Large-scale structure in a new deep IRAS galaxy redshift survey

Article (Published Version)


This version is available from Sussex Research Online: http://sro.sussex.ac.uk/25503/

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:
Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.
Large Scale Structure in a new deep IRAS galaxy redshift survey

S.J. Oliver¹, M. Rowan-Robinson¹, T.J. Broadhurst², R.G. McMahon³, W. Saunders⁴, A. Taylor⁵ A. Lawrence⁵, C.J. Lonsdale⁶, P. Hacking⁶, and T. Conrow⁶

¹ Imperial College of Science Technology and Medicine, Blackett Laboratory, Prince Consort Road. London SW7 2BZ.
² Johns Hopkins University, Johns Hopkins Road, Laurel, MD 20723-6099, USA.
³ Institute of Astronomy, Madingley Road, Cambridge CB3 0HA.
⁴ Department of Astrophysics, Keeble Road, Oxford OX1 3RH.
⁵ Institute for Astronomy, Blackford Hill, Edinburgh EH9 3HJ.
⁶ IPAC, California Institute of Technology, 770 South Wilson Avenue, Mail Code 100-22, Pasadena, CA 91125, USA.

Accepted 1995 November 16. Received 1995 November 1; in original form July 12

ABSTRACT

We present here the first results from two recently completed, fully sampled redshift surveys comprising 3703 IRAS Faint Source Survey (FSS) galaxies. An unbiased counts-in-cells analysis finds a clustering strength in broad agreement with other recent redshift surveys and at odds with the standard cold dark matter model. We combine our data with those from the QDOT and 1.2 Jy surveys, producing a single estimate of the IRAS galaxy clustering strength. We compare the data with the power spectrum derived from a mixed dark matter universe. Direct comparison of the clustering strength seen in the IRAS samples with that seen in the APM-Stromlo survey suggests \( b_0/b_1 = 1.20 \pm 0.05 \) assuming a linear, scale independent biasing. We also perform a cell by cell comparison of our FSS-z sample with galaxies from the first CfA slice, testing the viability of a linear-biasing scheme linking the two. We are able to rule out models in which the FSS-z galaxies identically trace the CfA galaxies on scales 5-20\( \,h^{-1}\)Mpc. On scales of 5 and 10\( \,h^{-1}\)Mpc no linear-biasing model can be found relating the two samples. We argue that this result is expected since the CfA sample includes more elliptical galaxies which have different clustering properties from spirals. On scales of 20\( \,h^{-1}\)Mpc no linear-biasing model with \( b_0/b_1 < 1.70 \) is acceptable. When comparing the FSS-z galaxies to the CfA spirals, however, the two populations trace the same structures within our uncertainties.

Key words: galaxies: clusters: general – galaxies: elliptical & lenticular, cD – galaxies: spiral – large-scale structure of Universe – infrared: galaxies

1 INTRODUCTION

Following the success of IRAS Point Source Catalog (PSC) galaxy redshift surveys in probing the large scale structure of the local Universe, which presented serious problems for the standard cold dark matter (CDM) theory (Efstathiou et al. 1990, Saunders et al. 1991, Kaiser et al. 1991, Moore et al. 1992, Strauss et al. 1992, Fisher et al. 1993), two new fainter redshift surveys have been undertaken to confirm these results at significantly greater depths.

The first of these surveys (FSS-z) was based on samples drawn from the IRAS Faint Source Data Base (FSDB, see Moshir et al. 1992). On the basis of deep IRAS coverage and freedom from cirrus contamination, three regions in the Northern Galactic Hemisphere were selected (these areas are defined in Table 1). Within these areas we believe the FSDB to be \( \sim 99 \) per cent complete for \( S_\nu \geq 0.2 \) Jy (Lonsdale et al. in preparation). We thus constructed a cat-

* These surveys have been referred to previously as ‘QCCOD’
Table 1. Definition of FSS-z areas. Areas N, A and E constitute FSS-z I while P and X constitute FSS-z II.

<table>
<thead>
<tr>
<th>Definition</th>
<th>Area/ sq deg</th>
</tr>
</thead>
<tbody>
<tr>
<td>N 150 deg ≤ l ≤ 210 deg, 50 deg ≤ b ≤ 70 deg</td>
<td>594</td>
</tr>
<tr>
<td>A 30 deg ≤ l ≤ 90 deg, 60 deg ≤ b ≤ 68 deg</td>
<td>178</td>
</tr>
<tr>
<td>E 70 deg ≤ l ≤ 90 deg, 50 deg ≤ b ≤ 55.5 deg</td>
<td>67</td>
</tr>
<tr>
<td>P 26.5 deg ≤ b ≤ 44.5 deg, b ≥ 70 deg if α ≤ 12h or b ≥ 68 deg if α &gt; 12h</td>
<td>476</td>
</tr>
<tr>
<td>X 32.5 deg ≤ b ≤ 38.5 deg, b &lt; 50 deg &amp; α ≥ 8h</td>
<td>94</td>
</tr>
</tbody>
</table>

dalogue of all FSDB sources within these areas having good to moderate 60µm fluxes greater than 0.2 Jy. (In area E an additional sample was selected with 0.15 ≤ S60 < 0.2 but this is not considered further in this paper.) The majority of such sources are galaxies [e.g. Rowan-Robinson et al. (1986), Lawrence et al. (1988)]. No formal colour-cuts were employed to exclude either stars or cirrus sources. The former were excluded upon inspection of the POSS plates or APM cartoons together with a case by case examination of the colours. In the combined FSS samples only five ‘stars’ lay close to the galaxy colour locus and they had the same colours as the 60µm excess stars. In addition 13 galaxies were excluded because a bright star lay in the field making identification and acquisition impossible, these are assumed to be random line of sight coincidences with little effect on our analysis. Cirrus sources were rare because of our careful selection of areas (four sources have been excluded upon examination of maps made from the raw IRAS data). This survey was particularly designed to clarify the evolutionary behaviour of the IRAS population; see contrasting conclusions of Saunders et al. (1991) and Fisher et al. (1992).

