Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 3 September 2003 (MN LATEX style file v1.4)

# A semi-blind second-order approach for statistical source separation in astrophysical maps

L. Bedini[1], S. Bottini[1], C. Baccigalupi[2,3], P. Ballatore[1], D. Herranz[1], E.E. Kuruoğlu[1], E. Salerno[1], and A. Tonazzini[1]

[1] *Istituto di Scienza e Tecnologie dell'Informazione. CNR, Area della Ricerca di Pisa, via G. Moruzzi 1, 56124 Pisa, Italy.*
[2] *International School for Advanced Studies, via Beirut 4, 34014 Trieste, Italy.*
[3] *Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley CA 94720*

**ABSTRACT**

This paper deals with a semi-blind source separation strategy which is applicable in the cases where the mixing operator can be described by just a few parameters. In these cases, unlike the independent component analysis approaches, our method only needs the covariance matrix of the data for learning the mixing operator, even if the original source processes to be separated are not totally uncorrelated. We are also able to estimate the probability density function of the source processes by a simple deconvolution procedure. Our algorithm has been tested with a database that simulates the one expected from the instruments that will operate onboard ESA's Planck Surveyor Satellite to measure the CMB anisotropies all over the celestial sphere. The assumption was made that the emission spectra of the galactic foregrounds can be parametrized, thus reducing the number of unknowns for system identification to the number of the foreground radiations. We performed separation in several sky patches, featuring different levels of galactic contamination to the CMB, and assuming several noise levels, including the ones derived from the Planck specifications. In all the cases, the CMB reconstruction was satisfactory on all the angular scales considered in the simulation; the average performance of the algorithm and its dispersion were checked and quantified against a large set of noise patterns.

**Key words:** methods: statistical – techniques: image processing – cosmic microwave background.

## 1 INTRODUCTION

Any radiometric measurement on sky emission results from a superposition of different radiating sources. Separating the individual radiations from the measured signals is thus a common problem in astrophysical data analysis (Tegmark et al. 2000). As an example, in cosmic microwave background anisotropy surveys, the cosmological signal is normally combined with radiation of both extragalactic and galactic origin, such as the Sunyaev-Zeldovich effects from clusters of galaxies, the effect of the individual radiogalaxies, the emission from galactic dust, the galactic synchrotron and free-free emissions. If one is only interested in estimating the CMB anisotropies, the interfering signals can just be treated as noise, and reduced by suitable cancellation procedures. However, the source signals have an interest of their own, and it could be useful to extract all of them from multichannel data, by exploiting their different emission spectra.

From measurements on different frequency channels, it is possible to extract a number of individual radiation data, assuming that the physical mixture model is perfectly known (Hobson et al. 1998, Bouchet et al. 1999). Unfortunately, such an assumption is rather unrealistic and could overconstrain the problem, thus leading to unphysical solutions. Attempts are being made to avoid this shortcoming by relaxing the strict requirement on the data model (Barreiro et al. 2003).

A class of techniques capable of estimating the source signals as well as identifying the mixture model has recently been proposed in astrophysics (Baccigalupi et al. 2000, Maino et al. 2002, Baccigalupi et al. 2002, Delabrouille et al. 2002). In digital signal processing, these techniques are referred to as blind source separation (BSS) and rely on statistical assumptions on the source signals. In particular, mutual independence and nongaussianity of the source processes are often required (Hyvärinen and Oja 2000). This totally blind approach, denoted as independent component analysis (ICA), has already given promising results, proving

to be a valid alternative to assuming a known data model. On the other hand, most ICA algorithms do not permit to introduce prior information. Since all available information should always be used, semi-blind techniques are being studied to make astrophysical source separation more flexible with respect to the specific knowledge often available in this type of problem (Kuruoğlu et al. 2003).

The first blind technique proposed to solve the separation problem in astrophysics (Baccigalupi et al. 2000) was based on ICA, allowing simultaneous model identification and signal estimation to be performed. In order to get independence, the statistics of all orders should be taken into account, and this can be achieved by explicit cumulant calculation or by suitable nonlinear transformations on the estimated source signals (Comon 1994, Cardoso 1998, Hyvärinen and Oja 2000).

The problem of estimating all the model parameters and source signals cannot be solved by just using second-order statistics, since these are only able to enforce uncorrelation. However, this has been done in special cases, where additional hypotheses on the spatial correlations within the individual signals are assumed (Tong et al. 1991, Belouchrani et al. 1997). Another possibility of solving the separation problem on the the basis of second-order statistics alone is when the relevant constraints are such that the total number of parameters to be estimated can be reduced. This is the case in the problem of separating astrophysical foregrounds from cosmic microwave background, provided that a structure for the source covariance matrix is assumed. Indeed, on the basis of a possible parametrisation of the mixing operator and the assumption of decorrelation between the different sources, it is possible to reduce the total number of model parameters. We will show that, under these hypotheses, a very fast model learning algorithm can be devised, based on matching the estimated zero-shift covariance matrix with the one obtained from the available data. This strategy also offers additional advantages, such as the possibility of estimating the source probability density functions and relaxing the strict assumption of uncorrelation between the source signals.

This paper is organised as follows. In Section 2, we formalise the problem and introduce the relevant notation. In Section 3, we describe our method, and, in Section 4, we present some experimental results. Some remarks and future directions are reported in the final section.

## 2 ASTROPHYSICAL SOURCE SEPARATION

As usual (Hobson et al. 1998, Baccigalupi et al. 2000), we assume that each radiation process $\tilde{s}_c(\xi,\eta,\nu)$ from the microwave sky has a spatial pattern $s_c(\xi,\eta)$ that is independent of its frequency spectrum $F_c(\nu)$:

$$\tilde{s}_c(\xi,\eta,\nu) = s_c(\xi,\eta)F_c(\nu) \tag{1}$$

Here, $\xi$ and $\eta$ are angular coordinates on the celestial sphere, and $\nu$ is frequency. The total radiation observed in a certain direction at a certain frequency is given by the sum of a number $N$ of signals (processes, or components) of the type (1), where subscript $c$ has the meaning of a process index. Assuming that the effects of the telescope beam on the angular resolution at different measurement channels have been equalised (Salerno et al. 2000), the observed signal at $M$ different frequencies can be modelled as

$$\mathbf{x}(\xi,\eta) = \mathbf{A}\mathbf{s}(\xi,\eta) + \mathbf{n}(\xi,\eta) \tag{2}$$

where $\mathbf{x}=\{x_d,\ d=1,\ldots,M\}$ is the $M$-vector of the observations, $d$ being a channel index, $\mathbf{A}$ is an $M\times N$ matrix whose entries, $a_{dc}$, are related to the spectra of the radiation sources and the frequency responses of the measuring instruments on the different frequency channels, $\mathbf{s}=\{s_c,\ c=1,\ldots,N\}$ is the $N$-vector of the individual source processes and $\mathbf{n}=\{n_d,\ d=1,\ldots,M\}$ is the $M$-vector of instrumental noise. Hereafter, matrix $\mathbf{A}$ will be referred to as the mixing matrix.

