Power Spectrum Analysis of the 2dF Survey
George Efstathiou January 2001

1 Introduction

This set of notes describes some preliminary work on power spectrum estimation of the 2dF Survey. The estimator applied in Section 3 is a straightforward FFT estimator and so is non-optimal. Work is in hand on applying maximum likelihood estimators. The power spectrum estimation code used here is almost identical to the code that I developed for the 2dF proposal in 1995, and which I used to analyse different survey geometries. Most of the new work that I have done over the last few months has been on analysing the completeness of the survey and in defining a survey mask.

In this analysis, I have decided to use only the SGP data. This is because I believe that the SGP photometry is reasonably well understood. I do not want, at this stage, to introduce zero point offsets in the NGP photometry. It is trivial to apply the analysis described here to the NGP data. I also think that we should analyse the NGP and SGP data *separately* and averaging the results because of a possible systematic photometric offset between the north and south.

I have used data from the odz files on the 2dF database as of 19/10/00. I decided to freeze the analysis to this data set. It is obviously straightforward to update the analysis to the latest set, but I wanted to get a thorough understanding of the data at one fixed time. I found some problems with the data base which are described in Section 2.2.

2 Defining a completeness mask

2.1 Input data

Starting from the odz files, I find all galaxies with redshifts which have an iqual flag greater than or equal to 2. Some galaxies have repeat measurements and so I search for multiple entries and choose the redshift with the highest iqual flag. This produces a unique catalogue of galaxies with measured redshifts. This catalogue is plotted in Figure 1 and contains 88788 galaxies.

Since the tiling strategy is based on overlapping tiles, the observed region in Figure 1 is incomplete. I define a complete mask by locating the area of intersecting tiles that has been completely observed. This region has a characteristic 'swiss cheese' appearance and is plotted in Figure 2. The complete region of the sky contains 71499 galaxies with measured redshifts, so although the surveyed area in Figure 2 is noticeably smaller than in Figure 1, we lose only about 10% of the galaxies.

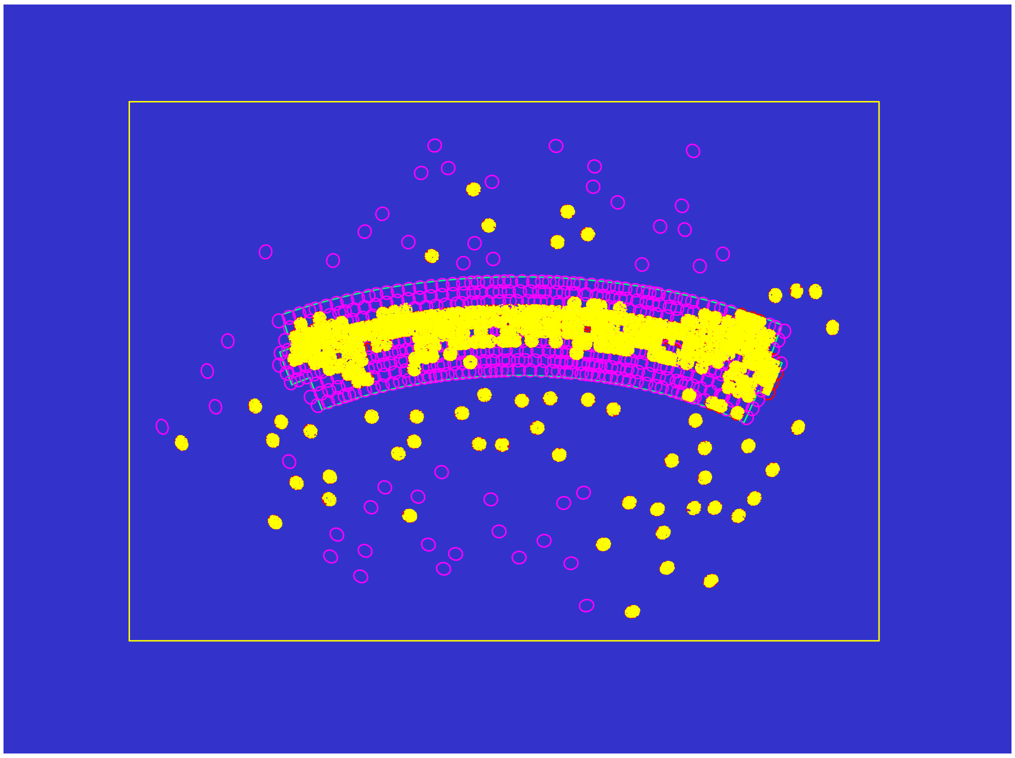

Figure 1: Equal area projection of SGP region. Distinct galaxies with measured redshifts are plotted as yellow dots. Unobserved fields are plotted as the violet circles. Observed 2dF fields are plotted as red circles. The blue lines demarcate the boundary of the contiguous 2dF SGP region.

