Mon. Not. R. Astron. Soc. **348,** 737–744 (2004) doi:10.1111/j.1365-2966.2004.07323.x

Fluctuation analysis of the far-infrared background – information from the confusion

Yasmin Friedmann[1,2⋆] and François Bouchet[3]

[1]*Racah Institute, The Hebrew University, Jerusalem 91904, Israel*
[2]*Departement de Physique Theorique, University of Geneva, 24 Quai E. Ansermet, 1211 Geneve 4, Switzerland*
[3]*IAP, 98bis Bd Arago, 75014 Paris, France*

Accepted 2003 October 21. Received 2003 October 9; in original form 2003 July 16

ABSTRACT

We investigate the extent to which one can use a $P(D)$ analysis to extract number counts of unclustered sources from maps of the far-infrared background. Such maps currently available, and those expected to emerge in the near future, are dominated by confusion noise resulting from poor resolution. We simulate background maps with an underlying two-slope model for $N(S)$, and we find that, in an experiment of FIRBACK type, we can extract the high-flux slope with an error of a few per cent, whereas other parameters are not so well constrained. However, we find that, in a *SIRTF*-type experiment, all parameters of this $N(S)$ model can be extracted with errors of only few per cent.

Key words: methods: data analysis – infrared: galaxies.

1 INTRODUCTION

Analysis of spatial fluctuations in the level of background radiation has been used by radio (Scheuer 1974; Condon 1974) and X-ray astronomers (Barcons & Fabian 1990) to gain information on the number count–flux relation below the detection limit. Recently (Lagache et al. 2000), fluctuations in the infrared background were first detected in maps of the FIRBACK survey at a wavelength of 170 μm.

The spatial resolution currently available in the far-infrared is of the order of arcminutes. As a result, observations at this wavelength are confusion-limited. This means that the dominant contribution to noise on sky maps at these wavelengths comes not from detector or photon noise but from the superposition of light originating from galaxies that are too close on the sky to be resolved individually. It has been shown (Puget et al. 1999) that the energy coming from resolved sources on the FIRBACK maps comprises only 10 per cent of the total energy while the rest is due to the unresolved background radiation. This means that other than fluctuation analysis ['$P(D)$ analysis'], not much else can be done to study the $N(S)$ of the unresolved infrared sources in the far-infrared.

This study investigates the conditions under which $P(D)$ analysis can usefully constrain galaxy evolution scenarios. Guiderdoni et al. (1998) have introduced a semi-analytical model of galaxy formation and evolution, and within this model suggested several scenarios including different amounts of ultraluminous infrared galaxies. For each of the scenarios, they calculated, among other things, faint galaxy counts. They show that at 175 μm the source counts at fluxes 10–100 mJy are quite sensitive to the details of the galaxy evolution; therefore, knowing $N(S)$ to high precision can help in choosing between the different scenarios of their galaxy evolution models. Similarly Takeuchi et al. (2001) show that at 175 μm the number counts at fluxes 10–100 mJy are very dependent on the galaxy evolution models that they propose.

When using a simple power-law parametrization for the source counts, errors of a few per cent in the parameters can give an error of 50 per cent in the number counts at fluxes of a few tens of mJy, i.e. one anticipates that such counts at these flux levels will easily constrain the parameters. Additionally, at least in the above-mentioned families of models in this flux range, the $N(S)$ due to the different models differ by an order of magnitude. This justifies attempting to measure the $N(S)$ parameters to high precision down to a few tens of mJy. Note that the flux range of a few tens of mJy is far below the detection limit of FIRBACK (180 mJy) and therefore, at the moment, can be probed only via $P(D)$ analysis.

Additionally one would also like to know how much information one can gain about the number counts from specific confusion-limited surveys like FIRBACK or *SIRTF* using a $P(D)$ analysis. This kind of study can be done using a Fisher matrix analysis where one calculates the minimal errors of extracted parameters given from the experiment and a parametrized theoretical model.

The analysis we carry out here does not take into account clustering of sources. Some clustering of the far-infrared sources is of course expected (Scott & White 1999; Haiman & Knox 2000; Knox et al. 2001), but its amplitude is not yet determined at 175 μm. The small area of the FIRBACK fields might not enable one to constrain the source clustering accurately, but this situation might change with the *SIRTF* observations, which cover larger areas of the sky (Dole, Lagache & Puget 2003). The resolved sources on the FIRBACK maps show a level of clustering consistent with zero. This is probably due to the small number of resolved sources (Guiderdoni and

⋆E-mail: yasmin@skopelos.unige.ch

Lagache, private communication). Hence we are assuming, at this stage, that sources are Poissonian distributed on the sky.

We find that we can constrain the slope of the number counts of sources with high fluxes ($\geqslant$20 mJy) at least as well as has been done by extracting individual strong sources. Other parameters (slope of the number counts at low fluxes, normalization and break flux) are not as well constrained in the FIRBACK type of experiment. However, in an experiment with smaller pixels and more of them (e.g. *SIRTF*), we can extract all the parameters to within several per cent. We also found some degeneracies between the different parameters and saw that a better experiment like *SIRTF* cannot resolve these degeneracies.

The paper is arranged as follows. We give an explanation of the nature of confusion in sky surveys in Section 2. In Section 3 we describe how we model the $N(S)$ in order to extract it from the data. In Section 4 we outline the method of analysing sky maps in order to extract the parameters of the model. We describe the implementation of the method and the results of analysing simulated skies in Section 5.

