Mon. Not. R. Astron. Soc. **318**, 769–780 (2000)

Neural networks and the separation of cosmic microwave background and astrophysical signals in sky maps

C. Baccigalupi,[1★] L. Bedini,[2★] C. Burigana,[3★] G. De Zotti,[4★] A. Farusi,[2★] D. Maino,[5★] M. Maris,[5★] F. Perrotta,[1★] E. Salerno,[2★] L. Toffolatti[6★] and A. Tonazzini[2★]

[1]*SISSA/ISAS, Astrophysics Sector, Via Beirut, 4, I-34014 Trieste, Italy*
[2]*IEI-CNR, Via Alfieri, 1, I-56010 Ghezzano, Pisa, Italy*
[3]*ITeSRE-CNR, Via Gobetti, 101, I-40129 Bologna, Italy*
[4]*Osservatorio Astronomia Padova, Vicolo dell'Osservatorio 5, 35122 Padova, Italy*
[5]*Osservatorio Astronomia Trieste, Via G. B. Tiepolo, 11, I-34131 Trieste, Italy*
[6]*Departamento de Física, c. Calvo Sotelo s/n, 33007 Oviedo, Spain*

Accepted 2000 June 9. Received 2000 May 22; in original form 2000 February 21

ABSTRACT

We implement an independent component analysis (ICA) algorithm to separate signals of different origin in sky maps at several frequencies. Owing to its self-organizing capability, it works without prior assumptions on either the frequency dependence or the angular power spectrum of the various signals; rather, it learns directly from the input data how to identify the statistically independent components, on the assumption that all but, at most, one of the components have non-Gaussian distributions.

We have applied the ICA algorithm to simulated patches of the sky at the four frequencies (30, 44, 70 and 100 GHz) used by the Low Frequency Instrument of the European Space Agency's *Planck* satellite. Simulations include the cosmic microwave background (CMB), the synchrotron and thermal dust emissions, and extragalactic radio sources. The effects of the angular response functions of the detectors and of instrumental noise have been ignored in this first exploratory study. The ICA algorithm reconstructs the spatial distribution of each component with rms errors of about 1 per cent for the CMB, and 10 per cent for the much weaker Galactic components. Radio sources are almost completely recovered down to a flux limit corresponding to $\simeq 0.7\sigma_{\rm CMB}$, where $\sigma_{\rm CMB}$ is the rms level of the CMB fluctuations. The signal recovered has equal quality on all scales larger than the pixel size. In addition, we show that for the strongest components (CMB and radio sources) the frequency scaling is recovered with per cent precision. Thus, algorithms of the type presented here appear to be very promising tools for component separation. On the other hand, we have been dealing here with a highly idealized situation. Work to include instrumental noise, the effect of different resolving powers at different frequencies and a more complete and realistic characterization of astrophysical foregrounds is in progress.

Key words: methods: numerical – techniques: image processing – cosmic microwave background – radio continuum: general.

1 INTRODUCTION

Maps produced by large-area surveys, aimed at imaging primordial fluctuations of the cosmic microwave background (CMB), contain a linear mixture of signals from several astrophysical and cosmological sources (Galactic synchrotron; free–free and dust emissions, both from compact and diffuse sources; extragalactic sources; the Sunyaev–Zeldovich effect in clusters of galaxies; or by inhomogeneous re-ionization, in addition to primary and secondary CMB anisotropies) convolved with the spatial and spectral responses of the antenna and the detectors. In order to exploit the unique cosmological information encoded in the CMB anisotropy patterns, as well as the extremely interesting astrophysical information carried by the foreground signals, we need to accurately separate the different components.

★ E-mail: bacci@sissa.it (CBa); bedini@iei.pi.cnr.it (LB); burigana@tesre.bo.cnr.it (CBu); dezotti@pd.astro.it (GDZ); farusi@iei.pi.cnr.it (AF); maino@ts.astro.it (DM); maris@ts.astro.it (MM); perrotta@sissa.it (FP); salerno@iei.pi.cnr.it (ES); toffol@pinon.ccu.uniovi.es (LS); tonazzini@iei.pi.cnr.it (AT)

A great deal of work has been carried out in recent years in this area (see de Oliveira-Costa & Tegmark 1999, and references therein; Tegmark et al. 2000). The problem of map denoising has been tackled with wavelets analysis on both the whole sphere (Tenorio et al. 1999) and on sky patches (Sanz et al. 1999b). Algorithms to single out the CMB and the various foregrounds have been developed (Tegmark & Efstathiou 1996; Hobson et al. 1998; Bouchet, Prunet & Sethi 1999). In these works, Wiener filtering (WF) and the maximum entropy method (MEM) have been applied to simulated data from the *Planck* satellite, taking into account the expected performances of the instruments. Assuming a perfect knowledge of the frequency dependence of all the components, as well as priors for the statistical properties of their spatial pattern, these algorithms are able to recover the the strongest components at the best *Planck* resolution.

We adopt a rather different approach. We consider the denoising and deconvolution of the signals on one side, and the component separation on the other, as separate steps in the data analysis process and focus here on the latter step only, presenting a 'blind separation' method based on 'independent component analysis' (ICA) techniques. The method does not require any a priori assumption of the spectral properties or the spatial distribution of the various components, but requires only that they are statistically independent and that all but, at most, one have a non-Gaussian distribution. It is important to note that this is in fact the physical system that we have to deal with: surely all the foregrounds are non-Gaussian, whereas the CMB is expected to be a nearly Gaussian fluctuation field for most of the candidate theories of the early universe.

The paper is organized as follows. In Section 2 we introduce the relevant formalism and briefly review methods applied in previous works. In Section 3 we outline the ICA algorithm in a rather general framework as it may be useful for a variety of astrophysical applications. In Section 4 we describe our simulated maps. In Section 5 we give some details on our analysis and present the results. In Section 6 we draw our conclusions and list some future developments.

2 FORMALISM AND PREVIOUS APPROACHES

We assume that the frequency spectrum of radiation components (referred to as sources) is independent of the position in the sky. As we deal here with relatively small patches of the sky, we adopt Cartesian coordinates (ξ,η). The function describing the ith source is then written

$$\tilde{s}_i(\xi,\eta,\nu) = s_i(\xi,\eta)\cdot\mathcal{F}_i(\nu) \qquad i=1,\ldots,N, \tag{1}$$

where N is the number of independent sources and $\mathcal{F}_i(\nu)$ is the emission spectrum.

The signal received from the point (ξ,η) in the sky is

$$\tilde{x}(\xi,\eta,\nu) = \sum_{i=1}^{N} s_i(\xi,\eta)\cdot\mathcal{F}_i(\nu). \tag{2}$$

Suppose that the instrument has M channels with spectral response functions $t_j(\nu)$ $(j=1,\ldots M)$ centred at different frequencies, and that the beam patterns are independent of frequency within each passband. Let beam patterns be described by the $h_j(\xi,\eta)$ of the space-invariant point spread function, so that the maps are produced by a linear convolutional mechanism. (Note that this is an additional simplifying assumption because in real experiments a position dependent defocussing, related to the chosen scanning strategy, may occur.) Then the map yielded by the jth channel is

$$\begin{aligned} x_j(\xi,\eta) &= \int h_j(\xi-x,\eta-y)t_j(\nu) \\ &\quad\times\sum_{i=1}^{N} s_i(x,y)\mathcal{F}_i(\nu)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\nu + \epsilon_j(\xi,\eta) \\ &= \tilde{x}_j(\xi,\eta)*h_j(\xi,\eta)+\epsilon_j(\xi,\eta), \quad j=1,\ldots,M, \end{aligned} \tag{3}$$

where

$$\tilde{x}_j(\xi,\eta) = \sum_{i=1}^{N} a_{ji}\cdot s_i(\xi,\eta), \qquad j=1,\ldots,M, \tag{4}$$

$$a_{ji} = \int \mathcal{F}_i(\nu)t_j(\nu)\,\mathrm{d}\nu, \quad j=1,\ldots,M, \quad i=1,\ldots,N, \tag{5}$$

$*$ denotes linear convolution and $\epsilon_j(\xi,\eta)$ represents the instrumental noise. Equation (4) can also be written in matrix form:

$$\tilde{\boldsymbol{x}}(\xi,\eta) = \mathbf{A}\boldsymbol{s}(\xi,\eta) \tag{6}$$

where the entries of the $M\times N$ matrix $\mathbf{A}$ are given by equation (5).

