Mon. Not. R. Astron. Soc. **346,** 1089–1102 (2003)

Multidetector multicomponent spectral matching and applications for cosmic microwave background data analysis

J. Delabrouille,[1⋆] J.-F. Cardoso[2] and G. Patanchon[1]

[1]*PCC – Collège de France, 11, place Marcelin Berthelot, F-75231 Paris, France*
[2]*CNRS/ENST – 46, rue Barrault, 75634 Paris, France*

Accepted 2003 August 1. Received 2003 July 24; in original form 2002 December 12

ABSTRACT

We present a new method for analysing multidetector maps containing several astrophysical components. Our method, based on matching the data to a model in the spectral domain, permits us to estimate jointly the spatial power spectra of the components and of the noise, as well as their mixing coefficients. It is of particular relevance for analysis of millimetre-wave maps of cosmic microwave background (CMB) anisotropies.

Key words: methods: data analysis – cosmic microwave background – cosmology: observations.

1 INTRODUCTION

Mapping sky emissions at millimetre wavelengths, and in particular cosmic microwave background (CMB) anisotropies, is one of the main objectives of ongoing observational effort in millimetre-wave astronomy. Sensitive balloon-borne and space-borne missions such as *Archeops* (Benoît et al. 2002), *BOOMERanG* (de Bernardis et al. 2000), *MAXIMA* (Hanany et al. 2000) and *WMAP* (Bennett et al. 1997) are currently yielding a large quantity of multidetector and multifrequency measurements. Within a few years, the *Planck* mission (Lamarre et al. 2000; Bersanelli & Mandolesi 2000), to be launched by ESA in 2007, will observe the full sky with ∼100 detectors distributed in nine frequency bands ranging from 30 to 850 GHz. The main objective of these observations is the determination of the spatial power spectrum of CMB anisotropies. A secondary objective is identifying and mapping the emission from all contributing astrophysical processes.

The availability of several detectors operating in several bands makes it possible to devise new powerful data processing schemes. In particular, by combining data from several detectors, it is possible to substantially improve the signal-to-noise ratio (by weighted averaging) and to separate several foreground components (possibly of astrophysical interest in their own right) from the CMB by component separation methods. Component separation, however, typically requires good knowledge of the transfer function connecting a multicomponent sky to multidetector maps.

This paper proposes spectral matching as a new approach to processing multidetector, multicomponent (MDMC) data, in which all the information needed to estimate the spatial power spectra of components *and/or* to separate them is sought in the data structure itself. The method works with or without prior detector calibration and gives access to spatial power spectra in a straightforward way; it is statistically efficient (being a maximum likelihood technique) and computationally efficient (working with a small set of statistics rather than with original maps).

This paper is organized as follows. The idea of spectral estimation via multidetector multicomponent spectral matching is introduced in Section 2. Section 3 describes the technique in more detail, connects it to a maximum likelihood method, and discusses specific implementations. Section 4 is devoted to evaluating the performance of the method on synthetic *Planck* HFI observations. We discuss the method and some possible extensions in Section 5.

2 THE MULTIDETECTOR MULTICOMPONENT FRAMEWORK

Multidetector CMB measurements can be modelled as resulting from the superposition of multiple components. Statistically efficient data processing should coherently exploit this MDMC structure.

2.1 Data model

The sky emission at millimetre wavelengths is well modelled at first order by a linear superposition of the emissions of a few processes: CMB anisotropies, thermal dust emission, thermal Sunyaev–Zel'dovich (SZ) effect, synchrotron emission, etc. The observation of the sky with detector d is then a noisy linear mixture of N_c components:

$$y_d(\theta, \phi) = \sum_{j=1}^{N_c} A_{dj} s_j(\theta, \phi) + n_d(\theta, \phi), \quad (1)$$

where s_j is the emission template for the jth astrophysical process, hereafter referred to as a *source* or a *component*. The coefficients A_{dj} reflect emission laws and detector properties while n_d accounts

⋆E-mail: delabrouille@cdf.in2p3.fr

for noise. For simplicity, we neglect for the moment beam effects, postponing their discussion to Section 5.

Quantities of prime interest are spatial power spectra. For the *j*th component, at frequency ℓ, this is

$$C_j(\ell) = \langle \, | \, s_j(\ell) \, |^2 \rangle, \tag{2}$$

where $\langle\cdot\rangle$ denotes the expectation operator and ℓ indexes either a Fourier mode or an (ℓ, m) mode.

In practice, power spectra are estimated by averages over bins

$$C_j(q) = \frac{1}{n_q} \sum_{\ell \in \mathcal{D}_q} C_j(\ell), \tag{3}$$

where $q = 1, \ldots, Q$ is the spectral bin index, $\mathcal{D}_q$ is the set of frequencies contributing to bin q and n_q is the number of such frequencies.[1] Typical bins can be bands $\ell_{\min} \leqslant \ell < \ell_{\max}$ extending over a range of one to tens of ℓ values.

Multidetector power spectrum. Since we focus on jointly processing the maps from all detectors, it is convenient to stack $y_1, \ldots, y_{Nd}$ into a single $N_d \times 1$ vector Y. Then the set of equation (1) for all N_d detectors is more compactly written in matrix vector form as

$$\boldsymbol{Y}(\theta, \phi) = \mathbf{A}\boldsymbol{S}(\theta, \phi) + \boldsymbol{N}(\theta, \phi) \tag{4}$$

with a so called $N_d \times N_c$ 'mixing matrix' $\mathbf{A}$. In Fourier space, this equation reads

$$\boldsymbol{Y}(\ell) = \mathbf{A}\boldsymbol{S}(\ell) + \boldsymbol{N}(\ell). \tag{5}$$

The power spectrum of process $\boldsymbol{Y}$ is represented by the $N_d \times N_d$ spectral density matrix $\langle \boldsymbol{Y}(\ell)\boldsymbol{Y}(\ell)^\dagger \rangle$ where † denotes transpose conjugation. Its average over bins

$$\mathbf{R}_Y(q) = \frac{1}{n_q} \sum_{\ell \in \mathcal{D}_q} \langle \boldsymbol{Y}(\ell)\boldsymbol{Y}(\ell)^\dagger \rangle \quad (q = 1, \ldots, Q) \tag{6}$$

will also be referred to as a spectral density matrix. According to the linear model (1), it is structured as

$$\mathbf{R}_Y(q) = \mathbf{A}\mathbf{R}_S(q)\mathbf{A}^\dagger + \mathbf{R}_N(q) \quad (q = 1, \ldots, Q) \tag{7}$$

with $\mathbf{R}_S(q)$ and $\mathbf{R}_N(q)$ defined similarly to $\mathbf{R}_Y(q)$. Statistical independence between components implies

$$\mathbf{R}_S(q) = \mathrm{diag}(C_1(q), \ldots, C_{N_c}(q)). \tag{8}$$

For the sake of exposition, we assume that the noise is uncorrelated, both across detectors and in space, so that the noise structure is described by N_d parameters:

$$\mathbf{R}_N(q) = \mathrm{diag}\left(\sigma_1^2, \ldots, \sigma_{N_d}^2\right). \tag{9}$$

Parameter extraction by spectral matching. The MDMC model, as defined by equations (7–9), depends on a set $\{\mathbf{R}_Y(q)\}$ of Q spectral density matrices, which in turn depend on $\{\mathbf{A},\ C_j(q),\ \sigma_d^2)\}$, amounting to $N_d \times N_c + Q \times N_c + N_d$ scalar parameters. However, the number of independent correlations in Q spectral density matrices is $Q \times N_d(N_d + 1)/2$ (since each matrix is real symmetric). This latter number is (in general) higher than the former.

With this in mind, our proposal can be summarized as 'MDMC spectral matching', meaning: *estimate all (or parts of) the parameters* $\theta = \{\mathbf{A}, C_j(q), \sigma_d^2)\}$ *by finding the best match between* $\{\mathbf{R}_Y(q)\}$, *as specified by (7–9), and a set of Q 'empirical spectral density matrices'* $\{\widehat{\mathbf{R}}_Y(q)\}$:

$$\widehat{\mathbf{R}}_Y(q) = \frac{1}{n_q} \sum_{\ell \in \mathcal{D}_q} \mathbf{Y}(\ell)\mathbf{Y}(\ell)^\dagger \quad (q = 1, \ldots, Q) \tag{10}$$

which are the natural non-parametric estimates of the corresponding $\mathbf{R}_Y(q)$.

2.2 Maximum likelihood spectral matching

Any reasonable measure of mismatch between the empirical density matrices $\{\widehat{\mathbf{R}}_Y(q)\}$ and their model counterparts $\{\widehat{\mathbf{R}}_Y(q;\theta)\}$ could be used to compute estimates of a θ parameter. In order to get good estimates, however, one should use a mismatch criterion derived from statistical principles. Such a derivation can be based on the statistical distribution of the Fourier coefficients of a stationary process which are (at least asymptotically in the data size) normally distributed, uncorrelated, with a variance proportional to the power spectrum (Whittle approximation, see Appendix B). Thus, the likelihood of the observations can be readily expressed in terms of spectral density matrices. Appendix B outlines how the (negative) log likelihood of the data then is (up to irrelevant factors and terms) equal to

$$\phi(\theta) = \sum_{q=1}^{Q} n_q D(\widehat{\mathbf{R}}_Y(q), \mathbf{R}_Y(q;\theta)) \tag{11}$$

where $D(\cdot, \cdot)$ is a measure of divergence between two positive $n \times n$ matrices defined by

$$D(\mathbf{R}_1, \mathbf{R}_2) = \mathrm{tr}\left(\mathbf{R}_1\mathbf{R}_2^{-1}\right) - \log\det(\mathbf{R}_1\mathbf{R}_2^{-1}) - n. \tag{12}$$

It can be seen[2] that $D(\mathbf{R}_1, \mathbf{R}_2) \geqslant 0$ with equality if and only if $\mathbf{R}_1 = \mathbf{R}_2$. Thus *spectral matching corresponds to maximum likelihood estimation in a stationary model*. The minimizer of $\phi(\theta)$ is then a maximum likelihood estimate, with the good statistical properties associated with it.

Only in an asymptotic framework can maximum likelihood procedures be proved to reach minimum estimation variance. This means that criteria which are equivalent to (11) are expected to have the same statistical quality as (11). In particular, criterion (11) can be replaced by a quadratic approximation: when each $\widehat{\mathbf{R}}_Y(q)$ is close to $\mathbf{R}_Y(q;\theta)$, a second-order expansion of $D(\widehat{\mathbf{R}}_Y, \mathbf{R}_Y)$ yields

$$D_2(\widehat{\mathbf{R}}_Y, \mathbf{R}_Y) = \mathrm{tr}\left(\widehat{\mathbf{R}}_Y^{-1}(\widehat{\mathbf{R}}_Y - \mathbf{R}_Y)\widehat{\mathbf{R}}_Y^{-1}(\widehat{\mathbf{R}}_Y - \mathbf{R}_Y)\right). \tag{13}$$

The resulting quadratic criterion is of particular interest when the unknown parameters enter *linearly* in $\mathbf{R}_Y(q;\theta)$ (for instance when $\mathbf{A}$ is known and θ only contains the binned power spectra of the components) since then criterion minimization becomes trivial. In this paper, however, we stick to using (11)–(12). Even though the divergence (12) may, in general, seem more difficult to deal with than its quadratic approximation (13), it actually lends itself to simple optimization via the EM algorithm (see Section 3.1) thanks to its connection to the likelihood.

2.3 Comments

Some preliminary comments about the MDMC spectral matching approach are in order.

[1] It is customary for CMB data analysis to weight the terms in sum 3 by $\ell(\ell + 1)$. For the sake of exposition, we use a flat weighting here (see Section 5 for weighted sums).

