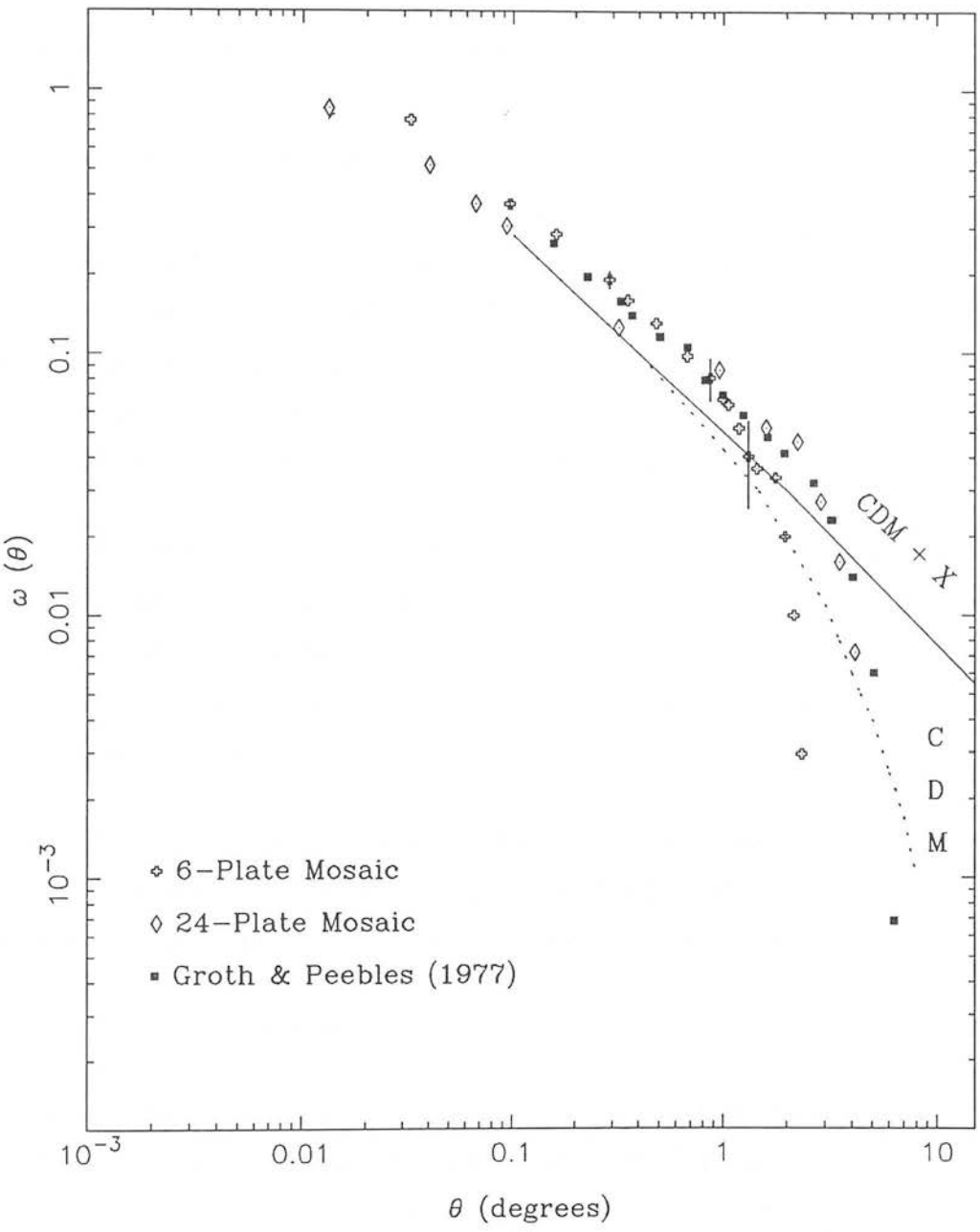


Figure 1.7: Theoretical angular two-point correlation functions for biased CDM and CDM+X models (Bond and Couchman 1988), together with the observed $\omega_{gg}(\theta)$ from Groth& Peebles (1977) (squares). The other data shown are from Collins *et al.* 1989.

1.4 Comparison of Theory and Observations

In the previous two sections the observed and theoretically predicted large-scale galaxy distributions have been described. In the latter section some comparisons between observations and theory were discussed to point out some of the shortcomings of HDM and CDM models. The biased CDM model was shown to be particularly successful at explaining many of the observed properties of the large-scale structure and of individual galaxies (Davis & Efstathiou 1987), and consequently this model is favoured over HDM, which produces an intrinsically over clustered galaxy distribution.

In Fig. 1.7 the observed $\omega_{gg}(\theta)$ (GP) calculated from the *Lick Catalogue* and that predicted by biassed CDM (Bond & Couchman 1988) are compared. The break in the observed function lends some support to the biassed CDM model which predicts a significant turn down on approximately the same scale. The reality of the break has however recently been brought into question.

Geller *et al.* (1984) and de Lapparent *et al.* (1986) have examined the possibility that residual errors in the correction factors applied by SSGP (Table 1.2) to the *Lick Catalogue* could be responsible for the break in $\omega_{gg}(\theta)$ observed by GP.

Geller *et al.* conclude that

(i) Individual plates are visible in both the SW and SSGP maps (Fig. 1.1) of the *Lick Catalogue*, as well as spurious structure associated with plate edges.

(ii) Zero or empty cells are not removed using the correction factors (Table 1.2). Such cells carry information on the plate scale as well as the position, in the survey, of erroneous (shallow) plates.

(iii) There is a break in the correlation function (ω_p) produced by applying the plate-to-plate factors to uniform datasets. As $\omega_{obs} = (1 + \omega_p)(1 + \omega_{gg}) - 1$, the break in

the observed function (ω_{obs}) maybe "indistinguishable from an artifact introduced by systematic variations in the effective magnitude limit".

(iv) The correlation introduced by systematic number density errors is:

$$\omega_p \sim \left(\frac{\delta n}{n}\right)^2 \sim (0{\cdot}6\ln(10)\Delta m_{lim})^2 \sim 2(\Delta m_{lim})^2, \tag{1.22}$$

where Δm_{lim} is the variation in plate depth. This is required to be much less than the intrinsic function, *ie.* $\omega_p \ll \omega_{gg}$, so for surveys of similar depth to the *Lick Catalogue* $\Delta m_{lim} < 0^{\rm m}.05$.

de Lapparent *et al.* examine possible causes of these residual number density variations. They conclude:

(i) Systematic variations in the counting efficiency of the two observers were sufficient to introduce both the filamentary appearance of the galaxy distribution and the break in $\omega_{gg}(\theta)$.

(ii) The counting gradient across plates in the raw data is dependent roughly on epoch. Application of averaged plate correction factors cannot, therefore, remove all the systematic errors.

(iii) The use of overlaps in the SSGP correction can propagate the effects of shallow plates over adjacent regions of the survey. A low density region close to the Coma cluster is an example of this. Such effects are particularly a problem where there is a group of shallow plates or where the plate depths alternate.

(iv) The residual rms systematic variation in the counts is estimated at 11% ($\equiv 0^{\rm m}.08$) of which 8% ($\equiv 0^{\rm m}.06$) is due to plate-to-plate variations.

The limits set in the above are probably a worst case example. The calculations assume that all the correction factors have a detrimental effect, when in fact they certainly improve the homogeneity over large regions of the survey. Groth and Peebles (1986a,b,

1988a,b) have re-examined the correlation function from the Lick data in response to the criticisms of Geller *et al.* . They calculated correlation functions separately for each observer and as a function of epoch. These function "agree extremely well with the average correlation function". Brown and Groth (1988) examined the affect of the plate variations on simulated datasets and then applied the SSGP correction procedure. The resultant correlation function from the corrected counts agreed with that "built-in" to the data. They also claim that the "corrected map is a good visual representation of the [simulated] galaxy distribution'.

Further to this work it is interesting to note that, subjective catalogues may contain many errors that it is impossible to quantify and which vary on very short times scales. For example it is very difficult to reproduce visual identifications of galaxies exactly, especially when working close to the plate limit. When the image classification procedure used in the EDSGC was compared with visual classifications, it was found that it was necessary to repeat the visual identifications in a double-blind way to ensure consistency. Phillips *et al.* (1981) have also suggested that the visual detection of galaxies produces a diameter or surface brightness limited catalogue, rather than a magnitude limited catalogue.

Shane and Wirtanen noted in their original work (1967) that there was "surprisingly... systematically more galaxies...counted on the east than on the west side of plates" and that the "most plausible explanation is some lack of symmetry in the optical system, though such is not evident from an inspection of the star images". The counts were however made in 10′ strips travelling from east to west and experience has shown that an observer tires very quickly and may not record as many objects when tired. It is pure conjecture, but it would seem reasonable that one strip would be counted per day or a rest (and a cup of coffee!) taken between counting strips. Variations of this kind are impossible to quantify. It should noted as well, that the visual homogeneity of a catalogue does not ensure that statistical measures are unaffected by systematic errors. Experiments show that number density changes of $\lesssim 15\%$ in otherwise random

distributions are very difficult to detect visually.

The inability of the biassed CDM model to reproduce the observed cluster-cluster correlation function was noted previously. One fault in these observations lies in the subjective nature of cluster detection procedures that have been used in the past. For example, though Abell's *statistical sample* only includes clusters if they meet the criteria in Table 1.3, the intial search was carried out by eye. Examination of Southern Abell Clusters in COSMOS datasets reveals that often the cluster is composed of several sub-units. One explanation is that as the observers know the approximate magnitude of galaxies in the region, they apply filtering on the scale of the m_{10} radius when deciding the candidate cluster centers. Abell *et al.* (1988) also note the difficulties involved in comparing clusters selected by different authors.

Sutherland (1988) has examined quantitatively a similar aspect of the Abell cluster detection procedure. He notes that the richness of distance clusters seen through the halos of nearby Abell clusters, may be overestimated compared to that observed if the cluster was seen in isolation. Such selection effects introduce the number of clusters detected in the vicinity of nearby clusters and so increase the amplitude of clustering. Sutherland estimates that the correct spatial function is:

$$\xi_{cc}(r) = \left(\frac{r}{14h^{-1}\ \mathrm{Mpc}} \right)^{-1.8}, \tag{1.23}$$

which should be compared with the Bachall & Soneira (1983) result in which the scale length for clustering in $30h^{-1}$ Mpc .

Postmann *et al.* (1988) compared Abell and Zwicky clusters and concluded that there was a factor of $\sim$ 10 difference in the amplitude of $\omega_{cc}(\theta)$ due to the different selection criteria. They also note that Abell richness may underestimated at low galactic latitudes. This suggests that the variation in clustering amplitude, as a function of richness, may partly be due to selection effects and not the result of biased galaxy formation.

Finally the theoretical predictions may also be dependent on the method of assigning clusters. Lumsden *et al.* (1988) have shown that the amplitude of the theoretical cluster-cluster correlation function is dependent on the form of filtering used to model the clusters from the density distribution. They obtain higher amplitudes for the theoretical functions than do White *et al.* (1987).

1.5 Summary

Since the 1950's a large amount of work has been carried out to determine the types of large-scale structure (*e.g.* clouds, filaments and voids) in which galaxies are distributed and to estimate the form of various statistical measures of galaxy clustering (*e.g.* the two-point correlation function). Most of this work is however based on galaxy and cluster catalogues (*e.g. The Lick Catalogue*) in which objects were selected and classified by visual scans of photographic plates. Systematic errors in the limiting magnitude and selection effects are difficult to quantify with such detection techniques. Recent work suggests that some of the observed features of the large-scale structure (*e.g.* filaments and the break in $\omega_{gg}(\theta)$) may be due to residual systematic errors rather than the intrinsic galaxy distribution.

Confirmation of the nature and effect of these systematic errors is important, as observations of the large-scale structure provide strong constraints on various theories of galaxy formation. In Hot Dark Matter scenarios it is difficult to reconcile both the amplitude and slope of $\omega_{gg}(\theta)$, as predicted by numerical simulations, with the observations. The form of $\omega_{gg}(\theta)$, including the break, lend some support to biased Cold Dark Matter scenarios. This model has been extremely successful, though other observations (*e.g.* $\omega_{cc}(\theta)$) predict more power on large-scales than can be accounted for by biased CDM. Hybrid CDM models with extra large-scale power added to the power spectrum remove the break from the galaxy two-point correlation function.

The Edinburgh/Durham Southern Galaxy Catalogue is based on modern high quality photographic plates and objective detection and classification techniques, rather than the subjective techniques employed previously. It will be possible, therefore, to quantify any residual systematics in this catalogue and accurately estimate their effects on the observed galaxy distribution.

Chapter 2

Raw Data for the EDSGC

2.1 Introduction

The largest previous survey of galaxies, the *Lick Catalogue*, is based on photographic plates taken with the 20 inch Carnegie Astrograph. One serious source of inhomogeneity in this work is the variation in plate quality, exposure times and observing conditions. The plates were also originally taken as part of a proper-motion study and so each each is accompanied by an objective grating image. Deep photographic plates are already crowded; adding further images can only further confuse image detection.

In this Chapter the raw data for the EDSGC are discussed. These are COSMOS digitised scans of *ESO/SERC Atlas* glass copies of the SERC J survey. This raw data has several advantages over that used in previous studies. The photographic material taken with the United Kingdom Schmidt telescope (UKST) using Kodak IIIa-J emulsion, is of a higher and more uniform quality, as well as being deeper than that used in the *Lick Catalogue* or produced as part of the Palomar Sky Survey. The detection of images by COSMOS is an objective process, which supplies accurate magnitude, shape and position information on each object, rather than the rough count per cell recorded (by eye) in the *Lick Catalogue*. The EDSGC raw data is therefore a better starting point for the production of a large-scale homogeneous galaxy catalogue, but still contains systematic variations that must be recognised and accounted for. Failure to do this will introduce systematic errors in image classification and galaxy photometry, which will in turn introduce spurious structures into the observed galaxy distribution.

This chapter is divided into two sections. The first, Section 2.2, deals with the photo-

graphic material used in the EDSGC, together with the processing and quality control techniques used. The second, Section 2.3, discusses the mechanics of COSMOS and details the methods of image detection and parameterisation. The COSMOS *deblending* or image de-merging software has been introduced to overcome incompleteness in the detection of images in crowded environments. This software is examined in Section 2.3.4 and changes are described to remove spurious images that would result from a less careful application of deblending techniques.

2.2 The ESO/SERC Atlas

2.2.1 SURVEY PRODUCTION

The *ESO/SERC Atlas* is a photographic atlas of the southern celestial hemisphere consisting of copies of *survey plates* taken with the United Kingdom Schmidt Telescope (UKST). The Atlas is divided into two passbands; blue or (J) and red (R). The J plates were taken as part of the SERC (J) survey which is now complete, while the equivalent survey in R is still in progress. The plate copying is carried out at the European Southern Observatory (ESO) onto both glass and film.

The plate material for the SERC (J) survey is Kodak IIIA-J emulsion, which is hypersensitised by soaking in either dry Nitrogen (plates before J1541) or Nitrogen and dry Hydrogen (plates after J1541). Hypersensitisation removes oxygen and water vapour from the emulsion, consequently improving emulsion sensitivity and low intensity reciprocity. Hypersensitisation is carried out at the UKST using an automated method that ensures repeatability and safety. The maximum speed of each batch of hypered plates varies by less than 20%. The J passband is defined by the response of hypersensitised IIIa-J, combined with a Schott GG395 filter which introduces a short wavelength cutoff. Image magnitudes in this passband are referred to as 'b_j' magnitudes. Photographic b_j, Cousins' photoelectric 'B', and CCD magnitudes are compared in Fig.

4.4.

The SERC (J) survey covers $\sim$ 4·25sr of the southern sky from $\delta = -17°$ to the south celestial pole and consists of 606 plates, centred on 606 of the 894 ESO/SERC southern sky standard field centres. The other 288 fields make up the Equatorial survey zone from $\delta = 0°$ to $-17°$. As each UKST plate subtends $6{\cdot}4° \times 6{\cdot}4°$ and the field centres are separated by 5°, there is $\frac{3}{4}°$ overlap between plates. The Field centres are numbered from 1 to 606, while plates are numbered sequentially as they were taken. For clarity field numbers will be prefixed 'F' and plate numbers will be prefixed 'J'. Full details of the SERC (J) survey, field centres and the UKST can be found in the UK Schmidt Telescope Unit (UKSTU) Handbook (1983).

The UKST is a classical Schmidt telescope with a 1·2m diameter aperture and corrector lens and a 1·8m diameter mirror. The main advantage of the Schmidt design is the wide field of view, though this does result in a small plate scale $67''{\cdot}12\ \mathrm{mm}^{-1}$. Vignetting is negligible within a radius of 2·7° from the field centre, but rises steeply after that (Fig. 2.1). An achromatic corrector was fitted in 1977 and used for all plates after J3153.

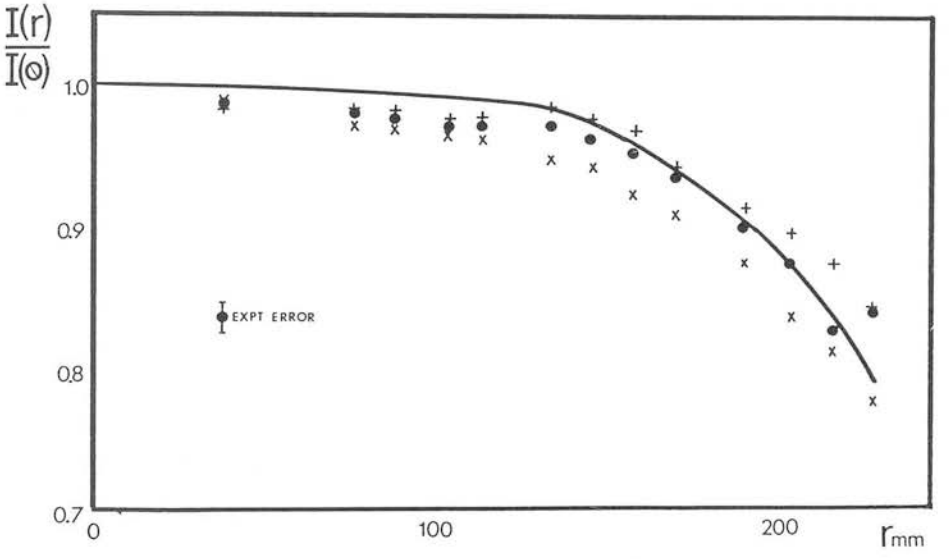
Survey plates are only taken in good seeing ($< 3''$), dark time (ie no moon), with the sun $> 18°$ below the horizon (astronomical twilight). Exposures are taken so the plate centre crosses the meridian and the polar axis is adjusted to reduce image trailing at the plate edges due to differential refraction (Wallace & Tritton 1979). A step-wedge and/or spot sensitometer are exposed at the plate edges during the sky exposure. The exposure time for survey plates is set between 60 and 80 minutes depending on the results of hypering and exposure tests. The resulting limiting magnitude is approximately $24^{\mathrm{m}}{\cdot}0$ in b_j, with a plate background of typically $22^{\mathrm{m}}{\cdot}3$ per arcsec^2. During exposure plates are held in a curved plate holder under a flat filter. The varying amount of air trapped between the plate and filter causes differential desensitisation and a steepening of the vignetting curve as seen in Fig. 2.1 (Dawe & Metcalfe 1982). Routine Nitrogen flushing of the plate holder was introduced in 1982 to cure this problem. Plates prior to J8279

were not nitrogen flushed and tend to have larger background variations, though the effect is a function of exposure time and humidity (Dawe, Coyte & Metcalfe 1984).

2.2.2 PLATE PROCESSING AND QUALITY CONTROL

The original glass plates from the UKST are processed at the Siding Springs Observatory (SSO). Full details of the careful processing procedure (developing, stopping, fixing, washing and drying) are given in Cannon *et al.* 1978. The process has been designed, along with careful experimentation, to be consistent and to ensure that the processing technique does not introduce non-uniformities. Examples of such non-uniformities that have been detected, include a "swirling" in the plate background due to operator-dependent methods of draining and a striped pattern due to non-uniform drying. This latter problem was solved by introducing turbulence into the laminar flow, previously used to dry plates, preventing the setting up of standing waves. Survey plates suffering from non-uniformities have been retaken for the *ESO/SERC Atlas*. The processing procedure is monitored by occasionally developing a uniformly exposed plate and examining it for non-uniformities.

Any prospective survey plate is only accepted if it passes a strict quality control, "meeting a minimum standard of resolution, image shape and exposure depth" (Cannon *et al.* 1978). Plates are given a score for each criterion and then graded as A,B or C, according to their total score. All Survey plates are grade A, i.e., have a total score of less than 3. Survey plates thus have faint images smaller than 45 μm and individual images elongated by less than 15 μm. The quality of plate exposure is assessed in terms of the image contrast at the background density level. The relationship between photographic density, $(-\log_{10}[\text{transmission}])$ and $\log_{10}(\text{exposure})$ is termed the characteristic curve. The contrast (slope) of the linear portion of the characteristic curve is found to be a function of exposure time. Setting a minimum value for the contrast ensures that the background density lies on the linear portion of the characteristic curve and optimises



continuous line: unhypered plates & geometric vignetting

+: hypered plates without filter, no N_2 flushing

×: hypered plates with filter, no N_2 flushing

•: hypered plates with filter, N_2 flushing

Figure 2.1: The Vignetting function for UKSTU plates. The data is taken from Dawe & Metcalfe (1982) and shows various combinations of hypersensitisation and N_2 flushing, for plates taken with or without the filter.

the detection of faint images and nebulousity.

Plates are also assessed for various faults such as processing marks, electrostatic marks and emulsion flaws. Plates may be rejected from the survey if these faults are considered (subjectively) to be unacceptable. A full list of these faults is given in Table 2.1. Localised faults, such as scratches or electrostatic marks, are only important for the EDSGC in that they produce spurious images that add to the number of noise images in the raw data (Section 2.3.3). Flaws that cover larger regions of plates may affect image classification and photometry (Chapters 3 & 4).

2.2.3 PLATE COPYING AND QUALITY CONTROL

The original plates accepted for the SERC (J) survey are kept in safe storage at the SSO. To limit the chance of damage these plates are not freely available. *ESO/SERC Atlas* plates are glass copies of grade A SERC (J) survey plates.

The copying process is designed to "reproduce as faithfully as possible the position, structure and contrast of all faint images (stellar and extended) uniformly over the whole of the original plate" (Standen & Tritton 1979). One immediate problem is that it is impossible to preserve the very wide dynamic range ($\sim$ 4 in density) present on IIIa-J original plates. The copying process has been optimised to faithfully reproduce (ensure linearity of the characteristic curve) images of the faintest objects, with the resultant saturation (reproduction at zero contrast) of brighter images. Independent evidence on the faithfulness of image reproduction during the copying process is provided by Brück & Waldron (1984). They conclude that "within the curtailed density range of the copies the information on the original plate is reproduced with remarkable fidelity." Copy-plates are produced so the background density is approximately constant (0·4 to 0·6) from plate to plate regardless of the background density of the original (0·8 to 1·8) and with linearity preserved over a range of $\sim$ 2 in density. The implications of the lower background for machine measurement are discussed in Section 2.3.3.

ESO/SERC Atlas Plate Faults.

Localised faults resembling astronomical objects:

Processing Faults. Streaky non-uniformities in the plate background. Strongly resembles faint nebulosity.

Electrostatic Marks. Round fuzzy spots resembling structureless galaxies.

Emulsion Faults. Scratches, gouges, holes, desensitised spots, poisoned spots, streaks of lower emulsion thickness, and shadows of dust particles (particularly doughnut-shaped shadows that may resemble globules).

Faults affecting larger regions of plates.

Fogging. Faults in manufacture, packaging or hypersensitisation may generate uneven fogging of the background.

Haze Halos. Atmospheric dust or thin haze produces fuzzy haloes around bright stars. Plates where more than 10% of the exposure is affected by cloud are rejected. Those affected for 5 to 10% of the exposure are rejected if they compare unfavourably with adjacent plates.

Table 2.1: Plate faults assessed during quality control of the *ESO/SERC Atlas*

Glass copies are produced at ESO and full details of the copying process are given in Standen & Tritton (1979) or West & Dumoulin (1974). Several positives (white stars on black sky) are produced from each original plate (negative). The procedure is carried out in a clean environment that ensures the removal of most (99·8%) dust particles larger than 0·3 μm and during exposure the negative and positive are held together using a vacuum. Thus the number of noise images, due to dust and defocussing during exposure, are kept to a minimum. The same care is taken during production of the Atlas negative copies from the intermediate positives.

The processing of positive and negative copies follows closely that used for processing the originals. A quality control process similar to that described above is also performed, as well as a comparison of the copies with the original. Film versions of the atlas are produced in greater number and in a production line process. Quality control of these copies is less strict and the quality is more variable.

2.2.4 PLATES USED FOR THE EDSGC

To ensure homogeneity and because of the limited availability of SERC (J) grade A duplicate original plates, the plates used for the EDSGC are all, bar one, glass Atlas plates from the version of the ESO/SERC atlas stored at the UKSTU at the Royal Observatory Edinburgh. Within the limits of the copying quality control this version of the atlas is identical to all other versions. Fig. 2.2 shows the extent of, and the field centres covered by the EDSGC. Table 2.2 lists the field centres with the corresponding plate numbers and plate grades of the original plates.

The Atlas plate for Field 411, J4606, has not been used in the EDSGC because the background density appears streaky and varies considerably across the plate. This was noted during quality control as "real nebulousity". Because problems were encountered during image classification, a new *original* plate for F411 was taken in 1987 by the UKST under the same observing criteria applied to survey plates (seeing was 2″). Processing

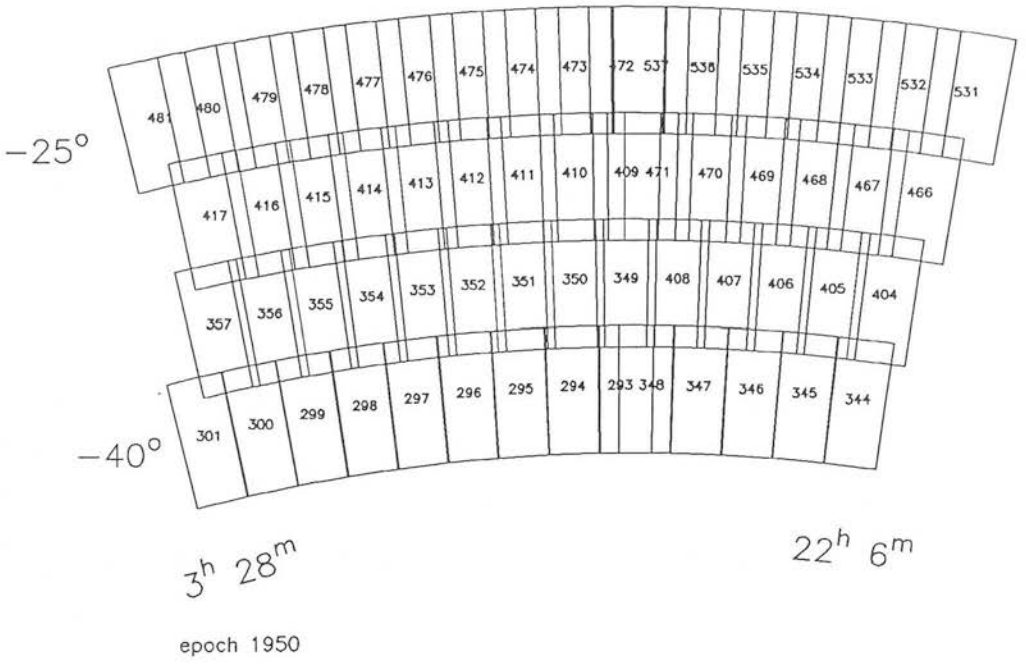


Figure 2.2: A map of the field centres covered by the EDSGC. For clarity not all fields are numbered.

Field	Plate	Grade	Field	Plate	Grade	Field	Plate	Grade
F293	J3578	AI2	F355	J5304	AI3	F469	J3508	AI3
F294	J3551	AI2	F356	J5462	A1	F470	J3515	A1
F295	J2712	A1	F357	J6470	AI2	F471	J6138	AI2
F296	J3524	AI2	F404	J2532	A2	F472	J3428	AI2
F297	J3593	A2	F405	J6231	AI2	F473	J3566	A1
F298	J3560	A2	F406	J9529	AUI3	F474	J6277	AI2
F299	J3588	A2	F407	J6468	A2	F475	J8031	A2
F300	J3536	A2	F408	J10342	AI3	F476	J8823	A2
F301	J2765	A1	F409	J2693	A2	F477	J3785	A1
F344	J6144	A1	F410	J2696	A2	F478	J2708	A1
F345	J3538	AT2	F411	J11341	A2*	F479	J5522	AI2
F346	J8022	AI3	F412	J3516	AI3	F480	J3649	AI2
F347	J2413	A1	F413	J3774	A1	F481	J3820	A2
F348	J3592	A2	F414	J3579	A2	F531	J8622	A1
F349	J6145	AI2	F415	J4607	A1	F532	J3420	AI2
F350	J4601	A2	F416	J4608	A1	F533	J6091	A1
F351	J8046	AI3	F417	J3818	AI3	F534	J6489	A1
F352	J6124	A1	F466	J5113	AI3	F535	J6114	A0
F353	J3596	A1	F467	J3514	AI2	F536	J3421	AI2
F354	J4630	A2	F468	J6436	AI3	F537	J8030	A2

Table 2.2: *ESO/SERC Atlas* J plates used in the EDSGC. Plate grades are described in the text. The following abbreviations are used: (I) Image size score ≥ 2. (T) Trailed, shape score ≥ 2. (U) Underexposed, contrast score ≥ 2. (*) not a survey plate.

and quality control was carried out in exactly as described above. The plate is grade A, with a total score of 2. A glass copy was made at the ROE Photolabs, with quality control performed at the UKSTU, in a similar manner to that applied to Atlas copies at ESO. As an internal check of consistency, an ROE Photolabs copy of a positive from a randomly chosen SERC (J) survey plate (J9529, F406) was also made at the same time. A comparison of the ROE and ESO copies of the chosen field is made in Section 3.5. The survey plate J7181 (F418) suffers from similar background variations to J4606, but as it lies on the edge of the EDSGC region it has been removed from the EDSGC.

2.3 COSMOS

2.3.1 MACHINE DETAILS

The COSMOS machine at the Royal Observatory Edinburgh is a plate scanning machine, or more precisely a flying spot microdensitometer. The acronym, COSMOS, stands for; *Co-Ordinates Shapes Magnitudes Orientations Sizes*, and refers to the information that can be obtained for each of the, on average, 2×10^5 images visible on a UKST survey plate (taken at the SGP).

The design of COSMOS, together with the various machine parameters and software used is described in detail in MacGillivray & Stobie (1984) (hereafter MGS) Some modifications have been introduced since this article, so brief details of COSMOS and the plate scanning procedure are given here. A full list of machine parameters used for scanning plates in the EDSGC are given in Table 2.3.

COSMOS uses a *spot* of light from a Cathode Ray Tube (see Fig. 2.3) to raster scan up to 287mm × 287mm (5·35° × 5·35°) of a photographic plate. The plate is scanned with south at the top (in order to avoid the step-wedge) in a series of 128 lanes. The

measurement area	287mm × 287mm 5·3° × 5·3° (UKST plate)
aperture diameter	32 μm
pixel spacing	16 μm
scan lane width	128 pixels
background cell size	1024 × 1024 μm
background determination	median value within cell
background filtering	2D Brownrigg filter
no. transmission levels	16384
usable density range	$D_{bg} + 1{\cdot}3$
detection threshold	7% to 9% of sky
detection isophote (b_j)	$\sim 25^{\mathrm{m}}{\cdot}0/\mathrm{arcsec}^2$
minimum image area	3 pixels
areal profile for deblending	1·07
deblend levels	8
max deblend area	60,000 pixels
min daughter area	3 pixels

Table 2.3: Parameters used by COSMOS when scanning and deblending images on UKST plates for the EDSGC.

CRT spot scans in the x direction while the plate is moved in the y direction. An autofocus system keeps the spot in focus to counteract plate sag. The transmitted spot intensity (S_t) is compared with a reference signal (S_r), to produce a 14-bit transmission value ($T = S_t/S_r$) for each pixel of the raster scan. This transmission value provides a measure of the emulsion density, which in turn is converted to a measured intensity value (I_p) for *each pixel* using a Baker Density calibration curve.

Photographic density is defined as:

$$D = -\log_{10}(T/T_c), \tag{2.1}$$

where T_c is the transmission of the clear plate, ie unexposed chemical fog only. There is however only a limited linear relation between density and log(exposure),so Baker Density, defined as:

$$BD = \log_{10}(\frac{T_c}{T} - 1{\cdot}0), \tag{2.2}$$

is often used because it is more nearly linear with log(exposure).

COSMOS Baker density is defined:

$$BD = \log_{10}\left[\left(\frac{T_c - T_b}{T - T_b}\right) - 1{\cdot}0\right]. \tag{2.3}$$

This accounts for the effects of an apparent 'transmission', T_b, caused by amplifier dark current. Calibration values for the Baker density curve for each plate are obtained from the step-wedge or spot sensitometer densities, which have known increments in exposure. T_c is measured for each plate, while T_b can be tuned to improve the linearity of the calibration data. A typical BD calibration curve is shown in MGS. A cubic spline is fitted to the sensitometer data to provide a BD-to-intensity look-up table, to convert each pixel BD value to a measured intensity value.

COSMOS can operate in one of two measurement modes. Mapping Mode (MM) records the intensity value for every pixel in the area of plate being measured. Because of the large storage requirements (620 Mbytes for the measurable area of a UKST plate), this

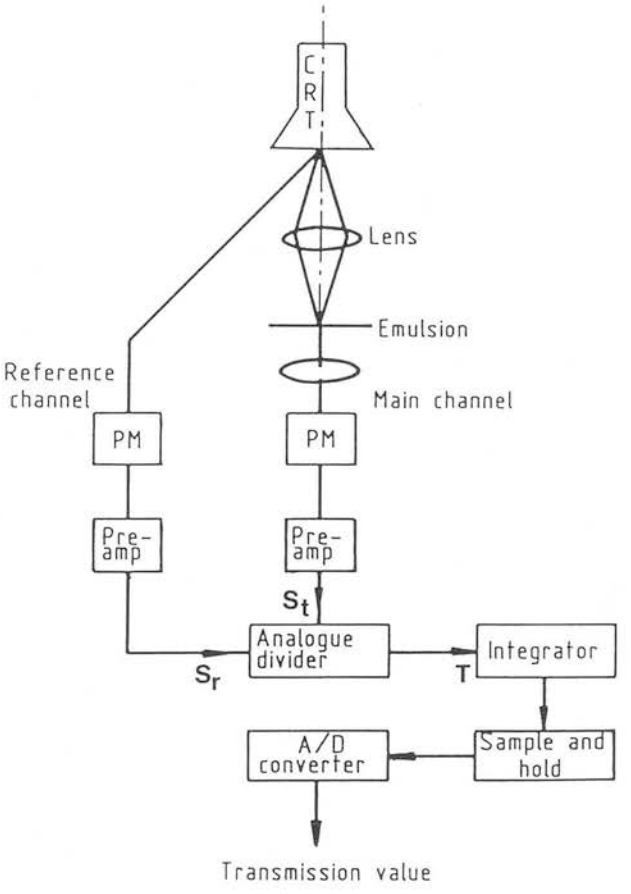


Figure 2.3: A schematic diagram showing the layout of the COSMOS scanning/detection system.

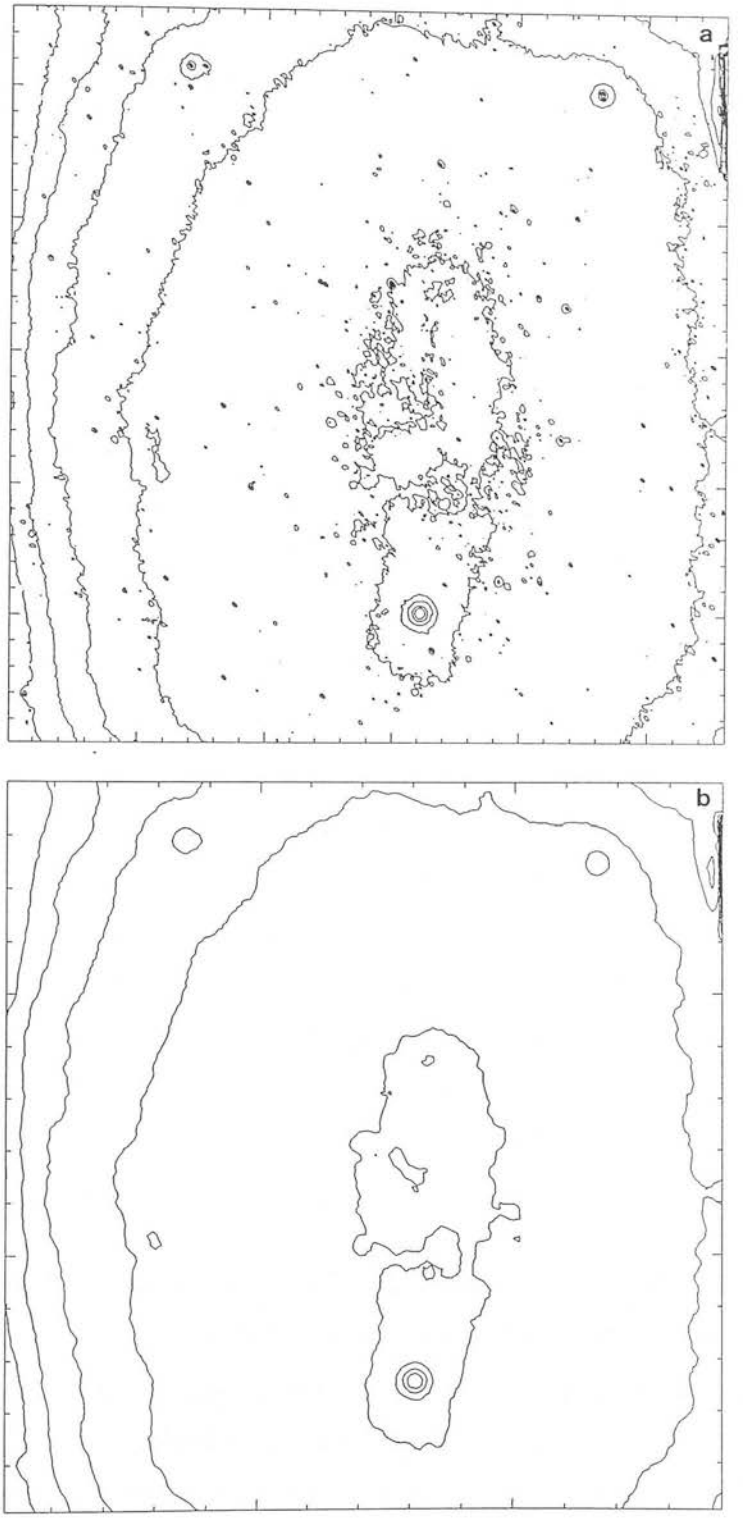


Figure 2.4: Contour maps of UKST plate backgrounds as measured by COSMOS, (a) before filtering, (b) after filtering. Data is from J6138 (F471).

mode is only used for small regions of a plate or individual images. In Threshold Mode (TM), pixel intensities (I_{pix}) and positions (x_p, y_p) are only recorded for those pixels lying above some threshold, which is set as a percentage above the plate background intensity. The variation of the background is mapped by *prescanning* the plate at low resolution (32 μm pixel spacing) and then determining a background value for each of an array of background cells (32, 64 or 128 prescan-pixels square). The background intensity value in each cell is the median of the prescan-pixel intensities, filtered across cells using a 2D Brownrigg filter (Brownrigg 1984). This procedure removes variations in the background caused by large objects. The local background intensity for each image pixel (I_{bg}) is then obtained by interpolation between cell values. Fig. 2.4a & b show the variation in plate background for J6138 (F471), before and after filtering. The two small circular of contours at the top-left and top-right are due to two sets of double stars. The densely grouped contours at top right are produced by the step wedge which protrudes into the scan region of this plate[1], while the variation in the centre is caused by an extremely bright star and its associated emulsion-corrector/filter-corrector ghost (UKSTU Handbook). This background map is often referred to as a *sky map* or as *sky intensity* variation. In fact the variation in background density ($\sim 0^{\rm m}.2$ Dawe & Metcalfe 1982), is primarily due to vignetting and differential desensitisation rather than intrinsic variation in the sky (Campbell 1982, Wesselink 1984). By comparison Zodiacal light, which is the brightest component of the night sky excluding moonlight and aurora, will vary by $\sim 0^{\rm m}.04$ across the 5° of a UKST plate (at 140°solar elongation, Roach & Gordon 1973).

The raw pixel data is passed to the COSMOS image analyser (Thanisch *et al.* 1984). In Image Analysis Mode (IAM) sets of connected pixels are defined, with each set constituting an image. Various image structure parameters, as well as image magnitude and position, can be obtained from the unweighted and intensity weighted moments of the pixel distribution within each image (Stobie 1980 & Section 2.3.2). Because of the variation in the plate background intensity across the plate all pixel intensities used in

[1]The plates are scanned with south at the top so that, for most plates, this does not occur

the moments calculation are background subtracted.

COSMOS image intensities and corresponding image magnitudes are isophotal. The value of the isophote (b_j magnitudes arcsec^{-2}) is a function of the detection threshold and background intensity. The latter is unknown for most plates at the time of scanning, but can vary by up to $\sim 0^{\rm m}.5$ from plate to plate (rms variation $0^{\rm m}.2$). Adopting a fixed detection threshold for each plate will introduce a systematic variation in magnitude for galaxies, dependent on their position in the survey. Assuming that the majority of images detected are faint galaxies and that the sky distribution of these galaxies ($b_j \simeq 22{\cdot}0$) is uniform on scales $\sim 5°$, then the number density of objects should be a fair estimate of the plate depth. The COSMOS detection threshold is set iteratively, between 8% and 10% above the background, so that the number of objects (2×10^5) detected is approximately constant from plate to plate. This choice of threshold corresponds to a nominal isophote of $25^{\rm m}.0$ arcsec^{-2} in b_j. For images brighter than $b_j \simeq 20{\cdot}5$ the COSMOS magnitudes are effectively total (MacGillivray & Dodd 1982), while the measured magnitudes of fainter images will be systematically fainter than their total magnitude. A lower threshold, *e.g.* 3%, would make fainter magnitudes $\sim$ total, but would increase the number of images dramatically. The 8% threshold restricts the number of noise images to a "practical level" (MacGillivray & Dodd 1982).

2.3.2 IAM PARAMETERS

There are 18 independent IAM parameters which are constructed from the zeroth, first and second order moments of the pixel positions and intensity distribution within each image. A full list of COSMOS parameters available to the EDSGC together with units and the measurement accuracies are given in Table 2.4.

The image area ($\mathcal{A}$) is the zeroth order unweighted moment, ie. N the number of pixels in the image ($\mathcal{A} = {\rm N} \times A_{pix}$, but for convenience conversion factors in this and subsequent formula are omitted). The background subtracted image intensity (I)

parameter	description	units	accuracy
RA	right ascension (1950)	decimal hrs	$0^{s}\cdot 8$
DEC	declination (1950)	decimal degs	$2{\cdot}0''$
x_{min}	minimum x coordinate	$0{\cdot}1\,\mu$m	$2{\cdot}0\,\mu$m
x_{max}	maximum x coordinate	$0{\cdot}1\,\mu$m	$2{\cdot}0\,\mu$m
y_{min}	minimum y coordinate	$0{\cdot}1\,\mu$m	$2{\cdot}0\,\mu$m
y_{max}	maximum y coordinate	$0{\cdot}1\,\mu$m	$2{\cdot}0\,\mu$m
$\bar{x}_i$	intensity weighted x centroid	$0{\cdot}1\,\mu$m	$1{\cdot}2\,\mu$m
$\bar{y}_i$	intensity weighted y centroid	$0{\cdot}1\,\mu$m	$1{\cdot}2\,\mu$m
$\mathcal{A}$	image area	arcsec^2	3 pixels
a_u	unweighted semi-major axis	arcsec	3%
b_u	unweighted semi-minor axis	arcsec	3%
a_i	intensity weighted semi-major axis	arcsec	2%
b_i	intensity weighted semi-minor axis	arcsec	2%
θ_u	orientation on the plate	degrees	2°
θ_{sky}	position angle on sky	degrees	2°
I_{max}	maximum intensity above sky	relative	—
I_{bg}	local background intensity	relative	—
m_{cos}	COSMOS calibrated magnitude	magnitudes	$0^{\mathrm{m}}{\cdot}03$
m_{cor}	COSMOS corrected magnitude	magnitudes	see Ch.3
b_j	image magnitude in b_j	magnitudes	see Ch.4
$\log_{10}(\mathcal{A})$	log image area in pixels	no units	—
$\mathcal{G}$	geometric parameter	no units	—
$\mathcal{S}$	Gaussian width	pixels$^{\frac{1}{2}}$	—
F	Field number	no units	—
id	identification number	no units	—

Table 2.4: Image parameters available within the EDSGC, together with units and measurement accuracies (for images with $m_{cos} \simeq -2{\cdot}0, \mathcal{A} \simeq 20 - 30$ pixels).

together with the corresponding magnitude ($m = -2{\cdot}5 \log_{10} \mathrm{I}$) is constructed from the weighted moment:

$$\mathrm{I} = \sum_{\text{image}} (\mathrm{I}_{pix} - \mathrm{I}_{bg}) \tag{2.4}$$

The intensity weighted centroid $(\bar{x}_i, \bar{y}_i)$ and unweighted centroid come from the first order moments:

$$\begin{aligned} \bar{x}_i &= \sum_{\text{image}} (\mathrm{I}_{pix} - \mathrm{I}_{bg})\, x_p \Big/ \mathrm{I}, \\ \bar{y}_i &= \sum_{\text{image}} (\mathrm{I}_{pix} - \mathrm{I}_{bg})\, y_p \Big/ \mathrm{I}, \\ \bar{x}_u &= \sum_{\text{image}} (x_p) \Big/ \mathrm{N}, \\ \bar{y}_i &= \sum_{\text{image}} (y_p) \Big/ \mathrm{N}. \end{aligned} \tag{2.5}$$

With the unweighted second order moments defined as:

$$\begin{aligned} \mathrm{U}_{xx} &= \sum_{\text{image}} (x_p - \bar{x}_u)^2 \Big/ \mathrm{N}, \\ \mathrm{U}_{yy} &= \sum_{\text{image}} (y_p - \bar{y}_u)^2 \Big/ \mathrm{N}, \\ \mathrm{U}_{xy} &= \sum_{\text{image}} (x_p - \bar{x}_u)(y_p - \bar{y}_u) \Big/ \mathrm{N}, \end{aligned} \tag{2.6}$$

The unweighted semi-major (a_u), semi-minor (b_u) axis and orientation (θ_u) of an ellipse fitted to the pixel distribution can be obtained as:

$$\begin{aligned} \tan 2\theta_u &= 2\mathrm{U}_{xx} / (\mathrm{U}_{xx} - \mathrm{U}_{yy}) \\ a_u^2 &= 2(\mathrm{U}_{xx} + \mathrm{U}_{yy}) + 2\left[(\mathrm{U}_{xx} - \mathrm{U}_{yy})^2 + 4\mathrm{U}_{xy}^2\right]^{\frac{1}{2}} \\ b_u^2 &= 2(\mathrm{U}_{xx} + \mathrm{U}_{yy}) - 2\left[(\mathrm{U}_{xx} - \mathrm{U}_{yy})^2 + 4\mathrm{U}_{xy}^2\right]^{\frac{1}{2}} \end{aligned} \tag{2.7}$$

The intensity weighted ellipse parameters a_i,b_i and θ_i, are constructed from the corresponding intensity weighted moments I_{xx},I_{yy} and I_{xy}. The other independent parameters are the minimum and maximum image coordinates $(x_{min}, x_{max}, y_{min}, y_{max})$ peak intensity (I_{max}) and local background intensity (I_{bg}).

For the purposes of image classification into stars and galaxies (Chapter 3.) and galaxy image photometry (Chapter 4.) extra image parameters are constructed from combinations of the IAM parameters. The geometric ($\mathcal{G}$) and Gaussian width ($\mathcal{S}$) parameters are defined:

$$\begin{aligned} \mathcal{G} &= \frac{\mathcal{A}}{\pi ab} \\ \mathrm{I}_{th} &= \mathrm{I}_{max} \exp\left(\frac{-\mathcal{A}}{2\pi\mathcal{S}^2}\right) \end{aligned} \tag{2.8}$$

and the COSMOS *calibrated* magnitude is constructed such that:

$$m_{cos} = -2{\cdot}5 \log\left(\frac{I_{pix} - I_{bg}}{I_{bg}/A_{pix}}\right) \tag{2.9}$$

where I_{th} is the detection threshold intensity and I_{max} the maximum image intensity.

The remaining parameters in Table 2.4 require other information outwith that contained in the IAM data. Celestial coordinates (RA,DEC) and the position angle of the image major-axis on the sky (θ_{sky}) are calculated from the plate position and orientation using a six-coordinate transformation obtained from known celestial and plate positions of catalogued SAO stars (*Smithsonian Astrophysical Observatory Star Catalog* 1960). Image magnitudes in b_j are constructed from m_{cos} and estimates of the sky background magnitude (m_{sky}), as described in Chapter 4. Field effect corrected magnitudes (m_{cor}) are used for image classification purposes (Section 3.2.1b). The parameter 'F' contains the field number of the plate scanned to obtain the data and *id* is the sequential list number of the galaxy in the catalogue.

2.3.3 LIMITATIONS ON COSMOS MEASUREMENTS

A possible source of contamination in COSMOS datasets comes from noise images. For the purposes of the EDSGC these are defined as any images that are not associated with a real star or galaxy. There are several sources of such images. For bright images noise images are produced (in order of increasing contamination) from localised plate faults (Table 2.1), dust or spurious images formed around bright star haloes. The quality control of ESO/SERC atlas plates ensures that images due to plate faults and dust

in the copying process are minimal. The presence of dust during the measurement process is reduced by cleaning the plates thoroughly before measuring and by housing the COSMOS machine in a clean environment (MGS). It is however inevitable that some dust will be deposited from the atmosphere during measurement, which typically takes several hours. The removal of spurious images around bright star halos, as well as satellite trails is presented in Section 2.3.4.

For faint images the main sources of noise images are grain clumps, emulsion granularity and machine noise. Stobie *et al.* (1984) use power spectrum analysis to investigate the noise characteristics of digitised data. High (spatial) frequency noise components (granularity) can be reduced by digitising using a large aperture, however a maximum aperture size is set by the need to retain information on the smallest images. For UKST plates the maximum practical digitising aperture is 30 μm, with a pixel spacing of 15 μm as dictated by sampling theory (COSMOS scans for the EDSGC use 32 μm aperture with 16 μm spacing). Stobie *et al.* also present a comparison between the noise characteristics of original and copy plates. They conclude that:

> with regard to the reproduction of photometric information [for faint images], which is what power spectrum analysis is describing, there is no significant degradation between the original and copies. Furthermore, as a copy normally has a sky background density lower than the original plate, measuring machine noise is likely to be less important.

The lower background of copy plates also implies that more image information will be accessible within the limited dynamic range of COSMOS density measurements (The maximum measurable density is set by the halo around the CRT spot and is $\sim$ 1·3 above the background density D_{bg}). This maximum sets the level at which images become saturated, ie attain the maximum measurable intensity. The magnitudes at which stars and galaxies saturate are $b_j \simeq$ 19·0 and 14·0 respectively (Section 4.2).

The number of faint noise images in the COSMOS datasets is restricted by adopting a 7% detection threshold (MacGillivray & Dodd 1982) and only using images with $\geq$ 3

pixels. An empirical estimate of the percentage of noise images (of all types) remaining can be made using scans of different plates covering the same area of sky. Repeat copies of grade A survey plates are in short supply, however the fields centred around RA= 0^{hr} have large overlaps ($\sim 3°$) and $\sim 10^4$ images in common. To investigate noise images, COSMOS datasets from the overlapping region of J3578 (F293) and J3592 (F348) were *paired.* Images common to both datasets are considered to be real stellar or galaxy images, while those present in only one dataset will be one of the types of noise images described above. Image pairing was carried out using celestial coordinates which have an rms measurement accuracy of $2''$. Pairing was carried out using a $3''$ pairing tolerance.

Fig. 2.5a & b show the percentage of noise images in each of the datasets as a function of image magnitude. For images brighter than $b_j \simeq 20{\cdot}5$ less than 2% of images are noise. The percentage of noise images only rises rapidly above this level within $0^m{\cdot}5$ of the plate limit, corresponding to the effects of measurement noise. At very bright magnitudes the apparent upturn in noise images is caused by variations in the positions of the centres of the images that are very much larger than the pairing tolerance. Further discussion on the effect of noise images on image classification is presented in Section 3.3.2 and estimates of the percentage of noise images that are classified as galaxies is presented in Section 3.4.3.