The unknowns of our problem are the N functions $s_i(\xi,\eta)$, and the data set is made of the M maps $x_j(\xi,\eta)$ of equation (3). Besides the measured data, we also know the instrument beam patterns $h_j(\xi,\eta)$ and, more or less approximately (depending on the specific source), the coefficients a_{ji} of equation (4).

Equation (3) can be easily rewritten in Fourier space:

$$X_j(\omega_\xi,\omega_\eta) = \sum_{i=1}^{N} R_{ji}(\omega_\xi,\omega_\eta)S_i(\omega_\xi,\omega_\eta) + \mathcal{E}_j(\omega_\xi,\omega_\eta), \tag{7}$$

where the capital letters denote the Fourier transforms of the corresponding lowercase functions, and

$$R_{ji}(\omega_\xi,\omega_\eta) = \mathcal{H}_j(\omega_\xi,\omega_\eta)a_{ji}, \tag{8}$$

where $\mathcal{H}_j$ is the Fourier transform of the beam profile h_j. Equation (7) can thus be rewritten in matrix form:

$$\boldsymbol{X} = \mathbf{R}\boldsymbol{S} + \boldsymbol{\mathcal{E}}. \tag{9}$$

The above equation must be satisfied by each Fourier mode (ω_ξ,ω_η) independently. The aim is to recover the true signals $S_i(\omega_\xi,\omega_\eta)$ that constitute the column vector $\boldsymbol{S}$. If the matrix $\mathbf{A}$ in equation (6) is known exactly then, in the absence of noise, the problem reduces to a linear inversion of equation (9) for each Fourier mode.

In practice, however, $\mathcal{H}_j$ vanishes for some Fourier modes. For these modes the entire jth row of the matrix R also vanishes, and R may become a non-full-rank matrix. An inversion based on statistical approaches built on a priori knowledge is thus needed.

In the following two subsections we briefly describe two such approaches, and in the third subsection we briefly recall a technique so far mostly exploited for the denoising problem and for the extraction of extragalactic sources.

2.1 The MEM approach

The MEM for the reconstruction of images is based on a Bayesian approach to the problem (Gull 1988; Skilling 1988, 1989). Let $\boldsymbol{X}$ be a vector of M observations, the probability distribution $P(\boldsymbol{X}|\boldsymbol{S})$ of which depends on the values of N quantities $\boldsymbol{S}=S_1,\ldots,S_N$.

Let $P(\boldsymbol{S})$ be the *prior* probability distribution of $\boldsymbol{S}$, which tells us

what is known about $\boldsymbol{S}$ without knowledge of the data. Given the data $\boldsymbol{X}$, Bayes' theorem states that the conditional distribution of $\boldsymbol{S}$ (the *posterior* distribution of $\boldsymbol{S}$) is given by the product of the likelihood of the data, $P(\boldsymbol{X}|\boldsymbol{S})$, with the prior:

$$P(\boldsymbol{S}|\boldsymbol{X}) = zP(\boldsymbol{X}|\boldsymbol{S})P(\boldsymbol{S}), \tag{10}$$

where z is a normalization constant.

An estimator $\hat{\boldsymbol{S}}$ of the true signal vector can be constructed by maximizing the posterior probability $P(\boldsymbol{S}|\boldsymbol{X}) \propto P(\boldsymbol{X}|\boldsymbol{S})P(\boldsymbol{S})$. However, although the likelihood in equation (10) is easily determined once the noise and signal covariance matrices are known, the appropriate choice of the prior distribution for the model considered is a major problem in the Bayesian approach: as Bayes' theorem is simply a rule for manipulating probabilities, it cannot by itself help us to assign them in the first place, so one has to look elsewhere. The MEM is a consistent variational method for the assignment of probabilities under certain types of constraints that must refer to the probability distribution directly.

The maximum entropy principle states that if one has some information I on which the probability distribution is based, one can assign a probability distribution to a proposition i such that $P(i|I)$ contains only the information I that one actually possesses. This assignment is performed by maximizing the entropy:

$$H \equiv -\sum_{i=1}^{N} P(i|I)\log P(i|I). \tag{11}$$

It can be seen that when nothing is known except that the probability distribution should be normalized, the maximum entropy principle yields the uniform prior. In our case the proposition i represents $\boldsymbol{S}$, and the information I is the assumption of signal statistical independence. The standard application of the method considered strictly positive signals (Gull 1988; Skilling 1988, 1989); the extension to the case of CMB temperature fluctuations, which can be both positive and negative, was worked out by Hobson et al. (1998).

The construction of the entropic prior requires, in general, the knowledge of the frequency dependence of the components to be recovered as well as that of the signal covariance matrix $\mathbf{C}(\boldsymbol{k}) = \langle \boldsymbol{S}(\boldsymbol{k})\boldsymbol{S}^{\dagger}(\boldsymbol{k})\rangle$, with the average taken on all the possible realizations.

2.2 The multifrequency WF

If a Gaussian prior is adopted, the Bayesian approach gives the multifrequency WF solution (Bouchet et al. 1999). In this case also, an estimator of the signal vector is obtained by maximizing the posterior probability in equation (10) given the signal covariance matrix $\mathbf{C}(\boldsymbol{k})$.

The Gaussian prior probability distribution for the signal has the form

$$P(\boldsymbol{S}) \propto \exp(-\boldsymbol{S}^{\dagger}\mathbf{C}^{-1}\boldsymbol{S}). \tag{12}$$

The estimator $\hat{\mathbf{S}}$ is linearly related to the data vector $\hat{\boldsymbol{X}}$ through the Wiener matrix $\mathbf{W} \equiv (\mathbf{C}^{-1} + \mathbf{R}^{\dagger}\mathbf{N}^{-1}\mathbf{R})^{-1}$, where $\mathbf{R}$ corresponds to the matrix in equation (9) and $\mathbf{N}(\boldsymbol{k}) = \langle \boldsymbol{\epsilon}(\boldsymbol{k})\boldsymbol{\epsilon}^{\dagger}(\boldsymbol{k})\rangle$ is the noise covariance matrix,

$$\hat{\boldsymbol{S}} = \mathbf{W}\boldsymbol{X}. \tag{13}$$

The $\mathbf{W}$ matrix has the role of a linear filter; again, its construction requires an a priori knowledge of the spectral behaviour of the signals.

This method is endangered by the clear non-Gaussianity of the foregrounds.

2.3 Wavelet methods

The development of wavelet techniques for signal processing has been very rapid in the last ten years (see e.g. Jawerth & Sweldens 1994). The wavelet approach is conceptually very simple: whereas the Fourier transform is highly inefficient in dealing with the local behaviour, the wavelet transform is able to introduce a good space–frequency localization, thus providing information on the contributions coming from different positions and scales.

In one dimension, we can define the *analysing* wavelet as $\Psi(x; R, b) \equiv R^{-1/2}\psi[(x-b)/R]$, which is dependent on two parameters: the dilation (R) and translation (b). $\psi(x)$ is a one-dimensional function satisfying the following conditions: (i) $\int_{-\infty}^{\infty} \mathrm{d}x\psi(x) = 0$; (ii) $\int_{-\infty}^{\infty} \mathrm{d}x\psi^2(x) = 1$; and (iii) $\int_{-\infty}^{\infty} \mathrm{d}k|k|^{-1}\psi^2(k) < \infty$, where $\psi(k)$ is the Fourier transform of $\psi(x)$. The wavelet Ψ operates as a mathematical microscope of magnification R^{-1} at the space point b. The wavelet coefficients associated to a one-dimensional function $f(x)$ are

$$w(R, b) = \int \mathrm{d}x f(x)\Psi(x; R, b). \tag{14}$$

The computationally faster algorithms for the wavelet analysis of two-dimensional images are the algorithms based on multiresolution analysis (Mallat 1989), or on 2D wavelet analysis (Lemarié & Meyer 1986) using tensor products of one-dimensional wavelets. The discrete Multiresolution analysis entails the definition of a one-dimensional *scaling* function ϕ, normalized as $\int_{-\infty}^{\infty} \mathrm{d}x\phi(x) = 1$ (Ogden 1997). Scaling functions act as low-pass filters whereas wavelet functions single out one scale. The 2D wavelet method (Sanz et al. 1999b) is based on two scales, and therefore provides more information on different resolutions (defined by the product of the two scales) than is provided by the multiresolution method.

Recently, wavelet techniques have been introduced in the analysis of CMB data. Denoising of CMB maps has been performed on patches of the sky of $12\overset{\circ}{.}8 \times 12\overset{\circ}{.}8$, using either multiresolution techniques (Sanz et al. 1999a) or 2D wavelets (Sanz et al. 1999b), as well as on the whole celestial sphere (Tenorio et al. 1999). As a first step, maps with the cosmological signal plus Gaussian instrumental noise have been considered.

