Blind Component Separation in Wavelet Space: Application to CMB Analysis

Y. Moudden

DAPNIA/SEDI-SAP, CEA/Saclay, 91191 Gif-sur-Yvette, France
Email: yassir.moudden@cea.fr

J.-F. Cardoso

CNRS, École National Supérieure des Télécommunications, 46 rue Barrault, 75634 Paris, France
Email: cardoso@tsi.enst.fr

J.-L. Starck

DAPNIA/SEDI-SAP, CEA/Saclay, 91191 Gif-sur-Yvette, France
Email: jstarck@cea.fr

J. Delabrouille

CNRS/PCC, Collège de France, 11 place Marcelin Berthelot, 75231 Paris, France
Email: delabrouille@cdf.in2p3.fr

It is a recurrent issue in astronomical data analysis that observations are incomplete maps with missing patches or intentionally masked parts. In addition, many astrophysical emissions are nonstationary processes over the sky. All these effects impair data processing techniques which work in the Fourier domain. Spectral matching ICA (SMICA) is a source separation method based on spectral matching in Fourier space designed for the separation of diffuse astrophysical emissions in cosmic microwave background observations. This paper proposes an extension of SMICA to the *wavelet* domain and demonstrates the effectiveness of wavelet-based statistics for dealing with gaps in the data.

Keywords and phrases: blind source separation, cosmic microwave background, wavelets, data analysis, missing data.

1. INTRODUCTION

The detection of cosmic microwave background (CMB) anisotropies on the sky has been over the past three decades a subject of intense activity in the cosmology community. The CMB, discovered in 1965 by Penzias and Wilson, is a relic radiation emitted some 13 billion years ago, when the universe was about 370 000 years old. Small fluctuations of this emission, tracing the seeds of the primordial inhomogeneities which gave rise to present large scale structures as galaxies and clusters of galaxies, were first discovered in the observations made by COBE [1] and further investigated by a number of experiments among which Archeops [2], boomerang [3], maxima [4], and WMAP [5].

The precise measurement of these fluctuations is of utmost importance to cosmology. Their statistical properties (spatial power spectrum, Gaussianity) strongly depend on the cosmological scenarios describing the properties and evolution of our universe as a whole, and thus permit to constrain these models as well as to measure the cosmological parameters describing the matter content, the geometry, and the evolution of our universe [6].

Accessing this information, however, requires disentangling in the data the contributions of several distinct astrophysical sources, all of which emit radiation in the frequency range used for CMB observations [7]. This problem of component separation, in the field of CMB studies, has thus been the object of many dedicated studies in the past.

To first order, the total sky emission can be modeled as a linear superposition of a few independent processes. The observation of the sky in direction (θ, φ) with detector d is then a noisy linear mixture of N_c components:

$$x_d(\vartheta, \varphi) = \sum_{j=1}^{N_c} A_{dj} s_j(\vartheta, \varphi) + n_d(\vartheta, \varphi), \tag{1}$$

where s_j is the emission template for the jth astrophysical process, herein referred to as a *source* or a *component*.

The coefficients A_{dj} reflect emission laws while n_d accounts for noise. When N_d detectors provide independent observations, this equation can be put in vector-matrix form:

$$X(\vartheta, \varphi) = AS(\vartheta, \varphi) + N(\vartheta, \varphi), \qquad (2)$$

where X and N are vectors of length N_d, S is a vector of length N_c, and A is the $N_d \times N_c$ mixing matrix.

Given the observations of such a set of independent detectors, component separation consists in recovering estimates of the maps of the sources $s_j(\vartheta, \varphi)$. Explicit component separation has been investigated first in CMB applications by [7, 8, 9]. In these applications, recovering component maps is the primary target, and all the parameters of the model (mixing matrix A_{dj}, noise levels, statistics of the components, including the spatial power spectra) are assumed to be known and are used to invert the linear system.

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 entries) 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 [10].

More recently, blind source separation or independent component analysis (ICA) methods have been implemented specifically for CMB studies. The work of Baccigalupi et al. [11], further extended by Maino et al. [12], implements a blind source separation method exploiting the non-Gaussianity of the sources for their separation, which permits to recover the mixing matrix A and the maps of the sources. Accounting for spatially varying instrumental noise in the observation model is investigated by Kuruoglu et al. in [13], as well as the possible inclusion of prior information about the distributions of the components using a generic Gaussian mixture model.

Snoussi et al. [14] propose a Bayesian approach in the Fourier domain assuming known spectra for the components as well as possibly non-Gaussian priors for the Fourier coefficients of the components. A fully blind, maximum likelihood approach is developed in [15, 16], with the new point of view that spatial power spectra are actually the main unknown parameters of interest for CMB observations. A key benefit is that parameter estimation can then be based on a set of band-averaged spectral covariance matrices, considerably compressing the data size.

Working in the frequency domain offers several benefits but the nonlocality of the Fourier transform creates some difficulties. In particular, one may wish to avoid the averaging induced by the nonlocality of the Fourier transform when dealing with strongly nonstationary components or noise. In addition, in many experiments, only an incomplete sky coverage is available. Either the instrument observes only a fraction of the sky or some regions of the sky must be masked due to localized strong astrophysical sources of contamination: compact radio sources or galaxies, strong emitting regions in the galactic plane. These effects can be mitigated in a simple manner thanks to the localization properties of wavelets.

Blind component separation (and in particular estimation of the mixing matrix), as discussed by Cardoso [17], can be achieved in several different ways. The first of these exploits non-Gaussianity of all, but possibly one, components. The component separation method of Baccigalupi [11] and Maino [12] belongs to this set of techniques. In CMB data analysis, however, the main component of interest (the CMB itself) has a Gaussian distribution and the observed mixtures suffer from additive Gaussian noise, so that better performance can be expected from methods based on Gaussian models. A second set of techniques exploits spectral diversity and works in the Fourier domain. It has the advantage that detector–dependent beams can be handled easily, since the convolution with a point spread function in direct space becomes a simple product in Fourier space. SMICA follows this approach in the context of noisy observations. Finally, a third set of methods exploits nonstationarity. It is adapted to situations where components are strongly nonstationary in real space.

It is natural to investigate the possible benefits of exploiting both nonstationarity and spectral diversity for blind component separation using wavelets. Indeed wavelets are powerful tools in revealing the spectral content of nonstationary data. Although blind source separation in the wavelet domain has been previously examined, the setting here is different. We should mention, for instance, the separation method in [18] which is based on the non-Gaussianity of the source signals but after a *sparsifying* wavelet transform and the Bayesian approach in [19] which adopts a similar point of view although with a richer source model accounting for correlations in the wavelet representation.

The paper is organized as follows. In Section 2, we first recall the principle of spectral matching ICA. Then, after a brief reminder of some properties of the *à trous* wavelet transform, we discuss in Section 3 the extension of SMICA to component separation in wavelet space in order to deal with nonstationary data. Considering the problem of incomplete data as a model case of practical significance for the comparison of SMICA and its extension wSMICA, numerical experiments and results are reported in Section 4.

2.   SMICA

Spectral matching ICA, or SMICA for short, is a blind source separation technique which, unlike most standard ICA methods, is able to recover Gaussian sources in noisy contexts. It operates in the spectral domain and is based on *spectral diversity*: it is able to separate sources provided they have different power spectra. This section gives a brief account of SMICA. More details can be found in [16]; first applications to CMB analysis are in [16, 20].

2.1.   Model and cost function

For a second-order stationary N_d-dimensional process, we denote by $R_X(\nu)$ the $N_d \times N_d$ spectral covariance matrix at frequency ν, that is, the (i, i)th entry of $R_X(\nu)$ is the power spectrum of the ith coordinate of X while the offdiagonal entries

of $R_X(\nu)$ contain the cross-spectra between the entries of X. If X follows the linear model of (2) with independent additive noise, then its spectral covariance matrix is structured as

$$R_X(\nu) = AR_S(\nu)A^\dagger + R_N(\nu) \qquad (3)$$

with $R_S(\nu)$ and $R_N(\nu)$ being the spectral covariance matrices of S and N, respectively. The assumption of independence between the underlying components implies that $R_S(\nu)$ is a diagonal matrix. We will also assume independence of the noise processes between detectors: matrix $R_N(\nu)$ also is a diagonal matrix.

In the definition of $R_X(\nu)$, we have not explicitly defined the frequency ν. This is because SMICA can be applied for the separation of components in many contexts: each observation X_d can be a time series (one-dimensional), an image (two-dimensional random fields), a random field on the sphere (as in full-sky CMB studies). In each case, the appropriate notions of frequency, stationarity, and power spectrum should be used.

SMICA estimates all (or a subset of) the model parameters

$$\theta = \{R_S(\nu_q), R_N(\nu_q), A\} \qquad (4)$$

by minimizing a measure of "spectral mismatch" between sample estimates $\widehat{R}_X(\nu)$ (defined below) of the spectral covariance matrices and their ensemble averages which depend on the parameters according to (3). More specifically, an estimate $\widehat{\theta} = \{\widehat{R}_S(\nu_q), \widehat{R}_N(\nu_q), \widehat{A}\}$ is obtained as $\widehat{\theta} = \mathrm{argmin}_\theta \phi(\theta)$ where the measure of spectral mismatch $\phi(\theta)$ is defined by

$$\phi(\theta) = \sum_{q=1}^{Q} \alpha_q \mathcal{D}\left(\widehat{R}_X(\nu_q), AR_S(\nu_q)A^\dagger + R_N(\nu_q)\right). \qquad (5)$$

Here, $\{\nu_q | 1 \le q \le Q\}$ is a set of frequencies, $\{\alpha_q | 1 \le q \le Q\}$ is a set of positive weights, and $\mathcal{D}(\cdot, \cdot)$ is a measure of mismatch between two positive matrices.