2 CONFUSION

The spatial fluctuations in the level of background radiation due to the spatial distribution of the discrete sources that contribute to the background are called 'confusion noise'. At far-infrared wavelengths the level of the confusion noise dominates over any photon or instrumental detector noise existing in today's instruments. Thus, while one can reduce the level of instrumental or photon noise by long integration times, the confusion remains a strong characteristic of far-infrared observations.

The full calculation of the probability distribution of the fluctuations [$P(D)$ in short] is classical (see Condon 1974; Barcons & Fabian 1990) and given in the Appendix. Here we only give the final result:

$$P(D) = \int_{-\infty}^{\infty} \Phi(\omega)\, \mathrm{e}^{-2\pi \mathrm{i}\omega D}\, \mathrm{d}\omega, \tag{1}$$

where D is the deflection from the mean level of flux and

$$\Phi(\omega) = \exp\left[A_{\mathrm{pix}} \int_0^{\infty} \mathrm{d}S \frac{\mathrm{d}N(S)}{\mathrm{d}S} \sum_i (\mathrm{e}^{2\pi \mathrm{i}\omega P_i S} - 1) \right]. \tag{2}$$

The shape of $P(D)$ depends on several inputs. First is the differential $N(S)$ relation, $\mathrm{d}N(S)/\mathrm{d}S$. This is the number of sources per steradian with fluxes in $[S, S + \mathrm{d}S]$. It also depends on the shape of the beam, which is described by P_i, the point spread function (PSF) of the telescope. Finally in (2) A_{pix} is the pixel size. In general $P(D)$ will also depend on clustering of the sources if it is strong enough (Barcons 1992), but, as mentioned before, we are going to deal with a source distribution that is not clustered but Poissonian.

It was shown in Scheuer (1974) that the width of the curve (the 1σ of the noise) is of the order of the flux for which there is one source per beam. The very faint sources do not contribute at all to the shape of the curve, but only to the mean level of the flux. This is because there are very many of them within each beam and the change of their number from beam to beam is relatively small and so does not contribute to the fluctuations. The very strong sources contribute only to the tail of the distribution. Typically the flux where there is one source per beam is much lower than the resolution limit.

3 LOG N – LOG S OF INFRARED SOURCES

Since this work is motivated by the FIRBACK survey, we will give in the following a short description of the survey and of the $\mathrm{d}N(S)/\mathrm{d}S$ found for the resolved sources. These details will guide us when we construct simulations to check our method of deriving $\mathrm{d}N(S)/\mathrm{d}S$ from the observed $P(D)$.

3.1 The FIRBACK survey

FIRBACK (Puget et al. 1999; Lagache & Dole 2001; Dole et al. 2001) is a deep survey of 4 deg^2 of the sky at 170 μm. The 4 deg^2 were chosen in such a way that the foreground cirrus contamination was as small as possible. Thus one can get information on the extragalactic radiation (Lagache & Puget 2000). In the FIRBACK survey there were 106 sources detected above the sensitivity limit of the experiment at 4σ, with fluxes between 180 and 2400 mJy. The slope of the log N – log S curve was measured by Dole et al. (2001) to be -3.3 ± 0.6 between 180 and 500 mJy.

3.2 Modelling source counts

In view of the above, we will assume a broken power law for the source-count model: the slope at low fluxes has to become shallower than 3.0 or the flux per pixel will diverge. Another motivation for this two-slope model is the predictions coming from galaxy evolution models discussed earlier. In all of the predictions, the number counts exhibit a relative flattening at low fluxes.

Therefore we write $\mathrm{d}N(S)/\mathrm{d}S$ as follows:

$$\frac{\mathrm{d}N(S)}{\mathrm{d}S} = \begin{cases} A_{\mathrm{norm}} S^{-\gamma_1} & \text{for } S \geqslant S_{\mathrm{break}}, \\ A_{\mathrm{norm}} S_{\mathrm{break}}^{\gamma_2-\gamma_1} S^{-\gamma_2} & \text{for } S_{\mathrm{min}} \leqslant S \leqslant S_{\mathrm{break}}. \end{cases} \tag{3}$$

The parameters to be determined are the normalization, A_{norm}, the flux of the break in the power law, S_{break}, and the two slopes, γ_1 and γ_2. Since S_{min} does not change the shape of the distribution and only affects the mean flux, and since we will be fitting for the shape of the distribution and not for the mean flux, S_{min} becomes irrelevant to the fitting process. It comes into play only in order to fine-tune the mean flux to its value from data. In the following we will refer to the four parameters commonly as θ, and the probability distribution of the deflection will be written as $P(D; \theta)$.

4 METHOD OF ANALYSIS

4.1 Minimum χ^2 method for binned data

Our data set is composed of several thousand measurements of incoming flux received by 46×46 arcsec^2 pixels that point to different directions in the sky. We will bin the data according to flux. In this way we can compare the experimental distribution of the fluctuations to a calculated $P(D; \theta)$.

Binning data may proceed in two ways: (i) one way is such that the bins are equal in length and the numbers of events vary from bin to bin; or (ii) one may bin the data such that there is the same number of events in each of the bins and the size of the bins changes accordingly. We use the former method, but we manually increase the bin size at the two tails of the distribution where there are very few events, so as not to have bins with zero events. We thus have three bins where there are around five events per bin, out of a total of 60 bins.