The separation problem consists in estimating the source vector $\mathbf{s}$ from the observed vector $\mathbf{x}$. Several estimation algorithms have been derived assuming a perfect knowledge of the mixing matrix. As already said, however, this matrix is related both to the instrumental frequency responses, which are known, and to the emission spectra $F_c(\nu)$, which are normally unknown. For this reason, relying on an assumed mutual independence of the source processes $s_c(\xi,\eta)$, some *blind* separation algorithms have been proposed (Baccigalupi et al. 2000, Maino et al. 2002, Patanchon et al. 2003), which are able to estimate both the mixing matrix and the source vector. For any procedure of this type, the estimation of $\mathbf{A}$ will be referred to as *system identification* (or *model learning*), and the estimation of $\mathbf{s}$ will be referred to as *source separation*. In this paper, we will only emphasise aspects related to learning; indeed, once the model has been identified, a number of standard reconstruction procedures are available to separate the individual sources. Assuming that the source signals are mutually independent, the $MN$ unknown coefficients can be estimated by finding a linear mixture that, when applied to the data vector, nullifies the cross-cumulants of all orders. If, however, some prior information allows us to reduce the number of unknowns, the identification problem can be solved by only using second-order statistics. This is the case with our approach, which is based on a parametrisation of matrix $\mathbf{A}$. This approach is described in the following section. Here we only introduce some additional notation that will be used. Learning the model on the basis of second-order statistics alone is also advantageous with respect to robustness against noise; indeed, estimation of second-order statistics is much more immune from erratic data than estimation of higher-order statistics.

The elements of $\mathbf{A}$ are related to the source spectra and to the frequency responses through the following formula:

$$a_{dc} = \int F_c(\nu) b_d(\nu) d\nu \tag{3}$$

where $b_d(\nu)$ is the instrumental frequency response on the $d$-th measurement channel, which is normally very well known. If we assume that the source spectra are constant within the passbands of the different channels, equation (3) can be rewritten as

$$a_{dc} = F_c(\nu) \int b_d(\nu) d\nu \tag{4}$$

The element $a_{dc}$ is thus proportional to the spectrum of the $c$-th source at the center-frequency $\nu_d$ of the $d$-th channel. While in a general source separation problem the

elements $a_{dc}$ are totally unknown, in our case we have some knowledge about them. In fact, the integral in (4) is related to known instrumental features and to the emission spectra of the single source processes, on which we do have some knowledge. As an example, if the observations are made in the microwave and millimeter-wave range, the dominant radiations are the cosmic microwave background, the galactic dust, the free-free emission and the synchrotron (see De Zotti et al., 1999). Another significant signal comes from the extragalactic radio sources. Here we assume that the latter has been removed from the data by one of the specific techniques proposed in the literature (Tenorio et al., 1999, Cayón et al., 2000, Vielva et al., 2001). As a matter of fact, the mentioned techniques cannot remove totally the extragalactic point sources, but they remove the brightest ones (that are the most important, since they are the ones that affect more the the study of the CMB, see Vielva et al. 2001). As far as the other signals are concerned, the emission spectrum of the cosmic microwave background is perfectly known, being a blackbody radiation. In terms of antenna temperature, it is:

$$F_{CMB}(\nu) = \frac{\tilde{\nu}^2 \exp(\tilde{\nu})}{[\exp(\tilde{\nu}) - 1]^2} \tag{5}$$

where $\tilde{\nu}$ is the frequency in GHz divided by 56.8. From (4) and (5), the column of $\mathbf{A}$ related to the CMB radiation is thus known up to an inessential scale factor. For the synchrotron radiation, we have

$$F_{syn}(\nu) \propto \nu^{-n_s} \tag{6}$$

Thus, the column of $\mathbf{A}$ related to synchrotron only depends on a scale factor and the spectral index $n_s$. For the thermal galactic dust, we have

$$F_{dust}(\nu) \propto \frac{\bar{\nu}^{m+1}}{\exp(\bar{\nu}) - 1} \tag{7}$$

where $\bar{\nu} = h\nu/kT_{dust}$, and where $h$ is the Planck constant, $k$ is the Boltzmann constant and $T_{dust}$ is the physical dust temperature. If we assume this temperature constant, the frequency law (7), that is, the column of $\mathbf{A}$ related to dust emission, only depends on a scale factor and the parameter $m$. In general, if we assume to have a perfectly known source (such as CMB) and $N-1$ sources with one-parameter spectra, the number of unknowns in the identification problem is $N-1$ instead of $NM$.

For the sake of simplicity, although other foregrounds (such as SZ and free-free) could be taken into account, in our experiments we only considered synchrotron and dust emissions, which are the most significant in the Planck frequency range.

## 3 A SECOND-ORDER IDENTIFICATION ALGORITHM

Let us consider the source and noise signals in (2) as realisations of two stationary vector random processes, whose components are mutually independent. The covariance matrices of these processes are, respectively,

$$\mathbf{C_s}(\tau,\psi) = \langle [\mathbf{s}(\xi,\eta) - \mu_\mathbf{s}] [\mathbf{s}(\xi+\tau,\eta+\psi) - \mu_\mathbf{s}]^\mathrm{T} \rangle \ , \tag{8}$$

$$\mathbf{C_n} = \langle \mathbf{n}\mathbf{n}^\mathrm{T} \rangle \tag{9}$$

where $\langle . \rangle$ denotes expectation under the appropriate joint probability, $\mu_s$ is the mean vector of process s, and the superscript T means transposition. As usual, the noise process is assumed signal-independent, white and zero-mean, with known variances. Thus, $\mathbf{C_n}$ is a known diagonal matrix whose elements are the noise variances in all the measurement channels. For the assumed mutual source independence, we know that $\mathbf{C_s}$ must be diagonal. In the following, however, we will see that this assumption can be somewhat relaxed.

By exploiting equation (2), the covariance of the observed data at shift $(\tau,\psi) = (0,0)$ can be written as:

$$\mathbf{C_x}(0,0) = \langle [\mathbf{x}(\xi,\eta) - \mu_\mathbf{x}] [\mathbf{x}(\xi,\eta) - \mu_\mathbf{x}]^\mathrm{T} \rangle = \mathbf{A}\mathbf{C_s}(0,0)\mathbf{A}^\mathrm{T} + \mathbf{C_n} \ . \tag{10}$$

Let us now define the matrix

$$\mathbf{H} = \mathbf{A}\mathbf{C_s}(0,0)\mathbf{A}^\mathrm{T} = \mathbf{C_x}(0,0) - \mathbf{C_n}. \tag{11}$$

As already proved (Belouchrani et al. 1997, Barros and Cichocki 2001), covariance matrices, i.e. second-order statistics, permit blind separation to be achieved when the sources show a spatial structure, namely, when they are spatially correlated. In these cases, the independence requirement of the ICA approach (Comon 1994) is replaced by an equivalent requirement on the spatial structure of the signal. In other words, finding matrices $\mathbf{A}$ and $\mathbf{C_s}$ is generally not possible from equation (11) alone; higher-order statistics or the covariance matrices at several shifts $(\tau,\psi)$ must be taken into account (Belouchrani et al. 1997).