This approach is a particular instance of moment matching. As such, if consistent estimates $\widehat{R}_X(\nu_q)$ of the spectral covariance matrices $R_X(\nu_q)$ are available and if the model is identifiable, then any reasonable choice of the weights α_q and of the divergence measure $\mathcal{D}(\cdot, \cdot)$ should lead to *consistent* estimates of the parameters. However, this does not mean that these choices should be arbitrary: in our standard implementation, we make specific choices (described next) in such a way that minimizing $\phi(\theta)$ is identical to maximizing the likelihood of θ in a model of Gaussian stationary processes. Hence, these choices guarantee a good statistical efficiency when the underlying processes are well modeled as Gaussian stationary processes. When this is not the case, though, the performance of SMICA may not be as good as (but not necessarily worse than) the performance of other methods designed to capture other aspects of the statistical distribution of the data, such as non-Gaussian features, for instance.

Given a data set, denote by $\widetilde{X}(\nu)$ its discrete Fourier transform at frequency ν and denote by $\{F_q | 1 \le q \le Q\}$ a set of Q frequency domains with F_q centered around frequency ν_q. Spectral covariance matrices are estimated nonparametrically by

$$\widehat{R}_X(\nu_q) = \frac{1}{n_q} \sum_{\nu \in F_q} \widetilde{X}(\nu)\widetilde{X}(\nu)^\dagger, \qquad (6)$$

where n_q denotes the number of Fourier points $\widetilde{X}(\nu)$ in the spectral domain F_q. We always use symmetric domains in the sense that frequency ν belongs to F_q if and only if $-\nu$ also does. This symmetry guarantees that $\widehat{R}_X(\nu_q)$ is always a real-valued matrix when X itself is a real-valued process.

In its standard form, the SMICA technique uses positive weights $\alpha_q = n_q$ and a divergence $\mathcal{D}$ defined as

$$\mathcal{D}_{\mathrm{KL}}(R_1, R_2) = \frac{1}{2}\left(\mathrm{trace}\,(R_1 R_2^{-1}) - \log\det(R_1 R_2^{-1}) - m\right) \quad (7)$$

which is the Kullback-Leibler divergence between two m-variate zero-mean Gaussian distributions with covariance matrices R_1 and R_2. These choices stem from the Whittle approximation according to which each $\widetilde{X}(\nu)$ has a zero-mean normal distribution with covariance matrix $R_X(\nu)$ and is uncorrelated with $\widetilde{X}(\nu')$ for $\nu \ne \nu'$. In this case, it is easily checked that $-\phi(\theta)$ evaluated with $\alpha_q = n_q$ and $\mathcal{D} = \mathcal{D}_{\mathrm{KL}}$ is (up to a constant) the log-likelihood for T data samples. This is actually true when the spectral domains are shrunk to just one DFT point ($n_q = 1$ for all q); when the spectral domains F_q are chosen to contain several (usually many) DFT points, then $-\phi(\theta)$ is the log-likelihood, in the Whittle approximation, of the Gaussian stationary model with constant power spectrum over each domain F_q. This approximation is at small statistical loss when the spectrum is smooth enough to show little variation over each spectral domain.

The major gain of assuming constant spectrum over each F_q is the resulting reduction of the data set to a small number of covariance matrices. This may be a crucial benefit in applications like astronomical imaging where very large data sets are frequent.

Regarding our application to CMB analysis, the hypothesized isotropy of the distribution of the sources leads to integrate over spectral domains with the corresponding symmetry. For sky maps small enough to be considered as flat, the spectral decomposition is the two-dimensional Fourier transform and the "natural" spectral domains are rings centered on the null frequency. For larger maps where curvature cannot be neglected, the spectral decomposition is over spherical harmonics and the natural spectral domains contain all the modes associated to a set of scales [20].

2.2. Parameter optimization

Minimizing the spectral mismatch $\phi(\theta)$ can be achieved using any optimization technique. However, ϕ being a likelihood criterion in disguise, one can also resort to the EM

algorithm. This is detailed in [16] in the case of spatially white noise, that is, $R_N(\nu)$ actually not depending on ν. Actually, this latter algorithm was slightly modified in order to deal with the case of colored noise N in (2). Another useful enhancement was to allow for constraints to be set on the model parameters so that prior information such as bounds on some entries of the mixing matrix A could be included. Details are given in the appendix.

The EM algorithm is straightforwardly implemented and does not require any tuning. It can quickly drive the spectral mismatch down to small values but is often unable to complete the optimization. Slow EM finishing is inherent to noisy models [21] and we have found it necessary to implement a mixed *ad hoc* strategy based on alternating EM steps and BFGS steps [16].

We have also found that initialization is critical: criterion (5) is probably multimodal for many data sets. This issue is not addressed in this paper though, since our prime interest is in the study of the statistical performances of different estimators of the model parameters θ. In the simulations reported below, the minimization of $\phi(\theta)$ is initialized at the true mixing matrix and with spectral covariance matrices estimated from the initial separate source and noise maps.

2.3. Component map estimation

When running SMICA, power spectral densities for the sources and detector noise are obtained along with the estimated mixing matrix. They are used in reconstructing the source maps *via* Wiener filtering in Fourier space: a Fourier mode $X(\nu)$ in frequency band $\nu \in F_q$ is used to reconstruct the maps according to

$$\widehat{S}(\nu) = \left(\widehat{A}^{\dagger}\widehat{R}_N(\nu)^{-1}\widehat{A} + \widehat{R}_S(\nu)^{-1}\right)^{-1}\widehat{A}^{\dagger}\widehat{R}_N(\nu)^{-1}X(\nu). \quad (8)$$

In the limiting case where noise is small compared to signal components, the Wiener filter reduces to

$$\widehat{S}(\nu) = \left(\widehat{A}^{\dagger}\widehat{R}_N(\nu)^{-1}\widehat{A}\right)^{-1}\widehat{A}^{\dagger}\widehat{R}_N(\nu)^{-1}X(\nu). \quad (9)$$

Note however that the above Wiener filter is optimal only in front of stationary Gaussian processes. For weak, point-like sources such as galaxy clusters seen *via* the Sunyaev–Zel'dovich effect (defined in Section 4.1), much better reconstruction can be expected from nonlinear methods.

3. SPECTRAL MATCHING IN WAVELET SPACE

The SMICA method for spectral matching in Fourier space has already shown significant success for CMB spectral estimation in multidetector experiments. It is in particular able to identify and remove residuals of poorly known correlated systematics and astrophysical foreground emissions contaminating CMB maps. However, SMICA suffers from several practical difficulties when dealing with real data.

Indeed, actual components are known to depart slightly from the ideal linear mixture model (2). The mixing matrix

(in particular those columns of A which correspond to galactic emissions) is known to depend somewhat on the direction of observation or on spatial frequency. Measuring the dependence $A(\vartheta, \varphi)$ is of interest for future experiments as Planck, and cannot be achieved directly with SMICA. Further, the components are known to be both correlated and nonstationary. For instance, galactic dust emissions are strongly peaked towards the galactic plane. A nonlocal spectral representation (*via* Fourier coefficients or *via* spherical harmonics) mixes contributions from high galactic sky, nearly free of foreground contamination, and contributions from within the galactic plane. Noise levels themselves may be quite nonstationary, with high SNR regions observed for a long time and low SNR regions poorly observed.

When there are sharp edges on the maps or gaps in the data, corresponding to unobserved or masked regions, spectral estimation using the smooth periodogram of (6) is not the most satisfactory procedure. Although apodizing windows may help cope with edge effects in Fourier analysis, they are not very straightforward to use in the case of arbitrarily shaped maps with arbitrarily shaped gaps, such as those encountered in the Archeops experiment [2].

Clearly, the spectral analysis of gapped data requires tools different from those used to process full data sets, if only because the hypothesized stationarity of the data is greatly disturbed by the missing samples. Common such methods often amount to using standard spectral estimators after the gaps were filled with estimates of the missing samples. However, the data interpolation stage is critical and cannot be completed without prior assumptions on the data. Another idea, applicable to CMB analysis, is to process gapped data as if they were complete but to correct afterwards the spectral estimates from the bias induced by the gaps [22]. We preferred to rely on methods intrinsically dedicated to the analysis of nonstationary data such as the wavelet transform, widely used to reveal variations in the spectral content of time series or images, as they permit to single out regions in direct space while retaining localization in the frequency domain. We see next how to reformulate (5) in the wavelet domain in order to deal with missing data. Note that, in the following, the locations of the missing samples are assumed to be known.