The measurement errors for each of the COSMOS IAM parameters are listed in Table 2.4. These were obtained (MGS) by scanning the same plate in two different orientations and so they represent the error associated with the rasterising of data, ie. they estimate binning noise associated with the single measurements of each image and refer to images with $\sim 20 - 30$ pixels ($\sim 2^m{\cdot}0$ above the sky). As part of this work on the EDSGC, the images common to overlapping plates have been used to measure the accuracy of individual parameters by employing the statistic:

$$\frac{\Delta P}{\overline{P}} = \left(\frac{P_1 - P_2}{P_1 + P_2}\right) \times 200 \qquad (2.10)$$

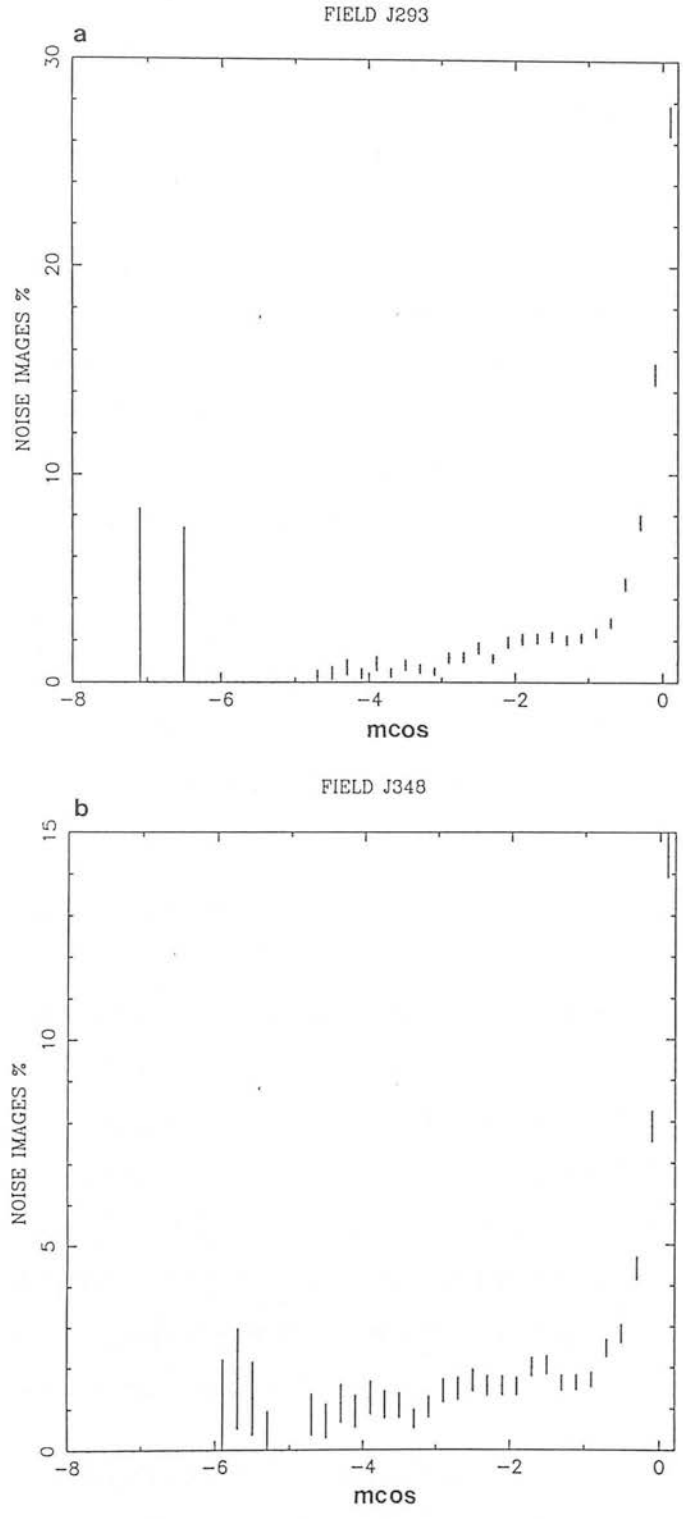


Figure 2.5: The variation in the percentage of noise images in COSMOS datasets as a function of image magnitude. (a) J3578, F293. (b) J3592, F348. m_{cos} is nominally $b_j - 22^{\rm m}.3$.

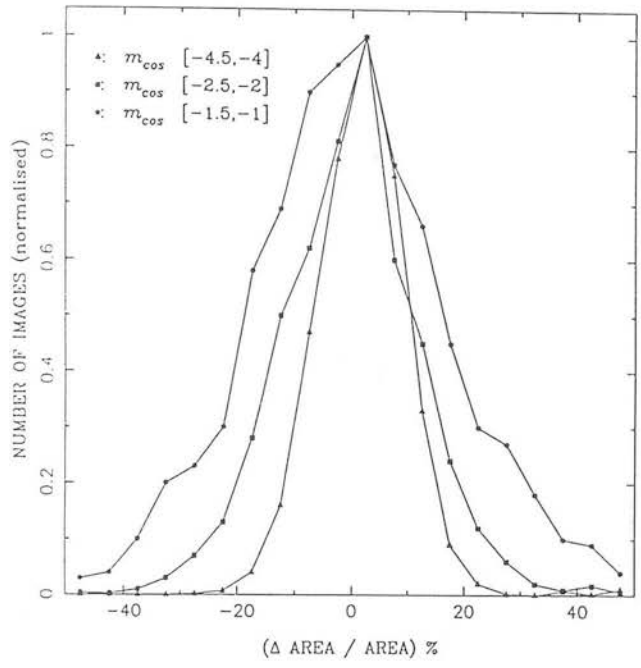
where P_1 and P_2 are the parameter values on the two overlapping plates. This method accounts for differences from plate to plate rather than just the effects of binning noise measured in MGS.

Fig. 2.6 shows the distribution of $\frac{\Delta P}{\overline{P}}$ with $P \equiv \mathcal{A}$ and $P \equiv \mathcal{S}$, for images in various magnitude slices. These distributions are peaked at and symmetric about zero (within a few percent) and so there is no systematic difference between the parameter values on the two plates. The width of the distributions in Fig. 2.6 does increase with fainter image magnitude. This increase in the error on parameter values is expected because the effects of noise become more important for smaller images and is quantified in Fig. 2.7 where the standard deviations (σ_A & σ_S) for the $\frac{\Delta P}{\overline{P}}$ distributions ($P \equiv \mathcal{A}$ and $P \equiv \mathcal{S}$) are plotted as a function of magnitude. At $\sim 2^{m}.0$ above the sky ($\mathcal{A}$ ~20-30 pixels), σ_A is $\sim 15\%$, which corresponds to a difference of 3 pixels between the image area on the two plates (σ_S at $m_{cos} \simeq 2{\cdot}0$ is $\sim 10\%$ which is again equivalent to the addition of ~ 3 pixels to the area).

2.3.4 CROWDED FIELD ANALYSIS

A potential problem for any digitised data arises because of overlapping images. Images from COSMOS TM data consist of connected pixels lying above a single detection threshold. Overlapping images are therefore often assigned to the same set of connected pixels. The resultant images are referred to as blends. At the SGP the average number density of objects ($b_j < 22{\cdot}3$) is ~ 2 arcmin^{-2} (~ 2.5 mm^{-2}). On UKST plates taken around the SGP the number of blended images is found to be approximately constant $\sim 10\%$ over a wide range of magnitudes, though blending is a particular problem for bright stars or for compact galaxy clusters.

The point spread function for stars brighter than $b_j = 16{\cdot}5$ is dominated by diffraction spikes, which substantially increase their area and consequently increase the probability that they are blended. Star-star blends present a particular problem for the EDSGC



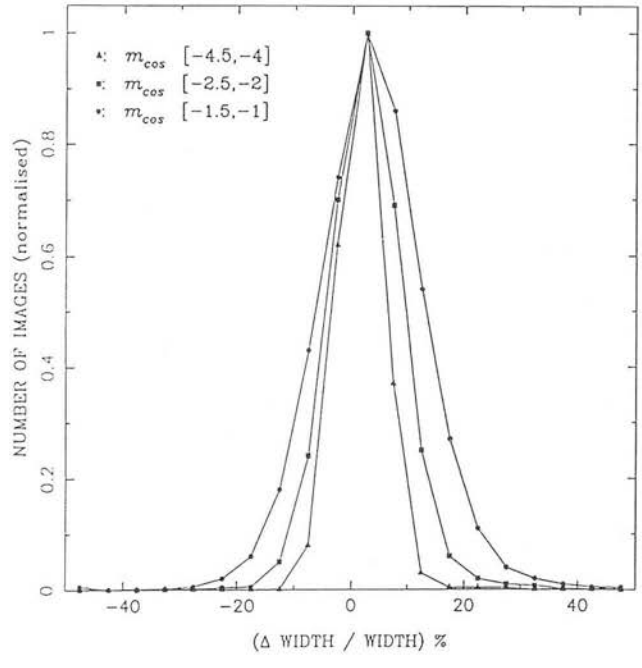


Figure 2.6: The distribution of the statistic $\frac{\Delta P}{P}$ with $P \equiv \mathcal{A}$ and $P \equiv \mathcal{S}$, for images, of various magnitudes, common to two different plates. $m_{cos} \simeq b_j - 22^{\rm m}.3$.

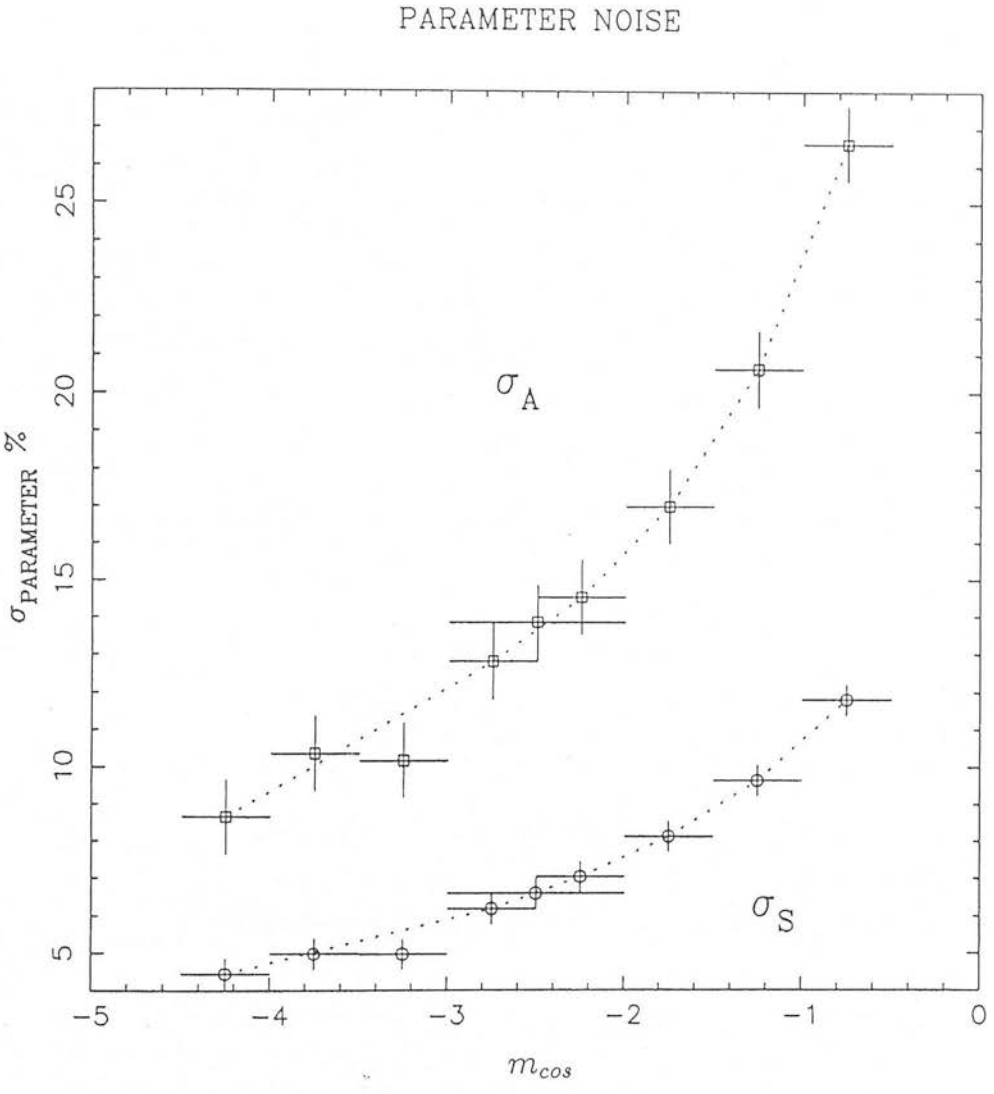


Figure 2.7: The variation with magnitude of the standard deviation (σ_P) of the statistic $\frac{\Delta P}{P}$ for $P \equiv \mathcal{A}$ and $P \equiv \mathcal{S}$. NB: $m_{cos} \simeq b_j - 22^m3$.

as their image parameters resemble those of galaxies because of their lower surface brightness and non-Gaussian profile. Unless detected these images will contaminate the EDSGC. In clusters the images are not only more crowded (surface densities can be up to 10 times higher than average), but the plate background density between objects is relatively high and varies rapidly due to intra-cluster light. This variation is often on scales less than the background filtering scale and so may not be accounted for in the background maps. Galaxies in compact cluster cores are sometimes blended into one large object, particularly in cD clusters. Fig. 2.8a shows a greyscale representation of the pixel configuration of the *single* blended image for such a cluster core. Such images represent a serious environmental incompleteness.

The COSMOS deblending software (Beard *et al.* 1989) has been introduced to remedy the problems caused by image blends. Given a set of connected pixels, the configuration of the pixel distribution is examined at a series of 8 higher detection isophotes. Fig. 2.9 shows a schematic representation of the re-thresholding of a blend image. To account for background variations across the plate these higher detection levels ($\{I_k\}$) are fixed multiples of the original detection level (I_{th}) with

$$I_k = A^k \times I_{th} \quad k = 1, 2....8 \tag{2.11}$$

The areal parameter (A) can be tuned depending on the types of images that prevail in the data. Empirically a value of 1.07, which spaces the levels close to the original threshold, optimises the deblending procedure for faint galaxies. A larger spacing will optimise the software for brighter images, eg. bright star blends.

At each higher isophote the image analyser connects pixels using the method described for a single detection threshold in Section 2.3.1. If a component image becomes resolved from the blend at a higher threshold then its pixels are no-longer connected to the original image. Each new image resolved from the blend is called a *daughter* image. Once all the daughter images are found, the pixels belonging to the original blend are assigned amongst them. Gaussian areal profiles are fitted to the pixels of each daughter image

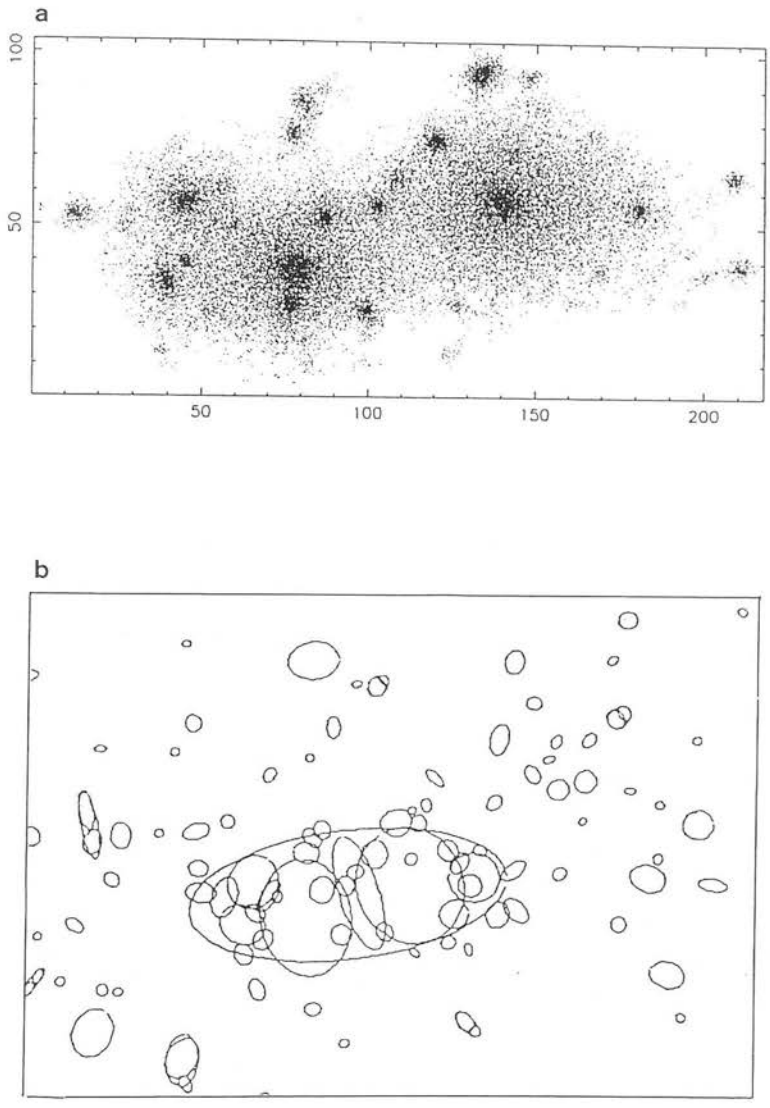


Figure 2.8: Deblended images in compact clusters. (a) A greyscale representation of the pixel distribution in the *single* blended image of a compact cluster core. (b) An ellipse plot from the IAM parameters of the resultant *images* after deblending.

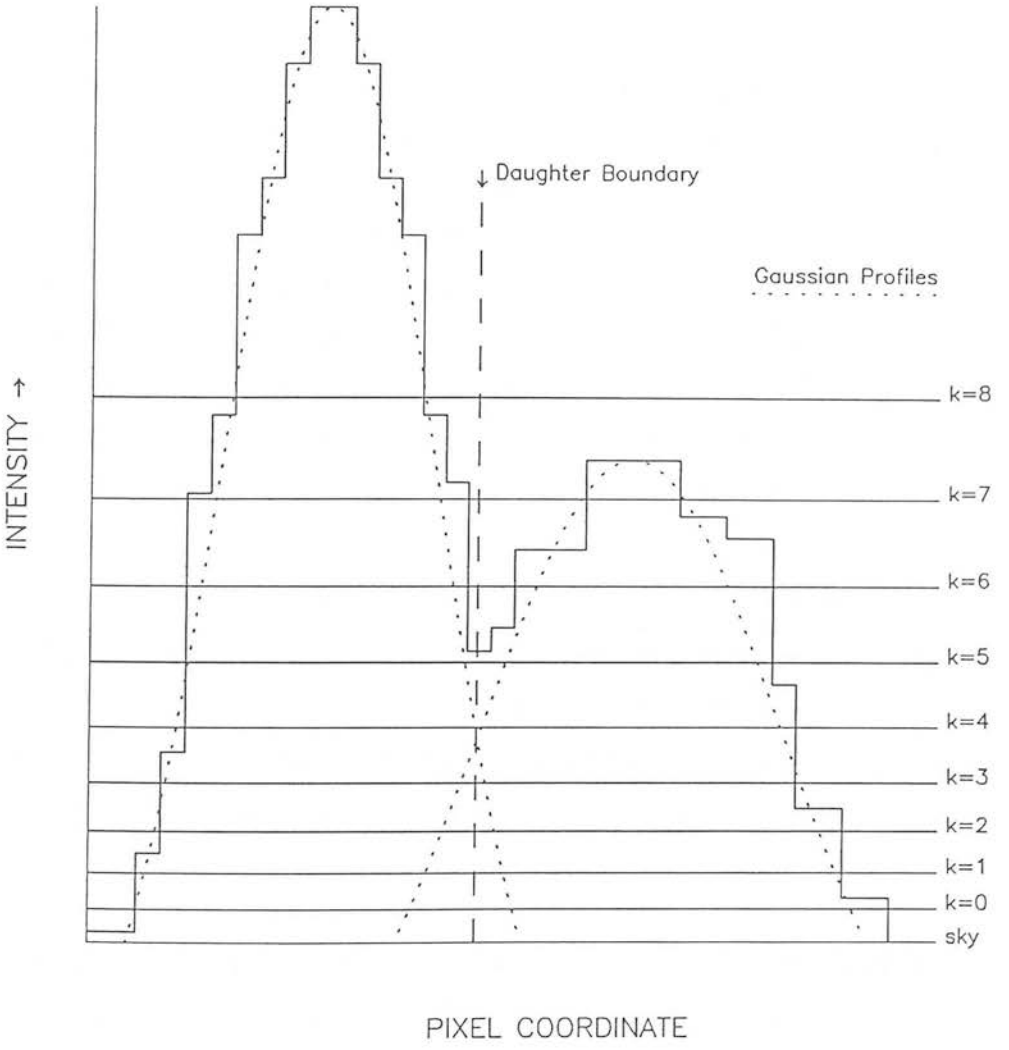


Figure 2.9: A schematic diagram showing the application of COSMOS deblending software to a merged image

isolated from the blend. At lower thresholds, the loci of intersection of profiles define the boundaries between daughter images. IAM parameters are constructed from the pixel distribution within these boundaries, with the original pixel intensities. Daughter images do not have Gaussian profiles. Fig. 2.8b shows a representation of the IAM data for the daughter images deblended from the cluster image in Fig. 2.8a. Most of these images are associated with galaxies, those that are spurious have low surface brightness.

The improved completeness of COSMOS datasets produced with the deblending software has been quantified by Hatzidimitriou (1989). COSMOS scans of UKST plates centred around the crowded fields of the Small Magellanic Cloud (SMC) were compared with larger scale CCD images. Fig. 2.10 shows the variation in completeness of COSMOS datasets as a function of local surface density. Completeness is defined to be the number of COSMOS images as a percentage of the number of images found on the CCD (to the same magnitude limit). At the average surface density of images on UKST plates COSMOS datasets are complete. The deblended datasets are $> 80\%$ complete to surface densities of ~ 20 arcmin^{-2}, compared to $\sim 50\%$ for the undeblended dataset. Also shown in Fig. 2.10 is the percentage of images that are still blends after application of the deblending software. This is effectively the percentage of images that are too close to be resolved at this plate scale and is an inherent incompleteness due to using UKST plates.

The deblending software is applied to all COSMOS datasets used in the EDSGC. The SMC data consists of faint stars and as such the structure of each image is the same. In the raw data used for the EDSGC the images have far more varied structure, not only because the data includes stars and galaxies, but also because image structure varies as a function of magnitude. To examine the effects of the deblending software on more varied data, the actual images on the plate corresponding to the daughter images from several trial plate scans were examined visually. Table 2.5 shows the percentage of daughter images classified according to the morphology of the parent blend image. The data is divided into daughter images brighter than $m_{cos} = -5{\cdot}5$, for which stellar images have

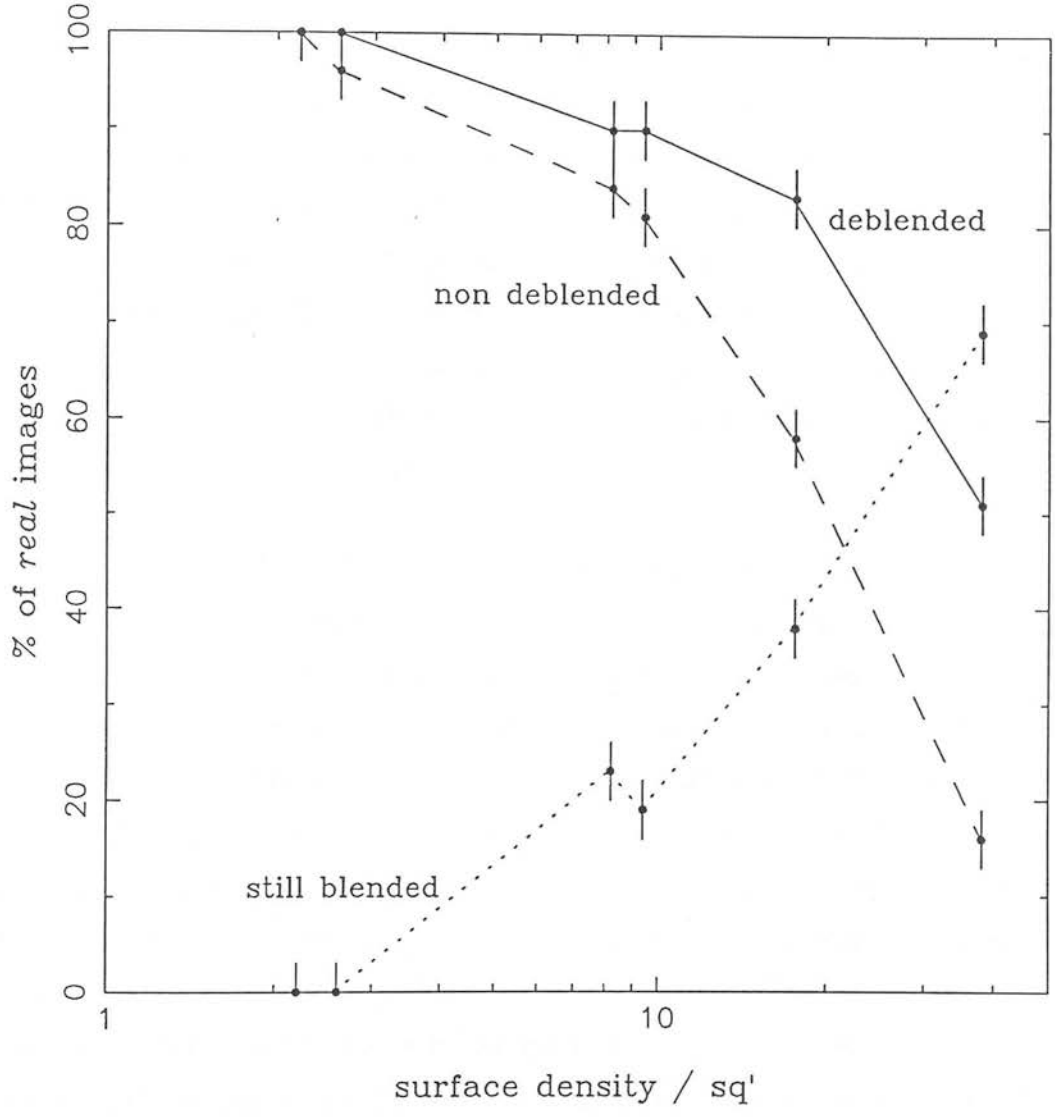


Figure 2.10: The completeness of COSMOS datasets as a function of local surface density, together with variation in contamination of images still merged after deblending. The error bars are $N^{\frac{1}{2}}$.

diffraction spikes, and those fainter than $m_{cos} = -5{\cdot}5$. For the brighter set, the majority of blends are star-star merges as expected because of the increased area of stellar images (diffraction spikes). For the fainter set, the majority of daughter images are equally distributed over star-star, star-galaxy and galaxy-galaxy parent images. There is however a high contamination ($\sim 25\%$) of spurious images produced from the breakup of bright star diffraction spikes, ghosts (concentric rings around bright stars caused by internal reflections in the telescope filter and corrector) and halos (light scattered by the emulsion). The variation in intensity across these images mimics that across faint overlapping images. Because the areal profile parameter, Eqn. 2.11, is optimised to deblend faint galaxies, such images are deblended.Spurious images produced by deblending would represent a serious contamination of the EDSGC as they tend to have low surface brightness and non Gaussian profiles.

Noise images associated with the break up of stellar halos (stars with $b_j \lesssim 15{\cdot}5$) are present in COSMOS datasets regardless of whether the deblending software has been applied. All such images within the halo radius (6·7 arcmin) of bright stars are removed from the EDSGC as part of the *drilling* procedure (Section 5.2). The break up of filter ghosts and diffraction spikes is mainly due to the application of the deblending software, so the Spurious Image Search Algorithm[2] (Table 2.6) has been developed to detect such contaminants. This algorithm searches for and finds virtually all of the daughter images produced by the spurious deblending of diffraction spikes and ghost images. These images are removed from the raw data and combined back into their parent image. Large images with $> 60,000$ pixels are also not deblended to prevent the spurious deblending of nearby galaxies and highly elongated images (parents and daughters with $\frac{b_u}{a_u} > 0{\cdot}05$) are removed at this stage as they are usually spurious, eg. satellite trails.

Given that spurious daughter images are removed from the catalogue the final completeness and contamination of the EDSGC will depend on how the IAM parameters

[2]See Section 3.3.1a for details of the $\mathcal{G}$ image classifier

morphology of blend	% of daughters $m_{cos} \lesssim -5{\cdot}5$	% of daughters $m_{cos} \gtrsim -5{\cdot}5$
star-galaxy	18%	24%
star-star	58%	19%
galaxy-galaxy	10%	26%
diffraction spike, ghosts, halos.	0%	24%
single star	9%	4%
single galaxy	4%	2%
noise (no object)	1%	1%

Table 2.5: The percentage of daughter images produced from various parent image morphologies, identified during visual inspection. Data is presented for daughters brighter and fainter than $m_{cos} = -5{\cdot}5$

Spurious Image Search Algorithm

IF parent or daughter has $\frac{b_u}{a_u} < 0{\cdot}05$ **THEN** remove from the dataset.

Image is spurious, usually associated with diffraction spikes and satellite trails

IF parent is larger than 60,000 pixels **THEN** do not deblend.

Image is a nearby galaxy and should not be deblended.

IF parent has $m_{cos} < -5{\cdot}5$ and $\mathcal{G} \leq 0{\cdot}9$ **THEN** do not deblend.

Image is bright star and should not be deblended.

IF parent has $m_{cos} < -5{\cdot}5$ and $\mathcal{G} \geq 0{\cdot}9$ **THEN:**

IF brightest daughter has $m_{cos} > -5{\cdot}5$ **THEN** use daughter images.

Parent image is a sum of faint images and no spurious images are images expected.

IF brightest daughter has $m_{cos} < -5{\cdot}5$ and $\mathcal{G} \geq 0{\cdot}9$ **THEN** use daughter images.

Parent image is a compact cluster core, daughters are galaxies.

IF brightest daughter has $m_{cos} < -5{\cdot}5$ and $\mathcal{G} \leq 0{\cdot}9$, **THEN** add the following spurious daughters back into the parent image:

a. Daughters within 600 μm of unweighted centroid.

b. Daughters with θ_u within 3°of a multiple of 90°.

Parent image is a bright star or bright star blend. Spurious filter ghost images (a) *and diffraction spikes* (b) *must be removed.*

Table 2.6: The EDSGC algorithm for the removal of spuriously deblended images. The $\mathcal{G}$ parameter is $< 0{\cdot}9$ for stars, $> 0{\cdot}9$ for galaxies. See Section 3.3.1a.

for deblended images compare to those the image would have if isolated. Hazidimitriou (1988) reports that if the overlapping objects are of comparable magnitude ($\Delta m \lesssim 1{\cdot}5$) then the result daughter image magnitudes show no systematic error. For overlapping with larger magnitude differences the daughter image, corresponding to the fainter object, is systematically brighter ($\sim 0^{\rm m}{\cdot}5$ brighter for $\Delta m = 4$). This is expected from the method of deblending, but because of the small number of images involved the effect is negligible. A discussion about the accuracy of IAM parameters of deblended images in terms of the automated image classification procedure is presented in Section 3.4.1 The classification statistics of daughter images are compared with those for isolated images and with statistics from visual classification. For classification purposes there is only a marginal difference between the IAM parameters of a daughter and those it would have if isolated.

2.4 Summary

The aim of this chapter was to describe the method of production of the raw plate data, *ie* the lists of images and image parameters, from the COSMOS scans of the *ESO/SERC Atlas* glass copy plates.

Careful quality control and processing techniques are used in the production of the SERC J survey and the *ESO/SERC Atlas* copy plates, which keep the number of plate flaws and variations in homogeneity to a minimum. The major sources of inhomogeneity across a plate are vignetting and de-hypersensitisation, which act as sensitivity variations. The latter of these effects is overcome by nitrogen flushing of the plate during exposure, but a large part of the survey was taken prior to this being introduced. The effects of sensitivity variations on image classification and image photometry will be dealt with in subsequent chapters.

COSMOS objectively finds $\sim 2\times10^5$ images on each photographic plate, with an image

defined as a set of connected pixels lying above a $\sim 25^{\mathrm{m}}.0$ per arcsec isophote. The detection threshold is set as a percentage above the local background intensity to account for the variations introduced by vignetting and de-hypersensitisation. Shape parameters are derived for each image from the weighted (with intensity) and unweighted moments of pixel distribution. Noise images caused by emulsion granularity and grain clumps are shown to produce $< 2\%$ contamination in the raw ($b_j < 20$) image datasets.

The COSMOS deblending software produces a significant increase in the completeness of images detected in crowded regions, over that produced by the normal image analyser. An algorithm has been implemented to detect spurious images, produced by the deblending of stellar filter ghosts and diffraction spikes, and to add them back into the original parent image. The use of an accurate method of image deblending is vital in the production of large galaxy catalogues and any subsequent cluster catalogues derived from them. Failure to apply this software would result in an incomplete cluster catalogue with compact clusters having spurious properties. The EDSGC will therefore provide an excellent basis to study the large-scale distribution of clusters.

Chapter 3

Image Classification

3.1 Introduction

In general star–galaxy image classification of digitised data from photographic plates can be performed because the two populations occupy different regions of some classification parameter versus magnitude scatter diagram (*classification plane*). Table 3.1 lists previous papers on image classification from photographic plates, together with a brief description of the method employed and the completeness and contamination statistics where available.

The need to ensure homogeneity when classifying images over large regions and many plates was discussed in Section 1.4. To extract the maximum information from a catalogue of this type it is also important to ensure the homogeneity as a function of magnitude. This is especially true for accurate galaxy number–magnitude counts. The purpose of this chapter is to discuss the various systematic errors associated with image classification from COSMOS datasets and to present the techniques employed to reduce them to within the limits set by Geller *et al.* (1984). It should be stressed that the techniques described here are specific to the raw data employed in the EDSGC. They are not generic and may not be applicable to other digitised data produced by different measuring machines from different plate material.

For the purposes of the EDSGC, three basic classification parameters are used to discriminate between stars and galaxies, together with a COSMOS image magnitude. The parameters used, $\mathcal{G}$, $\mathcal{A}$ and $\mathcal{S}$, defined in Section 2.3.2, are combinations of the zeroth and second order moments of the pixel distribution. These contain most of the useful

AUTHOR	DATA	CLASSIFICATION METHOD	*cmp*	*ctm*
MacGillivray *et al.* 1976	COSMOS IIIa-J UKST	Transmission *vs* area.	$\sim 90\%$	5–10%
Tyson & Jarvis. 1979	PDS IIIa-J KPNO	6 parameter classification using training samples.	n/a	n/a
Carter *et al.* 1979	PDS IIIa-J AAT	Profile fitting.	65–95%	3–10%
Peterson *et al.* 1979	PDS IIIa-J AAT	Gradient of profile *cf* visual classification	n/a	n/a
Shanks *et al.* 1980	COSMOS IIIa-J UKST	Transmission *vs* area & correction for seeing.	$\sim$90 %	>10%
Kron. 1980	PDS IIIa-J KPNO	R_2 statistic from image profile.	n/a	n/a
Butchins 1981	PDS IIIa-J PS	Profile fit to point spread func.	>90%	7–16%
Valdes 1982	PDS IIIa-J KPNO	Bayesian classification as resolved & unresolved	n/a	n/a
Hewett. 1983	COSMOS IIIa-J UKST	Area & peak surface brightness *vs* magnitude.	97%	47%
Shanks *et al.* 1984	COSMOS IIIa-J UKST	Area & Gaussian width *vs* magnitude.	n/a	n/a
Kurtz *et al.* 1985	PDS POSS	Analysis of moments & visual classification.	n/a	n/a
Oegerle *et al.* 1986	PDS 098-04 PS	6 parameter classification & Bayesian decision theory.	95%	n/a
Sebok. 1986	Caltech scanner IIIa-f PS	Sebok 'Phi' classifier bright objects by eye.	70%	20–50%
Dickey *et al.* 1987	APS POSS	Image width at various threshold levels.	70%	25%
Yamagata. 1987	PDS Kiso Schmidt	Principal component analysis.	n/a	n/a
Maddox *et al.* 1988	APM IIIa-J UKST	Surface brightness & profiles, bright images (< 17) by eye	$\sim 90\%$	n/a
Parker *et al.* 1988	COSMOS Prism plates	Comparison of prism plate and direct plate image magnitude	n/a	n/a
This work 1988	COSMOS IIIa-J UKST	Field effect correction & optimum parameter ranges	>95%	$\sim 10\%$

Table 3.1: Previous papers on image classification, together with completeness (*cmp*) and contamination (*ctm*). The following abbreviations are used: UKST, UK Schmidt Telescope. AAT, Anglo-Australian Telescope. KPNO, Kitt Peak National Observatory 4m. POSS, Palomar Observatory Sky Survey. PS, Palomar Schmidt. APS, Automatic Plate Scanner (Minnesota). APM, Automatic Plate Measuring (Cambridge).

structural information about each image compared with higher order parameters which suffer more from machine noise and consequently have show less *separation* between the two populations.

Fig. 3.1 illustrates the distribution of objects in each of the three classification planes. The magnitude parameter, m_{cor} shown in these figures is the *field effect corrected* version of m_{cos} (Section 3.2.1b). Because of the variation of image structure recorded on the plate for both stars and galaxies, the various parameters are used over different magnitude ranges:

(i) For relatively bright images the geometric parameter ($\mathcal{G}$, Eqn. 2.8) is used. This parameter measures how effectively an image fills the ellipse fitted to its unweighted major and minor axes (a_u, b_u). For stars brighter than $b_j \simeq 15{\cdot}5$ diffraction spikes increasingly dominate the fitted ellipse ($\pi a_u b_u$), so that $\mathcal{G} < 0{\cdot}9$, while for galaxies the image fills the fitted ellipse and so $\mathcal{G} \simeq 1{\cdot}0$.

(ii) For intermediate magnitudes, galaxy images have a lower surface brightness than the point spread function of stellar images, so that in the image area versus magnitude plane, galaxies lie above a tight stellar locus. In practice classification is carried out using $\log_{10}(\mathcal{A})$ due to the large range in image areas (3 to 10^3 pixels).

(iii) For fainter images the $\mathcal{S}$ parameter provides a means of classification. This is defined (Eqn. 2.8) as the width of a Gaussian fit to I_{max} and the image area at I_{th}. For stars the Gaussian width is set by the point spread function and, as with $\log_{10}(\mathcal{A})$ classification, stars occupy a tight locus below galaxy images.

The layout of this chapter is as follows. Section 3.2 is concerned with the causes and reduction of variations in image classification efficiency that are dependent on the image position in the survey. Such variations can occur across an individual plate or from plate-to-plate and are therefore treated separately. The techniques used to reduce magnitude dependent variations are presented in Section 3.3, together with a discussion of

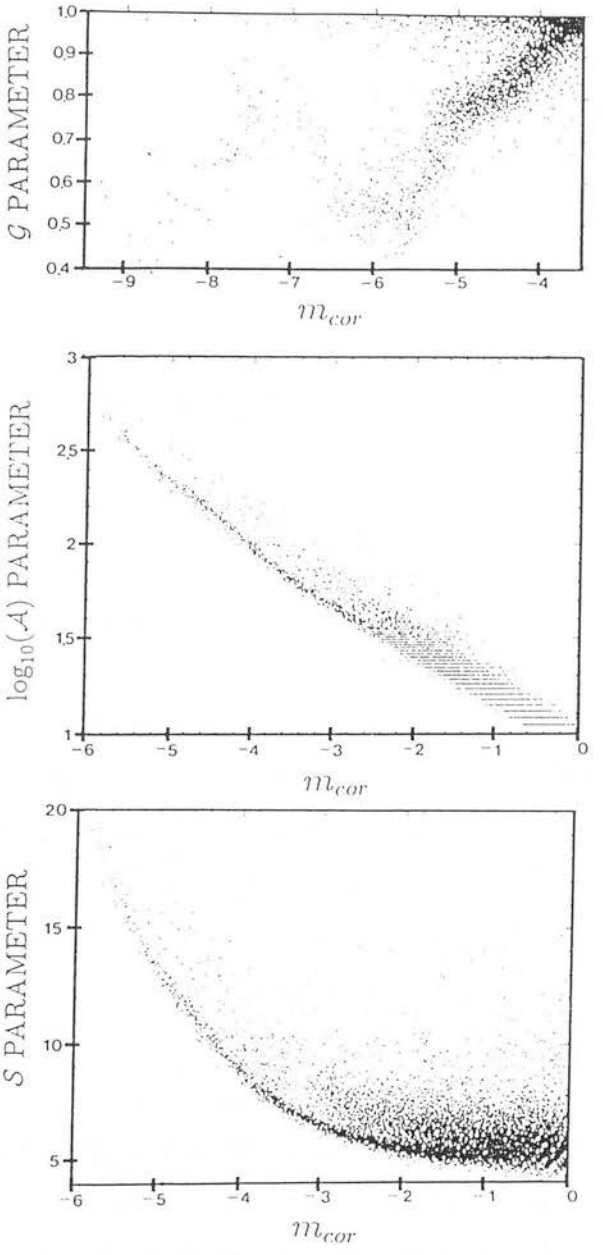


Figure 3.1: The distribution of images in the three main classification planes: (a) $\mathcal{G}$ verses m_{cor}. (b) $\log_{10}(\mathcal{A})$ verses m_{cor}. (c) $\mathcal{S}$ verses m_{cor}

the factors that set the limiting magnitude of the EDSGC. In Section 3.4 the resultant completeness, contamination and residual variations in classification efficiency are estimated using a series of tests on sample datasets from the EDSGC after classification using the algorithms presented in the previous sections. The classification of deblended and isolated images is also compared to estimate the accuracy of the algorithm in compact cluster cores. Finally possible sources of higher misclassification rates over localised areas of the survey are detailed in Section 3.5.

3.2 Reduction of position-dependent misclassification

3.2.1 THE VARIATION ACROSS INDIVIDUAL PLATES

3.2.1a *Field effects and image classification*

The major aspect of the SERC (J) survey which gives rise to plate position dependent misclassification on individual plates is the variation in the plate background density. This in turn leads to a variation in the measured properties of saturated point images as a function of their position on the plate, referred to as field effects. There are two primary causes of plate background variation across a UKST plate. The first cause is vignetting in the telescope optics and the second is differential desensitisation due to varying amounts of moist air in the curved plate holder (Dawe *et al.* 1984). Both effects, which are discussed in detail in Section 2.2.1, result in a radial decrease in the plate background density from the centre outwards. The latter of these two effects is worse for those plates taken before the introduction of routine nitrogen flushing (*ie.* J1541 to J8271). Unfortunately this is the case for most plates accepted for the *ESO/SERC Atlas*.

These background variations may change the characteristic image loci in the various classification planes and so affect image classification. The problem arises because the

relationships between the measured image magnitude (m_{cos}) and the actual object magnitude (b_j) depends on whether images are saturated on not and because the measured magnitude for saturated images varies as a function of position.

For most plates the intrinsic sky background variations such as zodiacal light, plate fogging and stray light are all negligible compared with the effects of vignetting and desensitisation (Section 2.2.1 and Section 2.3.1) . As a result of this the density variations evident on plates are equivalent to changes in emulsion sensitivity (Campbell 1982,Wesselink 1984). Under these circumstances the COSMOS magnitude m_{cos} is, in principle, independent of image position on the plate.

For unsaturated images and assuming that the variations in the background intensity are equivalent to sensitivity changes then, from Eqn. 2.9, m_{cos} is related to the image magnitude (b_j) by:

$$b_j = m_{cos} + m_{sky}, \tag{3.1}$$

where $m_{sky} = -2.5\log_{10}(\mathrm{I}_{sky}/A_{pix})$ and I_{sky} is the intrinsic sky background intensity. For the SERC (J) survey m_{sky} is typically $22^{\mathrm{m}}\!.3$ arcsec^{-2}. Fig. 3.2 shows total galaxy magnitudes and aperture star magnitudes, from the published photometric sequences of Kirshner *et al.* (1978) and Hawkins (1981), plotted against the image m_{cos}. The data plotted is colour corrected (Blair and Gilmore 1982) and is distributed over the whole of F411. The 45° slope fitted to the galaxy data, shows the validity of Eqn. 3.1 over the whole region of a scanned plate and over a wide range of magnitudes ($b_j \simeq 14{\cdot}0$ to $21{\cdot}0$). This is in agreement with the work of MacGillivray & Dodd (1983). COSMOS galaxy photometry is compared with CCD magnitudes in detail in Section 4.3.3.

Eqn. 3.1 does not apply for saturated images. A saturated image is defined as one for which the majority of pixels are limited to some constant density value. Fig. 3.2 shows that m_{cos} underestimates the image magnitude for stellar images brighter than $b_j \simeq 19{\cdot}5$, *ie.* they are saturated if brighter than this magnitude. Further to this the peak image intensity that is measurable is effectively constant across the plate, and

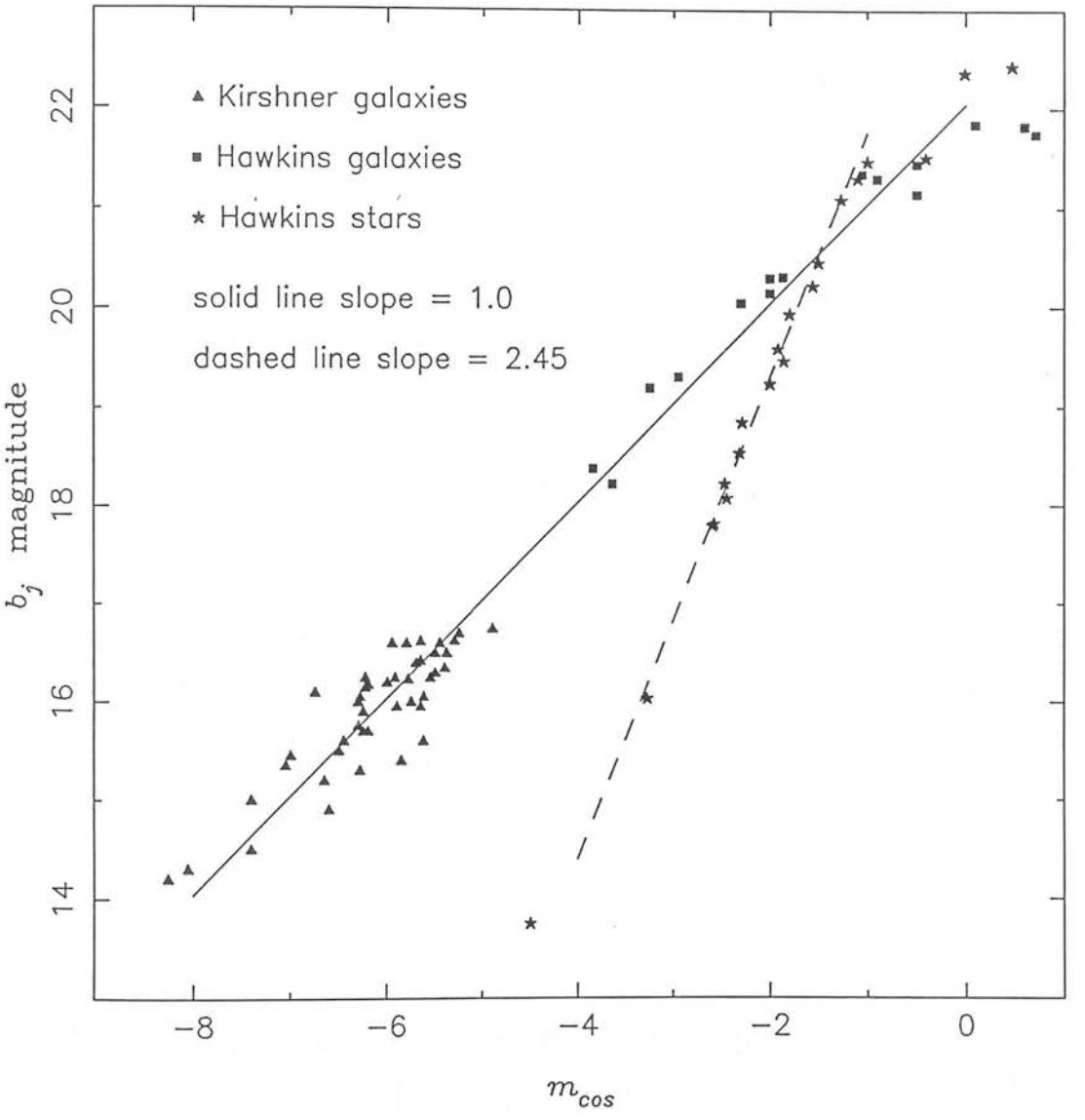


Figure 3.2: The relation between m_{cos} and b_j for stars and galaxies. The Hawkins (1981) data is electronographic, while the Kirshner *et al.* (1979) data is a combination of photographic and photoelectric measurements.

so from Eqn. 2.9 it can be seen that saturated images, of the same area, will have measured magnitudes:

$$m_{cos} = \text{CONSTANT} - m_{bg} \tag{3.2}$$

where $m_{bg} = -2{\cdot}5 \log_{10}(\mathrm{I}_{bg}/A_{pix})$.

Saturated images therefore have a value for m_{cos} that is dependent on I_{bg}, and which varies with position on the plate. Dawe & Metcalfe (1982) estimate that vignetting and desensitisation cause background density changes equivalent to $0^{\mathrm{m}}{\cdot}2$ between the plate centre and the plate edge. COSMOS scans confirm this as a typical value, though there are more extreme cases, that may be connected with the developing or copying process; the worst known case having a variation of $1^{\mathrm{m}}{\cdot}0$. Data from a small region of a typical plate indicates that the intrinsic width of the stellar locus in the classification plane is $< 0^{\mathrm{m}}{\cdot}2$. It follows that the composite stellar locus from distinct regions of a plate is usually wide enough to remove the separation between the star and galaxy populations for m_{cos} brighter than -3. Fig. 3.3a shows data from two different regions of a typical plate plotted in the $\log_{10}(\mathcal{A})$ versus m_{cos} plane. The stellar locus for data from a $0{\cdot}5^{\circ}$ strip at the western edge of the plate lies well to the left of data from a corresponding strip in the centre of a plate. Without field effect correction, uniform classification across the whole area of most SERC (J) survey plates is clearly impossible.

3.2.1b *Field effect correction*

In order to correct the magnitudes of images for field effects some independent estimate of m_{cos} for saturated images is required. One image parameter which is relatively independent of background variations is the image area ($\mathcal{A}$). This is defined by the COSMOS pattern analyser to be an isophotal area using a threshold level that is typically $\simeq 8\%$ above the sky level and so well below the COSMOS saturation level. This image property may therefore be used to provide an empirical correction for field effects to saturated m_{cos} values. Wallace & Tritton (1979) have investigated the change

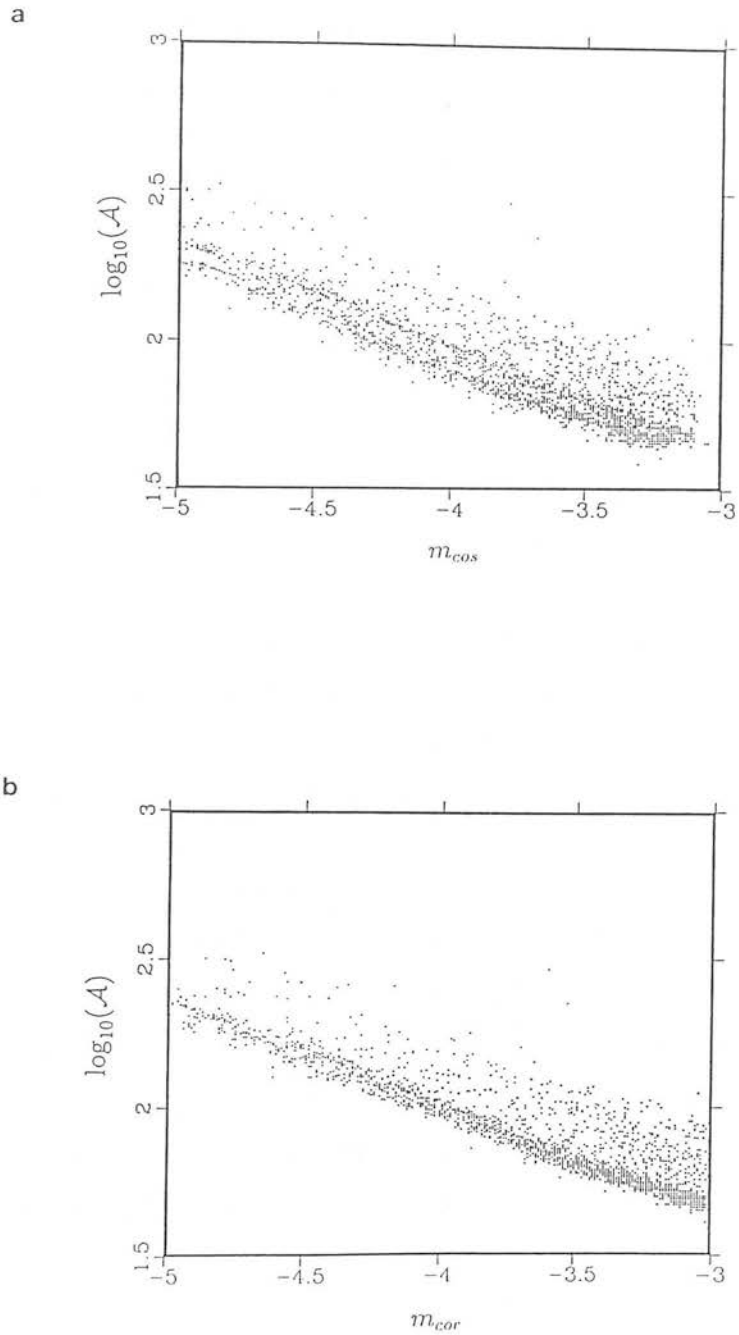


Figure 3.3: Field effects for SERC (J) plates: (a) Data from 0·5° strips at the western edge and centre of the plate, plotted in the $\log_{10}(\mathcal{A})$ verses m_{cos} plane. (b) The same data after field effect correction, *ie.* plotted in the $\log_{10}(\mathcal{A})$ verses m_{cor} plane.

in image area across SERC (J) survey plates, due to effects such as field rotation. Their results suggest these effects are negligible for classification purposes on survey plates.

