A MULTIFREQUENCY METHOD BASED ON THE MATCHED MULTIFILTER FOR THE DETECTION OF POINT SOURCES IN CMB MAPS

L.F. LANZ, D. HERRANZ, J.L. SANZ, M. LÓPEZ-CANIEGO
Instituto de Física de Cantabria (CSIC-UC),
Av. de Los Castros s/n, Santander, 39005, Spain

J. GONZÁLEZ-NUEVO
SISSA,
via Beirut 4, Trieste, I-34014, Italy

In this work we deal with the problem of simultaneous multifrequency detection of extragalactic point sources in maps of the Cosmic Microwave Background. We developed a linear filtering technique that takes into account the spatial and the cross-power spectrum information at the same time.

1 Introduction

A big effort has been devoted to the problem of detecting point sources in Cosmic Microwave Background (CMB) experiments. The main reasons are that the point sources contaminate the CMB radiation. It is therefore necessary to detect the maximum possible number of extragalactic point sources (EPS) and to estimate their flux with the lowest possible error. However, EPS are not just a contaminant that should be eliminated. They are a very important source of knowledge from the point of view of extragalactic astronomy.

The next generation of CMB experiments will allow one to obtain all-sky EPS catalogues that will fill in the existing observational gap in our knowledge of the Universe in the frequency range from 20 to roughly 1000 GHz. We expect to derive source number counts and spectral indices, to constrain evolutionary models, to study source variability and to discover rare objects such as inverted spectrum radio sources, extreme gigahertz peaked spectrum sources (GPS) and high-redshift dusty proto-spheroids (see for instance the *Planck Bluebook* [12]).

The detection and estimation of the flux of EPS are a difficult task. The main reason for this is that the many different types of EPS distributed in the sky form a very heterogeneous set of objects that do not have a common spectral behaviour. For this reason, in order to reduce the threshold detection level of point sources, we use multi-wavelength information: statistical information of the background and the spatial profile of the sources for both channels at the same time. We also take into account the spectral behaviour of the sources without making any a priori assumption (for a more detailed description of the method and the obtained results, see [7]).

2 Method

2.1 The single frequency approach

One of the standard single frequency point source detection methods in the literature is based on the *matched filter* [11,1,9]. The matched filter is the optimal linear detector for a single map in the sense that it gives the maximum signal to noise amplification. The matched filter can be expressed in Fourier space in the following way:

$$\psi_{MF}(q) = \frac{\tau(q)}{aP(q)}, \quad a = \int d\mathbf{q} \frac{\tau^2(q)}{P(q)}. \tag{1}$$

Here $\tau(q)$ and $P(q)$ are the Fourier transforms of the point source profile and the power spectrum, respectively; and a is a normalisation factor that preserves the source amplitude after filtering.

2.2 The Matched Multifilter

Let us assume a set of images corresponding to the same area of the sky observed simultaneously at N different frequencies:

$$y_\nu(\mathbf{x}) = f_\nu s_\nu(\mathbf{x}) + n_\nu(\mathbf{x}), \tag{2}$$

where $\nu = 1, \ldots, N$. At each frequency ν, y_ν is the total signal in the pixel $\mathbf{x}$ and s_ν represents the contribution of the point source to the total signal y_ν; for simplicity let us assume there is only one point source centred at the origin of the image; f_ν is the frequency dependence of the point source; and n_ν is the *background*. For simplicity, we assume Gaussian beams.

In the multi-frequency approach we take into account the statistical correlation of the noise between different frequency channels and the frequency dependence of the sources. Now let us model the background $n_\nu(\mathbf{x})$ as a homogeneous and isotropic random field with average value equal to zero and crosspower spectrum $P_{\nu_1\nu_2}$ defined by:

$$\langle n_{\nu_1}(\mathbf{q}) n^*_{\nu_2}(\mathbf{q}')\rangle = P_{\nu_1\nu_2}\delta^2_D(\mathbf{q}-\mathbf{q}'), \tag{3}$$

where $n_\nu(\mathbf{q})$ is the Fourier transform of $n_\nu(\mathbf{x})$ and δ^2_D is the 2D Dirac distribution. Let us define a set of N linear filters ψ_ν that are applied to the data

$$w_\nu(\mathbf{b}) = \int d\mathbf{x}\; y_\nu(\mathbf{x})\psi_\nu(\mathbf{x};\mathbf{b}) = \int d\mathbf{q}\; e^{-i\mathbf{q}\cdot\mathbf{b}} y_\nu(\mathbf{q})\psi_\nu(q). \tag{4}$$