3.1. Wavelet transform

The experiments described further down make use of the undecimated *à trous* algorithm with the 2D cubic B3 spline [23] as scaling function, for implementing a wavelet transform. Although, depending on the data analysis problem, it is possible that a different choice can lead to better results, for our specific application, the *à trous* wavelet transform has several favorable properties. First, it is a shift invariant transform, the wavelet coefficient maps on each scale are the same size as the initial image, and the wavelet and scaling functions have small compact supports on the data map. Hence, missing patches in the observed maps are easily handled. Second, the 2D wavelet and scaling functions are nearly isotropic which is best for the analysis of an isotropic Gaussian field such as the CMB, or of data sets such as maps of galaxy clusters,

which contain only isotropic features. The undecimated isotropic *à trous* wavelet transform has been shown to be well suited to the analysis of astrophysical data where translation invariance is desirable and where the emphasis is seldom on data compression [23]. Further, with this choice of scaling function, the so-called scaling equation is satisfied, and therefore fast implementations of the decomposition and reconstruction steps of the *à trous* transform are available [23].

Given a 2D data set $c_0(k, l)$, the *à trous* algorithm produces recursively a set of detail maps $w_i(k, l)$ on a dyadic resolution scale and a smooth approximation $c_J(k, l)$ [23]. We note that the lowest resolution $J_{\max}$ is obviously limited by the data map size. The transform is readily inverted by

$$c_0(k, l) = c_J(k, l) + \sum_{i=1}^{J} w_i(k, l), \qquad (10)$$

which is a simple addition of the smooth array with the detail maps.

3.2. *Spectral matching in wavelet space: wSMICA*

In order to define a sensible wavelet version of SMICA, we first rewrite the SMICA criterion (5) in terms of covariance matrices in the initial domain, where, for instance, the gaps are best described, rather than in the Fourier domain.

Consider a batch of T data samples $X_{t=1,T}$ where t is an appropriate index depending on the dimension of the data, and the set of Q ideal bandpass filters $\mathcal{F}_q$ associated with the nonoverlapping frequency domains F_q used in SMICA. Denoting by $X_q(t)$ the data filtered through $\mathcal{F}_q$, we define sample covariance matrices

$$\hat{R}_{\mathbf{T},X}(q) = \frac{1}{T} \sum_{t=1}^{T} X_q(t) X_q(t)^{\dagger} \qquad (11)$$

obtained by averaging in the original domain. Owing to the unitary property of the discrete Fourier transform, one has

$$\hat{R}_{\mathbf{T},X}(q) = \frac{n_q}{T} \hat{R}_X(\nu_q), \qquad (12)$$

where n_q was defined as the number of Fourier modes in spectral band F_q. These matrices are estimates of $R_{\mathbf{T},X}(q) = \mathrm{E}(X_q(t) X_q(t)^{\dagger})$, the covariance matrix of $X_q(t)$. Again, according to model (3), the covariance matrices are again structured as

$$R_{\mathbf{T},X}(q) = A R_{\mathbf{T},S}(q) A^{\dagger} + R_{\mathbf{T},N}(q), \qquad (13)$$

where $R_{\mathbf{T},S}(q)$ and $R_{\mathbf{T},N}(q)$ are defined similarly to $R_{\mathbf{T},X}(q)$. Hence, minimizing the SMICA objective function (5) is then equivalent to minimizing

$$\phi(\theta) = \sum_{q=1}^{Q} n_q \mathcal{D}_{\mathrm{KL}}\left(\hat{R}_{\mathbf{T},X}(q), A R_{\mathbf{T},S}(q) A^{\dagger} + R_{\mathbf{T},N}(q)\right) \qquad (14)$$

with respect to the new set of parameters $\theta = (A, R_{\mathbf{T},S}(q), R_{\mathbf{T},N}(q))$.
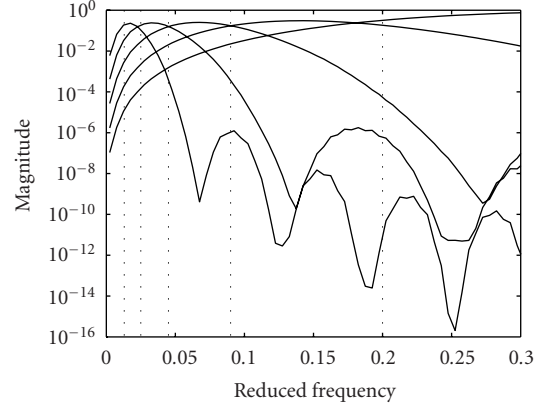

FIGURE 1: Magnitudes averaged over spectral rings of the nearly isotropic cubic spline wavelet filters $\psi_1, \psi_2, \dots, \psi_5$ used in the simulations described further down. The vertical dotted lines for $\nu = \{0.013, 0.025, 0.045, 0.09, 0.2\}$ delimit the five frequency bands used with SMICA in these simulations.

We now consider using another set of filters in place of the ideal bandpass filters used by SMICA. In dealing with nonstationary data or, as a special case, with gapped data, it is especially attractive to consider finite impulse response (FIR) filters. Indeed, provided the response of such a filter is short enough compared to data size T and gap widths, most of the samples in the filtered signal will be unaffected by the presence of gaps. Using exclusively these samples yields estimated covariance matrices which are not biased by the missing data, at the cost of a slight increase of variance due to discarding some data samples. In the following, we use filters $\psi_1, \psi_2, \dots, \psi_J, \phi_J$ (see Figure 1) and the wavelet *à trous* algorithm.

Consider again a batch of T regularly spaced data samples $X_{t=1,T}$. Possible gaps in the data are simply described with a mask μ, that is, an array of zeroes and ones of the same size as the data $X_{t=1,T}$ with ones corresponding to samples outside the gaps. Denoting by $W_1, W_2, \dots, W_J$ and C_J the wavelet scales and the smooth approximation of X, obtained with the *à trous* transform and $\mu_1, \dots, \mu_{J+1}$ the masks for the different scales determined from the original mask μ knowing the different filter lengths, wavelet covariances are estimated as follows:

$$\hat{R}_{\mathbf{W},X}(1 \le i \le J) = \frac{1}{l_i} \sum_{t=1}^{T} \mu_i(t) W_i(t) W_i(t)^{\dagger},$$

$$\hat{R}_{\mathbf{W},X}(J+1) = \frac{1}{l_{J+1}} \sum_{t=1}^{T} \mu_{J+1}(t) C_J(t) C_J(t)^{\dagger}, \qquad (15)$$

where l_i is the number of nonzero samples in μ_i. With source and noise covariances $R_{\mathbf{W},S}(i)$, $R_{\mathbf{W},N}(i)$ defined in a similar way, the covariance model in wavelet space becomes

$$R_{\mathbf{W},X}(i) = A R_{\mathbf{W},S}(i) A^{\dagger} + R_{\mathbf{W},N}(i). \qquad (16)$$
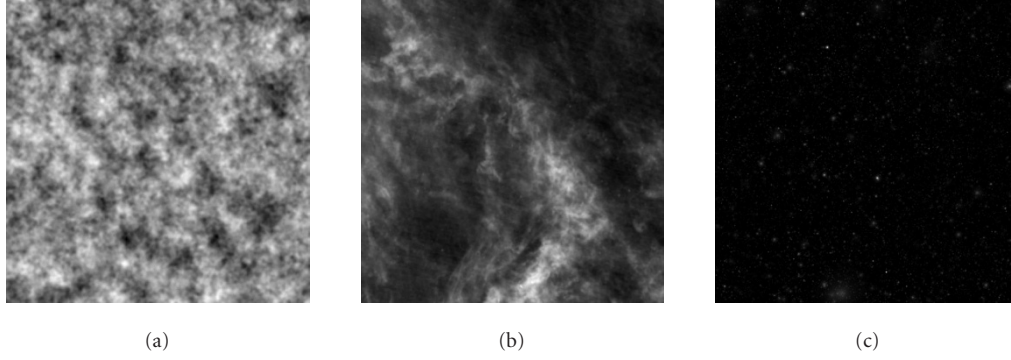
FIGURE 2: Samples of simulated component maps of CMB, dust, and SZ.

Our wavelet-based version of SMICA consists in minimizing the wSMICA criterion:

$$\phi(\theta) = \sum_{i=1}^{J+1} \alpha_i \mathcal{D}_{\mathrm{KL}}\left(\hat{R}_{\mathbf{W},X}(i), A R_{\mathbf{W},S}(i) A^\dagger + R_{\mathbf{W},N}(i)\right) \quad (17)$$

with respect to $\theta = (A, R_{\mathbf{W},S}(i), R_{\mathbf{W},N}(i))$ for some sensible choice of the weights α_i.

The weights in the spectral mismatch (17) should be chosen to reflect the variability of the estimate of the corresponding covariance matrix. Examining first (14), we see weights which are proportional to n_q, that is, to the number of DFT points used in computing the sample covariance matrix, because this is in fact the number of uncorrelated values of $\widetilde{X}(\nu)$ entering in the estimation of $\hat{R}_X(\nu_q)$. It is also proportional to the size of the frequency domain over which $\hat{R}_X(\nu_q)$ is evaluated. Since wSMICA uses wavelet filters with only limited overlap, we choose to use the same strategy as in SMICA since the latter amounts to using ideal bandpass filters. In other words, when no data points are missing, the weights for wSMICA are taken proportional to the size of the frequency domains covered at each scale. This is

$$\{\alpha_1, \alpha_2, \ldots, \alpha_J, \alpha_{J+1}\} = \left\{\frac{1}{2}, \frac{1}{4}, \ldots, \frac{1}{2^J}, \frac{1}{2^J}\right\} \quad (18)$$

in the one-dimensional case and

$$\{\alpha_1, \alpha_2, \ldots, \alpha_J, \alpha_{J+1}\} = \left\{\frac{3}{4}, \frac{3}{16}, \ldots, \frac{3}{4^J}, \frac{1}{4^J}\right\} \quad (19)$$

in the two-dimensional case.