The variation of the stellar locus in the $\log_{10}(\mathcal{A})$ versus m_{cos} plane is examined by dividing the plate into 10×10 cells and comparing the locus for each cell with that for data from a cell near the plate centre. In this way it is possible to define an array of magnitude correction factors that, as a function of image area, correct the value of m_{cos} for a saturated image to the value an image (of the same area) would have in the centre cell. To obtain a smooth variation in correction factors across the plate, the correction factor for a given image is obtained from bi-linear interpolation between the cell it occupies and adjacent cells. Corrected magnitudes are referred to as m_{cor}.

Fig. 3.3b shows the identical images as in Fig. 3.3b except that the corrected magnitudes m_{cor} have been used in place of the m_{cos} values. Only after field effect correction has been performed is it possible to define a single discrimination boundary for all images regardless of their position on the plate.

For the purposes of star–galaxy classification it is sufficient to apply the correction factors to all images. Since magnitudes are corrected to the plate centre, where the background is highest (see Fig. 2.4), corrections applied to unsaturated images move them away from the resultant stellar locus. For photometric purposes field effect corrected magnitudes are only applicable to saturated images. The uncorrected magnitudes are used for photometric calibration of galaxies, which, as shown above, are unsaturated to magnitudes at least as bright as $b_j = 15$.

3.2.2 THE VARIATION BETWEEN PLATES

3.2.2a *Interactive classification*

One key aspect to reducing misclassification variations between plates lies in the consistency with which a discrimination boundary between regions occupied by stars and galaxies can be determined. Several of the methods described in Table 3.1, used previously to classify images, rely on subjective fitting of a separation line between the different image populations in the classification plane. This line is usually a spline fit to a series of *separation points* chosen by eye.

Using the data available the stability of such subjective methods of classification has been examined, in terms of the completeness and contamination of the resulting galaxy catalogue. In these trials classification on each field was performed by plotting small subsets of the total number of images, in the $\log_{10}(\mathcal{A})$ versus m_{cor} plane, and fitting a separation locus. These subsets were either images from small regions of a plate or random samples of the data from whole plates. Separation points were defined using a cursor on an interactive graphics device. A spline fit was made to these points and then checked for other subsets or regions of the plate. Any necessary iterations were then made to achieve a 'best fit' separation locus. The procedure was repeated in the $\mathcal{S}$ versus m_{cor} plane for fainter images where there was little or no apparent separation in the $\log_{10}(\mathcal{A})$ classification plane. Further iteration of both separation loci was often required to remove discontinuities in the number counts at the changeover magnitude.

When comparing the results of this interactive separation method with visual classifications of images on plates from various epochs, it was obvious that, even after some iteration, the stellar contamination in the galaxy catalogue varied by as much as 20% from field to field. In addition, independent classifications of the same field by different observers often gave rise to variations $> 10\%$. This method will obviously introduce subjective classification errors which are unacceptable in the context of the limit set by

Geller *et al.*.

3.2.2b *An algorithm for automated classification*

A totally subjective definition of a separation locus can introduce rapidly varying contamination into a galaxy catalogue and so an automated procedure is used in the EDSGC to consistently and objectively define such a separation locus.

The first stage of the algorithm involves binning the data in the classification parameter versus magnitude plane. Binned distributions are shown, at various slices of m_{cor}, for the $\log_{10}(\mathcal{A})$ parameter in Fig. 3.4 and for the $\mathcal{S}$ parameter in Fig. 3.5. The distribution of images along each slice of constant magnitude is used to define a point of separation between the galaxy and stellar populations. This point is found by first calculating the mode of the local object distribution about the maximum bin. Using only objects near the mode reduces the effects of binning noise and prevents the estimate of the mode from being biased by the skew of the complete distribution. The position of a given percentile in the pure stellar distribution is then identified and used to define the separation point by reflecting about the mode of the stellar distribution. The procedure is repeated for each magnitude bin over the full magnitude range for which classification is required. A smooth separation locus is obtained by automatically fitting a cubic spline to the set of separation points (knots are set to be equally spaced at $\sim 1^{\mathrm{m}}.0$ intervals).

The main advantages of this algorithm are: (i) The only assumptions required are that the distribution of stars is compact and symmetric and that all images below the peak in the distribution are stars. No other assumptions about the shape or width of the stellar distribution are made. (ii) The algorithm provides stellar contamination estimates in the resulting EDSGC, both as a function of magnitude and for the total number of images classified (Appendix A). (iii) The separation percentile value is preset to apply the same separation criterion and thus ensure the same fraction of stars is

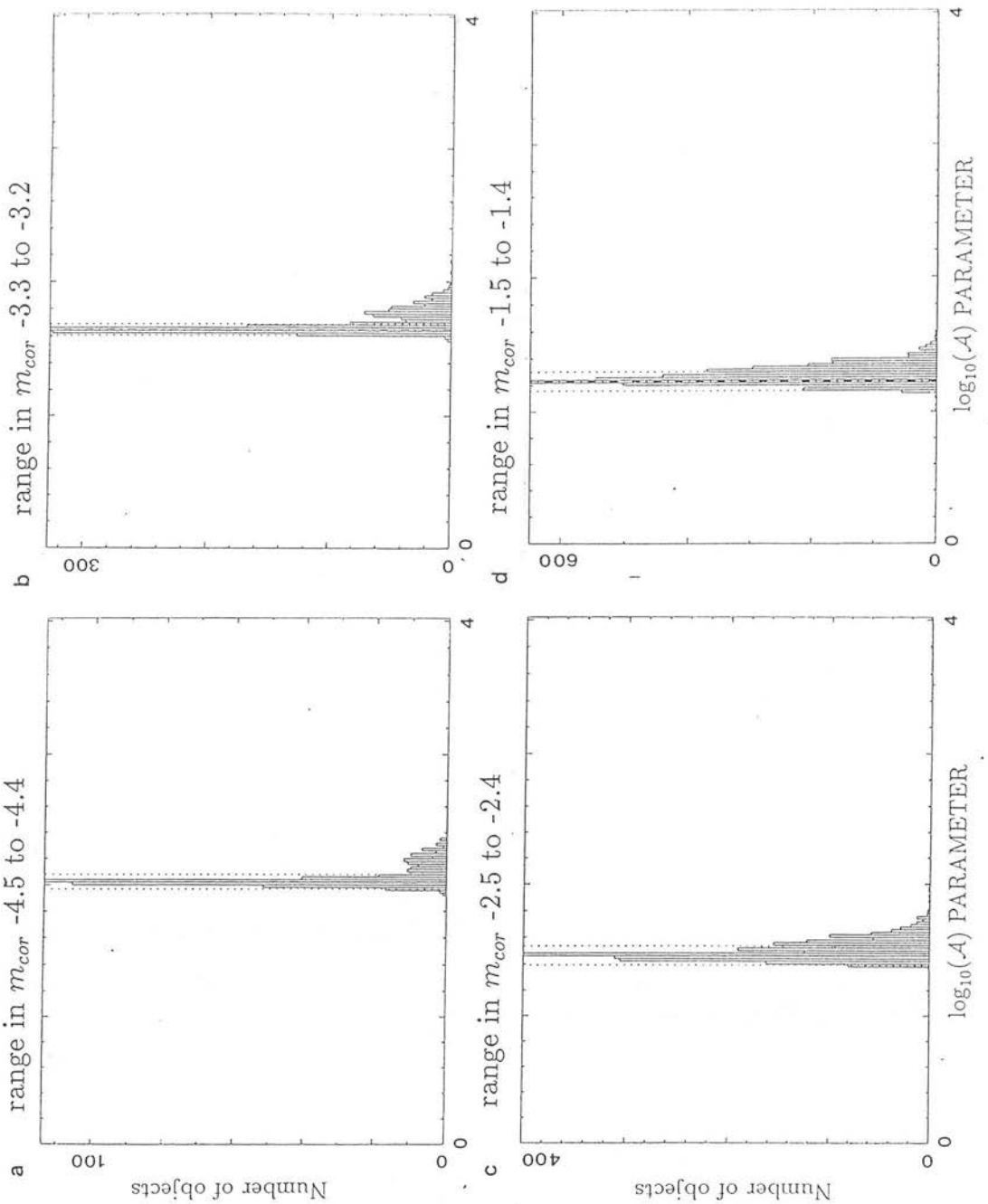


Figure 3.4: Binned distributions in the $\log_{10}(\mathcal{A})$ classification plane along slices of constant m_{cor}. The dotted line to the right of the peak represents the separation point: (a) Here some of the galaxy distribution lies below the stellar population due to diffraction spikes. (b) There is good separation at these magnitudes. (c) There is limited separation due to classifier breakdown and quantisation. (d) The two populations have merged.

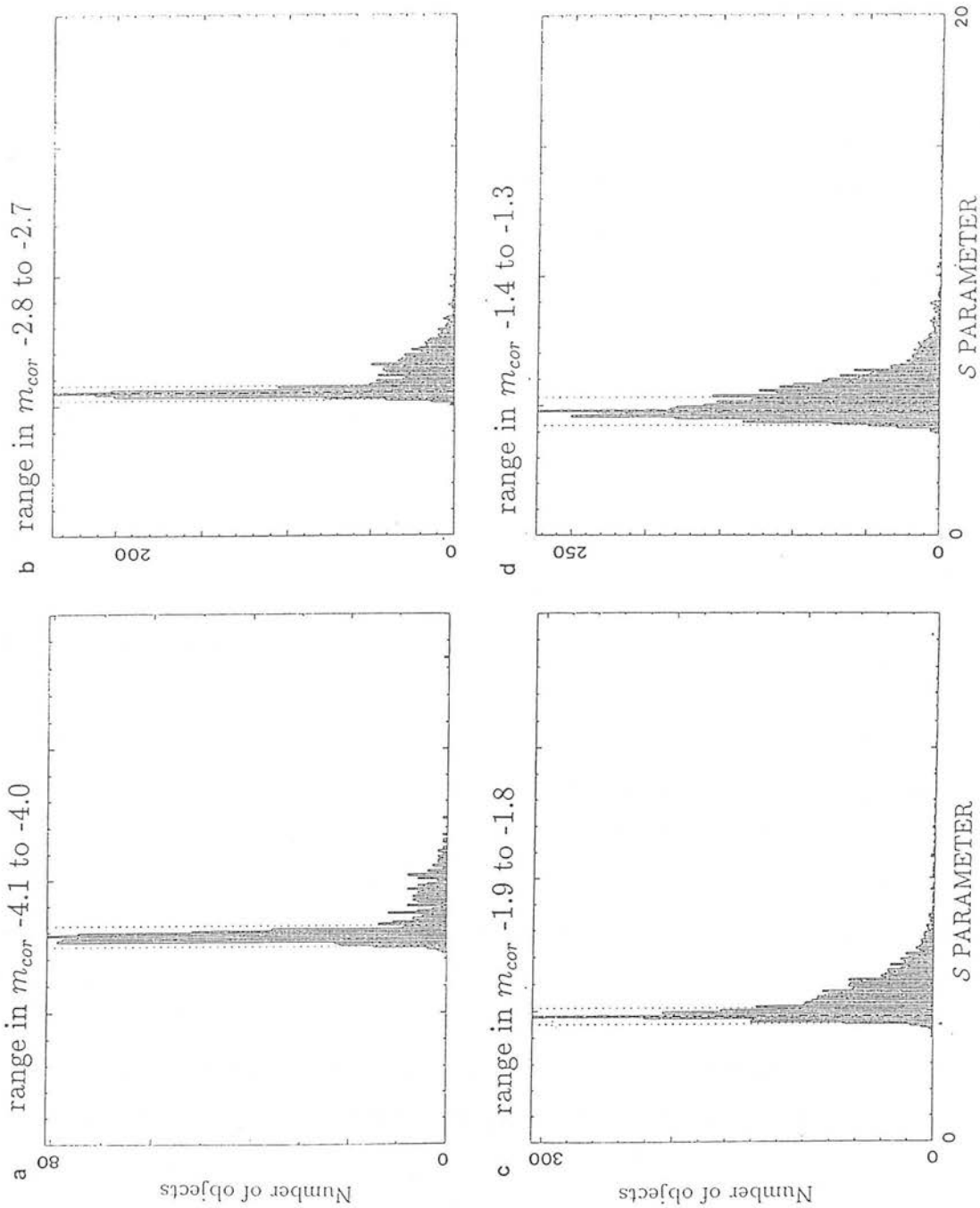


Figure 3.5: Binned distributions in the $\mathcal{S}$ classification plane along slices of constant m_{cor}. The dotted line to the right of the peak represents the separation point: (a) Here there is limited separation due to field effects. (b) There is good separation at these magnitudes. (c) There is limited separation due to information being lost below I_{th}. (d) The two populations have merged.

misclassified as galaxies for each survey field. This provides complete control over the final classification statistics of the catalogue. The optimal values adopted for the survey achieve a virtually complete catalogue of galaxies at the expense of $\simeq 10\%$ stellar contamination for plates below galactic latitude $-50°$ (the region of the EDSGC). (iv) The algorithm is quick. All 2×10^5 images on a typical plate can be classified on a VAX 11/780 in ~ 45 minutes (CPU time). Valdes (1982) quotes one hour (on the same computer) to classify 10^4 images using the FOCAS method (Jarvis & Tyson 1981) and six hours using the resolution classifier. The latter method would require a year of CPU time to classify all the images on the 60 plates used in the EDSGC.

3.3 Reduction of magnitude dependent misclassification

3.3.1 OPTIMUM MAGNITUDE RANGES FOR CLASSIFICATION

The automatic classification procedure described in Section 3.2.2b, has been used to investigate the efficiency of the most commonly used star–galaxy classification parameters available from COSMOS data (Section 3.1). Empirically it is found that each parameter has a limited magnitude range over which misclassification is acceptable. Outside this optimum range, the resultant catalogue is either very incomplete and/or very contaminated. Furthermore, the optimum ranges are not the same for each plate. They cannot be defined in terms of some constant range in image b_j or in terms of constant m_{cor} ranges. In order to achieve consistent classification image structure information, contained in the classification parameters themselves, is used to determine the optimum magnitudes.

3.3.1a *The $\mathcal{G}$ classifier*

Use of the $\mathcal{G}$ parameter for classification of bright images is discussed by MacGillivray & Stobie (1984). The parameter represents the filling factor of an ellipse fitted to the major and minor axes of the image. The characteristic loci of stars and galaxies in the $\mathcal{G}$ versus m_{cor} plane is shown in Fig. 3.1a The oscillation of the stellar locus is a result of the competition between the growth of diffraction spikes and the onset of both an image halo and two concentric image ghosts (filter ghost at stellar $b_j \simeq 11{\cdot}5$ and corrector ghost at $b_j \simeq 7{\cdot}5$, UKSTU Handbook). The basis for classification in this plane is that once stellar $\mathcal{G}$ values drop below 0·9, due to the onset of diffraction spikes, subsequent peaks never exceed this value. At magnitudes brighter than this point $\mathcal{G}$ is used to classify images, as 98 % of galaxies have $\mathcal{G} > 0{\cdot}9$. This is confirmed by visual identification of galaxies. The growth of diffraction spikes provides a convenient measure of image structure on each plate, to which the optimum magnitude ranges for all the classification parameters are anchored.

3.3.1b *The $\log_{10}(\mathcal{A})$ classifier*

Star–galaxy classification using $\log_{10}(\mathcal{A})$ versus magnitude was developed by MacGillivray *et al.* (1976) and has previously been used over a wide range of magnitudes (Shanks *et al.* 1980; Shanks *et al.* 1984; Fong *et al.* 1987). A characteristic $\log_{10}(\mathcal{A})$ versus m_{cor} plot is shown in Fig. 3.1b There are three limitations that restrict the use of this classification plane over a wide range of image magnitudes:

(i) The onset of diffraction spikes provide a bright limit for classification using measured surface brightness. Stellar magnitudes are dominated by the already saturated core, and so the faint diffraction spikes have little affect on the total magnitude. However, they do increase the image area, and consequently the stellar locus in the $\log_{10}(\mathcal{A})$ classification plane is "pulled up" into the region occupied by galaxies (Fig. 3.4a).

(ii) Once stellar images are unsaturated ($b_j \simeq 19{\cdot}0$) they are virtually indistinguishable in terms of surface brightness, from high surface brightness galaxies, although they may have very different image profiles.

(iii) For images with areas less than about 30 pixels (about 30 arcsec^2 on UKST plates, $b_j \sim 19{\cdot}5$ to $20{\cdot}0$) significant quantisation of the image area occurs due to the finite pixel size (Figs. 3.1b & 3.4c). This results in a large number of images from magnitude ranges that are broad, compared with the width of the stellar locus, being assigned identical areas. Consequently, reliable classification using the $\log_{10}(\mathcal{A})$ classifier is not possible at magnitudes corresponding to such image areas ($m_{cor} \simeq -2{\cdot}5$ and fainter). Fig. 3.6 reveals the results of such a classification. In this diagram images classified as galaxies by $\log_{10}(\mathcal{A})$ are plotted in the $\mathcal{S}$ versus m_{cor} plane. The saw-tooth feature is the result of successive rapid changes in completeness and contamination caused by the separation locus crossing area quantisation bins. Classification with $\log_{10}(\mathcal{A})$ at these faint magnitudes results in an extremely incomplete galaxy catalogue. This incompleteness is evident in Fig. 3.6 and has been confirmed by visual classification.

These limitations restrict the usefulness of the $\log_{10}(\mathcal{A})$ classifier to a range of $2^{\rm m}{\cdot}0$ fainter than the onset of diffraction spikes. Fainter than this the $\mathcal{S}$ parameter is used (Section 3.3.1c). The bright limit for $\log_{10}(\mathcal{A})$ classification discussed in (i) above, is not precise because of the gradual onset of diffraction spikes over a range of $0^{\rm m}{\cdot}5$ (stellar $b_j \simeq 16{\cdot}0$ to $15{\cdot}5$). This results in a transition region over which diffraction spikes reduce stellar surface brightness but fail to lower the $\mathcal{G}$ parameter below 0·9 For the worst case plates, 50% of galaxies can be misclassified as stars using either $\mathcal{G}$ or $\log_{10}(\mathcal{A})$ alone. This problem has been pointed out by other users of COSMOS (Shanks private communication). To ensure a complete catalogue in this small region information carried by both parameters is used. Misclassified galaxies are *sieved* from the *stars*, classified by $\log_{10}(\mathcal{A})$, by plotting them in the $\mathcal{G}$ versus m_{cor} plane. These galaxies could be recovered using a curved classification boundary, but for convenience a sieve is applied in the $\mathcal{G} \times \log_{10}(\mathcal{A})$ versus m_{cor} plane. Here (Fig. 3.7) there are two

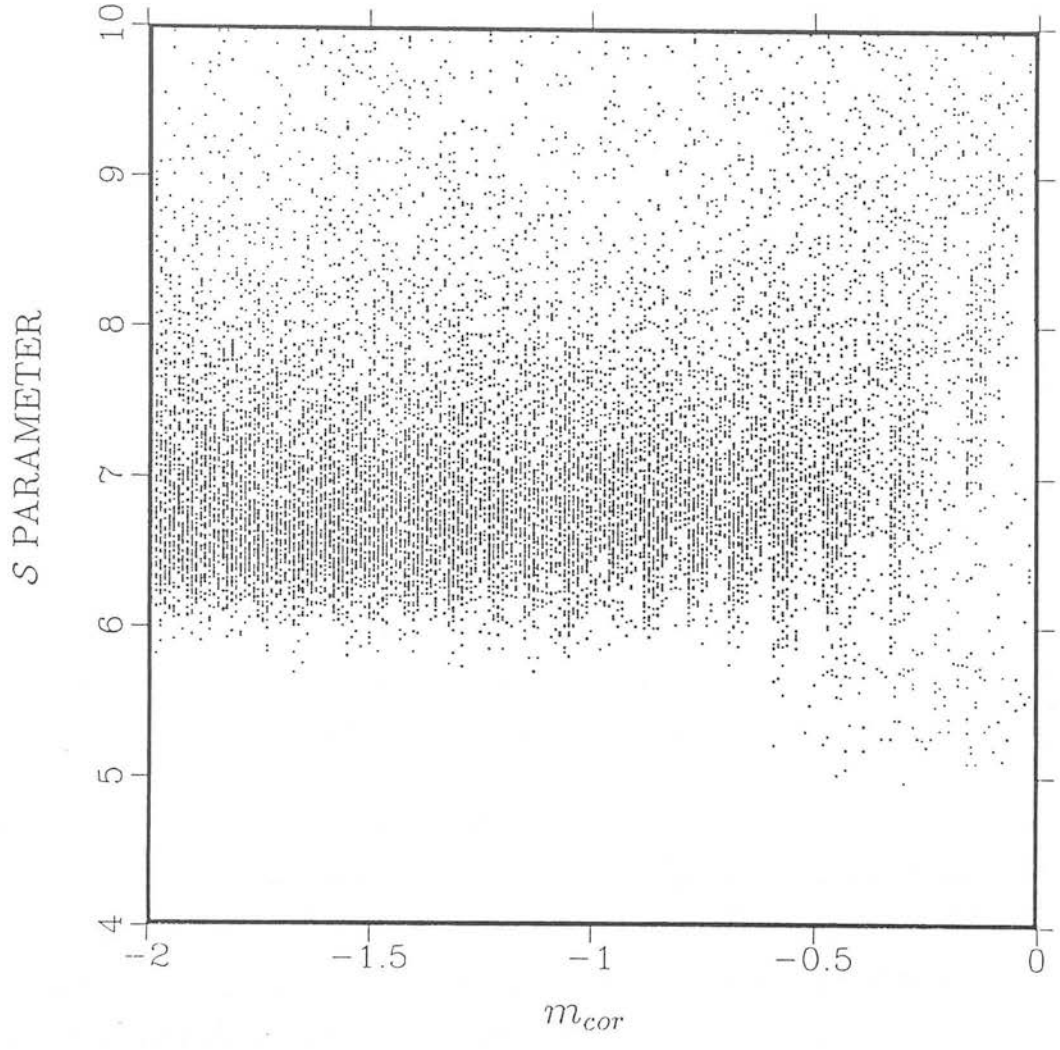


Figure 3.6: $\mathcal{S}$ values for images classified as galaxies using the $\log_{10}(\mathcal{A})$ classifier. The saw-tooth feature represents rapid changes in completeness and contamination due to area quantisation.

tight populations that can be easily separated with a straight line.

3.3.1c *The $\mathcal{S}$ classifier*

The $\mathcal{S}$ classifier defined in Eqn. 2.8 corresponds to the one-sigma width of a Gaussian profile fit to image parameters. At the same magnitude it is wider for galaxies than for the point spread function of stars. The distribution of images in the classification plane is shown in Fig. 3.1c. This parameter carries extra information at faint magnitudes compared to $\log_{10}(\mathcal{A})$, so it is less affected by pixel quantisation. The faint limit for the use of the $\mathcal{S}$ parameter sets the faint limit for the EDSGC and is discussed in Section 3.3.2. A fundamental limit on its use at bright magnitudes is discussed here.

The maximum image intensity parameter (I_{max}) suffers from field effects once images are saturated. Thus the $\mathcal{S}$ parameter of stars begins to be affected by field effects when $b_j < 19.0 - 19{\cdot}5$. Unlike the stellar locus shift in the $\log_{10}(\mathcal{A})$ versus m_{cos} plane (Section 3.2.1a), the shift in the $\mathcal{S}$ versus m_{cos} plane is in both directions. Here field effect correction of image magnitudes is not sufficient to prevent rapid widening of the stellar locus, once images become saturated (Fig. 3.5a).

The optimum m_{cor} for transition between the $\log_{10}(\mathcal{A})$ and $\mathcal{S}$ classifiers is therefore a compromise between the effects of quantisation on $\log_{10}(\mathcal{A})$ and the severe field effects on $\mathcal{S}$. It is fortunate that there is a transition region of $\sim 0^{\rm m}{\cdot}5$ over which both parameters work equally well. In this region there is less than 1% difference in the number of galaxies classified by either parameter.

3.3.2 THE COMPLETENESS LIMIT

The limiting magnitude to which the automated classification procedure can be used, has been investigated. For the majority of plate data sets classified, the galaxy catalogue

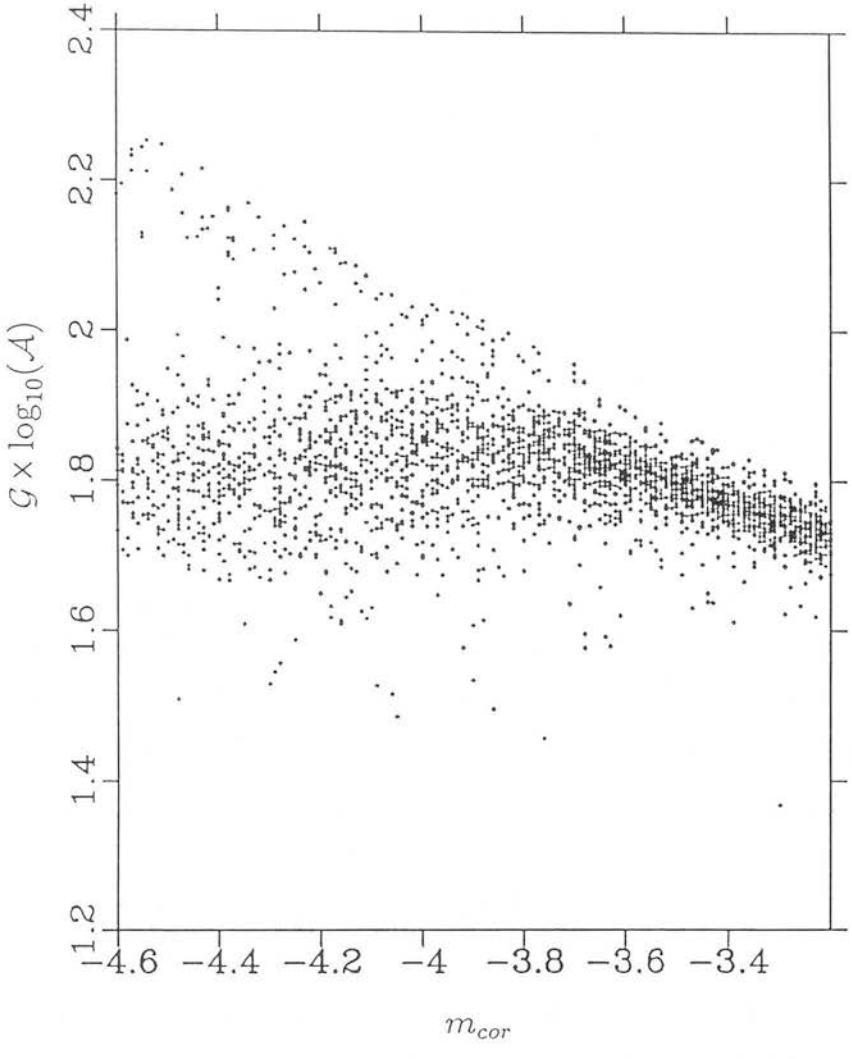


Figure 3.7: Objects classed as stars by $\log_{10}(\mathcal{A})$ plotted in the $\mathcal{G} \times \log_{10}(\mathcal{A})$ verses m_{cor} plane. Misclassified galaxies are easily identified as they lie above the true-star locus.

becomes rapidly incomplete fainter than $2^{\rm m}\!\cdot\!0$ above the m_{sky}, as determined from the slope of the differential number–magnitude counts, which rapidly becomes flat at fainter magnitudes (Fig. 3.8).

Examination of the distribution of images in the $\mathcal{S}$ classification plane, at $m_{cos} \sim -2\cdot0$ (Fig. 3.5c) shows that there is no major change in the shape of the distributions at this magnitude. The skew distribution discussed in Section 3.2.2b is still visible, with the stars determining the position of the peak. The incompleteness is therefore due to the galaxy and stellar populations merging, rather than the number of galaxies exceeding the number of stars and dominating the position of the peak.

The merging of the two populations could be due to either noise on the COSMOS image parameters and/or the emergence of a faint noise image population. In Section 2.3.3 the use of plates centred around $0^{\rm hr}$ to identify noise images was discussed. Images in the overlap region common to both plates are real, those detected on one plate but not the other are noise images. Noise images can be ruled out as the cause of the catalogue incompleteness since, for m_{cos} brighter than $-0\cdot5$, they constitute less than 2% of real images in any $0^{\rm m}\!\cdot\!2$ bin (Fig. 2.5). Application of the classification procedure on *clean* (images common to both plates) and *noisy* (data from one plate) datasets confirms that noise images do not affect the magnitude of the completeness limit (Fig. 3.8).

Fig. 2.7 shows the variation in the noise on the IAM parameters, $\mathcal{A}$ and $\mathcal{S}$, as a function of magnitude. Because the completeness of the galaxy catalogue drops rapidly fainter than $m_{cos} \simeq 2\cdot0$, then if this was due to noise on the $\mathcal{S}$ parameter, a correspondingly steep increase should also be seen in $\sigma_{\mathcal{S}}$. There is no such steep increase at $m_{cos} \simeq 2\cdot0$ just a continuation of the gradual increase seen at brighter magnitudes.

These investigations show that neither image noise on the classification parameters or noise images are responsible for setting the completeness limit of the classification procedure. The obvious conclusion is that there is insufficient information in the $\mathcal{S}$ parameter to distinguish between stars and galaxies at these faint magnitudes. In

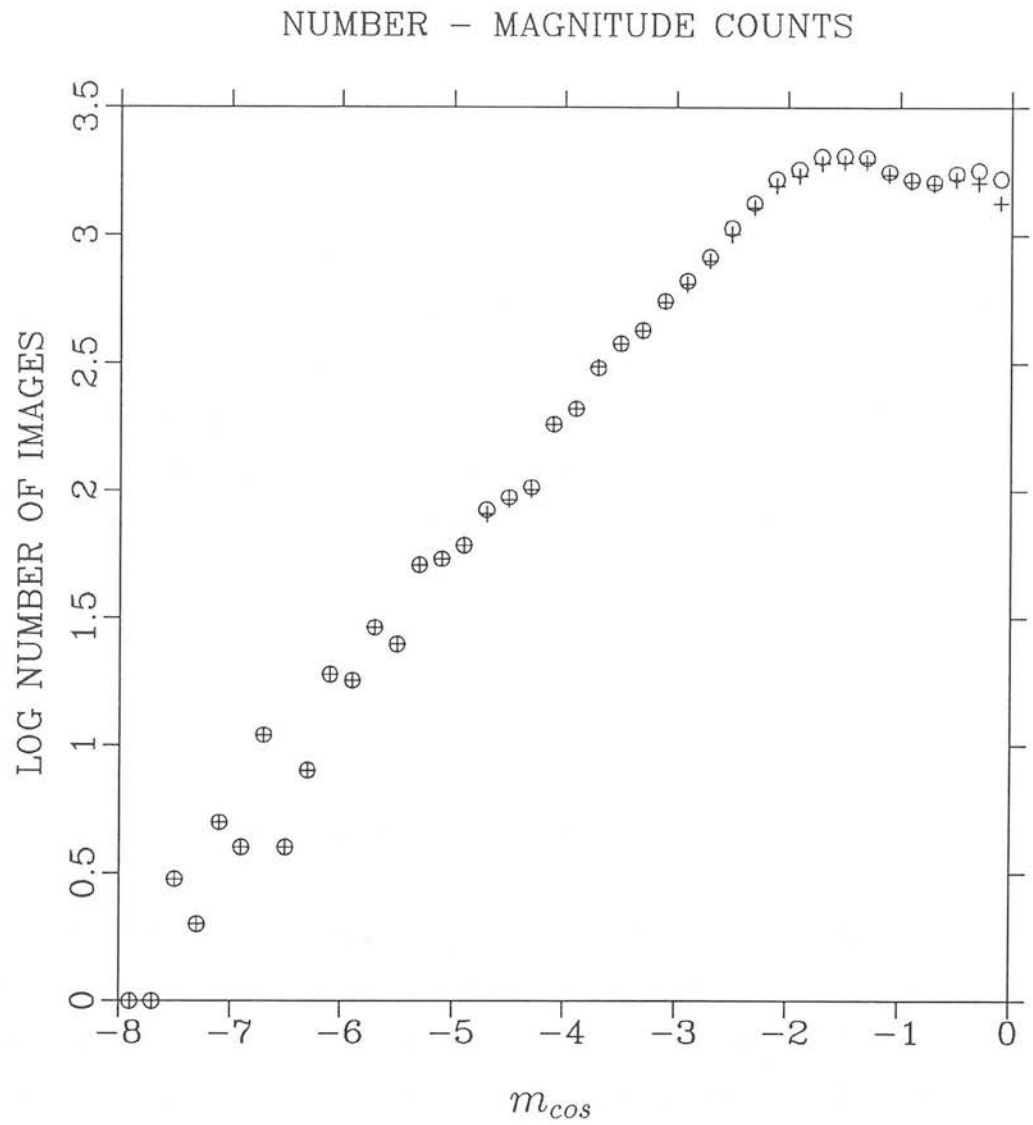


Figure 3.8: Differential number–magnitude counts for galaxies from the F293,F348 overlap. Crosses are for galaxies common to both plates (*clean* dataset). Circles are for galaxies on one plate only (*noisy* dataset).

addition, because all other classification parameters breakdown at brighter magnitudes, there is no parameter available to the EDSGC that can reliably classify images fainter than $m_{cos} \simeq -2{\cdot}0$. The variation of m_{sky} from plate to plate can be as large as $0^{\mathrm{m}}{\cdot}5$, so in order to ensure that the catalogue is at least 95% complete with $< 10\%$ contamination, the limiting magnitude of the survey is set at $b_j = 20{\cdot}0$.

It is interesting to note that the completeness limit is of the order of the magnitude at which isophotal effects become important for images detected at the $\sim 9\%$ threshold (Section 2.3.1). Choosing a lower isophote for the IAM data may recover information that at present is not detected as it lies below the threshold, and consequently extent the use of the $\mathcal{S}$ to fainter magnitudes. The resulting improvements may however be countered by the increased noise on the parameters and the resultant increase in the number of noise images.

3.4 Results

3.4.1 AUTOMATED VERSUS VISUAL CLASSIFICATION

Visual classification was carried out for 300 galaxies and 300 stars on each of 5 plates selected from the EDSGC. The plates chosen cover a range of different sky backgrounds and are fully representative of the plate grades (A1, A2, AI3) used and the epochs (with or without N_2 flushing) when plates were taken. Comparisons of the UKST logs indicate that these 5 plates represent the full range in plate quality expected in the completed EDSGC. The images themselves were randomly selected from 3 magnitude slices, one slice in each magnitude range where each of the classification parameters is used. The faintest of these slices ($\sim 3^{\mathrm{m}}{\cdot}0$ brighter than m_{sky}) is still bright enough for reliable visual classification. The results of these investigations are shown in Table 3.2 and can be summarised as follows:

CONTAMINATION & COMPLETENESS ESTIMATES

PLATE No.	J5304 {AI3}		J3579 {A2}		J8046 {AI3}		J2693 {A3}		J6124 {A1}	
	ctm	*cmp*	*ctm*	*cmp*	*ctm*	*cmp*	*ctm*	*cmp*	*ctm*	*cmp*
classification parameter(s)										
$\log(\mathcal{A})$ & $\mathcal{G} \times \log(\mathcal{A})$	2%	95%	5%	96%	11%	100%	2%	100%	2%	100%
$\log(\mathcal{A})$	10% (6%)	98%	2% (4%)	96%	11% (12%)	100%	11% (6%)	100%	7% (5%)	95%
$\mathcal{S}$	7% (6%)	96%	8% (3%)	98%	11% (12%)	100%	9% (5%)	96%	11% (6%)	98%

Table 3.2: Visual estimates of stellar contamination (*ctm*) and galaxy completeness (*cmp*) for a sample of plates used in the EDSGC. Plate grades are given in { }. Percentages in brackets, ( ), are those predicted by the automated classification algorithm.

(i) In all magnitude ranges the completeness exceeds 95%. The average completeness across all magnitudes is $98\% \pm 2\%$,

(ii) The contamination using $\log_{10}(\mathcal{A})$ is $8\% \pm 6\%$ and that using $\mathcal{S}$ (at the limiting magnitude) is $9\% \pm 2\%$.

(iii) At the depth of the *Lick Catalogue* ($b_j \simeq 18{\cdot}6$) the rms variation in contamination and completeness is 2%, which compares favourably with the limit discussed in Section 1.4.

(iv) The automated estimates of contamination are within 5% of the 'eyeball' estimates. Therefore, the automated procedure discussed in Section 3.2.2b appears to provide reliable quality control of the classification efficiency. These statistics will be provided for each field in the final version of the EDSGC (Appendix A).

(v) Further to Table 3.2, visual inspection of galaxies classified by the $\mathcal{G}$ parameter on F349, F350, F409, F410, F472 and F473 showed that the sample was $4 \pm 3\%$ contaminated with stars. The completeness limit is set at 98% by choosing a $\mathcal{G} = 0{\cdot}9$ discrimination boundary. During the determination of redshifts of a sample of such galaxies covering the whole of the survey it was noted that most of this contamination occurred at $m_{cos} \simeq -7$ (galaxy $b_j \simeq 15{\cdot}0$). This magnitude corresponds to to the point in the $\mathcal{G}$ classification plane where the sinusoidal stellar locus has a peak (Fig. 3.1a). Contamination $0^{\mathrm{m}}{\cdot}5$ either side of this point is marginally higher than average, but the number of objects involved is small (galaxies within this magnitude range have a surface density of ~ 1 per 3 square degrees).

The classification of daughter images produced by the COSMOS deblending software has also been examined by visual inspection. The examination of all bright images ($b_j < 16{\cdot}5$) before and after deblending shows that 96% of these images are correctly classified using the $\mathcal{G}$ parameter after deblending, compared with 94% of images successfully classified without deblending.

For fainter images two samples of daughter images, classified as stars and galaxies using the $\log_{10}(\mathcal{A})$ and $\mathcal{S}$ parameters were visually inspected. The resultant catalogue composed only of daughter images has a completeness of 96% with a 15% stellar contamination. Since the number of deblended images in the EDSGC is $\sim$ 10%, the overall classification efficiency is unchanged. To conclude, the statistics for bright images indicate that star-star blends are successfully removed from the catalogue and the parameters of deblended images are not altered to the detriment of image classification over the full magnitude range of the EDSGC.

3.4.2 CONTAMINATION ESTIMATES FROM NUMBER COUNTS

Misclassification checks 'by eye' can only be carried out over a restricted number of plates due to the sheer number of objects involved. In order to test the homogeneity of the catalogue over larger areas the affect of residual stellar contamination on the resulting galaxy differential number–magnitude counts has been simulated. The approach adopted involved testing the number–magnitude slope for various model populations consisting of galaxies contaminated with population of stars. Each trial consisted of selecting a random sample of stars from the full magnitude range of the EDSGC, transforming their magnitudes onto the b_j system as if they were galaxies, and then adding them to the galaxy population. The fraction of stars added at each magnitude is a constant and so closely follows the manner in which stars are expected to be misclassified by the automatic selection procedure. This accurately reproduces the magnitude distribution of stars expected in a contaminated galaxy catalogue.

Since stellar contamination flattens the galaxy number–magnitude slope, the most conservative assumption regarding the intrinsic galaxy magnitude distribution is that it follows the highest slope which is physically reasonable. For this reason a Euclidean slope is used for the galaxy only distribution, over the full magnitude range of the simulations ($b_j \simeq 13{\cdot}0$ to $b_j \simeq 19{\cdot}75$). This slope is consistent with that from

Mean stellar contamination	Gradient of least-squares fit
0 %	0.60 ±0.012
5 %	0.57 ±0.015
10 %	0.55 ±0.009
20 %	0.53 ±0.009
25 %	0.52 ±0.009
30 %	0.51 ±0.009

Table 3.3: Differential number–magnitude count slopes and their standard error for simulated catalogues with varying stellar contamination.

other galaxy samples in the same magnitude range, *eg.* Shanks *et al.* 1984 and Sebok 1986 (Differential counts for a sample region of the EDSGC[1], have a least squares slope of $0{\cdot}58 \pm 0{\cdot}01$ for $b_j \simeq 20{\cdot}0$). The results of the simulations are shown in Table 3.3. This presents the slope of the number–magnitude counts, estimated from least squares fitting, for model galaxy catalogues contaminated with different percentages of stars. From these figures a mean contamination of 10% gives rise to a slope that is inconsistent with our data at the $2{\cdot}5\sigma$ level, while a mean contamination of 20% or larger is ruled out at the $4{\cdot}5\sigma$ level of significance. These results provide strong constraints on the level of contamination in the EDSGC.

3.4.3 RESIDUAL PLATE-POSITION DEPENDENT MISCLASSIFICATION

In Section 3.2.1 the difficulties of classifying uniformly over the complete area of a plate were stressed. Since COSMOS scans are not centred exactly on the plate centre, the west side of the scan suffers considerably more from vignetting and desensitisation than the east side, causing stronger plate background density variations and field effects. In addition, image defocussing and fogging tend to be more extreme at the edges of plates. To investigate the severity of these effects and the extent to which the field effect correction algorithm has been successful, the misclassification statistics on the $\frac{1}{3}^{\circ}$ overlap region between adjoining plates were examined. Fig. 3.9 shows the percentage of unpaired objects, automatically classified as real galaxies, as a function of magnitude on the western edge of 5 plates, and across a 3° overlap centred at RA= $0^{\rm h}$. These are compared with a prediction of the percentage of unpaired objects expected from the 'eyeball' estimates of contamination and completeness (Table 3.2). This prediction was made using the assumption that after field effect correction has been applied, vignetting and desensitisation do not affect the misclassification rates. There is good agreement in Fig. 3.9 between the overlap statistics on the plate edges and those across a large overlap, and also with the predicted percentage. This indicates there is no evidence,

[1] presented in Heydon-Dumbleton *et al.* (1989)

after field effect correction, for greater misclassification at the edges of plates compared to their centres. The rms variation in the number of unpaired objects from plate to plate is 3% at $b_j = 20{\cdot}0$. This represents the systematic plate-to-plate variation in number density produced by application of the classification algorithm on plates of varying quality.

It should be noted that noise images, discussed in Section 2.3.3, are expected to have low surface brightness and non-Gaussian profiles and so are probably classified as galaxies. Only $\sim \frac{1}{3}^{\rm rd}$ of all $b_j \leq 20{\cdot}0$ images are galaxies, so noise image contamination in the EDSGC is expected to be $\sim 6\%$, given that 2% of images in the raw data are noise (Section 2.3.3). The percentage of unpaired images used in Fig. 3.9 will therefore have a non-negligible contribution due to noise images, which cannot be separated from that due to misclassification. Thus these figures are upper limits on the underlying level of contamination. The increase in the percentage of unpaired objects around $m_{cos} \simeq -2{\cdot}0$ may be due to the corresponding increase in the percentage of noise images seen in Fig. 2.5.

3.4.4 RESIDUAL SYSTEMATICS

As a final check on systematic number density variations across plates a 40 plate *stack* was obtained by summing the binned number density counts of galaxies ($m_{cos} < -1$) on each of the plates and hence removing real structure. Plotted in Fig. 3.10 are the total counts on 40 plates in each of 50 strips ($0{\cdot}1^\circ$ wide), running the length of the scanned region in approximately a north–south direction. The mean count level across these strips is also plotted, along with the 1 sigma rms fluctuation. This estimate of the systematic deviation from a totally uniform distribution corresponds to 3% of the average count per strip. This indicates that the combined systematic errors in both *intra-plate* photometry and misclassification in the data are confined to an acceptable level and are below the intra-plate gradients plates in the SSGP data ($\sim 7\%$), reported

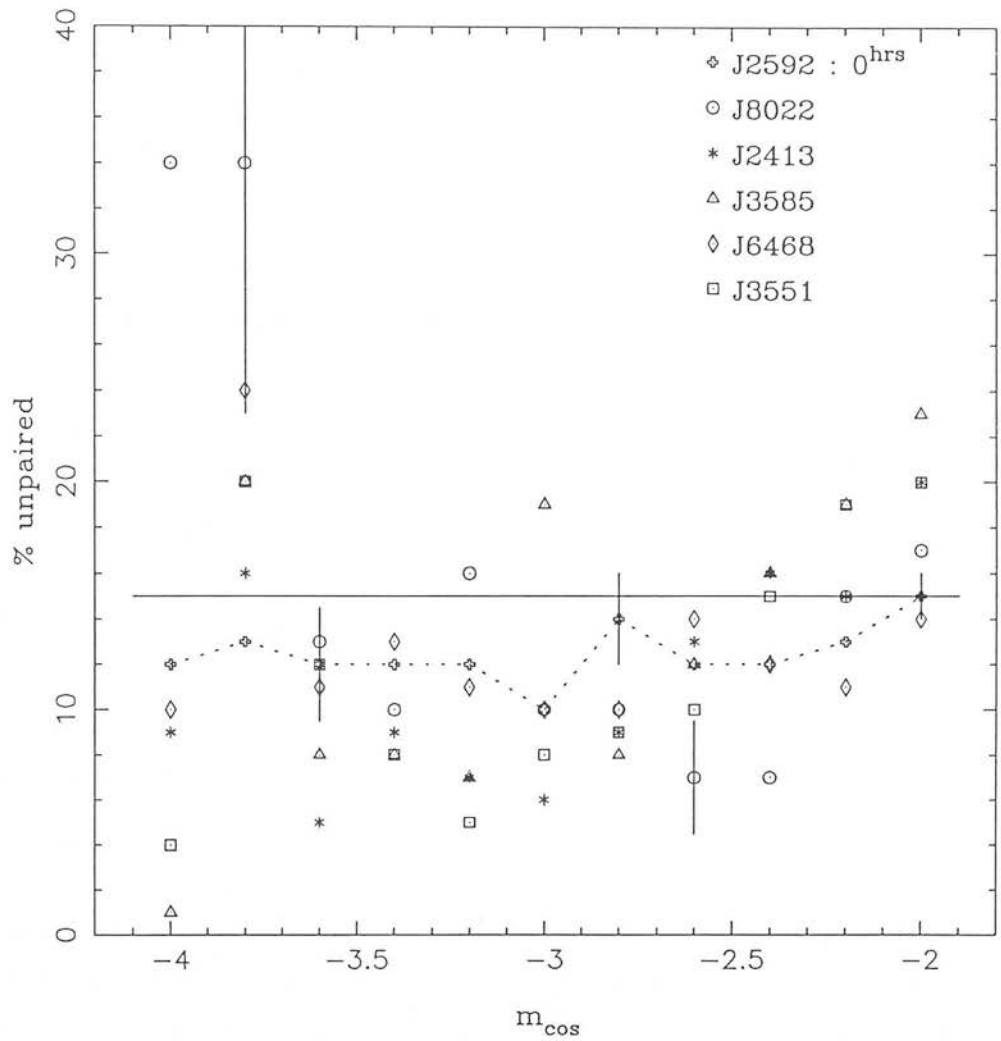


Figure 3.9: The percentage of unpaired galaxies from 6 plate overlaps plotted as a function of magnitude. The dotted line joins the points from a large (3°) overlap and the dashed line shows the percentage predicted from visual classification.

by Geller *et al.* (Section 1.4). A marginally higher number density ($2 \times$ rms) is visible on the western edge, but as this region is removed from each plate during construction of the *plate-mosaic* (Section 5.2), it does not present a problem. For comparison, Fig. 3.11 shows the residual variation across a plate in the raw Shane & Wirtanen data counted in 1° bins (rms $\simeq$ 7%) along with the EDSGC data smoothed on the same scale (rms $\simeq$ 2%).

3.5 Discussion

In Table 3.1 estimates of contamination and completeness from previous work are listed. Direct quantitative comparisons between most of these methods of classification are impossible because several of the classification parameters are not available from COSMOS data and the data sets produced by the different authors cover different plate material in different parts of the sky. However, the results presented in Section 3.4 and Table 3.2 are clearly a substantial improvement on the available statistics in Table 3.1.

In general the variation in classification efficiency can be kept within the limits discussed in Section 1.4, however as with any classification algorithm there are situations where it is possible for the algorithm to breakdown. It is important in the production of this large galaxy catalogue that the occurrence of these situations is noted and corrected where possible.

The largest images on UKST plates, such as dense star clusters, nearby galaxies, very bright stars and their associated halos and corrector ghosts, present a problem for the COSMOS analyser. These images are broken up and the resultant images are often classed as galaxies. Other spurious structures can occur in satellite trails and where the scanned region includes the stepwedge. The variation in the plate background due to the large emulsion-corrector/filter-corrector ghosts (diameters $\sim$ 60mm, see Fig. 2.4) may present a problem for the field effect correction software, because within

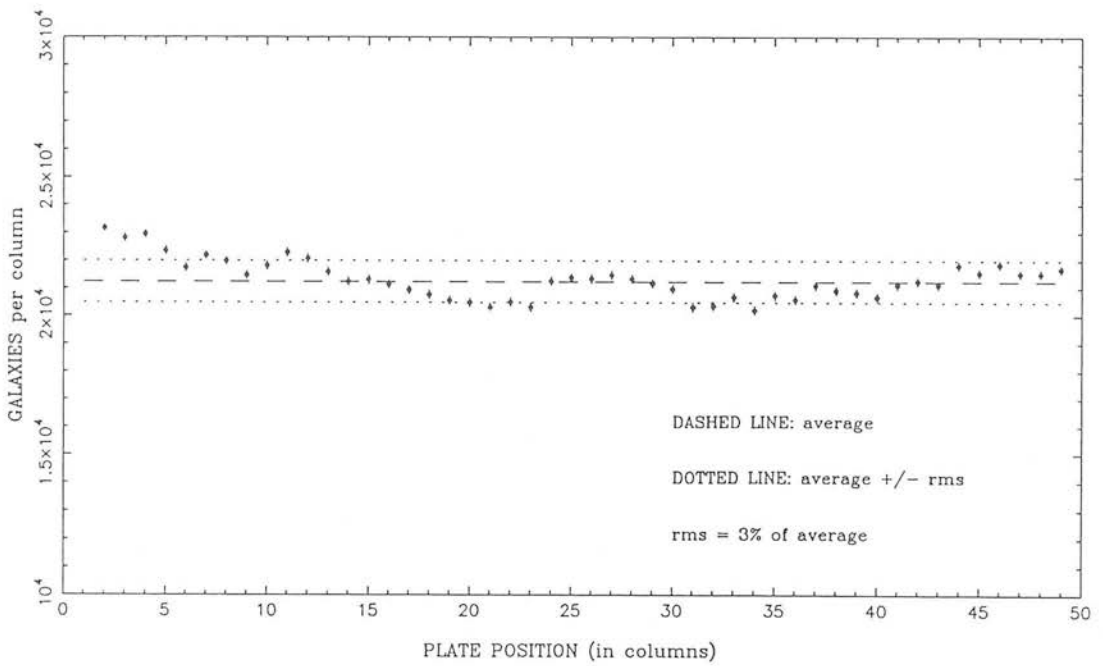


Figure 3.10: The number density variation in the W–E direction for a stack of 40 plates. Points were obtained by binning the galaxy distribution on each plate into 50 N–S strips ($\sim 0{\cdot}1°$ wide) and then adding the counts for all the plates. The systematic rms variation (dotted line) is 3%.