Denoising of CMB maps has been carried out by using a signal-independent prescription: the SURE thresholding method (Donoho & Johnstone 1995). The results are model independent and only a good knowledge of the noise affecting the observed CMB maps is required, whereas nothing has to be assumed about the nature of the underlying field(s). Moreover, wavelet techniques are highly efficient in localizing noise variations and features in the maps.

The wavelet method is able to improve the signal-to-noise ratio by a factor of 3–5; correspondingly, the error on the C_ℓ values derived from denoised maps is about two times lower than that obtained with the WF method.

Wavelets were also successfully applied to the detection of point sources in CMB maps in the presence of the cosmological signal and instrumental noise (Tenorio et al. 1999); more recently, successful results on source detection have also been obtained in the presence of diffuse galactic foregrounds (Cayón et al. 2000). The results are comparable to the results obtained with the filtering method presented by (Tegmark & de Oliveira-Costa

1998), which, however, rely on the assumption that all the underlying fields are Gaussian.

3 THE ICA APPROACH

We present here a rather different approach, which is characterized by the capability of working 'blindly' i.e. by working without prior knowledge of the spectral and spatial properties of the signals to be separated. The method is of interest for a broad variety of signal and image processing applications: whenever a number of source signals are detected by multiple transducers and the transmission channels for the sources are unknown, so that each transducer receives a mixture of the source signals with unknown scaling coefficients and channel distortion.

In this exploratory study we confine ourselves to the case of simple linear combinations of unconvolved source signals (Bell & Sejnowski 1995; Amari & Chichocki 1998). The problem can be stated as follows: a set of N signals is inputted to an unknown frequency dependent multiple-input-multiple-output linear instantaneous system, the M outputs of which are our observed signals. We use the term *instantaneous* to denote a system the output of which at a given point only depends on the input signals at the same point. Our objective is to find a stable *reconstruction system* to estimate the original input signals with no prior assumptions either about the signal distributions or about their frequency scalings. The problem in its general form is normally unsolvable, and a 'working hypothesis' must be made. The hypothesis we make is that our source signals are mutually *statistically independent*, whatever their actual distributions are. Several solutions have been proposed for this problem, each based on more or less sound principles, not all of which are typical of classical signal processing. Indeed, information theory, neural networks, statistics and probability have played an important part in the development of these techniques.

We do not consider here specific instrumental features like beam convolution and noise contamination, leaving the specialization of the ICA method to specific experiments for future work; this allows us to highlight the capabilities of this approach, which is able to work in conditions where other algorithms would not be viable. Therefore, we adopt equation (6) as our data model by just dropping the tilde accent on vector $\boldsymbol{x}$. Also, the instrumental noise term in equation (7) will be neglected.

It can be proved that, to solve the problem described above, the following hypotheses should be verified (Comon 1994; Amari & Chichocki 1998):

(i) all source signals are statistically independent;
(ii) at most one of the signals has a Gaussian distribution;
(iii) that $M \geqslant N$;
(iv) low noise.

The last two assumptions can be somewhat relaxed by choosing suitable separation strategies. As far as independence is concerned, roughly speaking, we may say that the search for an ICA model from non-ICA data (i.e. data not coming from independent sources) should give the most 'interesting' (namely, the most structured) projections of the data (Friedman 1987; Hyvärinen & Oja 1999). This is not equivalent to saying that separation is achieved; however, we have seen from our experiments that a good separation can be obtained even for sources that are not totally independent. The second assumption above tells us that Gaussian sources cannot be separated. More specifically, they can only be separated up to an orthogonal transformation. In fact, it can be shown that the joint probability of a mixture of Gaussian signals is invariant to orthogonal transformations. This means that if independent components are found from Gaussian mixtures, then any orthogonal transformation of them gives mutually independent components.

Many strategies have been adopted to solve the separation problem on the basis of the above hypotheses, all of which were based on looking for a set of independent signals that can be shown to be the original sources. A formal criterion to test independence, from which all the separating strategies can be derived, is described later in this section.

In order to recover the original source signals from the observed mixtures, we use a separating scheme in the form of a feed-forward neural network. The observed signals are input to an $N \times M$ matrix $\mathbf{W}$, referred to as the *the synaptic weight matrix*, the adjustable entries of which (w_{ij}, $i = 1, \ldots, N$ and $j = 1, \ldots, M$) are updated for every sample of the input vector $\boldsymbol{x}(\xi, \eta)$ (at step τ) following a suitable *learning algorithm*. The output of matrix $\mathbf{W}$ at step τ will be

$$\boldsymbol{u}(\xi, \eta, \tau) = \mathbf{W}(\tau)\boldsymbol{x}(\xi, \eta). \tag{15}$$

$\mathbf{W}(\tau)$ is expected to converge to a true separating matrix, that is a matrix the output of which is a *copy* of the inputs for every point (ξ,η). Ideally, this final matrix $\mathbf{W}$ should be such that $\mathbf{WA} = \mathbf{I}$, where $\mathbf{I}$ is the $N \times N$ identity. As an example, if $M = N$, we should have $\mathbf{W} = \mathbf{A}^{-1}$. There are, however, two basic indeterminacies in our problem: ordering and scaling. Even if we are able to extract N independent sources from M linear mixtures, we cannot know a priori the order in which they will be arranged, because this corresponds to unobservable permutations of the columns of matrix $\mathbf{A}$. Moreover, the scales of the extracted signals are unknown, because when a signal is multiplied by some scalar constant, the effect is the same as of multiplying by the same constant the corresponding column of the mixing matrix. This means that $\mathbf{W}(\tau)$ will converge, at best, to a matrix $\mathbf{W}$ such that

$$\mathbf{WA} = \mathbf{PD}, \tag{16}$$

where $\mathbf{P}$ is any $N \times N$ permutation matrix, and $\mathbf{D}$ is a non-singular diagonal scaling matrix. From equations (6), (15) and (16) we thus have

$$\boldsymbol{u} = \mathbf{W}\boldsymbol{x} = \mathbf{WA}\boldsymbol{s} = \mathbf{PD}\boldsymbol{s}. \tag{17}$$

That is, as anticipated, each component of $\boldsymbol{u}$ is a scaled version of a component of $\boldsymbol{s}$, not necessarily in the same order. This is not a serious inconvenience in our application because we should be able to recover the proper scales for the separated sources from other pieces of information, for example matching with independent lower resolution observations like those of *Cosmic Background Explorer* (*COBE*) in the case of *Microwave Anisotropy Probe* (*MAP*) and *Planck*. If $\mathbf{A}$ was known, the performance of the separation algorithm could be evaluated by means of the matrix $\mathbf{WA}$. If the separation is perfect, this matrix has only one non-zero element for each row and each column. In any non-ideal situation each row and column of $\mathbf{WA}$ should contain only one dominant element.

In all the cases treated here we assume $M \geqslant N$, but we consider the case where N, although smaller than M, is not known.

The mutual statistical independence of the source signals can be expressed in terms of a separable joint probability density function $q(\boldsymbol{s})$:

$$q(\boldsymbol{s}) = \prod_{j=1}^{N} q_j(s_j), \tag{18}$$

where $q_j(s_j)$ is the marginal probability density of the jth source.

Various algorithms can be used to obtain the matrix **W**. All these algorithms can be derived from a unified principle based on the Kullback–Leibler (KL) divergence between the joint probability density of the output vector $\boldsymbol{u}$, $p_{\boldsymbol{u}}(\boldsymbol{u})$, and a function $q(\boldsymbol{u})$, which should be suitably chosen among the functions of the type in equation (18). The KL divergence between the two functions mentioned above may be written as a function of the matrix **W**, and can be considered as a cost function in the sense of Bayesian statistics:

$$\mathbf{R}(\mathbf{W}) = \int p_{\boldsymbol{u}}(\boldsymbol{u}) \log \frac{p_{\boldsymbol{u}}(\boldsymbol{u})}{q(\boldsymbol{u})} \mathrm{d}\boldsymbol{u}. \tag{19}$$

It can be proved that, under mild conditions on $q(\boldsymbol{u})$, $\mathbf{R}(\mathbf{W})$ has a global minimum where **W** is such that $\mathbf{WA} = \mathbf{PD}$. The different possible choices for $q(\boldsymbol{s})$ lead to the different particular learning strategies proposed in the literature (Bell & Sejnowski 1995; Yang & Amari 1997; Amari & Chichocki 1998).