In the case of data with gaps, we must further take into account that some wavelet coefficients are discarded. Let β_i denote the fraction of wavelets coefficients which are unaffected by the gaps at scale i. The number of effective points is reduced by this fraction and one should use the weights

$$\{\alpha_1, \alpha_2, \ldots, \alpha_J, \alpha_{J+1}\} = \left\{\frac{\beta_1}{2}, \frac{\beta_2}{4}, \ldots, \frac{\beta_J}{2^J}, \frac{\beta_{J+1}}{2^J}\right\} \quad (20)$$

in the one-dimensional case and

$$\{\alpha_1, \alpha_2, \ldots, \alpha_J, \alpha_{J+1}\} = \left\{\frac{3\beta_1}{4}, \frac{3\beta_2}{16}, \ldots, \frac{3\beta_J}{4^J}, \frac{\beta_{J+1}}{4^J}\right\} \quad (21)$$

in the two-dimensional case. The fraction $1 - \beta_i$ of discarded points depends on scale i (even with the *à trous* algorithm) because the length of the wavelet filter itself depends on i. However, it is roughly scale independent, if the missing data are large patches of much bigger size than the length of the wavelet filters used at any scale in the wavelet decomposition.

Before closing, we note that the different wavelet filter outputs $W_i(t)$ are correlated due to the overlap between frequency responses (Figure 1). Optimal inference should take this correlation into account but we have chosen not to do so and rather to stick to a simple criterion like (17) which ignores the correlations between sample covariance matrices. No big loss is expected from this choice because the wavelet bands do not overlap very much.

4.   NUMERICAL EXPERIMENTS

4.1.   Simulation of realistic maps

We have simulated observations consisting of $m = 6$ mixtures of $n = 3$ components, namely, CMB, galactic dust, and SZ emissions for which templates were obtained as described in [16]; see Figure 2 for typical realizations.

Dust emission is the greybody emission of small dust particles in our own galaxy. The intensity of emission is strongly concentrated towards the galactic plane, although cirrus clouds at high galactic latitudes are present as well. The dust emission law is of the form $\nu^\alpha B_\nu(T_{\mathrm{dust}})$ where $\alpha \simeq 1.7$, $B_\nu(T)$ is the blackbody emission law, and $T_{\mathrm{dust}} \simeq 17\,\mathrm{K}$ is the typical dust temperature in the interstellar medium.

The Sunyaev-Zel'dovich effect (SZ) is a small distortion of the CMB blackbody emission law caused by inverse Compton scattering of CMB photons on free electrons in hot ionized gas, present mostly in clusters of galaxies. The energetic electron, in the interaction, gives a fraction of its energy to the scattered CMB photon, shifting its frequency to a higher value. As a result, the SZ effect causes a shift in CMB photon
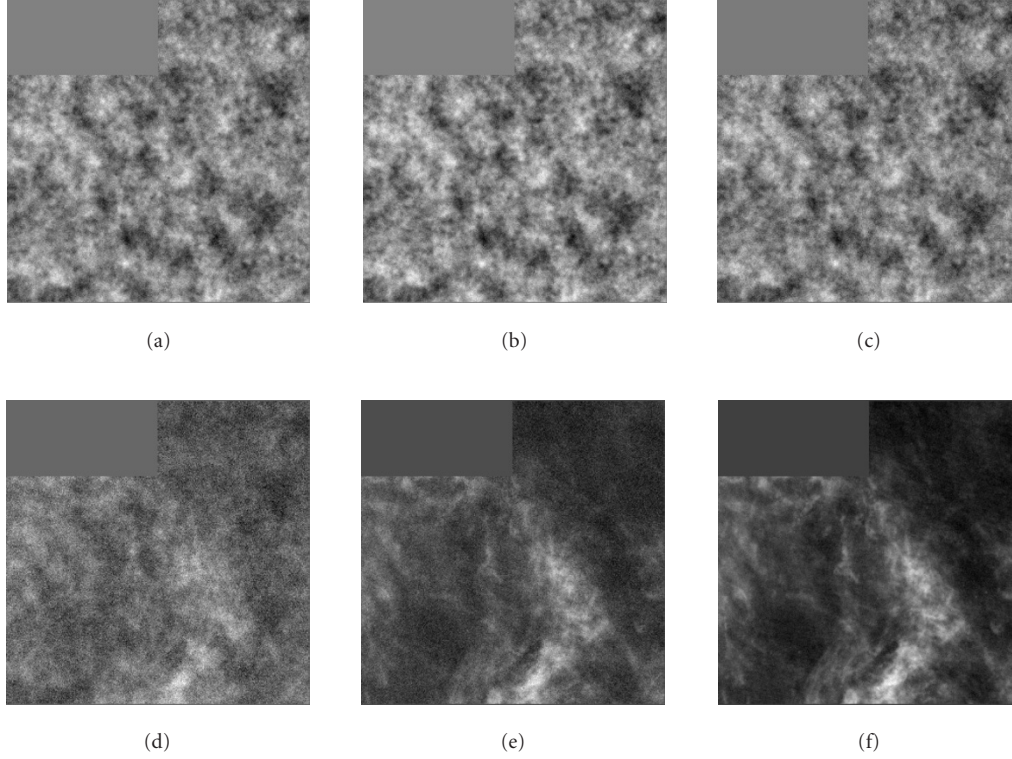
(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 3: Simulated observation maps based on the templates shown in **Figure 2** and the mixing matrix in **Table 1** for the nominal Planck HFI noise levels.

TABLE 1: Entries of A, the mixing matrix used in our simulations.

CMB	Dust	SZ	Channel
7.452×10^{-1}	3.654×10^{-2}	-8.733×10^{-1}	100 GHz
5.799×10^{-1}	7.021×10^{-2}	-4.689×10^{-1}	143 GHz
3.206×10^{-1}	1.449×10^{-1}	-2.093×10^{-3}	217 GHz
7.435×10^{-2}	3.106×10^{-1}	1.294×10^{-1}	353 GHz
6.009×10^{-3}	5.398×10^{-1}	2.613×10^{-2}	545 GHz
6.115×10^{-5}	7.648×10^{-1}	5.268×10^{-4}	857 GHz

energy distribution, depleting the occupation of low energy levels and populating high energy levels. The net effect, to first order, is a small additive emission, negative at frequencies below 217 GHz, and positive at frequencies above. A review on SZ effect can be found in [24].

The templates, and thus the mixtures in each simulated data set, consist of 300×300 pixel maps corresponding to a $12.5° \times 12.5°$ field located at high galactic latitude. The six mixtures in each set mimic observations that will eventually be acquired in the six frequency channels of the Planck HFI (**Figure 3**). The entries of the mixing matrix A used in these simulations actually are estimated values of the electromagnetic emission laws of each component at 100, 143, 217, 353, 545, and 857 GHz; see **Table 1**.

White Gaussian noise is added to the mixtures according to model (2) in order to simulate instrumental noise.

While the relative noise standard deviations between channels are set according to the nominal values of the Planck HFI, we also experiment with five *global* noise levels at -20, -6, -3, 0, and $+3$ dB from nominal values. **Table 2** gives the typical energy fractions that are contributed by each of the $n = 3$ original sources and noise, to the total energy of each of the $m = 6$ mixtures, considering Planck nominal noise variance. In fact, because SMICA and wSMICA actually work on spectral bands, a much better indication of signal-to-noise ratio in these simulations is given by **Figure 4** which shows how noise and source energy contributions distribute with respect to frequency in the six mixtures.

Finally, in order to investigate the impact of gaps in the data, and the benefits of using wSMICA in place of SMICA to deal with these gaps, the mask shown in **Figure 5** was applied onto the mixture maps. The case where no data is missing

TABLE 2: Energy fraction contributed by each source to the total energy of each mixture, for the nominal noise variance on the Planck HFI channels.

CMB	Dust	SZ	Noise	Channel
9.91×10^{-1}	1.18×10^{-4}	7.92×10^{-3}	2.53×10^{-6}	100 GHz
9.97×10^{-1}	7.25×10^{-4}	3.79×10^{-3}	5.17×10^{-7}	143 GHz
9.98×10^{-1}	1.01×10^{-2}	2.48×10^{-7}	1.34×10^{-7}	217 GHz
5.55×10^{-1}	4.8×10^{-1}	9.78×10^{-3}	7.47×10^{-8}	353 GHz
2.5×10^{-3}	1.0	2.75×10^{-4}	3.78×10^{-9}	545 GHz
1.29×10^{-7}	1.0	5.56×10^{-8}	1.24×10^{-10}	857 GHz

was also considered as a reference case. Spectral matching with wSMICA is conducted using the output of the five wavelet filters $\psi_1, \ldots, \psi_5$ associated to higher frequency details. For the sake of comparison, SMICA is run using five bands in Fourier space which are similar to the dyadic bands imposed by the wavelet transform, as shown in Figure 1. This latter choice of frequency bands is made to ease comparison between SMICA and wSMICA.