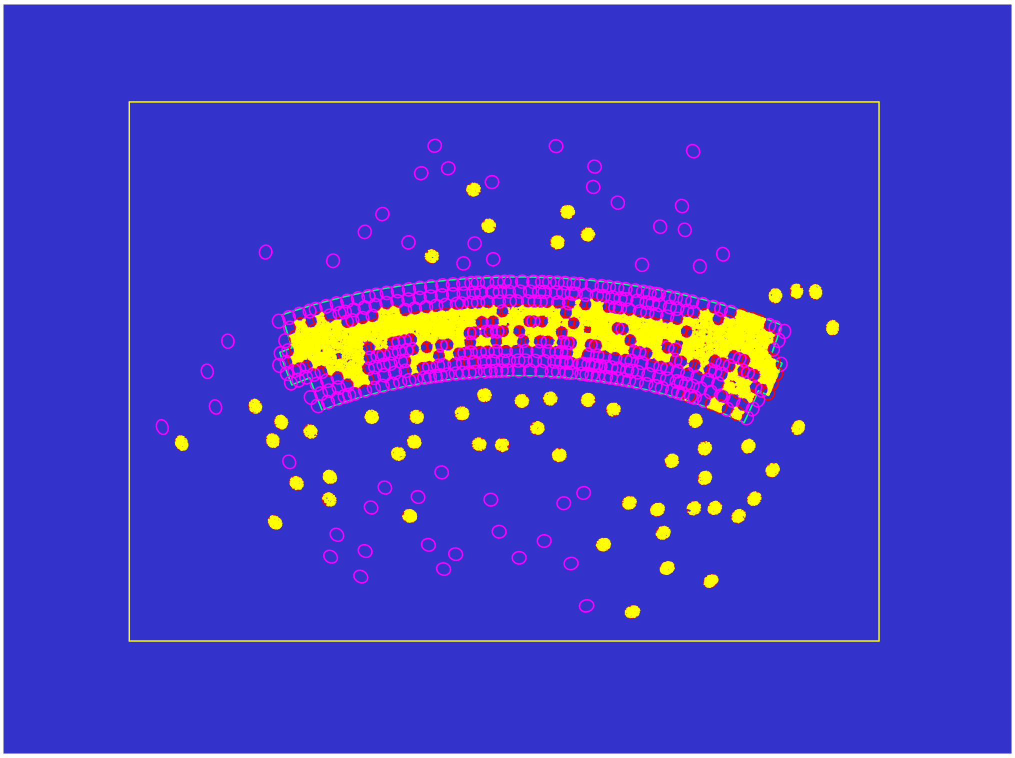

Figure 2: The mask for the SGP 2dF area. The colour coding of the fields is as in Figure 1. Here only galaxies within the area that has been completely observed are plotted, i.e. within the area that does not overlap with any violet circle.

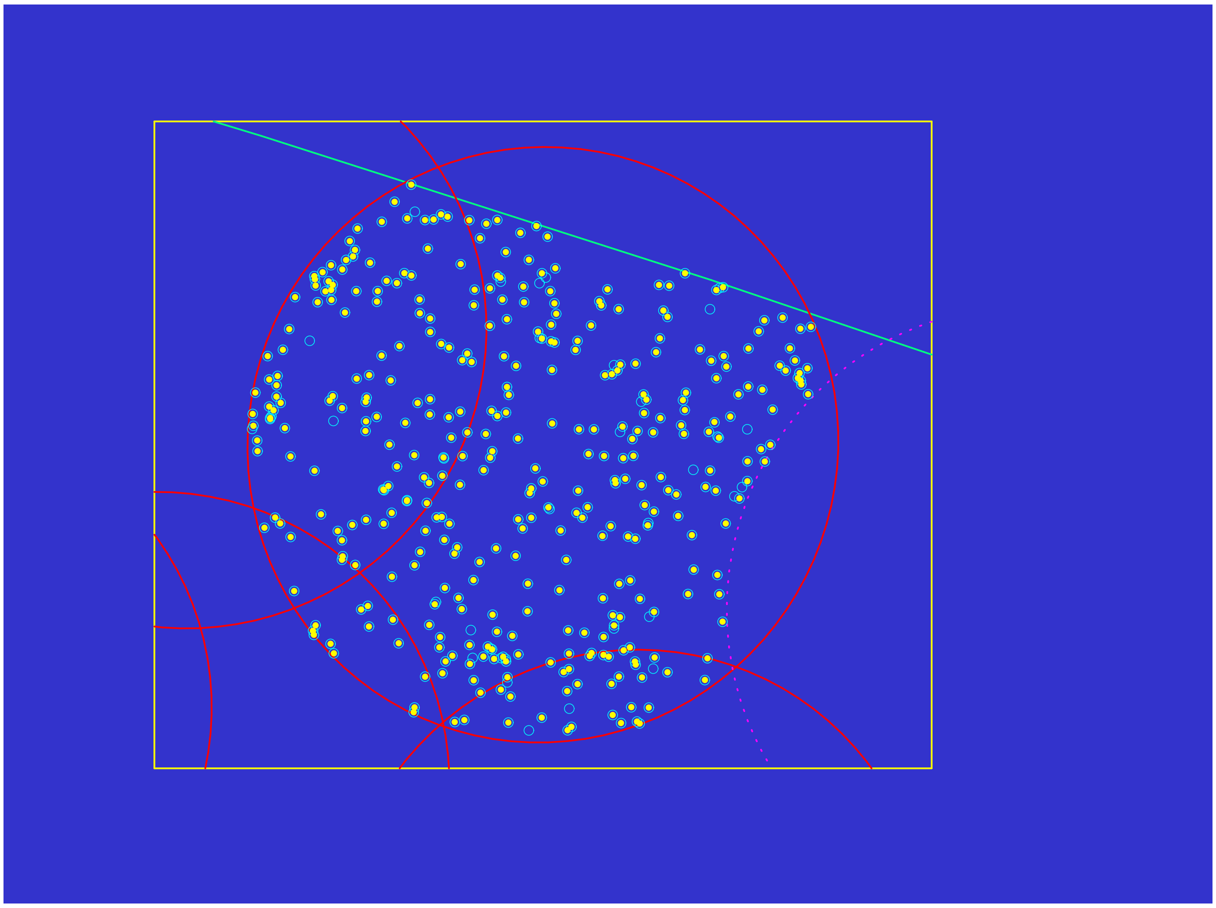

Figure 3: Example of field completeness. The figure is centred on SGP field number 4. All observed 2dF fields are plotted as red circles. Unobserved fields (one in this case) are plotted with violet dotted lines. The green line in the upper part of the diagram shows the SGP survey field boundary (field 4 lies across the edge of the survey). The completely observed region of field 4 contains 409 objects classified as galaxies in the parent catalogue (plotted as blue circles). 375 of these objects have high quality redshifts and are plotted as yellow dots. The completeness of field 4 is therefore 0.92.

The survey region shown in Figure 2 is not complete, of course, because our tiling and configuration algorithms do not target all galaxies and we fail to measure redshifts for some proportion of galaxies. An example is shown in Figure 3. This shows an estimate of the completeness of SGP field 4 *relative to the parent catalogue.* Thus, only those galaxies within the completely observed region are plotted (the overlaping field number 3 has yet to be observed). This particular field has a redshift completeness relative to the parent catalogue of 0.92. It is amongst the more complete of the 2dF fields, but not especially so.

Figure 4 shows a histogram of the completeness the SGP fields (not including the random fields). Since the fields overlap, there is no unique way of defining a completeness *relative to the parent catalogue.* [1] What I have done in Figure 4 is to compute the completeness of each

[1]We can define unambiguously the ratio of successful redshifts to targetted galaxies within a field (as listed in FieldDbase).