A histogram is in fact a multinomial distribution. This is the generalization of the binomial distribution to the case where it is possible to have more than two outcomes for the experiment. In our case the flux received by a pixel pointing in one of the directions in the sky will be one result of the experiment, and the outcome might

fall in any one of the bins. Let us call $r_1, \ldots, r_n$ the n possibilities for the outcome of the experiment (in our case n different flux ranges from the minimal to the maximal flux received in the map), and let p_i be the probability for a pixel to fall in bin i. The sum of all these probabilities is of course unity, since every pixel falls in one of the bins. We assume that the different pixels are independent (we will justify this assumption in Section 5.2). Then, after N experiments of measuring the incoming flux (N pixels in the map), the probability that the fluxes will distribute with $r_1, \ldots, r_n$ pixels falling in the bins will be given by

$$P(r_i, N, p_i) = \frac{N!}{r_1! r_2! \cdots r_n!} p_1^{r_1} p_2^{r_2} \cdots p_n^{r_n}. \quad (4)$$

Some important properties of this distribution are that $E(r_i)$, the expectation value for bin i, is given by $E(r_i) = Np_i$, and $V(r_i)$, the variance for bin i, is given by $V(r_i) = Np_i(1 - p_i)$. When the number of experiments becomes large, the multinomial distribution tends to the multinormal distribution. In the case when there are many bins and $p_i \ll 1$, the variance tends to be the expectation value. Thus when we bin the fluxes we must take care to have enough bins such that, on the one hand, $p_i \ll 1$ and, on the other, there are enough pixels per bin so that the distribution of the number of pixels per bin will be close to Gaussian. Then the overall likelihood of the data can be written as follows:

$$L(r_1, \ldots, r_n; \theta) = C \mathrm{e}^{-Q^2/2}, \quad (5)$$

where Q^2, the quadratic form, is

$$Q^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} \left(\frac{r_i - \mu_i}{\sigma_i} \right) \left(\frac{r_j - \mu_j}{\sigma_j} \right), \quad (6)$$

c_{ij} are the elements of the inverse covariance matrix C, given by

$$C = E[(\boldsymbol{r} - \boldsymbol{\mu})(\boldsymbol{r} - \boldsymbol{\mu})^{\mathrm{T}}], \quad (7)$$

r_i is the number of pixels within flux bin i, μ_i is the expected number of pixels in the bin according to the model, and σ_i is the square root of the variance, in our case $\sqrt{\mu_i}$. The correlation $\rho_{i,j}$ between bins i and j is given by

$$\rho_{i,j} = C_{i,j}/\sigma_i \sigma_j. \quad (8)$$

If we assume that there are no or only negligible correlations between the errors in the bins, then the covariance matrix will be almost diagonal. Therefore the quantity we need to minimize becomes

$$Q^2 = \sum_{i=1}^{N_{\mathrm{bins}}} \left(\frac{r_i - \mu_i}{\sqrt{\mu_i}} \right)^2. \quad (9)$$

This method is the 'minimum chi-square method' applied to histogram fits. When there are many events in each bin, then Q^2 has asymptotically a χ^2 distribution with [n − (number of fitted parameters)] degrees of freedom and the method is equivalent to a maximum likelihood estimation method. Therefore, from now on we will use the notation $\chi^2 = Q^2$. We are assuming that our way of binning allows Q^2 to behave close enough to χ^2.

4.2 Fisher matrix analysis

As described before, we have several thousand measurements of the flux S, and a model for $N(S)$ that leads to a $P(D; \theta)$ and we estimate the parameters θ using the maximum likelihood method with the binned data. There is a lower bound to the variance of an estimator, which is related to the Fisher matrix, F_{ij},

$$F_{ij} = E \left[-\frac{\partial \log L}{\partial \theta_i \partial \theta_j} \right]. \quad (10)$$

The Rao–Cramer–Frechet inequality states that, for any unbiased estimator, $\Delta\theta_i \geqslant (F^{-1})_{ii}^{1/2}$, where $\Delta\theta_i$ is the 1σ error of the parameter θ_i. This inequality was used by several authors (e.g. Jungman et al. 1996; Tegmark, Taylor & Heavens 1997; Efstathiou 1999) to assess how well different parameters may be estimated in future experiments.

In our case, where the the likelihood is multinormal, and the errors depend on the parameters, the Fisher matrix becomes (Tegmark et al. 1997):

$$F_{ij} = \frac{1}{2} \sum_{k=1}^{N_{\mathrm{bins}}} \frac{\partial \mu_k}{\partial \theta_i} \frac{\partial \mu_k}{\partial \theta_j} \frac{1}{\mu_k} \left(2 + \frac{1}{\mu_k} \right) \Bigg|_{\theta_{\mathrm{ML}}}. \quad (11)$$

As we can see, the Fisher matrix depends solely on the model and on the estimated parameters, $\{\theta_{\mathrm{ML}}\}$. We can now calculate the minimal variance for different experimental setups, namely as a function of the number of pixels in the maps, or as a function of the size of the pixel. In this way we can foresee how well we could extract the four parameters from future experiments.

The inverse of the Fisher matrix is an estimate of the covariance matrix, so by investigating F^{-1} one can locate degeneracies between parameters and see if they might be resolved in different experimental setups.

5 RESULTS

In this section we describe the analysis of simulated skies that we have made in order to check theoretically the limits of our method using the Fisher matrix analysis.

5.1 Simulations

We use mock images to check our algorithm. These images are built in two steps. First we build a mock sky. This is a high-resolution projected image of a given spatial distribution of sources. It is constructed by distributing point sources randomly on a 1.25×1.25 deg^2 field. Each source is assigned a flux such that the overall number counts are consistent with our predetermined two-slope $N(S)$. We chose to model $N(S)$ such that the main traits of the FIRBACK maps are realized, such as the number of sources above 180 mJy and the mean level of the background. Then we convolve it with the instrumental effects in order to produce an image as close as possible to real data. Finally we extract only the central 1 deg^2 as our map. A map produced in this way is presented in Fig. 1(a), and its histogram is shown in Fig. 1(b). The map includes some 2×10^8 sources with fluxes ranging from 0.01 to 2000 mJy, and the parameters chosen were $\gamma_1 = 3.3$, $\gamma_2 = 1.8$ and $S_{\mathrm{break}} = 14$ mJy and the normalization was 0.18×10^{11}.