The second survey (FSS-z II) was constructed from the Faint Source Catalogue Version 2 (Moshir et al. 1992). The area selected traversed the North Galactic Pole connecting the FSS-z I areas. In other respects the FSS-z II samples were selected in the same manner as FSS-z I. Designed principally for large scale structure studies, the flux completeness in areas P and (to a lesser extent) X is not as good as other areas and is estimated to be 90 per cent at 0.25 Jy. The combined surveys cover an area of 1310 deg2 (0.4 sr) and contain 3728 sources, more than 3600 of which are galaxies.

Redshifts were obtained from the literature for 872 sources; a number of redshifts were kindly provided in advance of publication by: John Huchra, Ray Walstencroft, Quentin Parker & Roger Clowes and Marc Davis & Michael Strauss. Major observation programmes were instigated to obtain the redshifts for the remaining sources using the FOS and FOS2 instruments on the INT and WHT facilities. Using automatic, optimal extraction techniques and line fitting procedures we obtained an average redshift accuracy of ∼ 190km s−1 (Oliver 1993, and Oliver et al. in preparation), c.f. an accuracy of ∼ 250 km s−1 for the QDOT survey (Lawrence et al., in preparation).

Of all 1931 FSS-z I galaxies we currently have 1769 redshifts giving an overall redshift completeness of 91.6 per cent; while for the FSS-z II project we have redshifts for 80.4 per cent of the galaxies. A mask has been constructed which, as well as defining the survey boundaries, excludes a number of sectors that have not been exhaustively followed up and those sources lying close to the IRAS coverage gap. Upon application of this mask the completeness statistics improve to 92.5 per cent and 86.8 per cent respectively. Above 0.25 Jy (where FSS-z II suffers less from flux incompleteness) we have 95 per cent of the FSS-z I redshifts and 90 per cent for FSS-z II. Detailed numbers are listed in Table 2. Of the failures the most interesting will be those with either very faint or no optical counterparts. VLA maps have been obtained for all sources in the FSS-z I that are either blank within the IRAS error ellipse (to the limit of the sky survey plates) or have very faint identifications. R-band CCD images have also been obtained for many of these sources. These have yielded a number of further identifications and redshifts have been obtained for several of these. Work is continuing to obtain the redshifts for the remainder.

This paper quantifies the large scale clustering seen in both surveys and directly compares these samples with optical surveys. These surveys are also being used for many other studies. It was during the first of these projects that the unique object F10214+4724 was discovered (Rowan-Robinson et al. 1991). The completeness and reliability of the FSS-z I survey will be detailed in future papers (Lonsdale et al. & McMahon et al., in preparation). Simple tests reveal significant evolution within this sample (Oliver et al. 1995). This evolution will be elaborated in greater detail by Broadhurst et al. (in preparation). The appearance of the Boötes void which overlaps with areas A & E will be discussed by Oliver et al. (in preparation). The data for both surveys will be presented by Oliver et al. (in preparation).

2 COUNTS-IN-CELLS

To assess the large scale structure seen in the FSS-z surveys quantitatively we have chosen to apply the counts-in-cells method described by Efstathiou et al. (1990) ; hereafter E90) and also used by Loveday et al. (1992). The two-point autocorrelation function is a useful descriptive statistic; indeed for Gaussian fields it provides a complete statistical profile of the field. The counts-in-cells method measures a volume integral of this function, the variance, and avoids the problem of assigning weight to galaxies in direct measurements of the two-point correlation function. This analysis is also relatively straightforward to perform in the presence of a mask covering a significant fraction of the sky, a considerable advantage over direct determinations of the power-spectrum in Fourier space.
2.1 Basic Method

We consider the moments of galaxy counts (Ni) over a lattice of M cells. Following E90 we define the statistics

\[
\bar{N} = \frac{1}{M} \sum_i N_i
\]

\[
S = \left( \frac{1}{M-1} \sum_i (N_i - \bar{N})^2 \right) - \bar{N}.
\]

The statistic \( \bar{N} \) has an expectation value \( nV \), where \( n \) is the mean galaxy density and \( V \) is the cell volume. If the cells are independent then the expectation value of \( S \), \( \langle S \rangle = n^2 \sigma^2 V^2 \), where \( \sigma^2 \) is a volume integral over the two-point autocorrelation function:

\[
\sigma^2(l) = \frac{1}{V^2} \int \xi(r_{12})dV_1dV_2.
\]

\( \sigma^2 \) is also closely related to the power spectrum as described in Section 2.4.

If the cells are independent and the number of cells is large then we can derive an expression for the variance of \( S \) \( \text{[var}(S)\text{]} \) in terms of the first four moments of the density field (equation 4 in E90). However, if we assume Gaussian fluctuations then the higher order reduced correlation functions vanish giving

\[
\text{var}(S) = \frac{2n^2V^2(1 + \sigma^2) + 4n^3V^3\sigma^2 + 2n^4V^4\sigma^4}{M}.
\]

This allows us to estimate var(\( S \)) for a given sample by using the empirical estimates \( nV \approx \bar{N} \) and \( \sigma^2 \approx S/N^2 \) in equation (3). For non Gaussian fluctuations this underestimates the errors and ideally we should use N-body simulations of a particular model to determine higher order moments and thus estimate the errors.

A survey volume is divided up into concentric shells and these are then divided into \( M \) cells. The mean number density of the cells is constant within a given shell allowing density estimations of the variance from each shell. These estimates can be combined using the maximum likelihood technique. The statistics \( N_i \) and \( S_i \) are calculated for each shell and we construct the likelihood function from all shells:

\[
\mathcal{L}(\sigma^2) = \prod_i \frac{1}{[2\pi\text{var}(S_i)]^{1/2}} \exp \left[ -\frac{(S_i - N_i^2\sigma^2)^2}{2\text{var}(S_i)} \right] .
\]

We can now estimate \( \sigma^2 \) by replacing \( n^2V^2 \) with \( \bar{N}_i^2 \) and maximizing the likelihood function numerically. We can justify not using a joint likelihood function for the two estimators, \( \bar{N}_i^2 \) and \( S_i \), since the errors in \( \bar{N}_i^2 \) are much smaller. The 68 per cent confidence limits to this maximum likelihood solution are the values of \( \sigma^2 \) where the log likelihood has dropped by a factor of 0.5 from its maximum value. This maximum likelihood technique automatically weights the shells, shells with few cells or sparse density receiving less weight.