The *uniform gradient* search method, which is a gradient-type algorithm, takes into account the Riemannian metric structure of our objective parameter space, which is the set of all non-singular matrices **W** (Amari & Chichocki 1998). In a general case, where the number N of sources is only known to be smaller than the number of observations, the following formula is derived:

$$\mathbf{W}(\tau+1) = \mathbf{W}(\tau) + \alpha(\tau) \times \{\mathbf{\Lambda} - \boldsymbol{u}(\tau)\boldsymbol{u}^{\mathrm{T}}(\tau) - \boldsymbol{f}[\boldsymbol{u}(\tau)]\boldsymbol{u}^{\mathrm{T}}(\tau)\}\mathbf{W}(\tau), \tag{20}$$

where $\mathbf{\Lambda}$ is an $M \times M$ diagonal matrix,

$$\mathbf{\Lambda} = \mathrm{diag}\{[u_1 + f_1(u_1)]u_1\} \ldots \{[u_M + f_M(u_M)]u_M\}. \tag{21}$$

Pixel by pixel, the $M \times M$ matrix **W** is multiplied by the M-vector $\boldsymbol{x}$, and gives vector $\boldsymbol{u}$ as its output. This output is transformed through the non-linear vector function $\boldsymbol{f}(\boldsymbol{u})$, and the result is combined with $\boldsymbol{u}$ itself to build the update to matrix **W** through equation (20). The process has to be iterated by reading the data maps several times. If N is strictly smaller than M, then $M - N$ outputs can be shown to rapidly converge to zero, or to pure noise functions.

The convergence properties of this iterative formula are shown to be independent of the particular matrix **A**, so that even a strongly ill-conditioned system does not affect the convergence of the learning algorithm. In other words, even when the contributions from some components are very small, there is no problem to recover the contributions. This property is called the *equivariant property* because the asymptotic properties of the algorithm are

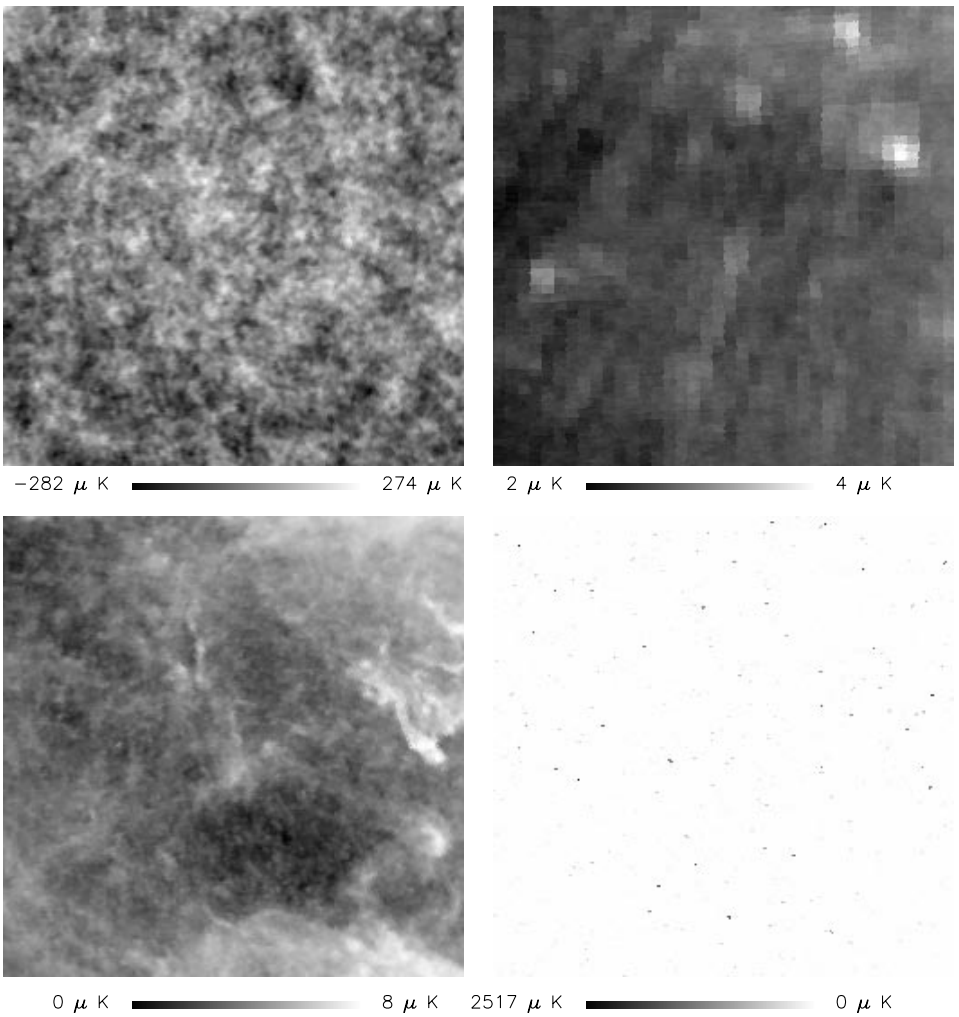

Figure 1. Input maps used in the ICA separation algorithm: from top left in a clockwise sense, simulations of CMB, synchrotron, radio sources and dust emission are shown. Radio sources and dust grey-scales are non-linear to better show the signal features.

independent of the mixing matrix. The τ-dependent parameter α is the *learning rate*; its value is normally decreased during the iteration. As far as the choice of $\alpha(\tau)$ is concerned, a strategy to learn it and its annealing scheme is given in (Amari & Chichocki 1998); we have chosen $\alpha(\tau)$ to decrease from 10^{-3} to 10^{-4} linearly with the number of iterations.

The final problem is how to choose the function $\boldsymbol{f}(\boldsymbol{u})$. If we know the true source distributions $q_j(u_j)$, the best choice is to make $f'_j(u_j) = q_j(u_j)$, because this gives the maximum likelihood estimator. However, the point is that when $q_j(u_j)$ are specified incorrectly, the algorithm gives the correct answer under certain conditions. In any case, the choice of $\boldsymbol{f}(\boldsymbol{u})$ should be made to ensure the existence of an equilibrium point for the cost function and the stability of the optimization algorithm. These requirements can be satisfied even though the non-linearities chosen are not optimal. A suboptimal choice for sub-Gaussian source signals (negative kurtosis) is

$$f_i(u_i) = \beta u_i + u_i|u_i|^2, \tag{22}$$

and for super-Gaussian source signals (positive kurtosis):

$$f_i(u_i) = \beta u_i + \tanh(\gamma u_i), \tag{23}$$

where $\beta \geqslant 0$ and $\gamma \geqslant 2$. If one source is Gaussian, the above choices remain viable as well. In our case, we verified that all the source functions except CMB are super-Gaussian, and thus we implemented the learning algorithm following equation (20) with the non-linearities in equation (23), $\beta = 0$ and $\gamma = 2$. As already stated, the mean of the input signal at each frequency is subtracted. In previous work (Yang & Amari 1997) the initial matrix was chosen as $\mathbf{W} \propto \mathbf{I}$; in that analysis, the image data consisted of a set of components with nearly the same amplitude. The initial guess for $\mathbf{W}$ affects the computation time as well as the scaling of the reconstructed signals and their order. Interestingly, we found that adjusting the diagonal elements so that they roughly reflect the different weights of the components in the mixture can speed-up the convergence. For the problem at hand, the results shown in Section 5 have been obtained starting from $\mathbf{W} = \mathrm{diag}[1, 3, 30, 10]$, for the case of a 4×4 $\mathbf{W}$ matrix, and using only 20 learning steps: the time needed was about 1 min on an UltraSparc machine (equipped with a 300-MHz UltraSparc processor with 256 Mb RAM, running on a SUN Solaris 7 operating system), which compiled the FORTRAN 90 code using SUN Fortran Workshop 5.0.