[2] For instance by expressing $D(\mathbf{R}_1, \mathbf{R}_2)$ in terms of the eigenvalues of $\mathbf{R}_2^{-1/2}\mathbf{R}_1\mathbf{R}_2^{-1/2}$.

Parameter choice. There is a lot of flexibility in the choice of parameters over which to minimize the spectral mismatch. By selecting different sets of parameters, different goals can be achieved. For instance, we may assume that matrix $\mathbf{A}$ and the noise spectrum $\mathbf{R}_N(q)$ are known so that the mismatch is minimized only with respect to the binned spectra $C_i(q)$ of all components: the method appears as a spectral estimation technique *which does not require the explicit separation of the observed maps into component maps*. Another important example, as illustrated in Section 4, consists of including matrix $\mathbf{A}$ among the free parameters. Then, the method works as the so-called 'blind techniques', and permits a determination of the emission law of the components, or the cross-calibration of detectors.

Degeneracies. A key issue in spectral matching is whether or not matrix $\mathbf{A}$ can be uniquely determined from the data only. When all parameters $\{\mathbf{A}, C_j(q), \sigma_d^2\}$ are allowed to vary, there are at least two obvious degeneracies. First, the *ordering* (or numbering) of the components in the model is immaterial: matrix $\mathbf{A}$ cannot be recovered to better than a column permutation on the sole basis of a spectral match. Second, a scalar factor can be exchanged, for each component j, between the jth column of $\mathbf{A}$ and $C_j(q)$. These scale factors cannot be determined from the data themselves.

Another trivial case of indetermination is when two columns of $\mathbf{A}$ corresponding to physically distinct components are proportional. In this case, the sum of the two appears in the model as one single component. The identifiability of the other components is not affected.

A more severe degeneracy occurs if any two components have proportional spectra. In this case, as is known from the noiseless case (Pham & Garat 1997), only the space spanned by the corresponding columns of $\mathbf{A}$ can be determined in a spectral match with $\mathbf{A}$ as a free parameter. In this case, however, the identifiability of the other components is unchanged, with no impact on the accuracy of component separation with a Wiener method (Section 5). The key point to remember is that spectral matching requires spectral diversity to separate components associated with unknown columns of $\mathbf{A}$.

Maximum likelihood. Section 2.2 explains why 'spectral matching' corresponds to maximum likelihood estimation. This occurs in a Gaussian stationary model with smooth (actually: constant over bins) spectra. In such a model the likelihood of the observations *is* a measure (11) of spectral matching. Since the likelihood then depends on the data *only* via the empirical spectral density matrices, the massive data reduction gained from replacing the observations by a (usually) much smaller set of statistics (the empirical spectral density matrices $\widehat{\mathbf{R}}_Y(q)$) is obtained without information loss.

Comparison with component separation. It is interesting to compare spectral matching with techniques based on prior explicit component separation.

Producing a CMB map as free as possible from foreground and noise contamination is the objective of the component separation step, in which maps obtained at different frequencies are combined to maximize the signal to noise ratio (where noise includes also foreground contamination).

The usual approach for taking advantage of multidetector measurements can be summarized as follows: first, form estimates $\widehat{s}_j(\ell)$ of component maps $s_j(\ell)$ (via component separation); secondly, estimate the spectrum of each component j by averaging within bins:

$$\widehat{C}_j(q) = \frac{1}{n_q} \sum_{\ell \in \mathcal{D}_q} |\,\widehat{s}_j(\ell)\,|^2, \tag{14}$$

with, possibly, some post-processing of the power spectrum estimates.

This method suffers from two difficulties. First, the best component separation methods typically require the prior knowledge of the statistical properties of the components (including the CMB power spectrum) and of the noise. Secondly, recovered maps contain residuals (including noise) which contribute to the total power, biasing the spectrum estimated on the map, unless the power spectrum of these residuals can be estimated accurately and subtracted.

In contrast our approach takes the reverse path. The first step is an estimation of the spectrum for the multidetector map (which takes the form of a sequence of spectral density matrices). This first step preserves all the joint correlation structure between maps. In essence, the second step (spectral matching) amounts to resolving the joint power spectrum into the spectra of individual components.

Hence, instead of first separating component *maps* and then computing power spectra, we first compute the multivariate power spectrum and then separate component *spectra*.

Applicability of spectral matching. The real sky emission is known to depart from the simple ideal model used for the development of the spectral matching method. It is important, then, to discuss how our method depends on some assumptions about the structure of the data.

In particular, the spectral adjustment is made under the following assumptions:

(i) the mixing matrix is position-independent,
(ii) the components are statistically independent,
(iii) noise is uncorrelated across detectors,
(iv) the components and the noise are normally distributed,
(v) the components and the noise are stationary,
(vi) the components and the noise are statistically isotropic,
(vii) their spectra are smooth functions of ℓ.

Assumptions i, ii, and iii are critical because they determine the structure (7) of the spectral matrices. Thus, matching estimates are no longer consistent if these assumptions are violated and for strong departures from the model, the spectral matching may fail completely. However, it is possible to refine the model as needed, for instance by including some correlation terms between components, by dividing the sky in patches where the mixing matrix may be assumed to be constant, by discarding regions of the sky where the emission is known to depart from the model (galactic regions of intense complex emission, strong point sources). The existence and applicability of such refined solutions depend on the exact sky emission model. If no structure can be assumed (for instance if the noise correlation matrix is completely arbitrary and unknown) there is not enough information to constrain the parameters uniquely.

Assumptions iv, v, vi and vii, in contrast, are not critical because they do not change the basic structure (7) of the spectral matrices and do not prevent the sample covariance matrices to converge to such a structure. Hence, violation of these assumptions only results in a loss of statistical efficiency with respect to techniques based on complex models. The Gaussianity assumption is not a requirement of the method because the sample covariance matrices (10) still have an expectation given by (7). If components or the noise are not Gaussian distributed, then some information is lost by only using second-order statistics (loss of efficiency), but consistency is preserved. Stationarity, isotropy, and spectral smoothness allow us to characterize in a meaningful way the distribution of a given component j via a simple smoothed spectrum $C_j(q)$. If they are violated then $C_j(q)$ may not have a meaningful statistical meaning; it is still

defined none the less and spectral matching will just produce an estimate of it. Again, the key point is that, regardless of the details of the distribution of $S_i(\ell)$, one can expect $\langle S_i(\ell)\bar{S}_j(\ell)\rangle = 0$ for $i \neq j$, that is, statistical independence is the true underlying inference engine. We also note that in the case of anisotropic spectra, it is always possible to use spectral domains $\mathcal{D}_q$ which are not ring-shaped.

Finally, spectral matching does not bias in any way the level of non-Gaussianity present in the CMB maps. Non Gaussian signals, even at low level, appear as an additional contribution to the CMB spectrum. They will show in maps as well, in a way depending on the method adopted to solve the linear system after the spectral matching step, once mixing coefficients and power spectra are known.

3 MDMC SPECTRAL MATCHING IN PRACTICE

The implementation of MDMC spectral matching will now be described in more detail; Section 3.1 describes the EM algorithm for its optimization; Section 3.2 describes a complementary technique for fast convergence.

3.1 The EM algorithm

The expectation maximization (EM) algorithm (Dempster, Laird & Rubin 1977) is a well-known technique for maximizing the likelihood of statistical models which include 'latent' or 'unobserved' variables. It is well suited to our purpose by taking the components as the latent variables. The EM algorithm is iterative: starting from an initial value of the parameters, it performs a sequence of parameter updates called 'EM steps'. Each step is guaranteed to increase the likelihood of the parameters.

The spectral matching criterion (11) actually being a likelihood function in disguise, the EM algorithm can be used for its minimization. Each EM step is guaranteed to improve the spectral fit by decreasing $\phi(\theta)$.

We consider the regular EM algorithm, based on the Gaussian likelihood described in Appendix B and taking as 'latent variables' the spectral modes $\mathbf{Y}(\ell)$. The form of the EM steps immediately follows, as sketched in Appendix C, and is summarized by the pseudo-code given in Table 1.

Table 1. The EM algorithm for minimizing the MDMC spectral mismatch $\phi(\theta)$ with respect to $\theta = \{\mathbf{A}, C_j(q), \sigma_d^2\}$.

Require: Spectral density matrices $\widehat{\mathbf{R}}_Y(1), \ldots, \widehat{\mathbf{R}}_Y(Q)$
Require: Initial value of $\theta = \{\mathbf{A}, C_j(q), \sigma_d^2\}$.
  Set $\widetilde{\mathbf{R}}_{yy}(q) = \widehat{\mathbf{R}}_Y(q)$ and $\widetilde{\mathbf{R}}_{yy} = \sum_q \frac{n_q}{n}\tilde{\mathbf{R}}_{yy}(q)$.
  repeat
    {_____ E-step. Compute conditional statistics:}
    Set $\mathbf{R}_S(q) = \mathrm{diag}(C_j(q))$ and $\mathbf{R}_N = \mathrm{diag}(\sigma_d^2)$
    for $q = 1$ to Q **do**
      $\mathbf{G}(q) = (\mathbf{A}^\dagger\mathbf{R}_N^{-1}\mathbf{A} + \mathbf{R}_S(q)^{-1})^{-1}$
      $\mathbf{W}(q) = \mathbf{G}(q)\mathbf{A}^\dagger\mathbf{R}_N^{-1}$
      $\widetilde{\mathbf{R}}_{ss}(q) = \mathbf{W}(q)\widehat{\mathbf{R}}_Y(q)\mathbf{W}(q)^\dagger + \mathbf{G}(q)$
      $\widetilde{\mathbf{R}}_{sy}(q) = \mathbf{W}(q)\widehat{\mathbf{R}}_Y(q)$
    end for
    $\widetilde{\mathbf{R}}_{ss} = \sum_{q=1}^{Q} \frac{n_q}{n}\widetilde{\mathbf{R}}_{ss}(q)$
    $\widetilde{\mathbf{R}}_{ys} = \sum_{q=1}^{Q} \frac{n_q}{n}\widetilde{\mathbf{R}}_{ys}(q)$
    {_____ M-step. Update the parameters:}
    $\mathbf{A} = \widetilde{\mathbf{R}}_{ys}\widetilde{\mathbf{R}}_{ss}^{-1}$
    $C_i(q) = \left[\widetilde{\mathbf{R}}_{ss}(q)\right]_{ii}$
    $\sigma_d^2 = \left[\widetilde{\mathbf{R}}_{yy} - \widetilde{\mathbf{R}}_{sy}^\dagger\widetilde{\mathbf{R}}_{ss}^{-1}\widetilde{\mathbf{R}}_{sy}\right]_{dd}$
    Rescale the parameters (see text).
  until a convergence criterion is satisfied

It is worth mentioning that EM steps take such a regular structure when the parameters are $\theta = \{\mathbf{A}, C_j(q), \sigma_d^2\}$. A slightly different form would result from a more constrained parameter set.

Recall that, as previously noted, there is a scale indetermination on each component's spectrum when $\theta = \{\mathbf{A}, C_j(q), \sigma_d^2\}$. We have found that this inherent indetermination must be explicitly fixed in order for EM to converge (this is the rescaling step in the last line of the pseudo-code). Our strategy is, after each EM step, to fix the norm of each column of $\mathbf{A}$ to unity and to adjust the corresponding power spectra accordingly. This is an arbitrary choice which happens to work well in practice.