To determine how well the data are modelled by our maximum likelihood solution we also calculate the \( \chi^2 \) of each fit,

\[
\chi^2 = \sum \frac{(S_i - N_i^2\sigma^2)^2}{\text{var}(S_i)}
\]

where the number of degrees of freedom (\( \nu \)) is one less than the number of shells.

2.2 Detailed Method

In this Section we discuss finer points of the counts in cells method. Specifically we propose two refinements to the method of E90.

In practice each cell does not have exactly the same volume owing to the non-uniform mask. We correct the estimators \( \bar{N} \) and \( S \) for incomplete volumes as prescribed by equation 9 in E90 (N.B. equation 9 in E90 has a minor typographical error) i.e.

\[
\bar{N} = \frac{\sum N}{\sum V}
\]

\[
S = \frac{\sum (N - \bar{N}V)^2 - \left(1 - \frac{\sum V^2}{\sum V} \right) \sum N}{\bar{N}^2 \left[ \sum V^2 - 2 \sum V^3 \sum V + \left( \sum V^2 \right)^2 \sum V^{-1} \right]}
\]

This correction only takes into account the fact that the expected number of galaxies in a restricted cell will be reduced. In addition, a masked cell will, in general, sample smaller scales. Since it is observed that clustering is stronger on smaller scales this means that the expected \( \sigma^2 \) increases. This bias is difficult to correct for since it depends in detail on the topology of the mask and the slope of the correlation function. We have elected simply to exclude those cells with more than half the volume masked.

At large distances the number of galaxies per cell has dropped significantly so that we no longer get a meaningful addition to our estimate of the variance. E90 showed that if \( \bar{N} \sigma^2 \gg 1 \) the variance is inversely proportional to the volume. Similarly it can be shown that if \( \bar{N} \sigma^2 \ll 1 \) and \( \bar{N} \ll 1 \) the variance on \( \sigma^2 \) actually increases with additional cells. Thus \( \bar{N} \sigma^2 = 1 \) represents some measure of useful depth of a survey (not surprisingly this is where the Poisson variance equals the clustering variance). Although the maximum likelihood analysis includes a weighting scheme there is little point in taking shells much beyond this limit. We estimate this depth using \( \sigma^2 \) from E90 and the luminosity function of Saunders et al. (1990) and only consider shells to roughly twice this depth.

It has been claimed that the counts-in-cells method as it stands is independent of the galaxy selection function [E90, Efstathiou (1993)]. This is not strictly true. Since the selection function is not constant across each shell the galaxies will tend to be concentrated towards the nearest side. This reduces the effective volume of the cells and (as with masking) will bias the variance upwards. With prior knowledge of the clustering strength and selection function one could estimate and remove this bias. Instead of this we eliminate this bias by volume limiting each shell (i.e. removing all galaxies that would not have been visible if placed at the far side of their shell). The removal of galaxies should not seriously affect our final errors since they are not sensitive to the mean.
density. As expected this correction reduces the variance in almost every case. For QDOT in particular the reduction was considerable, up to 1.7 times the quoted errors. Con- fortingly the $\chi^2$ usually decreased, showing that the fit was better.

We found that the variances obtained by the counts-in- cells method were quite sensitive to the specific grid pattern used. To reduce this sensitivity we use 16 different grids for each survey. First we divide the volumes into equal thick- ness concentric shells, then divide these shells with latitude and longitude cuts such that the resulting cells have equal volumes (in the absence of any mask). These latitude and longitude cuts are then offset by $1/2$ and $1/4$ spacing in both dimensions to produce eight different grids. This entire process is then repeated with the shells offset by $1/2$ a shell thickness. Each grid gives us $1\sigma$ upper and lower limits to the variance. We take the linear average of these over the 16 grids to produce a final estimate of the upper and lower limits on the variance. In the process we also produce an estimate of the scatter between the individual grids; in many cases it is comparable to our quoted errors. By averaging over 16 different grids we have thus undoubtedly reduced the errors. It would be possible to account for the reduction in systematic errors due to the re-gridding is not likelihood errors are underestimated. So a correction for the variances. This is a further indication that the maximum likelihood error, and this procedure would lead to negative cases our grid error is actually larger than our maximum limits on the variance. In the process we also produce an es- timate of the scatter between the individual grids; in many

In equation 5 we estimate $\chi^2$ with the unbiased estimator $\bar{N}^2$. We have investigated replacing $\bar{N}^2$ with the unbiased estimator $[(\sum_i N_i^2) - \sum_i N_i^2]/M(M - 1)$ but did not find any significant differences.

A further bias arises because the cells are not indepen- dent. As discussed by E90 and Loveday et al. (1992) this affects the estimate of var($S$). However, correlations between cells also bias the estimator $S$ downwards. Ideally each shell should include many uncorrelated volumes. A small area sur- vey will have fewer uncorrelated cells per shell than a large area survey of the same volume. We thus might expect our FSS-z to show less variance than the all sky surveys and since we do not see such an effect we assume that this bias is small.

2.3 Results

We apply the above method to our new surveys and also to existing IRAS redshift surveys: the latest version of the QDOT survey (Lawrence et al. in preparation); the 1.2 Jy survey (Fisher et al. 1993); a faint QDOT sample (0.6-1.2 Jy); a combined PSC (QDOT and 1.2 Jy) sample. We

also combine our FSS-z sample with the combined PSC sam- ple to produce a single IRAS sample.

For our FSS-z surveys we flux limit the whole sample at 0.25 Jy to allow for the varying completeness limits in different areas. The mask was described in Section 1.

For the QDOT and faint QDOT samples we use the standard QDOT mask. The faint QDOT sample has $0.6 < S_{50} < 1.2$ Jy so the volume limiting is rather harsh.