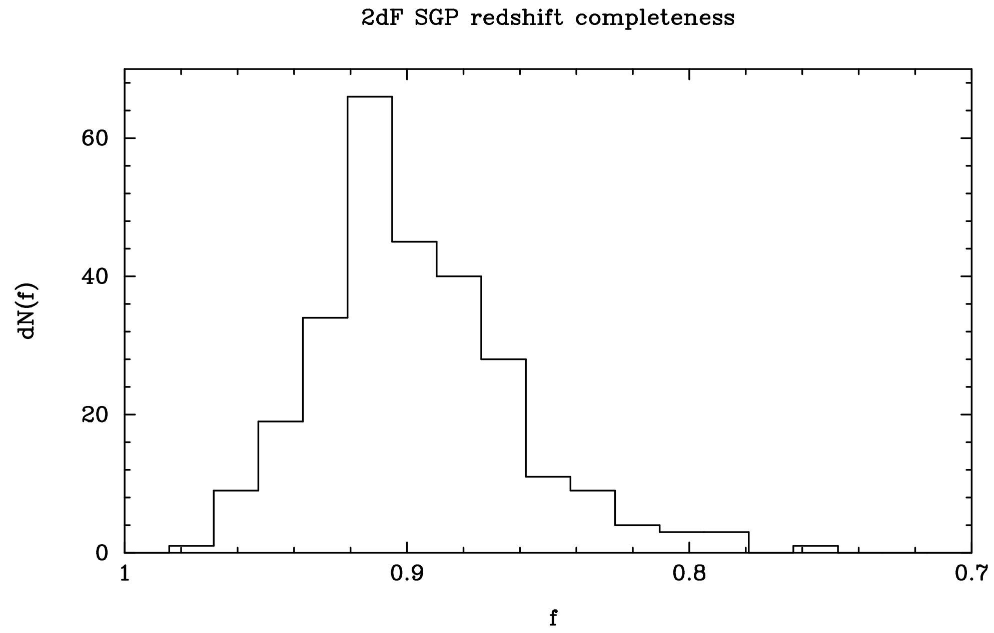

Figure 4: Histogram of completeness of 2dF SGP fields.

targetted field using all galaxies that overlap within the field. Refering back to Figure 3, this shows all galaxies within 2dF field 4, *but a significant fraction of these were observed in the exposures of the overlapping fields* 5, 6 *and* 61. If we define the completeness of field 4 as described above (0.92) what 'weight' do we assign to a galaxy that sits in an overlap region? In practice, completeness variations will correlate most strongly with the field number and so I will adopt the weight of the field *for which the redshift was measured.* However, one can see from Figure 4 that the completeness of the survey is remarkably high and uniform, and so weighting is not a critical issue.

2.2 Eliminated fields and possible data-base errors

There is a tail of fields with low completenesses. The following fields have been eliminated from the analysis:

Field	N_z	f	comment
130	33	0.53	only small fraction of field complete
155	198	0.53	cloudy weather
184	305	0.58	clear
203	313	0.67	clear
212	223	0.61	clear
418	135	0.68	some cloud

In each case *only data for one spectrograph* is given in the odz file and so the number of redshifts N_z is low leading to low completnesses f. Yet with the exception of field 155 the FieldDbase lists over 200 redshifts implying that there is data from both spectrographs. Note that in FieldDbase, the field listed as sgp556 is actually field sgp566.

2.3 The dN/dz distribution

Figure 5 shows the $dN(z)$ distribution for the complete SGP region. I have fitted this distribution by least squares to the following functional form:

$$\frac{dN}{dz} = Az^{\gamma}\exp\left[-\left(z/\overline{z}\right)^{\beta}\right], \tag{1}$$

with best fit parameters

$$\overline{z} = 0.135, \qquad \beta = 2.177, \qquad \gamma = 1.235. \tag{2}$$

Note that the peak of the distribution (1) occurs at redshift

$$z_{\rm peak} = \overline{z}\left(\frac{\gamma}{\beta}\right)^{1/\beta}. \tag{3}$$

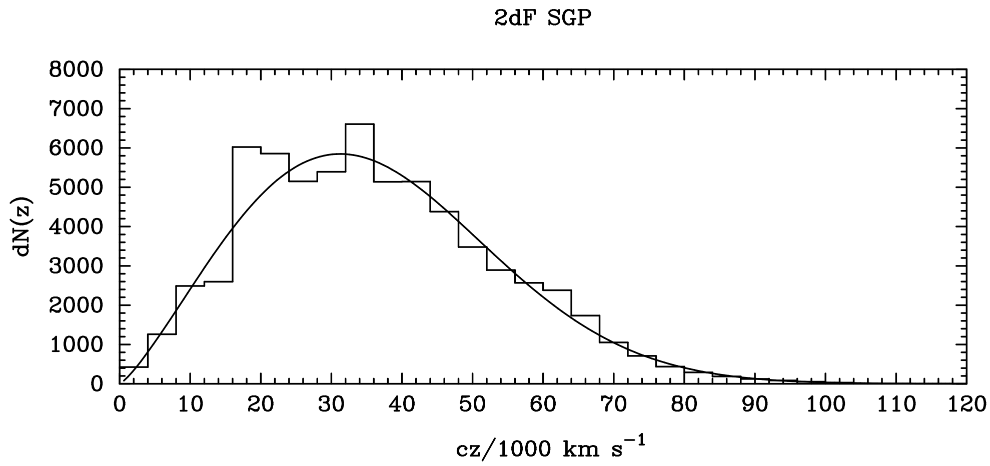

Figure 5: The redshift distribution of the SGP region. The solid line shows the best fitting function of equation (1).

For reference, I will define the radial coordinate of a galaxy to be the coordinate distance x, which for spatially flat universes is related to redshift by the integral,

$$H_0 x = c\int_0^z \frac{dz}{[\Omega_m(1+z)^3 + \Omega_\Lambda]^{1/2}}. \tag{4}$$

Unless otherwise stated, I will assume $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$. From the coordinate distance and the redshift distribution (1) we can compute the mean comoving galaxy number density as a function of redshift

$$\overline{n}(z) = \left(\frac{dN}{dz}\right) \frac{1}{\Omega_s} \frac{1}{x^2} \left(\frac{dz}{dx}\right), \tag{5}$$

where Ω_s is the solid angle of the unmasked survey area. For reference the solid angle of unmasked area of Figure 2, excluding the random fields, is 0.16sr.