5.2 Effect of different PSFs on correlation between flux bins

The total extent of the PSF of the FIRBACK instrument and of the upcoming *SIRTF* is bigger than the size of a pixel in these experiments. In FIRBACK the full width at half-maximum (FWHM) of the PSF is equal to the pixel size. For *SIRTF* it is twice as wide (Dole et al. 2003). In such cases we expect there to be some correlation between adjacent pixels on the map. Given this, we want to

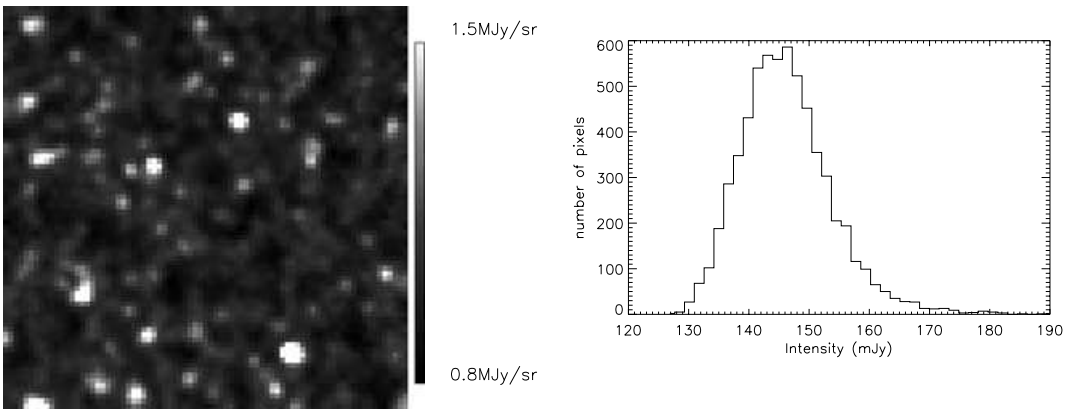

Figure 1. Left: (a) 1 deg^2 of simulated sky, with pixels of 46 × 46 arcsec^2. The sky contains some 2×10^8 sources, described by a model $N(S)$ given in equation (3), with parameters as follows: $\gamma_1 = 3.3$, $\gamma_2 = 1.8$, $S_{\rm break} = 14$ mJy and $A_{\rm norm} = 0.18 \times 10^{11}$. Right: (b) The histogram of the simulated map showing the fluctuations in the intensity. Note the skewed shape, with the tail due to the strong sources. This figure is available in colour in the on-line version of the journal on *Synergy*.

quantify the extent to which these correlations may be manifested as correlations between the different bins in the histogram.

To this end we should measure the correlation coefficients of the histogram given different sized PSFs. The different PSFs we used were constructed from the original one by rebinning again and again. We quantify their extent with respect to the pixel size by defining variable x, which is the ratio of the FWHM to the pixel size. The x we used are 2.7, 1.1, 0.5, 0.3 and 0.1. Following this we produced hundreds of sky maps based on the same underlying $N(S)$ relation. We convolved each with a PSF, produced a histogram, and measured the sample correlation matrix. The process was repeated for the different PSFs, and was done once without convolving with any PSF but instead we rebinned the sky to the size of the pixel, 46 × 46 arcsec^2.

If bins were not correlated at all, the correlation matrix would have been the identity matrix. We found that there is a negligible level of correlations with any of the PSFs, mostly less than 10 per cent. The largest correlation seen was between a few neighbouring bins for the largest PSF, near the centre of the histogram, at a level of 20 per cent. There is also no clear behaviour of ρ_{ij} as a function of the total coverage area of the PSF. This can be seen in Fig. 2, where three randomly selected elements of the correlation matrix are plotted as a function of the total coverage area of the PSF. In light of this we will neglect the correlations between the bins and estimate the parameters as described in the next section.

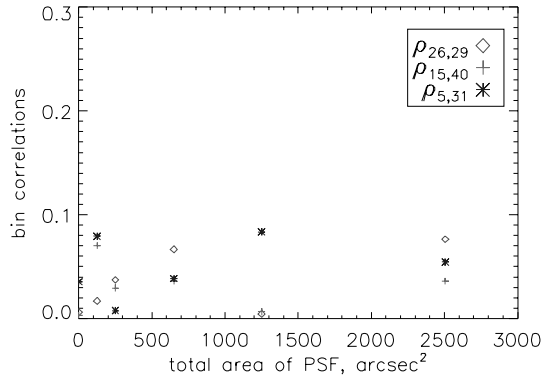

Figure 2. Three randomly chosen elements of the correlation matrix, plotted as a function of the total coverage area of the PSF. All correlations are less than 10 per cent, and there is no clear trend of behaviour as a function of the coverage.

5.3 Estimating parameters

In order to find estimates for the four parameters $A_{\rm norm}$, $S_{\rm break}$, γ_1 and γ_2, we first choose a small enough value for $S_{\rm min}$. On the one hand, this should be smaller than reasonable values of $S_{\rm break}$, and, on the other, it should be big enough to avoid numerical problems of integration. We then grid a large enough part of parameter space around a reasonable point found by trial and error. Each of these grid points will serve as a starting point for the minimization procedure. In this way we have more chance of catching the lowest minimum. The minimization procedure calculates the χ^2 and uses the derivatives of the model to go downhill in parameter space until a lowest local χ^2 is found. We looked at the estimated parameters and their errors to see whether they are all situated in the same part of the parameter space. We repeated the procedure with several different binnings to make sure that the results are not dependent on the binning. The best fit is shown in Fig. 3. The reduced χ^2 for this model is 0.96. The errors of the estimated parameters will be discussed below.