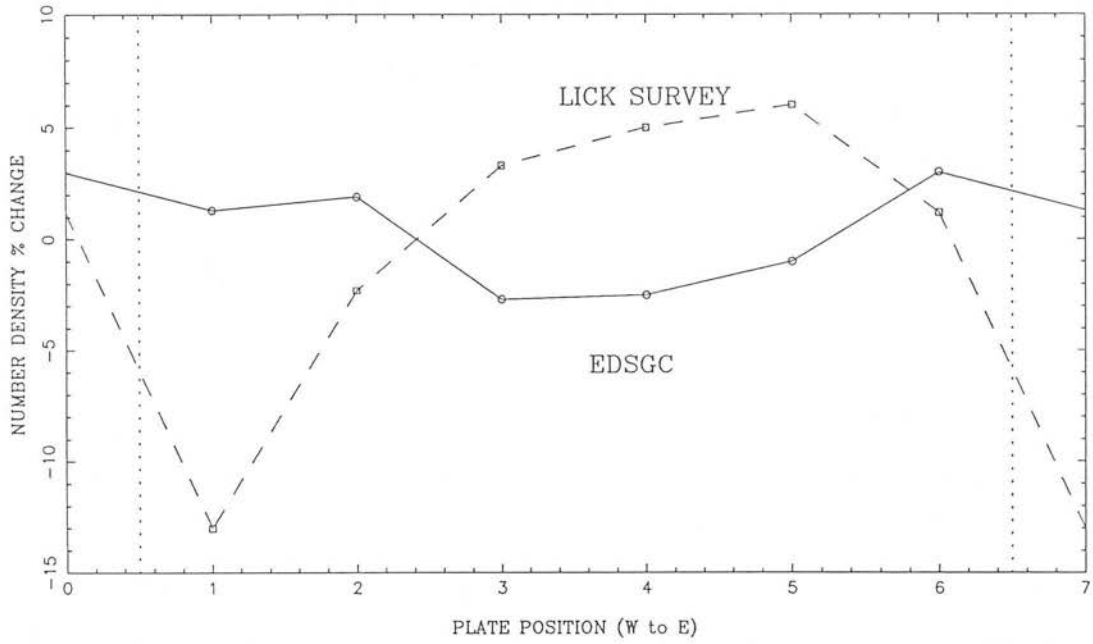


Figure 3.11: A comparison of the systematic number density variation from plate to plate across the *Lick Catalogue* and the EDSGC. Data is binned in $\sim 1^{\circ}$ cells and the dotted lines represent plate boundaries

the ghost there are background variations over $< 0{\cdot}5°$ (less than the resolution of the correction algorithm). The magnitude of the variation, which depends on the brightness of the image causing the ghost, will determine whether image classification is affected. All images within the vicinity of the above objects are routinely removed as well as any structure in the galaxy distribution that is obviously associated with such ghosts. Regions where there are no obvious effects are left in the survey. A list of all images and regions removed from the EDSGC is given in Appendix A.

The *ESO/SERC Atlas* plate for F411 (J4606) was removed from the survey (Section 2.2.4), because of large variations in the plate background ($> 1^{\rm m}{\cdot}0$), which could not be corrected by the field effect correction algorithm. A new original was obtained, because it was unclear whether the variations were confined to the ROE copy, or on the original, or whether they were due to real nebulosity. No such problems were encountered on the replacement copy or reported by Horstmann (1988), who used a PDS to scan a film copy of J4606. The ROE copy of J4606 may be defective.

The copy of the new F411 plate (J11341) was produced by the ROE Photolabs rather than at ESO. To check the consistency of the ROE copying process a copy of J9529 (F406) was also made. After image classification, Galaxies on the ESO and ROE copies of J9529 were paired with the method described in Section 3.4.3. There is no excess of unpaired images in the two J9529 galaxy catalogues above the level found on the F293, F348 overlap (Fig. 3.9) and so, as far as image classification is concerned, the ROE Photolabs copying process is of the same standard as the ESO process. Unless the copying process is highly variable, the ROE copy of J11341 is expected to be of a similar high quality.

3.6 Summary

An optimum procedure to objectively classify images from COSMOS scans of glass copies of the SERC (J) survey has been developed in order to construct the Edinburgh/Durham Southern Galaxy Catalogue. An extensive investigation has been carried out into both image position and magnitude dependent systematics and their affect on the image classification efficiency. For the first time, the breakdown of commonly used classification parameters has been quantified and optimum magnitude ranges have been defined, outside which each parameter should be used. In this chapter it has been demonstrated that, by using objective classification techniques, it is possible to produce a homogeneous galaxy catalogue, from machine measurements of plates of variable quality. The techniques used here can also successfully classify overlapping images resolved by the COSMOS deblending software, this is of particular importance for data from the cores of compact clusters. It is estimated that the EDSGC will be $> 95\%$ complete with $< 10\%$ stellar contamination, and that there is $\simeq 3\%$ rms residual variation in the number of objects classed as galaxies over the area covered by the entire catalogue.

Chapter 4

Image Photometry

4.1 Introduction

The second and equally important source of systematic error in galaxy catalogues is image photometry. Any systematic error in image photometry, particularly those which result in a variation in limiting magnitude (m_{lim}), will introduce spurious structure into the observed galaxy distribution (Eqn. 1.1). Such problems in the *Lick Catalogue* were noted in Section 1.4.

The purpose of this chapter is to describe the methods used in the EDSGC to ensure accuracy and homogeneity of the galaxy magnitudes across the whole survey region. It is important in the construction of a catalogue of this type that galaxy photometry is consistent across a particular plate (*intra*-plate photometry) and from plate to plate (*inter*-plate photometry). The interaction between image classification and image photometry is such that the techniques used to ensure consistent intra-plate photometry have already been described in Chapter 3. These will be re-iterated briefly in Section 4.2 and further estimates made of the associated errors.

The main topic of this chapter (Section 4.3) is the reduction of systematic errors associated with inter-plate photometry. The strategy adopted (Section 4.3.1) requires obtaining a large number of galaxy CCD sequences and arranging them so that those plates without sequences have 4 overlaps to plates with a sequence. The reduction of standard star and galaxy CCD sequences is described in Section 4.3.2. The details of the calibration of a 35-plate mosaic are presented in Section 4.3.3, together with the estimation of residual errors.

4.2 Intra-Plate photometry

Intra-plate photometry is concerned with the definition of a consistent magnitude system for all galaxies on the same plate. It is important that the measured magnitude of each galaxy can be related to its actual magnitude, regardless of position on the plate and consequently that the same limiting magnitude can be defined across the plate. The basic problem arises because of the effects of vignetting and desensitisation which act as if the emulsion sensitivity, and therefore m_{lim}, decreased radially from the plate centre outwards.

In the *Lick Catalogue* these effects are overcome by correcting the counts in each cell by a factor to ensure that the average number (over all plates) is the same regardless of the cell position. Such corrections are not required in the EDSGC as the COSMOS calibrated magnitude (m_{cos}) for unsaturated images is independent of the above variations and defines a consistent intra-plate magnitude system. It should be noted that if this was not the case, the application of average correction factors would not be totally effective as the amplitude of the background variation is dependent roughly on plate epoch, *ie* whether N_2 flushing was applied (Section 2.2.1). de Lapparent *et al.* (1986) suggest that the counting gradient across plates in the raw *Lick Catalogue* data is both observer and epoch dependent and consequently residual intra-plate gradients are responsible for some of the structure seen in the SSGP maps (Fig. 1.1).

m_{cos} is constructed (Eqn. 2.9) from the ratio of summed image intensity to local plate background intensity, *ie:*

$$m_{cos} = -2{\cdot}5 \log_{10} \left(\frac{\sum_{pix} \mathrm{I}_{pix} - \mathrm{I}_{bg}}{\mathrm{I}_{bg}/\mathrm{A}_{pix}} \right),$$

and can be related to the object magnitude (b_j) and the sky background ($m_{sky} = -2{\cdot}5 \log_{10}[\mathrm{I}_{sky}/\mathrm{A}_{pix}]$) by (Eqn. 3.1):

$$m_{cos} = b_j - m_{sky},$$

given the following assumptions:

(i) Intrinsic sky background variations are negligible compared with those caused by vignetting and desensitisation (Campbell 1982, Wesselink 1984). The latter affect I_{pix} and I_{bg} by the same factor, which cancels out in the ratio. I_{bg} can therefore be replaced by I_{sky}.

(ii) Images are unsaturated and therefore the measured pixel intensity still has a one-to-one relation with the incident intensity.

The validity of these assumptions can be seen in Fig. 3.2 which shows a plot of m_{cos} verses b_j for galaxies distributed across F411. The b_j magnitudes are obtained from Kirshner *et al.* (1978, PDS photographic and photoelectric magnitudes for galaxies distributed over the field) and Hawkins (1981, electronographic magnitudes for stars and galaxies near RA: $00^{hr}\ 55^{m}\ 48^{s}$, DEC: $-27°\ 53'$). Both these sequences consist of total magnitudes in B and V, which have been colour corrected to b_j using the colour equation of Blair and Gilmore (1982). The m_{cos} values for these objects were obtained from a COSMOS scan of the J4606 copy-plate in the ROE version of the *ESO/SERC Atlas*. This plate was later rejected from the EDSGC because of its severe background variations and so provides a good example to test the validity of the above assumptions.

There are several important features to note from this figure:

(i) $b_j = 1{\cdot}0 m_{cos} + m_{sky}$ for galaxies as faint as $b_j = 21{\cdot}0$ and at least as bright as 15·0 (possibly as bright as 14·0, MacGillivray & Dodd 1983).

(ii) Isophotal effects on m_{cos} cause it to underestimate galaxy magnitudes fainter than $b_j = 21{\cdot}0$ (MacGillivray & Dodd 1983).

(iii) m_{cos} badly underestimates stellar magnitudes brighter than $b_j = 19{\cdot}5$. The small change in m_{cos}, for a large change in b_j, results from the variation of the area of the point spread function with magnitude.

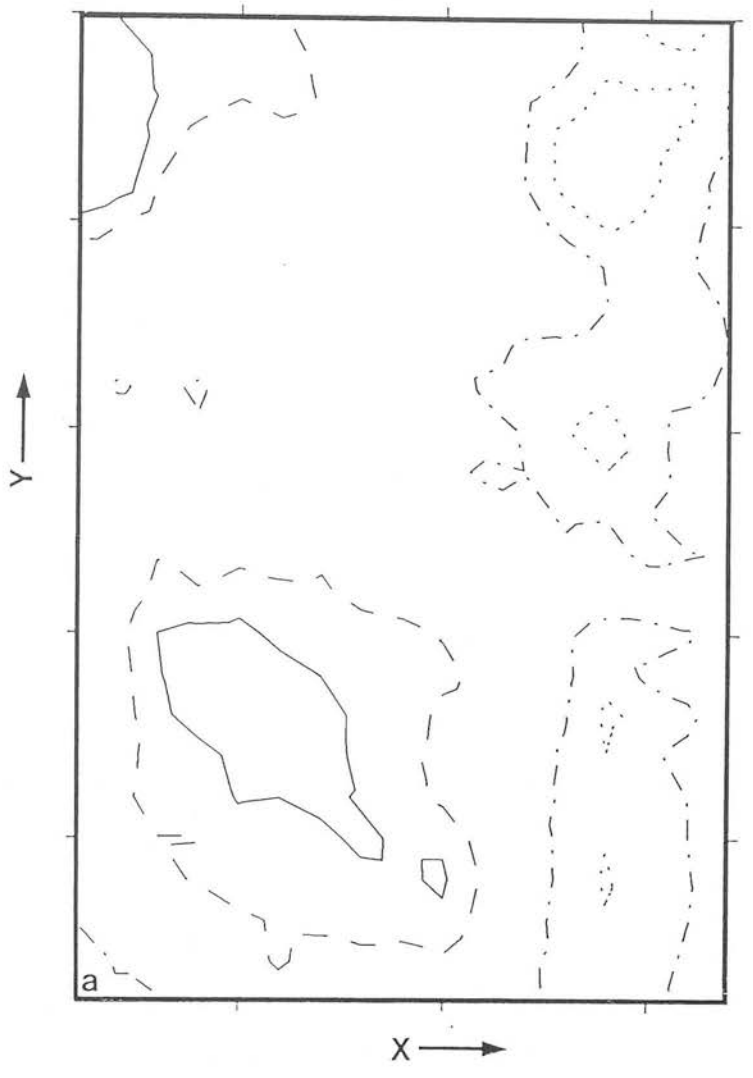


Figure 4.1: Smoothed contour values of the difference in m_{cos} values for images common to the F293,F348 overlap. Contours range from $-0^{m}.04$ to $0^{m}.04$ in $0^{m}.02$ intervals.The region shown covers $3° \times 5°$.

(iv) For stars brighter than $b_j \simeq 16$, the slope flattens again. This reflects the onset of diffraction spikes and stellar halos which carry unsaturated magnitude information.

(v) The Hawkins galaxy magnitudes with $b_j \lesssim 20{\cdot}5$ are systematically $\sim 0^{\rm m}{\cdot}2$ fainter than expected when compared to the Kirshner *et al.* data, which show an appreciable stochastic variation (rms $\sim 0^{\rm m}{\cdot}3$), but no systematic deviation from a 45° slope.

Further discussion on the validity of (i) is presented in Section 4.3.3. The Galaxy CCD sequences reveal a tendency for m_{cos} to underestimate b_j for some bright galaxies. These are thought to be high surface brightness elliptical galaxies which appear to saturate m_{cos} at magnitudes fainter than galaxies with lower surface brightnesses. The amount that m_{cos} underestimates the actual magnitude is dependant on the profile of each individual galaxy, and the effect is more prominent in the CCD sequences as they are centred on clusters. Arguments in 4.3.3 show that (i) is still valid for most regions of the plate, but with the caveat that care should be applied in the case of HSB galaxies (see also Metcalfe *et al.* 1989).

An estimate of the combined residual variation in image classification and galaxy photometry was presented in Section 3.4.4. Fig. 3.10 shows that the residual variation in galaxy number density, after application of an $m_{cos} \leq -1{\cdot}0$ cut off, is $\sim 3\%$ (rms). If this variation was due purely to systematic errors in image photometry, then from the number–magnitude relation ($2{\cdot}3\gamma\Delta m = \Delta N/N$) the amplitude of this variation would be $< 0^{\rm m}{\cdot}02$ ($\gamma = 0{\cdot}6$, Heydon-Dumbleton *et al.* 1989).

Data from the large overlaps at $0^{\rm hr}$ (Section 2.3.3) can be used to check for large gradients in image photometry and estimate the systematic variation across a plate. Fig. 4.1 shows smoothed contours of the difference in m_{cos} for images common to plates J3578 and J3592 (F293 and F348). A correction has been applied to account for the different plate m_{sky} values. Systematic differences are $\sim 0^{\rm m}{\cdot}03$ across the $3° \times 5°$ region.

Eqn. 3.1 is not expected to hold in the regions of very bright stars and their associated

filter-corrector/emulsion corrector ghosts, as here background variations are not equivalent to sensitivity changes. If the background changes are on scales > the COSMOS filter scale then:

$$m_{cos} = -2{\cdot}5 \log_{10} \left(\frac{\sum_{pix}(\mathrm{I}_{pix} - \mathrm{I}_{bg})}{(\mathrm{I}_{bg} + \delta \mathrm{I}_{bg})/\mathrm{A}_{pix}} \right), \tag{4.1}$$

where $\delta \mathrm{I}_{bg}$ is the extra background intensity contributed by the ghost or star halo. The COSMOS estimate of the image magnitude, b_j, is therefore too faint. Such bright stars and those ghosts, which affect image classification, have been routinely removed from the survey, so only very faint ghosts remain. Plate fogging and processing marks will also affect the image magnitudes in a similar way, though such flaws are limited to small localised regions of the plate because of the quality control procedure (Section 2.2.1).

An estimate of the accuracy of individual COSMOS magnitudes can be made from the CCD sequences used in the inter-plate calibration procedure described below. The rms variation between COSMOS b_j magnitude ($m_{cos} + m_{sky}$) and the CCD b_j magnitude is $0^{\mathrm{m}}{\cdot}3$, for galaxies brighter than $b_j = 20{\cdot}5$. This estimate is an upper bound on the accuracy of individual COSMOS magnitudes for two reasons. First, because of the uncertainties in defining galaxy CCD magnitudes (Section 4.3.2d) and secondly because some bright high surface brightness galaxy magnitudes are under estimated by COSMOS. Removing these saturated HSB galaxies ($b_j \simeq 17$) will reduce the scatter in $b_j(\mathrm{COSMOS}) - b_j(\mathrm{CCD})$ and so reduce the estimate of the error on individual COSMOS magnitudes.

4.3 Inter-Plate photometry.

4.3.1 ADOPTED STRATEGY.

Inter-plate photometry is concerned with the definition of a consistent magnitude system for galaxies on different plates. It is important that the measured magnitude of each galaxy can be related to its actual magnitude, regardless of the plate used and

consequently that a consistent limiting magnitude can be defined across the whole survey. Given Eqn. 3.1, a uniform inter-plate system can be defined within the EDSGC by obtaining a estimate of the depth, *ie.* m_{sky} for each plate. Though the quality control procedure applied to SERC (J) plates (Section 2.2.2) ensures that badly underexposed plates are rejected, the rms variation in m_{sky} values between uncalibrated plates is $0^{\rm m}.18$, with a peak-to-peak difference of $0^{\rm m}.8$, between F344 and F352 (m_{sky} = 21·77 and 22·57 magnitudes/arcsec^2 respectively). $\overline{m_{sky}}$ is $22^{\rm m}.28$/arcsec^2.

In the SSGP reduction of the *Lick Catalogue*, the variation in plate depth was corrected for, using the number-density in cells common to the overlap regions between plates. A similar method could be employed in the EDSGC, using the limited number of published standard galaxy magnitudes and by estimating Δm_{sky} from images common to plate overlaps. Geller *et al.* (1984) and de Lapparent *et al.* (1986) discuss the problems associated with global "adjustment" or correction techniques. In particular they suggest that the use of overlaps can propagate the effects of a shallow plate. They cite the example of the low density region visible to the south of Coma in the SSGP maps (Fig. 1.1). They claim that this is "in part due to propagation of the effects of correcting for the bad [shallow] plate immediately west of the Coma plate." The plate matching procedure is also likely leave uncorrected systematic error where there is a concentration of deep or shallow plates, or where there is an "alternation in the plate depths."

The aim in the production of the EDSGC has been to obtain a CCD galaxy calibration sequence in B and V for every second field (Fig. 4.2) and so remove the need for global correction. CCD pointing centres were chosen to coincide with compact clusters found by visual inspection of the relevant SERC (J) plate. If required, both a bright and faint cluster was observed to ensure a wide range of galaxy magnitudes was covered ($16^{\rm m}.0$ to $21^{\rm m}.0$). On average 15 galaxies were observed per SERC field.

From Fig. 4.2 it can be seen that those fields without a CCD sequence have 4 edge

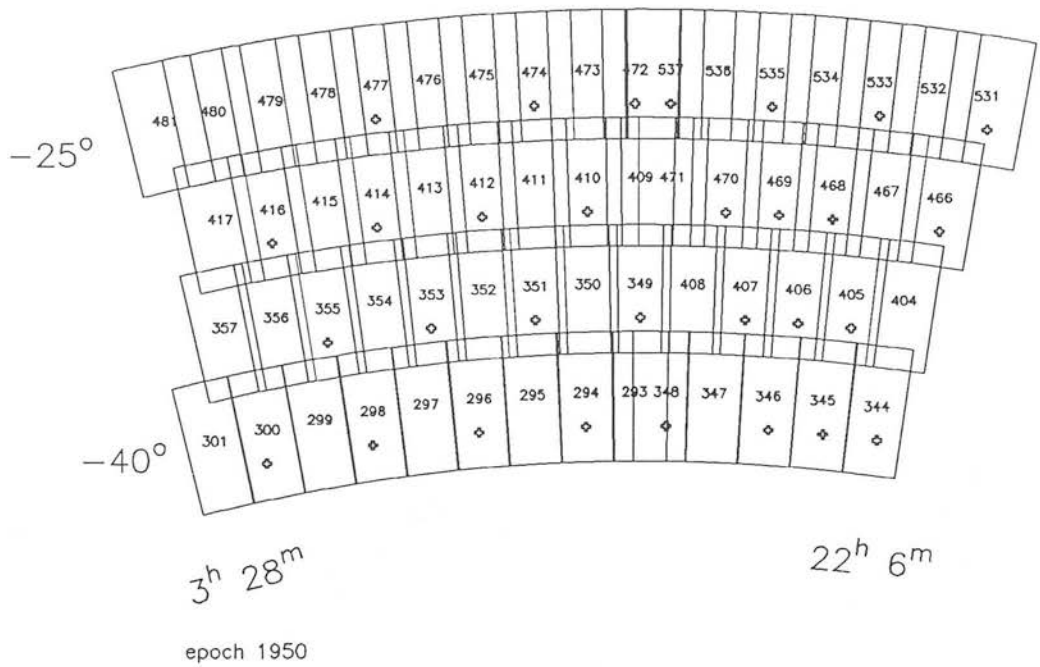


Figure 4.2: The layout of CCD calibration sequences in the EDSGC. Fields marked with a cross contain at least 1 compact cluster which has been observed in B and V.

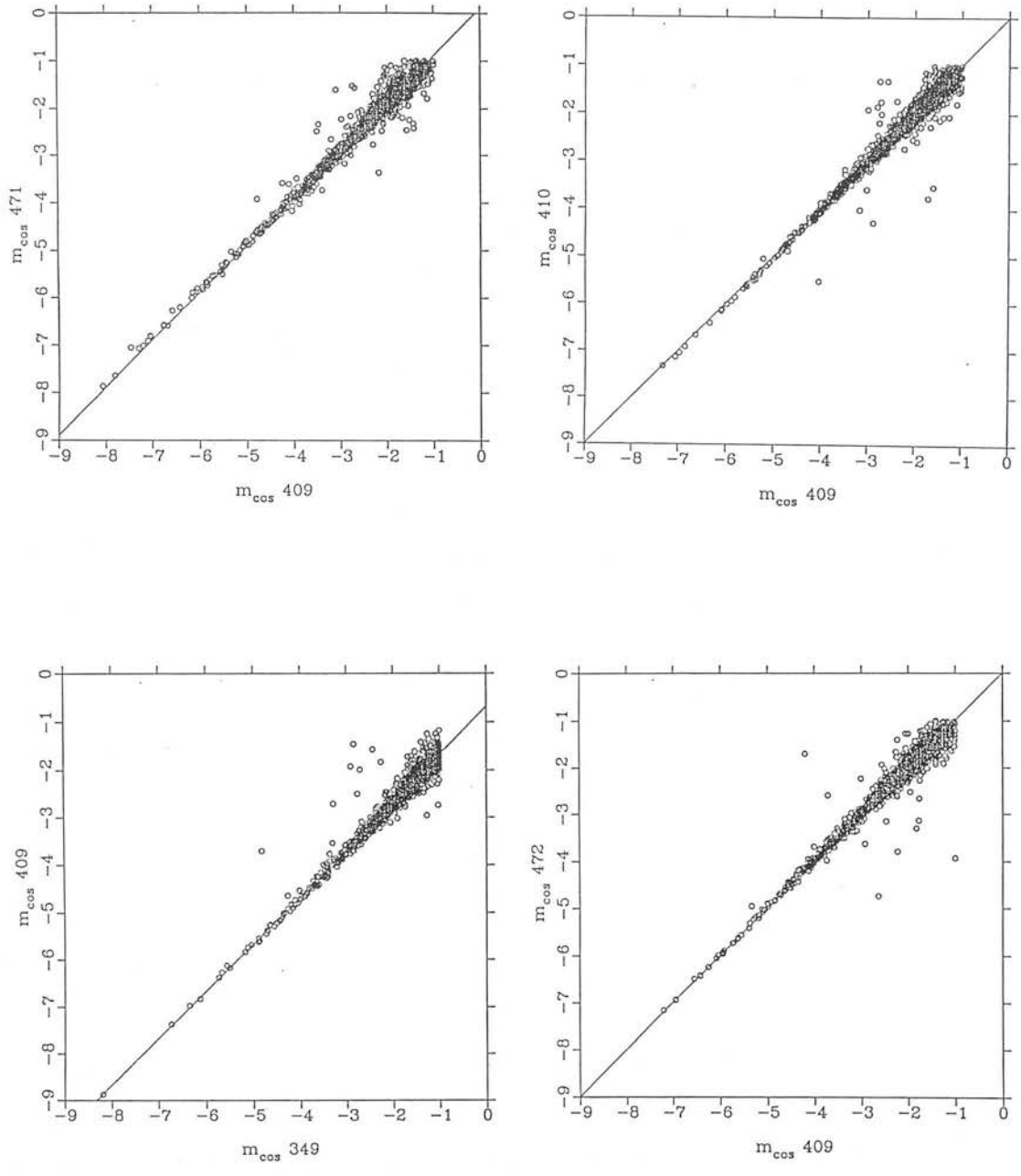


Figure 4.3: m_{cos} values on different different plates for objects common to overlaps between plates. Data from the 4 overlaps to F409 are shown, with least squares fits of 45° slopes. Standard errors on the gradient and intercept are typically $\sim$ 0·1 and $\sim 0^{m}.02$ respectively.

overlaps with 4 independently calibrated plates and so have 4 estimates of the plate m_{sky}. No reference to the number-density in the overlaps is required as the difference in plate depths, between overlapping plates (a and b), can be obtained directly from Eqn. 3.1. Δm_{sky} $(= m^a_{sky} - m^b_{sky})$ is the intercept of a plot of m_{cos} on plate b versus m_{cos} on plate a for all images common to both plates. Fig. 4.3 shows the 4 *overlap calibration plots* for F409. The slope of a least squares fit to the data in these and most other overlap calibrations is 45°, within the standard error. The number of images in an edge overlap ($\sim$ 1500) ensures accurate estimation of Δm_{sky} and because there are 4 estimates, any errant overlaps can easily be detected.

The one-in-two layout also provides a useful means of monitoring the accuracy of the CCD sequences using the diagonal overlaps between CCD calibrated plates, though the number of objects in some of these overlaps is small ($\sim$ few hundred). A further check is provided because of the 9 adjacent sequences around F406. Due to bad weather 3 CCD sequences (F479, F481, F357) are, at the present time, missing from the one-in-two layout. Special attention will be paid to their calibration.

4.3.2 CCD DATA.

4.3.2a *The Observations*

Most of the CCD sequences used in the EDSGC were obtained during 4 nights observing with the CCD Cassegrain Direct Imaging system on the 1·5m telescope at the Cerro Tololo Inter-American Observatory (CTIO) in Chile . Further data were obtained using the UCL–SAAO CCD camera system on the 1·0m telescope at the South African Astronomical Observatory (SAAO). It should be stressed that all observations used to calibrate the EDSGC were carried out in photometric weather and that any data from non-photometric nights has been discarded.

The CCD used on the 1·5m was the Observatory's RCA#5 chip (thinned and back-

illuminated). It has 320×512 pixels orientated in an E–W direction, subtending $\sim 3' \times 5'$ on the sky. The readout noise is $\sim$ 45 electrons/pixel or 9 ADU (analogue to digital units, conversion factor = 5·08 e^-/ADU). The saturation level is 32766 ADU, the bias is $\sim$ 204 ADU and the the dark-current is around 4 ADU/hour. The passbands for the CTIO B#3 and V#15 filters, used in the CTIO observations are shown in Fig. 4.4.

The CCD used in the UCL–SAAO camera system is also an RCA chip, with 320×512 pixels, orientated in a W–E direction. Because of the relatively high readout noise ($\sim$ 73 e^-/pixel, with = 11·5 e^-/ADU), the frames are pre-binned (2×2) to increase the signal-to-noise. The response begins to go non-linear around 22000 ADU and is saturated at 32768 ADU. The dark-current is typically around 20 ADU/hour. This chip suffers from a thresholding problem. At low signal levels (below 500 e^-/pixel) the change transfer efficiency drops dramatically and so a residual is left every time charge is transferred during the readout process. This results in images being trailed along the direction of charge transfer and a subsequent underestimation of aperture magnitudes. To ensure that the background level is high enough to avoid this effect, the CCD is *pre-flashed* prior to exposure. A 300 millisecond flash produces a background level of $\sim$ 550 e^-/pixel that is uniform to a few-percent. The Kron-Cousins B and V filters used during the SAAO Observations are shown in Fig. 4.4

The same observing procedure was followed for both sets of observations. Several bias frames, (*ie.* zero exposure frames), were taken at the beginning and end of each night and, at the SAAO, long pre-flash frames (30 second pre-flash) were taken in the same way. Focusing of the telescope was carried out at the start of each night, using the relevant procedure , and then checked several times during the night. CCD observations of *standard stars* of known magnitude (Graham 1982), at various airmasses, were made at the beginning, during and at the end of each night. Typically standard star frames would be taken after every third set (in B and V) of galaxy observations. The seeing and observing conditions are also monitored to ensure conditions remained photometric. Several flat fields (both twilight and dome flats in both passbands) were taken every

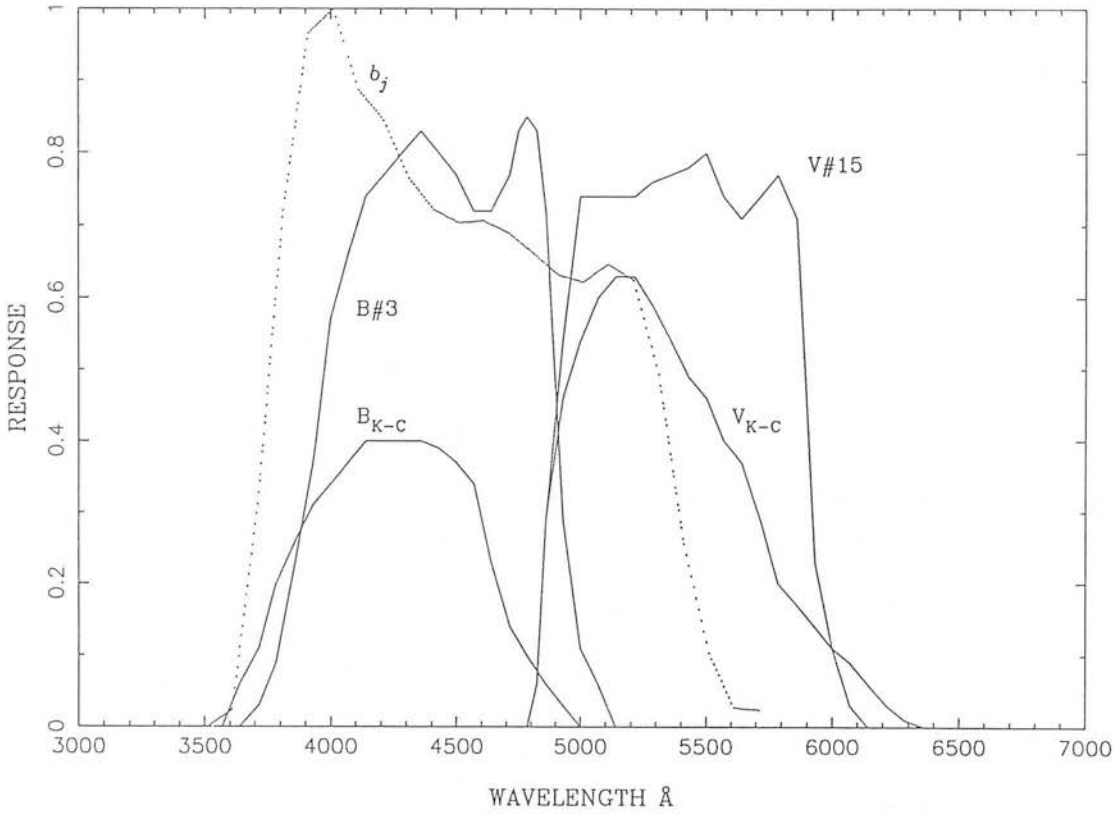


Figure 4.4: The filter passbands relevant to the calibration of the EDSGC: Kron-Cousins B and V filters (solid line). The CTIO's B#3 and V#15 filters (dashed line). The dotted line sketches the *normalised* combined filter transmission and emulsion response for b_j.

night and several 3 hour dark integrations were recorded during the observing runs.

4.3.2b *Initial Reduction Procedure*

The CCD data was reduced at the ROE using the Starlink FIGARO software. Some details of the reduction procedure differ from those described in the FIGARO documentation and so it is described briefly below. The term *science frame* refers to either a standard star or galaxy cluster observation.

(i) **Bias Subtraction.** The average frame of all bias frames, taken on a specific night is subtracted from all (pre-flash, flat and science) frames taken on the same night, such that for each pixel (P^{bs}) in the bias subtracted frame:

$$P^{bs} = P^{r} - \frac{m^{r}}{m^{b}} P^{b}. \tag{4.2}$$

The superscripts r and b refer to the raw and bias frames and m is the average value of the bias in the centre of the bias strip. Using the centre of the bias strip avoids problems associated with charge leakage from the image area to the bias strip.

(ii) **Pre-Flash Subtraction.** For the SAAO data the average of all long pre-flash frames (30 sec exposure), taken on a specific night, was subtracted from all (flat and science) frames taken on the same night. For each pixel (P^{fs}) in the flash-subtracted frame:

$$P^{fs} = P^{r} - \frac{t^{r}}{t^{pf}} P^{pf}, \tag{4.3}$$

with t equal the flash exposure time for both the raw and long pre-flash frames (superscript, pf).

(iii) **Dark Current Subtraction.** All the dark frames taken show no dark current in excess of that quoted above. This is negligible and was therefore ignored.

(iv) **fringe removal.** No fringes were observed on any of the science frames taken in either passband, at either observatory.

(iv) **Flat-Fielding.** The dome-flats taken at CTIO all showed significant gradients, that were not visible in the twilight flats or in the background of any science frame. All flat fielding was, therefore, carried out using the relevant twilight flat for each night. The flat-fielded science frames (superscript, ff) have:

$$\mathrm{P}^{ff} = \mathrm{P}^{s} \Big/ \frac{1}{m}\mathrm{P}^{f} \,, \tag{4.4}$$

with the superscripts s and f referring to the science frame and twilight flat respectively. Here m is the median value of the central region of the twilight flat. This is a better estimate of the sky value (mode) as it is less affected by bad pixels or cosmic-ray events.

4.3.2c *Standard Star Magnitudes*

Once bias subtraction and flat-fielding has been performed, the observations of standard stars with known magnitudes are used to empirically calculate airmass and colour corrections and to find the conversion factors from *instrumental magnitudes* (CCD counts) to object magnitudes.

For stars instrumental magnitudes are total magnitudes defined for each image by counting the ADU's within a series of flat, *ie.* un-weighted, apertures centred on the image centroid. A correction is applied to reduce all magnitudes to a 60 second exposure and background subtraction is carried out using the median count per pixel (m) in an annulus concentric with the apertures. The aperture magnitude is defined:

$$\mathrm{B}_{instr} = -2{\cdot}5 \log_{10} \left[\sum_{aperture} \left(\mathrm{P}^{ff} - m \right) \right] + 2{\cdot}5 \log_{10} \left(\frac{t}{60} \right) , \tag{4.5}$$

with t equal to the frame exposure time in seconds. V_{instr} is defined in the same way. The total magnitude for each image is the value the aperture magnitudes converge to, as the aperture radius increases. In crowded fields the aperture magnitudes occasionally converge and then increase again as other images come in to the aperture. In this

case the first converging value is taken for the instrumental magnitude. If the aperture magnitudes do not converge, the image is discarded. Providing a large enough annulus is taken, to reduce the effects of images, the instrumental magnitude is not changed significantly by altering the annulus position.

For the CTIO data the instrumental magnitudes can be converted to the Kron-Cousins system using:

$$\begin{aligned} \mathrm{B} &= \mathrm{B}_{instr} \\ &+ (0{\cdot}2 \pm 0{\cdot}02)(\mathrm{B}_{instr} - \mathrm{V}_{instr}) \\ &- (0{\cdot}2 \pm 0{\cdot}03)\sec(z) + (22{\cdot}96 \pm 0{\cdot}02) + \delta_b, \end{aligned} \tag{4.6}$$

$$\begin{aligned} \mathrm{V} &= \mathrm{V}_{instr} \\ &- (0{\cdot}03 \pm 0{\cdot}002)(\mathrm{B}_{instr} - \mathrm{V}_{instr}) \\ &- (0{\cdot}2 \pm 0{\cdot}03)\sec(z) + (22{\cdot}91 \pm 0{\cdot}02) + \delta_v. \end{aligned} \tag{4.7}$$

z is the zenith angle of the observation and δ_b and δ_v are small corrections dependent on the zenith point extinction for each night. $\delta_b = -0{\cdot}05$, $-0{\cdot}06$, $+0{\cdot}07$, and $\delta_v = -0{\cdot}03$, $-0{\cdot}03$, $+0{\cdot}05$, for each of the three nights at the CTIO. The CTIO B#3 filter has a significant colour correction to the Kron-Cousins system, due to its extra response longward of 4750Å (Fig. 4.4). The errors shown on each of these corrections are not necessarily independent. The repeatability of the standard star magnitudes are estimated as $\pm 0{\cdot}02$ and $\pm 0{\cdot}03$ in B and V respectively. These are the rms variations in the residuals (Kron-Cousins magnitude minus CCD magnitude) plotted against the time (UT) of observation (Fig. 4.5). This plot also shows there are no systematic changes to the CCD photometry during the course of observations taken in any one night.

For the SAAO data instrumental magnitudes can be converted to the Kron-Cousins system with:

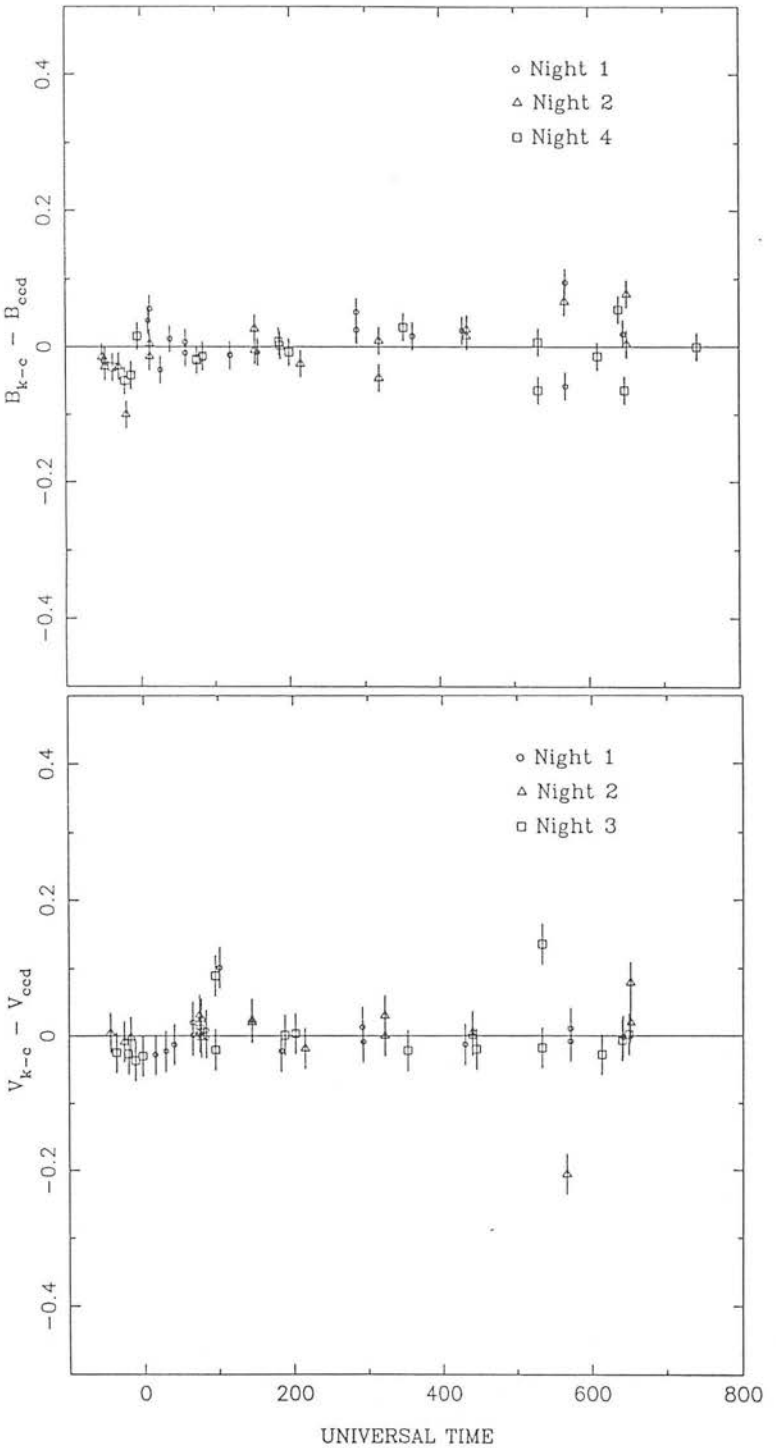


Figure 4.5: The residual difference between the standard star magnitudes (Kron-Cousins: B_{k-c} & V_{k-c}) and the CTIO CCD magnitudes (B_{ccd} & V_{ccd}, corrected with Eqn. 4.6 or Eqn. 4.7), plotted against the time of observation (Universal time in Mins, zero at 24:00:00).

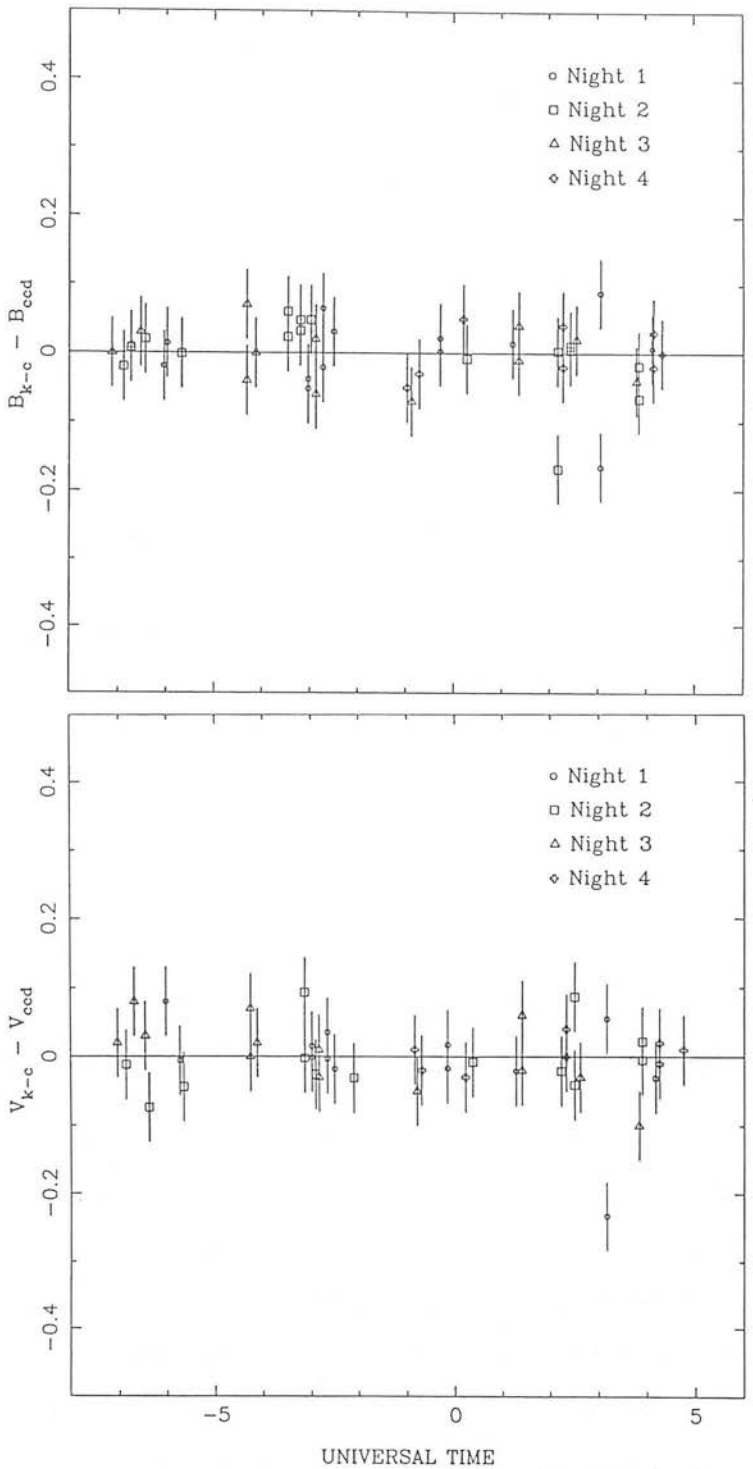


Figure 4.6: The residual difference between the standard star magnitudes (Kron-Cousins: B_{k-c} & V_{k-c}) and the SAAO CCD magnitudes (B_{ccd} & V_{ccd}, corrected with Eqn. 4.8 or Eqn. 4.9), plotted against the time of observation (Universal time in Hrs, zero at 24:00:00).

$$\begin{aligned} \mathrm{B} &= \mathrm{B}_{instr} \\ &+ (0{\cdot}054 \pm 0{\cdot}01)(\mathrm{B}_{instr} - \mathrm{V}_{instr}) \\ &- (0{\cdot}29 \pm 0{\cdot}1)\sec(z) + (25{\cdot}25 \pm 0{\cdot}03) + \delta_b, \end{aligned} \tag{4.8}$$

$$\begin{aligned} \mathrm{V} &= \mathrm{V}_{instr} \\ &- (0{\cdot}032 \pm 0{\cdot}01)(\mathrm{B}_{instr} - \mathrm{V}_{instr}) \\ &- (0{\cdot}23 \pm 0{\cdot}1)\sec(z) + (22{\cdot}91 \pm 0{\cdot}02) + \delta_v. \end{aligned} \tag{4.9}$$

Here the corrections for variation in zenith extinction from night to night are small, except for the first night. $\delta_b = -0{\cdot}39$, $0{\cdot}0$, $+0{\cdot}03$, $+0{\cdot}0$, and $\delta_v = -0{\cdot}21$, $0{\cdot}02$, $-0{\cdot}07$, $+0{\cdot}036$, for each of the four nights at the SAAO. There are only small colour corrections to account for the different response of the CCD and the photometer used to define the Kron-Cousins system. The repeatability of the SAAO standard star observations are estimated from Fig. 4.6 as 0·02 and 0·03 in B and V respectively. Again there are no systematic changes to the CCD photometry during course of observations taken in any one the night.

4.3.2d *Galaxy Magnitudes*

There is no one accepted method for calculating galaxy magnitudes. Different authors have used different techniques, *eg.* fixed aperture, metric aperture, total, profile or isophotal magnitudes. The magnitudes obtained from the CCD data have to be compared with the corresponding COSMOS calibrated magnitudes (m_{cos}). It is important that the methods used to calculate these magnitudes should be as consistent as possible. m_{cos} is an isophotal magnitude and so one possible method would involve applying the COSMOS analyser (Section 2.3.1) to calculate isophotal magnitudes from the CCD frames. There are two problems associated with this approach. Firstly the COSMOS detection threshold ($\sim 25^{\mathrm{m}}{\cdot}0/\mathrm{arcsec}^2$) varies from plate to plate, though this variation is limited as described in Section 2.3.1. Secondly the higher resolution of the CCD may result in a difference between the image defined by the analyser in each case. This may

be especially acute where images are merged. For images brighter than $b_j \simeq 20{\cdot}5$, m_{cos} is effectively total (Section 4.2) and so an alternative approach would involve calculating magnitudes in the same manner used for standard stars. However the galaxy CCD frames are centred on compact clusters, where images are close and often intra-cluster light is detected. This makes calculation of total magnitudes difficult as the aperture magnitudes often fail to converge.

The most consistent approach, which compares like with like, involves using the COSMOS IAM unweighted ellipse parameters (a_u, b_u, θ_u) to define an elliptical aperture, which is then applied to the CCD image (after correction for the different scales). This aperture is centred on the CCD image centroid and the galaxy instrumental magnitude calculated (in B and V) using Eqn. 4.5. This process can be further simplified (to work within the FIGARO package) by using a circular aperture with a radius equivalent to a_u. The difference in area between the elliptical and the circular apertures is small (average $b_u/a_u = 0{\cdot}7$ for CCD galaxies) and has little systematic effect (Section 4.3.3) because of the low light levels in the galaxy image outside the ellipse boundary. Variation in seeing is also not a problem as the images are extended and because the CCD observations were taken with seeing ($\sim 3''$) comparable to that in which SERC (J) plates are taken.

Once galaxy CCD instrumental magnitudes (B_{instr} & V_{instr}) have been obtained they are converted to Kron-Cousins B and V using the relevant conversion equation (Eqn. 4.6, 4.7, 4.8 or 4.9). The galaxy magnitude in b_j is then found using the colour equation of Blair and Gilmore (1982):

$$b_j = B - (0{\cdot}28 \pm 0{\cdot}04)(B - V) \tag{4.10}$$

This equation was calculated using standard stars with $-0{\cdot}1 \leq (B - V) \leq 1{\cdot}6$, so is appropriate for the galaxy sequences, where $\overline{(B - V)} \simeq 1{\cdot}0$.

CCD magnitudes in B, V and $b_j$have been obtained for $\sim$ 400 galaxies in order to calibrate the EDSGC. These will be published at a later date, together with the radius

of the aperture used, the galaxy celestial coordinates (epoch 1950), and the relevant ESO/SERC field. It should be stressed that these magnitudes are neither strictly total magnitudes or strictly isophotal magnitudes, though they approximate both well. The method used to calculate these magnitudes is designed specifically to accurately calibrate the light recorded (on the photographic plate) within the COSMOS IAM ellipse and so accurately calibrates m_{sky}.

4.3.3 CALIBRATION DETAILS

At the time of completion of this thesis, the EDSGC consists of a 35-plate mosaic (described fully in Section 5.3), which has been calibrated, as far as possible, in the manner described in the previous sections. Information from 18 CCD sequences was available for this calibration, so that most of the plates are covered by the one-in-two arrangement of CCD sequences. The data for the 3 sequences needed to complete this arrangement has not yet been reduced, so there are 3 plates (F296, F353, F355) where the calibration is performed using overlaps to plates without CCD sequences. The estimates of the accuracy of the calibration procedure employed in the EDSGC are (at the moment) based on these sequences and this mosaic. Final calibration details will be published elsewhere (together with the CCD magnitudes of individual galaxies), when the full 60 plate catalogue is completed.

Fig. 4.7 shows galaxy b_j versus m_{cos} plots for 6 sample plate sequences used in the calibration of the survey. Least-squares linear fits are shown, together with a least squares fit of a line with a 45° slope. The standard errors on the gradient and intercept (m_{sky}) are typically 0·4 and $0^{\mathrm{m}}.1$. Fig. 4.8 shows the galaxy CCD magnitudes (b_j(CCD)) from all 18 sequences plotted against the magnitudes predicted from the calibrated survey, *ie.* $b_j(\mathrm{COSMOS}) = m_{cos} + m_{sky}$. The value of m_{sky} is the value of the intercept of the 45° slope fit to the data for each sequence. A least squares fit to the data combined from all the sequences has a slope of $1{\cdot}08 \pm 0{\cdot}01$.

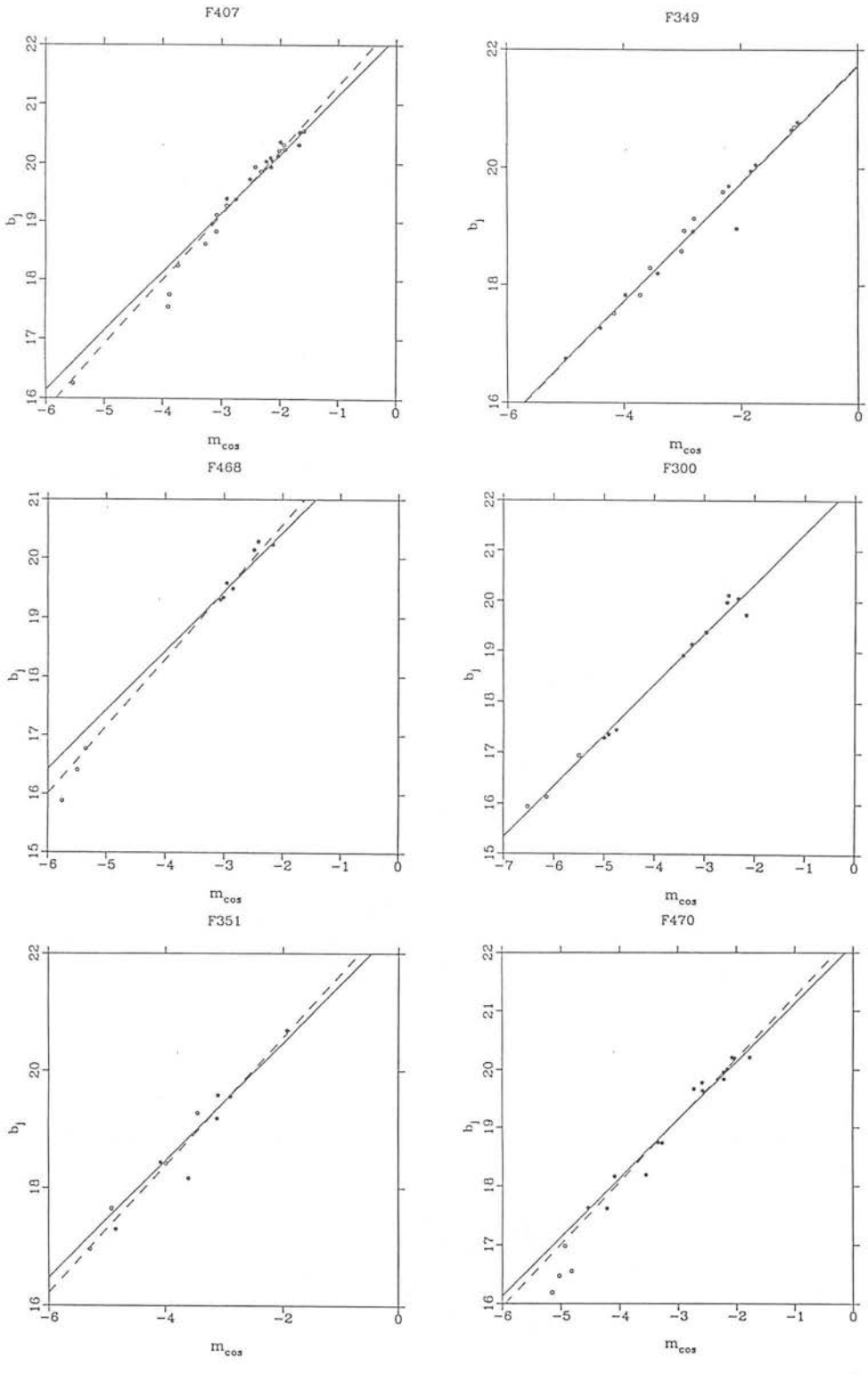


Figure 4.7: A sample of the CCD galaxy calibration sequences used to calibrate the 35-plate mosaic described in the text. Dashed lines are least squares fits, while solid lines are least squares fits with 45° slopes. Galaxies from different CCD frames are plotted with different symbols.