Here $\mathbf{b}$ defines a translation. The right part of equation (4) shows the filtering in Fourier space, where $y_\nu(\mathbf{q})$ and $\psi_\nu(q)$ are the Fourier transforms of $y_\nu(\mathbf{x})$ and $\psi_\nu(\mathbf{x})$, respectively. The quantity $w_\nu(\mathbf{b})$ is the the filtered map ν at the position $\mathbf{b}$. The *total filtered map* is the sum

$$w(\mathbf{b}) = \sum_\nu w_\nu(\mathbf{b}). \tag{5}$$

Therefore, the total filtered field is the result of two steps: a) filtering and b) fusion. During the first step each map y_ν is filtered with a linear filter ψ_ν; during the second step the resulting filtered maps w_ν are combined so that the signal s is boosted while the noise tends to cancel out. Note that the combination in eq. (5) is completely general, since any summation coefficients different than one can be absorbed in the definition of the filters ψ_ν. Then the problem consists in how to find the filters ψ_ν so that the total filtered field is *optimal* for the detection of point sources.

The total filtered field w is *optimal* for the detection of the sources if

1. $w(\mathbf{0})$ is an *unbiased* estimator of the amplitude of the source, so $\langle w(\mathbf{0})\rangle = A$ (A is the amplitude of the point source);

2. the variance of $w(\mathbf{b})$ is minimum, that is, it is an *efficient* estimator of the amplitude of the source.

If the profiles τ_ν and the frequency dependence of the point source f_ν are known and if the crosspower spectrum is known or can be estimated from the data, the solution to the problem is already known: the matched multifilter (MMF) [6]:

$$\mathbf{\Psi}(q) = \alpha\; \mathbf{P}^{-1}\mathbf{F}, \quad \alpha^{-1} = \int d\mathbf{q}\; \mathbf{F}^t\mathbf{P}^{-1}\mathbf{F}, \tag{6}$$

where $\mathbf{\Psi}(q)$ is the column vector $\mathbf{\Psi}(q) = [\psi_\nu(q)]$, $\mathbf{F}$ is the column vector $\mathbf{F} = [f_\nu \tau_\nu]$ and $\mathbf{P}^{-1}$ is the inverse matrix of the cross-power spectrum $\mathbf{P}$. Finally, we can obtain the variance of the total filtered field, given by the following expression:

$$\sigma_w^2 = \int d\mathbf{q} \mathbf{\Psi}^t \mathbf{P} \mathbf{\Psi} = \alpha. \tag{7}$$

The frequency dependence f_ν is modelled in the following way:

$$I(\nu) = I_0 \left(\frac{\nu}{\nu_0} \right)^{-\gamma}, \tag{8}$$

where $I(\nu)$ is the flux at frequency ν, ν_0 is a frequency of reference, I_0 is the flux at that frequency of reference and γ is the *spectral index*. This equation is widely used in the literature. By using eq. (8), the reference flux I_0 can easily be related to the reference amplitude A of the sources and the number of degrees of freedom is just one, the spectral index γ.

When we have the different simulated maps with the point sources (see next section), these images are iteratively filtered with different MMFs (in fact, we modify γ, but MMF depends on γ). The value of γ that maximises the SNR for a given source is an *unbiased* estimator of the real value of the spectral index of the source. After that, results are compared with the MF.

3 Simulations

In order to illustrate the MMF method described above and to compare the MMF multi-frequency approach with the single frequency approach, we have performed a set of basic, yet realistic, simulations.

For this example we take the case of the *Planck* mission [10]. We will consider the 44 GHz and 100 GHz *Planck* channels.

For the simulations we have used the *Planck* Sky Model[a] (PSM; Delabrouille et al., in preparation), a flexible software package developed by *Planck* WG2 for making predictions, simulations and constrained realisations of the microwave sky. The simulated data used here are the same as in [8].

For the purely descriptive purposes of this example, we take eight different regions of the sky located at intermediate Galactic latitudes. For each region we select a 512×512 pixel flat and square patch (at 44 and 100 GHz). Pixel size is 1.72 arcmin for the two frequencies. Therefore, each patch covers an area of 14.656 square degrees of the sky. Once the region have been selected, we add simulated extragalactic point sources with a spectral behaviour described by eq. (8). We take as frequency of reference $\nu_0 = 44$ GHz. The antenna beam is also taken into account: the full width at half maximum is FWHM=24 arcmin for the 44 GHz channel and FWHM=9.5 arcmin at 100 GHz. Finally, we have added to each patch uniform white noise with the nominal levels specified for *Planck* [12] and this pixel size.