The algorithm is initialized as follows. We take $\mathbf{R}_n = \mathrm{diag}(\widehat{\mathbf{R}}_Y)$ where $\widehat{\mathbf{R}}_Y = \sum_q (n_q/n)\widehat{\mathbf{R}}_Y(q)$. This is a gross overestimation since it amounts to assume no signal and only noise. The initial value of $\mathbf{A}$ is obtained by using the N_c dominant eigenvectors of $\widehat{\mathbf{R}}_Y$ as the N_c columns of $\mathbf{A}$. Again, this is nothing like any real *estimate* of $\mathbf{A}$, but rather a vague guess in 'the right direction'. Finally, the spectra $P_i(q)$ are taken as to be the diagonal entries of $\mathbf{A}^\dagger\widehat{\mathbf{A}}$. This ad hoc initialization procedure seems satisfactory. Note that it is a common rule of thumb to initialize EM with overestimated noise power.

Regarding the stopping rule, there is little point in devising a sophisticated stopping strategy (in practice, the algorithm is run for a pre-specified number of steps based on a few preliminary experiments with the data) because EM is only used 'halfway': its output is further fed to a specialized optimization algorithm (described at next section).

3.2 Non-linear optimization

When applied to our data, the EM algorithm demonstrates rapid convergence at first, and then enters a second phase of slower convergence. This is due to the fact that some parameters (e.g. subdominant power spectra in some spectral domains) have a very small effect on the criterion. In order to reach the true minimum of $\phi(\theta)$, it appears necessary to complement EM with another minimization technique. The strategy is to use the straightforward EM algorithm to quickly get close to the minimum of $\phi(\theta)$ and then to complete the minimization using a dedicated minimization algorithm. This complementary algorithm can use a simple design thanks to the good starting point provided by EM.

The spectral mismatch criterion (11) can, in theory, be minimized by any optimization algorithm. However, the same effect that slows down EM in its final steps also makes the minimization of the mismatch criterion (11) difficult for *any* algorithm. In particular, simple gradient algorithms are unacceptably slow. Actually, we found that even conjugate gradient techniques cannot overcome this problem and had to resort to a quasi-Newton method. We have used the classic BFGS (Broyden–Fletcher–Goldfarb–Shapiro) algorithm (Luenberger 1973). This technique minimizes an objective function by successive one-dimensional minimization (line searches). At each step, the direction for the line search is the gradient 'rectified' by the inverse of Hessian matrix. The BFGS technique is a rule to update an estimate of the inverse Hessian matrix at low computational cost.

4 TESTING AND PERFORMANCE

We now turn to illustrating the applications and performance of our multidetector multicomponent spectral-matching method on a simple set of synthetic observations: three-component noisy linear mixtures featuring contributions from CMB anisotropies, dust

emission, and SZ thermal emission. Bias and statistical uncertainties are investigated by a Monte Carlo technique.

Five implementations of the method for different applications will be discussed:

(i) a multicomponent spatial power spectrum estimation assuming the mixing matrix is known,

(ii) a blind approach in which spatial power spectra, noise levels, and the emission laws of components are jointly estimated on the data,

(iii) a semi-blind approach where CMB and SZ emission laws are assumed to be known, and the emission law of the dust component (in addition to spatial power spectra and noise levels for all components) is estimated from the data,

(iv) an application for detector cross-calibration,

(v) a Wiener-filter component separation using parameters estimated via blind spectral matching.

Finite beam effects are neglected in the present work, although they are not a fundamental limitation for our method (see Section 5). For definiteness, we also assume here that the noise is white, although this assumption can also be relaxed if needed.

4.1 Simulated data

Synthetic observations in six frequency bands identical to those of the *Planck* HFI are generated on 300×300 pixel maps corresponding to a $12.^{\circ}5 \times 12.^{\circ}5$ field located at high galactic latitude. For each mixture realization, synthetic components and noise are obtained as follows.

(i) The CMB component is a *COBE*-normalized, randomly generated realization of CMB anisotropies obtained using the spatial power spectrum C_ℓ predicted by the CMBFAST software (Zaldarriaga & Seljak 2000) with $H_0 = 65$ km s^{-1} Mpc^{-1}, $\Omega_{\rm m} = 0.3$, $\Omega_{\rm b} = 0.045$, $\Lambda = 0.7$.

(ii) The galactic dust emission template is obtained from the 100-μm *IRAS* data in the sky region located around $\alpha = 204^{\circ}$ and $\delta = 11^{\circ}$. Bright stars are removed using a point source extracting algorithm. Residual stripes are cut out by setting to zero the contaminated Fourier coefficients. The Fourier modes suppressed in this way are randomly regenerated with a distribution obtained, for each mode, from the statistics of the other modes at the same scale in the *IRAS* map. This method preserves the (assumed) statistical azimuthal symmetry and general shape of the spatial spectrum.

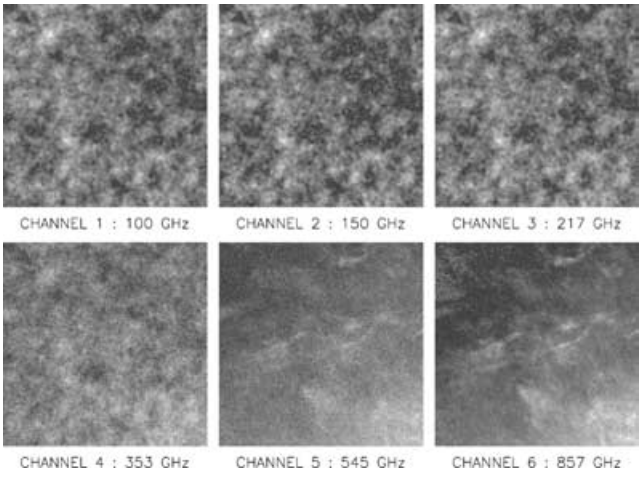

Figure 1. Simulated observations for six frequency bands. This figure can be seen in colour in the on-line version of the journal on *Synergy*.

Table 2. Fraction of the power in each of the components.

ν	100	143	217	353	545	857
CMB	0.889	0.926	0.896	0.275	0.0019	1.3×10^{-7}
dust	9×10^{-5}	6×10^{-4}	0.0082	0.215	0.687	0.938
SZ	0.0064	0.0032	2×10^{-7}	0.0044	0.00019	5.2×10^{-8}
noise	0.102	0.0727	0.108	0.536	0.320	0.0667

(iii) The thermal Sunyaev–Zel'dovich template is drawn at random from a set of 1500 SZ maps generated for this purpose using the software described in (Delabrouille, Melin & Bartlett 2002).

(iv) White noise at the level of the nominal per-channel *Planck* HFI values is added to the observations.

Synthetic observations are displayed in Fig. 1. The general common pattern which can be seen in the lowest frequency channels is simulated CMB anisotropies, whereas the pattern of emission of interstellar dust as observed with IRAS dominates our 857- and 545-GHz maps. The contribution of the SZ effect, very subdominant, is not obviously visible on these maps.

Table 2 gives, for each channel, the relative power of all components and of noise for a typical synthetic mixture (here, 'relative' means that the sum of all powers is normalized to unity). Typical input templates for the three components can be seen later (Fig. 6, left column).

4.2 Application 1: spectral estimation

The first application is the estimation of component spatial power spectra. It is assumed that the mixing matrix is known, but that the noise level for each map is not known precisely. The set of parameters to be estimated from the data then is $\theta = \{C_j(q), \sigma_d^2\}$.

Component spectra are estimated on 32 ring-shaped domains for 5000 different mixtures. The first 30 domains are equally spaced rings covering the lowest 60 per cent of the spatial frequencies ($0 < \ell/\ell_{\rm max} < 0.6$), and the remaining two cover respectively $0.6 < \ell/\ell_{\rm max} < 0.8$ and $0.8 < \ell/\ell_{\rm max} < 1$. This choice of spectral domains is adapted to the assumed azimuthal symmetry of the spectra by the choice of ring-shaped domains, and has a large number of rings in the region where the signal is strong and where information from source spectra is relevant.

The resulting estimation of the spatial power spectrum of the three components in the relevant frequency range is shown in Fig. 2. Errors on estimated spectra are obtained from the dispersion over the 5000 distinct simulated observations. For the SZ effect, the spatial power spectrum is averaged into larger bins after parameter estimation to reduce the scatter of the measurements. The figure shows that, as expected, a low-variance unbiased power spectrum is obtained for all components without explicit separation of the observations into component maps. For the CMB, the measurement is sample (cosmic) variance limited at small spatial frequencies. Such an effect does not appear on the dust spectrum estimate because we use only one dust map in the Monte Carlo simulation.

4.3 Application 2: blind parameter estimation

Let us now assume that the exact emission laws of all components are unknown. Then the full parameter set, to be estimated from the data, is $\theta = \{C_j(q), \sigma_d^2, \mathbf{A}\}$. Again, we estimate parameters on 5000 different simulated data sets. For each run, the scale freedom between mixing matrix columns and component power spectra is fixed by renormalizing to the true value of $\mathbf{A}$ at a single reference frequency

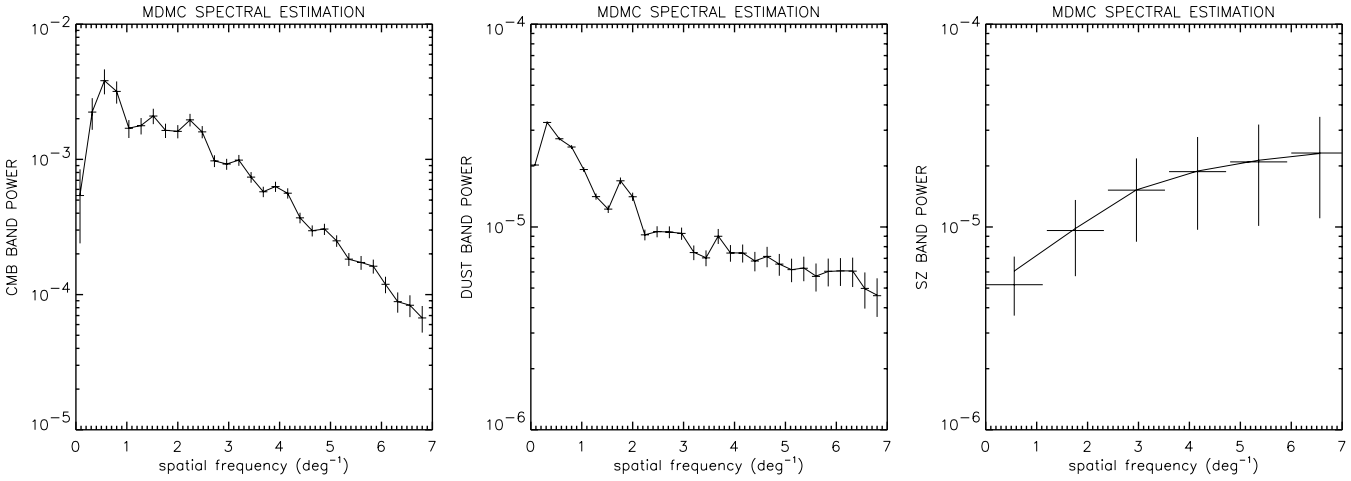

Figure 2. This figure shows the recovered spatial power spectrum $\ell^2 C_\ell$ (crosses) compared with the exact band-averaged spectra (solid lines) for CMB (left), dust (middle), and thermal SZ effect (right). These results correspond to a non-blind MDMC spectral estimation in which the mixing matrix **A** is known. Vertical bars show the 1σ errors, and horizontal bars the spatial frequency range of each bin.