4.2. Experiments with noise-free mixtures

Preliminary experiments were conducted in the case of vanishing instrumental noise variance. In this case, the blind component separation problem is "equivariant," entailing that the quality of separation on a given mixture does not depend at all on the mixing matrix A but only on the particular realization of the sources and on the algorithm used for separation. More specifically, in the case of SMICA and wSMICA, separation performance depends on the spectral diversity of the components and on the ability of both objective functions to exploit this diversity. Hence, the noise-free experiments in this section are indicative of the spectral diversity of the components, of the ability of (w)SMICA to capture it, and of the robustness of the (w)SMICA with respect to missing data.

Note that in a noise-free model, the spectral matching objective boils down to an objective of joint diagonalization of the covariance matrices, as shown in [25]. Hence, spectral matching can be implemented using an efficient dedicated algorithm [26].

The estimated components are related to the true one according to

$$\hat{S} = \mathscr{I} S, \tag{22}$$

where $\mathscr{I}$ is the product of the mixing matrix used in simulations and of the separating matrix obtained by joint diagonalization. It also includes any normalization needed for the components and their estimates to have total energy in all bands equal to 1. With this normalization, the square of any offdiagonal term $\mathscr{I}_{ij}$ is directly related to the residual level of contamination by component j in the recovered component i. Since performance in separating noise-free mixtures is independent of the mixing matrix, the choice of A in the simulations is irrelevant: it does not change the distribution of $\mathscr{I}$. In practice, our noise-free experiments are conducted

without any mixing, that is, we take A to be the 3×3 identity matrix. The following steps were repeated 1000 times.

(i) Randomly pick one of each component maps out of the available 200 CMB maps, 30 dust maps, and 1500 SZ maps.

(ii) Compute covariance matrices in the five wavelet or Fourier bands, both with and without masking part of the maps.

(iii) Normalize each source so that its total energy over the five bands is equal to one.

(iv) Estimate a separating matrix by joint diagonalization of the covariance matrices.

These noise-free experiments are complemented using "surrogate" data in order to assess the effect of any non-Gaussianity or nonstationarity in the source templates. We repeat the simulations on Gaussian stationary maps generated with the same spectra as the CMB, dust, and SZ components. The resulting distribution of $\mathscr{I}$ then only reflects the ability of (w)SMICA to exploit the spectral diversity of the components independently of the other aspects of their distribution.

The histograms on Figure 6 are for the offdiagonal term corresponding to the residual corruption of CMB by *Gaussian* dust in the second set of experiments (using surrogate data). In Tables 3 and 4, the results obtained with the synthetic component maps are given as well as those obtained with the surrogate Gaussian maps, in terms of the standard deviations of the offdiagonal entries $\mathscr{I}_{ij}$ defined by (22).

When working on surrogate Gaussian maps without masks, using covariance matrices in Fourier space or in wavelet space gives similar performances. It is also satisfactory, when covariances in wavelet space are used with surrogate Gaussian maps, that each computed standard deviation only slightly increases when a mask is applied on the data. Indeed, as a consequence of incomplete coverage, there are less samples from which to estimate the covariances. This increase is also observed when covariance matrices in Fourier space are used with the surrogate Gaussian maps but it can be as high as fivefold and it does not affect all the coefficients equally. Although this can again be attributed to the reduced data size, the lowered spectral diversity between components, because of the correlations and smoothing induced in Fourier space by the mask, is also part of the explanation. In fact, as shown on Figure 4, CMB and dust spatial power
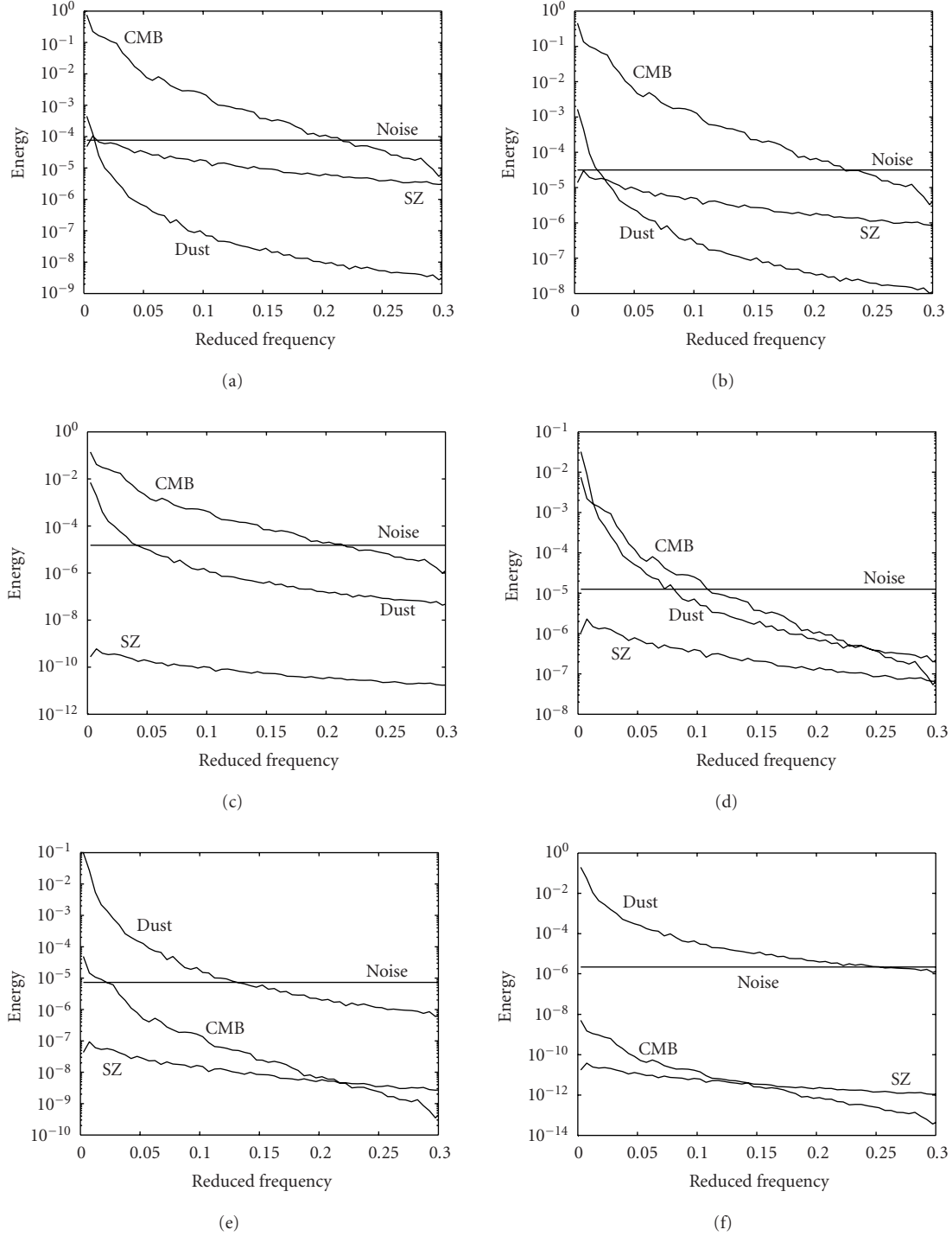
FIGURE 4: Energy contributed by each source and noise in each bolometer as a function of frequency, for the nominal noise variance on the Planck HFI channels: (a) 100 GHz, (b) 143 GHz, (c) 217 GHz, (d) 353 GHz, (e) 545 GHz, and (f) 857 GHz. Note how SZ is expected to be always below nominal noise, that CMB and dust strongly dominate in different channels, and that CMB and dust spectra, without being proportional, display the same general behavior dominated by low modes.

spectra are somewhat similar, that is, show low spectral diversity, and further smoothing can only degrade the performance of the source separation algorithm based on Fourier covariances.

In the case of realistic component maps, we note first that the comparison of the performance of component separation using wavelet-based covariance matrices with and without mask again agrees with the different data sizes, which is not

(a)   (b)

(c)   (d)

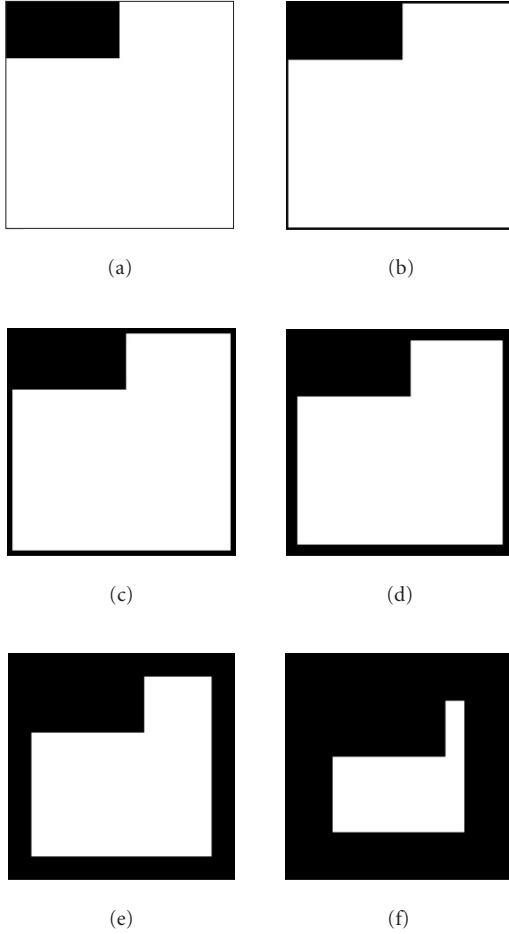(e)   (f)

FIGURE 5: (a) Mask used to simulate a gap in the data. (b)–(f) Modified masks at scales 1 through 5. The discarded pixels are in black.

the case with covariances in Fourier space. Next, whether covariance matrices are computed in Fourier space or in wavelet space, we note that the terms coupling CMB and dust are again much higher than with surrogate data, even on complete maps. This is probably to be attributed to the nonstationarity and/or non-Gaussianity of the Dust component. Another point is that the CMB and dust templates as in Figure 2 exhibit sharp edges compared to SZ and this inevitably disturbs spectral estimation using a simple DFT. To assess this effect, simulations were also conducted where the covariances in Fourier space were computed after an apodizing Hanning window was applied on the complete data maps. The results reported in Table 3, to be compared to Table 4, do indicate a slightly positive effect of windowing, but still the separation using wavelet-based statistics appears better. To further complete this preliminary study, we conducted similar experiments using JADE [27], an ICA algorithm based on fourth-order statistics. This algorithm does not use spectral information at all. As discussed earlier, such a method is not expected to work well on CMB data and the results reported in Table 5 do show lower performance in comparison to Tables 3 and 4.