In the analysis of the 1.2 Jy sample we apply a cut of $|b| > 5$ deg together with the QDOT mask. This is slightly more severe than the 1.2 Jy team used themselves.

Since the faint QDOT sample has no galaxies in common with the 1.2 Jy survey the two can be combined to construct a larger PSC sample. For this PSC sample we use the same mask as we used for the 1.2 Jy catalogue. Many few galaxies are excluded when volume limiting the com- bined PSC sample than are excluded when volume limiting the samples separately.

Finally, we take the individual variance estimates from each shell of our combined PSC sample, together with the individual variances from each shell of our FSS-z sample, to construct a single likelihood function (equation 5), as before. Maximizing this likelihood function gives us a single variance estimate from all samples. Since the volume of overlap be- tween the FSS-z and the PSC is small, any interdependence will not be a significant problem.

The results for all the various samples described above are given in Table 1.

In many cases the $\chi^2$ and $\nu$ values indicate that the goodness of fit for the individual grids is poor. As discussed in E90 the assumption of Gaussian fluctuations causes the errors to be underestimated which may explain this. Since we reduce the systematic errors by averaging over 16 grids, without adjusting our error estimates accordingly we have to some extent compensated for this.

The QDOT variances appear slightly smaller than those quoted by E90 and Efstathiou (1993). The main reason for this is that by volume limiting each shell we have removed the bias due to gradients in the selection function. The grids also play a part in this apparent discrepancy; without vol-
ume limiting, one grid did produce a variance as high as the E90 variance at 40 h\(^{-1}\) Mpc. Notice that the scatter between grids is always comparable to our quoted error (i.e. roughly due to differences in the masks.

The 1.2 Jy results, on the other hand, appear slightly larger than the results of Fisher \textit{et al.}(1994). This may be due to differences in the masks.

In comparing estimates of clustering from different surveys we can assume that the ratio of variances is independent of scale and use a weighted average of the log of this ratio (with associated \(\chi^2\) and \(\nu\) to estimate \(b_a/b_b = \sqrt{\sigma_a^2/\sigma_b^2}\). We find \(b_{\text{QDOT}}/b_{\text{1.2 Jy}} = 0.98 \pm 0.06\), \(b_{\text{FSS}}/b_{\text{1.2 Jy}} = 1.03 \pm 0.06\) and \(b_{\text{QDOT}}/b_{\text{QDOT}} = 1.06 \pm 0.07\). All have acceptable \(\chi^2\) values (indicating that the scale independence is justifiable) and on each scale the variances between surveys are consistent. All are consistent with \(b_a/b_b = 1\) and in addition there is no significant disagreement between the variances measured by the different surveys on any scale. This can be seen in Figure 1. This in turn justifies our combining the samples. (In these comparisons we have assumed that the errors from each survey are independent. The many galaxies in common between QDOT and 1.2 Jy suggest that they are not in fact independent so we have overestimated the errors in \(b_{\text{QDOT}}/b_{\text{1.2 Jy}}\). However, the true errors would need to be much smaller for the the surveys to disagree significantly.)

The agreement between QDOT and 1.2 Jy is slightly at odds with the conclusions of Tadros & Efstathiou’s (1993) power spectrum analysis of the two surveys. They find a discrepancy between the two surveys which disappears after the exclusion of the Hercules region. Their analysis uses a selection function derived from the 60\(\mu\)m luminosity function (Saunders \textit{et al.}1990). It is possible that their analysis is more sensitive to the Hercules anomaly as a result of errors in this selection function. At the distance of Hercules, errors in the evolutionary term of the Saunders \textit{et al.}1990 luminosity function contribute a systematic 7.5 per cent uncertainty to their selection function. Our analysis is independent of the selection function. Differences between the mask used by Tadros & Efstathiou and the one we use could also have affected this comparison, by excluding significant features in one or other analysis.

A more detailed comparison of the 1.2 Jy survey and QDOT survey shows that they are consistent with a single underlying density field (Efstathiou 1995).

We can apply the same test to compare the variances of these surveys with the standard CDM variances quoted by E90. For none of the surveys does the weighted ratio of galaxy to CDM variances provide an acceptable fit at around 99 per cent significance. Using the full IRAS sample we find a weighted average \(b_{\text{IRAS}}/b_{\text{CDM}} = 0.95 \pm 0.02\) (this average is dominated by the smallest scales). The fit is bad, \(\chi^2 = 17.0, \nu = 4\), allowing us to reject the CDM model at the \(> 99\) per cent level. This is clearly shown in Figure 2 where the slope of the variances with scale from the combined IRAS data is clearly at odds with the CDM slope.

### 2.4 Comparison With Model Power Spectra

Theoretical clustering predictions in the linear regime are frequently expressed in terms of the power spectrum \(P(k)\) or in dimensionless form \(\Delta^2 \sim k^3 P(k)\). The quantity \(\Delta^2\) is defined as the variance per ln \(k\) i.e. \(\Delta^2 \equiv \sigma^2 / d\ln k\). For a power law spectrum \((\Delta^2 \sim k^{n+3})\) the fluctuations within Gaussian spheres (radius \(R_G\)) is given by

\[
\sigma^2 = \Delta^2(k) \quad \text{and} \quad k = \left[\frac{1}{2}\left(\frac{n+1}{n+3}\right)\right]^{1/(n+3)} \frac{1}{R_G}. \quad (9,10)
\]
Figure 2 shows the predicted Fourier amplitudes from the FSS-z sample, QDOT, the 1.2 Jy sample and the PSC-FSS sample, calculated from equations (10) and (11) and using the results from Table 3. We also show the expected results from the APM power spectrum inversion of Baugh & Efstathiou (1994). We have assumed that the linear IRAS bias is $b_l = 0.7$ and the optical APM galaxies have a bias of $b_o = 0.9$, consistent with the results of Section 3, below.