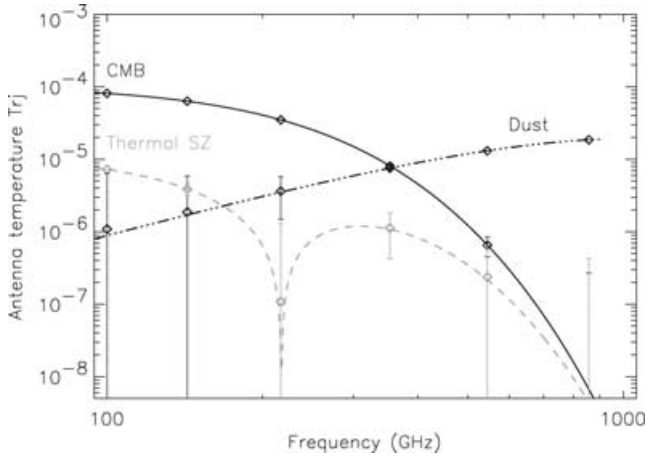

Figure 3. The figure shows the recovered emission laws of the components (diamonds) compared with the exact emission laws used in the simulations (solid lines). The errors are computed from the dispersion of the recovered values for 10 000 different synthetic mixtures.

(100 GHz for the CMB and thermal SZ effect, and 857 GHz for the dust). Error bars ($\pm 1\sigma$) for all parameters are computed from the distribution of the estimates over all simulated observations.

Fig. 3 displays recovered emission spectra (diamonds with 1σ error bars) as compared with exact emission spectra (lines). Emission laws of all components are recovered with no significant bias. The CMB emission law is recovered very accurately at all frequencies except 857 GHz. The dust emission law is recovered quite accurately at high frequencies, less accurately at frequencies where it is very subdominant. The SZ effect emission shape, subdominant at all frequencies, is recovered with larger relative error bars. Because of the renormalization, error bars for CMB and SZ vanish at 100 GHz, and the dust emission-law error bar vanishes at 857 GHz.

Spatial power spectra, in turn, are also estimated. As shown in Fig. 4, CMB and dust spatial power spectra are recovered with good accuracy and no significant bias, almost as well as for the non-blind spectral estimation. The SZ power spectrum is also significantly constrained, although error bars are significantly larger than in the non-blind spectral estimation.

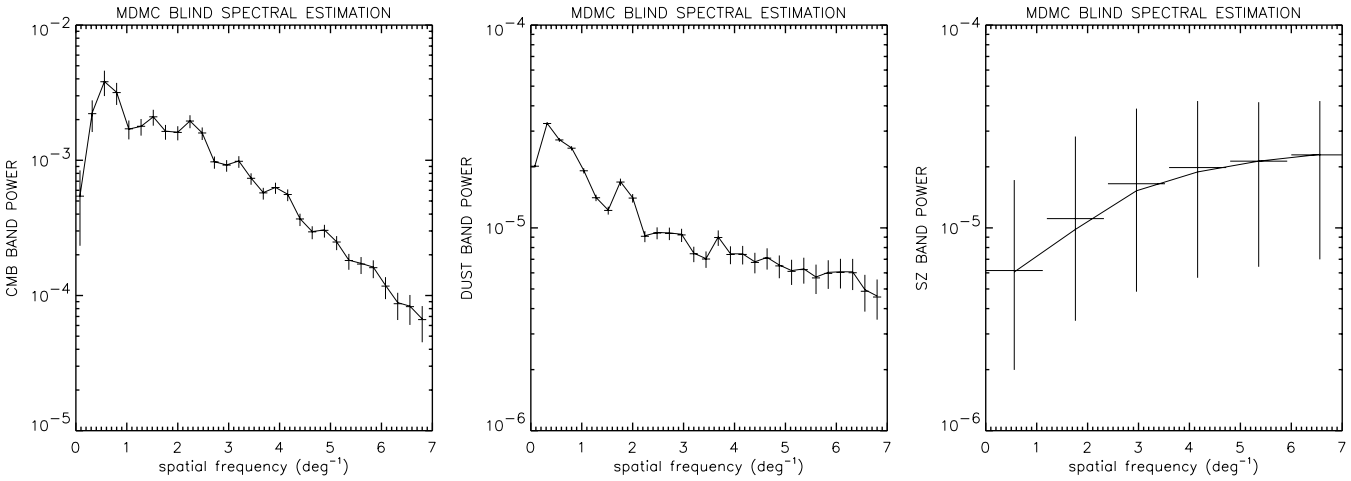

Figure 4. The figure – similar to Fig. 2 but for *blind* spectral matching – shows the recovered power spectra of the components compared to exact ones (solid lines). The errors are computed from the dispersion of the recovered values for 5000 different synthetic mixtures.

Table 3. Comparison of true and estimated noise levels (RMS). The errors are obtained from the dispersion of results obtained using 10 000 different mixtures.

Channel	100	143	217	353	545	857
RMS est. $\times 10^{-6}$	(29.1 ± 0.22)	(18.7 ± 0.13)	(12.85 ± 0.09)	(11.92 ± 0.07)	(8.98 ± 0.05)	(4.97 ± 0.06)
RMS true $\times 10^{-6}$	29.11	18.70	12.86	11.93	8.980	4.970

Table 4. Relative errors in dust emission-law estimation. In the first case, all the elements of the mixing matrix are estimated (blind approach). In the second case, the columns of the mixing matrix which corresponds to the CMB and the thermal SZ components are fixed (semi-blind approach). Although the semi-blind approach does not improve significantly the determination of the dust spectrum at 545 GHz, the improvement is very significant (factors of two to three) at other frequencies.

Channel	100	143	217	353	545	857
True dust em spectrum	0.3071	0.5902	1.2177	2.6106	4.5371	6.4288
Relative error, blind approach	6.229	2.469	0.634	0.0662	0.00790	No values
Relative error, semi-blind approach	2.623	1.056	0.285	0.0368	0.00725	No values

Finally, Table 3 shows the estimates of the noise RMS as compared to true levels. Relative errors are below 2.5 per cent for all channels.

4.4 Application 3: semi-blind parameter estimation

In our particular case, the emission laws of the CMB and of the SZ are known to almost perfect accuracy. Assume, however, that measuring the dust emission law is of particular interest. How much do we gain by forcing known emission laws to their true value, and estimating only the unknown dust emission spectrum?

We repeat the simulations described in 4.3, now fixing two columns of the mixing matrix, and estimating the third one (in addition to domain-averaged spatial spectra and noise levels). Table 4 compares quantitatively the relative errors on the resulting dust emission law. At low frequency (between 100 and 217 GHz), the accuracy of the estimation is improved by a factor of 2 to 3. At 353 GHz, the improvement is still noticeable, but at 545 GHz, where the dust emission begins to dominate, the blind and semi-blind approaches give similar errors. The use of partial prior information on the mixing matrix $\mathbf{A}$ is thus useful here to improve the estimation of the entries of $\mathbf{A}$ which contribute little relative power to the observations.

In addition to this substantial improvement in estimating the unknown 'dust column' of $\mathbf{A}$, the semi-blind approach is more efficient for estimating the SZ power spectrum than the full blind implementation. Fig. 5 shows the comparison of the quality of spectral estimation in the blind and semi-blind approaches relative to the non-blind. To the precision of our Monte Carlo tests (1–2 per cent level on error bars), the semi-blind result is as accurate for this particular mixture as the non-blind estimate, and significantly better than the blind result. As the semi-blind and the non-blind estimates give similar results, however, the actual enhancement in precision depends on details of the mixture and parametrization.

This comparison, however, shows that it is in general useful to exploit as much as possible reliable prior information. Our method is flexible enough to do so.

4.5 Application 4: detector calibration

The mixing matrix $\mathbf{A}$ depends not only on components (through emission spectra), but also on detectors (through frequency bands

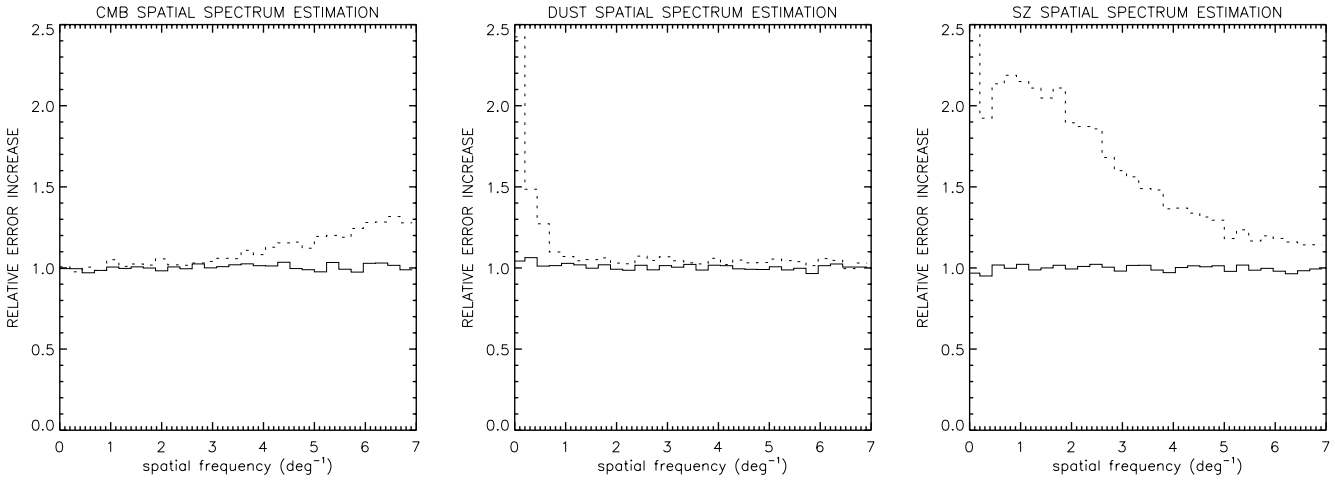

Figure 5. The figure shows the comparison of the quality of the blind and semi-blind power spectrum estimation of the components. The solid line displays the ratio between the size of the 1σ error in the semi-blind and in the non-blind spectral matching, showing that they are comparable to within simulation accuracy. In contrast, the dotted line shows the ratio between the size of the 1σ error in the blind and in the non-blind spectral matching, showing that some accuracy is lost when all components of the mixing matrix are adjusted as additional parameters.