Once we have a best-fitting set of parameters, we assume that it is a good enough approximation of the real parameters. Then we can continue to calculate the Fisher matrix elements for different

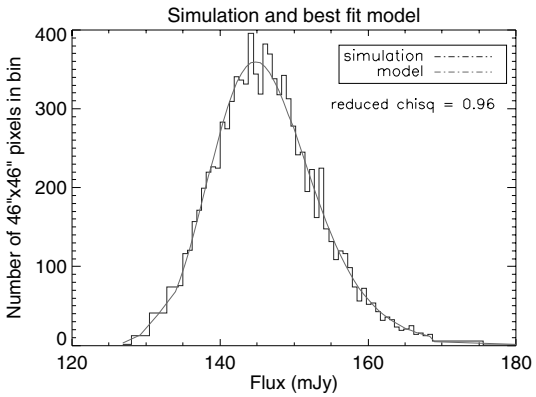

Figure 3. The histogram of the simulated map with the best-fitting model. This figure is available in colour in the on-line version of the journal on *Synergy*.

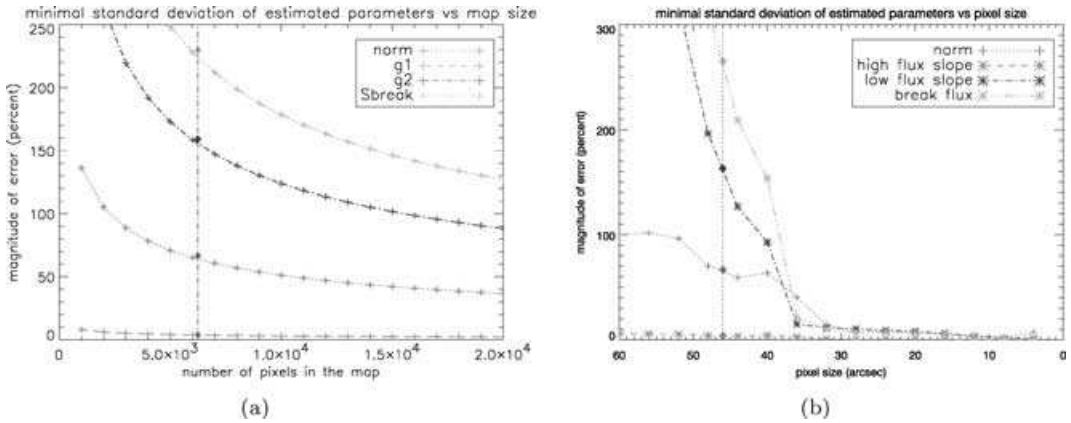

Figure 4. The errors on the estimated parameters (in per cent) as a function of (a) the number of pixels in the map and (b) the pixel size. The vertical line shows where the FIRBACK experiment is on these graphs. This figure is available in colour in the on-line version of the journal on *Synergy*.

experimental parameters. We will calculate the minimal errors of the different parameters as a function of pixel size and as a function of the number of pixels in the map. In Fig. 4(a) one sees how the errors reduce when we add more pixels to the map. In fact they reduce as (number of pixels)$^{-1}$ – and this behaviour is what we expect from the definition of the Fisher matrix, equation (11). In Fig. 4(b) one sees that the errors greatly reduce once we use smaller pixel sizes in the experiment, with all errors at the level of only a few per cent once we reach a pixel size of 15 arcsec^2. This is very encouraging because in the upcoming *SIRTF* experiment the pixels will be of size 16×16 arcsec^2.

Another point to look at is the comparison between the errors on the estimated parameters in our algorithm, and the minimal errors given from the Fisher matrix analysis. To this end we plotted, in both Figs 4(a) and (b), a vertical line indicating the properties of the FIRBACK and our simulations. On top of this line we put the parameter errors we got from the minimization procedure. It is encouraging to see that these errors are only very slightly larger than the minimal errors possible according to the Cramer–Rao–Frechet lower bound.

Another way to see the great improvement in the precision of the estimation is when one looks at the 1σ contours, which enclose 68 per cent of the joint distribution of several pairs of estimators, while we marginalize over the other two parameters, as in Fig. 5. The contours, independent of which plane we look at, encompass a shrinking patch of the parameter space as we add pixels to the map or use smaller pixels. We have plotted the original parameters as filled hexagons on top of the ellipses. It is important to note also that, except for the normalization, the true parameters of the simulation are enclosed within the 1σ contours of the original experiment.