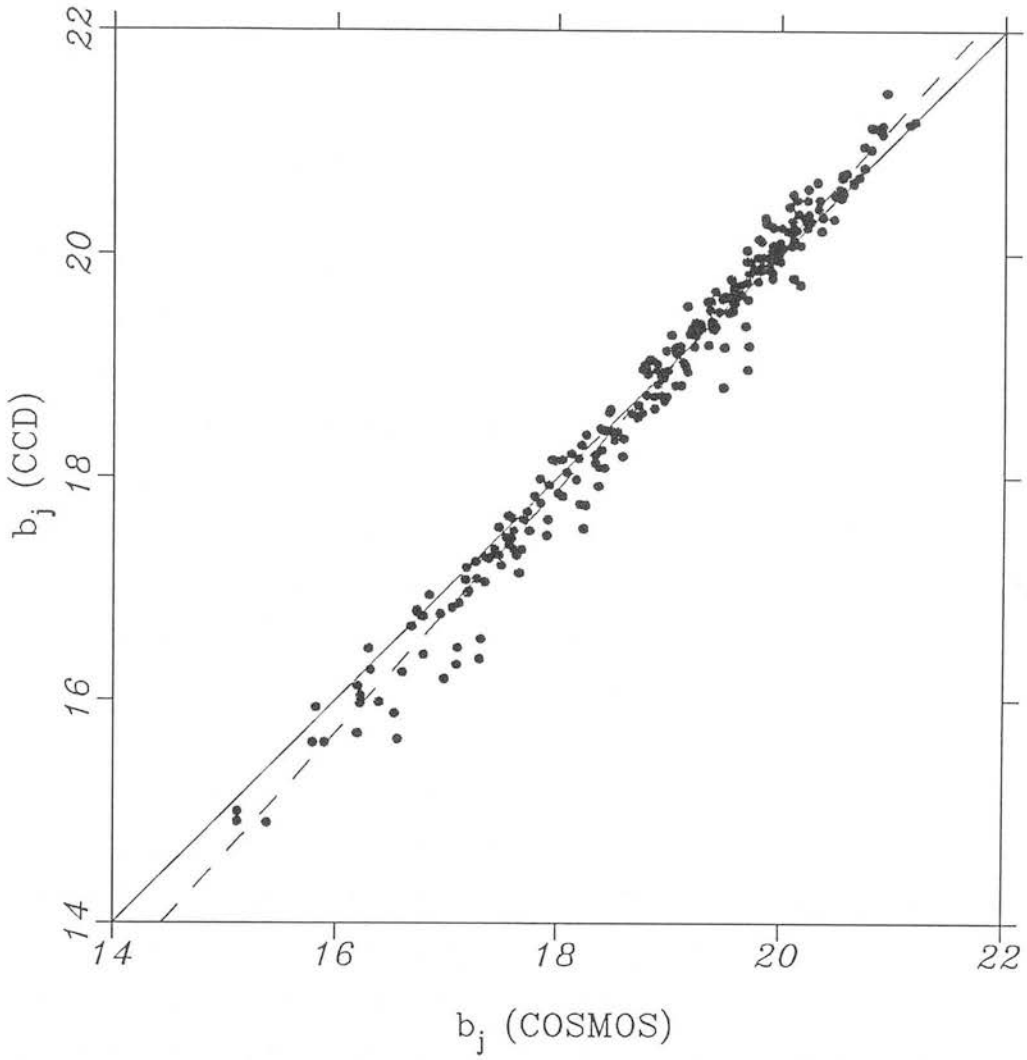


Figure 4.8: A comparison of Galaxy CCD magnitudes (b_j(CCD)) and the magnitude predicted by COSMOS, *ie.* b_j(COSMOS)= $m_{cos} + m_{sky}$, where m_{sky} is calculated from a 45° slope fit to the relevant calibration sequence. The least squares fit slope (dashed line) is 1·09.

The steeper slope for the calibration sequences indicated by the combined data is visible in some (*e.g.* F407, F351), though not all of the individual sequences (*e.g.* F300, F349) shown in Fig. 4.7. In some of the sequences, it is just the bright galaxies that lie below the 45° slope (*e.g.* F470A, F468). A steeper slope than expected from Eqn. 3.1 suggests that m_{cos} underestimates the magnitude of brighter images compared to the CCD values. There are several possible explanations for this effect:

(i) The use of circular apertures to define the CCD instrumental magnitudes could over estimate the (non-circular) isophotal magnitude defined by the COSMOS analyser. This effect would obviously depend on the morphology of individual galaxies, with highly elongated galaxies having a greater difference between the CCD and COSMOS predicted magnitude. Fig. 4.9a plots $b_j(\mathrm{CCD}) - b_j(\mathrm{COSMOS})$ against $-2.5\log(b_u/a_u)$, and Fig. 4.9b plots $-2.5\log(b_u/a_u)$ against $b_j(\mathrm{CCD})$. The least squares gradients for the data in each of these plots are $0{\cdot}25 \pm 0{\cdot}08$ and $0{\cdot}07 \pm 0{\cdot}02$ respectively. These figures indicate that there is a marginal tendancy for circular apertures to overestimate the isophotal magnitude and that the effect is greater for bright galaxies. Combining the empirical relations derived from these figures predicts that $b_j(\mathrm{CCD}) \propto (1{\cdot}02 \pm 0{\cdot}01) \times b_j(\mathrm{COSMOS})$. Thus the effect is small compared with that in Fig. 4.8.

(ii) COSMOS sky background following is filtered on scales $\sim$ the solid angle subtended by the CCD. Sky background estimates for the CCD magnitudes are, of course, taken within the CCD. In Fig. 2.4 background changes on scales $<$ the filtering scale can been seen, especially in the vicinity of large objects or dense regions. Differences between COSMOS and the CCD sky background subtraction procedures could result in an underestimation of $b_j(\mathrm{COSMOS})$. Such an effect may be responsible for the steeper slope seen in *e.g.* F470 and F468, where the bright and faint objects come from different CCD frames, but does not explain the more general trends seen in *e.g.* F407 or F351.

(iii) Saturation of the COSMOS magnitudes is a possible cause of the steeper slopes. This is the case for stars where m_{cos} drastically underestimates the magnitude for

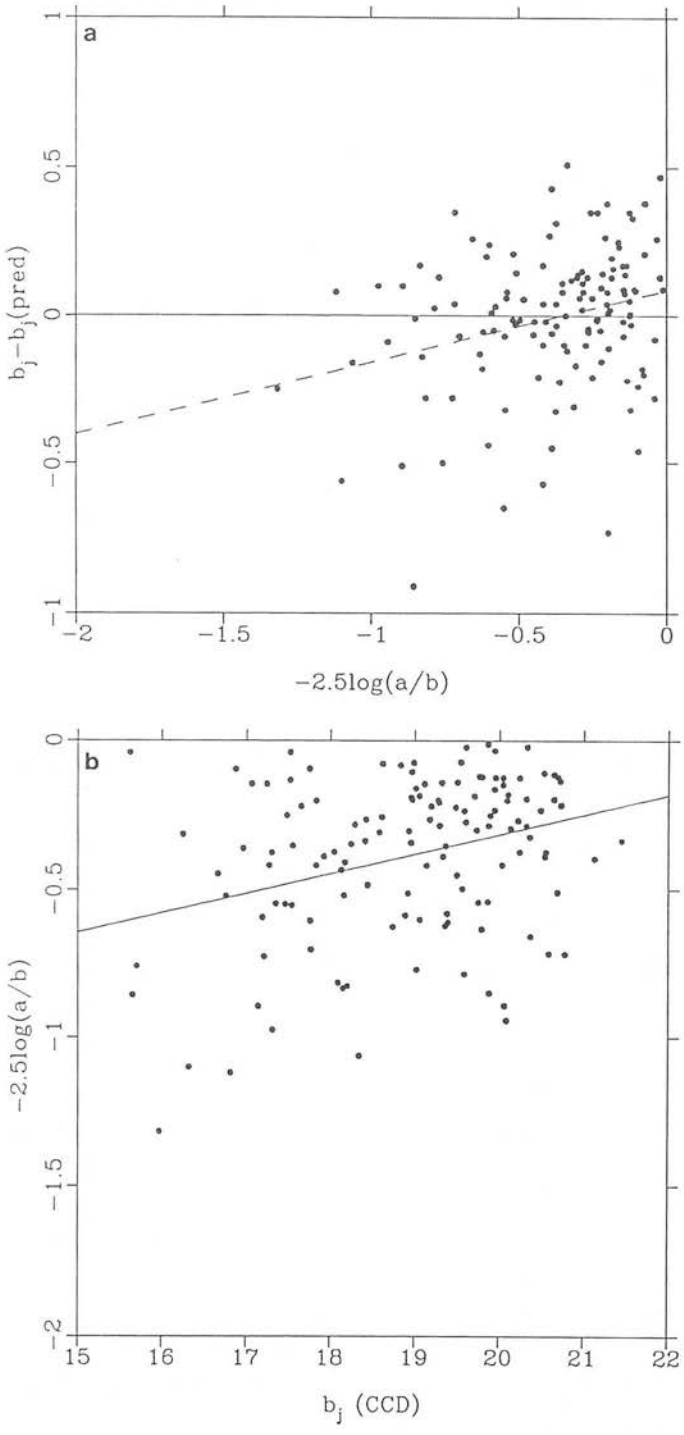


Figure 4.9: The effects of galaxy ellipticity on the CCD magnitude. (a) b_j(CCD) − b_j(COSMOS) plotted against $-2{\cdot}5\log_{10}(a_u/b_u)$. the least squares fit slope is $0{\cdot}25 \pm 0{\cdot}08$. (b) $-2{\cdot}5\log_{10}(a_u/b_u)$ plotted against b_j(CCD). The least squares fit slope is $0{\cdot}07 \pm 0{\cdot}02$.

$b_j < 19{\cdot}0$ (Fig. 3.2). The effect seen here for galaxies is considerably smaller than that noted for stars in Section 4.2. Here $m_{cos} = -4$ is underestimating the magnitude of $b_j = 18$ galaxies by $\sim 0^{\rm m}{\cdot}2$, whereas $m_{cos} = -4$ underestimates a stellar mag by $\sim 4^{\rm m}{\cdot}0$. These values are consistent with a few pixels in the core of the galaxy image being saturated and is not unexpected considering the smooth transition between stars and galaxies in the $\log_{10}(\mathcal{A})$ classification plane (Fig. 3.1). Such partial image saturation will be a function of image morphology, with high surface brightness galaxies (ellipticals) being affected more than low surface brightness (spirals). This would explain why the distribution of points about the linear fit in Fig. 4.8 is skew, *ie.* galaxies below the line are in general further away than those above the line.

The most likely and consistent cause of the non-45° slopes in Fig. 4.8 is (iii), *ie.* partial image saturation, though in some cases there may be problems with the background estimation(ii). The effect has been recently observed by Metcalfe *et al.* (1989), who draw the same conclusion. Steeper slopes are expected to be more predominant in the CCD calibration sequences as they are centred on reasonably compact clusters, where there are more HSB galaxies and where the backgrounds may be higher due to intra-cluster light. Given this, non-45° slopes are not expected to be appropriate for other regions of the plates, where galaxies are in less crowded environments and are predominantly of lower surface brightness. It is important to stress that, using 45° slope fits to the CCD data and to the overlap data (which, in the main have 45° slopes to within the standard error), it is possible to define a limiting magnitude that is consistent to within $0^{\rm m}{\cdot}05$.

The evidence at present is consistent with using Eqn. 3.1 to define the magnitude of galaxies in the EDSGC, but with a caveat that this value may underestimate the magnitude of high surface brightness galaxies. Further investigation of this effect will be possible once all the sequences (Fig. 4.2) are reduced, as there are sequences in some overlap regions between plates and also regions of the survey with a high concentration of CCD sequences (*eg.* around F406). This problem again emphasises the need for care

with photographic photometry. Extrapolating the calibration of a large survey from a limited amount of CCD data will result in erroneous galaxy magnitudes over sizeable areas of the survey.

The final m_{sky} values adopted for each plate in the 35-plate mosaic are listed in Table 4.1, together the error associated with each estimate. For plates calibrated with CCD sequences, the standard error on the intercept (m_{sky}) value is quoted, while for plates calibrated via overlaps, the standard error between the 4 estimates of m_{sky} is quoted The consistency of the survey limiting magnitude is estimated from the rms difference between the estimates from each overlap and the adopted m_{sky}. For plates calibrated using the 1-in-2 CCD sequence arrangement the accuracy on m_{sky} (and therefore m_{lim}) is $0^{\rm m}\!\cdot\!05$. Introducing plates calibrated by overlaps decreases the accuracy to $0^{\rm m}\!\cdot\!75$.

A preliminary calibration of the survey was carried out using just 4 galaxy sequences and extrapolating the calibration across large regions using the plate overlaps. Typically the calibration of a plate was carried out using different sets of overlaps and then the average m_{sky} was assigned to the plate. This preliminary calibration has been compared to the final 'one-in-two' calibration. The average difference in m_{sky} values between this preliminary calibration and the final calibration is $0^{\rm m}\!\cdot\!1$, with a peak difference of $0^{\rm m}\!\cdot\!52$. More dangerously 20% of the plates in the preliminary calibration are in error by $> 0^{\rm m}\!\cdot\!2$ ($\equiv$ 30% number density changes).

4.4 Summary

The photometric techniques used in the EDSGC remove the need for number-density corrections as a function of plate or survey position as used in the SSGP reduction of the *Lick Catalogue*. These global correction mechanisms are suspect because they may propagate systematic errors from one defective plate to the surrounding region and do

Field	m_{sky}	Error	Field	m_{sky}	Error
F293	22·16	±0·13	F356	22·47	±0·08
F294	22·31	±0·04	F404	22·36	±0·06
F295	22·48	±0·04	F405	22·22	±0·05
F296	22·49	±0·09	F406	22·04	±0·01
F297	22·29	±0·10	F407	22·14	±0·05
F298	22·41	±0·05	F408	22·19	±0·04
F299	22·24	±0·14	F409	22·48	±0·05
F300	22·34	±0·05	F410	22·54	±0·06
F344	21·77	±0·06	F411	21·92	±0·04
F345	22·06	±0·05	F412	22·44	±0·10
F346	22·19	±0·09	F413	22·42	±0·15
F347	22·33	±0·02	F414	22·29	±0·05
F348	22·25	±0·06	F415	22·40	±0·01
F349	21·77	±0·05	F466	22·10	±0·05
F350	22·14	±0·02	F467	22·16	±0·08
F351	22·48	±0·10	F468	22·29	±0·09
F352	22·57	±0·01	F469	22·20	±0·01
F353	22·36	±0·03	F470	22·13	±0·05
F354	22·54	±0·03	F471	22·38	±0·03
F355	22·39	±0·03			

Table 4.1: m_{sky} values (mag/arcsec^2) for plates in the EDSGC, together with the associated error. The method of calibration and error estimation is detailed in the text. All values available at the time of writing are listed.

not fully remove spurious structure from the raw data. A preliminary calibration of the EDSGC, with a limited number of CCD sequences and extrapolation of the calibration over large regions using plate overlaps, is found to result in 20%of the plates having an erroneous calibration of $> 0^{\mathrm{m}}.2$, which is equivalent to $> 30\%$ number density changes. Such changes would introduce unacceptable errors in the calculation of the correlation function.

COSMOS calibrated image magnitudes, m_{cos} are linearly related to the image b_j magnitude providing that the image is non-saturated and that background changes are equivalent to sensitivity changes. For the majority of galaxies brighter than $b_j = 21{\cdot}0$ and fainter than $b_j = 14{\cdot}0$, both these constraints apply and $m_{cos} = b_j - m_{sky}$. A caveat to this, is that high surface brightness galaxies (cluster ellipticals) may be partially saturated and so m_{cos} will underestimate b_j, though the magnitude of the effect ($\sim 0^{\mathrm{m}}.4 at b_j \sim 18{\cdot}0$) depends on the image profile. In general COSMOS magnitudes for individual galaxies are accurate to $< 0^{\mathrm{m}}.3$, while systematic variations in inter-plate photometry are estimated to be $\lesssim 0^{\mathrm{m}}.03$.

The uniformity of inter-plate photometry and hence the survey limiting magnitude is set using a CCD sequence for every second field. For each plate with a sequence m_{sky} is estimated directly, while those plates without sequences have 4 overlaps with CCD calibrated plates. These overlaps provide 4 independent estimates of the plate m_{sky}, which can be compared to check the CCD photometry on any of the surrounding plates and to detect any erroneous overlaps. This method has been used to calibrate a 35-plate mosaic, from which the accuracy of the inter-plate photometry is estimated to be $0^{\mathrm{m}}.05$ rms.

The variations in image photometry discussed here, compare favourably with the limits discussed in Section 1.4. A further advantage of the calibration strategy, adopted in the EDSGC, is that a strict quality control can be carried out. The one-in-two CCD arrangement can detect problems with individual CCD sequences and any erroneous

overlaps, thus preventing propagation of calibration errors across the survey.

Chapter 5

Construction and Application of the EDSGC

5.1 Introduction

The previous two chapters have discussed the methods used in the EDSGC, to classify images from the raw COSMOS scans (Chapter 2) and to calibrate the photometry of the resulting galaxies. As it has been described so far, the EDSGC consists of 60 separate plate datasets, though a single catalogue, covering a contiguous region of the sky, is required to investigate the large scale-structure.

Producing a single contiguous *plate-mosaic* is problematic, because of the plate overlaps. Data in the overlaps is repeated in two (or three) other plate datasets. This repeated data must be removed before appending the individual datasets, otherwise it will give rise to a spurious cellular structure. Removing it is non-trivial however, because the extent of the overlap region varies for each pair of plates. Leaving the EDSGC as a set of individual plate datasets is not acceptable, as this will limit the maximum angular scale over which the correlation function can be determined. Section 1.2.5 discusses limitations on estimates of $\omega_{gg}(\theta)$ from such small datasets.

The final chapter of this thesis begins (Section 5.2) with a description of the techniques used to construct a single galaxy catalogue from the 60 constituent datasets, together with the mechanics of removing spurious objects and regions that were identified in Section 3.5.

At the time of completing this thesis, about half of the final catalogue has been constructed. All 60 plates have been star-galaxy classified, with most plates combined

into several many-plate mosaics. CCD Data for 18 sequences have so far been reduced. Given the need for a one-in-two CCD arrangement (Section 4.3.1), the distribution of the available sequences allow the construction and calibration of a 35-plate mosaic covering $\sim$ 1000 deg^2. Section 5.3 describes this database in detail and includes a *subjective* account of the large-scale structure seen in this mosaic (based on various maps presented either here or in Appendix B), together with estimates of the overall homogeneity of the EDSGC. Some problem areas, where there appear to be spurious structures, are also noted and comparisons are made with other work.

Preliminary application of the catalogue is discussed in Section 5.4. Here estimates of the galaxy number-magnitude counts (Section 5.4.1) and the two-point angular correlation function (Section 5.4.2) are made, based on the data available in the 35-plate mosaic. Reference is made to Heydon-Dumbleton *et al.* (1989) and Collins *et al.* (1989) which contain estimates of these statistics from a small 6-plate region of the EDSGC. This region was constructed in the same manner as the 35-plate mosaic. Several projects have been started in collaboration with Chris Collins (Royal Observatory) and Luigi Guzzo (ESO, Garching and Osservatorio di Brera), for which telescope time has been awarded by the European Southern Observatory. These projects are described in Section 5.4.3, together with other areas of astronomy where the EDSGC may provide useful data.

The results presented in this chapter are based on the data available at the time of writing. Final results will be published, together with estimates of residual systematic errors and detailed models of their effect on clustering statistics, when the full 60 plate catalogue is completed.

5.2 Plate Drilling and Mosaic Construction

Section 3.5 discussed possible contamination of the EDSGC due to the break up of large images, with the resulting images being classed as galaxies because of their non-Gaussian profiles. The most frequent source of this problem is bright stars that have a bright halo (Section 2.9 and UKSTU Handbook), though other possible sources are dense star clusters, large galaxies, satellite trails, stellar corrector ghosts, extremely bright stars ($b_j < 6$) and their associated emulsion-corrector and filter-corrector ghosts. The reconstitution of stellar filter ghosts and diffraction spikes that are broken up by the deblending software was discussed earlier in Section 2.3.4.

The break up of stellar halos is countered by *drilling* all stellar images with $m_{cos} \lesssim -9$. This involves removing all images within the halo radius (6·7 arcmin) of the star's intensity weighted centroid. The exact magnitude, brighter than which stars are drilled, varies from plate to plate and is set by the distribution of stars in the $\mathcal{G}$ parameter classification plane. The sinusoidal shape of the stellar locus in Fig. 3.1a is caused by the combined growth of diffraction spikes (which lower $\mathcal{G}$) and the increasing brightness of the filter, halo and then corrector ghosts (which increase $\mathcal{G}$). Most problems associated with the break up of halos can be avoided by drilling stars that are brighter than the second (brightest) minima seen in Fig. 3.1a (typically $m_{cos} \lesssim -9$). Appendix A lists the number of such stars drilled on each plate of the 35-plate mosaic.

Other objects responsible for spurious structures are removed by visually inspecting the distribution of galaxies across each plate dataset. These structures either appear as extremely dense clusters (*e.g.* star clusters and emulsion-corrector/filter-corrector ghosts) or as holes surrounded by a ring of objects (*e.g.* galaxies and extremely bright stars). The centres of these objects are defined subjectively and then drilled with a large enough radius to remove all the spurious objects. A list of all such regions removed from the 35-plate mosaic is given in Appendix A.

After the plates have been drilled and visually inspected, they are combined together into the plate-mosaic. The plate datasets used in the mosaic include all objects with $m_{cos} < -1$. This limit, which is fainter than the completeness limit ($b_j = 20$) and corresponds to a varying limit in b_j, was chosen for several reasons. First, it allows changes to be made to the calibration of individual plates in the final catalogue without requiring reconstitution of the mosaic. Secondly, on most plates there is some separation (Fig. 3.1c) between stars and galaxies for images with m_{cos} between -1 and -2. Data at these magnitudes ($b_j \sim 20 - 21$) is incomplete (Section 3.3.2), but may be of use in some applications of the EDSGC. Data fainter than $m_{cos} = -1$ is increasingly contaminated by noise images and so removing it reduces the number of objects per plate to a manageable level.

The construction of the plate-mosaic involves determining the extent of the overlap region between each pair of plates and removing all repeated galaxies from one of the datasets. The western edge of each plate suffers the most from desensitisation and vignetting and slight over density of galaxies was detected here in Fig. 3.10. Because of this, galaxies from the plate where the western edge forms part of the (E–W) overlap will be removed. COSMOS scans the plates upside down (south at the top), so the west corresponds to the left edge of the plots in Appendix B.

An excess ($\times 2$) of faint noise images has been discovered on the north and western edges of the scan regions. This excess is in the raw data, so it is not connected with the image classification or calibration procedures, and is confined within the width of a scan-lane. It is not seen at the south (top) or east (right) edges of the scan region and so may be due to miscalculation of the background intensity at the start of a scan. Objects from the plate where the northern edge forms part of a (N–S) overlap will, therefore, be removed.

The mosaic construction is extremely complicated (and time consuming), so this section only includes brief details of the general procedure used to append the plates together.

Data manipulation is carried out using the STARLINK HAGGIS database software. The area (in deg^2) of each plate used in the mosaic is listed in Appendix A, while the extent of each area can be seen in the plots in Appendix B . To facilitate removal of the north and west overlap regions, the mosaic is constructed by starting from the north-west of the survey region and moving south-east.

Given that several plates have been combined (mosaic'*ed*) together, an additional plate is appended (to the south or east side) as follows. The additional plate dataset is *paired* (Section 2.3.3) with the existing mosaic, on the basis of image RA and DEC coordinates with a 6 arcsec pairing box. Galaxies in the additional dataset that are not in the existing mosaic (*rejected* objects) are either, (i) galaxies not in the overlap region, or (ii) noise images that lie in the overlap region. Because of the number density difference between (i) and (ii), there is a clear boundary between these two regions. The (x, y) distribution of rejected objects is plotted on an interactive graphic display and a polygon defined around area (i) using a cursor. Those galaxies lying inside this polygon (the non overlap region) are selected and appended to the existing mosaic. Galaxies outside the polygon lie in the west or north overlap regions and are therefore discarded. To speed up the production of the full mosaic, several small plate mosaics (consisting of $\sim$ 10 plates usually in strips of constant DEC) are constructed first and then these are combined together (across N–S overlaps) in a similar manner.

5.3 A 35 Plate Mosaic

5.3.1 DESCRIPTION OF THE MOSAIC

To date a 35-plate mosaic covering $\sim$ 950 deg^2 has been constructed from single plate datasets. Image classification and calibration has been carried out as described in Chapters 3 & 4. The following details are listed for each plate in Appendix A; the field and plate numbers, the plate grade, the percentage stellar contamination predicted

by the automated classification algorithm (using both $\log_{10}\mathcal{A}$ and $\mathcal{S}$ classifiers), m_{sky} values, the number of stars drilled and, finally, the area subtended in the mosaic by each plate dataset. Other objects and regions that have been removed from this mosaic are listed at the end of this Appendix.

The distribution of ($b_j < 19{\cdot}5$) galaxies across each individual plate are show in Appendix B. These plots only show the area of the scan region that has been appended to the plate mosaic, *ie.* overlaps have been removed, and have south at the top. Appendix C shows differential number–magnitude counts ($N/0^m{\cdot}5/deg^2$) for each of the plate datasets. These counts are for all galaxies brighter than $b_j = 20.0$ in the regions plotted in Appendix B. Points fainter than $b_j = 20$ are caused by rounding errors and should be added back into the bin centered on $b_j = 19{\cdot}75$.

The 35-plate mosaic covers ~ 950 deg^2, contains 293977 galaxies brighter than $b_j = 20{\cdot}0$ (including a 10% stellar contamination) and $\sim 9 \times 10^5$ galaxies with $m_{cos} < -1$. These figures imply that the final 60 plate EDSGC will contain $\sim 5 \times 10^5$ galaxies to $b_j = 20{\cdot}0$ and $1{\cdot}5 \times 10^6$ with $m_{cos} < -1$. 27 parameters (Table 2.4) are stored for each image, so the present dataset occupies $\sim$ 90 MBytes. These numbers should be compared with the *Lick Catalogue* that contains approximate positional information only on 8×10^5 galaxies. None of the differential number–magnitude counts shown in Appendix C show the incompleteness turn-over discussed in Section 3.3.2 or seen in Fig. 2.5. It may, therefore, be possible to extend the completeness limit of the final EDSGC to fainter than $b_j = 20$, with the final depth being set by the incompleteness limit of the shallowest plate.

Various sources of residual systematic errors that affect the overall homogeneity of the EDSGC in terms of limiting magnitude (m_{lim}) or spurious number density gradients, have been discussed in earlier chapters. They have been estimated as follows:

(i) Intra-plate number density variations, due to the combination of systematic errors in intra-plate photometry and image classification and variations in the percentage

of noise objects, are estimated (Section 4.2 and Section 3.4.4) at $\frac{\Delta N}{N} \simeq 3\%$ rms ($\equiv \Delta m_{lim} \simeq 0^{m}.022$ rms).

(ii) Variations in the efficiency of image classification from plate-to-plate result in $\sim 3\%$ rms inter-plate number density variations (Section 3.4.3). This estimate includes a contribution from the variation in the percentage of noise images on different plates.

(iii) Inter-plate calibration errors, *ie.* Δm_{sky} ($\equiv \Delta m_{lim}$), are estimated at $0^{m}.05$ rms ($\frac{\Delta N}{N} \simeq 6{\cdot}9\%$ rms).

Combining (ii) and (iii) gives the resultant plate-to-plate systematic error in number density as $\sim 7{\cdot}6\%$ ($\equiv \Delta m_{lim} \simeq 0^{m}.055$ rms), while the point-to-point systematic variation is estimated at $\sim 8\%$ ($\equiv \Delta m_{lim} \simeq 0^{m}.059$ rms) by combining (i), (ii) and (iii). This figures compare well with the limit of $\Delta m_{lim} \lesssim 0^{m}.05$ rms set by Geller *et al.* (1984). The point to point variation in the SSGP reduction of the *Lick Catalogue* is 11%($\equiv \Delta m_{lim} \simeq 0^{m}.08$ rms). Further to this the peak discrepancies will be less in the EDSGC, than in the *Lick Catalogue*, because the 1-in-2 CCD arrangement prevents the propagation of calibration errors from plate to plate.

Various maps of the large-scale structure are shown in Fig. 5.1 to Fig. 5.6. The bins in the greyscale maps subtend the same solid angle on the sky at all declinations. This solid angle ($10' \times 10'$) is the same as that used in the raw Shane & Wirtanen counts and in the SSGP maps (Fig. 1.1). The term *positive* implies that white pixels have the highest number of galaxies per pixel, while the term *negative* implies that black pixels have the highest number density.

Fig. 5.5 plots the distribution of bright galaxies ($b_j < 17{\cdot}0$) superimposed on the plate centres and boundaries plotted in Fig. 2.2.4. This is not an accurate projection, as the DEC scale has been stretched by a factor of ~ 2. Consequently structures seen in the galaxy distribution are also stretched along the DEC (vertical) axis. It should be noted that the western and northern overlap regions of each plate are removed in the mosaic

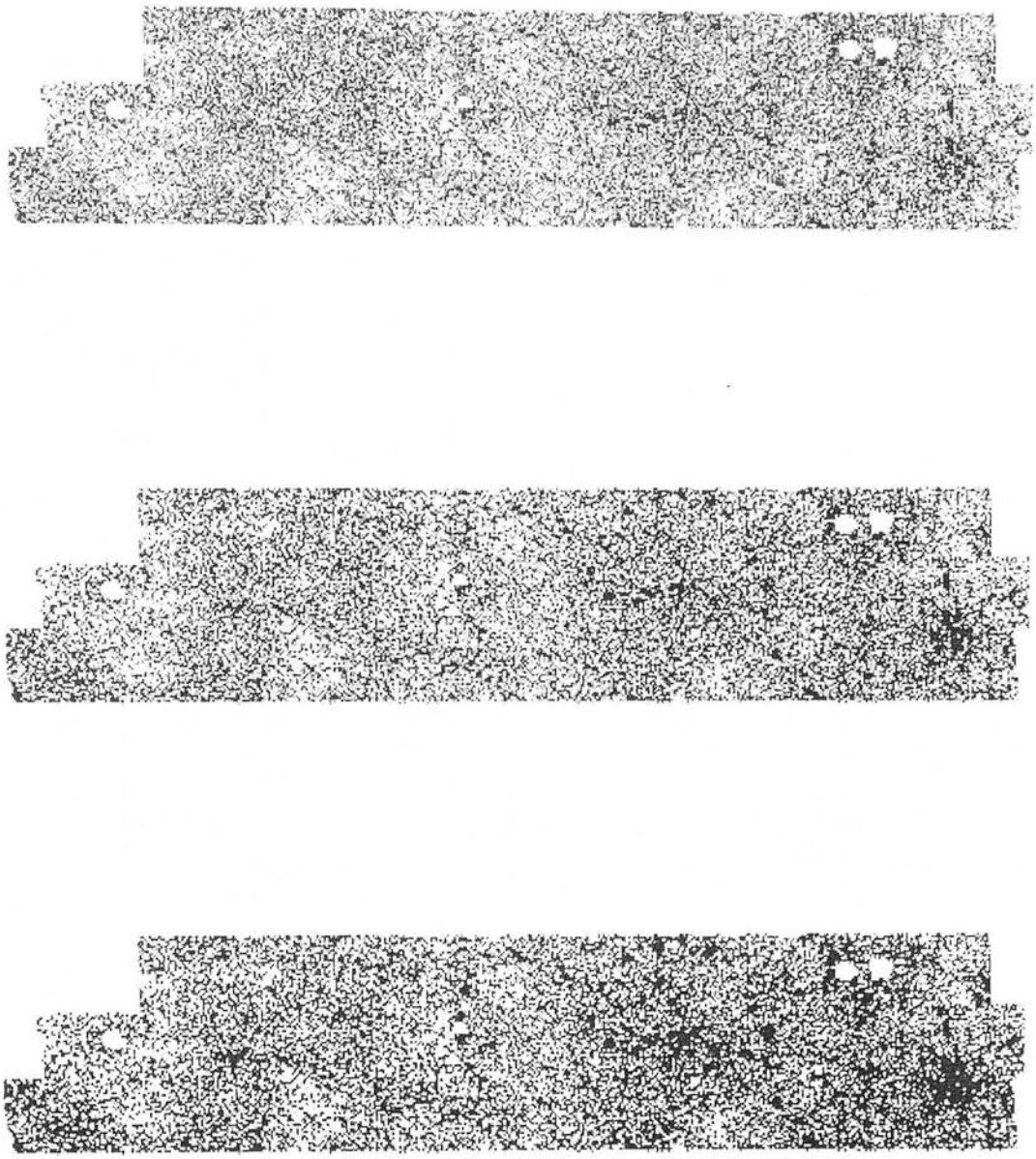


Figure 5.1: *Negative* greyscale maps of the ($b_j < 20$) galaxy distribution across the 35-plate mosaic. Bins subtend the same solid angle ($10' \times 10'$) at all declinations. (a) white < 180 deg^{-2}, black > 740 deg^{-2}. (b) white < 180 deg^{-2}, black > 540 deg^{-2}. (c) white < 180 deg^{-2}, black > 360 deg^{-2}.

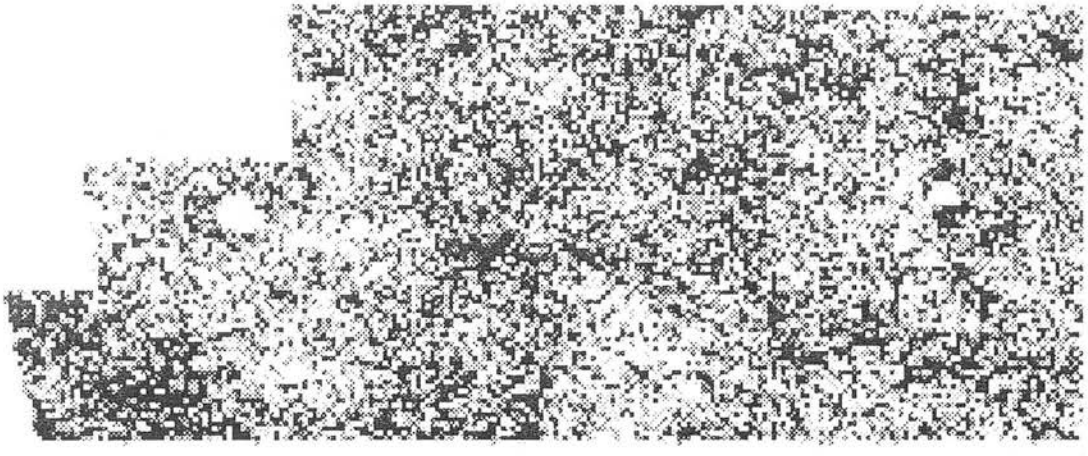


Figure 5.2: A blow-up of the eastern side of Fig. 5.1c. data is in $10' \times 10'$ bins. White pixels have < 180 galaxies deg^{-2}, black have > 360 deg^{-2}.

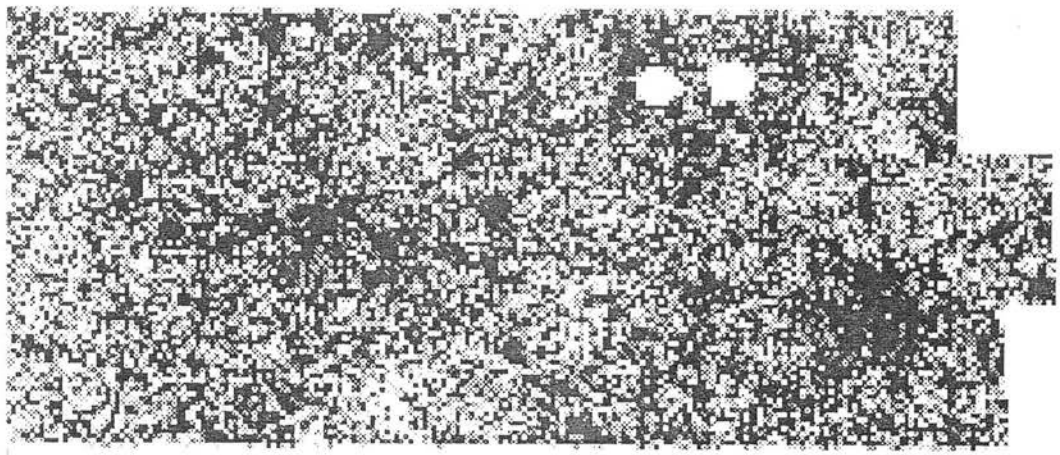


Figure 5.3: A blow-up of the western side of Fig. 5.1c. data is in $10' \times 10'$ bins. White pixels have < 180 galaxies deg^{-2}, black have > 360 deg^{-2}.

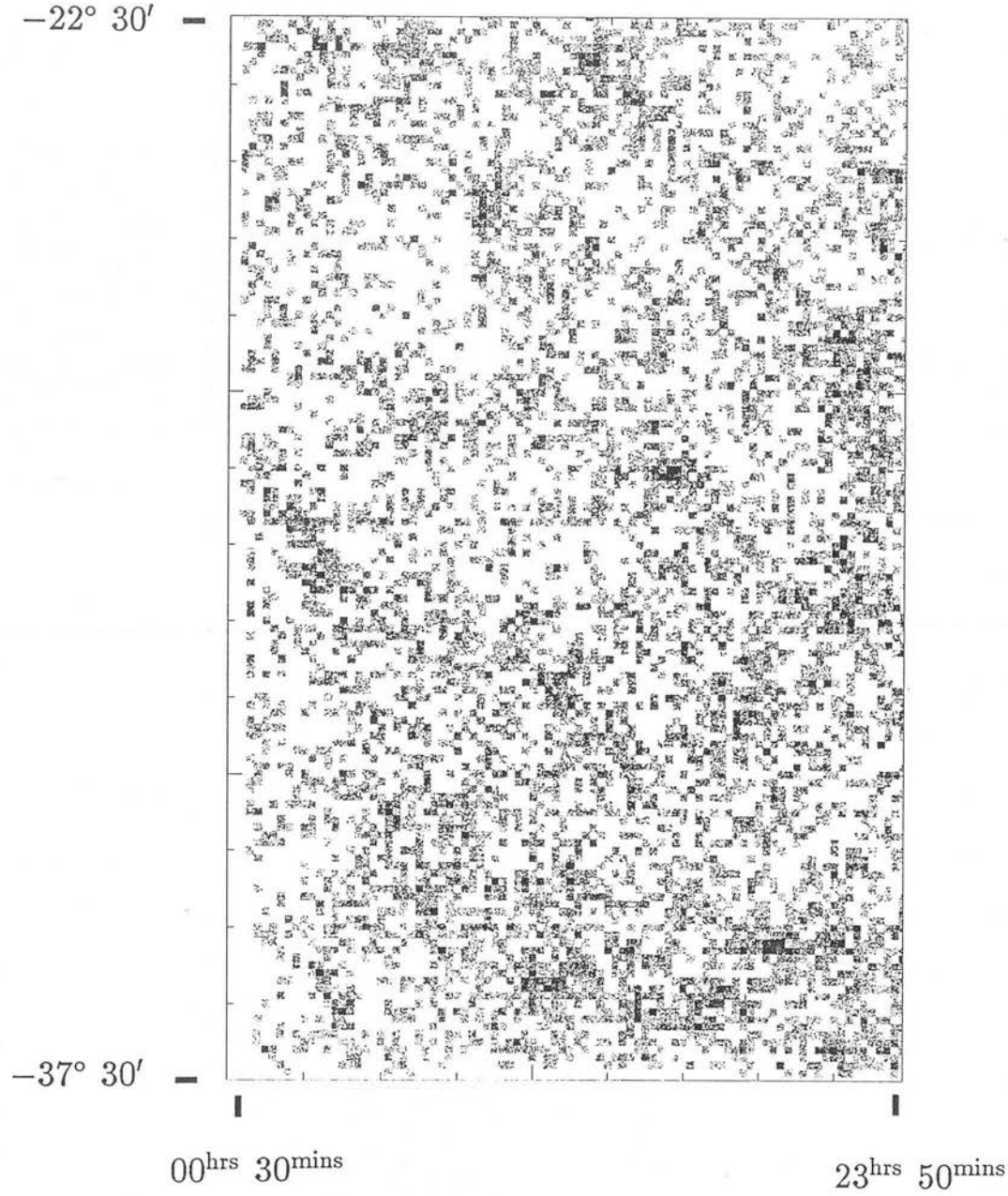


Figure 5.4: *Negative* greyscale maps of the ($b_j < 19{\cdot}5$) galaxy distribution across a 6-plate mosaic presented in Heydon-Dumbleton *et al.* (1988). Data from F349, F350, F409, F410, F472 and F473 is included.

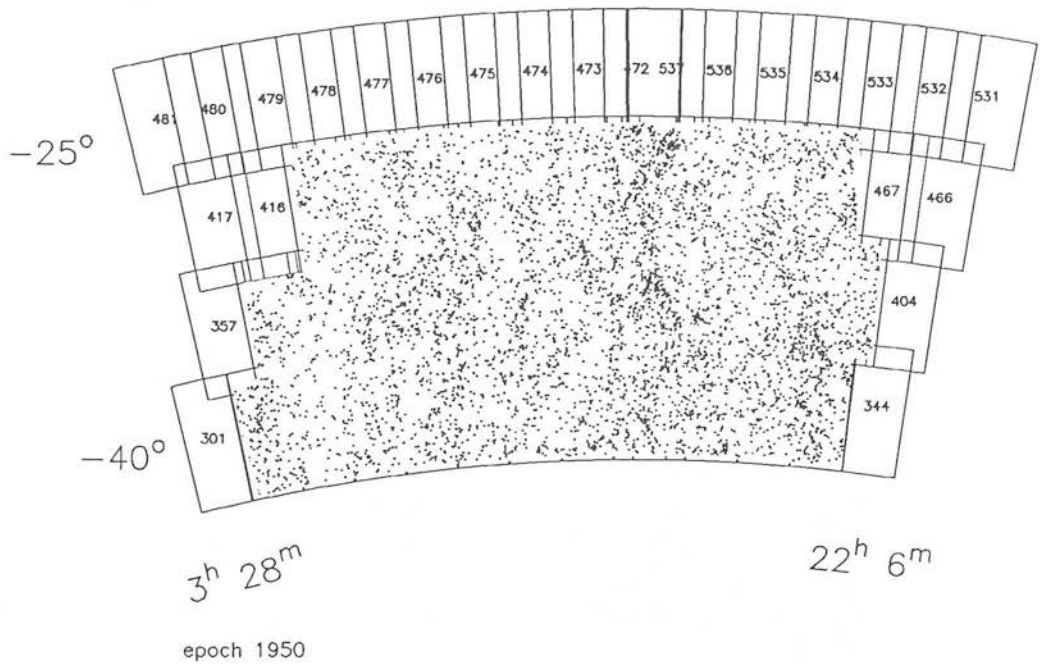


Figure 5.5: The distribution of bright ($b_j < 17$) galaxies across the 35-plate mosaic, superimposed on the plate centers. See the text for details.

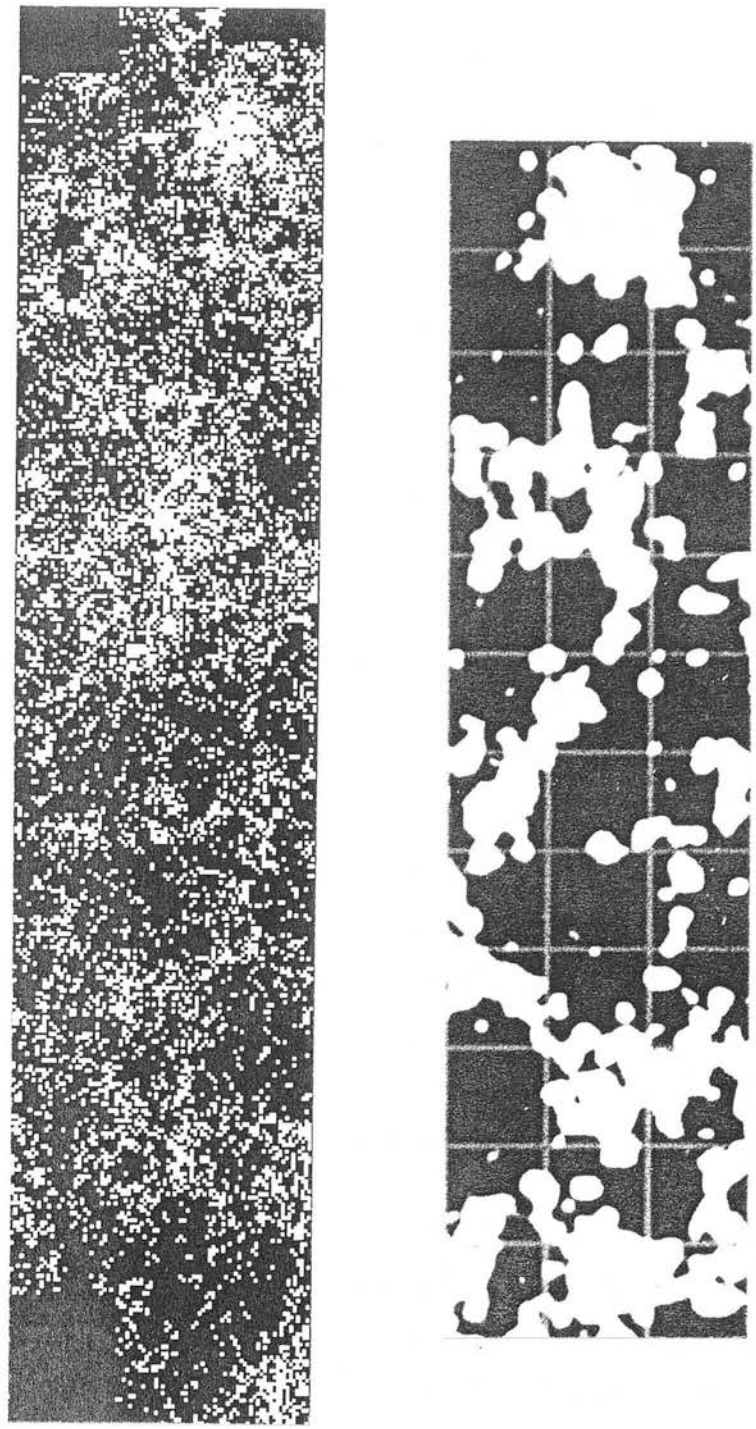


Figure 5.6: *Positive* greyscale maps from the EDSGC 35-plate mosaic and the SSGP maps. The maps are shown at approximately the same scale, but cover different areas of the sky. The SSGP map is reproduced from de Lapparent *et al.* (1986). Greyscale levels are discussed in the text.

construction.

A positive high contrast (white ≥ 10, black < 10 galaxies per pixel) greyscale map is shown in Fig. 5.6, together with a smoothed high contrast map of a part of the SSGP maps. These maps have approximately the same scale to facilitate comparison of the types of structures seen in the EDSGC and the *Lick Catalogue*, but cover different parts of the sky. The EDSGC map has black or white pixels depending on whether the number density of galaxies is $<$ or $>$ 360 deg^{-2}. The SSGP data is in the northern hemisphere and is reproduced from de Lapparent *et al.* (1986). These counts have been smoothed and the highest 18% of the pixels are shaded white.

5.3.2 LARGE-SCALE STRUCTURE IN THE MOSAIC

The previously observed large-scale distribution of galaxies in the Universe was discussed in Section 1.2. Galaxies were described as lying in clusters, clouds, superclusters and filaments, while avoiding certain void regions. All these structures can be seen in the various maps of the EDSGC 35-plate mosaic. There is a favourable comparison between the structures seen in this work and those seen in the SSGP maps (Fig. 1.1). in both cases there are large complex structures covering several fields linked by filaments. A more objective comparison between these two datasets will be made in Section 5.4. Here some of the structures seen in the EDSGC are described (subjectively).

Two large supercluster complexes, separated by $\sim 15 - 20°$, can clearly be seen in the data (Fig. 5.1). One of these is linear and stretches (W to E) across $\sim 10°$ from F408 through F349 to F350. The other supercluster is more spherical in shape (diameter $\sim 5°$) and is positioned at the corner overlap of F345, F346, F405 and F406. This structure was reported in Collins *et al.* (1988) and has a core cluster at a redshift of 0·16. The diameter of this supercluster is estimated at $\sim 15h^{-1}$ Mpc . Both of these structures are visible in the plot of bright ($b_j < 17{\cdot}0$) galaxies in Fig. 5.5, though the orientation of the linear structure has changed. Here it is more prominent, in a N to

S orientation, on F408 and F471. This change suggests that the structure is a large filament that twists along the line of sight. A further, though less obvious structure, can be seen covering F412, F413, F353, F354 and F298 in both the bright and faint galaxy distribution. This structure has a dense core on F413.

Several filaments can be seen in the $b_j < 20{\cdot}0$ greyscale maps, which can also be traced across several of the plate datasets plotted in Appendix B. In particular two arc structures can be seen *snaking* out of the F408-F349-F350 complex. These arcs (*subjectively*) appear to have clusters spaced almost regularly along them at approximately 2 to 3° intervals, much like knots on a piece of string.

There is a number density gradient across the 35-plate mosaic. This is visible in the various greyscale maps and in Fig. 5.7, which plots the average number density ($N/0^m{\cdot}5/deg^2$ at $b_j = 17{\cdot}75$) on each plate against the RA at its centre. The majority of plates to the west of 0·5 hrs have a higher than average number density, while those to east have a lower than average number density. Comparison of the EDSGC with the APM Galaxy Survey (Maddox *et al.* 1988a,b) suggests that this gradient is associated with real large-scale structure. Coupled with the observed clustering on scales $\sim 20°$, this gradient suggests that the present 35-plate mosaic (~ 1000 deg^2) is not yet a fair-sample of the large-scale galaxy distribution.

5.3.3 DISCUSSION

Possible sources of spurious structures that can be introduced during the production of the EDSGC have been identified in Chapters 3 and 4. The purpose of this section is to note regions of the 35-plate mosaic where the observed structure in *e.g.* Fig. 5.1 may be connected with calibration or classification errors.