4.3.   Realistic experiments

The results of the previous section show that, in the noiseless case, using wavelet-based covariance matrices provides a simple and efficient way to cancel the bad impact that gaps actually have on the performance of estimation using Fourier-based statistics. We move on to investigate the effect of additive noise on SMICA and wSMICA.

Picking at random one of each component maps out of the available 200 CMB maps, 30 dust maps, and 1500 SZ maps, 1000 sets of six synthetic mixture maps were generated as previously described, for each of the 5 noise levels chosen. Then, component separation was conducted using the spectral matching algorithms SMICA and wSMICA both with and without part of the maps being masked. A typical run of SMICA or wSMICA in the setting considered here (i.e., 300 by 300 maps, 6 mixtures, 3 sources, 5 wavelet scales, no constraints on the mixing matrix) takes only a few seconds on a 1.25 Ghz Mac G4 when coded in IDL. The same optimization techniques are used for SMICA and wSMICA since the two criteria have the same form.

Each run of SMICA and wSMICA on the data returns estimates $\hat{A}_f$ and $\hat{A}_w$ of the mixing matrix. These estimates are subject to the indeterminacies inherent to the instantaneous linear mixture model (2). Indeed, in the case where optimization is over all parameters θ, any simultaneous permutation of the columns of A and of the lines of S leaves the model unchanged. The same occurs when exchanging a scalar possibly negative factor between any column in A and the corresponding line in S. Therefore, columnwise comparison of $\hat{A}_f$ and $\hat{A}_w$ to the original mixing matrix A requires first fixing these indeterminacies. This is done "by hand" after $\hat{A}_f$ and $\hat{A}_w$ have been normalized columnwise.

The results we report next focus on the statistical properties of $\hat{A}_f$ and $\hat{A}_w$ as estimated from the 1000 runs of the two competing methods in the several configurations retained. In fact, the correct estimation of the mixing matrix in model (2) is a relevant issue, for instance, when it comes to dealing with the cross-calibration of the different detectors. Figures 7, 8, and 9 show the results obtained, using the quadratic norm

$$QE_j = \left(\sum_{i=1}^{m} \left(A_{ij} - \hat{A}_{ij} \right)^2 \right)^{1/2} \tag{23}$$

with $\hat{A} = \hat{A}_f$ or $\hat{A}_w$ and $j =$ CMB, dust, or SZ, to assess the residual errors on the estimated emissivities of each component. The plotted curves show how the mean of the above positive error measure varies with increasing noise variance. For the particular case of CMB, Table 6 gives the estimated standard deviations of the relative errors $(A_{ij} - \hat{A}_{ij})/A_{ij}$ on the estimated CMB emission law in the six channels of Planck's HFI in the different configurations retained.

Closer to our source separation objective, a more significant way of assessing the quality of $\hat{A}_f$ and $\hat{A}_w$ as estimators of the mixing matrix A would be to use the following
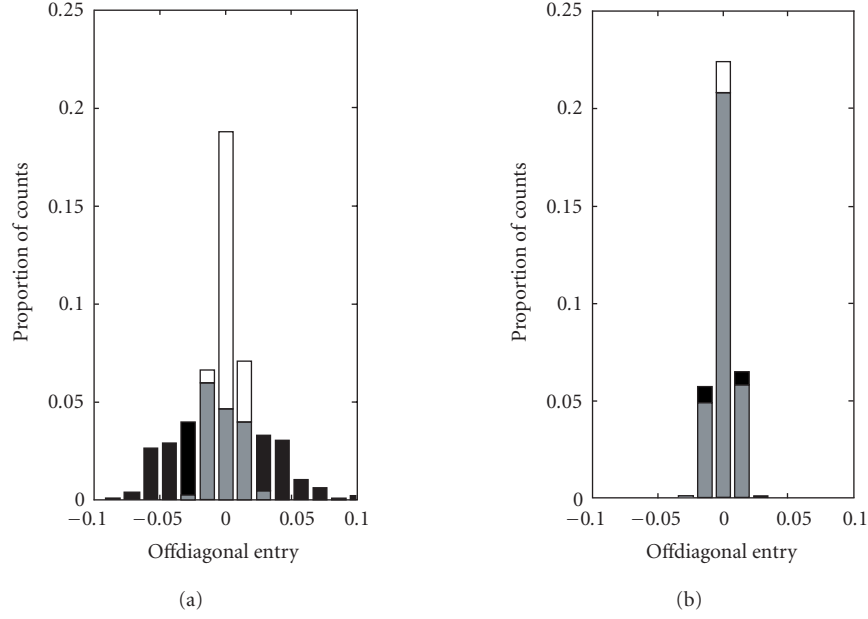
(a)

(b)

FIGURE 6: Histograms of the offdiagonal term of ℓ, defined in (22), corresponding to the residual corruption of "CMB" by "dust" while separating Gaussian maps generated with the same power spectra as the astrophysical components, by joint diagonalization of covariance matrices in (a) Fourier and (b) wavelet spaces, with *(black, which appears grey when seen through white)* and without *(white)* masking part of the data. The dark widest histogram on the left highlights the impact of masking on source separation based on Fourier covariance matrices.

TABLE 3: Standard deviations of the offdiagonal entries ℓ_{ij} defined by (22) obtained while separating realistic component maps by joint diagonalization of covariance matrices in *Fourier* space, with (M) or without masking (NM) part of the data, or applying an apodizing Hanning window (Han). Components 1, 2, and 3, respectively, stand for CMB, dust, and SZ. The numbers in *italic* were obtained with Gaussian maps and the underlined numbers correspond to the histograms in Figure 6.

Offdiag. entry	NM		M		Han
$\ell_{1,2}$	0.097	*0.0076*	0.074	*0.038*	0.024
$\ell_{1,3}$	0.0049	*0.0044*	0.005	*0.006*	0.0094
$\ell_{2,1}$	0.017	*0.0066*	0.018	*0.01*	0.017
$\ell_{2,3}$	0.0064	*0.0077*	0.0066	*0.0096*	0.011
$\ell_{3,1}$	0.0024	*0.0026*	0.0028	*0.0037*	0.0039
$\ell_{3,2}$	0.0054	*0.0071*	0.0054	*0.0079*	0.01

TABLE 4: Standard deviations of the offdiagonal entries ℓ_{ij} defined by (22) obtained while separating realistic component maps by joint diagonalization of covariance matrices in *wavelet* space, with (M) and without masking (NM) part of the data. Components 1, 2, and 3, respectively, stand for CMB, dust, and SZ. The numbers in *italic* were obtained with Gaussian maps and the underlined numbers correspond to the histograms in Figure 6.

Offdiag. entry	NM		M	
$\ell_{1,2}$	0.015	*0.0071*	0.018	*0.0079*
$\ell_{1,3}$	0.0025	*0.0029*	0.0028	*0.0031*
$\ell_{2,1}$	0.016	*0.0077*	0.019	*0.0089*
$\ell_{2,3}$	0.0041	*0.0051*	0.0048	*0.0075*
$\ell_{3,1}$	0.0024	*0.0029*	0.003	*0.0039*
$\ell_{3,2}$	0.0039	*0.0054*	0.0053	*0.0085*

TABLE 5: Standard deviations of the offdiagonal entries $\mathcal{l}_{ij}$ defined by (22) obtained while separating realistic component maps using JADE, with (M) and without masking (NM) part of the data. Components 1, 2, and 3, respectively, stand for CMB, dust, and SZ.

Offdiag. entry	$\mathcal{l}_{1,2}$	$\mathcal{l}_{1,3}$	$\mathcal{l}_{2,1}$	$\mathcal{l}_{2,3}$	$\mathcal{l}_{3,1}$	$\mathcal{l}_{3,2}$
NM	0.021	0.25	0.022	0.02	0.31	0.02
M	0.023	0.29	0.025	0.018	0.34	0.018

interference-to-signal ratio:

$$\mathrm{ISR}_j = \frac{\sum_{i \neq j} \mathcal{l}_{j,i}^2 \sigma_i^2}{\mathcal{l}_{j,j}^2 \sigma_j^2}, \tag{24}$$

where the σ_j are the source variances and

$$\mathcal{l} = \left(\hat{A}^\dagger \hat{R}_N^{-1} \hat{A} \right)^{-1} \hat{A}^\dagger \hat{R}_N^{-1} A \tag{25}$$

with R_N the noise covariance. The plots on Figures 10, 11, and 12 show how the mean ISR from the 1000 runs of SMICA and wSMICA in different configurations varies with increasing noise. Figure 13 typically estimated component maps obtained using SMICA and wSMICA. For the sake of comparison, component maps estimated using the JADE source separation method are also included.