2.4 Generation of random catalogues

In power spectrum estimation it is straightforward to account for the survey window by generating a random catalogue with the same $dN(z)$ distribution as the actual survey. Figure 6 shows a random catalogue projected on the sky, showing that the mask software produces an exact match to the complete survey area shown in Figure 2.

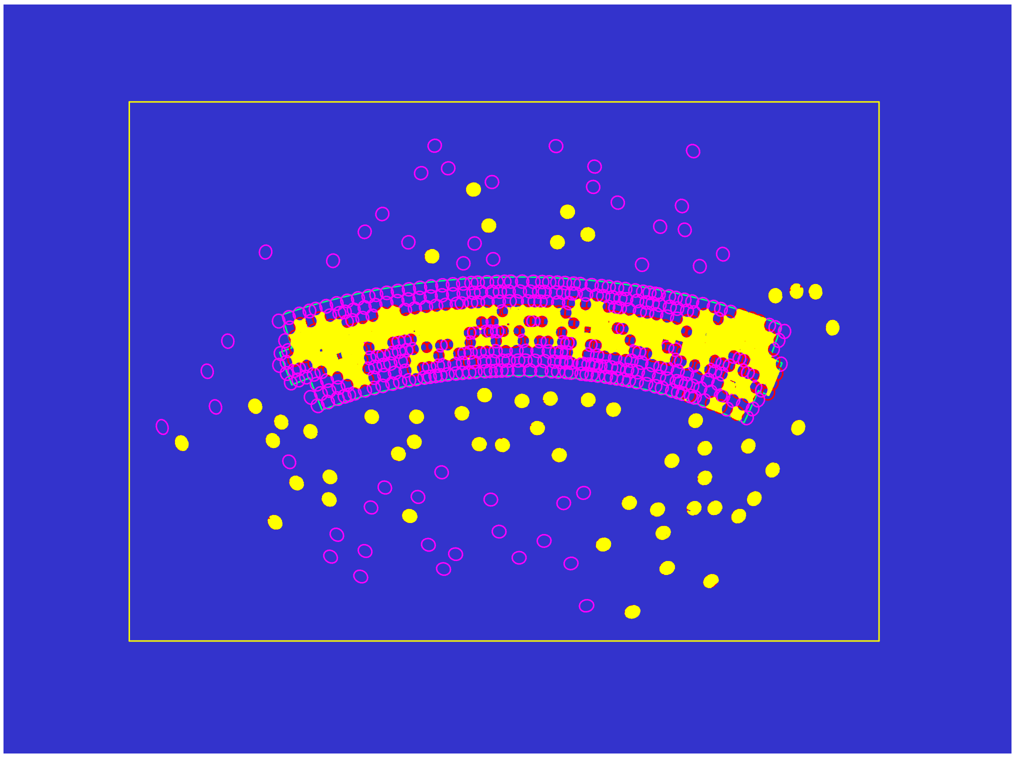

Figure 6: Equal area projection of a random catalogue generated with the same sky mask as in Figure 2.

The $dN(z)$ distribution for the random catalogue is shown in Figure 7 and the solid line shows the fit to the $dN(z)$ of the galaxy survey shown in Figure 5. I generated random points to a redshift of $z = 0.2$ and I will use this upper redshift limit in the power spectrum analysis unless otherwise stated.

I will adopt the following approach to correcting for incompleteness. I generate a uniform random catalogue with a fixed $dN(z)$ distribution as in Figures 6 and 7. The mean $dN(z)$ is therefore assumed to be uniform across the sky. Each galaxy has a weight given by

$w_g = 1/f_i$ where f_i is the completeness factor of the field i on which the galaxy was observed as calculated in Figure 4. In computing the power spectrum, one can weight the galaxies by the factors w_g if desired, and one can correct the shot noise contribution accordingly. We could implement a more elaborate correction for incompleteness. For example, we could attempt to estimate an effective magnitude limit (or joint magnitude/surface brightness limit) for each field and generate a random catalogue that includes the resulting modulations in the $dN(z)$ distribution. However, by construction, the completeness of the survey is high and strongly peaked at around $f_i = 0.9$. I very much doubt whether we can gain much from a more elaborate correction for incompleteness that the simple corrections that I have adopted here.

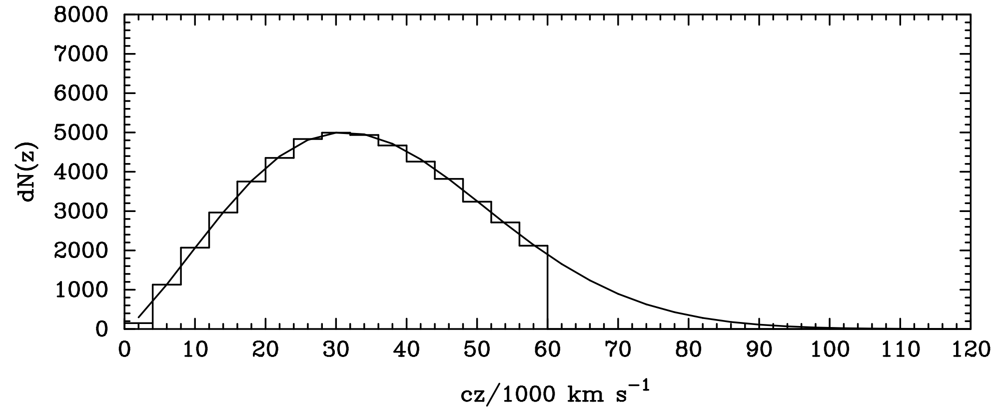

Figure 7: The redshift distribution of the random catalogue. The solid line shows the fit to $dN(z)$ for the galaxy survey.

To compute the power spectrum, I have gridded the data and the random catalogues on a cubical grid in the coordinate variable x. (Thus wavenumber $k \equiv 2\pi/x$). In this summary, I have performed FFTs on a 256^3 grid, with an effective resolution of $3.98h^{-1}$Mpc per cell. This is so that I could compute FFTs on my PC. I could increase the size of the grid to a maximum of 2048^3 using the COSMOS machine if required. Figure 8 shows the distribution of the data and the random catalogue within the grid.

3 Numerical Simulations

Before presenting results for the 2dF survey, it is worth analysing some simulated 2dF samples. The analysis described here is similar to the analysis that I did for the 2dF proposal, except now I can analyse the mask of the real 2dF survey.