We are interested in comparing the performance of the multi-frequency approach with that of the single-frequency matched filter. In particular, we expect to be able to detect fainter sources with the MMF than with the MF. From recent works [9,8] we know that in this kind of *Planck* simulations, the MF can detect sources down to fluxes ~ 0.3 Jy (the particular value depends on the channel and the region of the sky). Here we will simulate sources in the interval $[0.1, 1.0]$ Jy plus a few cases, that will be described below, where even lower fluxes are necessary. Regarding the spectral index of the sources, according to [5], most radio galaxies observed by WMAP at fluxes ~ 1 Jy show spectral indices that lie in the range $(-1.0, 1.4)$.

We sample the interesting intervals of flux and spectral index by simulating sources with fluxes at 44 GHz $I_0 = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$ Jy and spectral indices $\gamma =$

[a] http://www.apc.univ-paris7.fr/APC_CS/Recherche/Adamis/PSM/psky-en.php

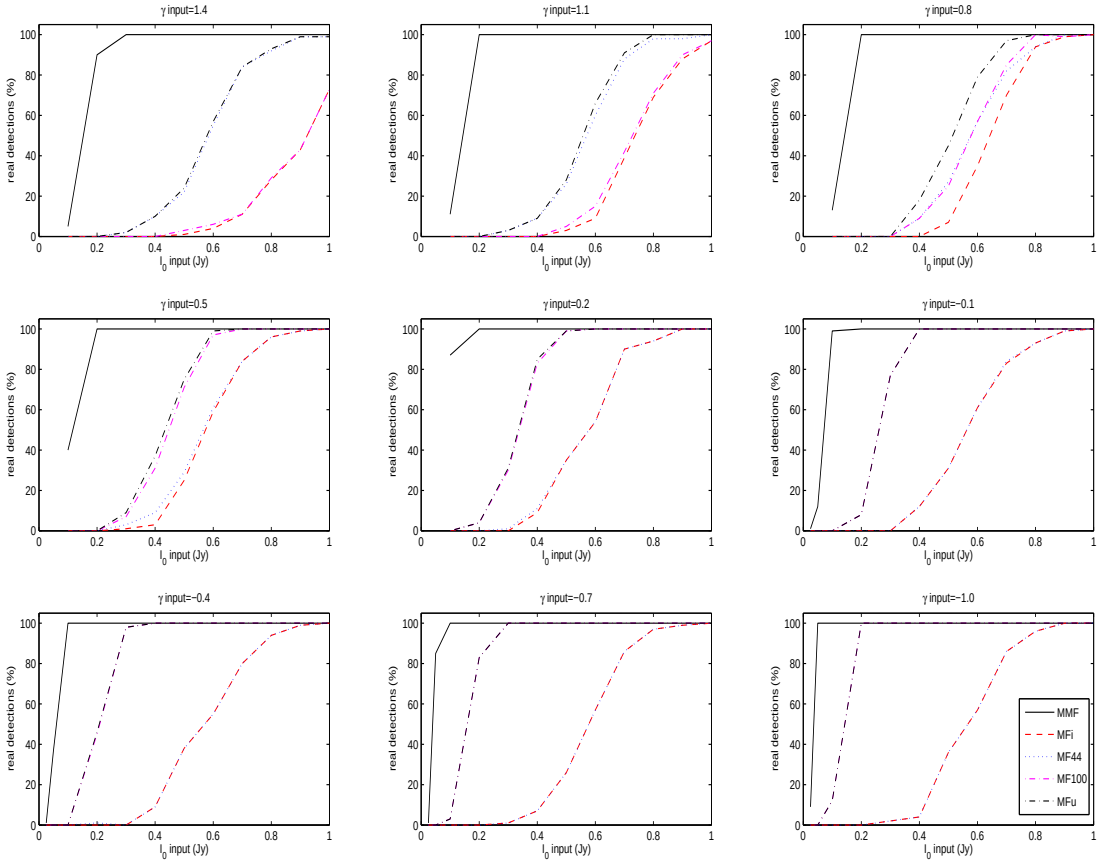

Figure 1: Number of detections against the input value of I_0 for different values of the spectral index γ. MF44 represents the sources detected with the matched filter at 44 GHz. MF100 the same but at 100 GHz. MFi is the intersection of MF44 and MF100, and MFu is the union of MF44 and MF100.