We note again that the performance of wSMICA behaves as expected when noise increases and if part of the data is missing. However, this is not always the case with SMICA. Finally, this set of simulations, conducted in a more realistic setting with respect to ESA's Planck mission, again confirms the higher performance, over Fourier analysis, that we indeed expected from the use of wavelets. The latter are able to correctly grab the spectral content of partly masked data maps and from there allow for better component separation.

5. CONCLUSION

This paper has presented an extension of the spectral matching ICA algorithm to the wavelet domain, motivated by the need to deal with components which exhibit spatial correlations and are nonstationary. Maps with gaps are a particular instance of practical significance. Substituting covariance matching in Fourier space by covariance matching in wavelet space makes it possible to cope with gaps of any shape in a very straightforward manner. Mainly, it is the finite length of the wavelet filters used here that allows the impact of edges and gaps on the estimated covariances and hence on component separation to be lowered. Optimally choosing the FIR filter bank regarding a particular application is a possible further enhancement.

Our numerical experiments, based on realistic simulations of the astrophysical data expected from the Planck mission, confirm the benefits of correctly processing existing gaps. Clearly, other possible types of nonstationarities in the collected data such as spatially varying noise or component variance, and so forth could be dealt with very simply in a similar fashion using the wavelet extension of SMICA.

Regarding future work, a few points are in order. First, we note that possible correlations between the components are not accounted for in SMICA or wSMICA as presented here. However, it is not difficult in principle to handle such known or suspected correlations by adding offdiagonal parameters in the model spectral covariances of the sources. Still, in the case of CMB analysis from high frequency observations which contain only one galactic component (dust) as in our simulations, spatial correlations between components should not be a problem.

We note that the proposed wavelet-based approach, as implemented with the standard *à trous* wavelet transform, offers little flexibility in the spectral bands available for wSMICA while the Fourier approach gives complete flexibility in this respect. But it is possible, even straightforward, to use other transforms such as the *à trous* wavelet packet transform, or the *continuous* wavelet transform, or in fact any set of linear filters, preferably FIR filters. This in turn raises the question of optimally choosing this set of filters, keeping in mind that higher resolution in Fourier space requires longer filters, which is not desirable in the case of incomplete or nonstationary data. In fact, the optimal selection of bands is clearly a meaningful question both for SMICA and wSMICA.

We also note that in the CMB application, the components have quite different statistical properties: some are expected to be very close to Gaussian (like the CMB) whereas others are strongly non-Gaussian (like SZ). The non-Gaussianity of some components does not affect the consistency of our estimator but, for a given spectrum, it does affect the distribution of the estimates although this impact is not easily predicted. It is clear, however, that ignoring the strong non-Gaussianity of some components is a loss of information. Devising a method able, with reasonable complexity, to exploit jointly non-Gaussianity (as in traditional ICA techniques) and spectral information (as in Fourier or wavelet SMICA) appears as a difficult challenge.

APPENDIX

EM ALGORITHM WITH CONSTRAINTS ON THE MIXING MATRIX

Considering Q separate frequency bands of size n_q with $\sum n_q = 1$, the EM functional derived for the instantaneous mixing model (2) with independent Gaussian stationary sources S and noise N is

$$\Phi(\underline{\theta}, \theta) = \mathcal{E} \left\{ \log p(X, S | \underline{\theta}) | \theta \right\} \tag{A.1}$$

with $\theta = (A, R_{S,1}, \ldots, R_{S,Q}, R_{N,1}, \ldots, R_{N,Q})$ and $\underline{\theta} = (\underline{A}, \underline{R}_{S,1}, \ldots, \underline{R}_{S,Q}, \underline{R}_{N,1}, \ldots, \underline{R}_{N,Q})$. The maximization step of the EM algorithm seeks then to maximize $\Phi(\underline{\theta}, \theta)$ with respect to $\underline{\theta}$ and the optimal $\underline{\theta}$ is used as the value for θ at the next EM step, and so on until satisfactory convergence is reached. Explicit expressions are easily derived for the optimal $\underline{\theta}$ in the white noise case where an interesting decoupling occurs between the reestimating equations for noise variances, source variances, and the mixing matrix [15].
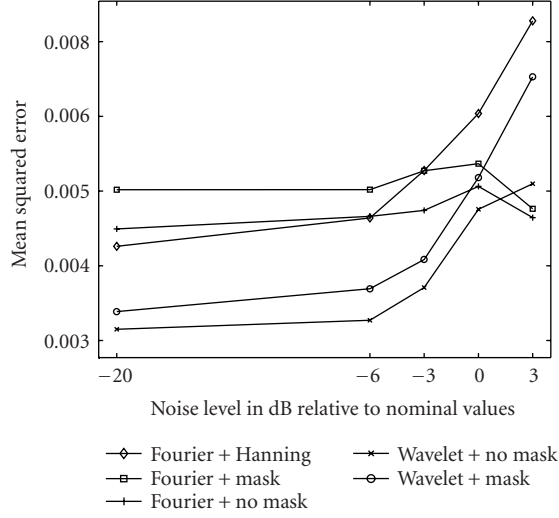
FIGURE 7: Comparison of the mean squared errors on the estimation of the emission law of CMB as a function of noise in five different configurations: wSMICA without mask, wSMICA with mask, fSMICA without mask, and fSMICA with mask, and fSMICA with Hanning apodizing window.
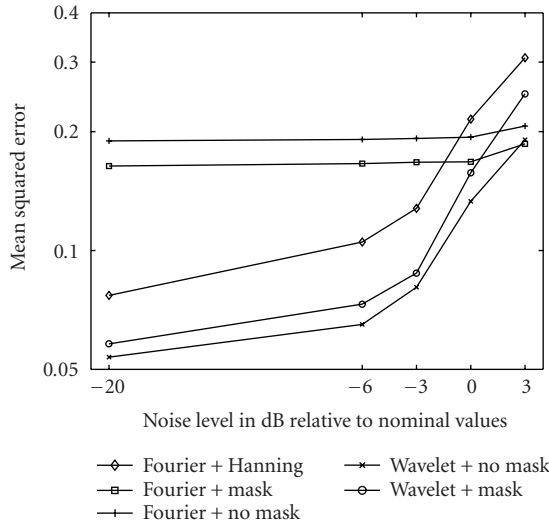

FIGURE 8: Comparison of the mean squared errors on the estimation of the emission law of *dust* as a function of noise in five different configurations: wSMICA without mask, wSMICA with mask, fSMICA without mask, and fSMICA with mask, and fSMICA with Hanning apodizing window.

Linear equality constraints

When A is subject to linear constraints, the joint maximization of the EM functional with respect to all model parameters is no longer easily achieved in general. In fact, one cannot simply decouple the reestimating rules for the noise parameters and the mixing matrix and these have to be optimized separately. We give next the modified reestimating equations for the mixing matrix and the source variances in the case of constant noise (i.e., $\theta = (A, R_{S,1}, \dots, R_{S,Q})$).
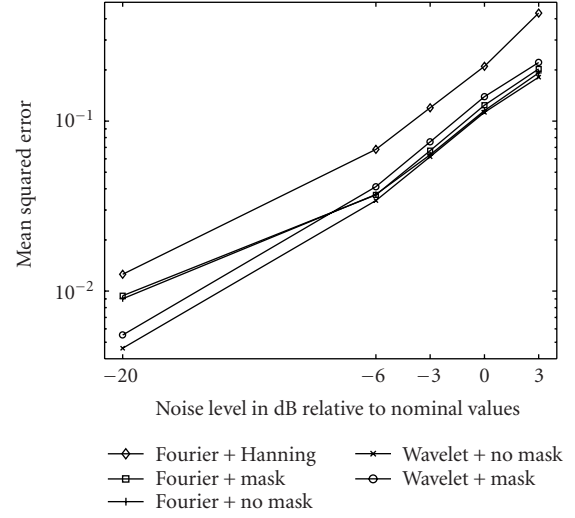

FIGURE 9: Comparison of the mean squared errors on the estimation of the emission law of SZ as a function of noise in five different configurations: wSMICA without mask, wSMICA with mask, fSMICA without mask, fSMICA with mask, and fSMICA with Hanning apodizing window.