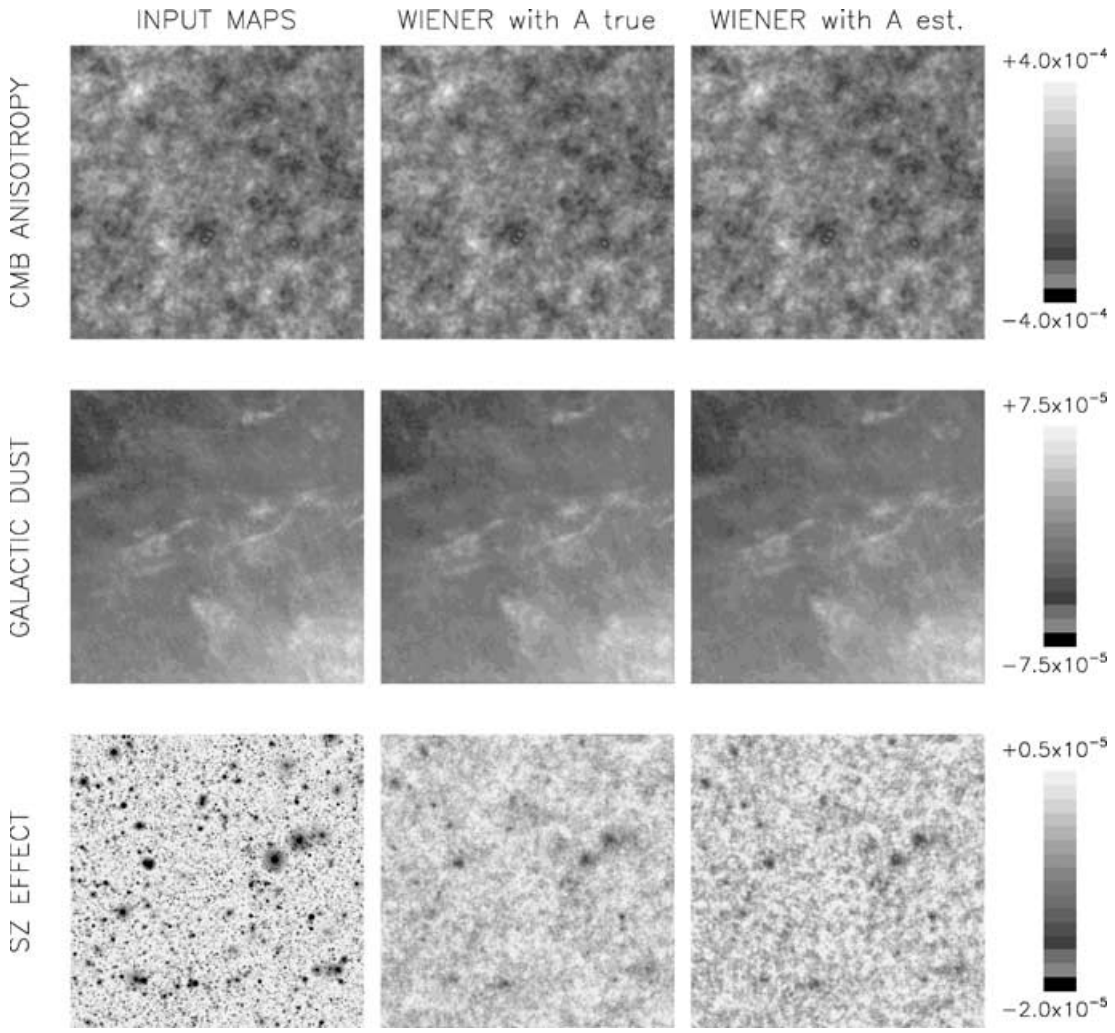

Figure 6. Left: true templates used as inputs. Middle: templates recovered using the 'true' Wiener filter. Right: templates recovered using the blind separation. This figure can be seen in colour in the on-line version of the journal on *Synergy*.

and optical efficiency). Mixing matrix coefficients A_{dj}, expressed in readout (rather than physical) units can be approximated by the product of a detector-dependent calibration coefficient α_d and an emission law $\epsilon_j(\nu)$:

$$A_{dj} \simeq \alpha_d \epsilon_j(\nu_d), \tag{15}$$

where ν_d is the central observing frequency of detector d. Used on a data set from detectors observing in the same frequency band, the estimation of $\mathbf{A}$ for any astrophysical component gives relative calibration coefficients between detectors. If in addition the emission law of at least one of the components is known (e.g. CMB anisotropies), the estimation of the mixing matrix provides a relative calibration across frequency bands. Finally, if among the components there is one with known emission spectrum and known amplitude (or known spatial power spectrum), absolute calibration can be obtained in the same way. For instance, it is not excluded that in the not-so-far future, a high-resolution experiment dedicated to a wide-field point source survey in the millimetre range can be calibrated on CMB anisotropies(!).

4.6 Application 5: component separation

The separation of astrophysical components by some kind of inversion of the linear system of equation (1) has been the object of extensive previous work. Popular linear methods are listed in Appendix **A**. In a Gaussian model, the best inversion is obtained by the Wiener filter. This filter, however, requires prior knowledge of the mixing matrix $\mathbf{A}$, component spatial power spectra, and noise power spectra. As discussed by Cardoso et al. (2002), our spectral-matching method yields all the parameters needed to implement a Wiener-based component separation *on maps*.

We compare the quality of component reconstruction using either the estimated parameter set $\theta = \{C_j(q), \sigma_d^2, \mathbf{A}\}$ or 'true' best-knowledge values.

Reconstructed maps. Fig. 6 illustrates the quality of map reconstruction by Wiener inversion. The first column displays the input components, the second column shows components recovered with the exact Wiener filter (computed from the true mixing matrix, ensemble averages of the noise, ensemble averages of CMB and SZ power spectra, and a k^{-3} fit of the spatial power spectrum of the dust template). The third column displays the components recovered by Wiener inversion using estimated parameters. In both cases, CMB and dust emissions are recovered satisfactorily, but the SZ effect – strongly peaked and hence poorly suited to processing in Fourier space – remains noisy. Visually, both methods perform about as well.

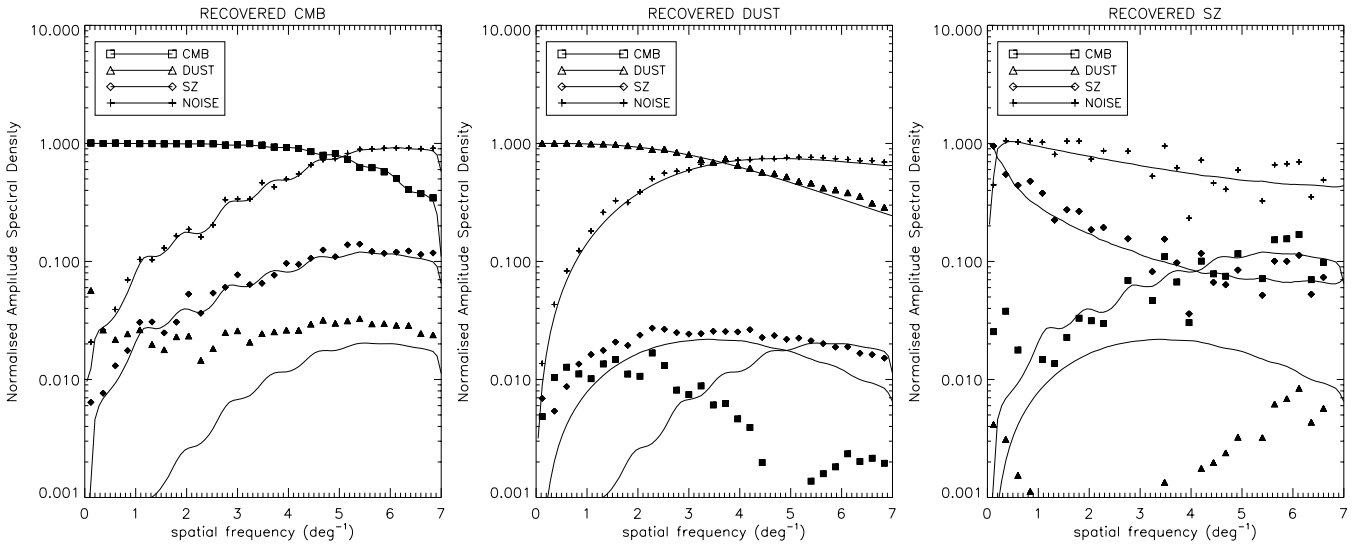

Figure 7. Contributions to the output map as a function of spatial frequency, relative to the true level of that component map. The left panel is for CMB, the middle panel for dust, and the right panel for the SZ effect. Results obtained with exact values of the mixing matrix, the spectra and noise levels are plotted as plain lines, and results obtained with the Wiener implementation using estimated parameters as diamonds.

Contamination levels. The quality of the separation can be assessed by a measure of contamination levels, i.e. how much of the other components get into a given component map after separation.

The Wiener matrix, $\mathbf{W} = [\mathbf{A}^t\mathbf{R}_N^{-1}\mathbf{A} + \mathbf{R}_S^{-1}]^{-1}\mathbf{A}^t\mathbf{R}_N^{-1}$, obtained with exact values of $\mathbf{A}$, $\mathbf{R}_N$ and $\mathbf{R}_S$, differs slightly from its estimate $\widehat{\mathbf{W}}$, computed with estimates $\widehat{\mathbf{A}}$, $\widehat{\mathbf{R}}_N$ and $\widehat{\mathbf{R}}_S(q)$.$\widehat{\mathbf{R}}_S(q)$ differs from $\mathbf{R}_S$ not only because it is an estimate, but also because it is a flat band-power approximation.

At each frequency, off-diagonal terms of $\widehat{\mathbf{W}}\mathbf{A}$ correspond to leakage of other components into a given component's estimate at spatial frequency k. Each panel of Fig. 7 refers to one component (CMB, dust and SZ), and shows the relative contribution of all components and of noise to the recovered map. Levels are relative to the true map, so that the contribution of a component to its own recovered map illustrates the spatial filtering induced by the Wiener inversion. The figure illustrates that the inversion done with blindly estimated parameters performs almost as well as the separation using exact values of the spectra and mixing matrix. Differences are typically much smaller than noise contamination, which is comparable with the blind and the non-blind approaches.

5 DISCUSSION

5.1 Related work on component separation

Explicit component separation was investigated first in CMB applications by Tegmark & Efstathiou (1996), Bouchet & Gispert (1999) and Hobson et al. (1998). In these applications, all the parameters of the model (mixing matrix, noise levels, statistics of the components, including the spatial power spectra) are assumed to be known.

Recent research has addressed the case of an imperfectly known mixing matrix. It is then necessary to estimate it (or at least some of its components) directly from the data. For instance, Tegmark et al. assume power-law emission spectra for all components except CMB and SZ, and fit spectral indices to the observations (Tegmark et al. 2000).

More recently, it has been proposed to resort to 'blind source separation' or 'independent component analysis' (ICA) methods. The work of Baccigalupi et al. (2000), further extended by Maino et al. (2002), implements a blind source separation method exploiting the non-Gaussianity of the sources for their separation. This infomax method, unfortunately, is not designed for noisy mixtures and cannot deal with a frequency-dependent beam.

The idea to use spectral diversity and an EM algorithm for the blind separation of components in CMB observations was proposed first by Snoussi et al. (2001). This approach exploits the spectral diversity of components as in our MDMC spectral matching, but assumes the prior knowledge of the spatial power spectra of the components. Our approach extends further on this idea, with significantly more flexibility, and the new point of view that spatial power spectra are actually the main unknown parameters of interest for CMB observations.

Other reports of blind component separation in Astronomical data include Nuzillard & Bijaoui (2000) and Funaro, Oja & Valpola (2001).

5.2 Comments on the spectral matching approach

Robustness. Our approach assumes that the data are collected in the form of a linear mixture of a known number of components that are independent, have different spatial power spectra, and different laws of emission as a function of frequency. These assumptions are valid in the three-component mixtures used in our simulations. Applying this method to real data obtained with the *Archeops* experiment (Benoît et al. 2003) gave us the opportunity to verify that the method is quite robust, with satisfactory performance even when the noise is neither white nor stationary, and when some residual systematic effects remain in the data. Of course, the exact impact of large departures from the model remains to be tested on a case by case basis.

Detector-dependent beams. It is quite usual in CMB observations that, because of the diffraction limit, the resolution of the available maps depends on frequency. For *Planck*, the resolution ranges from about 30 arcmin at 30 GHz to 4.5 arcmin at 350 GHz

and higher. It is mandatory that a method combining all observations benefit from the full resolution of the highest frequency channels. MDMC spectral matching, being implemented in Fourier (or spherical harmonics) space, permits one to take beam effects into account straightforwardly by including in the model the effect of a transfer function.

Identifying components. In practice, MDMC spectral matching runs with a fixed number of components. This number might not be well known (or even not very well defined), and must be guessed (or assumed). For CMB applications, an educated guess can be made (as usual for all component separation methods).

A practical way to handle this issue consists of applying the method several times with a increasing number of expected components. Comparing successive results permits one to identify 'stable' components, which remain essentially unchanged when more components are sought. Too few components result in unsatisfactory identification and poor adjustment of the model to the empirical spectrum. Too many components result in the separation of artificial components, either very weak or single-detector noise maps.