$\{-1.0, -0.7, -0.4, -0.1, 0.2, 0.5, 0.8, 1.1, 1.4\}$. For each pair of values (I_0, γ), we have simulated 100 point sources. The point sources are randomly distributed in the maps (the same source is placed in the same pixel in both frequencies), with only one constraint: it is forbidden to place a source closer than $FWHM_{44}/2$ pixels from any other. In this way we avoid source overlapping. Image borders are also avoided.

4 RESULTS AND DISCUSSION

We will compare the performance of the two methods in terms of the following aspects: spectral index estimation, source detection and flux estimation.

4.1 Source detection

Figure 1 shows the real sources (in %) that we detect above a 5σ level detection whose intrinsic fluxes (values introduced by us in the simulations) in the reference frequency (I_0) are the corresponding values in the horizontal axis. We can observe several interesting aspects. The first one is the fact that the matched multifilter improves the level of detection with respect to the matched filter level for all the values of γ we have inserted.

The second one is a natural selection effect. We detect more flat/inverted sources ($\gamma \sim 0$/ negative values of γ) at low fluxes than steep ones (positive values of γ). Keeping in mind that the reference frequency ν_0 is equal to 44 GHz, and according to eq. (8), it can be seen that for $\gamma > 0$, the simulated sources appear less bright at 100 GHz ($I_{100} < I_{44}$). In these conditions it is quite difficult to detect sources at 100 GHz, and for this reason we are not able to give the spectral indices for most of these point sources by means of the matched filter method when γ is strongly positive (for instance, $\gamma \gtrsim 1$). On the other hand, for the opposite reason, we have added two additional bins ($I_0 = 25, 50$ mJy) for the case $\gamma = -0.1$, because the matched filter seems to perform better at 100 GHz for low values of γ.

Additionally, we observe that the matched multifilter is capable to detect sources whose $I_0 < 0.1$ Jy for $\gamma \lesssim -0.1$. It is interesting to compare this with the matched filter which does not detect sources below 0.1 Jy in the conditions of this work. The method presented here, allows us to detect point sources whose I_0 is too low to be detected with the traditional matched filter.

To summarise, we can say that the MMF improves the detection level. Specially remarkable are the cases where the sources are near to be *flat* (central row of Figure 1). At 100 GHz, the MF recovers the 100% of the sources for $I_0 \sim 0.4 - 0.6$ Jy. Meanwhile, the MMF reaches this level for $I_0 \sim 0.1$ Jy. This particular case is really interesting because high frequency surveys show that most of the sources have this spectral behaviour.

4.2 Spectral index estimation

In Figure 2 we see how the spectral indices are recovered by means of the MMF and by the matched filter. In general, we observe that the MMF is able to recover the value of γ with more accuracy and less uncertainty than the traditional matched filter.

Another aspect is that the error bars increase when I_0 is smaller. It seems logical, because we have fainter sources and a smaller number of detections (see Figure 1). Then, at $I_0 = 0.1$ Jy, we can see that the estimation of γ is not as good as we wish, because it has a great uncertainty. The main reason is that the signal to noise ratio is close to the threshold level we have imposed.

In the case of the matched filter, the spectral indices are correctly estimated for $I_0 \gtrsim 0.7$ Jy at $\gamma \leq 0.8$. At higher values of γ we find the same problem that we have mentioned before: there are few detections at 100 GHz below 0.7 Jy (Figure 1). Since the detected sources are close to the noise level, the fluxes recovered present an overestimation with respect to the input value due to the Eddington bias [2].

Finally, we can observe an interesting aspect of the matched filter. When we do not have the sufficient detections in at least one channel (the sources detected are below the $\sim 40\%$ of the total number of sources), the estimation of the spectral index is not good. For $\gamma < 1.1$, the sources at 100 GHz are much fainter and the number of detections at this channel is really small. Then, because of the Eddington bias, the flux at this frequency is overestimated, and consequently, the value of γ is underestimated. The Eddington bias explains as well the overestimation of γ in the other cases. The only difference is that now it is at 44 GHz where we have a smaller number of detections. If we also see Figure 1, we observe that for values of $I_0 \lesssim 0.6$ Jy, we are pretty close to the noise level. It means that the noise fluctuations in the maps produce an overestimation in the flux at 44 GHz (I_0) and, as a consequence, an overestimation in γ too. In summary, for $\gamma = 1.4, 1.1$, the Eddington bias appears at 100 GHz (underestimation of the spectral index). For the rest of γ values, this bias appears at 44 GHz (overestimation of the spectral index).