4 SIMULATED MAPS

We produced simulated maps of the antenna temperature

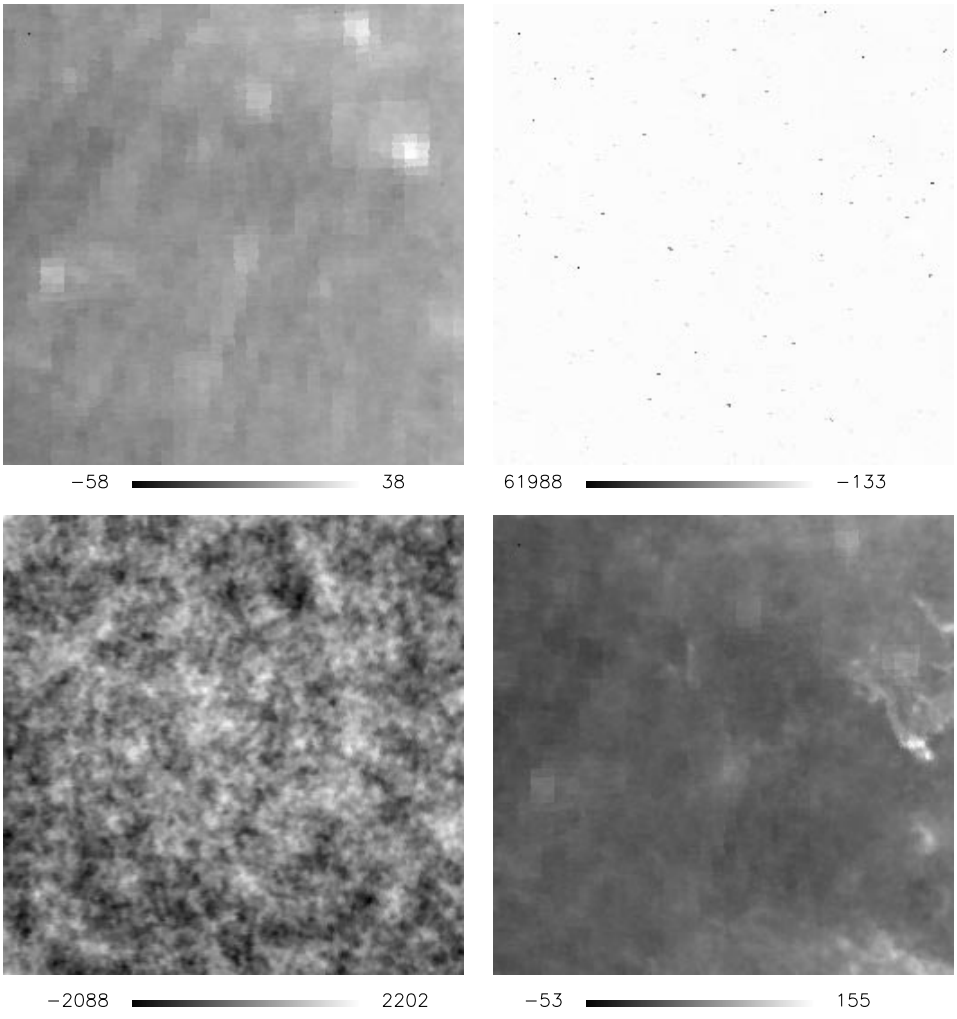

Figure 2. Reconstructed maps produced by the ICA method; the initial ordering has not been conserved in the outputs. From top left, in a clockwise sense, we can recognize synchrotron, radio sources, dust and CMB. Radio sources and dust grey-scales are non-linear as in Fig. 1.

distribution, using a pixel size of 3.5 arcmin for a $15^\circ \times 15^\circ$ region centred at $l = 90^\circ$ and $b = 45^\circ$, at the four central frequencies of the *Planck* Low Frequency Instrument (LFI) channels (Mandolesi et al. 1998), namely 30, 44, 70 and 100 GHz (Fig. 1). The HEALPix pixelization scheme (see Górski et al. 1999) was adopted. The maps include CMB anisotropies, Galactic synchrotron and dust emissions, and extragalactic radio sources.

CMB fluctuations correspond to a flat cold dark matter (CDM) model ($\Omega_{\rm CDM} = .95$, $\Omega_b = .05$ and three massless neutrino species), normalized to the *COBE* data (see Seljak & Zaldarriaga 1996). As it is well known, the CMB spectrum, in terms of antenna temperature, is written:

$$s^{\rm antenna}_{\rm CMB}(\xi, \eta, \nu) = s^{\rm therm}_{\rm CMB}(\xi, \eta) \frac{\tilde{\nu}^2 \exp(\tilde{\nu})}{[\exp(\tilde{\nu}) - 1]^2}, \quad (24)$$

where $\tilde{\nu} = \nu/56.8$, ν is the frequency in GHz and $s^{\rm therm}_{\rm CMB}(\xi, \eta)$ is frequency independent (Fixsen et al. 1996).

As for Galactic synchrotron emission, we have extrapolated the 408-MHz map with about 1° resolution (Haslam et al. 1982), assuming a power law spectrum, in terms of antenna temperature:

$$\mathcal{F}_{\rm syn} \propto \tilde{\nu}^{-n_{\rm s}}, \quad (25)$$

with spectral index $n_{\rm s} = 2.9$.

The dust emission maps with about 6 arcmin resolution constructed by Schlegel, Finkbeiner & Davies (1998) combining *IRAS* and *DIRBE* data have been used as templates for Galactic dust emission. The extrapolation to *Planck*/LFI frequencies was performed assuming a grey-body spectrum:

$$\mathcal{F}_{\rm dust} \propto \frac{\tilde{\nu}^{m+1}}{\exp(\tilde{\nu}) - 1}, \quad (26)$$

with $m = 2$, $\tilde{\nu} = h\nu/kT_{\rm dust}$, $T_{\rm dust}$ being the dust temperature. Although in general $T_{\rm dust}$ varies across the sky, it turns out to be approximately constant at about 18 K in the region considered here; we have therefore adopted this value in the above equation.

Because of the lack of a suitable template we have ignored here free–free emission, which may be important particularly at 70 and 100 GHz. This component needs to be included in future work.

The model by Toffolatti et al. (1998) was adopted for extragalactic radio sources, which were assumed to have a Poisson distribution. An antenna temperature spectral index $n_{\rm rs} = 1.9$ was adopted ($\mathcal{F}_{\rm rs} \propto \tilde{\nu}^{-n_{\rm rs}}$).

5 BLIND ANALYSIS AND RESULTS

As it is well known, the strongest signals at the *Planck*/LFI frequencies come from the CMB and from radio sources (although the latter show up essentially as a few high peaks), whereas synchrotron emission and thermal dust signals are roughly one or two orders of magnitude reduced in strength, depending on the frequency. Thus we are testing the performances of the ICA algorithm with four signals exhibiting very different spatial patterns, frequency dependences and amplitudes.

As we are interested in the fluctuation pattern, the mean of the total signal (sum of the four components) is set to zero at each

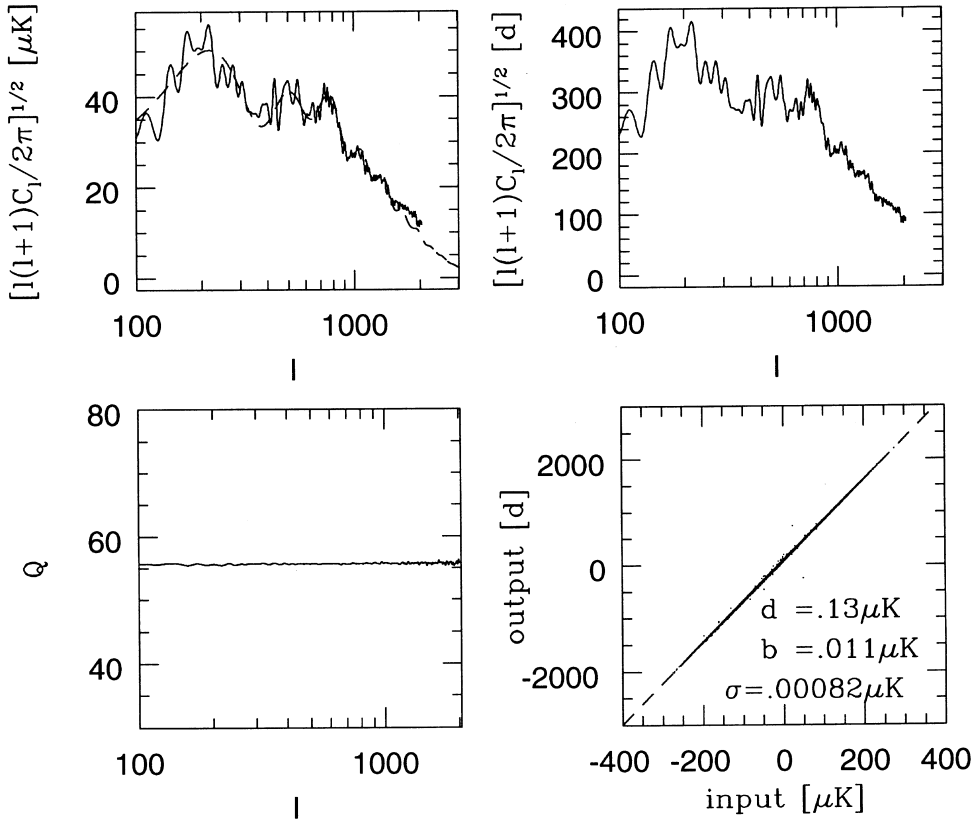

Figure 3. Top left: input angular power spectra – simulated (solid line) and theoretical (dashed line, see text). Top right: the angular power spectrum of the reconstructed CMB patch. Bottom left: quality factor relative to the input/output angular spectra. Bottom right: scatter plot and linear fit (dashed line) for the CMB input/output maps.

frequency. We adopt a 'blind' approach: no information on either the spatial distribution or the frequency dependence of the signals is provided for the algorithm.