I have analysed two sets of simulated 2dF surveys with the completeness mask of Figure 2. The first set is based on 10 simulations with $\Gamma = 0.2$, $\Omega_\Lambda = 0.8$, $\Omega_m = 0.2$ (LCDM ensemble) and the second is based on 10 simulations with $\Gamma = 0.5$, $\Omega_\Lambda = 0$, $\Omega_m = 1$ (SCDM ensemble). Each simulation contains 4×10^6 particles within a computational cube of side

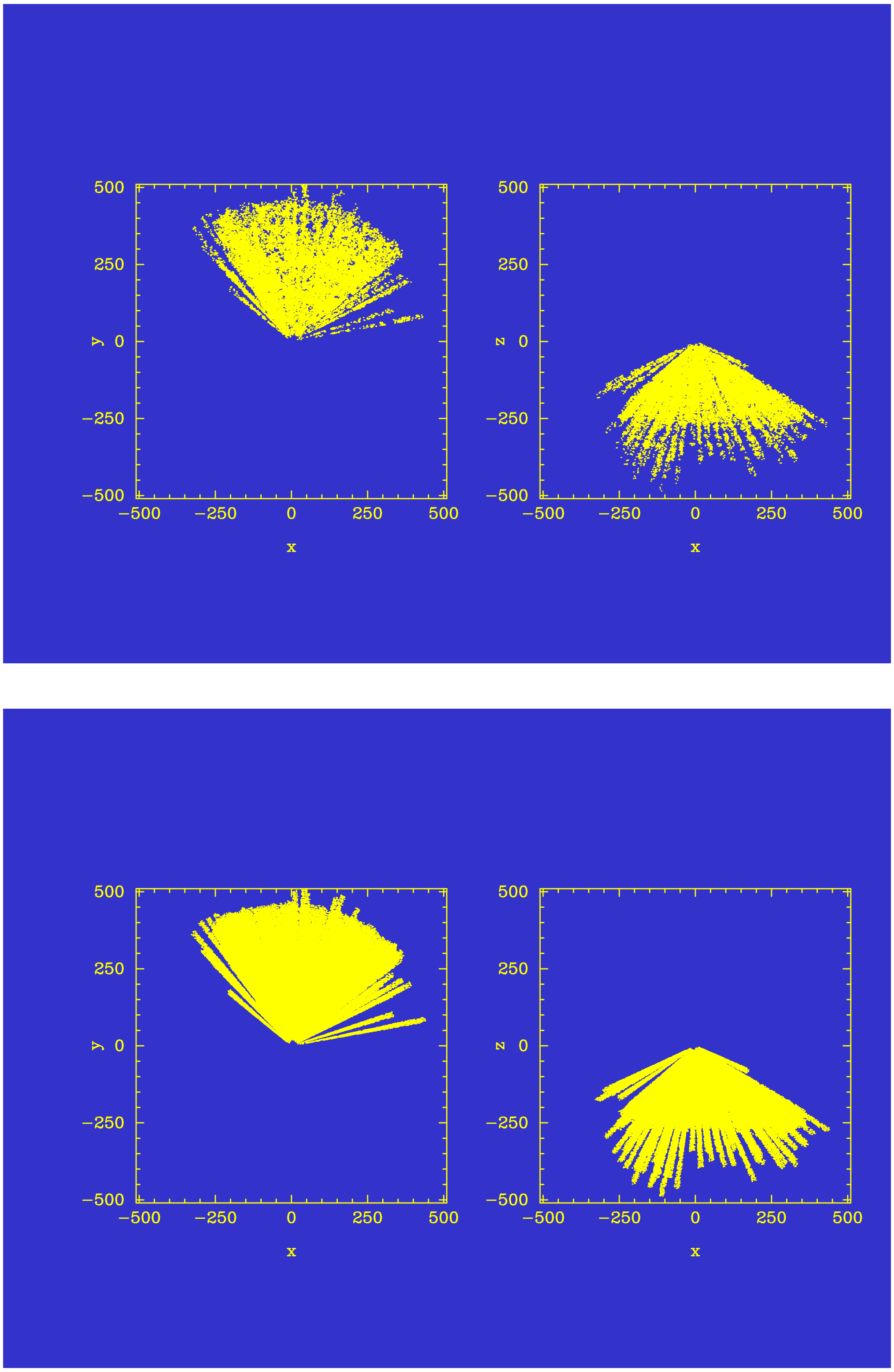

Figure 8: The upper picture shows the distribution of the redshift data (including random fields) withing the cubical box used to compute the FFT. The lower picture shows the distribution of the random catalogue.

$L_b = 600h^{-1}$Mpc normalised to $\sigma_8 = 1$ at the present day. To generate the simulated 2dF surveys, I select points with the $dN(z)$ distribution of the actual survey, but since the simulations have a mean density of $0.019h^3\text{Mpc}^{-3}$, the simulated surveys are semi-volume limited at $z < 0.06$. Since the simulations are on a constant time-slice, I use Euclidean geometry for both the simulations and random catalogues, relating coordinate distance to redshift via $x = cz/H_0$. I generate one simulated 2dF SGP hemisphere with a maximum redshift of $z = 0.2$ from each of the simulations. Each simulated 2dF area is strictly independent and does not require any replication of the periodic simulation boxes.