4.3 Flux estimation

Figure 3 shows the recovered flux at the reference frequency (44 GHz) for a given value of the spectral index. The error bars recovered with the matched filter are, in general, larger than the

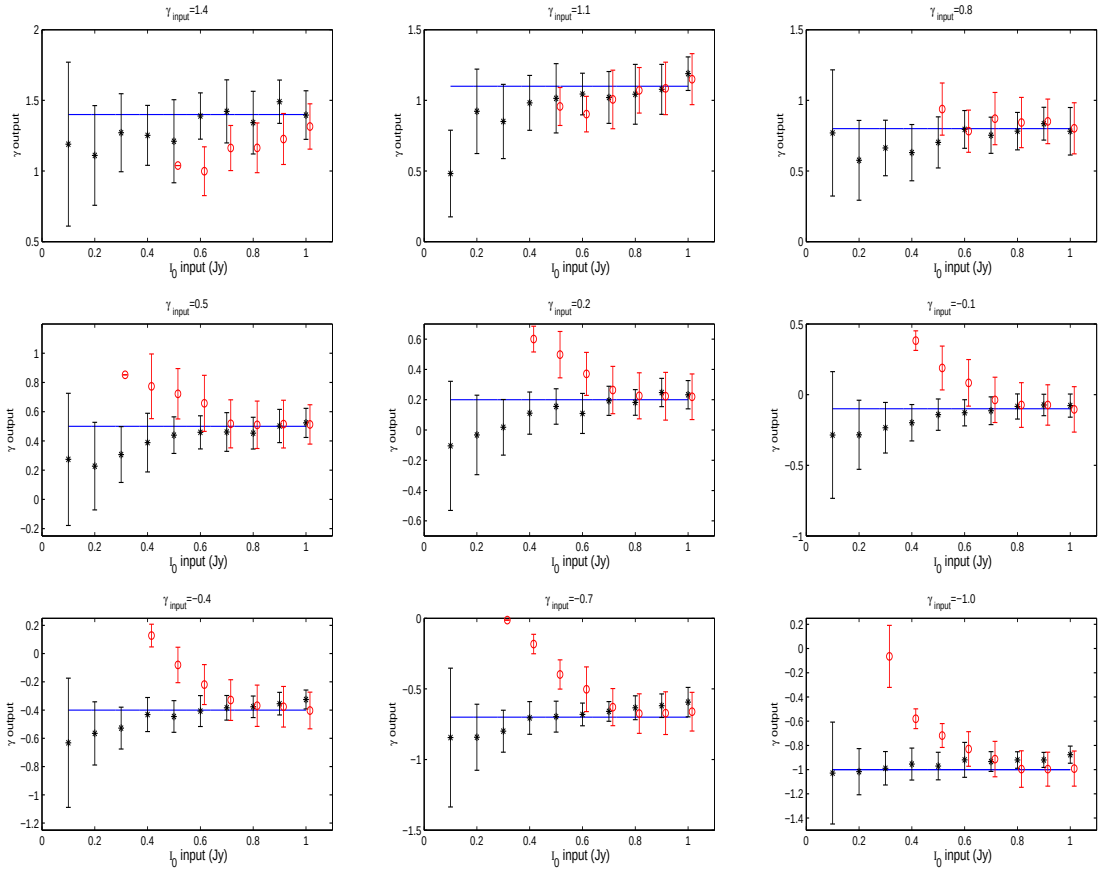

Figure 2: Values of γ recovered by means of the MMF (asterisks) and the MF (circles). The line indicates the ideal recovering of the input. The circles corresponding to the MF are slightly displaced in the horizontal axis in order to distinguish the results.