These results are compared with the Mixed Dark Matter (MDM) model, which has previously been shown to be in good agreement with large scale clustering data (Taylor & Rowan-Robinson 1992, Schafer & Shafi 1992). The model shown has $\Omega_{HDM} = 0.15$ and $h = 0.5$, taken from the results of van Dalen & Schafer (1992), and is normalized to the COBE spectrum assuming $n = 1$, yielding an effective quadrupole of $19.9 \pm 1.6\mu$K derived by Gorski et al. (1994). We have also plotted the scale invariant $n = 1$ Harrison-Zeldovich spectrum.

Figure 3 again shows that there is now broad agreement between QDOT, 1.2 Jy and the PSC-FSS surveys. The MDM model is a good fit to the APM data, although it appears to have a slightly steeper slope than the IRAS data. This fit could not be improved by increasing the proportion of hot dark matter (HDM) because, with the same COBE normalization, we would then predict too much power on all these scales. Improvements could be made, either by adjusting the primordial spectral index, or by introducing a scale dependent bias. There are a number of reasons why the small discrepancies between the MDM model and the IRAS data do not justify such fine tuning. First, there are a number of debatable assumptions that have been made in order to predict the non-linear, redshift-space power spectrum from a linear theory, real-space power spectrum. Secondly, the assumption of Gaussian fluctuations in the data analysis causes us to underestimate our errors as described above. Thus we would conclude that our data are not incompatible with the MDM model.

3 COMPARISON WITH OPTICAL

An interesting question in the study of large scale structure is how the clustering properties of different galaxy populations are related. Since the various structure formation theories such as CDM predict the mass distribution, whereas redshift surveys probe only the galaxy distribution, we need to know how these are related. If different classes of galaxies do not trace the same structures then clearly at best only some of these classes can be faithfully tracing the mass. With an increased understanding of how different classes are related we can hope to gain some understanding of how the galaxies relate to the mass. This question is particularly pertinent to the interpretation of flux limited redshift surveys where the mix of populations changes with both the flux limit and the redshift. Any survey that contains two (or more) populations with different clustering properties is inevitably going to produce ambiguous results dependent on selection functions of the classes involved and the scales being considered.

A very simple possible relationship between galaxy populations is that they linearly trace some underlying density
Figure 3. Square root of the Power spectrum, $\Delta(k)$, for the FSS-z (stars), QDOT (circles), 1.2 Jy (triangles) PSC-FSS (squares) and APM (open circles). The data points have been offset along the $k$ axis for clarity. The curved solid line is the Mixed Dark Matter model with 15 per cent hot dark matter and $h = 0.5$, normalized to COBE, and corrected for nonlinear effects. The broken line is the linear spectrum. The straight solid line is the extrapolated scale invariant Harrison-Zeldovich spectrum. The data has been corrected for linear redshift space distortion in the small angle regime, and a linear bias of $b_1 = 0.7$ for the IRAS galaxies and $b_0 = 0.9$ for the optical APM galaxies.

The optical parent catalogue is the only published strip in the second CfA survey (de Lapparent et al.1986); i.e. 26.5 deg $\leq \delta < 32.5$ deg, $8^m < \alpha < 16^m$. The IRAS parent catalogue comprises the FSS-z samples discussed above, including all galaxies down to 0.2 Jy.

Since our aim is to compare the structures traced by different galaxies within the same volumes we must first restrict the two parent surveys to have the same areal profile. This means excluding CfA galaxies which lie within our FSS-z mask and also any FSS-z galaxies outside the CfA boundaries. These cuts leave us with 976 FSS-z galaxies and 854 CfA galaxies.

Of course many galaxies that appear in the CfA catalogue will also be found in the FSS-z. We identify all such pairs using a 3′ search radius. If more than one partner is found for a given source the match is made on the basis of brightness, nearness etc. (the details of the duplicate removals are relatively unimportant since most duplicates are neighbouring galaxies). One of these pairs was found to have a much larger redshift in one catalogue and, since the angular separation was also large, this match was rejected as a chance projection. This matching left us with three catalogues: 554 galaxies seen only in the CfA 2 survey; 300 galaxies seen in both surveys; 676 galaxies seen only in the FSS-z surveys. These three catalogues are statistically inde-
dependent, in the sense that no galaxy appears in more than one sample.

We now require a more physically meaningful classification of ‘IRAS’ and ‘optical’ galaxies. The mix of galaxy types in our three samples depends to a large extent on the survey selection functions, for example a nearby elliptical may be seen in both surveys but a similar galaxy would be unlikely to be detected in the FSS-z sample at a greater distance. To obviate this problem we have introduced a 60μm, B magnitude colour cut which neatly divides the common sample into two. The colour chosen is simply the ratio of the respective survey limits i.e.

\[
\frac{S_\text{B}}{S_\text{60}} = 10^{0.4(15.5-m_B)}.
\]

(Naturally all objects seen only in the FSS-z survey will have a colour that is more infrared than this limit and vice versa. The only dependence on redshift that is now present in our classification enters through the difference between the slopes of the spectral energy distributions around the 60μm and B-bands. Applying reasonable K-corrections at z = 0.3 this flux criterion corresponds to a cut at L_B > 2L_60 and although this cut is essentially arbitrary, it provides a much more appealing definition of an IRAS galaxy or optical galaxy than its presence or absence in a flux limited catalogue.

After applying this cut we arrive at two samples:

- 866(735) IRAS galaxies,
- 664(639) optical galaxies,

where the numbers in parentheses indicate those sources that have redshifts.

We already know that elliptical galaxies do not trace the same density field as other galaxies (Dressler 1980) and we also know that ellipticals are under-represented in IRAS sample (de Jong et al. 1984). So a priori we would expect some difference between the clustering of these two surveys simply from this morphological effect. This will be particularly important in this comparison since the CfA strip is strongly dominated by the Coma cluster. We therefore create a further sample in which we restrict the optical galaxies to exclude all elliptical galaxies. To identify the ellipticals we use the morphological classifications in the CfA catalogue and exclude all types earlier than Sa. This gives us an optical-spiral sample with 269(267) galaxies.