With this strategy, the method can be viewed as a component *discovery* tool, which can be useful in particular to uncover and separate out instrumental effects behaving as additional components.

Connected to the issue of component identification is the uniqueness (or identifiability) problem. As discussed above, MDMC spectral matching uses spectral diversity as the 'engine' of blind separation: components with proportional spatial power spectra (or nearly so) are not (or poorly) separated. In the current test, the three components are different enough that no such problem arises. In richer mixtures, containing contributions from several galactic components, it is quite possible that spectral diversity does not hold. If, for instance, several galactic components had a spatial power spectrum proportional to $1/k^3$, the method would satisfactorily estimate parameters relevant to the CMB and the SZ effect, but would be unable to un-mix Galactic contributions. A way out is to use a semi-blind approach in which some entries of the mixing matrix are forced to zero when the contribution of a particular component at a particular frequency is known to be negligible. This is the object of forthcoming research.

5.3 Comments on the Wiener inversion

After adjusting the parameters of the model to the data, the recovered mixing matrix, spectra, and noise levels can be used for component separation by Wiener inversion.

Quite interestingly, the Wiener filter can be implemented for identified components even if some sub-mixtures are not identified (for instance by lack of spectral diversity). It can be shown straightforwardly that the Wiener form

$$\mathbf{W} = \left[\mathbf{A}^\dagger \mathbf{R}_N^{-1} \mathbf{A} + \mathbf{R}_S^{-1}\right]^{-1} \mathbf{A}^\dagger \mathbf{R}_N^{-1} \tag{16}$$

can be rewritten equivalently as

$$\mathbf{W} = \mathbf{R}_S \mathbf{A}^\dagger [\mathbf{A}\mathbf{R}_S \mathbf{A}^\dagger + \mathbf{R}_N]^{-1} \tag{17}$$

or

$$\mathbf{W} = \mathbf{R}_S \mathbf{A}^\dagger \mathbf{R}_Y^{-1}. \tag{18}$$

Thus, the Wiener inversion for component j requires only an estimate of $\mathbf{R}_Y$ (readily available as $\widehat{\mathbf{R}}_Y$), of the spatial power spectrum of component j, and of the column of the mixing matrix $\mathbf{A}$ corresponding to component j. Therefore, *it is not necessary to identify all components, nor to know all spatial power spectra, nor to know noise levels, to separate the CMB from the other components*. We just need to know the CMB emission law (which we do) and its spatial power spectrum (which can be estimated blindly with our method).

As a final note, we stress that the Wiener method has the property of filtering the data spatially – an unpleasant fact when power spectra are estimated on separated maps. In contrast, MDMC spectral matching adjusts domain-averaged spatial power spectra on the data prior to component map separation (bypassing the need for power-spectrum estimation on output maps).

5.4 Comments on spectral estimation

In the above discussion, we have assumed for simplicity that the noise is spatially white for all detectors. This assumption, however, can be relaxed if needed, without (in general) loosing identifiability.

If the noise is uncorrelated between detectors, noise terms appear only on the diagonal of the multivariate power spectrum of the observations $\mathbf{R}_Y$. Off-diagonal terms contain only contributions from the off-diagonal terms of $\mathbf{A}\mathbf{R}_S\mathbf{A}^\dagger$. If noise power spectra are completely free, off-diagonal terms of $\widehat{\mathbf{R}}_Y$ constrain $\mathbf{A}\mathbf{R}_S\mathbf{A}^\dagger$, and diagonal terms serve to measure $\mathbf{R}_N$.

For instance, if the mixing matrix $\mathbf{A}$ is known, it is possible to adjust simultaneously the spatial power spectrum of the components and that of the noise on the data, as long as enough observations are available, which is generically the case.

If data from several experiments are analysed jointly, however, no correlated noise of instrumental origin is expected between data from detectors belonging to different experiments. This provides strong consistency checks, which ultimately provides an additional handle on the assessment of errors in the final results.

With a MDMC approach in Fourier (or spherical harmonic) space, data at different frequencies and with different beam sizes can be analysed jointly. This joint analysis can be done straightforwardly by stacking all observations from different instruments in the same vector of observations $\boldsymbol{Y}$, as long as they cover the same area of the sky. This is bound to become of major importance for the future scientific exploitation of multiscale and multifrequency data.

5.5 Using single-detector maps

For a well-calibrated instrument, the linear mixture model can be written in physical units, and the mixing matrix $\mathbf{A}$ depends only on the emission laws of components. Traditionally then, component separation is implemented on a set of maps per frequency channel (data from all detectors in each single frequency channel are combined into a single map). This approach should be preferred if good maps cannot be obtained independently for each detector (for sampling reasons, or because of striping...), and if all detector data at the same frequency can be combined (with some optimality) into one single map.

An alternate solution, when calibration coefficients and noise properties for individual detectors (levels, correlations between the noise of different detectors) are not known precisely, is to estimate parameters directly using single detector maps in readout units (e.g. microvolts), which can be done naturally with our spectral-matching method.

5.6 Comment on domain averaging

We have considered band-averaged spectra as in definition (3). In CMB studies, one may be more interested in quantities like $\ell(\ell+1)C_j(\ell)$ which are expected to vary more slowly than $C(\ell)$

itself. In this case, it may be more appropriate to perform bin averages as

$$\widetilde{\mathbf{R}}_Y(q) = \left(\sum_{\ell \in \mathcal{D}_q} \ell(\ell+1) \right)^{-1} \sum_{\ell \in \mathcal{D}_q} \ell(\ell+1)\, \mathbf{Y}(\boldsymbol{\ell}) \mathbf{Y}(\boldsymbol{\ell})^\dagger. \quad (19)$$

Spectral matching on such statistics would then yield estimates of

$$\widetilde{C}_j(q) = \left(\sum_{\ell \in \mathcal{D}_q} \ell(\ell+1) \right)^{-1} \sum_{\ell \in \mathcal{D}_q} \ell(\ell+1) C_j(\boldsymbol{\ell}). \quad (20)$$

This weighted band-averaging can be used in our MDMC spectral-matching method as well.

5.7 Extensions and application

The spectral matching method described in this paper resolves several problems relevant to the analysis of multidetector observations of noisy linear mixtures.

To a large extent, the efficiency of the method is due to the assumed structure of the data. In real observations, the model holds only approximately. For instance, some components may display anomalous behaviour in some regions of the sky. In this case, it may be worth cutting those regions out of the observations and treating them independently.

If the emission laws of some of the components vary slowly with position on the sky, then the method may be used on small maps rather than all-sky maps. A similar approach can be adopted in the case where spatial power spectra or noise levels are very dependent on sky coordinates.

Point sources induce correlated emissions across frequency channels, which cannot be modelled as a small set of additional components on the sky. One way to handle this problem is by detecting them with a matched filter, and blanking them out in the observations prior to applying the spectral matching.

To summarize, while MDMC spectral matching is a very powerful tool for CMB data analysis, it is not by itself a full reduction pipeline, and should be used for the most appropriate application, in conjunction with any other tool relevant to the problem at hand.

6 CONCLUSION

This paper describes a spectral matching method for blind source identification in noisy mixtures. The method adjusts a simple model of the data to the observations. We estimate a physically relevant set of parameters (fundamental parameters of the model: the mixing matrix, domain-averaged spatial power spectra of the sources and of noise) by maximum likelihood. Only unknown parameters are estimated, as the method lends itself easily to the modifications necessary to exploit partial prior information. Thanks to a Gaussian stationary model, the likelihood depends only on a reduced set of statistics (average spectral density matrices of the observations). An efficient, dedicated algorithm can adjust the parameters in just a few minutes on a modest workstation.

Our method is of particular relevance for CMB data analysis in a multidetector, multichannel mission such as *Planck*.

First, the method permits the blind separation of underlying components, hence, of emissions coming from different astrophysical sources. Obtaining clean maps of emissions resulting from distinct astrophysical processes is crucial to understanding their properties.

Secondly, the blind method permits one to estimate the number of components (by repeating the adjustment with a varying number of sources). This will be of utmost importance for analysing data from sensitive missions such as *Planck*, in particular for the identification and characterization of subdominant processes of foreground emission (e.g. free–free emission, non-thermal dust emission), or to track down systematic effects in the data.

Thirdly, the blind method can estimate the entries of the mixing matrix. This may be used, if needed, to constrain the emission law (electromagnetic spectrum) of the different components contributing to the mixture, which is essential for understanding their physical properties and possibly the emission processes.

Fourthly, if strong sources, for which the mixing matrix is well recovered, contribute to the mixture, the method can provide a useful tool for the intercalibration (or the absolute calibration) of the different detectors or of the different channels.

Fifthly, as our method is essentially a spectral matching method, which adjusts the spectra of a number of components to the observational data, it provides a direct measurement of the spatial power spectrum of the components in the mixture, of particular relevance for the CMB.

Sixthly, even though the above discussion assumed for simplicity a spatially white noise for all detectors, this assumption can be relaxed if needed, without (in general) loosing identifiability. The EM algorithm has to be adapted accordingly.

As a final word, let us emphasize that the method can be applied to sets of data from different experiments. As the MDMC spectral matching approach, implemented in Fourier space, straightforwardly accounts for beam effects, it can also be used for the blind, joint analysis of multiexperiment, multichannel, multidetector, multiresolution data as long as they cover the same area of the sky. The method may become an essential tool for mapping and analysing sources of emission observed with present and upcoming submillimetre experiments.

ACKNOWLEDGMENTS

We thank James G. Bartlett for a critical reading of the original manuscript. We acknowledge useful discussions with Mark Ashdown, Mike Hobson, Juan Macias, Ali Mohammad-Djafari, Hichem Snoussi, and the *Archeops* collaboration. This work was made possible by a grant from the French Ministère de la Recherche to stimulate interdisciplinary work for applying state-of-the-art signal and image processing techniques to CMB data analysis.

REFERENCES

Baccigalupi C. et al., 2000, MNRAS, 318, 769
Bennett C. L. et al., 1997, Am. Astron. Soc. Meeting, 191, 8701
Benoît A. et al., 2002, Astroparticle Phys., 17, 101
Benoît A. et al., 2003, A&A, 399, L23
Bersanelli M., Mandolesi R., 2000, Astrophys. Lett. Commun., 37, 171
Bouchet F. R., Gispert R., 1999, New Astron., 4, 443
Cardoso J.-F., Snoussi H., Delabrouille J., Patanchon G., 2002, in Proc. EUSIPCO Vol. 1, XI European Signal Processing Conf. Available online at http://www.carte-blanche.fr/∼eusipco2002/, p. 561
de Bernardis P. et al., 2000, Nat, 404, 955
Delabrouille J., Melin J., Bartlett J., 2002, in Chen L., Ma C., Ng K., Pen U., eds, ASP Conf. Ser. Vol. 257, AMiBA 2001: High-Z Clusters, Missing Baryons, and CMB Polarization. Astron. Soc. Pac., San Francisco, p. 81
Dempster A., Laird N., Rubin D., 1977, J. R. Statistical Soc. B, 39, 1
Funaro M., Oja E., Valpola H., 2001, in Lee J., Makeig S., eds, Proc. ICA2001. 3rd Int. Conf. Independent Component Analysis and Blind Signal Separation. Available on-line at http://ica2001.ucsd.edu

Hanany S. et al., 2000, AJ, 545, L5
Hobson M. P., Jones A. W., Lasenby A. N., Bouchet F. R., 1998, MNRAS, 300, 1
Lamarre J.-M. et al., 2000, Astrophys. Lett. Commun., 37, 161
Luenberger D. G., 1973, Introduction to Linear and Nonlinear Programming. Addison-Wesley, Reading, MA
Maino D. et al., 2002, MNRAS, 334, 53
Nuzillard D., Bijaoui A., 2000, A&AS, 147, 129
Pham D.-T., Garat P., 1997, IEEE Tr. SP, 45, 1712
Snoussi H., Patanchon G., Maciaz-Perez J., Mohammad-Djafari A., Delabrouille J., 2001, in Fry R. L., ed., Proc. MAXENT Workshop, Bayesian Interference and Maximum Entropy Methods. AIP, New York, p. 125
Tegmark M., Efstathiou G., 1996, MNRAS, 281, 1297
Tegmark M., Eisenstein D. J., Hu W., de Oliveira-Costa A., 2000, ApJ, 530, 133
Zaldarriaga M., Seljak U., 2000, ApJS, 129, 431

APPENDIX A: LINEAR COMPONENT SEPARATION

The separation of astrophysical components relies on the key assumption that the total sky emission at frequency ν is a linear superposition of a number of components as in equation (1). In principle, then, the observation of the sky emission at several frequencies $(\nu_1, \nu_2, \ldots)$ recovers estimates $\widehat{\boldsymbol{S}}_j(\theta, \phi)$ of the component templates $\boldsymbol{S}_j(\theta, \phi)$ by inverting equation (1). There are several methods for a linear inversion of the system when the mixing matrix $\mathbf{A}$ is known.