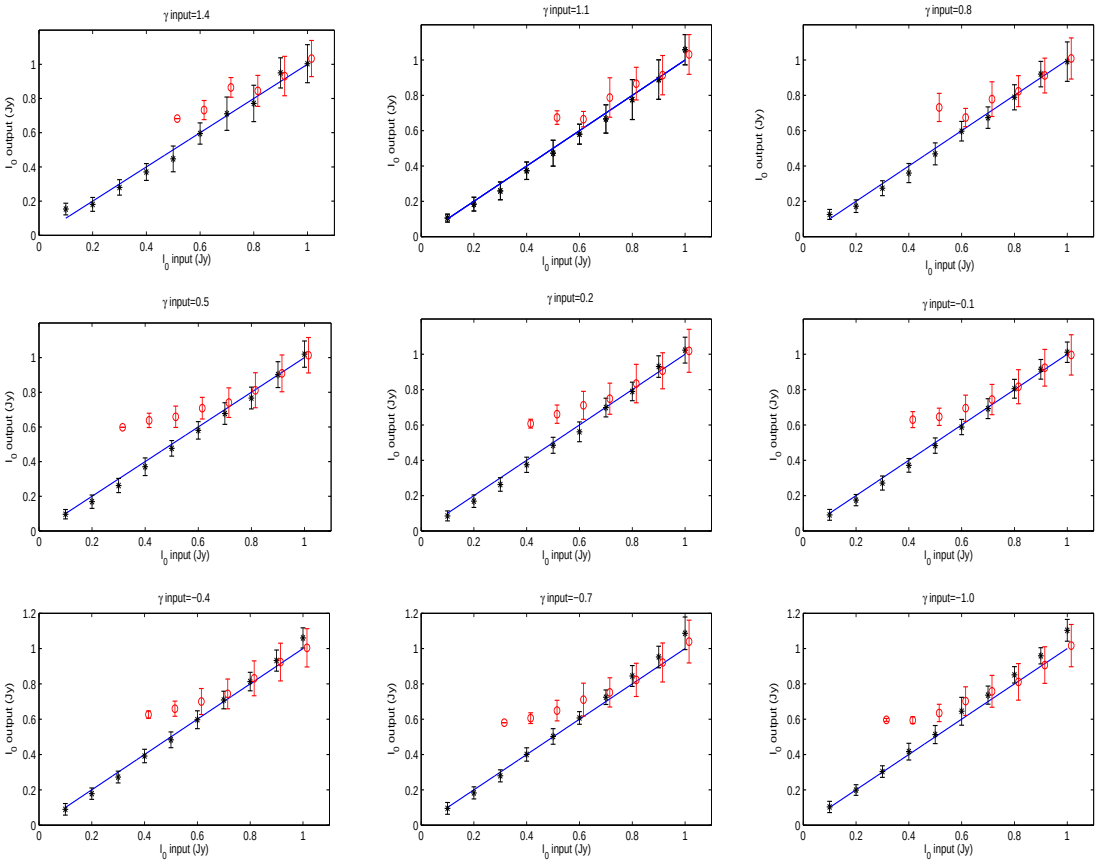

Figure 3: Values of I_0 (flux at 44 GHz) recovered by means of the MMF (asterisks) and the MF (circles). The line indicates the ideal recovering of the input. The circles corresponding to the MF are slightly displaced in the horizontal axis in order to distinguish the results.

ones we obtain with the matched multifilter. It is particularly notorious at small values of I_0, where the recovered values of the flux have a good agreement with respect to the input values, with small error bars.

In general, for all the values of γ that we have studied, the matched multifilter is a suitable and effective tool to estimate the I_0 of the sources. For the matched filter we observe a good determination of I_0 for input values above 0.7 Jy. For smaller values, I_0 has a higher value than its real one. That is due to the Eddington bias at 44 GHz. At this frequency, in Figure 1 we observe that for values smaller than 0.6 Jy, we only detect a $\sim 40\%$ of the total sources. That means that many of these sources are close to this noise level. And for the correct estimation of I_0 and the spectral index, we need a good detection of the sources at the two channels. For low values of I_0 the number of detected objects is small and we have few statistics.

The same conclusions are applicable to the estimation of the fluxes at the second frequency (100 GHz in our example), even if in the MMF case these fluxes are estimated using eq. (8) with the consequent propagation of the detection errors.

4.4 Reliability

In order to study the MMF in terms of reliability and spurious detections, we produce a new set of more realistic simulations (100) with the following characteristics:

- We used as a background the same eight regions described in the previous sections.
- The sources were simulated with an almost uniform Poissonian distribution (see[4] for more details about the method) at 44 GHz, with fluxes that follow the source number counts model of[3].
- The fluxes at 100 GHz were estimated assuming random spectral indices from the[5] distribution.
- The point source maps were filtered with the same resolution as the background maps and randomly added to them.