5.4 Parameter degeneracies

Fig. 5 also shows us that there are degeneracies between the different parameters. These degeneracies are not broken when we use a more accurate experiment; they are 'built in' through the definition of the model $N(S)$. For example, as the model is defined, the normalization is the number of sources with fluxes greater than 1 mJy. Naturally, if we increase γ_1 we should increase the normalization in order to remain within the error bars for $N(S)$. In the following we will look further into the degeneracies by looking at the derivatives of the model with respect to the parameters in which we are interested (Tegmark et al. 1997). The reason for this can be seen if we look again at the expression for the elements of the Fisher matrix (equation 11): they have the structure of a dot product between vectors μ_k/θ_i and μ_k/θ_j. If one of these vectors is a linear combination of another, the Fisher matrix elements will be singular, and the errors of the estimated parameters will be infinite. If the vectors are completely orthogonal, then F and F^{-1} will be diagonal and thus there will be no correlation between different parameters and their errors. Usually there will be some level of correlation, which will be manifested by a somewhat similar shape between the functional

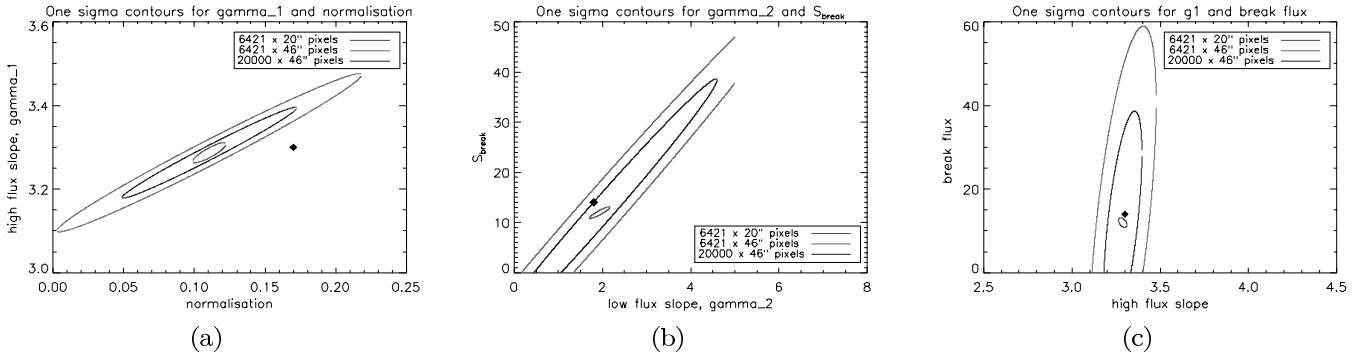

Figure 5. The 1σ contours in the two-dimensional plane of three different pairs of parameters, while marginalizing over the rest. These contours are calculated according to F^{-1}. The filled hexagon gives the original parameters of the simulation on these planes. This figure is available in colour in the on-line version of the journal on *Synergy*.

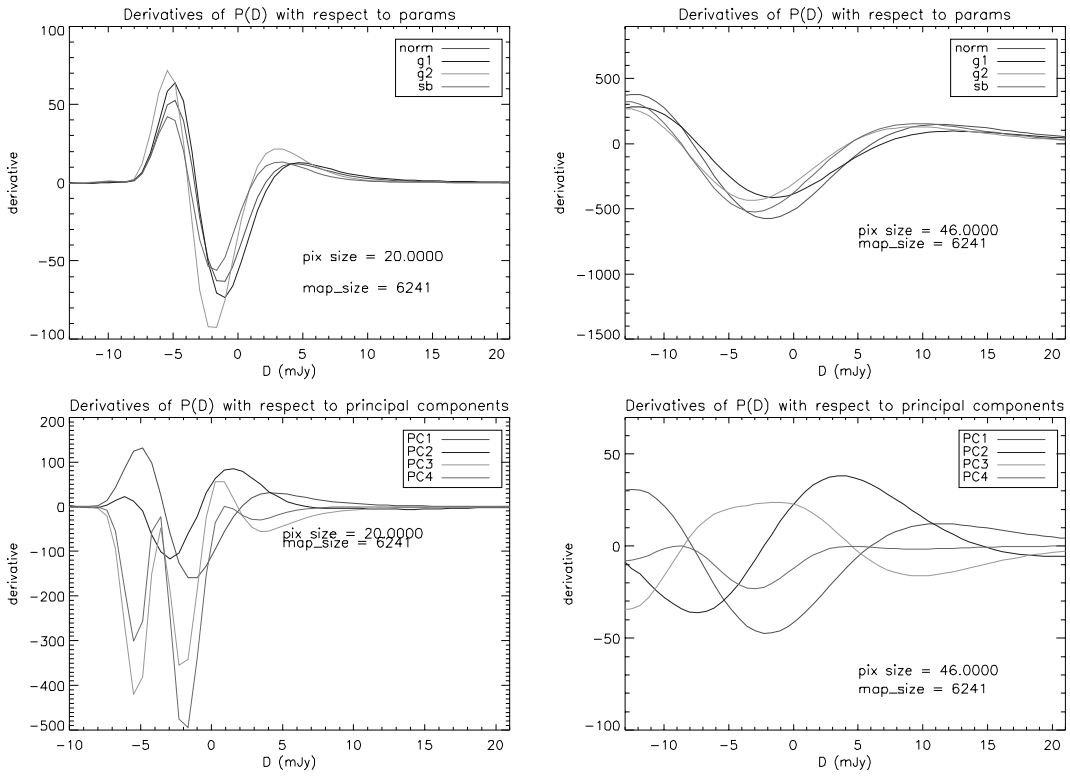

Figure 6. Top: The derivatives of the model with respect to the four parameters at two pixel sizes, 46×46 arcsec^2 and 20×20 arcsec^2. Bottom: The derivatives of the model with respect to the principal components. It is possible to see how the degeneracies disappear. This figure is available in colour in the on-line version of the journal on *Synergy*.

shape of the different derivatives. By looking at the functional shape of the derivatives, one may recognize such degeneracies. The way to avoid any degeneracies between parameters will be to diagonalize the Fisher matrix, thus changing the parameters into new ones that are linear combinations of the originals. These parameters, however, are not always physical, and thus do not have much meaning unless they are very similar to the old ones, with not much 'contamination' from other parameters.