### 3.2 Qualitative comparison

The Figure 4 shows the cone diagrams for the three subsamples described in Section 3.1. In the optical sample the most striking structure is the Coma cluster and ‘Great Wall’. In the IRAS sample the Coma cluster is considerably less pronounced. The difference in appearance of the Coma cluster between the two samples is easily explained; the optical sample contains more ellipticals which are preferentially found in rich clusters. Clearly the IRAS sample probes much deeper rich clusters. Notice that the peak corresponding to the ‘Great Wall’ is a single broad peak in the optical but has two peaks in the IRAS data. This arise from the difference in selection functions.

The cone diagrams above are dominated by the selection functions of the various samples. Dividing the volume into cells within concentric shells we can estimate \( \rho_\text{O}/\rho_\text{I} = n_\text{O}/\sum n_i \) for each class within each cell. In Figures 5 and 6 we have plotted the densities of each cell as seen in the optical sample against the densities as seen in the IRAS sample. Since there is no overlap between cells all the data points are independent. We have not plotted error bars which would have confused the plot. It should be noted that the under-densities will usually have fewer galaxies and therefore larger errors. The general trend in these plots seems to be a steeper slope than would be expected for identical clustering, or even a linear-bias model with \( b_O/b_I = 1.3 \). On larger scales this appears to be true both above and below the mean density indicating that, in the optical, not only are the clusters more dense but also the voids are less dense. On scales of \( 5h^{-1}\) Mpc the IRAS density appears to be higher than the optical density for densities larger than the mean density. This might be explained by fingers of god which would tend to smear out the optical clusters.

### 3.3 Contingency Table

We now turn to a quantitative comparison between the two samples. This method was first discussed by Oliver (1993). Recently a similar but statistically more rigorous method was brought to our attention (Efstathiou 1993).

#### 3.3.1 Method

The most obvious scheme to test is that the IRAS and optical galaxies cluster in exactly the same way (i.e. \( b_O/b_I = 1 \)). Our null hypothesis is thus that the over-density in any given cell is independent of galaxy class. We divide our survey volume into shells and cells as we did for the counts in cells analysis. Under our null hypothesis we are at liberty to combine the samples to give a single estimate of the ‘galaxy’ over-density in any cell. From this estimated over-density and the observed over-densities in each individual class we construct a statistic, \( X^2 \), (Kendall & Stuart 1973) which we compute across the whole shell:

\[
X^2 = \sum_{ij} \frac{(n_{ij} - n_\text{av}n_{ij}/n)^2}{n_\text{av}n_{ij}/n}.
\]
Here \( n = \sum_{ij} n_{ij} \) represents the total number of galaxies from both surveys in a given shell. The subscript \( i \) refers to the individual cells in the shell and \( j \) to the galaxy class. We have employed the notation \( n_\ast = \sum_j n_{ij} \) to represent the marginal distributions. Here we have implicitly assumed that the observed galaxies give us an estimate of the underlying galaxy-density field, and that the measured differ-

**Figure 4.** Cone Diagram of FSS-z/CfA2 Strip 1 intersection. Top – IRAS Galaxies. Middle – optical galaxies. Bottom – optical-spiral galaxies.

**Figure 5.** \( N(z) \) diagrams. The top diagram shows the \( N(z) = \rho(z) dV \) for the IRAS sample while the middle diagram shows the same for the optical Sample, the bottom diagram is the ratio between the two (together with a power-law fit) this gives us \( \bar{\rho}_I / \bar{\rho}_O \) as a function of \( z \), independent of clustering (if there is no relative biasing).
Figure 6. $\rho_O/\bar{\rho}_O$ v $\rho_I/\bar{\rho}_I$ diagram of FSS-z/CIA2 Strip 1 intersection. Densities are calculated in nearly cubical cells of sizes 5 (top) and 10 $h^{-1}$ Mpc (bottom). The straight line would indicate identical clustering, the curved line would be followed in a linear-biasing model with $b_O/b_I = 1.3$. All points are independent so the scatter is real, to avoid confusion we have not plotted error bars.

Figure 7. As for Figure 6 but for scales 20 (top) and 30 $h^{-1}$Mpc (bottom).

3.3.2 Selection Function

As with the counts-in-cells analysis this analysis is affected by the selection functions of the samples. If the relative density of the two samples $\bar{\rho}_I/\bar{\rho}_O$ changes significantly across a shell then this will introduce spurious discrepancies between the optical and IRAS clustering. As before, we could volume limit each sample within each shell. This discards more galaxies than necessary and for this analysis our statistics are sensitive to the mean density. To avoid any bias we require $\bar{\rho}_I/\bar{\rho}_O$ to be constant over each shell. We can estimate this quantity neatly, without calculating either se-
lection function independently, and (in the case of no relative bias) independently of clustering. It is simply the ratio of the two $\rho(z)$ distributions as shown in Figure \ref{fig:ratio}. We have fitted the observed ratio with a power law giving us 
$$\hat{\rho}/\rho_0(z) \propto (1 + z)^{27}.$$ In any given shell we then randomly throw away galaxies from the most dense sample to give us a constant $\hat{\rho}/\rho_0$ over that shell. This is very efficient since the errors are dominated by the least dense sample and in any case the gradients across a shell are small. Our results were not significantly affected by this correction.

3.3.3 Cells with few galaxies

Technically we can calculate $X^2$ for all cells where there is at least one galaxy in each class but the $X^2$ statistic is only distributed as $\chi^2$ in the limit of large numbers so it is unwise to take those results at face value. To avoid the problem of poorly sampled cells requires some care. If we simply exclude cells with fewer than $n_{\text{lim}}$ galaxies in any class we would end up biasing our results. This is because we would be excluding cells with a higher over density in the class which has the lower mean density, and thus concluding that the other population had more under-dense regions. It is thus clear that any thresholding has to be imposed on $\delta \rho/\bar{\rho}$ rather than $n$. We thus impose the limit

$$\frac{n_{\text{lim}}}{n} > \frac{n_{\text{lim}}}{\min(n_{s})},$$

where we have replaced $\delta \rho/\bar{\rho}$ by the estimator $n_{s}/n$. Now our limit, $n_{\text{lim}}$, determines the smallest number of galaxies of the sparse class that we would expect to appear in any cell. The actual $\delta \rho/\bar{\rho}$ limit varies from shell to shell, allowing us to extract the maximum information from the surveys. These sparse cells still contain useful information, especially about the voids. Fortunately our null hypothesis allows us to group cells together. We choose to group sparse cells with their neighbour but one, rather than excluding them totally. We present results for $n_{\text{lim}} = 1, 2, 3, 4, 5$.