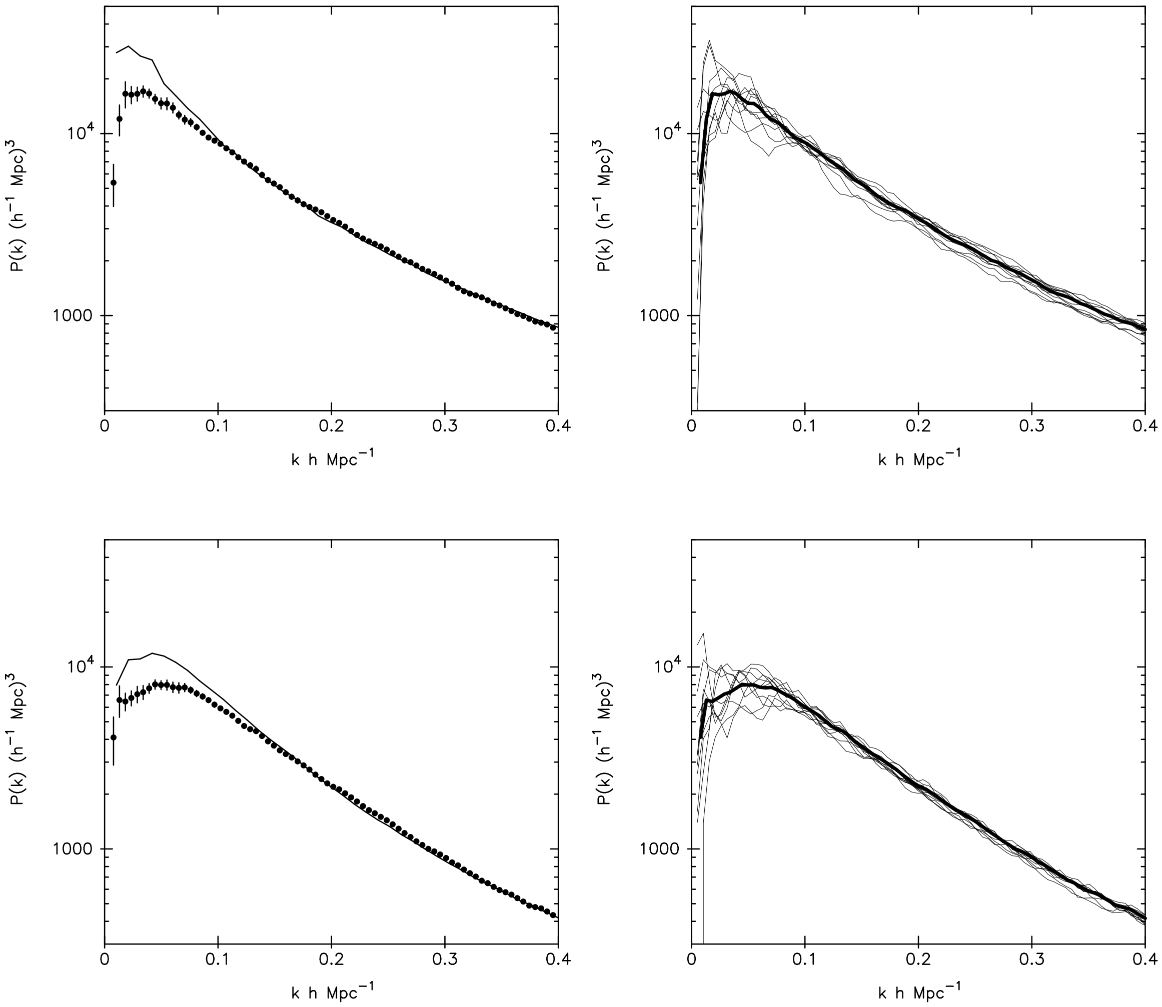

Figure 9: Power spectra from simulated 2dF SGP surveys generated from numerical simulations. Top two pictures show the LCDM simulations and the bottom two pictures show the SCDM simulations. Figures to the left show the spectra averaged over 10 simulations (filled circles). The error bars on these points show one-σ errors on the mean. The solid lines show the mean (redshift space) distorted power spectra estimated from the parent simulations. Figures to the right show the esimated power spectra from each of the 10 simulated 2dF surveys (thin lines) together with the mean (thick lines).

Each galaxy and random point is assigned a weight

$$w(z) = \frac{1}{(1 + \overline{P}\overline{n}(z))}, \tag{6}$$

where $\overline{n}(z)$ is the mean galaxy density computed from equation (5) for redshifts larger than the semi-volume limit $z = 0.06$ and is equal to $0.019h^3\mathrm{Mpc}^{-3}$ at lower redshifts. In equation (6) $\overline{P}$ is a 'mean' power spectrum weighting factor which I take to be $\overline{P} = 8000h^{-3}\mathrm{Mpc}^3$. Note that a subtle, but important point is that the ratio of galaxy to random mean densities is computed via the weighted summations

$$\alpha = \frac{\sum_{i=1}^{N_G} 1/(1 + 4\pi J_3 \overline{n}(z_i))}{\sum_{i=1}^{N_R} 1/(1 + 4\pi J_3 \overline{n}(z_i))}, \qquad (7)$$

for which I adopt $4\pi J_3 = 10000h^{-3}\mathrm{Mpc}^3$.

In the results shown in Figure 9, I have exclude the random fields. The figures to the left show the effect of the convolution of the survey window function and the 'integral constraint' arising from the density estimates of equation (7) on the shape of the power spectra. The effects are as described in the 1995 proposal. In fact, in the 1995 proposal I wrote a code which explicitly convolved an input spectrum with the exact anisotropic window function of the survey and which exactly reproduced the type of smoothing shown in Figure 9. I will update that programme and run it on the new simulations shortly. In a paper with Tadros (MNRAS 1996, 282, 1381) I managed to derive simple formulae for the integral constraint bias for the case of volume limited surveys. However, the formulae for more general weightings are extremely complicated and difficult to evaluate. One can see from Figure 9 that the integral constraint becomes significant at wavenumbers $\lesssim 0.03h\mathrm{Mpc}^{-1}$.

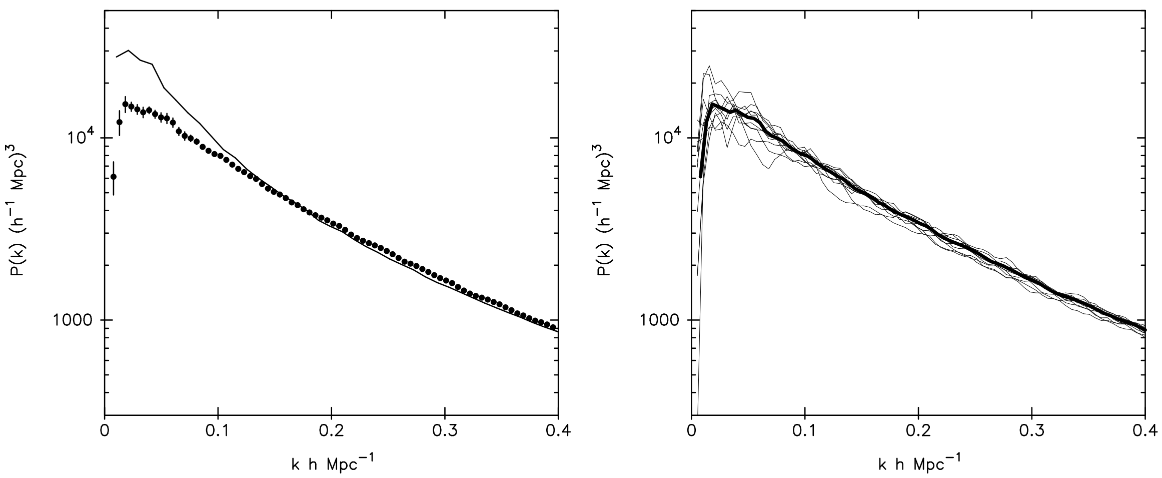

Figure 10: Power spectra of LCDM simulations including random fields (compare with the upper two panels of Figure 9.