Comparison of the bright galaxy distribution with the plate boundaries in Fig. 5.5 is on the whole favourable. Most high and low density regions can be traced across several

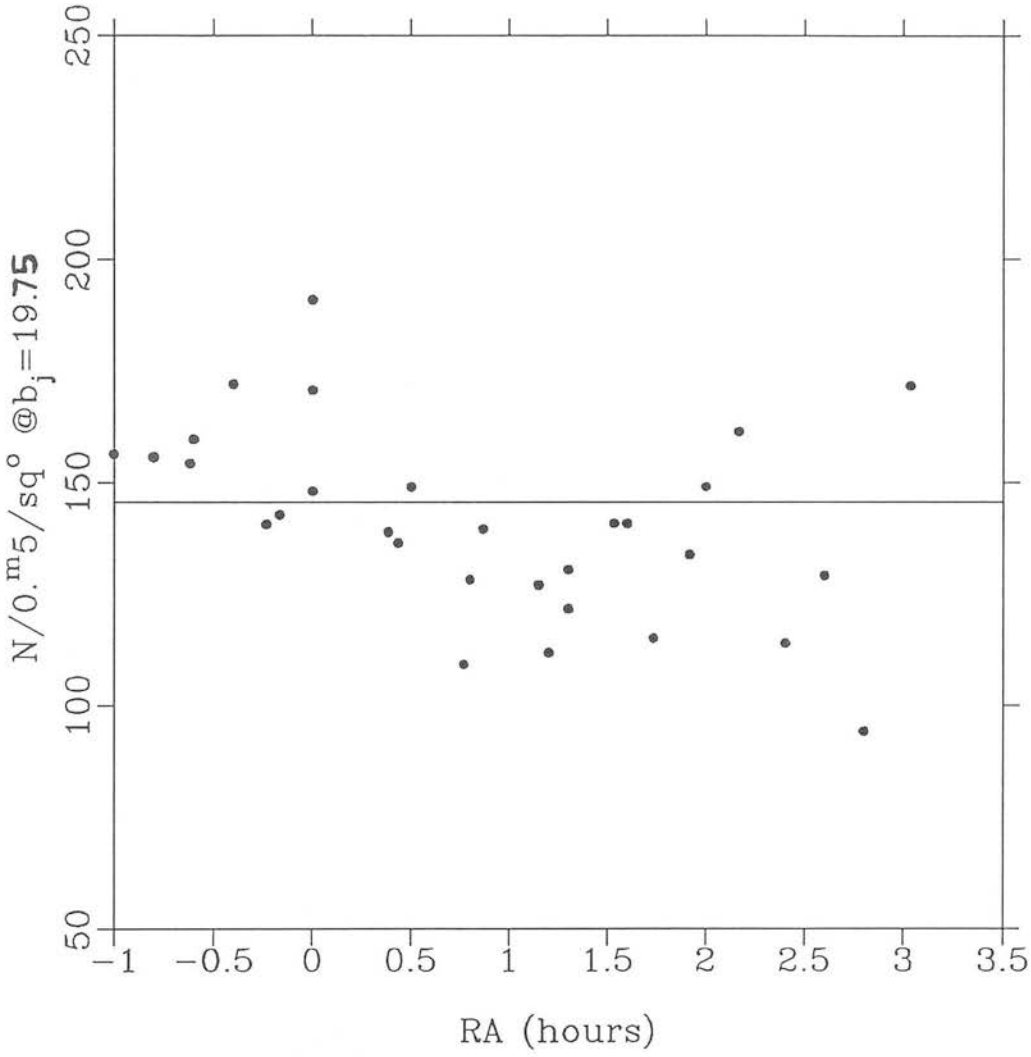


Figure 5.7: Number density on each plate ($N/0^{m}.5/deg^2$ at $b_j = 17.75$) plotted against the plate RA coordinate. The solid line shows the average number density. Plates west of 0^{hrs} are assigned negative RA.

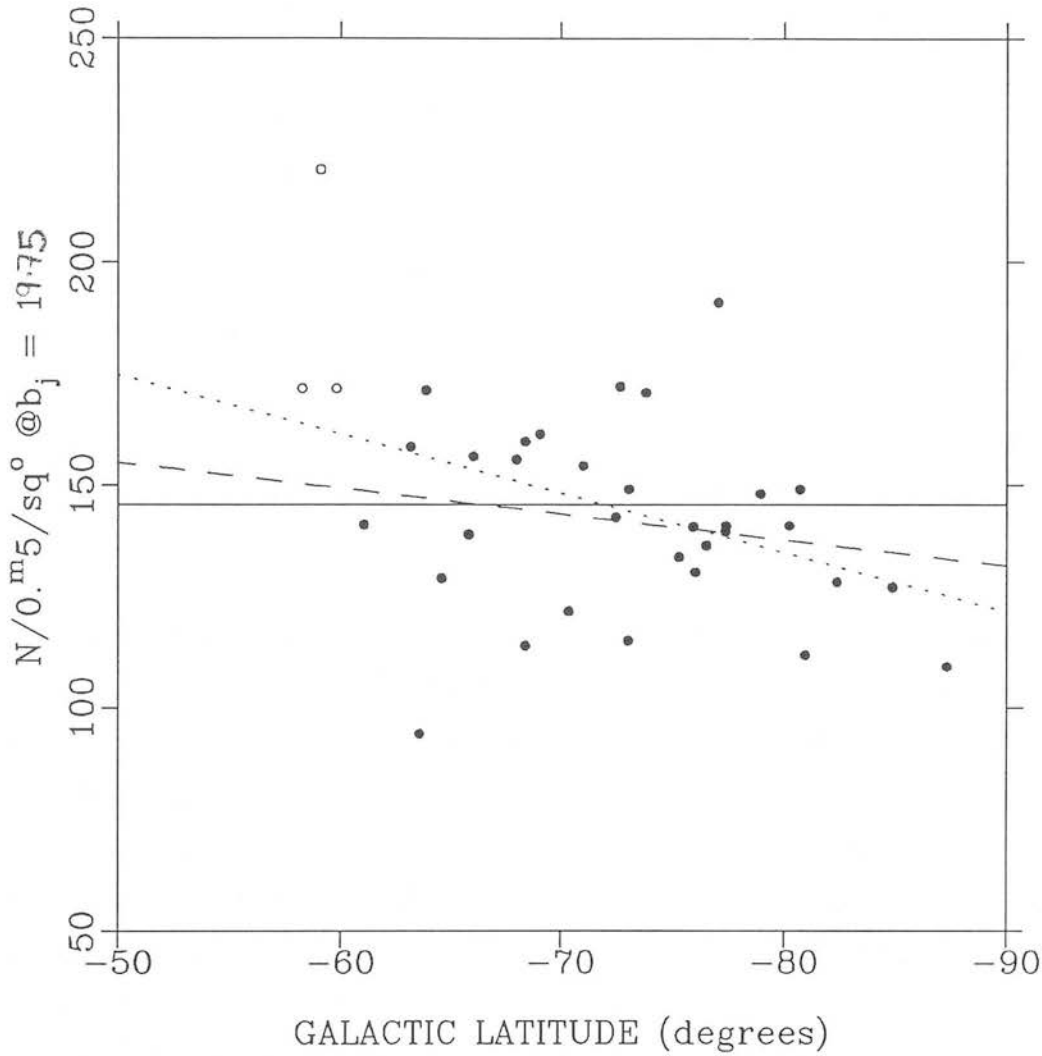


Figure 5.8: Number density on each plate ($N/0^{m}5/deg^2$ at $b_j = 17{\cdot}75$) plotted against the galactic latitude of the plate centre. The solid line shows the average number density. The dashed and dotted lines are referred to in the text.

plates and are not aligned with a plate boundary, though there is some suspicion about the density variation on F299, F300 and F355. The extent and boundaries of these fields can *just* be distinguished in Fig. 5.5, though structures can be traced in Fig. 5.2 from F299 to F300 and from F354 to F355. The variations in number density on these fields are equivalent to F299 and F355 being calibrated too shallow and F300 being too deep. These plates are in the region of the 35-plate mosaic where the one-in-two CCD arrangement has not been used. Inspection of the plots in Appendix B reveals that there is a defocussed region on the top left of F406 and an un-drilled star at the top left of F408. The first of these two flaws is in the mosaic because a non-survey plate was appended by mistake. The ESO copy of the survey plate is now available and has been classified and calibrated. The erroneous plate will be replaced in the final version of the EDSGC. The structure at the bottom left of F406 has been confirmed by visual inspection of the survey plate.

Number density variation as a function of galactic latitude (b) has been investigated. Fig. 5.8 shows the average number density ($N/0^{m}.5/deg^2$) on each plate plotted against the galactic latitude of the plate centre. The data from all 35 fields gives a slope of $1{\cdot}33 \pm 0{\cdot}52$ (dotted line). This trend is probably due to the W–E gradient, commented on in the previous section, as the 35-plate mosaic data is in a thin strip that runs from $b = -50°$, though the SGP (on F411), back to $b = -50°$. Removal of F345, which contains a supercluster structure, reduces the slope to $0{\cdot}5 \pm 0{\cdot}48$, while removal of the 3 fields with $b < 60$ (F345, F405 & F300, which may have a calibration error: open circles in Fig. 5.8) results in a slope of $0{\cdot}58 \pm 0{\cdot}56$ (dashed line). Neither of these slopes is significant, so there is no evidence for significantly higher stellar contamination at lower galactic latitudes.

5.3.4 COMPARISON WITH THE APM SURVEY

The APM Galaxy Survey (Maddox *et al.* 1988a,b) is an objective galaxy survey similar to the EDSGC, based on scans of SERC J survey plates by the Automatic Plate Machine in Cambridge. This survey is still in production and final details of the techniques used or estimates of systematic errors have not been published, so it is not possible to comment on this work in detail. The construction of the APM survey is however different to the EDSGC in many details and so comparison of the two surveys will provide useful constraints on the validity of observed structures, as well as independent measure of the errors in the clustering statistics. Some difference between the APM Survey and EDSGC are listed below:

(i) The APM Survey covers (~ 4400 deg^2) 176 plates around the SGP and is estimated to contain 4×10^6 galaxies to $b_j < 21$. The APM scans in density space and produces a summed image density for all image pixels lying a given constant above the local *sky-density*. Measured image densities are converted to *optical density* using a look-up table based on the fog and saturation levels of "an average copy plate". APM magnitudes (M) are obtained using a linear optical density to intensity calibration, with $M = z - b_j$, where z is the plate *zero point* ($\sim 29 - 30$). Maddox *et al.* (1988b) contains no details of whether an attempt is made to account for plate-to-plate variations in the contrast, fog and saturation levels (see Section 2.2.2 and Section 2.3.1).

(i) For images with $b_j < 16{\cdot}3$ classification as stars or galaxies is carried out visually. Bunclark (1984) reports that the APM is unable to detect images larger than 2mm and produces incomplete datasets for images larger than 1mm ($b_j \sim 16 - -17$). No solution to this problem has been reported in the published work on the APM Survey, so it is unclear whether bright images ($b_j < 16{\cdot}3$) are also detected by eye. Merged images are flagged but not *deblended* as in the EDSGC. The APM Survey is probably incomplete in the cores of compact clusters.

(iii) For images with $16{\cdot}5 < b_j < 21$, classification was originally carried out using the APM automated classification program (STATS), which is based on surface brightness and size criteria. Later work uses profile information (Maddox *et al.* 1988b), with a confidence limit set according to the results of visual classifications. No details of stellar contamination are given, but with fixed classification parameters a value of 70% completeness for $b_j < 20{\cdot}5$ is quoted. Publication of number–magnitude counts would be a good indication of the uniformity of the classification techniques as a function of magnitude. Maddox *et al.* state that work to *match* the classification parameters across plate overlaps is in progress and that this will ensure consistent completeness and contamination statistics between plates. No attempt has been made to account for variations within each plate, *e.g.* field effects (Section 3.2.1).

(iv) Inter-plate calibration is achieved using an algorithm similar to that applied to the *Lick Catalogue* by SSGP. Maddox *et al.* (1988b) claim a residual zero point scatter of $0^{\rm m}{\cdot}02$. In addition to this, 66 CCD sequences are used as "tie points" in the calibration algorithm to reduce large-scale scale gradients that may result from global calibration techniques (Section 1.4). The number of and arrangement of CCD sequences in the APM survey is considerably less than the one-in-two arrangement used in the EDSGC. Attention is drawn to Section 4.3.1 and Section 4.3.3 which discuss some of the problems associated with this approach and the need for a large number of calibration sequences. Estimates from preliminary calibration of the EDSGC and from de Lapparent *et al.* suggest that global calibration techniques result in larger calibration errors ($\sim 0^{\rm m}{\cdot}06$ Section 1.4) and that individual plates can have considerably larger errors ($> 0^{\rm m}{\cdot}2$, Section 4.3.3).

A greyscale map of the large-scale galaxy distribution from the APM Survey is given in Maddox *et al.* (1988b). This data is calibrated but has not had the classification matching procedure (iii) applied. No details on the procedure for mosaic construction and removal of overlaps are given. The region of the EDSGC 35-plate mosaic can be aligned with this map as F406, F408 and F357 have not been appended to the APM

Survey. The 3 large structures visible in the EDSGC, described in Section 5.3.2, can also be seen in the APM map. There is also a low density region around F355 and F356, with a high density region around F299 and F300. This is in agreement with F299 being calibrated too shallow in the EDSGC. On larger scales, the distribution of galaxies in the the APM Survey, over the region of the 35-plate mosaic, is consistent with the decrease in galaxy number density from west to east seen in the EDSGC. This effect would appear to be due to the position of the EDSGC with respect to the two super-cluster structures described previously, rather than any systematic error.

There are a few spurious structures in the APM map that are not in the EDSGC data. The compact star cluster on the F355,F354 boundary, which has been drilled from the EDSGC, is extremely conspicuous in the APM survey as a high peak in an otherwise low density region. The bright star and its associated ghost on F469 have also not been removed. A low density square structure is clearly visible to the left of F408 in the APM greyscale, but is not visible in the EDSGC (Fig. 5.2). This may be due to F349, which has a bright m_{sky} value, being incorrectly calibrated in the APM survey. Finally comparison of Fig. 10 (profile classification) in Maddox *et al.* 1988b with Fig. 2 (no classification) of the same reference and with Fig 4 (STATS classification) in Maddox *et al.* 1988a shows that the number density of objects in the APM Survey is dependent on galactic latitude. Plates below $b = -50°$ have a higher average number density than those above. The profile classifier has only partially removed this dependence from the APM Survey.

5.4 Application of the EDSGC

5.4.1 NUMBER-MAGNITUDE COUNTS

The differential counts for all data in the 35-plate mosaic are plotted in Fig. 5.9, together with the counts from previous authors (Ellis 1987) and data from the 6-plate

mosaic described in Heydon-Dumbleton *et al.* (1989). The solid line in this diagram is the variation expected from a no-evolution model (Ellis 1987), while the error bars represent the rms variation in the amplitude of the counts from individual plate datasets. These variations are 16%, 24% and 44% at b_j = 19·75, 18·75, and 15·75 respectively. The counts from individual plates, which are shown separately in Appendix C, are superimposed on the data from previous authors in Fig. 5.10.

Although the counts presented here are not as deep as those of previous workers, they do provide an independent estimate of galaxy counts in the range b_j = 18·0 to 20·0, where there is already a significant difference between authors. In addition, the large area of the EDSGC survey region, which is 300 times the area of any previous survey to similar depths, provides the most accurate estimate, yet produced, of the number–magnitude relation at bright magnitudes ($b_j < 18{\cdot}0$). The EDSGC counts show a least squares fit slope of $0{\cdot}598 \pm 0{\cdot}005$ for galaxies with $13{\cdot}5 \leq b_j \geq 19{\cdot}0$. Data from the 6-plate mosaic has a slope of $0{\cdot}6 \pm 0{\cdot}1$ in the range b_j = 14·0 to 17·0 and a slope of $0{\cdot}58 \pm 0{\cdot}01$ in the range b_j = 17·0 to 19·5. The form of these counts and their agreement with those of previous authors provides strong evidence for the uniformity, across a wide range of magnitudes, of the classification algorithms detailed in Chapter 3.

At the bright end our counts agree well with the Durham/AAT redshift survey, which indicates our stellar contamination is negligible at these magnitudes (this redshift survey does not contain any stars). Some care should be taken with the counts at $b_j \simeq 14-15{\cdot}0$, as a contamination spike could have been introduced at these magnitudes by the $\mathcal{G}$ parameter classification method. This was noted in (v) of Section 3.4.1. This contamination spike is visible to varying degrees in the plots in Appendix C and in the data from the 6-plate mosaic at $b_j \simeq 14{\cdot}5$. It is not apparent in the counts for the whole region and may have been washed out because of the variation in m_{sky}. Removal of this contamination spike will steepen the slope of the number-magnitude relation for $b_j \lesssim 15{\cdot}5$ data, as is seen in the 6-plate mosaic data.

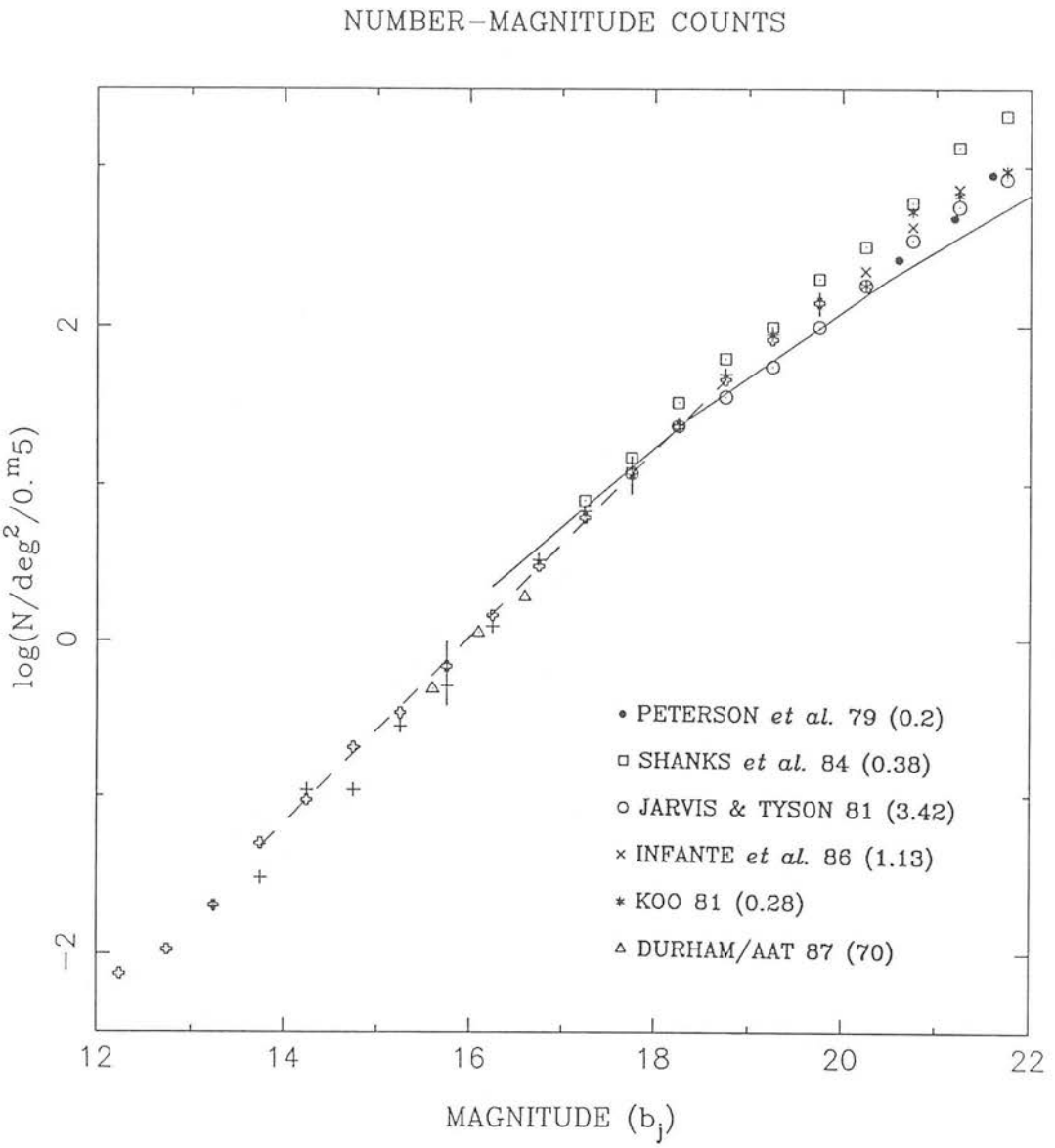


Figure 5.9: Differential number–magnitude counts from the EDSGC 35-plate (open crosses) and 6 plate (crosses) mosaics. The dashed line is a least squares fit to the 35-plate mosaic data with $13{\cdot}5 \leq b_j \geq 19{\cdot}0$. The slope is $0{\cdot}598 \pm 0{\cdot}005$. The errors bars represent the rms variation in the counts from individual plates.

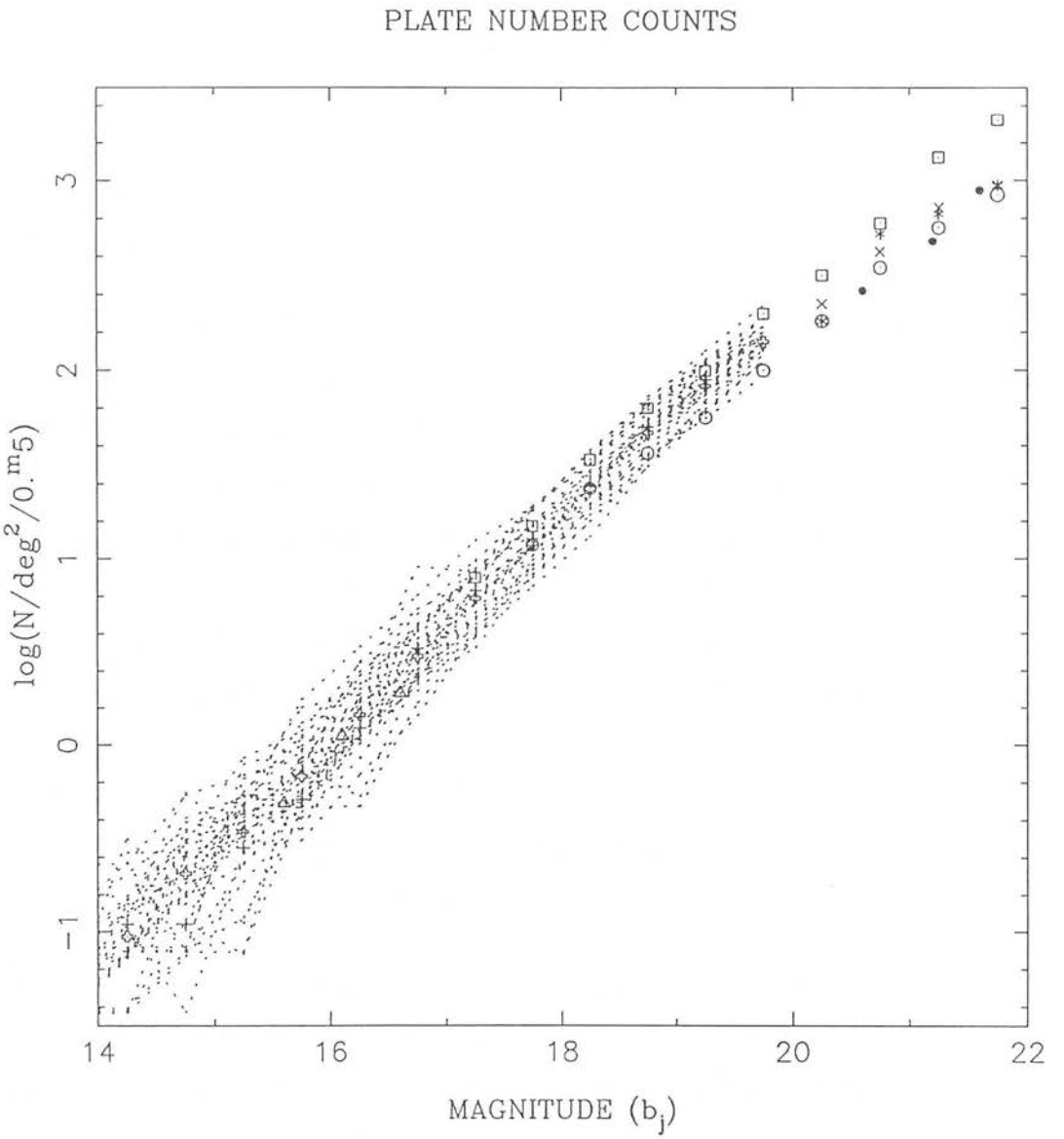


Figure 5.10: Variation in number–magnitude counts across the 35-plate mosaic. Counts for individual plates (Appendix C) are plotted as dotted lines, on top of the data shown in Fig. 5.9.

The EDSGC data agree well with the counts of Jarvis & Tyson (1981) at intermediate magnitudes (b_j = 17·5 to 18·75), but are inconsistent with these counts and the no-evolution model of Ellis (1987) fainter than b_j = 18·75. For instance, we count 1·5 times as many galaxies as Jarvis & Tyson (1981) at $b_j \simeq 19{\cdot}75$. This data supports the claim of higher count estimates than expected from a non-evolved model, indicating moderate galaxy evolution (e.g. Shanks *et al.* 1984).

The reason for the discrepancy between the amplitudes quoted by different authors is unclear and has been a source of contention for some time. It could obviously be caused by the different detection, classification and photometric techniques used, though Ellis (1987) concludes that the scatter in slope and amplitude is due to large-scale clustering of faint galaxies. Fong *et al.* (1987) compare galaxy counts with IRAS and HI data and rule out galactic extinction as the source of these fluctuations. The results in Section 3.4 show that the form of EDSGC counts are not due to either stellar contamination or galaxy incompleteness and Section 5.3.1 estimates the plate-to-plate variation in number-density, due to systematic error, as 8% rms. The variation in the individual plate counts at b_j = 19·75 is 16% and so cannot be due to systematic error. Results from the EDSGC agree with the conclusion of Ellis that these fluctuations are due to the large-scale clustering of galaxies.

Finally, as can be seen in Fig. 5.10, the data from Shanks *et al.* (1984) and Jarvis & Tyson (1981) lie at the extreme (high and low) limits of the variation in counts from single plate counts. Though this discrepancy may be due to the different methods of image classification or image photometry, it should be noted that both these surveys cover regions that are much smaller[1] than that covered by a single EDSGC plate. It is, therefore, not unexpected that they should show larger variations that those for the single plate datasets.

[1]The Shanks *et al.* data is a sum of several small regions

5.4.2 CORRELATION FUNCTIONS

One of the major clustering statistics for galaxies is the two-point angular galaxy-galaxy correlation function ($\omega_{gg}(\theta)$). This quantity measures the excess number of galaxies above random, at a given angular separation, seen in projection on the sky. Despite the smaller clustering contrast seen in two dimensions, due to projection effects, compared to the equivalent spatial quantity $\xi(r)$, there has been renewed interest recently in two-dimensional clustering studies. This is primarily due to the large number of galaxies available in deep wide area digitised surveys. This makes it possible to trace $\omega_{gg}(\theta)$ out to cosmologically interesting scales ($> 10h^{-1}$ Mpc) for the first time.

The definition and some of the properties of $\omega_{gg}(\theta)$ where discussed in Section 1.2.5. The first major determination of $\omega_{gg}(\theta)$ was carried out by Groth & Peebles (1977) using data from the *Lick Catalogue*. The main result of this study was a power-law correlation function of index ~ -0.734 on small scales, with a break at $\sim 2.5°$, corresponding to a scale $\sim 10h^{-1}$ Mpc. The amplitude of power law form and the existence of a break put strong constraints on various theories of galaxy formation. These were discussed in Section 1.3.4, though the possibility that the break may be due to residual systematic errors in plate-to-plate calibration, as suggested by Geller *et al.* (1984), was noted in Section 1.4.

The correlation function has been determined for 3 magnitude limited samples ($b_j >$ 18·0, 19·0 and 20·0) of the 35-plate dataset. Two methods of calculation have been used, as described in Section 1.2.5. For the $b_j < 18$ datasets and for $\theta < 0{\cdot}1°$ in the $b_j < 19$ and $b_j < 20$ datasets, the separation for all pairs of galaxies is calculated and $\omega_{gg}(\theta)$ obtained from the number of pairs per angular interval (Eqn. 1.3). Random datasets are constructed to cover the same region in RA and DEC as the 35-plate datasets and have the same areas drilled, though no other variations (*e.g.* for galactic extinction) are introduced. For separations larger than 0·1°, the galaxy and random data is binned into $10' \times 10'$ cells (as shown in Fig. 5.1) and the correlation estimated from the counts

in disjoint cells separated by $\theta \pm \delta\theta$

Fig. 5.11 shows the computed correlations for the 3 magnitude limited samples. Least squares fits to the data with $\theta < 1{\cdot}2°$ give the following amplitudes and exponents:

$b_j < 18$: $\omega_{gg}(\theta) = 0{\cdot}106 \pm 0{\cdot}002\theta^{-0{\cdot}703\pm0{\cdot}016}$,

$b_j < 19$: $\omega_{gg}(\theta) = 0{\cdot}075 \pm 0{\cdot}002\theta^{-0{\cdot}603\pm0{\cdot}01}$,

$b_j < 20$: $\omega_{gg}(\theta) = 0{\cdot}051 \pm 0{\cdot}003\theta^{-0{\cdot}518\pm0{\cdot}02}$.

These power-law fits are shown as solid lines in the Fig. 5.11. At larger scales each of the functions show a break away from this power-law form, though the position of the breaks are difficult to estimate because of various *bumps* on scales $\gtrsim 1°$. The functions go negative at approximately 18°, 9° and 7° for the $b_j < 18$, < 19, and < 20 data respectively.

To compare the EDSGC data with that from Groth and Peebles (1977), the functions shown in Fig. 5.11 have been scaled (see Section 1.2.5) to the depth of the *Lick Catalogue*. These scaled functions are plotted in Fig. 5.12. The amplitude and angular scalings are carried out separately using the method described in Groth and Peebles (1977), with:

$$\begin{aligned} \theta_l &= \theta \times x \times \left(\frac{\mathrm{D}}{\mathrm{D}_l}\right) \quad \text{and,} \\ \omega_l &= \omega \times y \times \left(\frac{\mathrm{D}}{\mathrm{D}_l}\right), \end{aligned} \tag{5.1}$$

where the subscript, l, indicates the value of the amplitude (ω) and angle (θ) at the Lick depth. The factors x and y account for the variation of the selection function and the luminosity function with depth and for cosmological effects. $(\mathrm{D}/\mathrm{D}_l)$ is the ratio of the effective depth of the EDSGC and Lick data.

The values for x and y used for each of the EDSGC samples are given in Table 5.1

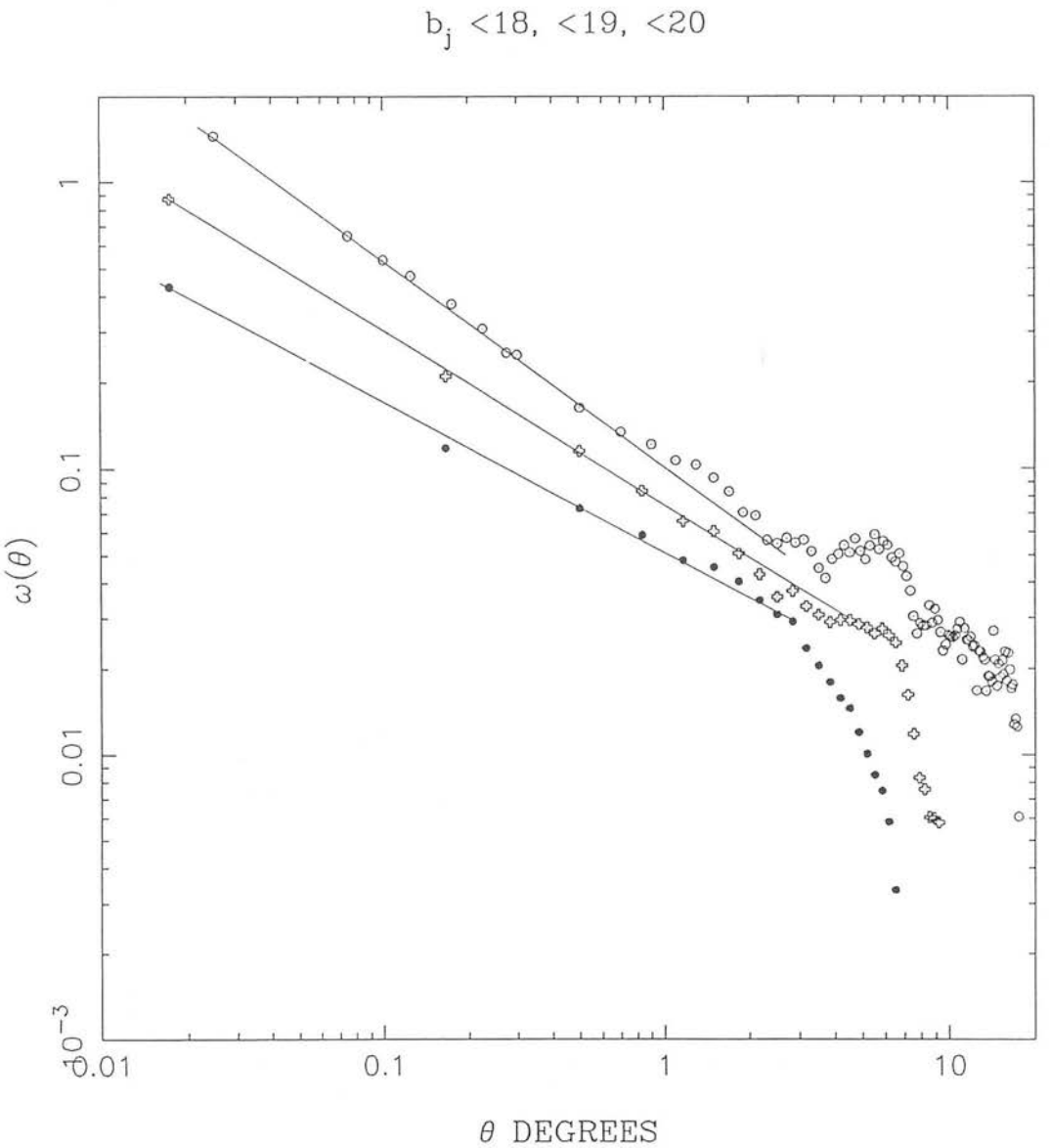


Figure 5.11: Correlation functions from the EDSGC 35-plate mosaic. Data is shown for $b_j < 18$ (open circles), $b_j < 19$ (open crosses) and $b_j < 20$ (filled circles) magnitude limited samples. The solid lines are least squares fits to the functions for $\theta < 1{\cdot}2°$.

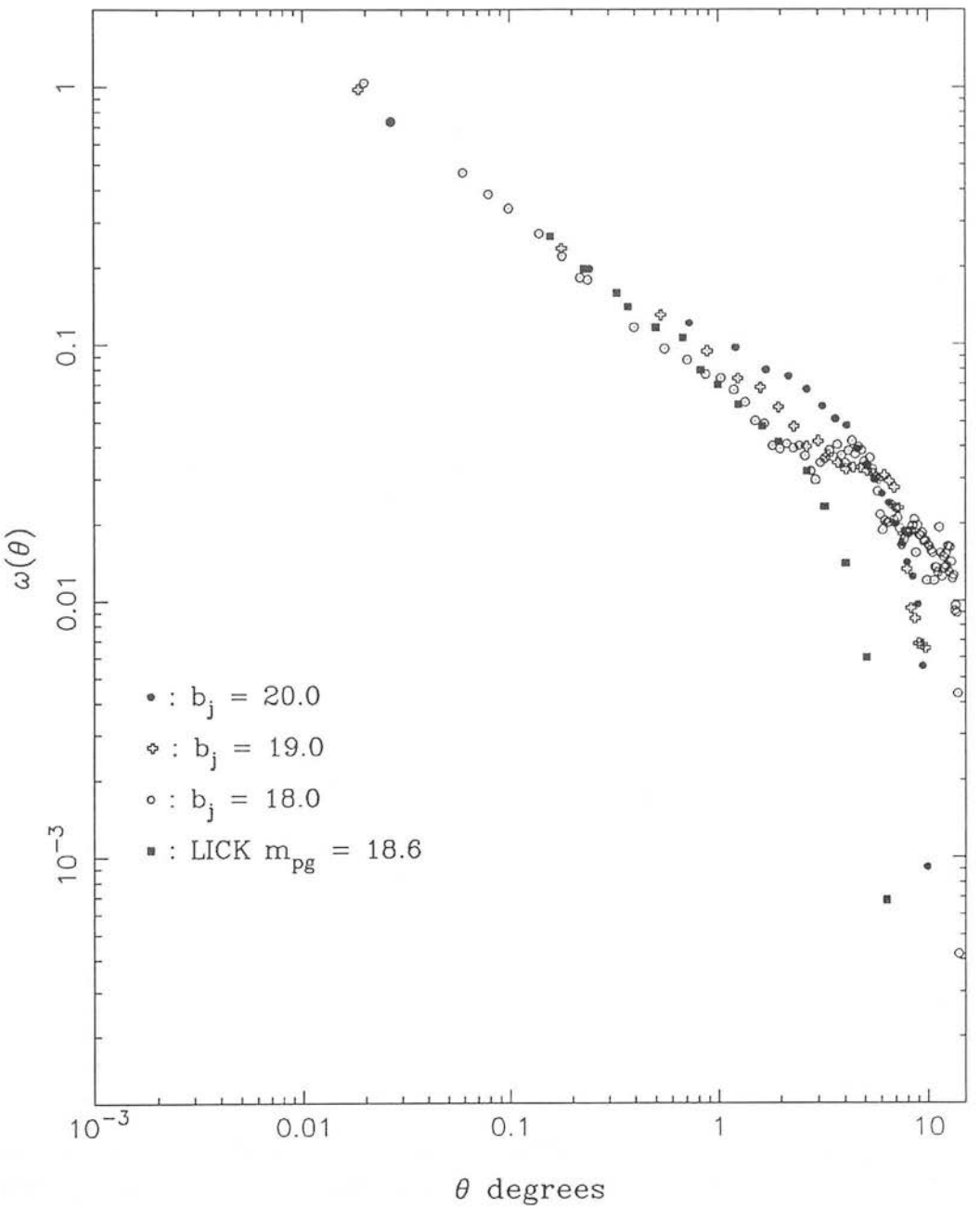


Figure 5.12: Correlation functions from the EDSGC 35-plate mosaic scaled to the depth of the *Lick Catalogue*. The symbols are as Fig. 5.11. The Groth and Peebles correlation function is also shown (filled squares).

m_{lim}	x	y	(D/D_l)
18·0	1·10	1·00	0·715
19·0	0·94	1·00	1·124
20·0	0·86	0·99	1·678

Table 5.1: The parameters used in scaling the EDSGC correlation functions from the various limiting magnitudes to the depth of the *Lick Catalogue*. See Eqn. 5.1.

and are estimated from those quoted in Groth and Peebles (1977) and Shanks *et al.* (1984)[2]. The ratio of effective depths (also given in Table 5.1) is calculated from the observed number densities with:

$$\left(\frac{\mathrm{D}}{\mathrm{D}_l}\right) = \left(\frac{(0{\cdot}9) \times \mathrm{N}}{1{\cdot}77 \times 10^5}\right)^{\frac{1}{3}}, \tag{5.2}$$

where N is the number per steradian in the EDSGC samples. The factor of 0·9 accounts for the 10% stellar contamination in the EDSGC data.

On small scales ($\theta < 0{\cdot}6$) the EDSGC scaled correlation functions are in good agreement with the work of Groth & Peebles, but at large scales the data is confusing. The EDSGC data scaled from $b_j < 19$ and $b_j < 20$ scale to give a break from power-law at $\sim 5°$, while the data scaled from $b_j < 18$ gives a break at $\sim 7°$, though the positions of the break scales are difficult to estimate because of the bumps. Similar behavior is seen in the scaling of the point at which each of these functions goes through zero. The data scaled from $b_j < 19$ and $b_j < 20$ goes negative at 10°, while the data scaled from $b_j < 18$ goes through zero at 14°. The positions should be compared with the break and zero-crossing point of the Groth & Peebles data, which occur at 2·5° and $\sim 7°$ respectively. At the Lick depth ($m_{pg} \sim 18{\cdot}6$), the break scale in the Groth & Peebles data corresponds to $10h^{-1}$ Mpc , while the EDSGC data indicates that the break occurs at $\sim 20h^{-1}$ Mpc .

Comparisons can also be drawn with data from smaller regions of the EDSGC and from Shanks *et al.* (1980). Scaled correlation functions are plotted in Fig. 1.7 for data from a 6-plate mosaic ($b_j < 19{\cdot}5$) and a *partially* calibrated 24-plate mosaic ($b_j \lesssim 20{\cdot}5$), previously published in Collins *et al.* (1989). The 6-plate mosaic is calibrated from CCD sequences and plate overlaps and so the error on m_{lim} is estimated as $0^{\mathrm{m}}{\cdot}05$ rms. The 24-plate mosaic is calibrated using overlaps only, so has a larger error ($\sim 0^{\mathrm{m}}{\cdot}2$ rms Section 4.3.3) in the limiting magnitude. The position of the breaks for these (scaled) functions was estimated at 2°($\sim 7h^{-1}$ Mpc) for the 6-plate dataset and 2·5° for the

[2]note the factors in these references are for scaling to the Zwicky depth. The x and y factors in Shanks *et al.* are printed the wrong way round.

24 plate mosaic. Shanks *et al.* (1980) obtained correlation functions from COSMOS scans of the central 14 deg^2 of individual UKST plates. They report a break in the correlation function (scaled from 20·5 to the Lick depth), but at 0·8° ($\sim 3h^{-1}$ Mpc). The areas subtended by the 6-plate region and by the Shanks *et al.* data correspond to radii $\theta_R \sim 5°$ and $\theta_R \sim 2°$ (Section 1.2.5). These dimensions are small enough so that the (unscaled) correlation functions could not detect the power on scales ~3° to 5°, that is seen in the 35-plate $b_j < 20$ sample.

Criticism of the 33-plate correlation functions centres on whether the bump features are real or due to systematic errors. At the moment it is impossible to discriminate between these two possibilities. Further detailed examination of the residual systematics in the final EDSGC is required together with modelling the effect that such systematics have on $\omega_{gg}(\theta)$. There are, however, several features that indicate that the bumps may be due to enhanced clustering:

(i) Bump features occur at scales both larger and smaller than the plate size ($\sim 5°$), and so they cannot be due to residual plate to plate errors, but may be due to intra-plate errors. The lower amplitude of bump features in the $b_j < 20$ sample, where most of the data comes from intra-plate pairs, argues against this.

(ii) Bump features are seen in some correlation functions derived from individual plate datasets (Fig. 5.13), taken from the 6-plate mosaic. The plate datasets with large bump features do seem to show enhanced clustering. F349, which is part of the linear supercluster structure described in Section 5.3.2, has the largest bump *amplitude* of all the correlation functions shown in Fig. 5.13.

(ii) The 24-plate mosaic has the largest plate-to-plate variation in limiting magnitude of any of the EDSGC datasets presented. This however shows no bumps in the derived correlation function. (Fig. 1.7). The region chosen for this sample is located on the east side of the 35-plate mosaic (Fig. 5.2). It has already been noted that there is less clustering observed in this region than on the west side (Fig. 5.3).

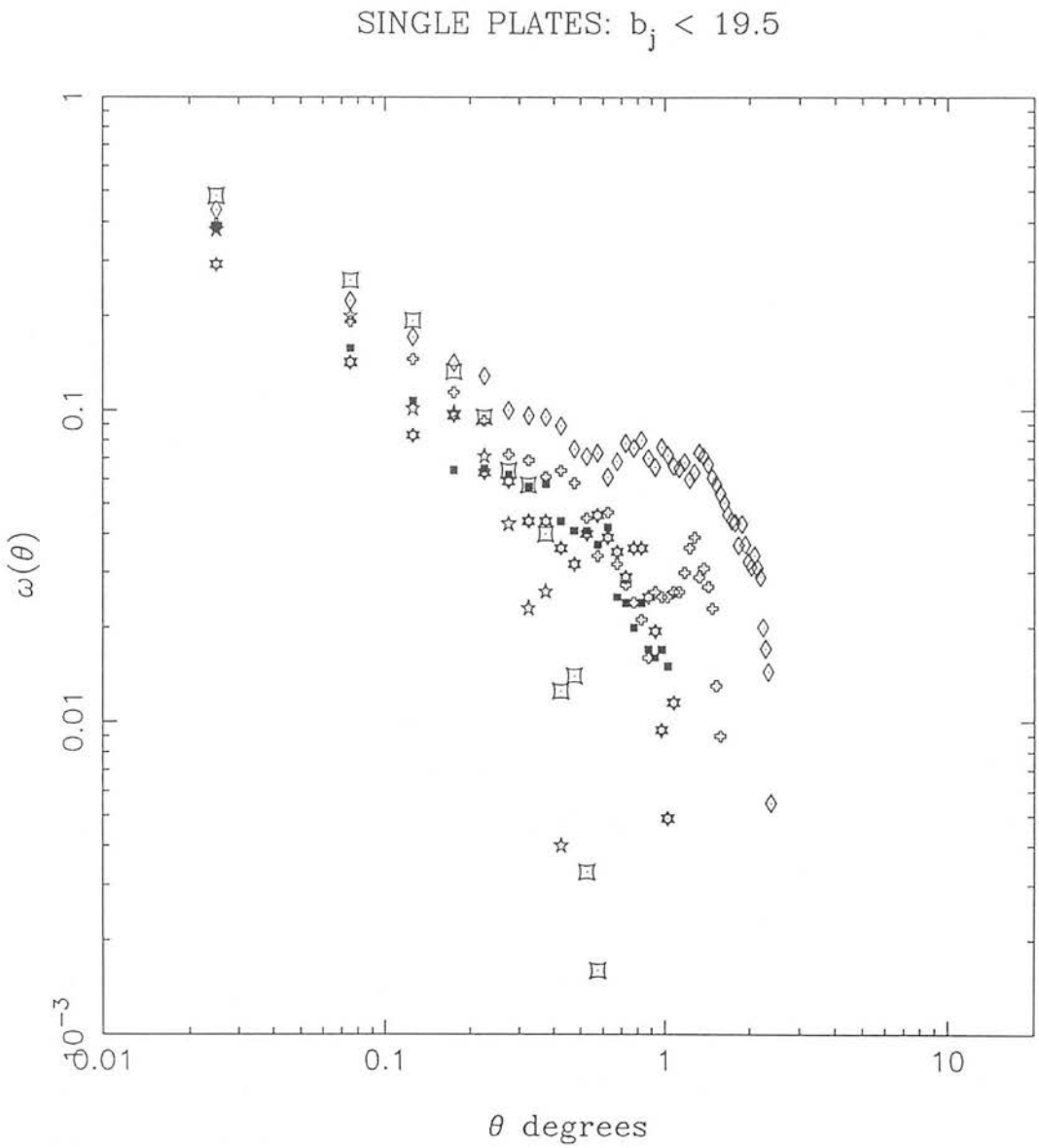


Figure 5.13: Correlation functions from the single plate datasets that make up the 6-plate mosaic. Data from F349 (diamonds), F350, F409, F410, F472 and F473 is shown. The functions are calculated using the all pairs technique, Eqn. 1.3, with bright stars drilled from the random datasets.

If the bumps are real they could be related to the wealth of structure visible around the two superclusters discussed previously. The fact that the bump amplitude is largest for the $b_j < 18{\cdot}0$ may indicate the distance of the clusters that cause the bumps ($b_j = 18{\cdot}5$ implies an effective depth $\sim 360h^{-1}$ Mpc). If confirmed the variation in the correlation function (between the 24- and 35-plate mosaics) strengthens the conclusion that the 1000 deg^2 region survey is not yet a fair sample of the galaxy distribution.

The correlation functions from the EDSGC show significant clustering on scales larger than that observed in the Groth & Peebles data, though the existence of a break from power-law is confirmed contrary to the work of Geller *et al.* (1984). Integral constraints (Section 1.2.5) on the 35-plate estimates of the correlation function can be calculated from Eqn. 1.9, if the intrinsic function is assumed to be a pure power-law, or alternatively from:

$$\mathrm{B} = \left(\frac{2\mathrm{A}}{2-\gamma}\right) \times \frac{\theta_{\mathrm{c}}^{2-\gamma}}{\theta_{\mathrm{R}}^{2}}, \tag{5.3}$$

if the intrinsic function has a break such that $\omega_{gg} = 0$ for $\theta \geq \theta_{\mathrm{c}}$. The 35-plate mosaic has an area of $\sim$ 1000 deg^2, which is equivalent to a radius (θ_{R}) of 18°. At $b_j = 20_{\cdot}$ the Groth and Peebles correlation function has an amplitude of $\sim 0{\cdot}025$ (Shanks *et al.* 1984). If the intrinsic function has a pure power-law form (slope $\sim -0{\cdot}71$), then the integral constraint for the $b_j < 20$ data is $\sim 0{\cdot}005$. This value is insufficient to remove the break from the $b_j < 20$ data. The corrections for the integral constraint calculated with Eqn. 5.3 are negligible ($\lesssim 0{\cdot}002$) for all the 35-plate datasets, and cannot remove the observed break. It is interesting to note that though the Groth & Peebles correlation function does not show significant clustering on scales larger than 7°, the SSGP maps do show structures separated on larger scales. It is therefore possible that though the the existence of uncorrected zero cells in the Groth and Peebles data (Section 1.4) does not create the break, it does move it to smaller scales.

Theoretical correlation functions for CDM and CDM+X models (Section 1.3.4) from the work of Bond and Couchman (1988) are shown in Fig. 1.7. Confirmation of a break in $\omega_{gg}(\theta)$ lends some support to the CDM model, which does predict a significant turn-

down in amplitude on these scales, though the angular correlation function for biased CDM calculated by Bond and Couchmann has significantly less power than the EDSGC function. The theoretical function drops away slowly whereas the observed function breaks sharply. The most important conclusion that can be drawn from a comparison with theory is that confirmation of the break limits the amount of large-scale power in the mass fluctuation spectrum in any theory of galaxy formation. Such power may be required to explain other observations of the large-scale structure, *e.g.* streaming flows (Collins *et al.* 1986, James *et al.* 1987) and the cluster-cluster correlation function (Bahcall & Soneira 1983). If these measurements are correct (c.f. Sutherland 1988) then extra large-scale power in the CDM fluctuation spectrum is required, if CDM is to remain a viable galaxy formation model (Bardeen *et al.* 1987). Bond & Couchman (1987) calculated the effect such a hybrid power spectrum has on the predicted $\omega_{gg}(\theta)$. Their result is reproduced as the dashed line marked CDM + X in Fig. 1.7. Since the main effect of the extra large-scale power is to remove the break in $\omega_{gg}(\theta)$ its confirmation here contradicts models with a standard CDM fluctuation spectrum and extra large–scale power.

5.4.3 FUTURE APPLICATIONS

The final version of the EDSGC is expected to contain information of $\sim 1{\cdot}5 \times 10^6$ galaxies, or $\sim 2 \times 10^5$ to the completeness limit ($b_j < 20{\cdot}0$). The careful data reduction techniques employed in the production of the EDSGC have produced a database that can be used over a wide range of astronomical applications. Results on the differential galaxy number–magnitude counts and the two-point angular correlation function have already been discussed. Here some of the possible future applications of the EDSGC are presented. Work on some of the applications is already in progress and, where possible, observational constraints are noted. Once the EDSGC has been finished and final estimates of the homogeneity calculated, it is hoped that the data will be made freely available to the astronomical community. This will be done either by supplying

data on request, or by including the EDSGC in a common-user database such as the STARLINK STADAT system.

The EDSGC supersedes the *Lick Catalogue* because positions, magnitude and shape parameters (27 parameters in all) are stored for each individual image. It is therefore possible to use the EDSGC to probe clustering on scales smaller than the $10' \times 10'$ resolution of the *Lick Catalogue* and to examine the clustering properties of galaxies with different properties. Examples of the first of these applications include the clustering of close pairs of galaxies or the distribution of galaxies within clusters. These are made possible in the EDSGC because the deblending software has significantly improved the completeness of the COSMOS datasets in crowded regions. Investigation of the alignment of galaxies and galaxy position angles in and around clusters is an example of the second of these applications (see Binggeli 1982, MacGillivray & Dodd 1984, Argyres *et al.* 1986 and Kindl 1987). The EDSGC will be able to investigate such alignments further because it contains objectively defined parameters for a sample of galaxies considerably larger than those studied before. A further advantage is that it is possible with the EDSGC database to use objective techniques to define cluster centres and axes, rather than rely on subjective methods used previously.

There are several alternative clustering statistics to the two-point correlation function described in this thesis. Different statistics have different advantages and disadvantages. For example the two-point correlation function is not very sensitive to clustering in filamentary structures, though the ease with which it is calculated is a big advantage over other methods. Rudnicki (1988) list several of the available statistics and comments on how sensitive each is to various forms of clustering. As the EDSGC retains the position of each individual galaxy, it is possible to use it to calculate any clustering statistic. This is not the case for the binned data of the *Lick Catalogue.*

It is possible to cross correlate the distributions of galaxies in the EDSGC with other catalogues, *eg.* clusters (Abell *et al.* 1988) or IRAS galaxies. Cross-correlation of those

IRAS point sources identified as galaxies (Wolstencroft *et al.* 1986) with the EDSGC data will yield useful information on the clustering environments of such galaxies. The co-added IRAS point source catalogue is soon to be made available (C Lonsdale, private communication) and is $\sim 2^{\mathrm{m}}.5$ deeper than the previous point source catalogue. The EDSGC can provide optical identifications for some of these sources.