As assumed in the previous section, in this application we are able to reduce the number of unknowns by parametrising the mixing matrix. This allows us to rely on zero-shift covariance alone to solve the problem. Let $\mathbf{A}$ only depend on $N-1$ parameters $\alpha_i$, $i = 1,\ldots,N-1$, and $\mathbf{C_s}$ on its $N$ diagonal elements $\sigma_{ii}$, $i = 1,\ldots,N$. Once $\mathbf{H}$ is given, relationship (11) yields $M(M+1)/2$ nonlinear equations, one for each element of the symmetric matrix $\mathbf{H}$, and $2N-1$ unknown parameters. Note that the number of equations is larger than the number of unknowns as soon as $M = N$ and $N > 2$. Actually, we do not possess matrix $\mathbf{H}$ as defined by (10) and (11), but we are able to get an estimate of it from the known matrix $\mathbf{C_n}$ and the following estimate of $\mathbf{C_x}$(0,0):

$$\mathbf{C_x} = \frac{1}{N_p} \sum_{\xi,\eta} [\mathbf{x}(\xi,\eta) - \mu_\mathbf{x}] [\mathbf{x}(\xi,\eta) - \mu_\mathbf{x}]^\mathrm{T} , \tag{12}$$

where the summation is extended over all the $N_p$ available data samples, and $\mu_x$ is the mean vector of the data sequence. Let

$$\hat{\mathbf{H}} = \mathbf{C_x} - \mathbf{C_n} \tag{13}$$

be our estimate of matrix $\mathbf{H}$. Matrices $\mathbf{A}$ and $\mathbf{C_s}$ can be estimated from

$$(\alpha_1,\ldots,\alpha_{N-1},\sigma_{11},\ldots,\sigma_{NN}) = \arg\min \left\| \mathbf{A}(\alpha_1,\ldots,\alpha_{N-1})\mathbf{C_s}(\sigma_{11},\ldots,\sigma_{NN})\mathbf{A}^\mathrm{T}(\alpha_1,\ldots,\alpha_{N-1}) - \hat{\mathbf{H}} \right\|, \tag{14}$$

where the minimisation is performed over all the possible values of $\alpha_1,\ldots,\alpha_{N-1},\sigma_{11},\ldots,\sigma_{NN}$, and an appropriate matrix norm should be selected. Our present strategy to find the minimiser is to alternate a componentwise minimisation in $\alpha_i$ with fixed $\mathbf{C_s}$, and the evaluation of $\mathbf{C_s}$, whose elements, $\{\sigma_{11},\ldots,\sigma_{NN}\}$, can be calculated exactly once $\mathbf{A}$

is fixed. A more accurate minimisation strategy is now being studied.

Even though the sources are not totally uncorrelated, we still have possibilities of estimating $\mathbf{A}$, provided that we know which sources are actually correlated and the number of available equations is still sufficient. More precisely, for each pair of significantly correlated sources, we can add an extra unknown in matrix $\mathbf{C_s}$, and this can be evaluated in the same way as for the diagonal elements. Since in our case the foreground maps are correlated, we will see a realistic example of this possibility in the experimental section below.

By exploiting the results of model learning, we are able to estimate the individual source maps, and, also, their probability density. As already said, many choices are possible for source estimation. We now show a strategy to estimate their densities as well. Let us assume of having an estimate of $\mathbf{A}$. Let $\mathbf{B}$ be its Moore-Penrose generalised inverse. In our case we have $M \geq N$, thus, as is known,

$$\mathbf{B} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}. \tag{15}$$

From (2) we have

$$\mathbf{Bx}=\mathbf{s}+\mathbf{Bn} \tag{16}$$

Let us denote each of the $N$ rows of $\mathbf{B}$ as an $M$-vector $b_i$, $i = 1, \ldots, N$, and consider the generic element $y_i$ of the $N$-vector $\mathbf{Bx}$,

$$y_i := \mathbf{b}_i^{\mathrm{T}} \cdot \mathbf{x} = s_i + \mathbf{b}_i^{\mathrm{T}} \cdot \mathbf{n} := s_i + n_{t_i} \tag{17}$$

The probability density function of $y_i$, $p(y_i)$, can be estimated from $\mathbf{b}_i$ and the data record $\mathbf{x}(\xi, \eta)$, while the probability density function of $n_{t_i}$, $p(n_{t_i})$, is a Gaussian, whose parameters can be easily derived from $\mathbf{C_n}$ and $\mathbf{b}_i$. The pdf of $y_i$ is the convolution between $p(s_i)$ and $p(n_{t_i})$:

$$p(y_i) = p(s_i) * p(n_{t_i}). \tag{18}$$

From this relationship, $p(s_i)$ can be estimated by deconvolution. As is well known, deconvolution is normally an ill-posed problem and, as such, it lacks a stable solution. In our case, we can regularise it by the obvious positivity constraint and the normalisation condition for pdfs.

Equation (16) already gives a rough estimate of the source processes. As can be seen, applying the generalised inverse to our data gives the original source vector, corrupted by amplified noise. A simple source estimation strategy could be to apply first equation (16) and then suitably filter the result, to reduce the influence of noise. More sophisticated strategies can also be adopted, such as different types of probabilistic estimates. To this purpose, the availability of the source distributions can be very useful. Indeed, any probabilistic estimation approach can exploit their knowledge to regularise the solution. In other words, being the linear estimate of $\mathbf{s}$ in (16) corrupted by Gaussian noise, a Bayesian procedure can refine the estimation by exploiting functions $p(s_i)$ as prior source distributions. In the case examined here, the source distributions should be estimated from the data, since they are normally unknown, save for the one of the CMB, which should be Gaussian. This strategy has been attempted for astrophysical source separation by Kuruoğlu et al. (2003), using the Independent Factor Analysis (IFA) approach (Moulines et al., 1997, Attias 1999). The source densities are modelled as mixtures of Gaussians, and the related parameters are estimated by means of a very expensive algorithm. Minimisation in (14) followed by deconvolution of equation (18) is not very expensive computationally, thus, if included in the IFA strategy, our learning procedure can reduce the overall computational cost of the whole identification-separation task.