3.3.4 Results

The results of our contingency table analysis are presented in Table \ref{table:contingency}. The probability $P(X^2, \nu)$ is that of obtaining a value of $X^2$ higher than that observed assuming our null hypothesis and that $X^2$ is distributed as $\chi^2$ with $\nu$ degrees of freedom. One can see from this Table that the hypothesis that the optical galaxies and the IRAS galaxies are identically tracing the same underlying field is ruled out on scales $5 - 20 h^{-1}\text{Mpc}$ with a high level of significance (\geq 99\%). On the $30 h^{-1}\text{Mpc}$ with $n_{\text{lim}} = 5$ it is impossible to rule out the null hypothesis, although the high $\chi^2$ at smaller $n_{\text{lim}}$ may hint that our statistics are too poor for this to be informative.

As stated before, we expect some discrepancy because of the over-representation of elliptical galaxies in optical surveys. In Table \ref{table:contingency} we present the results after having excluded the CfA ellipticals. Here the story is markedly different. We are unable to exclude our null hypothesis for scales 5, 10, 20 $h^{-1}\text{Mpc}$. At $30 h^{-1}\text{Mpc}$ we can formally rule out our null hypothesis at the 95 per cent level. With only two degrees of freedom this result should be treated with caution.

<table>
<thead>
<tr>
<th>Scale /h^{-1}\text{Mpc}</th>
<th>$X^2$</th>
<th>$\nu$</th>
<th>$P(X^2, \nu)$</th>
<th>$n_{\text{lim}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>86.6</td>
<td>49</td>
<td>0.07470</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>70.6</td>
<td>31</td>
<td>0.00640</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>49.1</td>
<td>18</td>
<td>0.01036</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>25.7</td>
<td>10</td>
<td>0.41861</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>26.3</td>
<td>8</td>
<td>0.09185</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>86.4</td>
<td>47</td>
<td>0.04046</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>68.3</td>
<td>33</td>
<td>0.02920</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>57.4</td>
<td>27</td>
<td>0.05782</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>39.8</td>
<td>19</td>
<td>0.34293</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>44.1</td>
<td>15</td>
<td>0.01071</td>
<td>5</td>
</tr>
<tr>
<td>20</td>
<td>41.7</td>
<td>17</td>
<td>0.07288</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>38.4</td>
<td>11</td>
<td>0.00677</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>37.4</td>
<td>9</td>
<td>0.00226</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>30.7</td>
<td>7</td>
<td>0.00702</td>
<td>4</td>
</tr>
<tr>
<td>20</td>
<td>28.8</td>
<td>6</td>
<td>0.00662</td>
<td>5</td>
</tr>
<tr>
<td>30</td>
<td>12.8</td>
<td>7</td>
<td>7.60401</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>11.5</td>
<td>6</td>
<td>7.37997</td>
<td>2</td>
</tr>
<tr>
<td>30</td>
<td>12.4</td>
<td>6</td>
<td>5.32677</td>
<td>3</td>
</tr>
<tr>
<td>30</td>
<td>12.9</td>
<td>5</td>
<td>2.43994</td>
<td>4</td>
</tr>
<tr>
<td>30</td>
<td>5.7</td>
<td>4</td>
<td>22.12649</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 4. Contingency table analysis results. Under the null hypothesis that the optically selected galaxies trace the same field as the IRAS galaxies $P(X, \nu)$ gives the probability of obtaining $X^2$ higher than the observed value, assuming that $X^2$ is distributed as $\chi^2$ (the statistic $X^2$ is defined in equation (3)). $n_{\text{lim}}$ is the threshold used in merging cells with few galaxies (equation (4)).

We thus conclude that the detectable differences in clustering between these two surveys are mainly due to the higher proportion of ellipticals in the optically selected samples. It is nevertheless surprising that significant differences persist to large scales where we might have expected the overabundance of ellipticals in rich clusters to be of less importance. This could be because of the Coma ‘finger of god’ which is quite extensive in redshift space. Alternatively it might suggest that the morphology-density relation extends to large scales.

3.4 Relative Bias Ratio

The contingency table analysis described in Section 3.3.3 tests only a very particular linear-biasing scenario (where $b_0 = b_1$). A more general model allows for a relative bias between the different samples (equation (3)). An extension of the above method allows us to test this more general model. To allow comparison with estimates of $b_0/b_1$ from other methods (see e.g. Section 3.3.1), we will consider a model where the optical over-density is a constant multiple $(b_0/b_1)$ of the IRAS over-density.

We thus construct a series of null hypotheses that the two galaxy distributions are linearly biased with a given ratio of bias parameters. Since we have no direct measures of the underlying mass distribution we can only investigate the ratio of $b_0/b_1$. For clarity we will discuss the biasing of the two populations relative to an underlying mass distribution.
The variance is determined from Poisson statistics, neglecting errors and covariance terms arising from $\bar{n}_{ij}$ (in agreement with our contingency table analysis if $b_O/b_I = 1$):

$$\text{var}(\frac{1}{b_I} \frac{\Delta n_{ij}}{\bar{n}_{ij}}) = \frac{\Delta_i + 1/b_I}{b_I \bar{n}_{ij}}.$$  

(20)

In the previous analysis we merged cells with an estimated under-density low enough that, for the class of galaxies with the lowest mean density at that shell, we would expect fewer than $n_{\text{lim}}$ galaxies in that cell. With our new approach a similar argument introduces the following threshold

$$\Delta_i > \max_j \left( \frac{1}{b_I} \frac{n_{\text{lim}} - \bar{n}_{ij}}{\bar{n}_{ij}} \right).$$  

(21)

The results presented here have taken $n_{\text{lim}} = 5$. The $\Delta_i$ limit varies with $b_O/b_I$, as does $\Delta_i$ itself so the number of cells merged depends on the value of $b_O/b_I$ being tested.