There is also another interesting quantity commonly used in the study of the performance of a source detector: the number of spurious sources. Spurious sources are fluctuations of the background that satisfy the criteria of the detection method and therefore are considered as detected sources. It is clear that the best method will be the one that has the best detections vs. spurious ratio. The maps are filtered using the MMF and the MF at both frequencies. We estimate the position and flux of the sources above 3σ level and, by comparing them with the input source simulations, we count the number of real and spurious sources that we are able to detect. We have also changed the detection level from 5σ to 3σ in order to increase the number of spurious sources to make the analysis.

In Figure 4 we observe the number of real sources that both methods are capable to detect, whose intrinsic fluxes are higher than the corresponding value in the horizontal axis. As we can see at 44 GHz, MMF detects a higher number of real sources for fluxes below $\sim 0.4 - 0.5$ Jy, being this difference very important at lower fluxes. At 100 GHz we observe a similar behaviour, but in this case the differences between the MF and the MMF start at ~ 0.2 Jy. If we observe Figure 1, we notice that the number of sources detected with the MF is higher at 100 GHz than at 44 GHz for values of the spectral index between 0 and 0.5. These values of γ, according to the model used to simulate the point sources in this section, are the most frequent ones. This gives us an idea about why the detection level of the MF is higher at 100 GHz.

In Figure 5 the *reliability* of both methods at 44 and 100 GHz is compared. Reliability above a certain recovered flux is defined as $r = N_d/(N_d + N_s)$, where N_d is the number of real sources

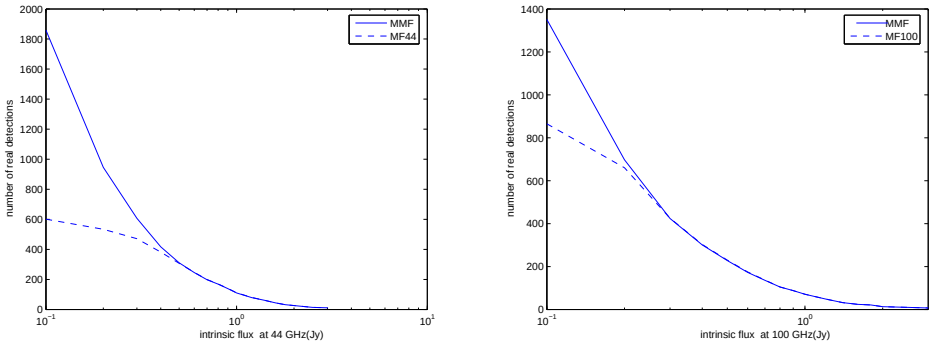

Figure 4: Number of real sources recovered by the MMF (solid line) and the MF (dashed line) at 44 GHz (left panel) and 100 GHz (right panel) whose intrinsic fluxes are higher than the corresponding value in the x axis.

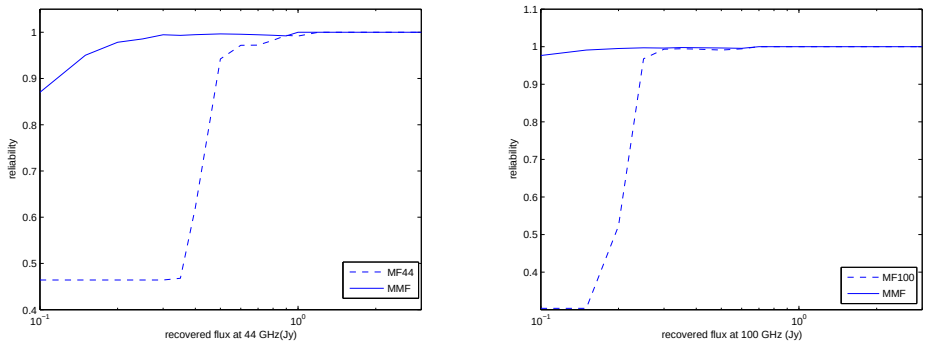

Figure 5: Reliability versus recovered flux for the MMF (solid line) and the MF (dashed line) at 44 GHz (left panel) and 100 GHz (right panel).

above that flux, and N_s is the number of spurious sources above the same flux. At 44 GHz we reach a $\sim$ 100% of reliability at fluxes of $\sim$ 0.3 Jy with the MMF. However, the MF at this frequency reaches this level of reliability only for $\sim$ 1 Jy. On the other hand, at 100 GHz we obtain better levels of reliability. For instance, with the MMF we have at 0.1 Jy more than 95% of reliability ($\sim$ 0.3 Jy for the MF). Therefore, we can say that the MMF is more reliable than the MF, specially at lower fluxes.