From the 1995 proposal, the window function for the random fields will produce a still broader convolution of the real power spectrum, will will reduce the errors in $P(k)$. This is illustrated in Figure 10. As one can see, including the random fields reduces the variance at wavenumbers $\lesssim 0.1hmpci$, but also decreases the amplitude because the window function of the random fields is extremely anisotropic.

Figure 11 shows a clearer picture of the sampling fluctuations. Here I have simply the first four of the LCDM simulated 2dF surveys (expanding the scales of the ordinate and abscissa)

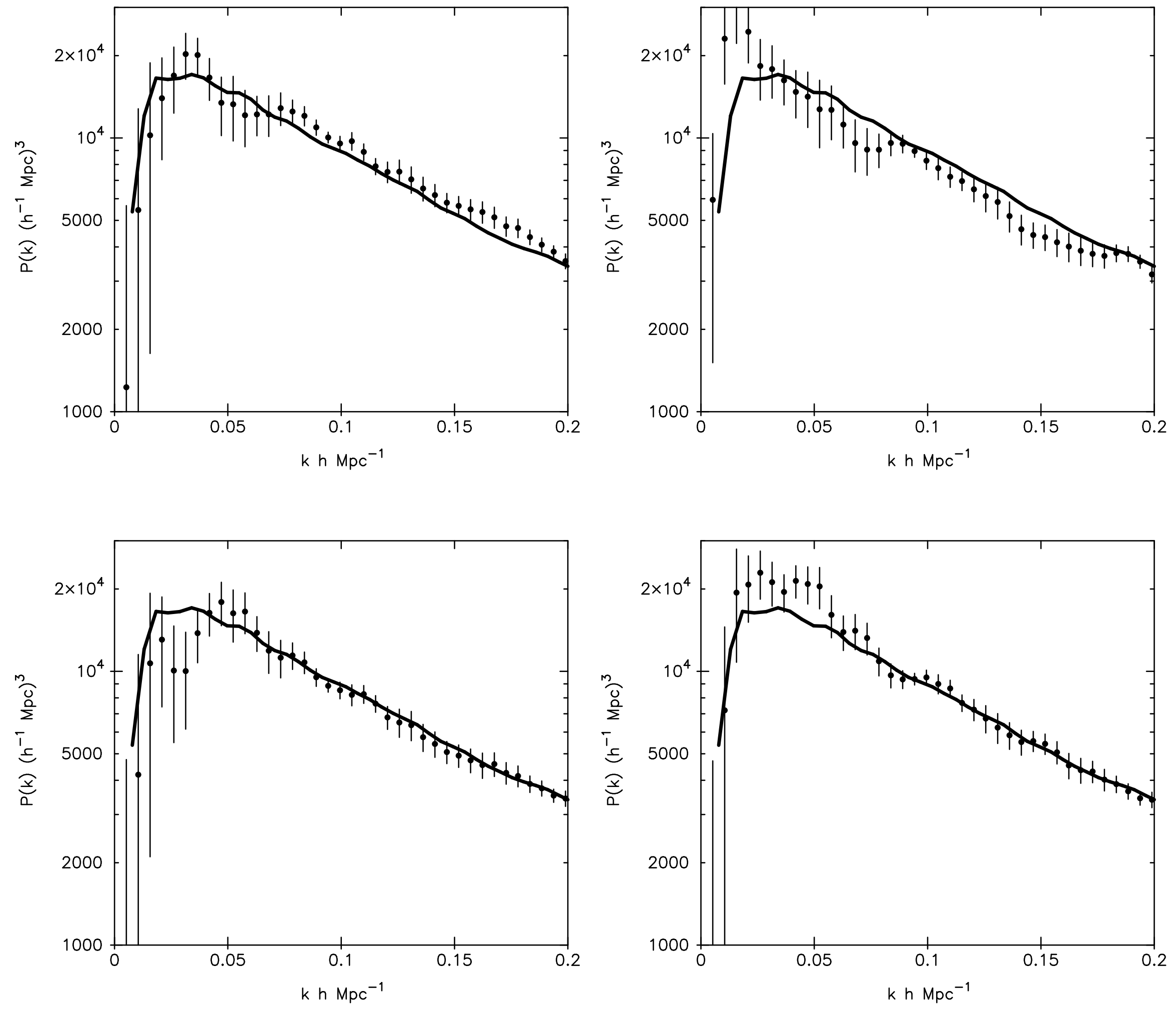

Figure 11: Power spectra (and 1σ errors) of four LCDM 2dF simulations (excluding random fields). The thick line shows the mean over 10 simulations.

and plotted their power spectra and 1σ error bars. The thin lines show the mean power spectrum of the 10 simulations. Notice that the power spectra from each single simulation have 'wiggles' with a characteristic width of $\Delta k \sim 0.05 hmpci$. This simply reflects the coherance scale of the survey window function. Power spectra points are highly correlated over this typical width. The oscillatory features are therefore an artefact of the survey window function. *We should therefore be extremely skeptical of interpreting wiggles in the power spectra as baryonic features.* Not only do we expect to see wiggles, as shown in Figure 11, but the convolution of the power spectrum with the survey window function will depress the amplitude of any real wiggles (as discussed in the 1995 proposal and shown in Figures 9 and 10).

4 Application to the 2dF Survey

Figure 12 shows two estimates of the power spectrum of the 2dF SGP region (excluding random fields). The only difference in these two estimates are the redshift ranges: the filled circles show the redshift range $0.01 < z < 0.2$ and the open circles show the redshift range $0.05 < z < 0.15$. This gives an indication of the sensitivity of the analysis to the redshift range which may be caused by biases in the fit to the $dN(z)$ distribution, luminosity dependent biases etc. These effects clearly need to be investigated in detail.