The Zwicky catalogue (CGCG) has provided the basis for the CFA redshift surveys. This catalogue is however based primarily in the northern hemisphere and and only includes galaxies brighter than $m_{pg} = 15{\cdot}5$. The EDSGC could be used as the basis for comparative redshift surveys in the southern hemisphere. Using the EDSGC such surveys could be extended to significantly fainter limiting magnitudes and, therefore, sample the spatial distribution of galaxies with fainter intrinsic luminosities.

Recent developments in fibre optic technology enable redshifts to be obtained for several tens of galaxies per observation. The accuracy of the plate and sky coordinates of galaxies in the EDSGC ($1{\cdot}2\,\mu$m and $2{\cdot}0''$ respectively), make it an ideal source of data for observations with fibre systems. The FLAIR2 fibre optic system on the UKST will have 64 fibres that can be arranged over the telescope field of view ($\sim 6{\cdot}4° \times 6{\cdot}4°$). It is estimated (Watson *et al.* 1989, in preparation) that redshifts can be obtained for 80% of the targeted galaxies ($b_j < 16{\cdot}75$) in a 3 hour exposure. A proposal has been made (T Shanks, private communication) to survey 4000 galaxies from a $b_j < 17{\cdot}0$ magnitude limited sample of the EDSGC. It is estimated that this project could be completed in two years given $\sim$ 60 grey/dark nights per year. Preliminary work has also been carried to assess the possibility of using the OCTOPUS fibre system on the ESO 4m system. This system has 30-50 fibres covering a $0{\cdot}25$ deg^2 field, so can optimally survey the redshifts of galaxies down to $b_j = 19$. L. Guzzo (private communication) reports that with manual positioning of the fibres, it is possible to observe 6 fields per night (1hr exposures, redshifts accurate to 30 km/s). If automatic positioning of the fibres is available it will be possible to observe about 9 fields per night, *ie.* obtain $\sim$ 400 galaxies redshifts per night. A proposal (L. Guzzo & M. Tarenghi, private

communication) has been made to survey a slice of data from the EDSGC covering $\sim 3° \times 70°$, which contains $\sim 2 \times 10^4$ galaxies. An advantage of this survey over the FLAIR proposal, is that it will enable the mapping of the distribution of low-luminosity galaxies (absolute magnitude ≤ 18) out to a distance of $250h^{-1}$ Mpc . This will provide a test of some biased galaxy formation scenarios, which predicted that such galaxies should be uniformly distributed.

Redshift surveys are useful for determining the spatial topology of the large-scale structure, but as shown by Kaiser (1986), the estimation of the power in $\xi_{gg}(r)$ on large scales is carried out more efficiently if redshifts are obtained for a random sample of galaxies. The optimum *sparse-sampling* rate for a $b_j < 17$ magnitude limited sample from the EDSGC is 1-in-20, which implies that redshifts should be obtained for ~ 650 from the EDSGC. This will provide as much information on the correlation function on large scales as would be obtained by surveying all the $b_j < 17$ galaxies in the EDSGC, but in $1/20^{\text{th}}$ of the time. To date redshifts have been obtained for a complete sparse-sample of $b_j < 15{\cdot}0$ galaxies.

The previously observed distribution of galaxy clusters and their correlation function ($\omega_{cc}(\theta)$) was discussed in Section 1.3.4 together with some of the implications of these observation for galaxy formation theories. It has, however, been suggested that some of the observed clustering properties are due to selection effects in the cluster detection mechanism (Section 1.4), rather than the intrinsic distribution of galaxies. As many previous cluster catalogues are based on visual searches, such selection effects are difficult to quantify.

It will be possible with the EDSGC to create several different cluster catalogues based on different detection techniques, *e.g.* Turner & Gott (1976), Shectman (1984), MacGillivray & Dodd (1986) and Murtagh (1986). In this way, it will be possible to quantify the effect of the various selection techniques on the observed distribution. Sutherland (1988) has proposed that the amplitude of $\omega_{cc}(\theta)$ derived from the Abell clusters is

too high, because of the over estimation of the richness of distant clusters that lie behind the halos of nearby clusters. With the EDSGC it will be possible to map the distribution of foreground and background galaxies and hence quantify this and other such projection effects.

A cluster detection algorithm has been developed to produce an objective catalogue of Abell type clusters from the EDSGC data. This algorithm applies the COSMOS analyser to the binned galaxy distribution to detect candidate clusters. These candidates can either be peaks above some global threshold number density or peaks above some threshold set according to local background number density. The distribution of galaxies within each candidate (a set of connected pixels), is used to define a cluster centroid and an Abell radius (Table 1.3). Galaxies lying within this radius define the Abell cluster with the richness estimated using maps of the background number density. This algorithm has only been applied to a small region of the EDSGC, so no detailed results have been obtained. Most of the Abell clusters in this region were detected by this algorithm, though it occasionally splits the Abell cluster into separate sub-clusters (see Section 1.4).

Preliminary work has been carried out to estimate the cluster–cluster spatial correlation function, $\xi_{cc}(r)$, from an objective sample of clusters (in collaboration with L. Guzzo and C. Collins). The feasibility of obtaining redshifts from EFOSC observations with the ESO 4m telescope has been tested using EDSGC data from around Abell clusters. One advantage of EDSGC data is that it is possible to prepare accurate pointing centres and integration times in advance. With favourable observing conditions it is possible to measure redshifts for 10 clusters per night, with at least 15 galaxies observed per cluster. To date redshifts for 35 clusters (richness class ≥ 1, distance class ≤ 4) have been obtained. This number of cluster redshifts will not provide useful information on $\xi_{cc}(r)$, but coupled with the EDGSC galaxy magnitudes, will provide useful information on the galaxy luminosity function.

5.5 Thesis Summary

In this final chapter, the construction of a large part of the EDSGC has been described and some preliminary results presented, based on its application to the galaxy number–magnitude counts and two-point correlation function. To sum up the work presented in this thesis, the conclusions of each of the preceeding chapters are listed briefly below:

Chapter 1: Previous galaxy and cluster catalogues have been constructed from visual scans of photographic plates (*e.g. The Lick Catalogue*, Shane & Wirtanen 1967). Some of the features in these catalogues may be due to residual systematic errors, rather than the intrinsic large-scale galaxy distribution. In particular Geller *et al.* (1984) and de Lapparent *et al.* (1986) have suggested that the filamentary appearance of the Seldner *et al.* (1977) maps and the break in the two-point angular correlation function (Groth & Peebles 1977) are due to variations in the counting efficiency of Shane and Wirtanen. The break feature in $\omega_{gg}(\theta)$ lends some support to the biased CDM model of galaxy formation, but is not visible in theoretical correlation functions from CDM models with extra large-scale power (CDM+X) in the mass fluctuation spectrum. Confirmation of the existence of a break implies that such hybrid models cannot be used to account for other observations such as the amplitude of $\omega_{cc}(\theta)$ or streaming flows.

Chapter 2: The raw photographic data for the *Lick Catalogue* vary considerably in quality and exposure from plate to plate and contain extra confusing images due to the use of an objective prism grating. The raw photographic data for the EDSGC are *ESO/SERC Atlas* glass copy plates of SERC J survey plates taken with the UKST. The care taken during the production, processing, copying and quality control of these plates removes those that are underexposed, defocussed, trailed or have an unacceptable number of plate flaws. These plates are, therefore, more uniform than any set used previously to construct a galaxy catalogue.

Image detection in the EDSGC is carried out using the COSMOS plate measuring machine, rather with than visual searches of plates. 27 parameters, including position and magnitude, are recorded for on average 2×10^5 images per plate. Noise images constitute $< 2\%$ of this dataset ($b_j < 20$). Image deblending techniques have been implemented to improve the completeness of COSMOS datasets in crowded regions (*e.g.* clusters), while not breaking up stellar diffraction spikes or filter ghosts whose parameters mimic those of merged galaxies. The EDSGC will therefore provide an excellent starting point for the construction of a catalogue of clusters and their properties.

Chapter 3: The methods employed in the EDSGC to perform accurate star–galaxy image classification are discussed. The techniques adopted involve correcting the measured magnitudes of saturated (stellar) images for field effects (variation across the plate) and then automatically classifying images according their position in one of several classification parameter versus magnitude planes. Classification using each parameter is carried out only over the magnitude range where it efficiently classifies images. Use of a parameter outside its optimum magnitude range (determined separately for each plate) would produce a galaxy catalogue that is incomplete and/or has a large stellar contamination. Using these techniques, it is possible to produce a galaxy catalogue ($b_j \leq 20$) that is $\gtrsim 95\%$ complete, with $\lesssim 10\%$ stellar contamination and $\sim 3\%$ residual variation from plate to plate in the number of objects classed as galaxies.

Chapter 4: The global calibration method employed by SSGP in their correction of the raw Lick counts can propagate errors from a defective plate to the surrounding region and may not fully remove spurious structures associated with systematic errors. These techniques are not used in the calibration of the EDSGC. The COSMOS calibrated magnitude, m_{cos}, is linearly related to galaxy b_j magnitudes by $m_{cos} = b_j - m_{sky}$, where m_{sky} is the plate background (mag/arcsec^2). This is expected to hold for most regions and galaxies on a plate, but m_{cos} may underestimate the magnitude of high surface brightness galaxies or where variations in the plate background are not equivalent to sensitivity changes, *e.g.* in filter-corrector/emulsion corrector ghosts associated with

extremely bright stars. Intra-plate photometry has been estimated to be uniform to $\lesssim 0^{\rm m}.03$ across a plate, while the magnitudes of individual galaxies are accurate to better than $0^{\rm m}.3$ rms.

The uniformity of inter-plate photometry and hence the limiting magnitude of the EDSGC is set by the accuracy with which m_{sky} can be estimated for each field. This has been carried out by obtaining a CCD galaxy sequence for every second field. Those fields without a sequence therefore have 4 overlaps to fields with a sequence. The four independent estimates of m_{sky} for such plates can be used to detect any erroneous overlaps and check the consistency of CCD measurements on the adjacent fields. The variation of the limiting magnitude from plate to plate is estimated to be $\sim 0^{\rm m}.05$ rms.

Chapter 5: The construction of a database covering a contiguous region of the sky is problematic, because the extent of the overlap regions between each pair of plates varies. The production of the plate-mosaic is carried out by determining the extent of the north and west overlap regions of each plate, removing them and then appending the resulting dataset to the mosaic. Spurious regions of the mosaic associated with bright stars, nearby galaxies and star clusters are also drilled out at this stage. To date a 35-plate mosaic has been constructed covering ~ 1000 deg^2 and containing $\sim 300,000$ galaxies to $b_j \leq 20{\cdot}0$. The variation in the limiting magnitude of this mosaic, due to systematic variations in image classification and photometry, is estimated at $\sim 0^{\rm m}.055$ rms from plate to plate and $\sim 0^{\rm m}.06$ from point to point. These compare favourably with the plate-to-plate variation limit of $\sim 0^{\rm m}.05$ set by Geller *et al.* and the $\sim 0^{\rm m}.08$ point-to-point variation in the *Lick Catalogue.*

The large-scale structure visible in the maps produced from this 35-plate mosaic is dominated by two supercluster structures separated by $\sim 15^\circ - 20^\circ$ on the western half of the survey. One of these superclusters is a linear structure that can be traced over several fields, while the other is more spherical with a diameter of $\sim 15h^{-1}$ Mpc . Several arc structures can be seen, with clusters distributed along them like knots on

a piece of string.

The differential galaxy number–magnitude counts for the 35-plate mosaic have $\gamma \simeq 0{\cdot}6$ for $b_j < 18{\cdot}75$. The counts for $18{\cdot}75 \leq b_j \leq 20$ have a shallower slope, but more galaxies are detected than estimated by no-evolution models. The variation in amplitude of the counts from plate to plate is larger than can be accounted for by systematic variations in the limiting magnitude. Hence the scatter in amplitude of the counts from plate to plate and from author to author is probably due to the large-scale clustering of faint galaxies.

Preliminary estimates of the two-point angular correlation function show a break from power-law and therefore contradict the theoretical function predicted from CDM+X models of galaxy formation. The position of the break is at larger scales ($\sim 20h^{-1}$ Mpc) than estimated by Groth & Peebles ($\sim 10h^{-1}$ Mpc), though the exact position is difficult to determine due to bumps on the EDSGC correlation function on scales $\simeq$ the break scale. These bumps may be associated with the enhanced clustering around the previously mentioned superclusters, but detailed modelling of the effects of systematic errors on $\omega_{gg}(\theta)$ is required to confirm this.

When finished the EDSGC will contain information on $\sim 1{\cdot}5 \times 10^6$ galaxies and thus will provide an ideal basis to investigate many different aspects of the angular and spatial distribution of galaxies. An important future application is the construction of a catalogue of clusters based on objective detection techniques. Work has already been started to estimate $\xi_{cc}(r)$ from Abell clusters and $\xi_{gg}(r)$ from a sparse sample of galaxies.

Finally it seems appropriate to close this thesis with a quote from Argyres *et al.* (1986):

> "Obtaining an adequate sample of deeper galaxy counts would be very useful but an exceedingly large project"

Appendix A

Summary of the 35-Plate Mosaic

At the time of completing this thesis, the EDSGC consists of a mosaic of 35 plates. A full discussion of the classification and photometric techniques is given in Chapters 3 & 4 respectively, while the construction of a contiguous mosaic and the removal of bright stars is described in Section 5.2. Here the following details and statistics are presented for each individual plate in the mosaic:

FIELD: The ESO/SERC Field centre number (Section 2.2.1), prefixed 'F'.

PLATE: The SERC J survey plate number (Section 2.2.1), prefixed 'J'.

GRADE: The quality control grade assigned to the original plate. See Section 2.2.2 and Table 2.2.4

$\mathcal{A}$ *ctm*: The percentage stellar contamination estimated by the automatic classification algorithm (Section 3.2.2b), for images classed as galaxies using the $\log_{10}(\mathcal{A})$ classifier (Section 3.3.1b).

$\mathcal{S}$ *ctm*: The percentage stellar contamination estimated by the automatic classification algorithm (Section 3.2.2b), for images classed as galaxies using the $\mathcal{S}$ classifier (Section 3.3.1c).

m_{sky} : The value of the plate background (mag/arcsec^2) derived for each plate from either CCD sequences or plate overlaps (Section 4.3.3)

DRILLS: The number of bright stars drilled from each plate (Section 5.2).

AREA: The area (deg^2) subtended by each plate after addition to the mosaic, *ie.* after removal of the west and north overlaps (Section 5.2).

A list of larger regions (Section 3.5) removed from the mosaic are listed at the end of this appendix, together with coordinates and drill radii.

FIELD	PLATE	GRADE	$\mathcal{A}$ *ctm*	$\mathcal{S}$ *ctm*	m_{sky}	DRILLS	AREA
F293	J3578	AI2	9	5	22·16	7	12·3
F294	J3551	AI2	8	6	22·31	9	26·1
F295	J2712	A1	7	5	22·48	7	25·5
F296	J3524	AI2	7	5	22·49	7	26·4
F297	J3593	A2	9	5	22·29	8	25·4
F298	J3560	A2	10	5	22·41	9	25·2
F299	J3588	A2	10	6	22·24	9	25·6
F300	J3536	A2	10	5	22·34	13	26·4
F345	J3538	AT2	9	5	22·06	9	27·5
F346	J8022	AI3	9	6	22·19	4	25·7
F347	J2413	A1	8	5	22·33	7	26·6
F348	J3592	A2	8	5	22·25	4	24·9
F349	J6145	AI2	8	4	21·77	6	25·4
F350	J4601	A2	9	5	22·14	6	27·1

FIELD	PLATE	GRADE	$\mathcal{A}$ *ctm*	$\mathcal{S}$ *ctm*	m_{sky}	DRILLS	AREA
F351	J8046	AI3	18	9	22·48	5	21·2
F352	J6124	A1	9	5	22·57	6	24·9
F353	J3596	A1	9	4	22·36	8	25·9
F354	J4630	A2	9	5	22·54	8	25·8
F355	J5304	AI3	10	5	22·36	7	25·8
F356	J5462	A1	8	5	22·47	11	24·5
F405	J6231	AI2	12	6	22·22	7	27·9
F406	J9529	AUI3	11	5	22·04	13	25·3
F407	J6468	A2	13	6	22·14	3	25·1
F408	J10342	AI3	10	5	22·19	9	24·5
F409	J2693	A2	8	5	22·48	9	15·4
F410	J2692	A2	7	5	22·54	7	25·6
F411	J11341	A2	9	5	21·92	8	23·5
F412	J3516	AI3	8	4	22·44	5	28·4

FIELD	PLATE	GRADE	$\mathcal{A}$ *ctm*	$\mathcal{S}$ *ctm*	m_{sky}	DRILLS	AREA
F413	J3774	A1	8	4	22·42	8	26·9
F414	J3579	A2	7	5	22·29	9	27·1
F415	J4607	A1	8	5	22·40	7	27·0
F468	J6436	AI3	11	6	22·29	6	28·4
F469	J3508	AI3	8	5	22·20	12	27·0
F470	J3515	A1	9	5	22·13	8	27·1
F471	J6138	AI2	12	7	22·38	8	26·9

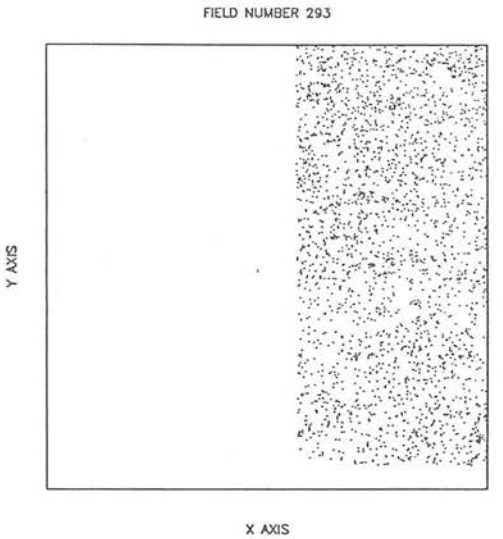
FIELD	RA (1950)	DEC (1950)	RADIUS	OBJECT
F469	23:05:06.1	-30:02:38	41′	Star
F469	23:54:51.3	-29:53:38	41′	EC/FC Ghost
F352	00:57:44.7	-34:00:00	25′	Star Cluster
F295	00:52:26.4	-37:44:06	18′	Galaxy
F356	02:37:45.1	-34:42:49	40′	Star Cluster

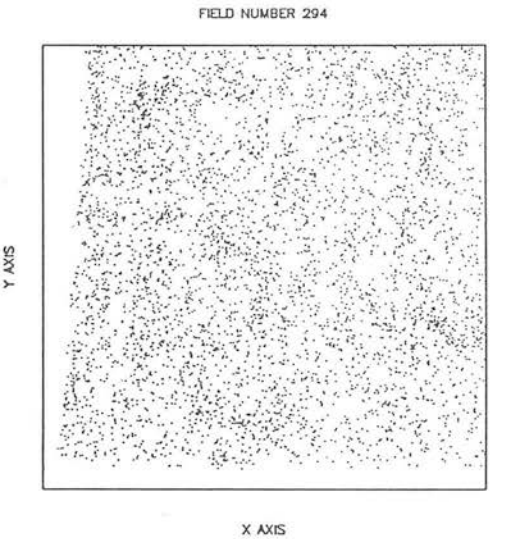
Appendix B

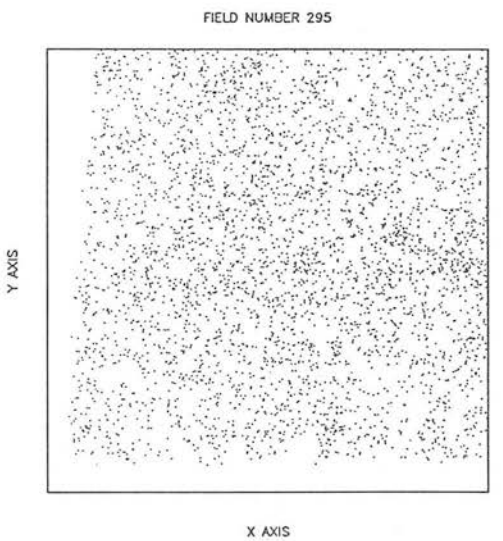
The Distribution of Galaxies on Individual Plates

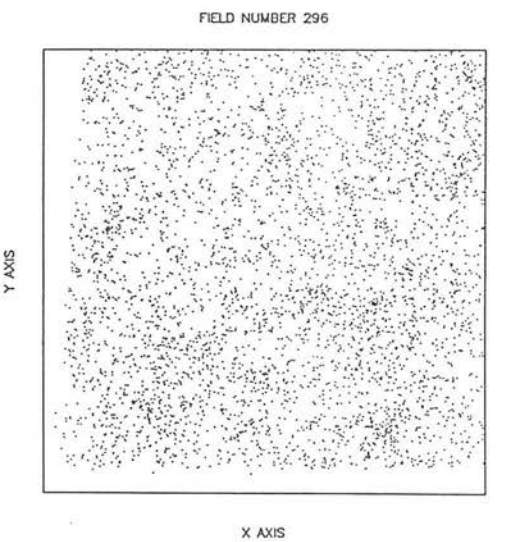
In the following appendix, maps are plotted showing the distribution of galaxies ($b_j \leq$ 19·5) across each constituent plate of the 35-plate mosaic. The regions plotted correspond to the area of each plate scanned by COSMOS, but only galaxies appended to the mosaic are shown, *ie.* the north and west overlaps with adjacent plates have been removed (Section 5.2). Galaxies are plotted in (x_i, y_i) coordinates (Section 2.3.2), with $14546\,\mu\text{m} \leq x_i \leq 301266\,\mu\text{m}$ and $39162\,\mu\text{m} \leq y_i \leq 325802\,\mu\text{m}$. North is at the top.

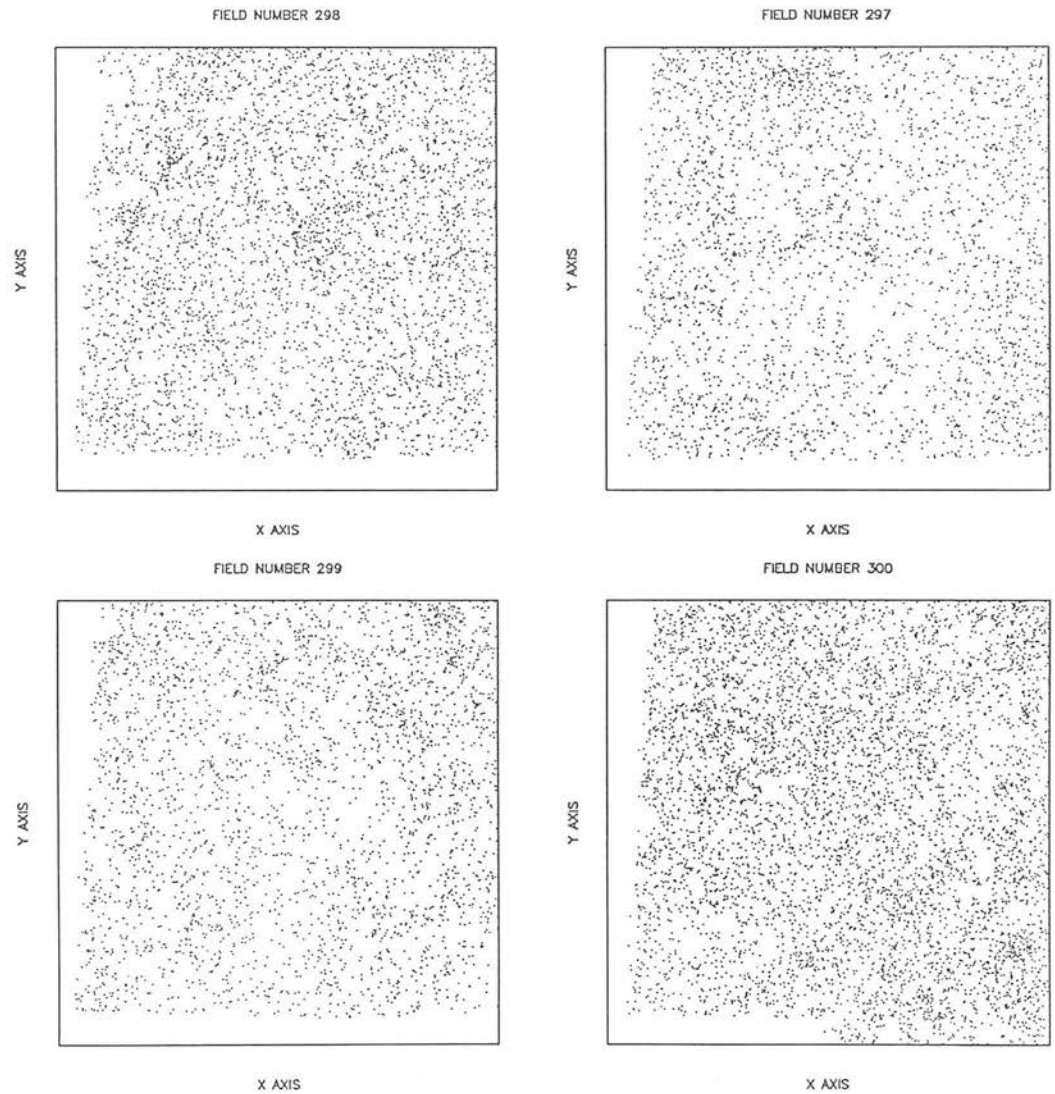
F293, F349 & F409 have larger regions removed from the west of the plates, because they are centred on RA= 0^{hrs} and have large overlaps with adjacent plates.

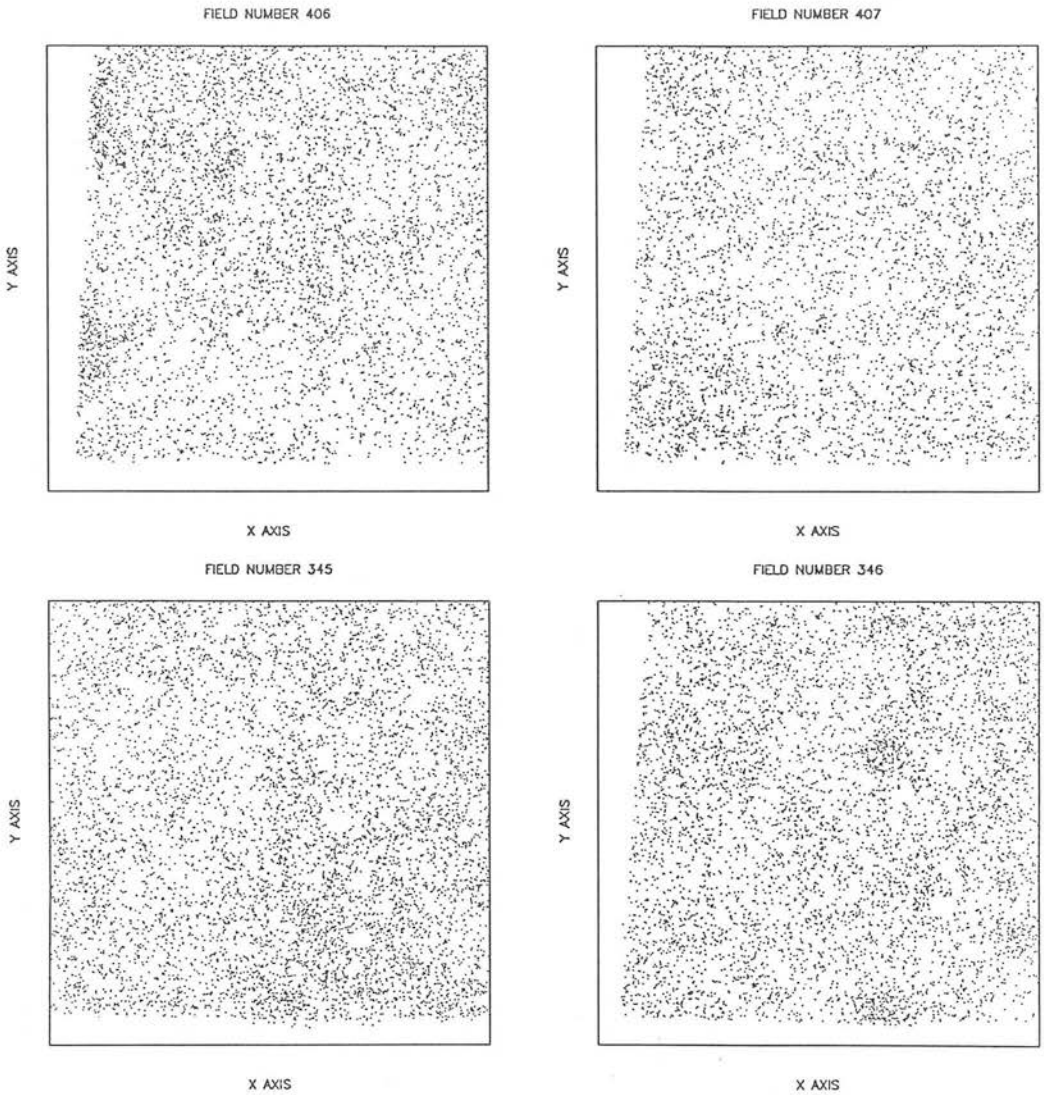

FIELD NUMBER 293
Y AXIS
X AXIS

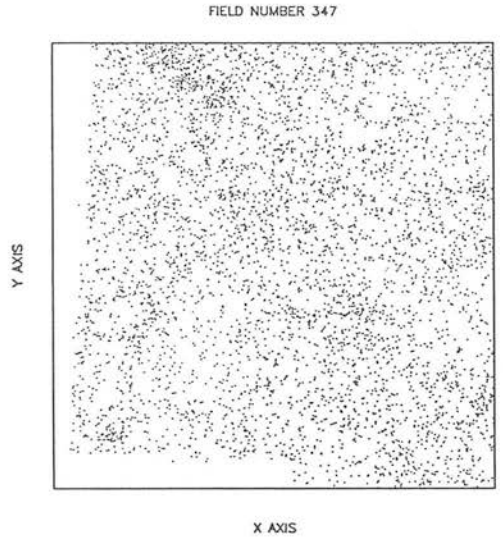

FIELD NUMBER 294
Y AXIS
X AXIS

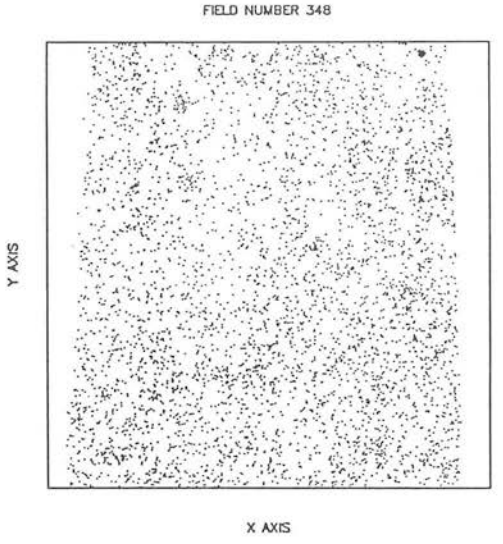

FIELD NUMBER 295
Y AXIS
X AXIS

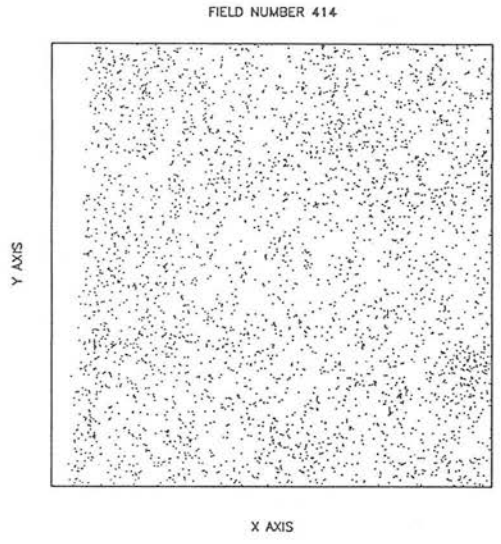

FIELD NUMBER 296
Y AXIS
X AXIS

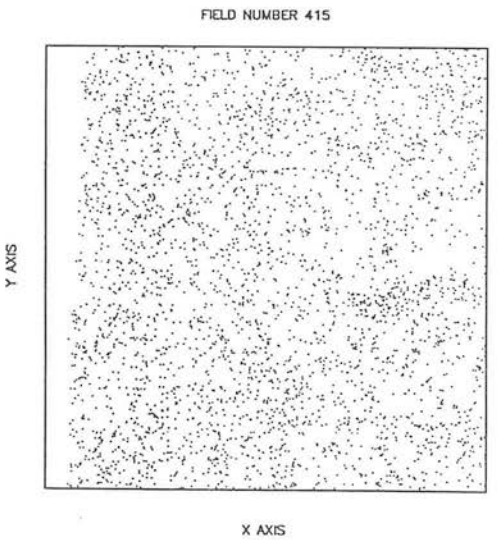

FIELD NUMBER 298
Y AXIS
X AXIS
FIELD NUMBER 297
Y AXIS
X AXIS
FIELD NUMBER 299
Y AXIS
X AXIS
FIELD NUMBER 300
Y AXIS
X AXIS

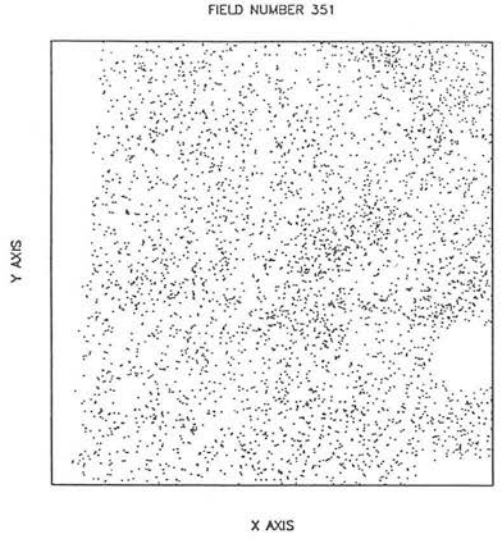

FIELD NUMBER 406
Y AXIS
X AXIS
FIELD NUMBER 407
Y AXIS
X AXIS
FIELD NUMBER 345
Y AXIS
X AXIS
FIELD NUMBER 346
Y AXIS
X AXIS

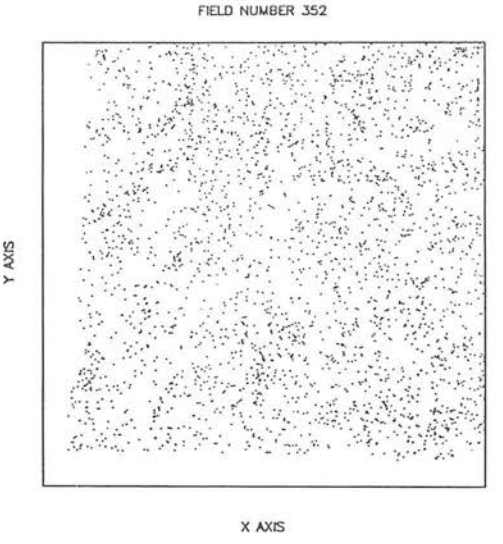

FIELD NUMBER 347
Y AXIS
X AXIS

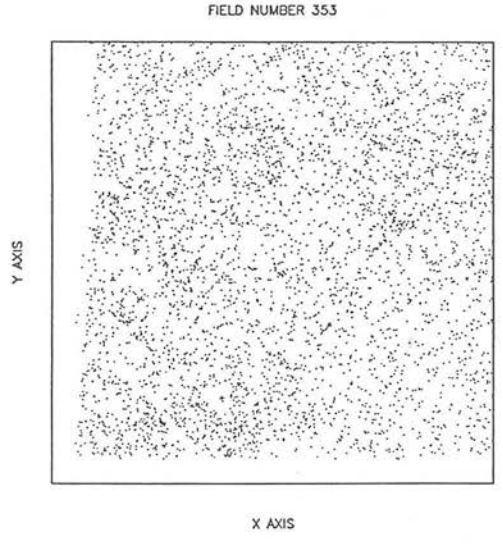

FIELD NUMBER 348
Y AXIS
X AXIS

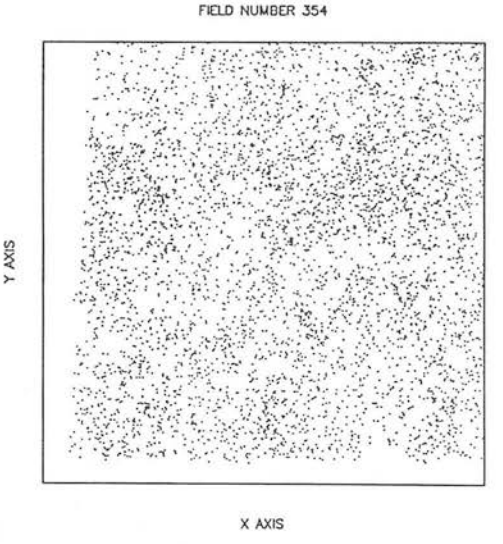

FIELD NUMBER 414
Y AXIS
X AXIS

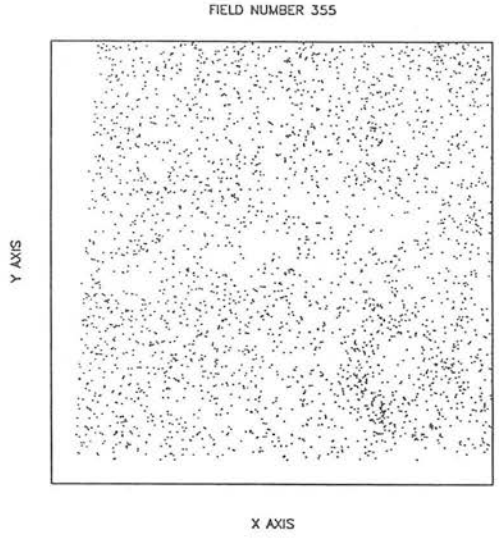

FIELD NUMBER 415
Y AXIS
X AXIS

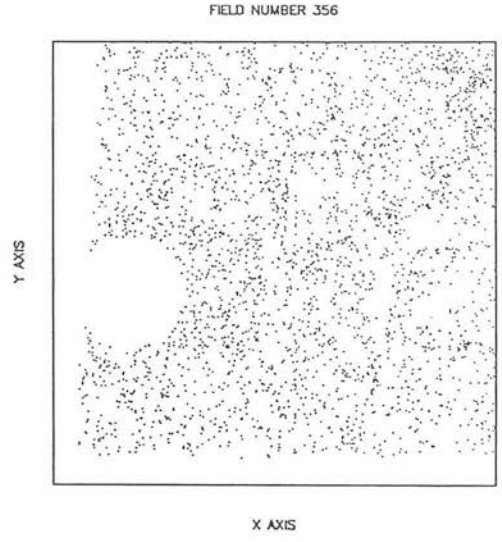

FIELD NUMBER 351
Y AXIS
X AXIS

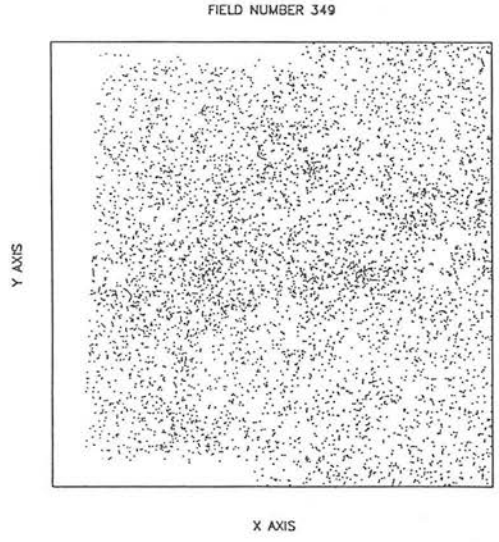

FIELD NUMBER 352
Y AXIS
X AXIS

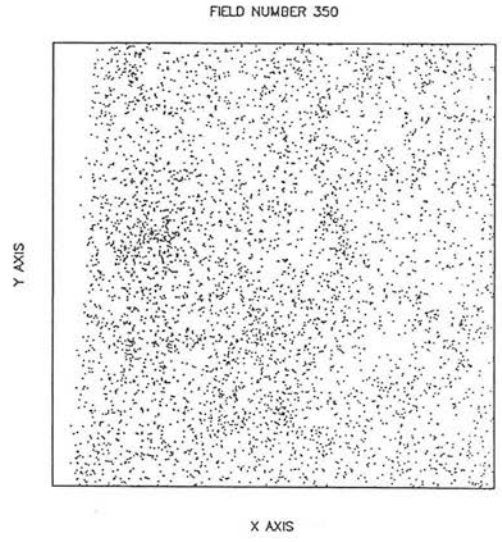

FIELD NUMBER 353
Y AXIS
X AXIS

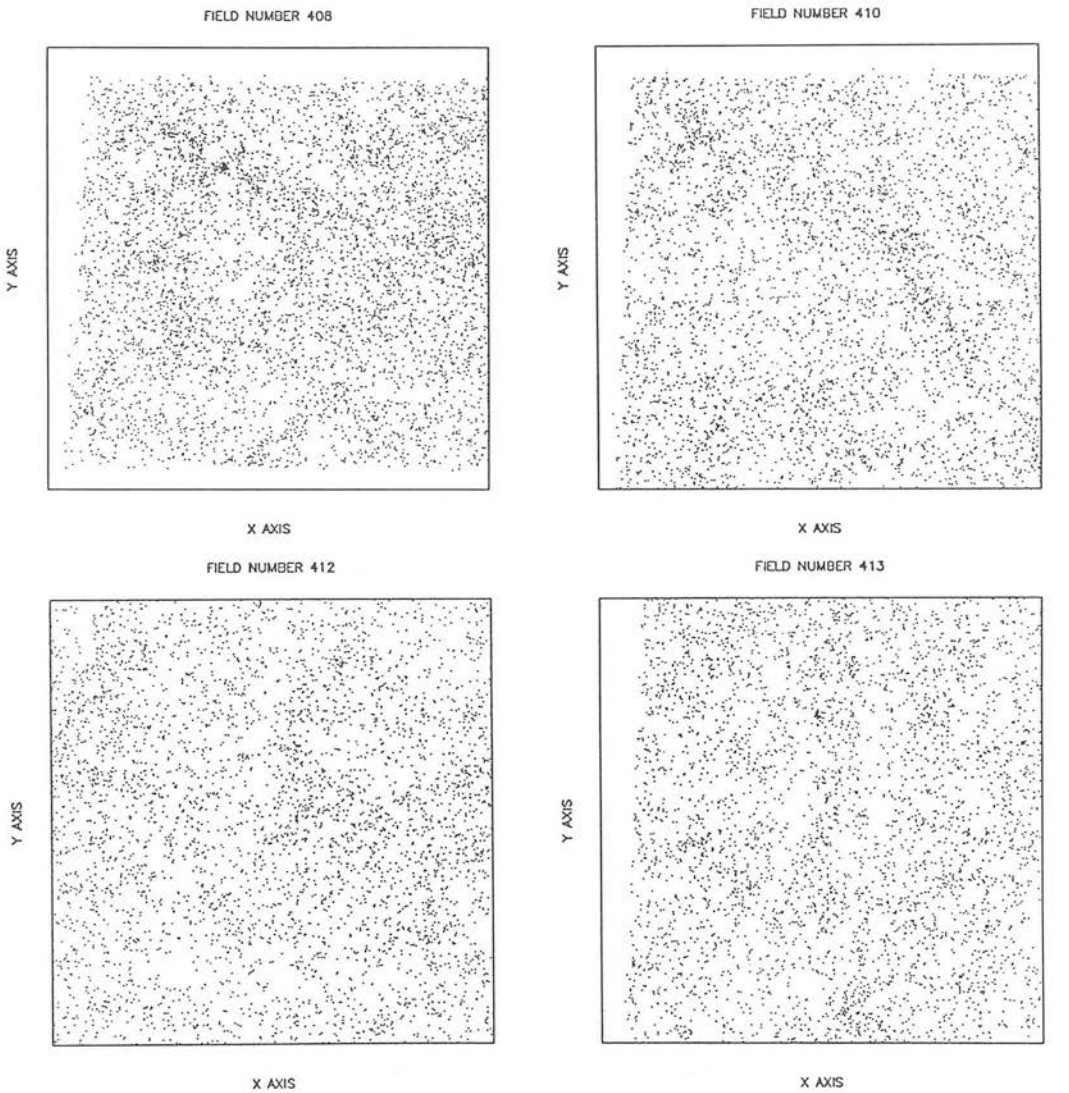

FIELD NUMBER 354
Y AXIS
X AXIS

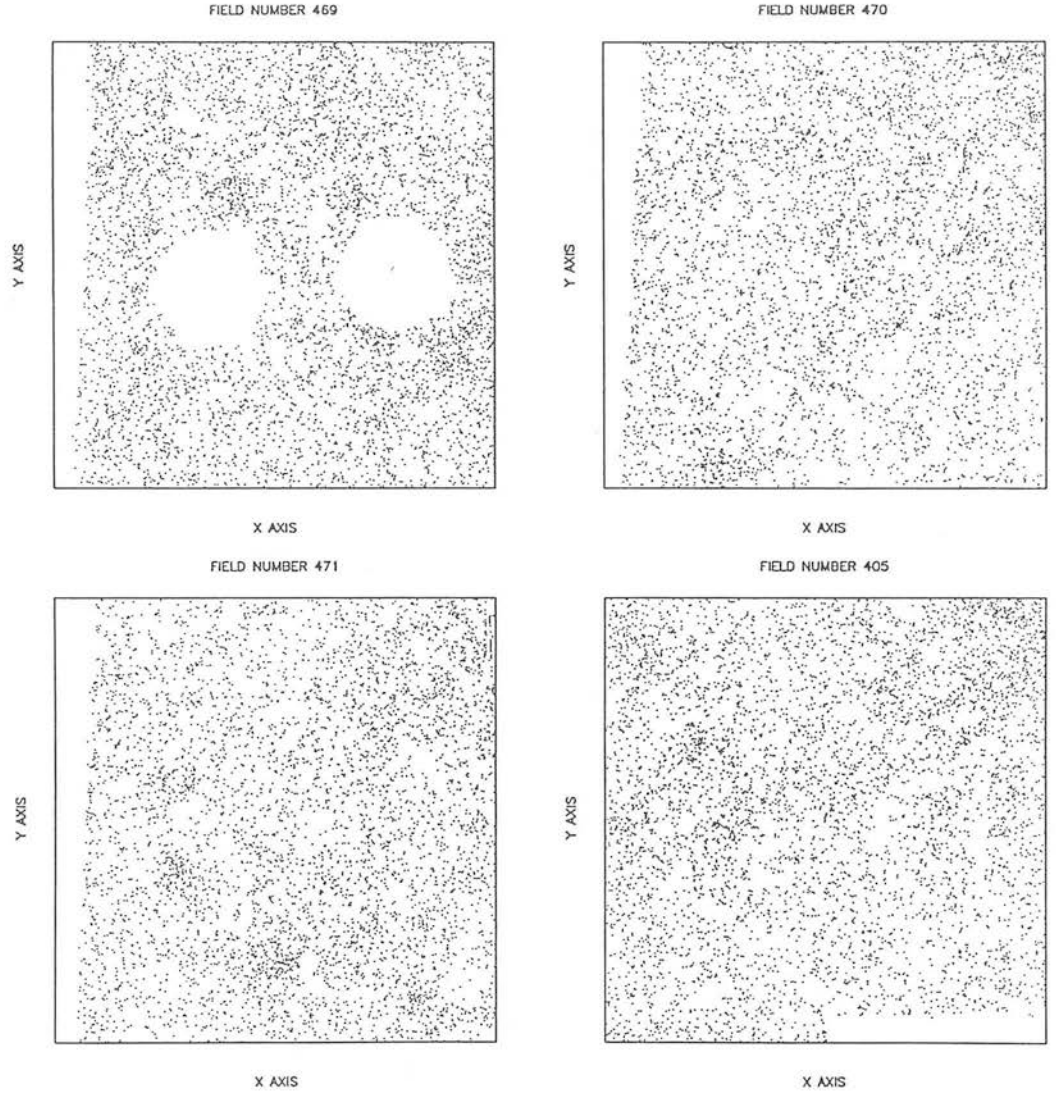

FIELD NUMBER 355
Y AXIS
X AXIS

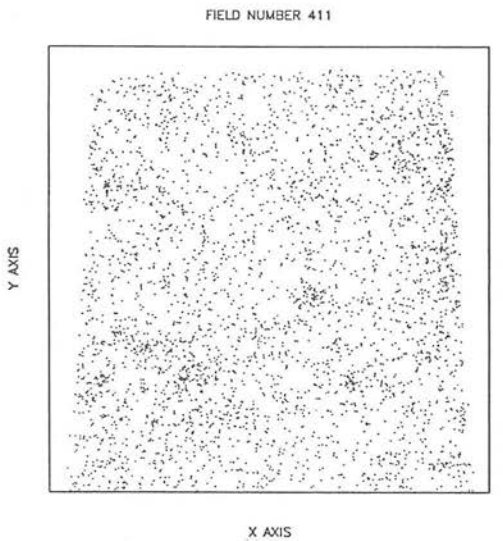

FIELD NUMBER 356
Y AXIS
X AXIS

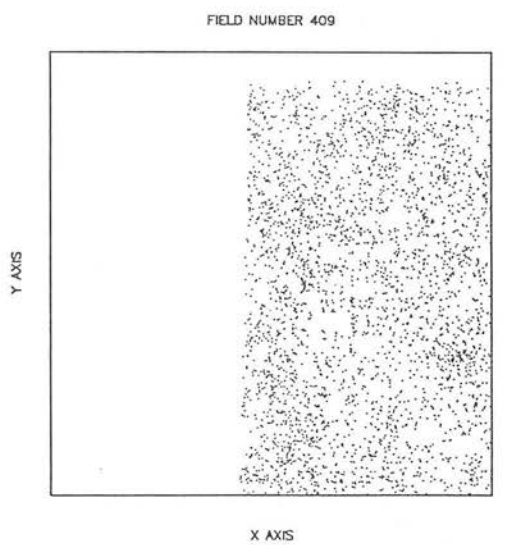

FIELD NUMBER 349
Y AXIS
X AXIS

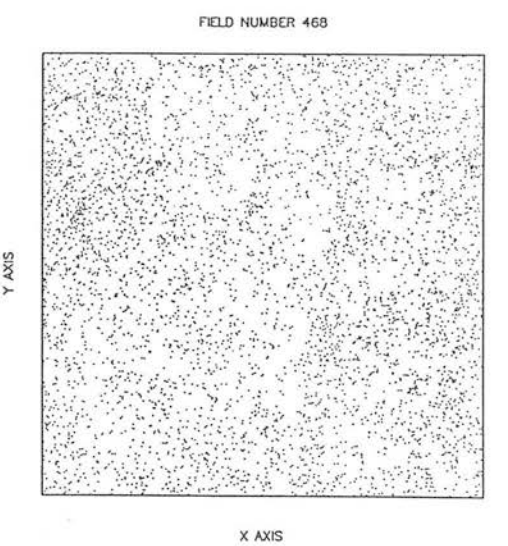

FIELD NUMBER 350
Y AXIS
X AXIS


FIELD NUMBER 408
Y AXIS
X AXIS
FIELD NUMBER 410
Y AXIS
X AXIS
FIELD NUMBER 412
Y AXIS
X AXIS
FIELD NUMBER 413
Y AXIS
X AXIS


FIELD NUMBER 469
Y AXIS
X AXIS
FIELD NUMBER 470
Y AXIS
X AXIS
FIELD NUMBER 471
Y AXIS
X AXIS
FIELD NUMBER 405
Y AXIS
X AXIS


FIELD NUMBER 411
Y AXIS
X AXIS

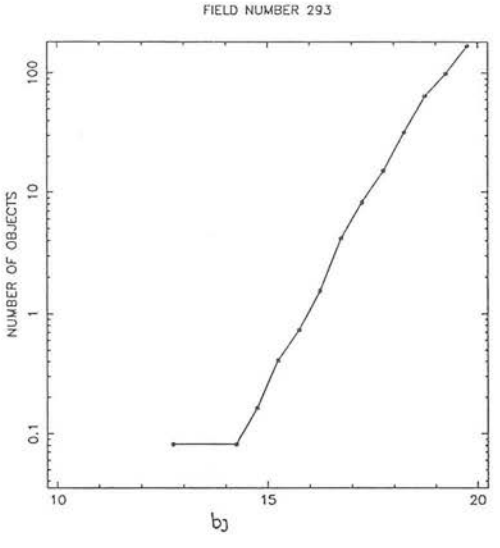

FIELD NUMBER 409
Y AXIS
X AXIS

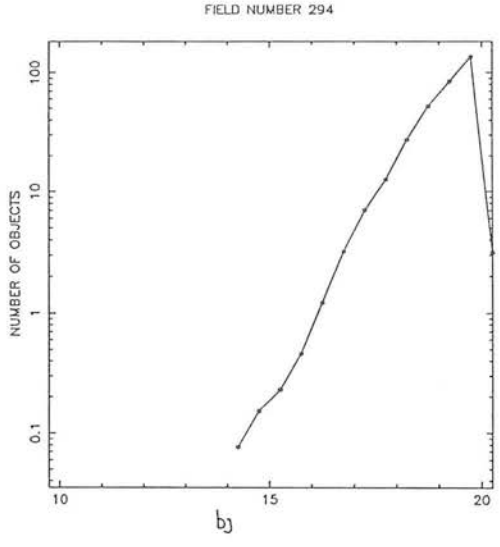

FIELD NUMBER 468
Y AXIS
X AXIS

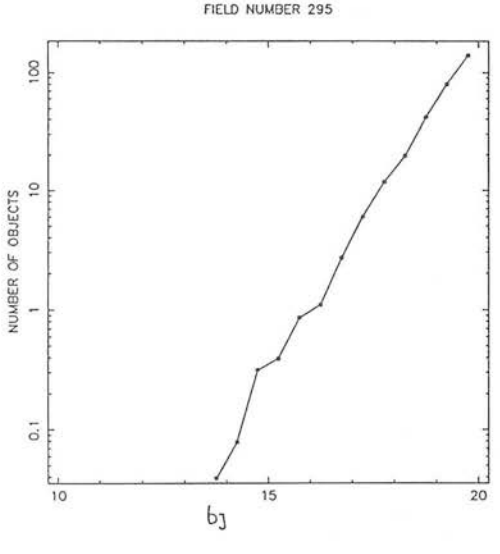
Appendix C

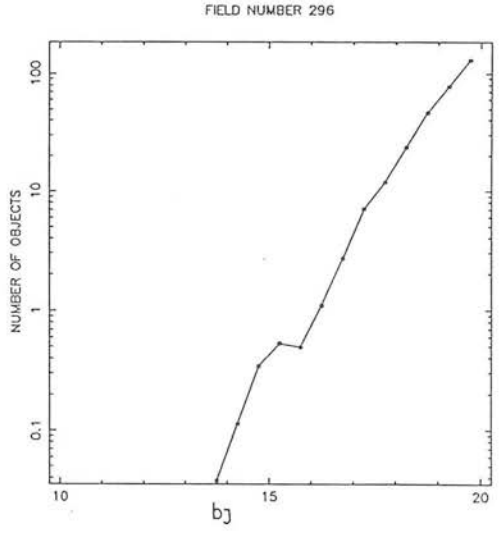
Number–Magnitude Counts for Individual Plates

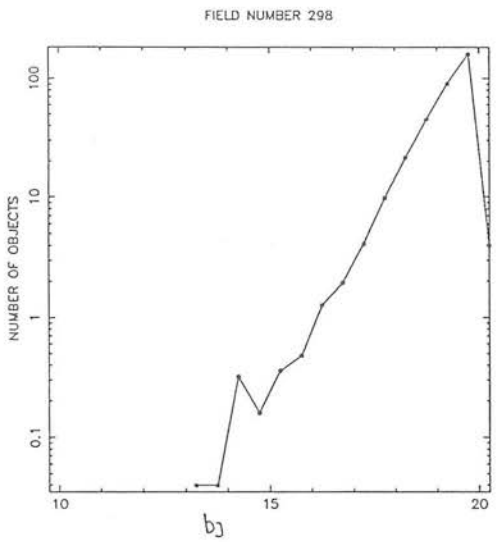
In the following appendix differential number–magnitude counts ($N/0^{m}.5/deg^2$) are shown for each of the constituent plates of the 35-plate mosaic. Counts are shown for all galaxies with $b_j \leq 20{\cdot}0$ in the regions plotted in appendix B, *ie.* the overlap regions have been removed. Points in the following plots at $b_j = 20{\cdot}25$ are due to rounding errors and the data should be added back into the bin centred on $b_j = 19{\cdot}75$. The data presented here is discussed in Section 5.4.1.