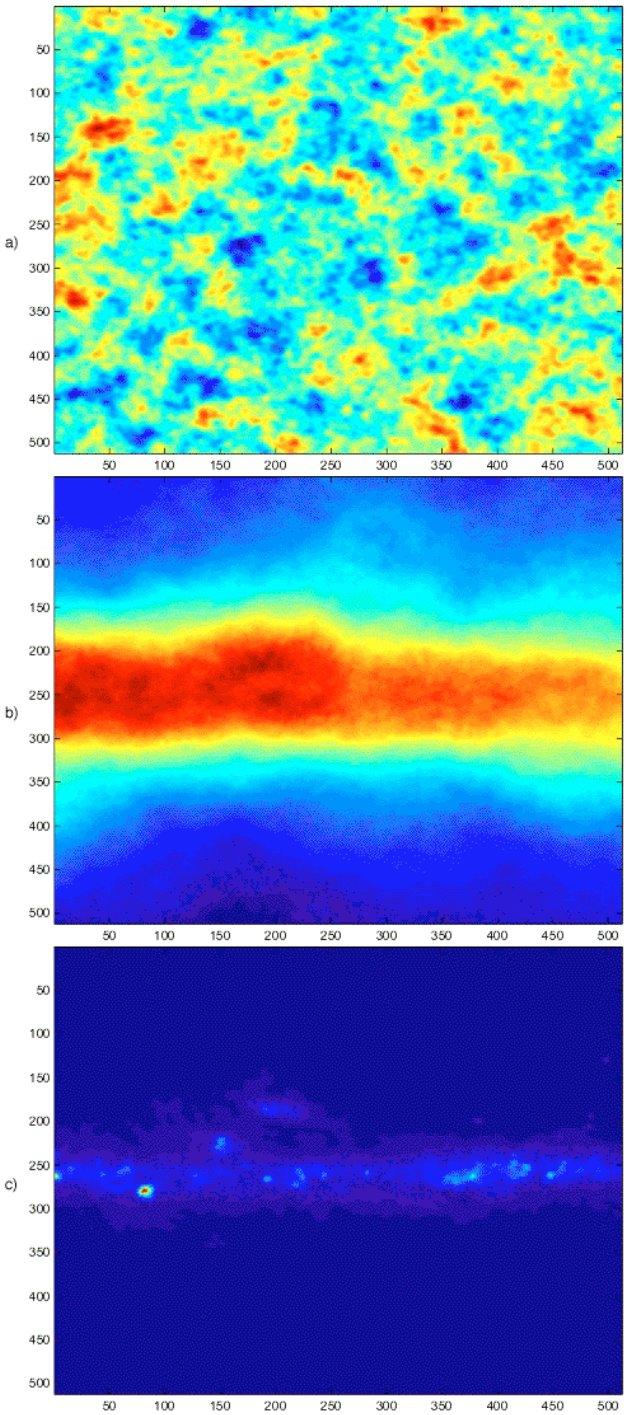


**Figure 1.** Source maps from a $15° \times 15°$ patch centered at $0°$ galactic latitude and $20°$ galactic longitude, at 100 GHz: a) CMB; b) synchrotron; c) thermal dust.

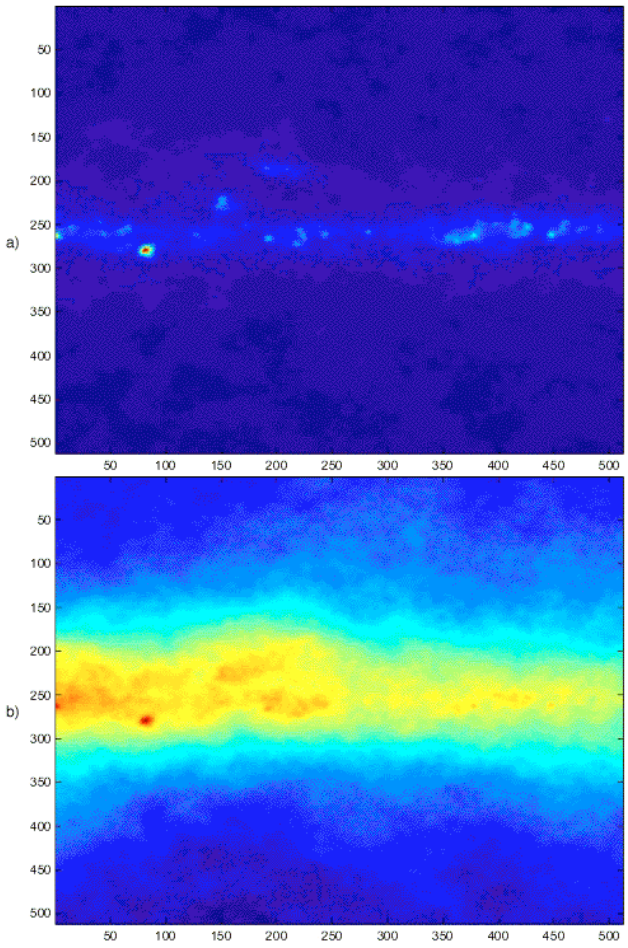


**Figure 2.** Data maps containing CMB, synchrotron and dust emission as well as instrumental noise at a) 100 GHz; b) 30 GHz.

## 4 EXPERIMENTAL RESULTS

We tested our technique against a simulated data set that is a simplified version of the one expected from the Planck surveyor satellite (Mandolesi et al., 1998, Puget et al., 1998). The source maps we considered were the CMB anisotropy, the galactic synchrotron and thermal dust emissions over the four measurement channels centred at 30 GHz, 44 GHz, 70 GHz and 100 GHz. The test data maps have been generated by extracting several sky patches at different galactic coordinates from the simulated database, scaling them exactly according to formulas (5)-(7), generating the mixtures for the channels chosen, and adding realisations of Gaussian, signal independent, white noise. Several noise levels have been used, from a ten percent to more than one hundred percent of the CMB standard deviation. The range chosen contains noise levels within the Planck specifications (see the Planck homepage*).

It is to remark that model learning and separation failed with sky patches taken at high galactic latitudes, where the only dominant signal is the CMB, and the foregrounds are often well below the noise signal. Some other techniques, such as ICA (see Baccigalupi et al. 2000), did obtain good

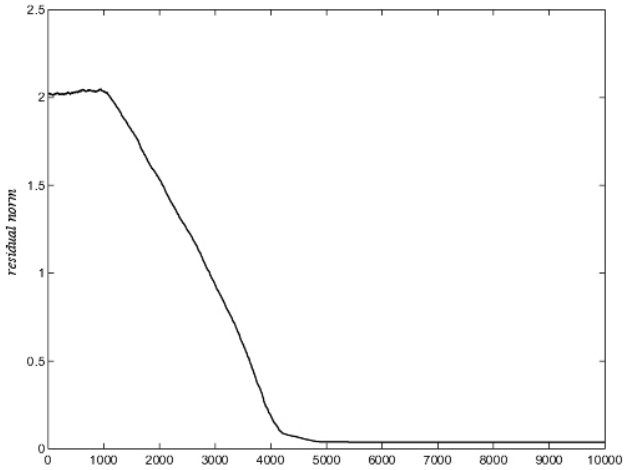


**Figure 3.** Norm of the residual in eq. (14) as a function of the iteration number.

results even in these regions, but the noise levels introduced in those cases were much lower than the ones we have used in this work. In these regions, however, CMB is almost the only measured radiation at the considered frequencies, and is estimated very well with all the assigned signal-to-noise ratios. At lower galactic latitudes, conversely, the situation is rather different. Here, the dust emission is stronger than CMB, and separation is strictly necessary if CMB is to be distinguished from the foregrounds. Our method performed very well with these data, and all the relevant parameters were satisfactorily estimated even with the strongest noise components.