Moreover, we make an additional plot where we represent, for both frequencies, the number of real sources detected vs. the number of the spurious sources (Figure 6). In this way, what we represent is the number of sources that a method detects given a number of spurious sources. If we compare both plots, we can see that the curve of the MMF is always above the MF. It means that, when we have a fixed number of spurious detections, the MMF method detects more real sources.

Finally, we have to point out that the plots that we have introduced here are not directly comparable to Figure 1, due to three basic but important differences:

- A different way to simulate the point sources.
- A different level of detection (in this case, a 3σ level).
- In Figure 1 we represent the number of sources with the corresponding flux in the horizontal axis. In the plots of this section, what we represent is the number of sources whose fluxes are higher than the corresponding value in the horizontal axis.

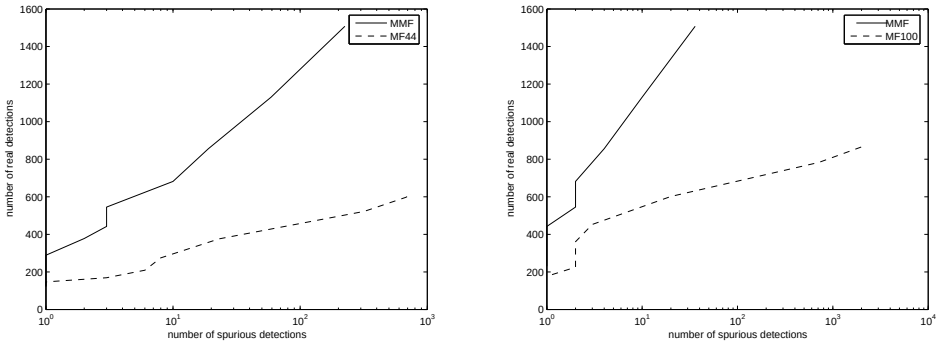

Figure 6: Number of real sources recovered by the MMF (solid line) and the MF (dashed line) at 44 GHz (left panel) and 100 GHz (right panel) vs. the number of spurious sources.

Acknowledgment

The authors acknowledge partial financial support from the Spanish Ministry of Education (MEC) under project ESP2004-07067-C03-01 and the joint CNR-CSIC research projects 2006-IT-0037 and 2008IT0059. LFL acknowledges the Spanish CSIC for a JAE-Predoc fellowship. Partial financial support for this research has been provided to JLS by the Spanish MEC and to JG-N by the Italian ASI (contracts Planck LFI Activity of Phase E2 and I/016/07/0 COFIS) and MUR. JG-N also acknowledges a research position grant at the SISSA (Trieste). ML-C acknowledges a post-doctoral fellowship from EGEE-III (FP7 INFSO-RI 222667). The authors are grateful to Raquel Fraga Encinas for reading carefully this document and her style and grammatical suggestions and acknowledge the use of the PSM, developed by the Component Separation Working Group (WG2) of the Planck Collaboration.

References

1. R.B. Barreiro *et al*, MNRAS **342**, 119 (2003).
2. A.S. Eddington, MNRAS **73**, 359 (1913).
3. G. De Zotti *et al*, A&A **431**, 893 (2005).
4. J. González-Nuevo *et al*, ApJ **621**, 1 (2005).
5. J. González-Nuevo *et al*, MNRAS **384**, 711 (2008).
6. D. Herranz *et al*, MNRAS **336**, 1057 (2002).
7. L.F. Lanz *et al*, MNRAS **403**, 2120 (2010).
8. S.M. Leach *et al*, A&A **491**, 597 (2008).
9. M. López-Caniego *et al*, MNRAS **370**, 2047 (2006).
10. J.A. Tauber, in Lasenby A.N., Wilkinson A., eds, Proc. IAU Symp. **201**, New Cosmological Data and the Values of the Fundamental Parameters. Astron. Soc. Pac., S. Francisco (p.86)2005.
11. M. Tegmark & A. de Oliveira-Costa, ApJ **500**, L83 (1998).
12. The Planck Collaboration, arXiv:astro-ph/0604069v1 , (2006).