The reconstructed maps of the the four components are shown in Fig. 2. Several interesting features may be noticed. The order of the plotted maps is permuted with respect to the input maps in Fig. 1, reflecting the order of the ICA outputs: the first output is synchrotron, the second represents radio sources, the third is CMB and the fourth is dust. All the output maps look very similar in comparison with the true ones; even synchrotron lower resolution pixels have been reproduced. In Figs 3, 4, 5 and 6 we analyze the goodness of the separation by comparing power spectra and showing plots of the scatter between the inputs and the outputs.

5.1 Signal reconstruction

For each map we have computed the angular power spectrum defined by the expansion coefficients C_ℓ of the two point correlation function in Legendre polynomials. As is well known, it can conveniently be expressed in terms of the coefficients of the expansion of the signal S into spherical harmonics, $S(\theta, \phi) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\theta, \phi)$:

$$C_\ell = \frac{1}{2\ell + 1} \sum_m |a_{\ell m}|^2. \tag{27}$$

Such coefficients are useful because, from elementary properties of the Legendre polynomials, it can be seen that the coefficient C_ℓ quantifies the amount of perturbation on the angular scale θ given by $\theta \simeq (180/\ell)^\circ$.

The panels on the top of Figs 3, 4, 5 and 6 show the power spectra of the input (left) and output (right) signals. The CMB exhibits the characteristic peaks on subdegree angular scales as a result of acoustic oscillations of the photon–baryon fluid at decoupling; the dashed line represents the theoretical model from which the map was generated, whereas the solid line is the power spectrum of our simulated patch. The difference between the two curves is caused by the sample variance corresponding to the CMB Gaussian statistics. Radio sources are completely different because of their point-like structure and shot noise spatial distribution (Mandolesi et al. 1998; Puget et al. 1998). The bottom left-hand panels show the quality factor, defined as the ratio between true and reconstructed power spectrum coefficients, for each multipole ℓ. Owing to the limited size of the analysed region, the power spectrum can be defined on scales below roughly 2°. The bottom right-hand panels are scatter plots of the ICA results: for each pixel of the maps, we plotted the value of the reconstructed image versus the corresponding input value.

The reconstructed signals have a mean of zero and are in units of the constant d produced during the separation phase, as described in Section 3: the scale of each signal is unreproducible for a blind separation algorithm such as the ICA. Nevertheless, a lot of information is encoded into the spatial pattern of each signal, and ultimately the overall normalization of the signal could be recovered by exploiting data from other experiments. Therefore, the relation between each true signal and its reconstruction is

$$s_i^{\rm in} = d s_i^{\rm out} + b, \qquad i = 1, \ldots, N_{\rm pixels}, \tag{28}$$

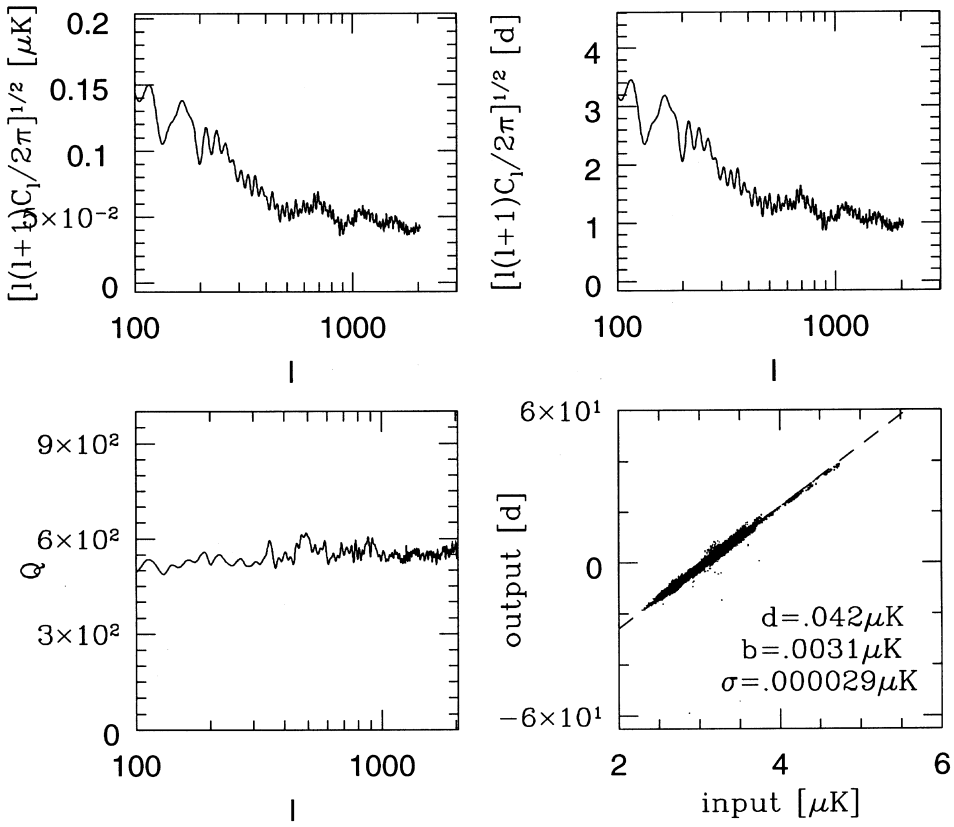

Figure 4. Top panels: angular power spectra for the simulated input (left) and reconstructed (right) synchrotron maps. Bottom left: quality factor relative to the input/output angular spectra. Bottom right: scatter plot and linear fit (dashed line) for the synchrotron input/output maps.

where b merely represents the mean of the input signal, which is zero for the CMB and some positive value for the foregrounds.

To quantify the quality of the reconstruction, we have recovered d and b by performing a linear fitting of output to input maps ($\boldsymbol{s}^{\rm in}$, $\boldsymbol{s}^{\rm out}$) for each signal:

$$d = \frac{\sum_i s_i^{\rm in} s_i^{\rm out} - \bar{s}^{\rm in}\sum_i s_i^{\rm out}}{\sum_i (s_i^{\rm out})^2 - \bar{s}^{\rm out}\sum_i s_i^{\rm out}}, \qquad b = \bar{s}^{\rm in} - d\bar{s}^{\rm out}, \tag{29}$$

where the sums run over all the pixels and the bar indicates the average value over the patch; the values of d and b, as well as the linear fits (dashed lines), are indicated for all the signals in the scatter plot panels. Also, in the same panels, we show the standard deviation of the fit, that is

$$\sigma = \left[\frac{1}{N_{\rm pixels}}\sum_i (s_i^{\rm in} - d s_i^{\rm out} - b)^2\right]^{1/2}. \tag{30}$$

A comparison of this quantity with the input signals (bottom right-hand panels) gives an estimate of the goodness of the reconstruction. CMB and radio sources are recovered with 1 and 0.1 per cent precision, respectively, whereas the accuracy drops roughly to 10 per cent for the (much weaker) Galactic components, synchrotron and dust. Also, the latter appear to be slightly mixed; this is likely caused by the fact that they are somewhat correlated so that the hypothesis of statistical independence is not properly satisfied.

We have also tested to what extent the counts of radio sources are recovered. This was performed in terms of the relative flux

$$\Delta s = s/s_{\rm max}, \tag{31}$$

$s_{\rm max}$ being the flux of the brightest source.