We calculate the statistic $X^2$ for a range of 0.5 < $b_O/b_I$ < 3 and, under the assumption that this statistic is distributed as $\chi^2$ with $M - N$ degrees of freedom, we calculate the probability $P(X^2, \nu)$ as before.

The results are presented in Figure 8 which shows the probability of obtaining a higher value of $X^2$ than that observed as a function of the assumed biasing ratio $b_O/b_I$. The curves in Figure 8 are for the four scales 5, 10, 20 and 30$h^{-1}$ Mpc. The curves are not smooth because our cell merging threshold depends on the assumed bias ratio and jumps occur when new cells become merged. On scales of 5 to 10$h^{-1}$ Mpc all linear-biasing models relating optical and IRAS galaxies can be ruled out with around 99 per cent significance. At the 5$h^{-1}$ Mpc scale this is perhaps unsurprising since we are in the non-linear regime ($b > 0$) so a naive justification for a simple linear-biasing breaks down. Also on these smaller scales we expect that the ‘fingers of god’ (more prominent for the ellipticals in clusters) will have a non-linear impact on our analysis over and above any discrepancies in the over-densities. However, on scales of 10$h^{-1}$ Mpc we might have expected that a linear-biasing model would have been reasonable. On larger scales (20 and 30$h^{-1}$ Mpc) we do find that there are linear-biasing models that we cannot reject. A bias ratio of 1.7 < $b_O/b_I$ < 3.3 cannot be rejected with more than 95 per cent confidence for 20$h^{-1}$ Mpc cells and likewise 1.1 < $b_O/b_I$ < 2.0 is acceptable at 30$h^{-1}$ Mpc. It is interesting that we can reject all $b_O/b_I$ < 1.7 with more than 95 per cent confidence since it is within this region that most estimates of $b_O/b_I$ lie (Section 3.1). Taken at face value this would imply that optical and IRAS density fields are not linearly related even on scales of 20$h^{-1}$ Mpc.

Since we were unable to reject the hypothesis that $b_O/b_I = 1$ after the exclusion of optical ellipticals, we expect a wider range of acceptable bias ratios when comparing optical-spirals with the IRAS sample. This is indeed the case and for brevity we do not show the $P(X^2, \nu)$ plots for these samples. Suffice to say we cannot rule out 0.8 < $b_O/b_I$ < 3 between 5 and 20$h^{-1}$Mpc (all linear-biasing models could be ruled out with 95 per cent confidence at 30$h^{-1}$ Mpc but as before the few degrees of freedom at this scale cause us to be sceptical of this results).

4 CONCLUSION

We have completed two new large, deep IRAS redshift surveys. The large scale clustering properties of these surveys...
have been analysed and found to be consistent with previous IRAS redshift surveys, strengthening previous conclusions that the IRAS galaxy distribution is inconsistent with the standard CDM predictions.

We noted two particular shortcomings in the counts-in-cells analysis. Firstly, the variances determined using different grid patterns on the same data show a scatter comparable with the quoted errors. Secondly the method is not completely independent of the selection function and we demonstrate one way of overcoming this bias.

We are able to combine all the IRAS samples together to obtain a single estimate of the variance on large scales with errors that will not be matched until the completion of the full PSC-z survey (Saunders et al. 1995a; Saunders et al. 1995b). Converting these variances into the form of a dimensionless power-spectrum (correcting for redshift-space distortions), we compare them with a mixed dark matter (MDM) model. The slope of this model spectrum appears slightly steeper than that observed but the assumptions involved in transforming theory to data are sufficiently uncertain that we can not read too much into this.

A comparison of our variance estimates with those from the APM-Stromlo redshift survey suggests $b_O/b_I = 1.20 \pm 0.05$, assuming a linear-biasing model with ratio independent of scale. This is higher than that found by the APM-Stromlo team themselves (Loveday et al. 1992) because we find smaller IRAS variances than E90.

Using our FSS-z sample and the first CfA strip we directly compare the distributions of optical and IRAS galaxies (using a physically meaningful definition of the two classes). This is done on a point-by-point basis so we are able to test the hypothesis that the two populations are related by a linear-bias ratio ($b_O/b_I$).

We are able to reject the hypothesis that $b_O/b_I = 1$ with a very high degree of significance on the scales 5, 10 and 20 $h^{-1}$ Mpc. This hypothesis is, however, acceptable if we exclude ellipticals and S0 galaxies from the optical sample, suggesting that the discrepancies are due to the morphology density relation and the under-representation of ellipticals in IRAS samples. On scales of 30 $h^{-1}$ Mpc our conclusions would be reversed but we suspect that our statistics are too poor at this scale for this to be very meaningful.

Allowing $b_O/b_I$ to vary away from 1 we find that only on the largest scales are any optical/IRAS linear-biasing models acceptable, if elliptical galaxies are included in the optical samples. Even on scales of 20 $h^{-1}$ Mpc, linear-biasing with $b_O/b_I < 1.7$ is ruled out with more than 95 per cent confidence. It should of course be stressed that in this particular volume the optical galaxy distribution is dominated by the Coma cluster and so the distorting effect of the elliptical galaxies has to some extent been amplified.

This non-linear clustering relation between ellipticals and IRAS galaxies may present complications for the interpretation of any galaxy surveys that include both elliptical and spiral galaxies. As yet we have been unable to determine any differences in the clustering properties of IRAS galaxies and spiral galaxies and so surveys composed of either of these can be meaningfully compared.

ACKNOWLEDGMENTS

Figure 8. $P(\chi^2, \nu) v b_O/b_I$ for IRAS and optical galaxies on various scales.
SJO and ANT acknowledge PPARC support. WS was supported by a PPARC Advanced Fellowship. We thank John Huchra, Ray Wolstencroft, Quentin Parker & Roger Clowes and Marc Davis & Michael Strauss for providing redshifts in advance of publication. We acknowledge use of the ING telescopes and thank the staff for their invaluable assistance. Much use has been made of the STARLINK resources.

REFERENCES

Efstathiou, G. MNRAS(in press)