In this section, we report a specific example and describe the results of extensive trials on different sky patches. In figure 1, we show the three source maps we used for a typical experiment. These maps were extracted from a sky patch of $15^\circ \times 15^\circ$, centred at $0^\circ$ of galactic latitude and $20^\circ$ galactic longitude. Their size is $512 \times 512$, and thus each pixel is about 1.7 arcminutes wide. The standard deviations of the source processes at 100 GHz were 0.084 mK for CMB, 0.043 mK for synchrotron, and 0.670 mK for dust, in antenna temperature. We assigned the sources $s_1$ to CMB, $s_2$ to synchrotron and $s_3$ to dust, and the signals $x_1$, $x_2$, $x_3$ and $x_4$ to the measurement channels at 100, 70, 44 and 30 GHz, respectively. Therefore, the columns of the mixing matrix will be related to CMB, synchrotron and dust in the order mentioned above; the rows will be related to the four measurement channels, in the order mentioned above. The true mixing matrix, $\mathbf{A}_o$, has been derived from equations (5)-(7), with spectral indices $n_s = 2.9$ and $m = 1.8$ (see for example Banday and Wolfendale, 1991, and Giardino et al. 2002):

$$\mathbf{A}_o = \begin{pmatrix} 1 & 1 & 1 \\ 1.135 & 2.813 & 0.548 \\ 1.224 & 10.814 & 0.246 \\ 1.257 & 32.836 & 0.126 \end{pmatrix}. \qquad (19)$$

The noise covariance matrix we adopted, in $\mathrm{mK}^2$, is:

* http://astro.estec.esa.nl/SA-general/Projects/Planck/

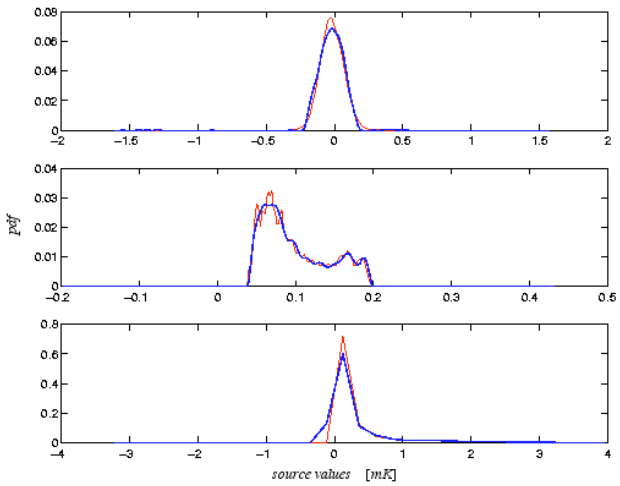


**Figure 4.** Real (blue) and estimated (red) source density function. Top panel: CMB; middle panel: synchrotron; bottom panel: dust.

$$\mathbf{C}_{\rm n} = \begin{pmatrix} 0.0846^2 & 0 & 0 & 0 \\ 0 & 0.1128^2 & 0 & 0 \\ 0 & 0 & 0.0982^2 & 0 \\ 0 & 0 & 0 & 0.1306^2 \end{pmatrix}. \qquad (20)$$

Note that the noise standard deviation at 100 GHz is the same as the CMB standard deviation. The other values have been established by scaling the standard deviation at 100 GHz in accordance with the expected Planck sensitivity at each channel. We only retained the relative scalings among the sensitivities, and not the expected noise variances, in order to analyse our algorithm with a wide range of noise levels. In figure 2, we show the noisy data maps for the 100 GHz and the 30 GHz channels. Also, note that the case examined does not fit the assumptions of the independent component analysis approach to blind source separation. Indeed, there is a significant correlation, of the order of 62%, between the dust and synchrotron maps, whereas the empirical correlations between CMB and synchrotron and CMB and dust are, respectively, 0.9% and 1.9%. The original source covariance matrix at zero shift is:

$$\mathbf{C}_{\rm s} = \begin{pmatrix} 0.00716 & 0.00003 & -0.00112 \\ 0.00003 & 0.00183 & 0.01778 \\ -0.00112 & 0.01778 & 0.44990 \end{pmatrix}. \qquad (21)$$

For the data described above, we ran our learning algorithm for 500 different noise realisations; for each run, 10000 iterations of the minimisation procedure described in the previous section were performed. The unknown parameters were the spectral indices $n_s$ and $m$ and the covariances $\sigma_{11}$, $\sigma_{22}$, $\sigma_{33}$ and $\sigma_{23}$, assuming a negligible correlation between the CMB and the foregrounds and a significant correlation between the synchrotron and dust radiations. In implementing our algorithm, we adopted the Frobenius norm in the objective function (14). The behaviour of this norm as a function of the iteration number in a particular run is shown in figure 3. The typical elapsed times per run were tens of seconds on a 2 GHz CPU computer, with a Matlab interpreted code. In a typical case, we estimated $n_s = 2.8904$ and $m = 1.8146$, corresponding to the mixing matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 1.135 & 2.804 & 0.546 \\ 1.224 & 10.729 & 0.244 \\ 1.257 & 32.459 & 0.124 \end{pmatrix}. \qquad (22)$$

As a quality index for our estimation, we adopted the matrix $\mathbf{Q}=(\mathbf{A}^{\rm T}\mathbf{C}_{\rm n}^{-1}\mathbf{A})^{-1}(\mathbf{A}^{\rm T}\mathbf{C}_{\rm n}^{-1}\mathbf{A}_{\rm o})$, which, in the ideal case, should be the $M \times M$ identity matrix $\mathbf{I}$. For the matrices in (19), (20) and (22), we have:

$$\mathbf{Q} = \begin{pmatrix} 1.0000 & -0.0516 & 0.0042 \\ -0.0000 & 1.0134 & -0.0001 \\ 0.0000 & 0.0400 & 0.9960 \end{pmatrix}. \qquad (23)$$

The Frobenius norm of matrix $\mathbf{Q}-\mathbf{I}$ should be zero in the case of perfect model learning. In this case, it is 0.0669. The estimated source covariance matrix, to be compared with the matrix in (21), is

$$\mathbf{C}_{\rm s}(0,0) = \begin{pmatrix} 0.00717 & 0 & 0 \\ 0 & 0.00188 & 0.01802 \\ 0 & 0.01802 & 0.44803 \end{pmatrix}, \qquad (24)$$

where, of course, the values of $\sigma_{12}$ and $\sigma_{13}$ were kept fixed to zero. The accordance between the estimated and the original matrices is apparent.