It is worth while to check whether or not, under improved experimental conditions, these degeneracies, if they exist, are removed. In Fig. 6 we plot the derivatives with respect to the usual four parameters and the derivatives with respect to the new, 'diagonalized' parameters (called principal components, PC1–4). On the left panel we have the case for a pixel of 20×20 arcsec^2, and on the right, 46×46 arcsec^2. Again we can see that the degeneracies are not removed by improving the experiment. Only diagonalizing the Fisher matrix allows us to remove them. We also add a table that specifies how the new parameters are built from the old ones. In the first row we have the four parameters and in the first column we have the four principal components. In each of the four rows we have the coefficients of the linear combination. The strongest coefficients are marked in italics.

$$\begin{array}{c} \\ \text{PC1} \\ \text{PC2} \\ \text{PC3} \\ \text{PC4} \end{array} \begin{array}{c} \begin{array}{cccc} K & \gamma_1 & \gamma_2 & S_{\text{break}} \end{array} \\ \begin{pmatrix} 0.32 & \textit{0.91} & 0.24 & -0.03 \\ \textit{0.94} & -0.34 & 0.03 & -0.004 \\ -0.11 & -0.21 & \textit{0.96} & -0.09 \\ 0.0015 & 0.002 & 0.098 & \textit{0.995} \end{pmatrix} \end{array}$$

PC1 has the highest weight from γ_1 with some contribution from K and γ_2. PC2 has the highest contribution from K and some contribution from γ_1. PC3 is mostly γ_2 and some γ_1, and PC4 is almost exclusively S_{break}. The degeneracies are to be expected since, as mentioned before, $P(D)$ analysis cannot give information at fluxes much below the one source per beam flux level. In our simulation this level is of the order of 7 mJy, and therefore it is reasonable that for example γ_2, which is the slope of the counts below a few mJy, is degenerate with the other parameters.[1]

6 DISCUSSION

In this work we have explored the extent to which one can use a $P(D)$ analysis to gather information from far-infrared sky maps. These maps are characterized by a very high level of confusion noise, which arises as a result of the relatively poor resolution power available at these wavelengths.

We created a simulated map of the sky with an underlying modelled $N(S)$. The model consisted of a two-slope model with a high-flux slope greater than 3 and a shallower low-flux slope. It was chosen this way following the finding of a steep slope of the number counts of resolved objects in the FIRBACK maps and in agreement with predictions of galaxy evolution models. The parameters of the model are the two slopes, the break flux (where the slope changes) and the normalization.

We then created the histogram of the simulated map and used it to find the best-fitting parameters of the $N(S)$ model and their

[1] We thank the referee for pointing out this cause for the degeneracies.

errors. After finding the best-fitted parameters, we used the Fisher matrix analysis to calculate the minimal errors possible of these parameters in experiments with different pixel sizes and in experiments with different total number of pixels, including those of the FIRBACK.

We found that our algorithm gives fitted parameters with almost the minimal errors possible theoretically. The underlying parameters of the simulated map were within 1σ of the best-fitted ones (except for the normalization). This means that the tool we have constructed in order to find number counts is quite reliable.

The Fisher matrix analysis shows that in an experiment with pixel size of the order of 10 arcsec we will be able to find all parameters with errors of only about a few per cent. The situation of the FIRBACK experiment is quite different: it is only the high-flux slope that can be found with a small error bar of $\approx$4 per cent. This is somewhat better than what was found by individual source extraction from the maps (error of $\approx$20 per cent).

The advantage of the $P(D)$ analysis that it is sensitive down to the flux for which there is one source per beam – in the FIRBACK case this is around a few mJy, much below the detection limit, which is 180 mJy. Also, the extraction of even the high-flux slope is straightforward and does not warrant an a priori extraction of sources or other manipulation of the maps.

In order to be able to extract the other parameters of the number counts, we will have to wait for the *SIRTF* experiment. In that experiment, the pixel size is 16×16 arcsec^2 and the number of different points measured on the sky is up to an order of magnitude larger than for FIRBACK. In this case the precision of estimation is enhanced due to both factors: smaller pixels and more of them.

ACKNOWLEDGMENTS

We would like to thank Jeremy Blaizot for coding the instrumental effects of ISOPHOT and for many discussions. YF thanks Michel Fioc, Stephane Colombi, Bruno Guiderdoni, Guilaine Lagache, Herve Dole, Ofer Lahav, Roberto Trotta, Jean-Pierre Eckmann, Andreas Malaspinas, Tsvi Piran, Avishai Dekel and Yehuda Hoffman for helpful discussions. YF was funded by a Chateaubriand fellowship, a Marie Curie grant and a Marie Heim-Voegtlin grant.

REFERENCES

Barcons X., 1992, ApJ, 396, 460
Barcons X., Fabian A. C., 1990, MNRAS, 243, 366
Condon J. J., 1974, ApJ, 188, 279
Dole H. et al., 2001, A&A, 372, 364
Dole H., Lagache G., Puget J.-L., 2003, ApJ, 585, 617
Efstathiou G., 1999, MNRAS, 310, 842
Guiderdoni B., Hivon E., Bouchet F. R., Maffei B., 1998, MNRAS, 295, 877
Haiman Z., Knox L., 2000, ApJ, 530, 124
Jungman G., Kamionkowski M., Kosowsky A., Spergel D. N., 1996, Phys. Rev. Lett., 76, 1007
Knox L., Cooray A., Eisenstein D., Haiman Z., 2001, ApJ, 550, 7
Lagache G., Dole H., 2001, A&A, 372, 702
Lagache G., Puget J. L., 2000, A&A, 355, 17
Lagache G. et al., 2000, in Lemke D., Stickel M., Wilke K., eds, Lect. Notes Phys. Vol. 548, Proc. Ringberg Workshop, ISO Survey of a Dusty Universe. Springer, Berlin, p. 81
Puget J. L. et al., 1999, A&A, 345, 29
Scheuer P. A. G., 1974, MNRAS, 166, 329
Scott D., White M., 1999, A&A, 346, 1
Takeuchi T. T., Ishii T. T., Hirashita H., Yoshikawa K., Matsuhara H., Kawara K., Okuda H., 2001, PASJ, 53, 37
Tegmark M., Taylor A. N., Heavens A. F., 1997, ApJ, 480, 22