In Fig. 7 we show the cumulative number of input (dashed) and output (solid line) pixels exceeding a given value of Δs. The algorithm correctly recovers essentially all sources with $\Delta s \geqslant 2\times10^{-2}$, corresponding to a signal of $T_s \simeq 50\,\mu$K, or to a flux density $S = (2k_{\rm B}T_s/\lambda^2)\Delta\Omega \simeq 15$ mJy, where $k_{\rm B}$ the Boltzmann constant, λ the wavelength and $\Delta\Omega$ the solid angle covered by a pixel that is $3.5\times3.5\,{\rm arcmin}^2 \simeq 10^{-6}$ sr. At fainter fluxes the counts are overestimated; this is probably caused by contamination from the other signals. In any case, the flux limit for source detection is surprisingly low, even lower than the rms CMB fluctuations ($\sigma_{\rm CMB} \simeq 70\,\mu$K at the resolution limit of our maps), and substantially lower or at least comparable to that achieved with other methods that require stronger assumptions (Hobson et al. 1999; Cayón et al. 2000). This high efficiency in detecting point sources illustrates the ability of the method in taking the maximum advantage of the differences in frequency and spatial properties of the various components.

On the other hand, we stress that our approach is idealized in a number of aspects: beam convolution and instrumental noise have not been taken into account, and the same frequency scaling has been assumed for all radio sources. Therefore, more detailed investigations are needed to estimate a realistic source detection limit.

Finally, note that the quality of the separation is similar on all scales, as shown by the bottom left-hand side panels of Figs 3, 4, 5

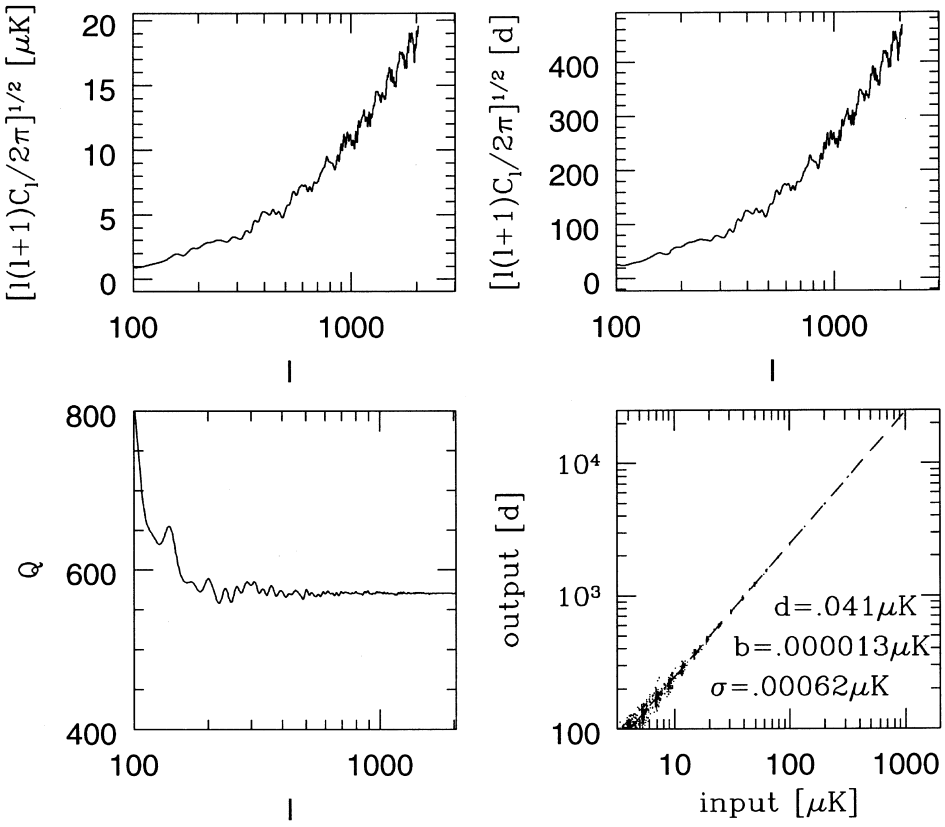

Figure 5. Top panels: angular power spectra for the simulated input (left) and reconstructed (right) dust emission maps. Bottom left: quality factor relative to the input/output angular spectra. Bottom right: scatter plot and linear fit (dashed line) for the dust input/output maps.

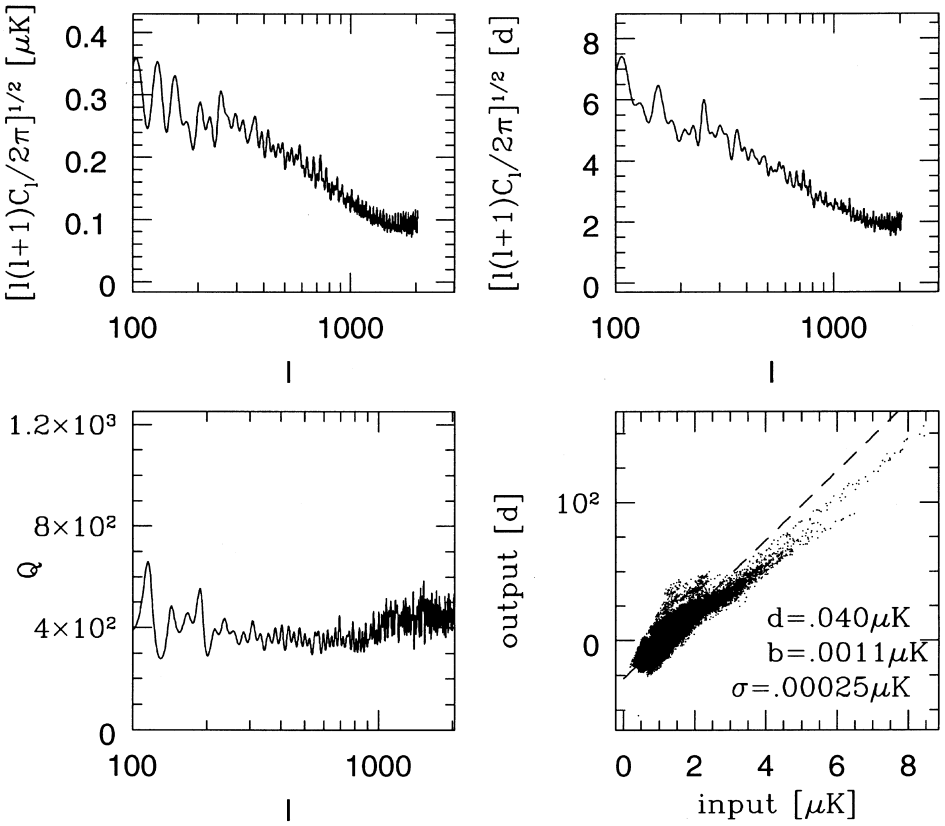

Figure 6. Top panels: angular power spectra for the simulated (left) and reconstructed (right) radio source map. Bottom left: quality factor relative to the input/output angular spectra. Bottom right: scatter plot and linear fit (dashed line) for the radio source emission input/output maps.

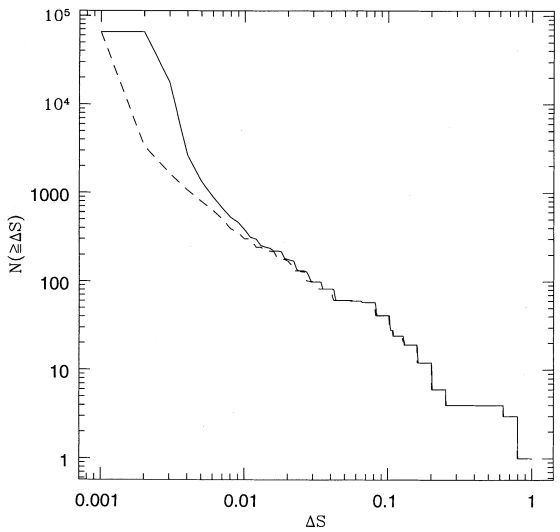

Figure 7. Cumulative number of pixels as a function of the threshold Δs (see text for more details): input (dashed line) versus output (solid line).

and 6. The exception are radio sources, the true power spectra of which go to zero at low ℓs more rapidly than the spectrum of the reconstructed one.

5.2 Reconstruction of the frequency dependence

Another asset of this technique is the possibility of recovering the frequency dependence of individual components. The outputs can be written as $\boldsymbol{u} = \mathbf{W}\boldsymbol{x}$, where $\boldsymbol{x} = \mathbf{A}\boldsymbol{s}$. As previously mentioned, in the ideal case $\mathbf{WA}$ would be a diagonal matrix containing the constants d for all the signals, multiplied by a permutation matrix. It can be easily seen that, if this is true, the frequency scalings of all the components can be obtained by inverting the matrix $\mathbf{W}$ and performing the ratio, column by column, of each element with the one corresponding to the row corresponding to a given frequency. However, as pointed out in Section 3, if some signals are much smaller than others the above reasoning is only approximately valid. This is precisely what happens in our case: we are able to accurately recover the frequency scaling of the strongest signals, CMB and radio sources, whereas the others are lost (see Table 1).