The reconstructed probability density functions of the source processes, estimated from equations (17) and (18), are shown in figure 4. The reconstructed densities are often indistinguishable from the original functions. In any case, the ranges of the individual processes have been identified very sharply. Although estimates of the prior densities are available, we separated the sources simply by multiplying the data matrix by the Moore-Penrose generalised inverse, as in (16). As already said, this is not the best choice reconstruction algorithm at all, especially when the data are particularly noisy. However, the results we obtained are visually very good, as shown in figure 5. Part of the noise still present in the results of figure 5 was then reduced by Wiener filtering the reconstructed maps. The final result is shown in figure 6. To evaluate more quantitatively the results of the whole learning-separation procedure, we compared the power spectrum of the CMB map with the one of the reconstructed map. This comparison is shown in figure 7, where we also show the possibility of correcting the reconstructed spectrum for the known theoretical spectrum of the noise component $n_{t_1}$, obtained as in (17). As can be seen, the reconstructed spectrum is very similar to the original within a multipole $l = 2000$.

Averaging the learned results over the 500 runs, we have found mean values 2.8894 for $n_s$ and 1.8146 for $m$, with standard deviations of 0.0015 and 0.0019, respectively. The slight bias of the results with respect to the assumed real values is currently under study, and it could depend on a lack of convergence of the minimisation algorithm we have been using so far. We compared all the reconstructed sources with the original ones, and computed the averaged power spectrum of the residual maps. The result for 100 runs with different noise realizations is shown in figure 8, where the averaged residual spectrum almost coincides with the theoretical noise spectrum. The detected bias in the results is thus not visible in the scale displayed in figure 8.

We repeated this analysis for several patches and different noise levels. Besides the already mentioned failure of the algorithm at high galactic latitudes, we have found that, in