APPENDIX A: PROBABILITY DISTRIBUTION OF BACKGROUND FLUCTUATIONS

The shape of the curve describing the probability that a pixel will accumulate a certain level of flux depends on a few factors. These are the point spread function of the telescope, the pixel size and shape, and the source number counts $N(S)$.

We are assuming some properties for the sources that make up the background: the sources are point-like, they are not clustered, and once a source is 'found' its flux is a random variable distributed according to the number count–flux relation, $N(S)$.

The integrated flux seen by a pixel pointing in direction $\boldsymbol{n}$ will be

$$S_{\rm pix}(\boldsymbol{n}) = \int {\rm d}\Omega_{\boldsymbol{n}'} G(\boldsymbol{n}-\boldsymbol{n}')S(\boldsymbol{n}'), \tag{A1}$$

where $\mathrm{d}\Omega_n$ is the solid angle in direction $\boldsymbol{n}$; $G(\boldsymbol{n}-\boldsymbol{n}')$ is the beam profile, i.e. the convolution of the point spread function of the imaging instrument and the pixel profile; and $S(\boldsymbol{n})$ is the flux per unit solid angle coming from direction $\boldsymbol{n}$. Now we can calculate the probability density of $S_{\rm pix}$ (the probability that a certain pixel will receive flux $S_{\rm pix}$). In order to do that, one first calculates the characteristic function of $S_{\rm pix}$, which is given by

$$\Phi(\omega) = \left\langle {\rm e}^{2\pi {\rm i}\omega S_{\rm pix}} \right\rangle. \tag{A2}$$

Then the probability distribution for $S_{\rm pix}$ will simply be the Fourier transform of Φ.

Say there are K sources. Then the total flux received by a pixel should be written as the following discrete sum:

$$S_{\rm pix}(\boldsymbol{n}) = \sum_{i=1}^{k} S_i G(\boldsymbol{n}-\boldsymbol{n}_i). \tag{A3}$$

The probability of finding K sources in the sky, with fluxes $S_1, \ldots, S_k$ and in directions $\boldsymbol{n}_1, \ldots, \boldsymbol{n}_k$, is

$$F^{(k)}(\boldsymbol{n}_1, S_1, \ldots, \boldsymbol{n}_k, S_k) = \frac{(4\pi\mu)^k}{k!} {\rm e}^{-4\pi\mu} (4\pi)^{-k} \prod_{i=1}^{k} f(S_i), \tag{A4}$$

where μ is the mean number of sources per unit solid angle, and $f(S)$ is the probability that a source has a flux in the range $[S, S+\mathrm{d}S]$. This is a product of the Poissonian probability to find K sources while the mean number of sources is expected to be μ and the probability to find a source with flux S_i given the model $N(S)$, all this per steradian [hence the term $(4\pi)^{-k}$]. The normalization condition is that the sum of the probabilities to find any number of sources in any direction is unity:

$$\sum_{k=0}^{\infty} \int {\rm d}\Omega_{\boldsymbol{n}_1} \int_0^{\infty} {\rm d}S_1 \cdots \int {\rm d}\Omega_{\boldsymbol{n}_1} \int_0^{\infty} {\rm d}S_k F^{(k)}(\boldsymbol{n}_1, S_1; \ldots \boldsymbol{n}_k, S_k) = 1. \tag{A5}$$

Using the normalization condition and the form of $S_{\rm pix}$, we may now calculate the characteristic function, which becomes

$$\Phi(\omega) = \exp\left[A_{\rm pix} \int_0^{\infty} {\rm d}S \frac{{\rm d}N(S)}{{\rm d}S} \sum_i \left({\rm e}^{2\pi i \omega P_i S} - 1 \right) \right], \tag{A6}$$

where P_i is the pixelized PSF and $A_{\rm pix}$ is the area of the pixel in steradians. The integration over the angles is separate from that over

flux and gives $A_{\rm pix}$. This is justified in our case because, although the PSF is quite extended, outside of the pixel it has very small values. So now we may write the expression for $P(S_{\rm pix})$ as the Fourier transform of the characteristic function:

$$P(S_{\rm pix}) = \int_{-\infty}^{\infty} \Phi(\omega)\, \mathrm{e}^{-2\pi i \omega S_{\rm pix}}\, \mathrm{d}\omega. \tag{A7}$$

The above expression gives the probability that a pixel will receive an amount of flux equal to $S_{\rm pix}$. We will be working with a slightly different expression – the probability to get a certain flux above or below the mean flux on the map, D, defined as $S_{\rm pix} = D + \langle S_{\rm pix} \rangle$,

$$P(D) = \int_{-\infty}^{\infty} \Phi(\omega)\, \mathrm{e}^{-2\pi i \omega D}\, \mathrm{d}\omega. \tag{A8}$$

Once we have $P(D)$ we may integrate between the bins and multiply by the total number of pixels in the map to get the expected number of pixels falling within each bin.

This paper has been typeset from a TEX/LATEX file prepared by the author.