(i) If there are as many noiseless observations as there are astrophysical components contributing to the total emission, by simple inversion of the square matrix $\mathbf{A}$, so that the recovered components, $\widehat{\boldsymbol{S}}$, are given by $\widehat{\boldsymbol{S}} = \mathbf{A}^{-1}\boldsymbol{Y}$.

(ii) If there are more observations than astrophysical components, the system can be inverted using the pseudo-inverse, $\widehat{\boldsymbol{S}} = [\mathbf{A}^\dagger\mathbf{A}]^{-1}\mathbf{A}^\dagger\boldsymbol{Y}$.

(iii) For optimal signal-to-noise ratio under Gaussian statistics, without other prior assumptions on the astrophysical components, one can use a generalized least-squares solution, $\widehat{\boldsymbol{S}} = [\mathbf{A}^\dagger\mathbf{R}_N^{-1}\mathbf{A}]^{-1}\mathbf{A}^\dagger\mathbf{R}_N^{-1}\boldsymbol{Y}$, where $\mathbf{R}_N$ is the noise correlation matrix.

(iv) The choice $\widehat{\boldsymbol{S}} = [\mathbf{A}^\dagger\mathbf{R}_N^{-1}\mathbf{A} + \mathbf{R}_S^{-1}]^{-1}\mathbf{A}^\dagger\mathbf{R}_N^{-1}\boldsymbol{Y} = \mathbf{W}\boldsymbol{Y}$ is the Wiener solution. It is the linear solution which minimizes the variance of the error, but requires knowledge of both the noise autocorrelation, $\mathbf{R}_N$, and of the component autocorrelation, $\mathbf{R}_S$. As $[\mathbf{WA}]_{ii} \leqslant 1$, this solution modifies the spatial spectra of the components since different weights are given to different spatial frequencies of a component map.

(v) The renormalized Wiener solution, $\widehat{\boldsymbol{S}} = \mathbf{\Lambda}\mathbf{W}\boldsymbol{Y}$, where $\mathbf{\Lambda} = [\mathrm{diag}(\mathbf{WA})]^{-1}$, is the Wiener solution under the constraint $[\mathbf{WA}]_{ii} = 1$. This solution renormalizes the Wiener solution at each spatial frequency, so that no spatial filtering is applied to the data.

In the above list, solution (i) is the special case of (ii) when $\mathbf{A}$ is square and regular; (ii) is the special case of 3 when the noise is white ($\mathbf{R}_N \propto \mathrm{Id}$); (iii) the special case of (iv) when the signal is much stronger than the noise; and (v) a constrained version of (iv) that does not modify the relative importance of different spatial frequencies in a component map after inversion. Depending on the chosen method, one or more of $\mathbf{A}$, $\mathbf{R}_N$ and $\mathbf{R}_S$ (which can be considered as parameters of the model) is needed to implement the inversion.

Realizing the fact that optimal component separation requires prior knowledge of a set of parameters of the model is one of the driving ideas behind our MDMC spectral-matching approach: we implement the *joint estimation* of all such parameters that are not necessarily known a priori.

APPENDIX B: SPECTRAL MATCHING AND LIKELIHOOD

This section shows that minimizing the spectral matching criterion (11) is equivalent to maximizing the likelihood of a simple model.

Gaussian likelihood and covariance matching. We first show how criterion (11) is related to a Gaussian likelihood. If y is a real $n \times 1$ zero mean Gaussian random vector with covariance matrix $\mathbf{R}$, then

$$-2\log p(\boldsymbol{y}) = \boldsymbol{y}^\dagger\mathbf{R}^{-1}\boldsymbol{y} + \log\det(2\pi\mathbf{R}). \tag{B1}$$

If $\mathbf{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T]$ is an $n \times T$ matrix made of T such vectors, independent from each other, with $\mathrm{Cov}(\boldsymbol{y}_t) = \mathbf{R}_t$, then

$$-2\log p(\mathbf{Y}) = \sum_{t=1}^{T} \boldsymbol{y}_t^\dagger\mathbf{R}_t^{-1}\boldsymbol{y}_t + \log\det(2\pi\mathbf{R}_t). \tag{B2}$$

Assume further that the index set $[1, \ldots, T]$ can be decomposed in Q subsets $I_1, \ldots, I_Q$ such that R_t is constant with value $\mathbf{R}(q)$ over the qth subset, that is, $\mathbf{R}_t = \mathbf{R}(q)$ if $t \in I_q$. Then, equation (B12) can be rewritten, using $\boldsymbol{y}^\dagger\mathbf{R}^{-1}\boldsymbol{y} = \mathrm{tr}(\mathbf{R}^{-1}\boldsymbol{y}\boldsymbol{y}^\dagger)$, as

$$-2\log p(\mathbf{Y}) = \sum_{q=1}^{Q} n_q\left[\mathrm{tr}\widehat{\mathbf{R}}(q)(\mathbf{R}(q)^{-1}) + \log\det(\mathbf{R}(q))\right] + \text{constant}$$

where $\widehat{\mathbf{R}}(q) = \frac{1}{n_q}\sum_{t\in I_q}\boldsymbol{y}_t\boldsymbol{y}_t^\dagger$ and n_q is the number of indices in I_q. This last expression also reads

$$-2\log p(\mathbf{Y}) = \sum_{q=1}^{Q} n_q D(\widehat{\mathbf{R}}_y(q), \mathbf{R}_y(q)) + \text{constant} \tag{B3}$$

where the constant term is a function of the data Y via $\widehat{\mathbf{R}}_y(q)$ but not of any $\mathbf{R}(q)$. This form makes it clear that the mismatch (11) corresponds to the log-likelihood of a sequence of zero mean Gaussian vectors which are modelled as having blockwise identical covariance matrices.

Whittle approximation. The statistical distribution of the Fourier coefficients of a stationary time series is a well researched topic. If T samples $y(1), \ldots, y(T)$ of an n-variate discrete time series are available, the Fourier transform is:

$$\widetilde{y}(f) = \frac{1}{\sqrt{T}}\sum_{t=0}^{T-1} y(t)\exp -2\mathrm{i}\pi f t. \tag{B4}$$

For a stationary time series with spectral covariance matrix $R(f)$, simple asymptotic (for large T) results are available. In particular, the Whittle approximation consists of approximating the distribution of the Fourier transform $\widetilde{y}(f)$ at DFT points $f = q/T$ as follows.

(i) The real part and the imaginary part of $\widetilde{y}(f)$ are Gaussian, uncorrelated, with the same covariance matrix and $E\widetilde{y}(f)\widetilde{y}(f)^\dagger = R(f)$.

(ii) For $0 < p \neq p' < T/2$ (assuming T even and for p, p' integers), $\widetilde{y}(p/T)$ is uncorrelated with $\widetilde{y}(p'/T)$.

This is a standard approximation: it has been used for the blind separation of noise free mixtures of components by Pham & Garat

(1997) and in the context of astronomical component separation by e.g. Bouchet & Gispert (1999) and Tegmark & Efstathiou (1996).

Expression (B3) thus shows[3] that the minimization of (11) is equivalent to maximizing (the Whittle approximation to) the likelihood provided we model the spectra of the sources as being constant over spectral domains.

APPENDIX C: AN EM ALGORITHM IN THE SPECTRAL DOMAIN

The expectation maximization (EM) algorithm (Dempster et al. 1977) is a popular technique for computing maximum likelihood estimates. This section first briefly reviews the general mechanism of EM and then shows its specific form when applied to our model.

The EM algorithm. Consider a probability model $p(y, s \mid \theta)$ for a pair (y, s) of random variables with θ a parameter set. If the variable s is not observed, the log-likelihood of the observed y is

$$l(\theta) = \log p(y \mid \theta) = \log \int p(y, s \mid \theta)\, \mathrm{d}s. \tag{C1}$$

For some statistical models, the maximization of the log-likelihood $l(\theta)$ can be made easier by considering the EM functional:

$$l(\theta, \theta') = \int \log(p(y, s \mid \theta))\; p(s \mid y, \theta')\, \mathrm{d}s. \tag{C2}$$

The EM algorithm is an iterative method which computes a sequence of estimates according to

$$\theta^{(n)} \rightarrow \theta^{(n+1)} = \arg\max_{\theta} l\left(\theta, \theta^{(n)}\right). \tag{C3}$$

It can be shown that

$$l(\theta'', \theta') > l(\theta', \theta') \Rightarrow l(\theta'') > l(\theta'), \tag{C4}$$

meaning that every step of the algorithm can only increase the likelihood. Actually, a stationary point of the algorithm is also a stationary point of the likelihood since

$$\frac{\partial l(\theta)}{\partial \theta} = \left.\frac{\partial l(\theta, \theta')}{\partial \theta}\right|_{\theta'=\theta}. \tag{C5}$$

The EM algorithm is an interesting technique for maximizing the likelihood if (i) the computation of the conditional expectation in definition (C2) (E step) and (ii) the maximization (C3) of the functional (M step) are both computationally tractable.

Both the E step and the M step turn out to be straightforward because one elementary EM step amounts to solving

$$0 = \int \frac{\partial \log\left(p\left(y, s \,\middle|\, \theta^{(n+1)}\right)\right)}{\partial \theta}\, p\left(s \,\middle|\, y, \theta^{(n)}\right) \mathrm{d}s. \tag{C6}$$

In our model, the partial derivative in (C6) turns out to be a simple function of y and s, allowing the conditional expectation to be easily computed and equation (C6) to be easily solved. This is sketched in the following.

[3] Actually some care is required in dealing with the fact that the Fourier coefficients are complex-valued and that $\widetilde{y}(-f) = \widetilde{y}(f)^{\star}$. This introduces some minor complications in the computations, but does not affect the final result.