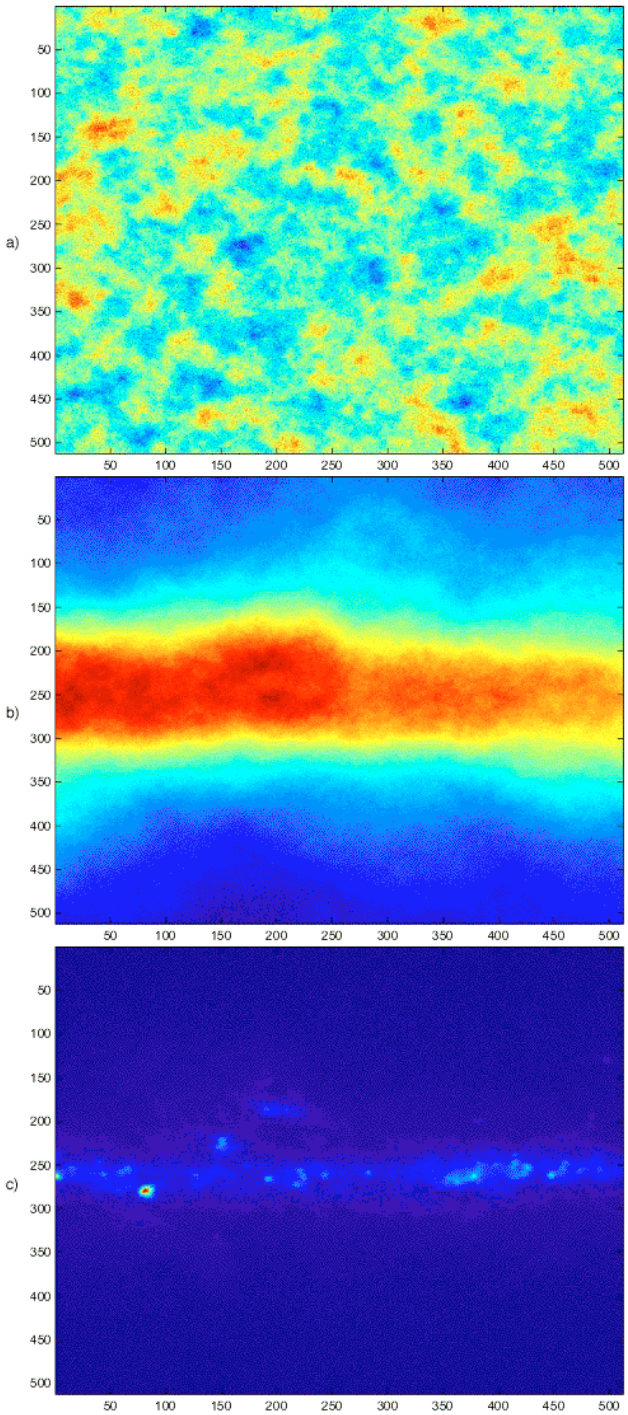


**Figure 5.** Pseudo-inverse reconstructed maps: a) CMB; b) synchrotron; c) dust.

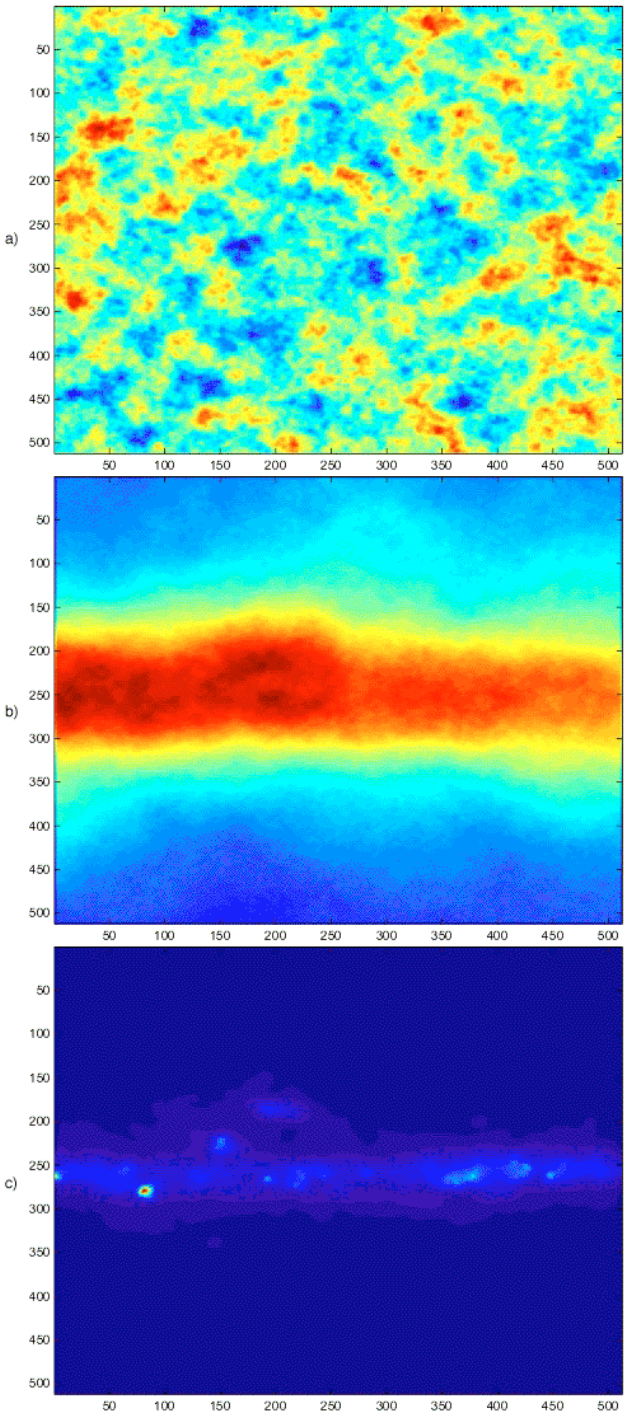


**Figure 6.** Wiener-filtered versions of the maps in figure 5.

all the patches taken into account, the average parameters estimated are the same as the ones reported above. As an example, the analysis carried out on a patch centered at a galactic latitude of $15^{\circ}$ and longitude $20^{\circ}$, with the noise level specified by the matrix in (20), gave average values 2.8897 and 1.8157 for $n_s$ and $m$, respectively, almost identical to the previously shown values. However, note that, since the noise covariance is the same as before but the signal power is smaller than the one we had for the patch in the galactic plane, the signal-to-noise ratio is in this case significantly smaller than before. This fact makes the single estimations much more dispersed than in the previous case. In fact, we have standard deviations of 0.0094 for $n_s$ and 0.1959 for $m$, about 9 and 100 times, respectively, larger than the values found on the previous patch. In figure 9, we show the averaged residual spectrum for this case.

As a general observation, we can say that, obviously, the noise on the estimated source maps will increase as the noise on the data maps increases. Wiener filtering can be used to

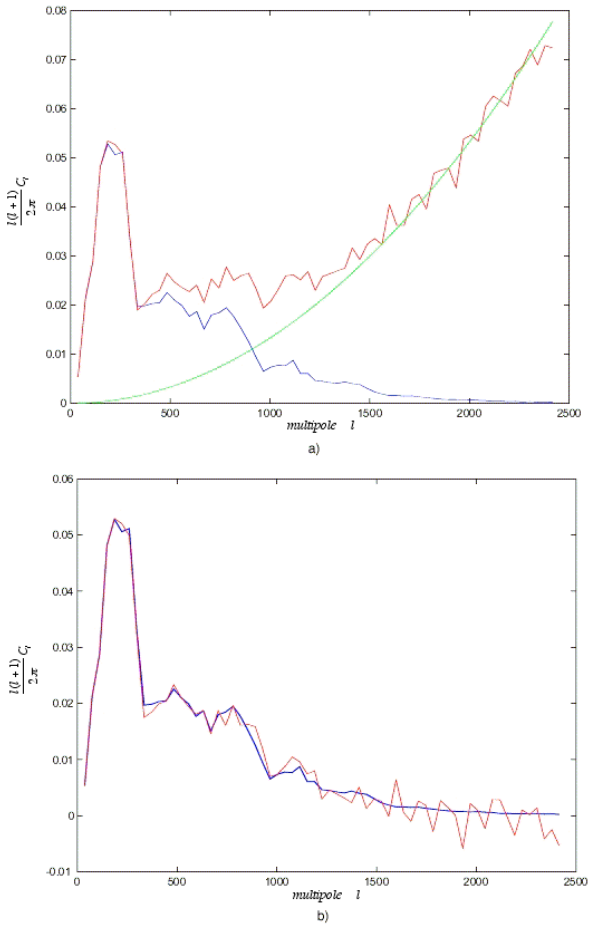


**Figure 7.** a) Real (blue) and estimated (red) CMB power spectra. The green line represents the theoretical power spectrum of the noise component $n_{t_1}$ in (17), evaluated from the noise covariance and the Moore-Penrose pseudoinverse of the estimated mixing matrix. b) Real (blue) and estimated (red) CMB power spectrum, corrected for theoretical noise.

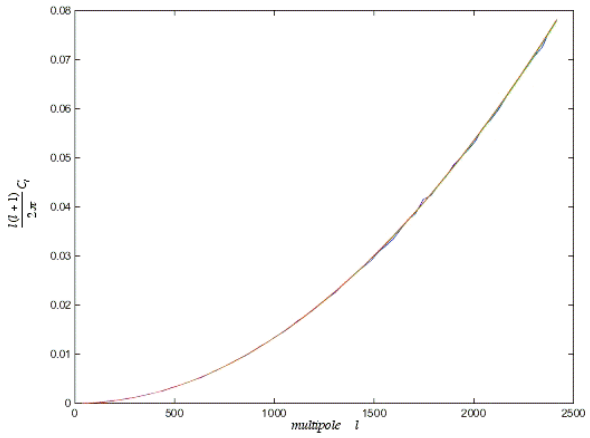


**Figure 8.** Averaged power spectrum of 100 CMB residual maps on the $(0°, 20°)$ sky patch. The curve is almost indistinguishable from the theoretical noise spectrum and the dispersion band.

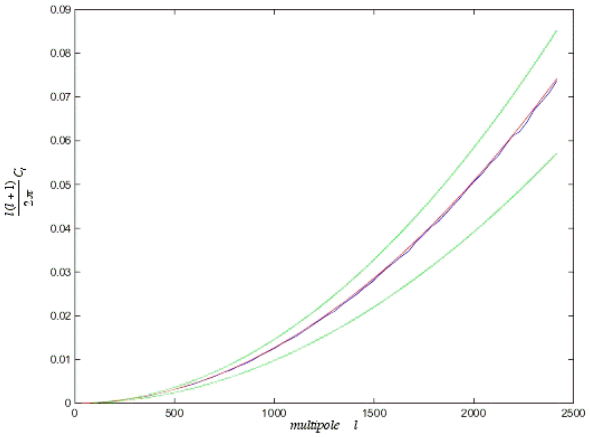


**Figure 9.** Averaged power spectrum of 100 CMB residual maps on the $(15°, 20°)$ sky patch (blue curve). The red curve represents the theoretical noise spectrum and the green curves delimit the dispersion band of the individual spectra.

improve the appearance of the final maps. Conversely, spectrum estimation has been found fairly insensitive to system noise, and this is a good point in favour of our algorithm. All the power spectra have been calculated on the output maps before the application of the Wiener filter.