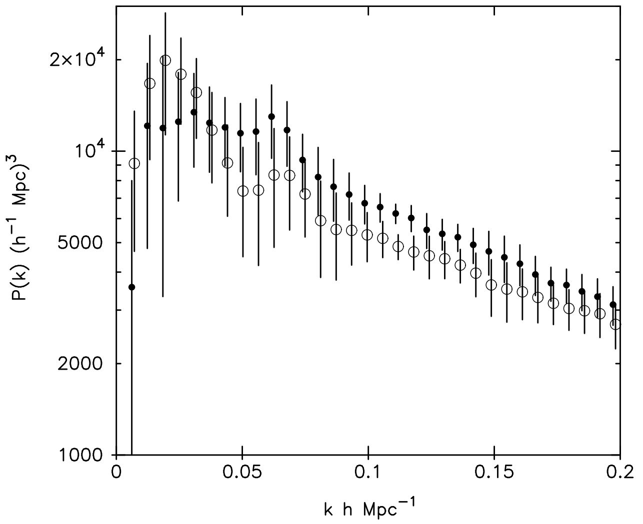

Figure 12: Power spectra of the 2dF SGP survey (excluding random fields). Filled circles show the redshift range $0.01 < z < 0.2$ and open circles show the redshift range $0.05 < z < 0.15$.

Figure 13 shows a comparison of my results (filled circles, as plotted in Figure 12) with those calculated by independently by Percival and Peacock (PP). The agreement between the twp estimates is striking (and must be fortuitous, given the variation seen in Figure 13). Notice that the 'wiggle' in the power spectrum at $k \sim 0.07h\mathrm{Mpc}^{-1}$ is actually more prominent in my estimate that in PPs. Note also that PP analyse both NGP and SGP fields.

In Figure 14, I plot my estimate of $P(k)$ against three-dimensional power spectrum estimated for the parent APM survey using maximum-likelihood (see Efstathiou and Moody, astro-ph/0010478). The maximum-likelihood estimate is an optimal inversion of the 2-d power spectrum to 3-d and so is unaffected by redshift space distortions. The agreement between the two estimates is spectacular and actually surprised this author. The agreement in amplitude is consistent with a small distortion between redshift-space and real-space. More importantly, however, we can see that the shapes agree and that the error bars of the redshift survey are already considerably smaller that the error bars of the inverted estimate. This is very interesting, because (a) I have used only a small fraction of the 2dF redshift survey ($\sim 70,000$ galaxies); (b) the $P(k)$ analysis described here is non-optimal and it should be possible to reduce the error bars by applying an ML estimator.

5 What needs to be done?

- I need to resurrect and check my window convolution code and use it to verify the convolution in Fig. 9 and to fit theoretical models to the estimated $P(k)$.

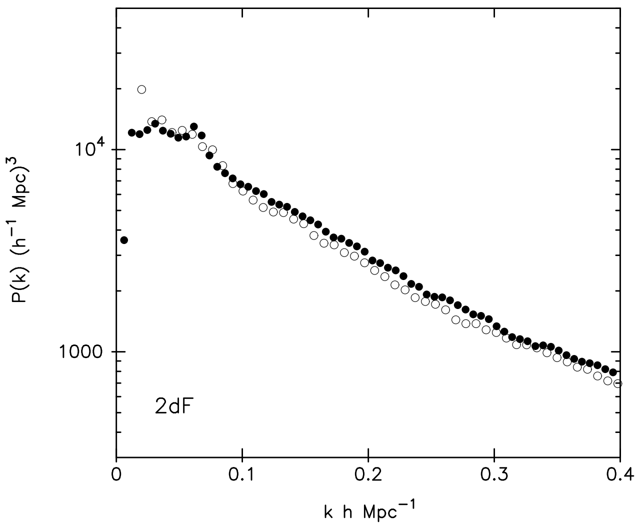

Figure 13: Filled circles show my estimate of $P(k)$ for the SGP region. Open circles show Percival and Peacock's estimate for the whole survey.

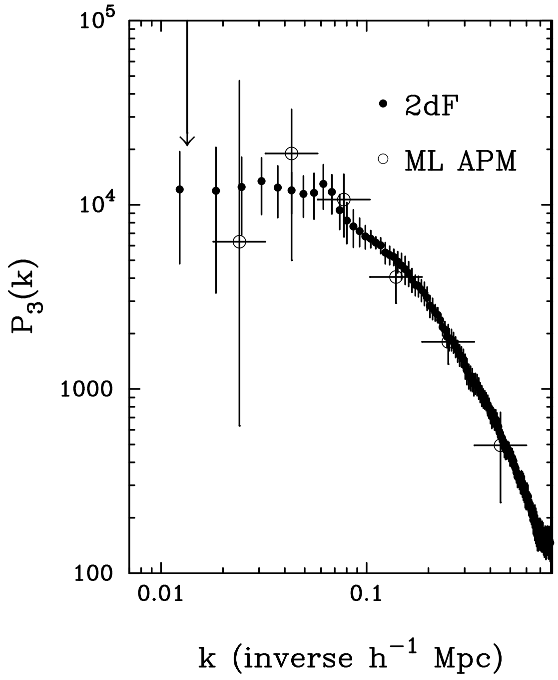

Figure 14: Maximum likelihood bandpower estimates of $P(k)$ from the 2-d APM survey (open circles and errors) as computed by Efstathiou and Moody compared with the redshift space power spectrum of the 2dF SGP region (filled circles).

- If we are going to fit restricted ranges of $P(k)$ to theoretical models, we *must* investigate the sensitivity of the shape of $P(k)$ to the sample, weighting scheme and completeness correction. For example, how do the results depend on the fit to $dN(z)$, inclusion of random fields, variations between north and south etc.

- We should *not* perform a joint analysis of north and south. Since the photometry in the north is suspect, and is not matched with the photometry in the south, we should analyse north and south regions *independently* and compare the results. If one fits $dN(z)$ for the two samples independently one is filtering out cross-correlations anyway.

- I am now going to investigate a straightforward adaptation of the maximum likelihood method that I applied to the 2 dimensional parent APM catalogue. It should be possible to get some interesting results using a modification of the existing code. However, I will need to spend some time implementing new methods to apply maximum likelihood methods at high spatial resolution. This, incidentally, will automatically solve for β (or β as a function of scale).