A single Gaussian vector. In order to introduce the necessary notations, we start by considering a simple case where $\boldsymbol{y} = \mathbf{A}\boldsymbol{s} + \boldsymbol{n}$ where $\boldsymbol{s}$ and $\boldsymbol{n}$ are independent Gaussian vectors with zero-mean and covariance matrices equal to $\mathbf{R}_s$ and $\mathbf{R}_n$ respectively. Then the parameter set is $\theta = (\mathbf{A}, \mathbf{R}_s, \mathbf{R}_n)$ and one has

$$-2\log p(\boldsymbol{y} \mid \boldsymbol{s}, \theta) = (\boldsymbol{y} - \mathbf{A}\boldsymbol{s})^{\dagger}\mathbf{R}_n^{-1}(\boldsymbol{y} - \mathbf{A}\boldsymbol{s}) + \log|\mathbf{R}_n| + \text{constant},$$

$$-2\log p(\boldsymbol{s} \mid \theta) \quad = \boldsymbol{s}^{\dagger}\mathbf{R}_s^{-1}\boldsymbol{s} + \log|\mathbf{R}_s| + \text{constant}.$$

Using $p(\boldsymbol{y}, \boldsymbol{s}) = p(\boldsymbol{y} \mid \boldsymbol{s})p(\boldsymbol{s})$, the log derivatives of the joint density with respect to the components of θ are

$$\frac{\partial \log p(\boldsymbol{y}, \boldsymbol{s} \mid \theta)}{\partial \mathbf{A}} = \mathbf{R}_n^{-1}\left[(\boldsymbol{y} - \mathbf{A}\boldsymbol{s})\boldsymbol{s}^{\dagger}\right], \tag{C7}$$

$$\frac{\partial \log p(\boldsymbol{y}, \boldsymbol{s} \mid \theta)}{\partial \mathbf{R}_n^{-1}} = -\frac{1}{2}\left[(\boldsymbol{y} - \mathbf{A}\boldsymbol{s})(\boldsymbol{y} - \mathbf{A}\boldsymbol{s})^{\dagger} - \mathbf{R}_n\right], \tag{C8}$$

$$\frac{\partial \log p(\boldsymbol{y}, \boldsymbol{s} \mid \theta)}{\partial \mathbf{R}_s^{-1}} = -\frac{1}{2}\left[\boldsymbol{s}\boldsymbol{s}^{\dagger} - \mathbf{R}_s\right]. \tag{C9}$$

Thus, in this simple model, computing the conditional expectations as in equation (C6) would boil down to evaluating the conditional expectations of the random variables $\boldsymbol{s}\boldsymbol{s}^{\dagger}$, $\boldsymbol{s}\boldsymbol{y}^{\dagger}$, $\boldsymbol{y}\boldsymbol{s}^{\dagger}$ and $\boldsymbol{y}\boldsymbol{y}^{\dagger}$. This is a routine matter in a Gaussian model $\boldsymbol{y} = \mathbf{A}\boldsymbol{s} + \boldsymbol{n}$ for which one finds

$$E(\boldsymbol{s}\boldsymbol{s}^{\dagger} \mid \boldsymbol{y}, \theta) = \mathbf{W}(\theta)\boldsymbol{y}\boldsymbol{y}^{\dagger}\mathbf{W}(\theta)^{\dagger} + \mathbf{C}(\theta), \tag{C10}$$

$$E(\boldsymbol{s}\boldsymbol{y}^{\dagger} \mid \boldsymbol{y}, \theta) = \mathbf{W}(\theta)\boldsymbol{y}\boldsymbol{y}^{\dagger}, \tag{C11}$$

$$E(\boldsymbol{y}\boldsymbol{s}^{\dagger} \mid \boldsymbol{y}, \theta) = \boldsymbol{y}\boldsymbol{y}^{\dagger}\mathbf{W}(\theta)^{\dagger}, \tag{C12}$$

$$E(\boldsymbol{y}\boldsymbol{y}^{\dagger} \mid \boldsymbol{y}, \theta) = \boldsymbol{y}\boldsymbol{y}^{\dagger}, \tag{C13}$$

with the following definitions for matrices $\mathbf{C}(\theta)$ and $\mathbf{W}(\theta)$:

$$\mathbf{C}(\theta) = \left(\mathbf{A}^{\dagger}\mathbf{R}_n^{-1}\mathbf{A} + \mathbf{R}_s^{-1}\right)^{-1}, \tag{C14}$$

$$\mathbf{W}(\theta) = \left(\mathbf{A}^{\dagger}\mathbf{R}_n^{-1}\mathbf{A} + \mathbf{R}_s^{-1}\right)^{-1}\mathbf{A}^{\dagger}\mathbf{R}_n^{-1}. \tag{C15}$$

Note that $\mathbf{C}(\theta) = \mathrm{Cov}(\boldsymbol{s} \mid \boldsymbol{y}, \theta)$ and that $\mathbf{W}(\theta)$ is the Wiener filter, that is $E(\boldsymbol{s} \mid \boldsymbol{y}, \theta) = \mathbf{W}(\theta)\boldsymbol{y}$.

The EM algorithm in the Whittle approximation. In our model, according to the Whittle approximation, the DFT points $y(k)$ are independent so that the EM functional (C2) for the whole data set simply is a sum over DFT frequencies of elementary functionals. Thus an EM step $\theta' \rightarrow \theta$ consists of solving

$$0 = \sum_k E\left\{\frac{\partial}{\partial\theta}\log p\left(\boldsymbol{y}(k), \boldsymbol{s}(k) \mid \theta\right) \mid \boldsymbol{y}(k), \theta'\right\}. \tag{C16}$$

To proceed further, equation (C16) is specialized to the case of interest by using two ingredients. First, we use the relation $y(k) = As(k) + n(k)$ and the Gaussianity of each pair $(y(k), s(k))$; this is expressed via equations (C7)–(C9). Secondly, we use the approximation that the power spectra are constant over each spectral domain. Combining these properties, the cancellation (C16) of the gradient with respect to $\mathbf{A}$, $\mathbf{R}_n$ and each $\mathbf{R}_s(q)$ yields

$$0 = \widetilde{\mathbf{R}}_{ys}(\theta') - \mathbf{A}(\theta)\widetilde{\mathbf{R}}_{ss}(\theta'), \tag{C17}$$

$$\begin{aligned} 0 = {} & \widetilde{\mathbf{R}}_{yy}(\theta') - \mathbf{A}(\theta)\widetilde{\mathbf{R}}_{sy}(\theta') - \widetilde{\mathbf{R}}_{ys}(\theta')\mathbf{A}(\theta)^{\dagger} \\ & + \mathbf{A}(\theta)\widetilde{\mathbf{R}}_{ss}(\theta')\mathbf{A}(\theta)^{\dagger} - \mathbf{R}_n, \end{aligned} \tag{C18}$$

$$0 = \widetilde{\mathbf{R}}_{ss}(\theta', q) - \mathbf{R}_s(\theta, q) \quad (q = 1, \ldots, Q), \tag{C19}$$

where we have defined the matrix

$$\widetilde{\mathbf{R}}_{ss}(\theta, q) = \frac{1}{n_q} \sum_{k \in \mathcal{D}_q} E\left(\boldsymbol{s}(k)\boldsymbol{s}(k)^\dagger \mid \boldsymbol{y}(k), \theta\right) \tag{C20}$$

and its weighted average over all domains

$$\widetilde{\mathbf{R}}_{ss}(\theta) = \sum_{q=1}^{Q} \frac{n_q}{n} \widetilde{\mathbf{R}}_{ss}(\theta, q). \tag{C21}$$

The same definitions hold for $\widetilde{\mathbf{R}}_{sy}(q, \theta)$ [resp. $\widetilde{\mathbf{R}}_{yy}(q, \theta)$] as an averaged conditional expectation of $\boldsymbol{s}(k)\boldsymbol{y}(k)^\dagger$ [resp. $\boldsymbol{y}(k)\boldsymbol{y}(k)^\dagger$] and $\widetilde{\mathbf{R}}_{sy}(\theta)$ [resp. $\widetilde{\mathbf{R}}_{yy}(\theta)$] as its weighted average over spectral domains.

Equations (C17)–(C19) are readily solved for unconstrained $\mathbf{A}$, $\mathbf{R}_n$ and $\mathbf{R}_s(q)$. Recall however that our model involves *diagonal* covariance matrices so that the actual parameter set is $(\mathbf{A}, C_j(q), \sigma_d^2))$. This constraint, however, preserves the simplicity of the solution of the M step since it suffices to use the diagonal parts of the solutions of (C17)–(C19). Thus, the M step boils down to

$$\mathbf{A} = \widetilde{\mathbf{R}}_{ys}(\theta')\widetilde{\mathbf{R}}_{ss}(\theta')^{-1}, \tag{C22}$$

$$\sigma_i^2 = \left[\widetilde{\mathbf{R}}_{yy}(\theta') - \widetilde{\mathbf{R}}_{ys}(\theta')\widetilde{\mathbf{R}}_{ss}(\theta')^{-1}\widetilde{\mathbf{R}}_{sy}(\theta')\right]_{ii}, \tag{C23}$$

$$P_i(q) = [\widetilde{\mathbf{R}}_{ss}(\theta', q)]_{ii}. \tag{C24}$$

The E-step of the algorithm essentially consists of computing the conditional covariance matrices $\widetilde{\mathbf{R}}_{\times\times}(q)$. In this step, the linearity and the Gaussianity of the model, together with the domain approximation, again provides us with significant computational savings. Indeed, matrices $\mathbf{C}$ and $\mathbf{W}$ defined at equations (C14) and (C15) are actually constant over each spectral domain so that the E-step is implemented by the following computations which directly stem from (C14)–(C15) and from equations (C10)–(C13):

$$\mathbf{C}(q) = \left(\mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{A} + \mathbf{R}_s(q)^{-1}\right)^{-1}, \tag{C25}$$

$$\mathbf{W}(q) = \left(\mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{A} + \mathbf{R}_s(q)^{-1}\right)^{-1} \mathbf{A}^\dagger \mathbf{R}_n^{-1}, \tag{C26}$$

$$\widetilde{\mathbf{R}}_{ss}(q) = \mathbf{W}(q)\hat{\mathbf{R}}_y(q)\mathbf{W}(q)^\dagger + \mathbf{C}(q), \tag{C27}$$

$$\widetilde{\mathbf{R}}_{sy}(q) = \mathbf{W}(q)\hat{\mathbf{R}}_y(q). \tag{C28}$$

From this, one easily reaches the EM algorithm as described at algorithm 1. The description of this procedure is completed by specifying the initialization, the rescaling of the parameters and the stopping rule, as briefly discussed next.

Some comments on EM implementation. Rescaling is required because, as noted above, the model is not completely identifiable: the spectral density matrices $\mathbf{R}_Y$ are unaffected by the exchange of a scalar factor between each column of $\mathbf{A}$ and each component's power spectrum. We have found that this inherent indetermination must be fixed in order for EM to converge. Our strategy is, after each EM step, to fix the norm of each column of $\mathbf{A}$ to unity and to adjust the corresponding power spectra accordingly. This is an arbitrary choice which happens to work well in practice.

The algorithm is initialized with the following parameters. We take $\mathbf{R}_n$ to be $\mathrm{diag}(\widehat{\mathbf{R}}_y)$ where $\widehat{\mathbf{R}}_y = \sum_q \frac{n_q}{n} \widehat{\mathbf{R}}_y(q)$. This is a gross overestimation since it amounts to assuming no signal and only noise. The initial value of $\mathbf{A}$ is obtained by using the N_c dominant eigenvectors of $\widehat{\mathbf{R}}_y$ as the N_c columns of $\mathbf{A}$. Again, this is nothing like any real *estimate* of $\mathbf{A}$, but rather a vague guess in 'the right direction'. Finally, the spectra $P_i(q)$ are taken as the diagonal entries of $\mathbf{A}^\dagger \widehat{\mathbf{R}}_y(q)\mathbf{A}$ which would be a correct estimate in the noise free case if $\mathbf{A}$ itself was. This ad hoc initialization procedure seems satisfactory. Note that it is a common rule of thumb to initialize EM with overestimated noise power.

Regarding the stopping rule, recall (from Section 3.2) that the EM algorithm is only used 'halfway' to the maximum of the likelihood and maximization is completed by a quasi-Newton technique. For this reason, there is little point in devising a sophisticated stopping strategy: in practice, the algorithm is run for a pre-specified number of steps (based on a few preliminary experiments with the data).

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