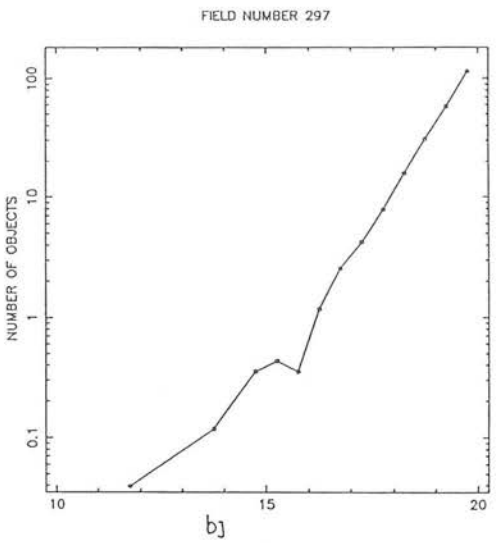

FIELD NUMBER 293
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bJ

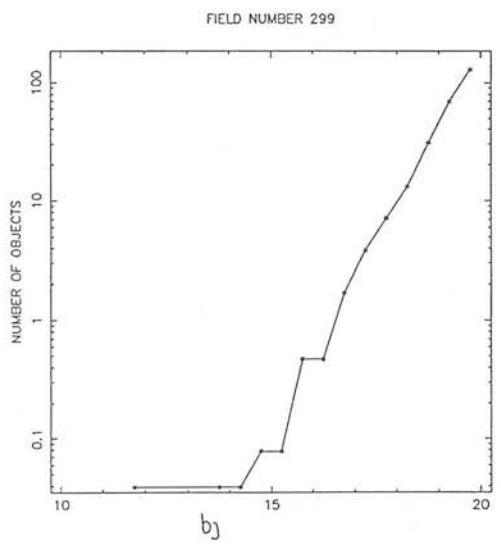

FIELD NUMBER 294
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bJ

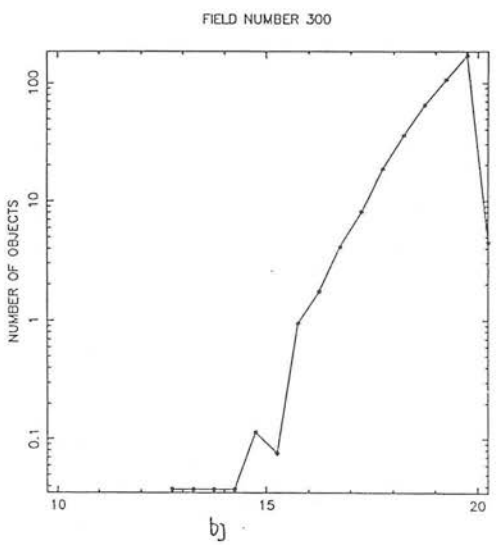

FIELD NUMBER 295
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bJ

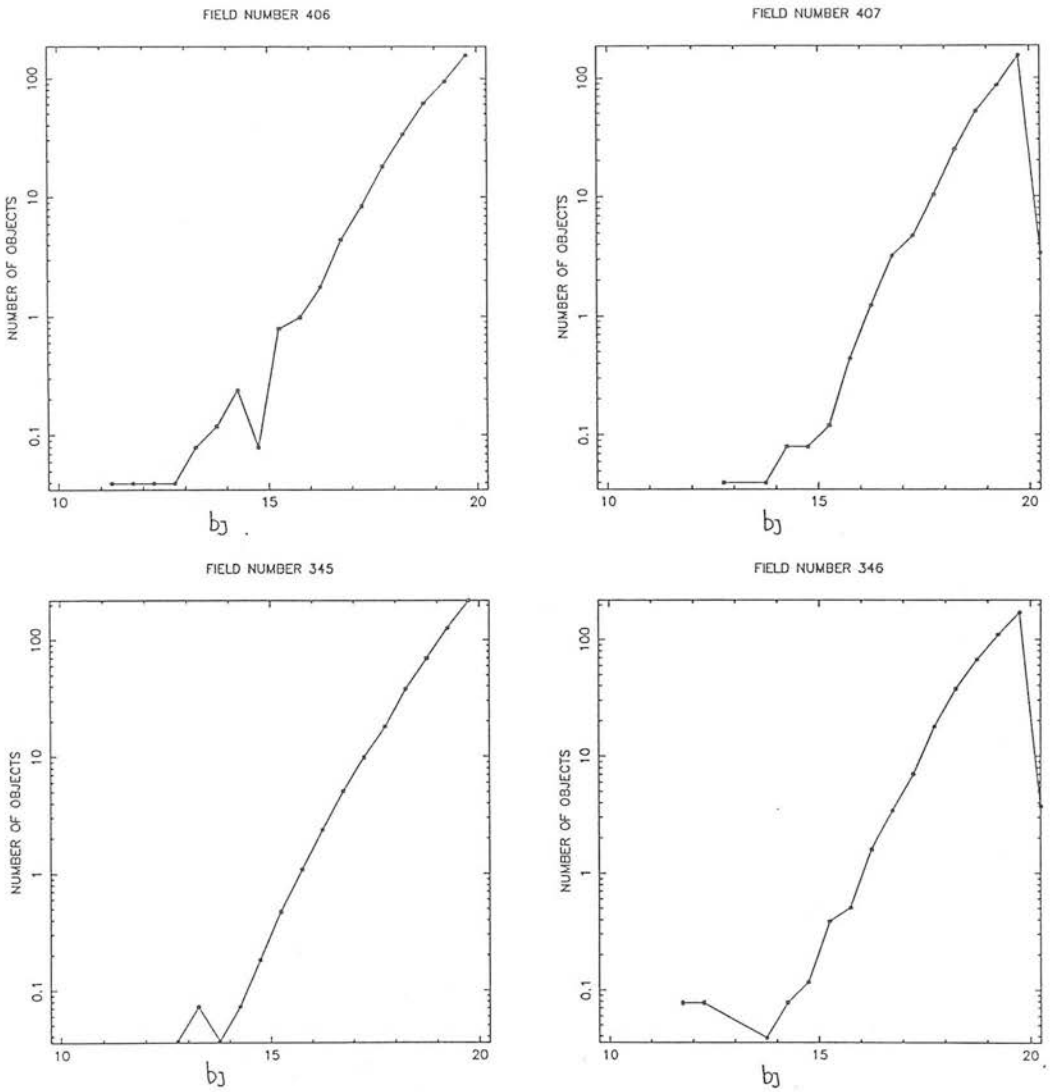

FIELD NUMBER 296
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bJ

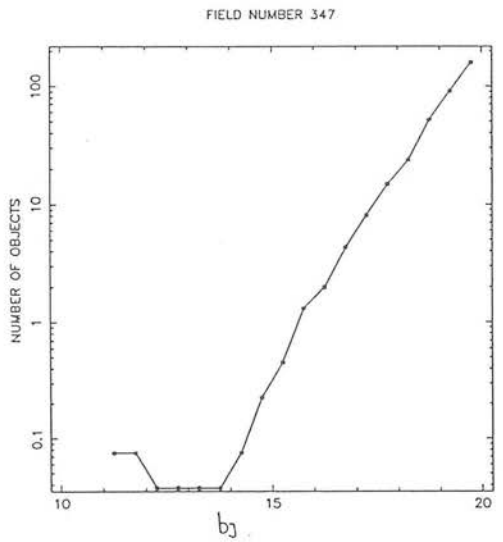

FIELD NUMBER 298
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bJ

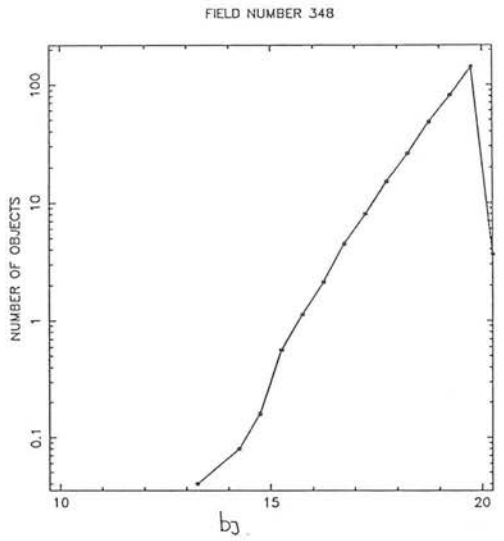

FIELD NUMBER 297
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bJ

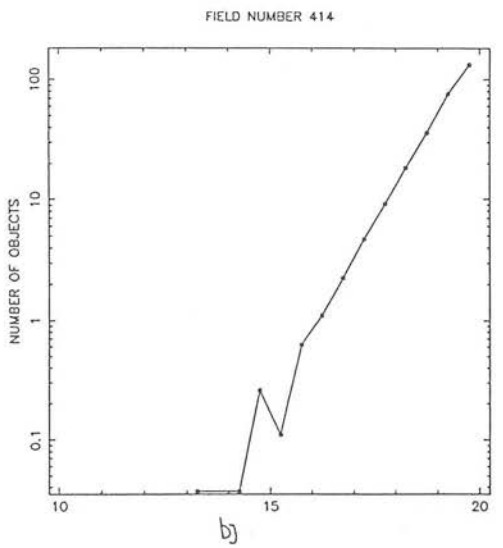

FIELD NUMBER 299
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bJ

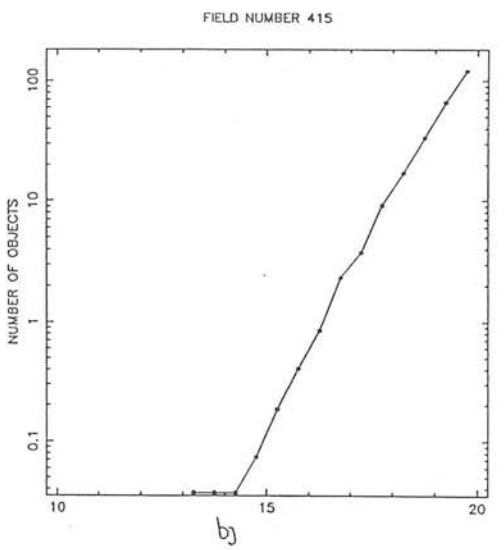

FIELD NUMBER 300
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bJ

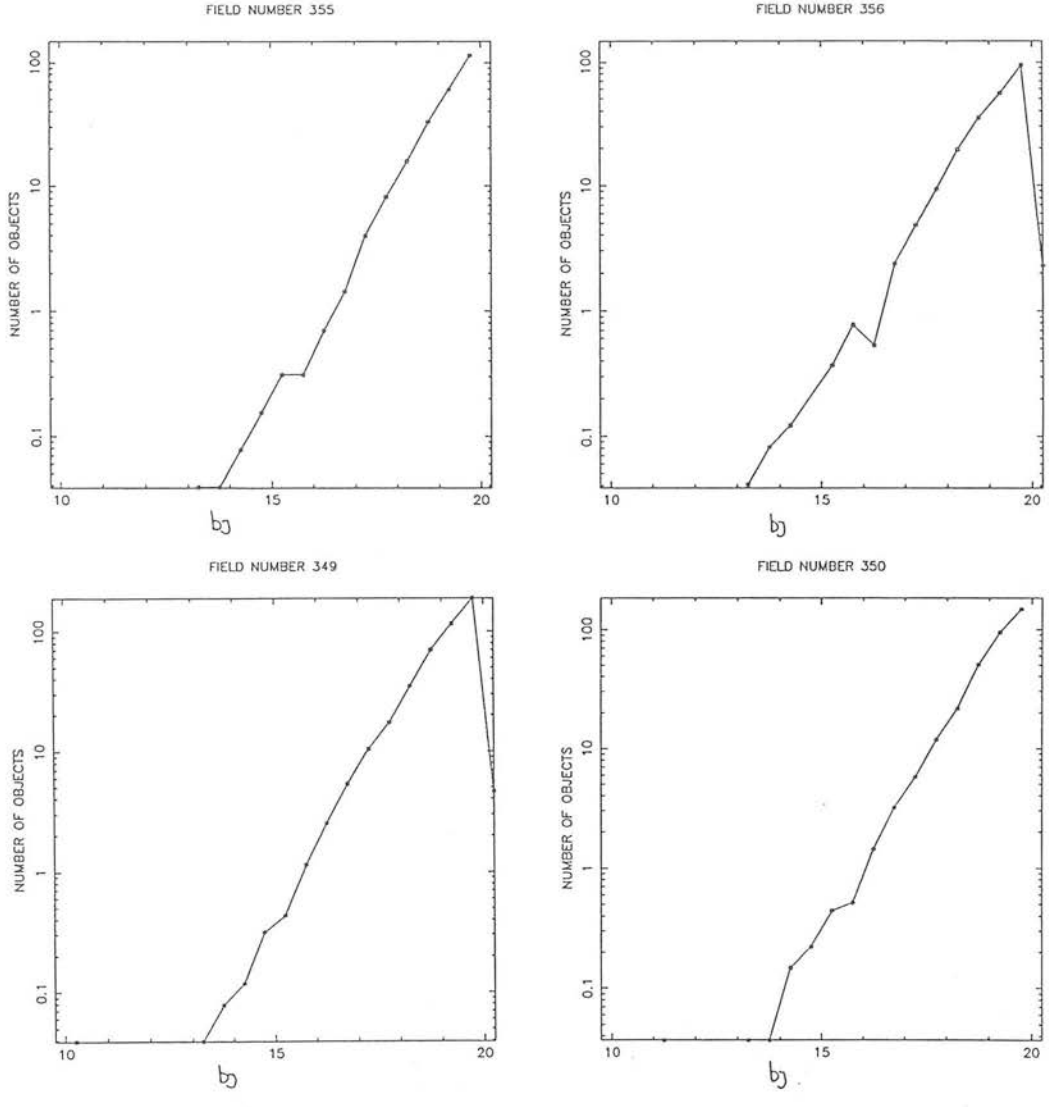

FIELD NUMBER 406
FIELD NUMBER 407
FIELD NUMBER 345
FIELD NUMBER 346
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bJ

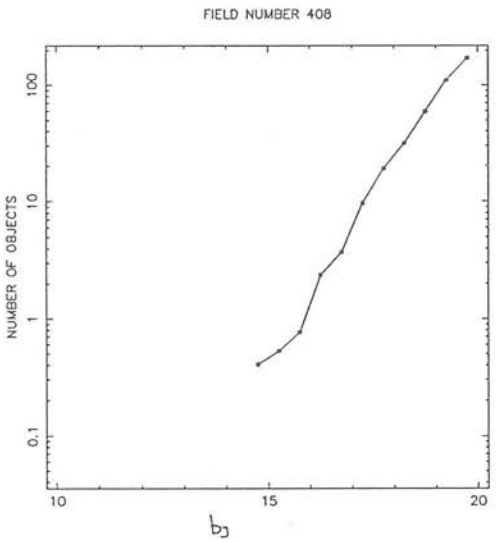

FIELD NUMBER 347
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bJ

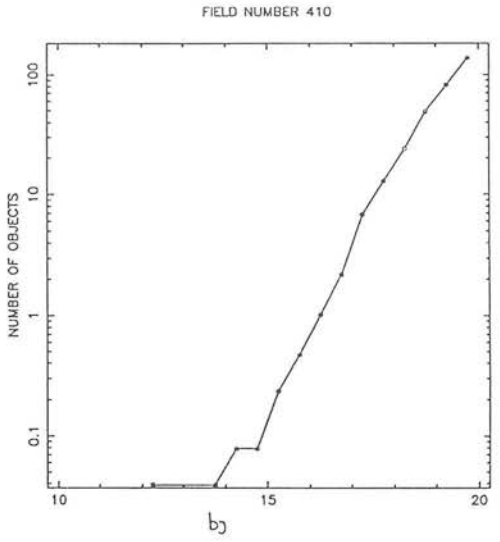

FIELD NUMBER 348
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bJ

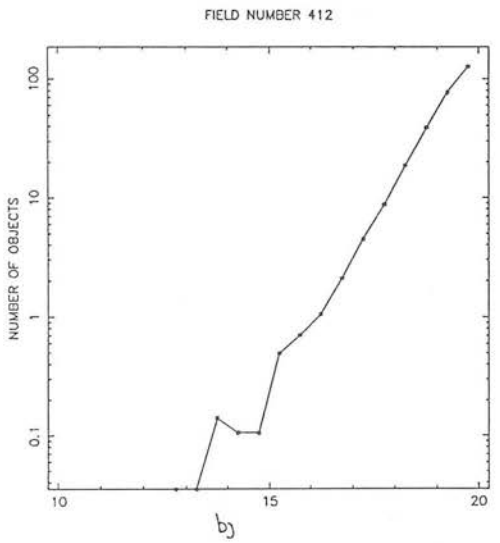

FIELD NUMBER 414
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bJ

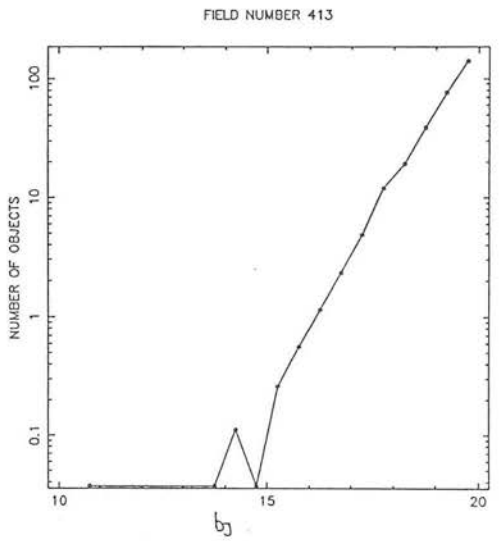

FIELD NUMBER 415
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bJ

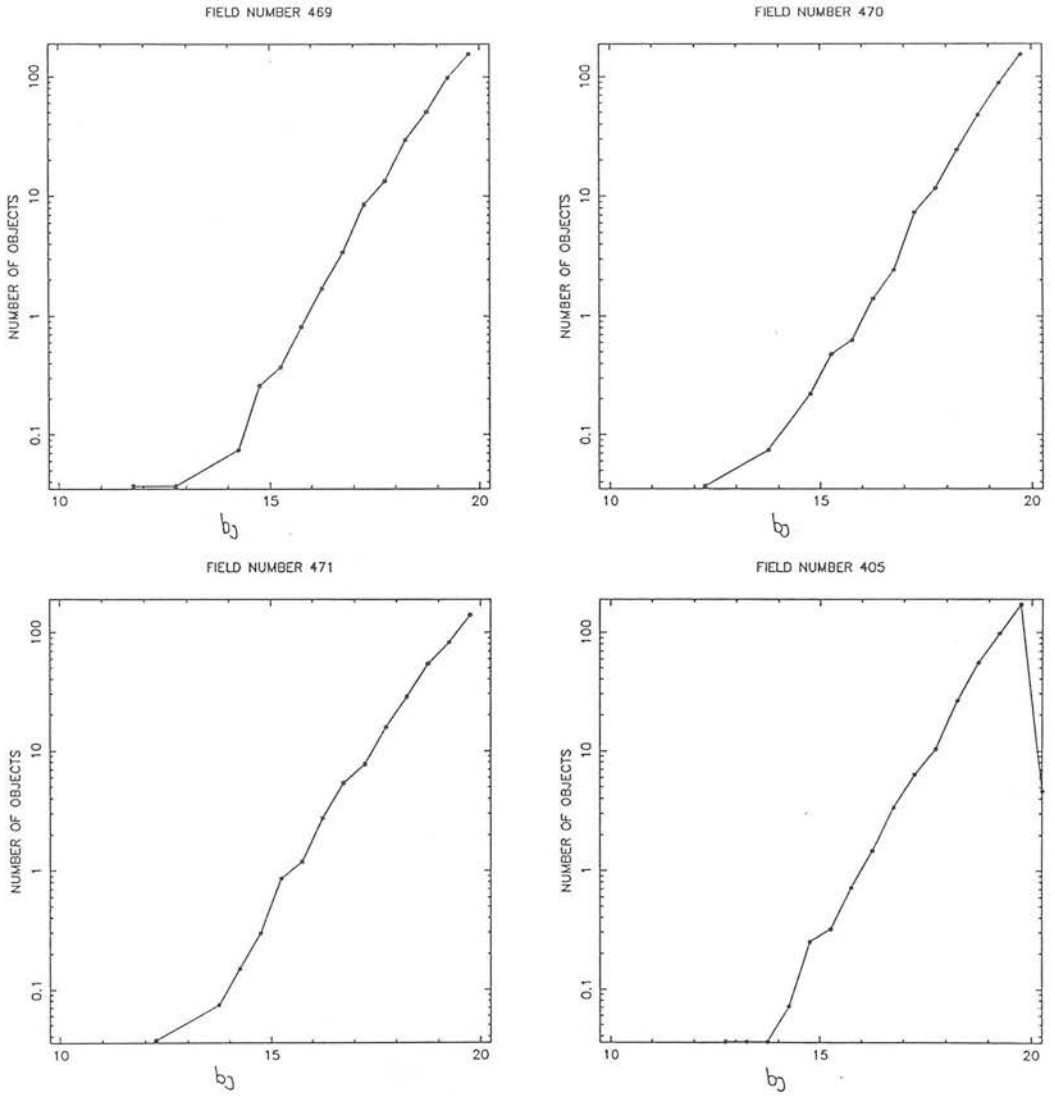

FIELD NUMBER 355
FIELD NUMBER 356
FIELD NUMBER 349
FIELD NUMBER 350
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bj

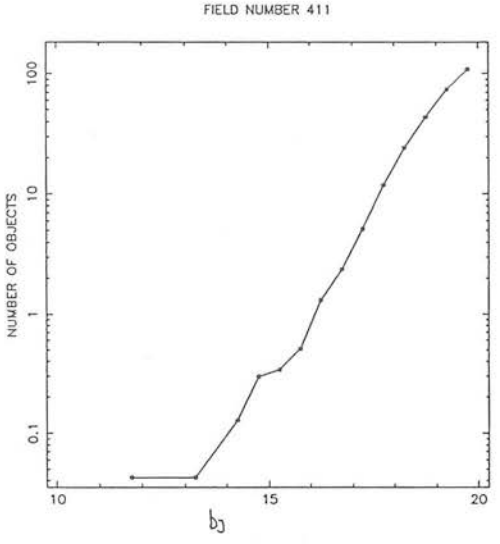

FIELD NUMBER 408
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bj

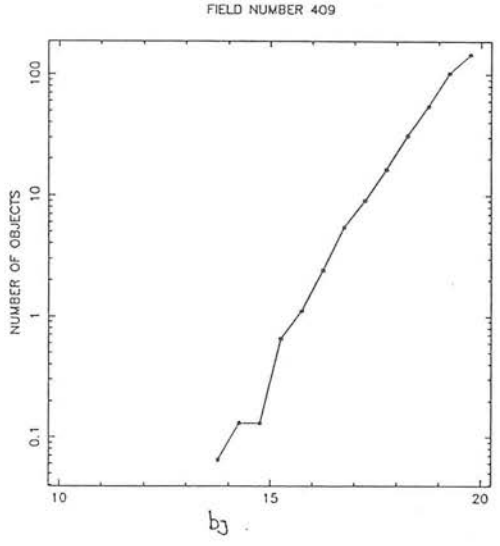

FIELD NUMBER 410
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bj

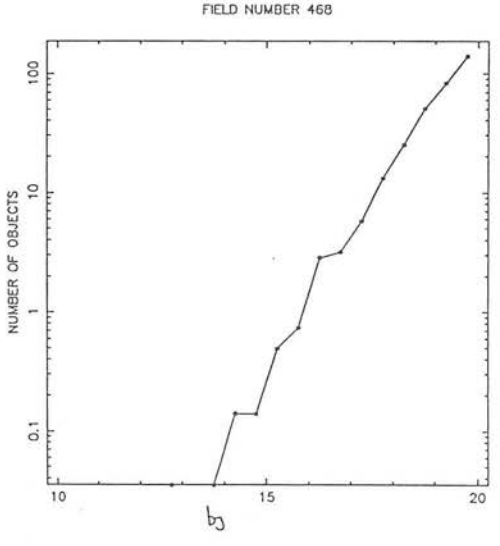

FIELD NUMBER 412
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bj

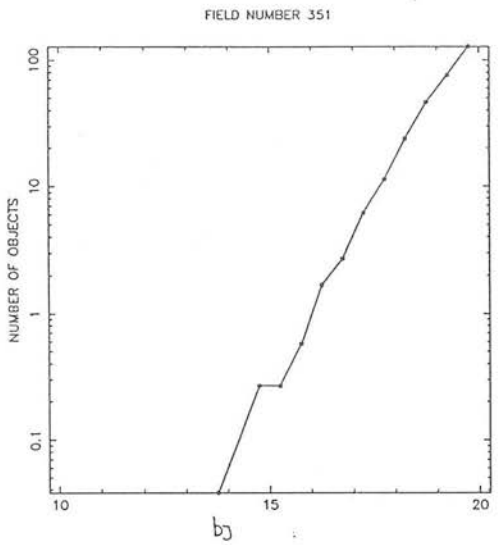

FIELD NUMBER 413
NUMBER OF OBJECTS
0.1
1
10
100
10
15
20
bj

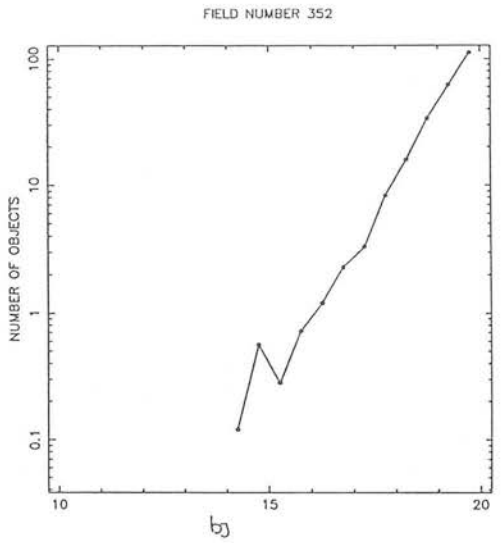

FIELD NUMBER 469
FIELD NUMBER 470
FIELD NUMBER 471
FIELD NUMBER 405
NUMBER OF OBJECTS
100
10
1
0.1
10
15
20
bJ

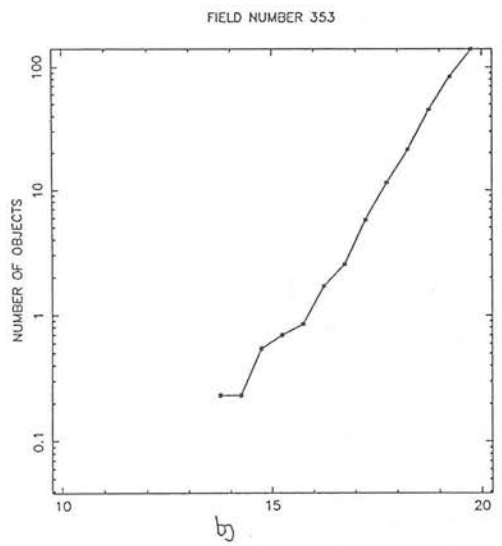

FIELD NUMBER 411
NUMBER OF OBJECTS
100
10
1
0.1
10
15
20
bJ

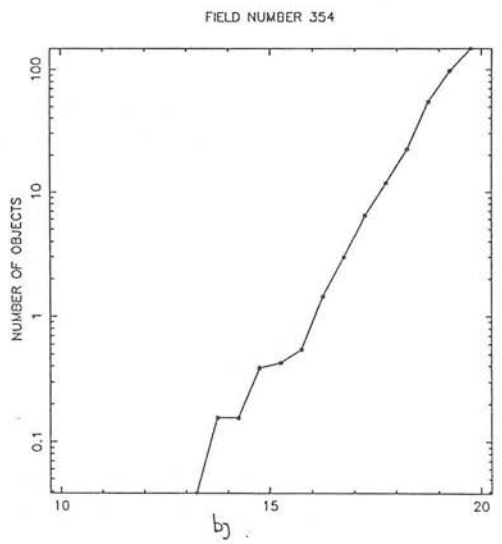

FIELD NUMBER 409
NUMBER OF OBJECTS
100
10
1
0.1
10
15
20
bJ


FIELD NUMBER 468
NUMBER OF OBJECTS
100
10
1
0.1
10
15
20
bJ


FIELD NUMBER 351
NUMBER OF OBJECTS
100
10
1
0.1
10
15
20
bJ


FIELD NUMBER 352
NUMBER OF OBJECTS
100
10
1
0.1
10
15
20
bJ


FIELD NUMBER 353
NUMBER OF OBJECTS
100
10
1
0.1
10
15
20
bJ


FIELD NUMBER 354
NUMBER OF OBJECTS
100
10
1
0.1
10
15
20
bJ

References

Abell, G.O., 1958. *Astrophys. J. Suppl.*, **3**, 211.

Abell, G.O., Corwin, H.G.Jr. & Olowin, R.P., 1988. *Astrophys. J. Suppl.*, submitted.

Argyres, P.C., Groth, E.J., Peebles, P.J.E. & Struble, M.F., 1986. *Astr. J.*, **91**, 471.

Bahcall, N.A. & Soneria, R.M., 1984. *Astrophys. J.*, **277**, 27.

Bahcall, N.A. & Soneria, R.M., 1982. *Astrophys. J.*, **262**, 419.

Batuski, D.J., & Burns, J.O., 1985. *Astrophys. J.*, **299**, 5.

Beard, S.M., MacGillivray, H.T. & Thanisch, P., 1989, *Mon. Not. R. astr. Soc.*, in preparation

Blair, M. & Gilmore, G., 1982, *Publ. astr. Soc. Pacific*, **94**, 742.

Bicay, M.D., & Giovanelli, R., 1987 *Astr. J.*, **96**, 1326.

Binggeli, B. 1982. *Astr. Astrophys.*, **107**, 338.

Bond, J.R., 1988. *Large Scale Structures of the Universe, IAU Symp. 130,* p93, eds Audouze, J., Pelletan, M.& Szalay, A., Kluwer, Dordrecht, Holland.

Bond, J.R. & Couchman, H., 1987, *Proceedings of the Second Canadian Conference in General Relativity and Relativistic Astrophysics.* eds. Coly, A. & Dyer, C., World Scientific Press, Singapore.

Brownrigg, D.R.K., 1984. *Communication of the Association for Computing Machinery,* **27**, 807.

Brown, M.E. & Groth E.J., 1988. preprint.

Brück, M.T. & Waldron, J.D., 1984. *Astronomical Photography,* Occ. Rep. R. Obs. Edin., **14**, 185, eds Sim, M. E. & Ishida, K., Royal Observatory Edinburgh.

Bunclark, P.S., 1984. *Ph. D. Thesis,* University of St Andrews.

Burns, J.O., Ward Moody, J., Brodie, J.P. & Batuski, D.J., 1988. *Astrophys. J.*, **350**, 542.

Butchins, S. A., 1982. *Astr. Astrophys.*, **109**, 306.

Campbell, A., 1982. *Observatory,* **102**, 195

Canon, R.D., Hawarden, T.G., Sim, M.E. & Tritton, S.B., 1978. *The UK* 1·2 *m Schmidt Telescope & Southern Sky Survey II, – Photographic Techniques,* Occ. Rep. R. Obs. Edin., **4**, Royal Observatory Edinburgh.

Carter, D. & Godwin, J.G., 1979. *Mon. Not. R. astr. Soc.*, **187**, 711.

Chincarini, G., Rood, H.J., & Thompson, L.A., 1981. *Astrophys. J.*, **249**, L47.

Collins, C.A., Heydon-Dumbleton, N.H. & Shanks, T., 1988. *Large Scale Structures of the Universe, IAU Symp. 130,* eds Audouze, J., Pelletan, M. & Szalay, A., Kluwer, Dordrecht, Holland.

Collins, C.A., Heydon-Dumbleton, N.H. & MacGillivray, H.T., 1989. *Mon. Not. R. astr. Soc.*, in press.

Collins, C.A., Joseph, R.D. & Robertson, N.A., 1986 *Nature* **320**, 506.

Corwin, H.G. Jr., de Vaucouleurs, A. & de Vaucouleurs, G., 1985. *Southern Galaxy*

Catalogue, Univeristy of Texas Monographs in Astronomy No. 4, Austin, Texas.
Cousins, A.W.J., 1980, *Mon. Notes Astr. Soc. Southern Africa,* **39**, 22.
Davis, M. & Djorgovski, S., 1985. *Astrophys. J.,* **299**, 15.
Davis, M. & Efsathiou, G., 1988. *Proceedings of the Vatican Study Week on Large-Scale Motions in the Universe,* preprint.
Davis, M. & Huchra, J.P., 1982. *Astrophys. J.,* **254**, 437.
Dawe, J.A. & Metcalfe, N., 1982. *Proc. astr. Soc. Australia,* **4**, 466.
Dawe, J.A., Coyte, E. & Metcalfe, N., 1984. *Astronomical Photography,* Occ. Rep. R. Obs. Edin., **14**, 59, eds Sim, M. E. & Ishida, K., Royal Observatory Edinburgh.
Dickey, J.M., Keller, D.T., Pennington, R. & Salpeter, E.E., 1987. *Astr. J.,* **93**, 788.
Dodd, R.J. & MacGillivray, H.T., 1986. *Astr. J.,* **92**, 706.
Dreyer, J.L.E., 1888. *Mem. R. A. S.,* **49**, 1. (NGC)
Dreyer, J.L.E., 1893. *Mem. R. A. S.,* **51**, 185. (Index 1)
Dreyer, J.L.E., 1910. *Mem. R. A. S.,* **59**, 105. (Index 2)
Ellis, R.S., 1987. *Observational Cosmology, IAU Symp. No. 124,* 367, eds Hewitt, A., Burbidge, G., & Fang, L. Z., Reidel, Dordrecht, Holland.
Fong, R., Jones, L.R., Shanks, T., Stevenson, P.R.F., Strong, A.W., Dawe, J.A. & Murray, J.D., 1987. *Mon. Not. R. astr. Soc.,* **224**, 1059.
Frenk, C.S., 1986. *Phil. Trans. R. Soc. Lond.* **330**, 517.
Geller, M.J., de Lapparent, V. & Kurtz, M.J., 1984. *Astrophys. J.,* **287**, L55.
Giovanelli, R., Haynes, M.P. & Chincarini, G.L., 1986 *Astrophys. J.,* **300**, 77.
Gott, J.R.III, Melott, A.L. & Dickinson, M., 1986. *Astrophys. J.,* **366**, 341.
Graham, J.A., 1982. *Publ. astr. Soc. Pacific,* **94**, 244.
Gregory, S.A. & Thompson, L.A., 1978. *Astrophys. J.,* **222**, 784.
Gregory, S.A. & Thompson, L.A. & Tifft, W.G., 1981. *Astrophys. J.,* **243**, 411.
Groth, E.J. & Peebles, P.J.E., 1977. *Astrophys. J.,* **217**, 385.
Groth, E.J. & Peebles, P.J.E., 1986a. *Astrophys. J.,* **310**, 409.
Groth, E.J. & Peebles, P.J.E., 1986b. *Astrophys. J.,* **310**, 507.
Groth, E.J. & Peebles, P.J.E., 1988a,b., preprint.
Groth, E.J. & Peebles, P.J.E., 1988b. *Astrophys. J.,* in press. Gunn, J.E. & Oke, J.B., 1975. *Astrophys. J.,* **195**, 255.
Hamilton, A.J.S., 1988. *Astrophys. J.,* **331**, L59.
Hatzimdimitriou, D., 1989. *Ph. D. Thesis,* University of Edinburgh.
Hawkins, M.R.S., 1981 *Mon. Not. R. astr. Soc.,* **194**, 1013.
Hauser, M.G., & Peebles, P.J.E., 1973. *Astrophys. J.,* **185**, 757.
Haynes, M.P. & Giovanelli, R., 1986. *Galaxy Distances and Deviations from Universal Expansion,* p117, eds. Madore, B.F. & Tully, R.B., D Reidel, Dordrecht, Holland.
Hewett, P.C., 1983. *Ph.D. Thesis,* University of Edinburgh.
Heydon-Dumbleton, N.H., Collins, C.A. & MacGillivray, H.T., 1989. *Mon. Not. R. astr. Soc.,* inpress
Heydon-Dumbleton, N.H., Collins, C.A. & MacGillivray, H.T., 1988. *Large-Scale Structure in the Universe — Observational and Analytical Methods,* p71, eds. Seitter,

W.C., Duerbeck, H.W. & Tacke, M., Lecture Notes in Physics, **310**, Springer-Verlag.

Holmberg, E., 1937. *Ann. Obs. Lund,* No. 6.

Humason, M.L., Mayall, M.O. & Sandage, A.R., 1956. *Astr. J.,* **61**, 97.

Hubble, E.P., 1934. *Astrophys. J.,* **79**, 8.

Huchra, J.P. & Geller, M.J., 1982 *Astrophys. J.,* **257**, 423.

Infante, L., Pritchet, C., & Quintana, H., 1986. *Astr. J.,* **91**, 217.

James, P.A., Joseph, R.D. & Collins, C.A., 1987. *Mon. Not. R. astr. Soc.,* **229**, 53.

Jarvis, J.F. & Tyson, J.A., 1981. *Astr. J.,* **86**, 476.

Kaiser, N., 1984. *Astrophys. J.,* **284**, L9.

Kaiser, N., 1986. *Mon. Not. R. astr. Soc.,* **219**, 785.

Kant, Immanuel, *Allgemeine Naturgeschichte und Theorie des Himmels,* in *Immanuel Kants Werke*, E. Cassirer (.ed.), vol. I, Berlin 1912; translated by W. hashe as *Kant's Cosmology,* revised and edited by Willy Ley (New York: Greenwood), 1968.

Kindl, E., 1987. *Astr. J.,* **93**, 1024.

Kirshner, R.P., Oemler, A. & Schechter, P.L., 1978. *Astr. J.,* **83**, 1549.

Kirshner, R.P., Oemler, A., Schechter, P.L., & Shectman, S.A., 1981. *Astrophys. J.,* **248**, L57.

Kirshner, R.P., Oemler, A., Schechter, P.L., & Shectman, S.A., 1987. *Astrophys. J.,* **314**, 493.

Klemola, A.R., 1969. *Astr. J.,* **74**, 804.

Koo, D.C., 1981. *Ph.D. thesis*, University of California at Berkeley.

Kron, R.G., 1980. *Astrophys. J. Suppl.,* **43**, 305.

Kurtz, M.J., Huchra, J.P., Beers, T.C., Geller, M.J., Gioia, I.M., Maccacaro, T., Schild, R.E. & Stauffer, J.R., 1985. *Astr. J.,* **90**, 1665.

de Lapparent, V., Kurtz, M.J. & Geller, M.J., 1986. *Astrophys. J.,* **304**, 585.

de Lapparent, V., Geller, M.J. & Huchra, J.P., 1986. *Astrophys. J.,* **302**, L1.

Lauberts, A. 1982. *The ESO/Uppsala survey of the ESO (B) Atlas,* European Southern Observatory.

Lugger, P.M., 1976. *Senior Thesis,* Harvard Univ.

Lumsden, S.L., Heavens, A.F. & Peacock, J.A., 1989. *Mon. Not. R. astr. Soc.,* in press.

Maddox, S.J., Efstathiou, G. & Loveday, J., 1988a. *Large Scale Structures of the Universe, IAU Symp. 130,* eds Audouze, J., Pelletan, M.& Szalay, A., Kluwer, Dordrecht, Holland.

Maddox, S.J., Loveday, J., Sutherland, W.J. & Efstathiou, G., 1988b. *Large-Scale Structure in the Universe — Observational and Analytical Methods,* p90, eds. Seitter, W.C., Duerbeck, H.W. & Tacke, M., Lecture Notes in Physics, **310**, Springer-Verlag,

MacGillivray, H.T. & Dodd, R.J., 1983 *Workshop on Astronomical Measuring Machines,* Occ. Rep. R. Obs. Edin., **10**, eds. Stobie R.S. & McInnes, B., Royal

For Horstmann H. see p 214

Observatory Edinburgh.
MacGillivray, H.T. & Dodd, R.J., 1984 *Astr. Astrophys.*, **145**, 269.
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.
MacGillivray, H.T. & Stobie, R.S., 1984. *Vistas Astr.*, **27**, 433.
Messier's Catalogue of Nebulae 1781. see *Messier's Nebulae and Starclusters*, Jones, K.G., 1968. Faber, London.
Metcalfe, N., Fong, R., Shanks, T. & Kilkenny, D. 1989. *Mon. Not. R. astr. Soc.*, **236**, 207.
More, J.G., 1988. *Ph. D. Thesis*, University of Edinburgh.
Murtagh, F., 1986. *International Workshop on data Analysis and Astronomy*, Plenum Press.
Nilson, P., 1973. *Uppsala General Catalogue of Galaxies.* Uppsala Obs. Ann. 6.
Oegerle, W.R., Hoessel, J.G. & Ernst, R.M., 1986. *Astr. J.*, **91**, 697.
Oort, J.H., 1983 *Ann. Rev. Astron. Astrophysics*, **21**, 373.
Parker, Q.A., Beard, S.M. & MacGillivray, H.T., 1988 *Astr. Astrophys.*, **173**, L5.
Parker, Q.A. & MacGillivray, H.T., 1987 *Astr. Astrophys.*, **199**, 364.
Peebles, P.J.E., 1980. *Large-scale Structure of the Universe*, Princeton University Press, Princeton.
Peebles, P.J.E., 1987. *Astrophys. J.*, **277**, L1.
Peterson, B.A., Ellis, R.S., Kibblewhite, E.J., Bridgeland, M., Hooley, A. & Horne, D., 1979. *Astrophys. J.*, **233**, L109.
Peterson, B.A, Ellis, R.S., Efstathiou, G., Bean, A.J., Shanks, T., Fong, R. & Zhen-Long, Z., 1986. *Mon. Not. R. astr. Soc.*, **221**, 223.
Phillips, S., Ellis, R.S. & Strong, A.W., 1981 *Mon. Not. R. astr. Soc.*, **197**, 151.
Postmann, M., Geller, M.J. & Huchra, J.P., 1986. *Astr. J.*, **91**, 1267.
Primack, J.R., 1984. *Lectures presented at the International School of Physics "Enrico Fermi"*, SLAC-PUB-3387, Stanford University.
Primack, J.R. & Blumenthal, G.R., 1984. *Clusters and Groups of Galaxies*, p435, eds. Mardirossian, F., Giuricin, G. & Mezzetti, M., Reidel, Dordrecht, Holland.
Roach, F.E., & Gordon, J.l., 1973. *Light of the Night Sky*, Geophysics & Astrophysics Monographs 4, D Reidel, Dordrecht, Holland.
Rose, J.A., 1976. *Astron. Astrophys. Suppl.*, **23**, 109.
Rudnicki, K., 1988. *Large-Scale Structure in the Universe — Observational and Analytical Methods*, p306, eds. Seitter, W.C., Duerbeck, H.W. & Tacke, M., Lecture Notes in Physics, **310**, Springer-Verlag.
Rudnicki, K., Dworak, T.Z., Flin, P., Baranowski, B. & Sendrakowski, A., 1972. *Acta Cosmologica*, **1**, 7.
Sandage, A. & Tammann, G., 1975. *Astrophys. J.*, **197**, 265.
Sandage, A., Kristian, J. & Westphal, J.A., 1976. *Astrophys. J.*, **205**, 688.
Sebok, W.L., 1986. *Astrophys. J. Suppl.*, **62**, 301.

Seitter, W.C., 1988. *Large-Scale Structure in the Universe — Observational and Analytical Methods,* p9, eds. Seitter, W.C., Duerbeck, H.W. & Tacke, M., Lecture Notes in Physics, **310,** Springer-Verlag.

Seldner, M., Siebers, B., Groth, E.J. & Peebles, P.J.E., 1977. *Astr. J.,* **82,** 249.

Sersic, J.L., 1974. *Astrophys. Space Science,* **28,** 365.

Shane, C.D. & Wirtanen, C.A., 1967. *Publs Lick Obs.* **Vol 22,** Part 1.

Shanks, T., Bean, A.J., Efstathiou, G., Ellis, R.S., Fong, R. & Peterson, B.A., 1983. *Astrophys. J.,* **274,** 529.

Shanks, T., Fong, R., Ellis, R.S. & MacGillivray, H.T., 1980. *Mon. Not. R. astr. Soc.,* **192,** 209.

Shanks, T., Stevenson, P.R.F., Fong, R. & MacGillivray, H.T., 1984. *Mon. Not. R. astr. Soc.,* **206,** 767.

Shapley, H. & Ames. A., 1932. *Harvard Obs. Ann.,* **88(4),** 43.

Shapley, H., 1933. *Proc. Nat. Acad. Sci. U.S.,* **19,** 591.

Shectman, S.A., 1985. *Astrophys. J. Suppl.,* **57,** 77.

Smithsonian Astrophysical Observatory Star Catalog, 1966, in 4 parts (reprinted 1971), Smithsonian Publication 4652, Boston.

Snow, T.P, 1970. *Astr. J.,* **75,** 237.

Standen, P.R. & Tritton, K.P., 1979. *The UK 1·2 m Schmidt Telescope & Southern Sky Survey III, - Plate Copying,* Occ. Rep. R. Obs. Edin., **5,** Royal Observatory Edinburgh.

Stevenson, P.R.F., 1985. *Ph. D. Thesis,* University of Durham.

Stobie, R.S., 1980. *Journal Brit. Interplanetary Soc.,* **33,** 323.

Stobie, R.S., Okamura S., Davenhall A.C. & MacGillivray, H.T., 1984. *Astronomical Photography,* Occ. Rep. R. Obs. Edin., **14,** 219, eds Sim, M. E. & Ishida, K., Royal Observatory Edinburgh.

Sutherland, W., 1988. *Mon . Not. R. astr. Soc.,* **234,** 159.

Tago, E., Einasto, J. & Saar, E., 1984. *Mon. Not. R. astr. Soc.,* **206,** 559.

Tarenghi, M., Tifft, W.G., Chincarini, G., Rood, H.J. & Thompson, L.A., 1979. *Astrophys. J.,* **234,** 793.

Tarenghi, M., Chincarini, G., Rood, H.J. & Thompson, L.A., 1980. *Astrophys. J.,* **235,** 724.

Thanisch, P., McNally, B.V. & Robin, A., 1984. *Image & Vision Computing,* **2,** 4.

Tully, R.B., 1987. *Catalog of Nearby Galaxies,* Cambridge University Press, Cambridge.

Tully, R.B. & Fisher, J.R., 1987. *Atlas of Nearby Galaxies,* Cambridge University Press, Cambridge.

Turner, E.L. & Gott, J.R., 1976, *Astrophys. J. Suppl.,* **32,** 409.

Tyson, J.A., & Jarvis, J.F., 1979. *Astrophys. J.,* **230,** L153.

United Kingdom Schmidt Teloscope Unit (UKSTU) Handbook, 1983. Royal Observatory Edinburgh.

Valdes, F., 1982. *Instrumentation in Astronomy IV,* SPIE **331,** 465.

de Vaucouleurs, G. & de Vaucouleurs, A., 1964. *Reference Catalog of Bright galaxies.*

Univ. Texas Press, Austin.

de Vaucouleurs, G., 1976. in *Galaxies and the Universe,* eds. Sandage, A., Sandage, M., & Kristian, J., p557, Univ. Chicago Press, Chicago.

Wallace, P.T., & Tritton, K.P. 1979 *Mon . Not. R. astr. Soc.,* **189,** 115.

Wesselink, T., 1984. *Astronomical Photography,* Occ. Rep. R. Obs. Edin., **14,** 261, eds Sim, M. E. & Ishida, K., Royal Observatory Edinburgh.

West, R.M. & Dumoulin, B., 1974. *Photographic Reproduction of Large Astronomical Plates,* Report of the ESO Sky Atlas Laboratory, European Southern Observatory.

White, S.D.M., 1986. *Inner Space/Outer Space,* p228, eds. Kolb, E.W., Turner, M.S., Lindly, D, Olive, K. & Seckel, D., University of Chicago Press, Chicago.

White, S.D.M., Frenk, C.S., Davis, M. & Efstathiou, G., 1987. *Astrophys. J.,* **313,** 505.

Wolstencroft, R.D., Savage, A., Clowes, R.G., MacGillivray, H.T., Leggett, S.K. & Kalafi, M., 1986. *Mon. Not. R. astr. Soc.,* **223,** 279.

Yamagata, T., 1987. preprint, Tokyo Astron. Obs., University of Tokyo.

Yang, J., Turner, M. S., Steighan, G., Schramm, D. N. & Olive, K. A., 1984. *Astrophys. J.,* **281,** 493.

Zwicky, F., 1933. *Helv. Phys. Acta.,* **50,** 218.

Zwicky, F., 1938. *Publ. astr. Soc. Pacific,* **50,** 218.

Zwicky, F., Herzog, E., Wild, P., Karpowicz, M., Kowal, C.T., 1961–1968. *Catalogue of Galaxies & Clusters of Galaxies,* 6 volumes, Calif. Inst. Technol., Pasadena.

Horstmann H., 1988 *Large-Scale Structure in the Universe – Observational and Analytical Methods,* p111, eds Seitter, W.C., Duerbeck, H.W. & Tacke, M., Lecture notes in Physics, 310, Springer Verlag.