6 CONCLUDING REMARKS AND FUTURE DEVELOPMENTS

We have developed a neural network suitable to implement the ICA technique for separating different emission components in maps of the sky at microwave wavelengths. The algorithm was applied to simulated maps of a $15° \times 15°$ region of sky at 30, 44, 70 and 100 GHz, corresponding to the frequency channels of the *Planck*/LFI.

Simulations include the CMB, extragalactic radio sources and Galactic synchrotron and thermal dust emission. The various components have markedly different angular patterns, frequency dependences and amplitudes.

The technique exploits the statistical independence of the different signals to recover each individual component with no prior assumption either on their spatial pattern or on their

Table 1. Input and output frequency scalings of the various components.

Frequency	Radio sources		CMB		synchrotron		dust	
(GHz)	input	output	input	output	input	output	input	output
100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
70	1.97	1.95	1.14	1.14	2.81	1.36	0.68	0.93
44	4.76	4.70	1.22	1.23	10.8	1.72	0.35	1.93
30	9.86	9.70	1.26	1.26	32.8	−12.0	0.19	3.77

frequency dependence. The great virtue of this approach is the capability of the algorithm to *learn* how to recover the independent components in the input maps. The price of the lack of a priori information is that each signal can be recovered multiplied by an unknown constant produced during the learning process itself. However, this is not a substantial limitation, as a lot of physics is encoded in the spatial patterns of the signals, and ultimately the correct normalization of each component can be obtained by resorting to independent observations.

The results are very promising. The CMB map is recovered with an accuracy at the 1-per-cent level. The algorithm is remarkably efficient also in the detection of extragalactic radio sources: almost all sources brighter then 15 mJy at 100 GHz (corresponding to $\simeq 0.7\sigma_{\rm CMB}$, $\sigma_{\rm CMB}$ being the rms level of CMB fluctuations on the pixel scale) are recovered; on the other hand, it must be stressed that this is not directly indicative of what can be achieved in the analysis of *Planck*/LFI data because the adopted resolution (3.5×3.5 arcmin^2) is much better than that of the real experiment, instrumental noise has been neglected and the same spectral slope was assumed for all sources.

Also, the frequency dependences of the strongest components are correctly recovered (error on the spectral index is 1 per cent for the CMB and extragalactic sources).

Maps of subdominant signals (Galactic synchrotron and dust emissions) are recovered with rms errors of about 10 per cent; their spectral properties cannot be retrieved by our technique.

The reconstruction has equal quality on all the scales of the input maps down to the pixel size.

All this indicates that this technique is suitable for a variety of astrophysical applications, i.e. whenever we want to separate independent signals from different astrophysical processes occurring along the line of sight.

Of course, much work has to be performed to better explore the potential of the ICA technique. It has to be tested under more realistic assumptions, taking into account instrumental noise and the effect of angular response functions, as well as including a more complete and accurate characterization of the foregrounds.

In particular, the assumption that the spectral properties of each foreground component is independent of position will have to be relaxed to allow for spectral variations across the sky. Also, it will be necessary to deal with the fact that Galactic emissions are correlated.

The technique is flexible enough to offer good prospects in this respect. In the learning stage, the ICA algorithm makes use of non-linear functions that, case by case, are chosen to minimize the mutual information between the outputs; improvements could be obtained by specializing the ICA inner non-linearities to our specific needs. Also, it is possible to take into account properly our prior knowledge on some of the signals to recover, while still taking advantage as far as possible of the ability of this neural network approach to carry out a 'blind' separation. Work in this direction is in progress.

ACKNOWLEDGMENTS

We warmly thank Luigi Danese for original suggestions. We also thank Krzysztof M. Górski and all the people who collaborated to build the HEALPix pixelization scheme that was extensively used in this work. The work was supported in part by ASI (Italian Space Agency) and MURST (Ministry for University, Scientific and Technological Research). LT acknowledges financial support from the Spanish DGES, projects ESP98–1545–E and PB98–0531–C02–01.

REFERENCES

Amari S., Chichocki A., 1998, Proc. IEEE, 86, 2026
Bell A. J., Sejnowski T. J., 1995, Neural Comp., 7, 1129
Bouchet F. R., Prunet S., Sethi S. K., 1999, MNRAS, 302, 663
Cayón L.,et al., 2000, MNRAS, 315, 757
Comon P., 1994, Signal Processing, 36, 287
de Oliveira-Costa A., Tegmark M., 1999, in de Oliveira-Costa A., Tegmark M., eds, ASP Conf. Ser. Vol. 181, Microwave Foregrounds. Astron. Soc. Pac., San Francisco
Donoho D. L., Johnstone I. M., 1995, J. Am. Statistics Assoc., 90, 1200
Fixsen D. J., Cheng E. S., Gales J. M., Mather J. C., Shafer R. A., Wright E. L., 1996, ApJ, 473, 576
Friedman J. H., 1987, J. Am. Statistics Assoc., 82, 249
Górski M., Wandelt B. D., Hansen F. K., Hivon E., Banday A. J., 1999, preprint (astro-ph/9905275; also see the HEALPix web page at `http://www.eso.org/~kgorski/healpix/`)
Gull S. F., 1988, in Erickson G. J., Smith C. R., eds, Maximum Entropy and Bayesian Methods in Science and Engineereng. Kluwer, Dordrecht, p. 53
Haslam C. G. T., Stoffel H., Salter C. J., Wilson W. E., 1982, A&AS, 47, 1
Hobson M. P., Jones A. W., Lasenby A. N., Bouchet F. R., 1998, MNRAS, 300, 1
Hobson M. P., Barreiro R. B., Toffolatti L., Lasenby A. N., Sanz J. L., Jones A. W., Bouchet F. R., 1999, MNRAS, 306, 232
Hyvärinen A., Oja E., 1999, Independent Component Analysis: A Tutorial, `http://www.cis.hut.fi/~aapo/papers/IJCNN99_tutorialweb/`
Jawerth B., Sweldens W., 1994, SIAM Rev., 36, 377
Lemarié P. G., Meyer Y., 1986, Rev. Mat. Ib., 2, 1
Mallat S. G., 1989, IEEE Trans. Pat. Anal. Mach. Int., 11, 674
Mandolesi N. et al., 1998, *Planck* Low Frequency Instrument. A Proposal Submitted to the ESA for the FIRST/*Planck* Programme
Ogden R. T., 1997, Essential Wavelets for Statistical Applications and Data Analysis. Birkhauser, Boston, p. PAGE
Puget J. L. et al., 1998, High Frequency Instrument for the *Planck* Mission, A Proposal Submitted to the ESA for the FIRST/*Planck* Programme
Sanz J. L., Argüeso F., Cayón L., Martínez-González E., Barreiro R. B., Toffolatti L., 1999a, MNRAS, 309, 672
Sanz J. L., Barreiro R. B., Cayón L., Martínez-González E., Ruiz G. A., Díaz F. J., Argüeso F., Silk J., Toffolatti L., 1999b, A&A, 140, 99
Schlegel D. J., Finkbeiner D. P., Davies M., 1998, ApJ, 500, 525
Seljak U., Zaldarriaga M., 1996, ApJ, 469, 437
Skilling J., 1988, in Erickson G. J., Smith C. R., eds, Maximum Entropy

and Bayesian Methods in Science and Enginereeng. Kluwer, Dordrecht, p. 173
Skilling J., 1989, in Skilling J., eds, Maximum Entropy and Bayesian Methods. Kluwer, Dordrecht, p. 53
Tegmark M., de Oliveira-Costa A., 1998, ApJ, 500, L83
Tegmark M., Efstathiou G., 1996, MNRAS, 281, 1297
Tegmark M., Eisenstein D. J., Hu W., de Oliveira-Costa A., 2000, ApJ, 530, 133
Tenorio L., Jaffe A. H., Hanany S., Lineweaver C. H., 1999, MNRAS, 310, 823
Toffolatti L., Argüeso F., De Zotti G., Mazzei P., Franceschini A., Danese L., Burigana C., 1998, MNRAS, 297, 117
Yang H. H., Amari S., 1997, Neural Comp., 9, 1457

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