## 5 REMARKS

Obviously, the semi-blind identification technique described here cannot be seen as a general approach to separation. However, it relies on a parametric knowledge of the emission spectra, a fairly common assumption in astrophysical data analysis. Our approach permits a very fast and robust model identification, thus enabling an accurate source estimation procedure to be implemented. We also envisage a method to estimate the source probability densities, which, in their turn, can help the separation task.

It has been observed that it does not make sense to try source separation in those regions where the foreground emissions are much smaller than CMB anisotropy and well below the noise level. As a matter of fact, in many of our experiments, we introduced a noise component which is stronger than the one expected from the Planck radiometers. In any case, the CMB angular power spectrum has always been estimated fairly well up to a multipole $l = 2000$. The estimation of the source densities has also given very good results. Source separation by our method has been particularly interesting with data from low galactic latitudes, where the foreground level is often higher than the CMB signal. Note that many separation strategies, both blind and non-blind, have failed their goal in this region of the celestial sphere. As an example, WMAP data analysis (see Bennett et al., 2003) was often performed by using pixel intensity masks that exclude the brightest sky portion from being considered. Another interesting feature is that, when a significant cross-correlation between a pair of sources is expected, this can be estimated as well by introducing an extra unknown in the problem. Our method is suitable to work directly with all-sky maps, but it could be necessary to apply it to small patches, as is shown in the above experimental section, to cope with the expected variability of the spectral indices and the noise variances in different sky regions.

We will further investigate this approach in the future, aiming at improving its identification and separation performances. One of the directions of research will be the use of covariance matrices at nonzero shift values, as already made by Belouchrani et al. (1997). This strategy should both improve the performance of the method with respect to uncorrelated noise and make it possible to relax the assumptions on the mixing matrix to be estimated. One problem that is still open with the expected Planck data is the different resolution of the data maps in some of the measurement channels. The identification part of our method can work with maps whose resolution has been degraded in order to be the same in all the channels. The result should be an estimate of the mixing matrix, which can be used in any non-blind separation approach with channel-dependent resolution, such as maximum entropy. However, the possible asymmetry of the telescope beam patterns should be taken into account in verifying this possibility.

## ACKNOWLEDGEMENTS

This work has been partially supported by the Italian Space Agency, under contract ASI/CNR 1R/073/01. DH is supported by the European Community's Human Potential Programme under contract HPRN-CT-2000-00124 CMBNET. The authors adopted the simulated sky templates provided by the Planck Technical Working Group 2.1 (diffuse component separation). In particular, the authors are grateful to Martin Reinecke (MPA), Vlad Stolyarov (Cambridge), Andrea Moneti, Simon Prunet and François Bouchet (IAP) for setting up and distributing the database. Extensive use of the HEALPix scheme (Hierarchical, Equal Area and iso-latitude pixelisation of the sphere, http://www.eso.org/science/healpix), by Krysztof M. Górski et al, has been made throughout this work.

## REFERENCES

Amari S. and Chichocki A., 1998, Proc. IEEE, 86, 2026
Attias H., 1999, Neural Computation, 11, 803
Baccigalupi C., Bedini L., Burigana C., De Zotti G., Farusi A., Maino D., Maris M., Perrotta F., Salerno E., Toffolatti L. and Tonazzini A., 2000, MNRAS, 318, 769
Baccigalupi C., Perrotta F., De Zotti G., Smoot G., Burigana C., Maino D., Bedini L. and Salerno E., 2002, astro-ph/0209591
Banday A. J. and Wolfendale A. W., 1991, MNRAS 252, 462
Barreiro R. B., Hobson M. P., Banday A. J., Lasenby A. N., Stolyarov V., Vielva P. and Górski K. M., 2003, astro-ph/0302091
Barros A. K. and Cichocki A., 2001, Neural Computation, 13, 1995
Belouchrani A., Abed-Meraim K., Cardoso J.-F. and Moulines E., 1997, IEEE Trans. on SP, 45, 434
Bennett C., Hill, R. S., Hinshaw G., Nolte M. R., Odegard N., Page L., Spergel D. N., Weiland J. L., Wright E. L., Halpern M., Jarosik N., Kogut A., Limon M., Meyer S. S., Tucker G. S. and Wollack E., 2003, ApJ, to appear, astro-ph/0302208
Bouchet F. R., Prunet S. and Sethi S. K., 1999, MNRAS, 302, 663
Cardoso J.-F., 1998, Proc. IEEE, 86, 2009
Cayón L., Sanz J. L., Barreiro R. B., Martínez-González E., Vielva P., Toffolatti L., Silk J., Diego J. M. and Argüeso F., MNRAS, 315, 757
Comon P., 1994, Signal Processing, 36, 287
Delabrouille J., Cardoso J.-F. and Patanchon G., 2002, astro-ph/0211504
De Zotti G., Toffolatti L., Argüeso F., Davies R. D., Mazzotta P., Partridge R. B., Smoot G. F., Vittorio N., 1999, 3K Cosmology, Proc. of the EC-TMR Conference, Woodbury, NY, American Institute of Physics, 476, 204
Giardino G. A., Banday A. J., Górski K. M., Bennett K., Jonas J. L. and Tauber J., 2002, A&A 387, 82
Hobson M. P., Jones A. W., Lasenby A. N. and Bouchet F. R., 1998, MNRAS, 300, 1
Hyvärinen A. and Oja E., 2000, Neural Networks, 13, 411
Kuruoğlu E.E., Bedini L., Paratore M. T., Salerno E. and Tonazzini A., 2003, Neural Networks, 16, 479
Maino D., Farusi A., Baccigalupi C., Perrotta F., Banday A. J., Bedini L., Burigana C., De Zotti G., Górski K. M. and Salerno E., 2002, MNRAS, 334, 53
Mandolesi N. et al., 1998, proposal submitted to ESA for the Planck Low Frequency Instrument
Moulines E., Cardoso J. F. and Gassiat E., 1997, Proc. ICASSP'97, 5, 3617
Puget J. L. et al., 1998, proposal submitted to ESA for the Planck High Frequency Instrument
Patanchon G., Snoussi H., Cardoso J.-F., Delabrouille J., 2003, astro-ph/0302078
Salerno E., Baccigalupi C., Bedini L., Burigana C., Farusi A., Maino D., Maris M., Perrotta F. and Tonazzini A., 2000, IEI-CNR, Pisa, Italy, Technical Report B4-04
Tegmark M., Eisenstein D. J., Hu W. and de Oliveira-Costa A., 2000, ApJ, 530, 133
Tenorio L., Jaffe A. H., Hanany S. and Lineweaver C. H., 1999, MNRAS, 310, 823
Tong L., Liu R., Soon V. C. and Huang Y.-F., 1991, IEEE Trans. on CAS, 38, 499
Vielva P., Martínez-González E., Cayón L., Diego J. M., Sanz J. L. and Toffolatti L., 2001, MNRAS, 326, 181
Yeredor A., 2002, IEEE Trans. on SP, 50, 1545