First, we exhibit the quadratic dependence of the EM functional $\Phi(\underline{\theta}, \theta)$ on $\underline{A}$:

$$
\begin{aligned}
\Phi(\underline{\theta}, \theta) \\
= -\frac{1}{2} \sum_q n_q \, \text{trace} \left(\underline{A} R_q^{ss} \underline{A}^\dagger R_{N,q}^{-1} - \underline{A} R_q^{xs\dagger} R_{N,q}^{-1} - R_q^{xs} \underline{A}^\dagger R_{N,q}^{-1} \right) \\
+ \text{const}_{\underline{A}},
\end{aligned}
\tag{A.2}
$$

where

$$
\begin{aligned}
C_q &= \left(A^\dagger R_{N,q}^{-1} A + R_{S,q}^{-1} \right)^{-1}, \\
W_q &= \left(A^\dagger R_{N,q}^{-1} A + R_{S,q}^{-1} \right)^{-1} A^\dagger R_{N,q}^{-1}, \\
R_q^{xs} &= \widehat{R}_{X,q} W_q^\dagger, \\
R_q^{ss} &= W_q \widehat{R}_{X,q} W_q^\dagger + C_q.
\end{aligned}
\tag{A.3}
$$

In the white noise case, $R_{N,q} = R_N$, (A.2) becomes

$$
\Phi(\underline{\theta}, \theta) = -\frac{1}{2} \, \text{trace} \left(\left(\underline{A} - R^{xs} R^{ss-1} \right) R^{ss} \left(\underline{A} - R^{xs} R^{ss-1} \right)^\dagger R_N^{-1} \right) \\
+ \text{const}_{\underline{A}},
\tag{A.4}
$$

where

$$
R^{xs} = \sum_q n_q R_q^{xs}, \qquad R^{ss} = \sum_q n_q R_q^{ss}.
\tag{A.5}
$$

Again, this can be rewritten as

$$
\Phi(\underline{\theta}, \theta) = -\frac{1}{2} (\underline{\mathcal{A}} - \mathcal{M})^\dagger \mathcal{Q} (\underline{\mathcal{A}} - \mathcal{M}) + \text{const}_{\underline{A}},
\tag{A.6}
$$

TABLE 6: *Standard deviations* of the relative errors on the estimated emission laws $\mathcal{A}_{il}$ of CMB in Planck's HFI six channels. The column labels WNM, WM, FNM, FM, FHan are for the different configurations, respectively: wSMICA without mask, wSMICA with mask, fSMICA without mask, fSMICA with mask, and fSMICA with Hanning apodizing window. The five figures in each box are for noise variance -20, -6, -3, 0, and 3 dB from nominal Planck values.

Emission law	WNM	WM	FNM	FM	FHan
	$4.4 * 10^{-4}$	$5.0 * 10^{-4}$	$6.2 * 10^{-4}$	$7.3 * 10^{-4}$	$7.2 * 10^{-4}$
	$5.4 * 10^{-4}$	$7.5 * 10^{-4}$	$7.1 * 10^{-4}$	$8.5 * 10^{-4}$	$9.5 * 10^{-4}$
$\mathcal{A}_{11}$	$6.6 * 10^{-4}$	$9.2 * 10^{-4}$	$8.2 * 10^{-4}$	$8.9 * 10^{-4}$	$1.3 * 10^{-3}$
	$9.4 * 10^{-4}$	$1.2 * 10^{-3}$	$1.0 * 10^{-3}$	$1.0 * 10^{-3}$	$1.7 * 10^{-3}$
	$1.2 * 10^{-3}$	$1.7 * 10^{-3}$	$1.2 * 10^{-3}$	$1.4 * 10^{-3}$	$2.3 * 10^{-3}$
	$1.6 * 10^{-4}$	$2.1 * 10^{-4}$	$2.1 * 10^{-4}$	$2.0 * 10^{-4}$	$2.7 * 10^{-4}$
	$5.3 * 10^{-4}$	$7.8 * 10^{-4}$	$5.6 * 10^{-4}$	$5.7 * 10^{-4}$	$1.0 * 10^{-3}$
$\mathcal{A}_{21}$	$7.0 * 10^{-4}$	$1.1 * 10^{-3}$	$7.6 * 10^{-4}$	$8.4 * 10^{-4}$	$1.4 * 10^{-3}$
	$1.0 * 10^{-3}$	$1.6 * 10^{-3}$	$1.0 * 10^{-3}$	$1.0 * 10^{-3}$	$2.1 * 10^{-3}$
	$1.4 * 10^{-3}$	$2.2 * 10^{-3}$	$1.5 * 10^{-3}$	$1.7 * 10^{-3}$	$3.1 * 10^{-3}$
	$1.5 * 10^{-3}$	$1.8 * 10^{-3}$	$2.2 * 10^{-3}$	$2.5 * 10^{-3}$	$2.3 * 10^{-3}$
	$1.7 * 10^{-3}$	$2.1 * 10^{-3}$	$2.3 * 10^{-3}$	$2.6 * 10^{-3}$	$2.9 * 10^{-3}$
$\mathcal{A}_{31}$	$2.1 * 10^{-3}$	$2.6 * 10^{-3}$	$2.6 * 10^{-3}$	$2.8 * 10^{-3}$	$3.7 * 10^{-3}$
	$2.7 * 10^{-3}$	$3.0 * 10^{-3}$	$2.9 * 10^{-3}$	$3.0 * 10^{-3}$	$4.2 * 10^{-3}$
	$3.3 * 10^{-3}$	$4.6 * 10^{-3}$	$3.3 * 10^{-3}$	$3.5 * 10^{-3}$	$6.1 * 10^{-3}$
	$1.8 * 10^{-2}$	$2.0 * 10^{-2}$	$2.7 * 10^{-2}$	$3.0 * 10^{-2}$	$2.5 * 10^{-2}$
	$1.9 * 10^{-2}$	$2.1 * 10^{-2}$	$2.7 * 10^{-2}$	$2.1 * 10^{-2}$	$2.7 * 10^{-2}$
$\mathcal{A}_{41}$	$2.1 * 10^{-2}$	$2.4 * 10^{-2}$	$2.8 * 10^{-2}$	$3.1 * 10^{-2}$	$2.9 * 10^{-2}$
	$2.7 * 10^{-2}$	$2.8 * 10^{-2}$	$3.1 * 10^{-2}$	$3.0 * 10^{-2}$	$3.5 * 10^{-2}$
	$3.0 * 10^{-2}$	$4.1 * 10^{-2}$	$2.5 * 10^{-2}$	$2.7 * 10^{-2}$	$4.9 * 10^{-2}$
	$4.0 * 10^{-1}$	$4.5 * 10^{-1}$	$6.1 * 10^{-1}$	$6.6 * 10^{-1}$	$5.6 * 10^{-1}$
	$4.2 * 10^{-1}$	$4.7 * 10^{-1}$	$6.1 * 10^{-1}$	$6.5 * 10^{-1}$	$5.8 * 10^{-1}$
$\mathcal{A}_{51}$	$4.5 * 10^{-1}$	$5.0 * 10^{-1}$	$6.1 * 10^{-1}$	$6.7 * 10^{-1}$	$6.4 * 10^{-1}$
	$5.7 * 10^{-1}$	$5.9 * 10^{-1}$	$6.7 * 10^{-1}$	$6.7 * 10^{-1}$	$7.5 * 10^{-1}$
	$6.2 * 10^{-1}$	$8.4 * 10^{-1}$	$5.0 * 10^{-1}$	$5.5 * 10^{-1}$	1.0
	$5.7 * 10^{1}$	$6.2 * 10^{1}$	$8.5 * 10^{1}$	$9.2 * 10^{1}$	$7.8 * 10^{1}$
	$5.8 * 10^{1}$	$6.5 * 10^{1}$	$8.6 * 10^{1}$	$9.1 * 10^{1}$	$8.1 * 10^{1}$
$\mathcal{A}_{61}$	$6.2 * 10^{1}$	$6.9 * 10^{1}$	$8.6 * 10^{1}$	$9.4 * 10^{1}$	$8.9 * 10^{1}$
	$7.9 * 10^{1}$	$8.2 * 10^{1}$	$9.3 * 10^{1}$	$9.2 * 10^{1}$	$1.0 * 10^{2}$
	$8.6 * 10^{1}$	$1.2 * 10^{2}$	$6.9 * 10^{1}$	$7.7 * 10^{1}$	$1.4 * 10^{2}$

where

$$\underline{\mathcal{A}} = \text{vect}\,\underline{A}, \qquad \mathcal{Q} = \underline{R_N}^{-1} \otimes \sum_q n_q R_q^{ss},$$

$$\mathcal{M} = \text{vect}\left(\left(\sum_q n_q R_q^{xs} \right) \left(\sum_q n_q R_q^{ss} \right)^{-1} \right). \tag{A.7}$$

Here "vect" builds a column vector with the entries of a matrix taken along its rows. Now we consider linear constraints on the mixing matrix, specified as follows:

$$\mathcal{C}^{\dagger}\left(\underline{\mathcal{A}} - \mathcal{A}_0\right) = 0, \tag{A.8}$$

where $\mathcal{C}$ is a matrix with as many columns as constraints, and the columns of $\mathcal{C}$ have the same size as $\mathcal{A}$. The maximum of the EM functional with respect to $\underline{\theta}$ subject to the specified linear constraints is then reached for

$$\underline{\mathcal{A}} = \mathcal{M} - \mathcal{Q}^{-1}\mathcal{C}\left(\mathcal{C}^{\dagger}\mathcal{Q}^{-1}\mathcal{C}\right)^{-1}\mathcal{C}^{\dagger}(\mathcal{M} - \mathcal{A}_0),$$

$$\underline{R}_{S,q} = \text{diag}\left(R_q^{ss}\right), \tag{A.9}$$

where "diag" returns a diagonal matrix with the same diagonal entries as its input argument.

In the free noise case, things are quite similar except that the noise covariance matrices $R_{N,q}$ do not factor out as nicely.
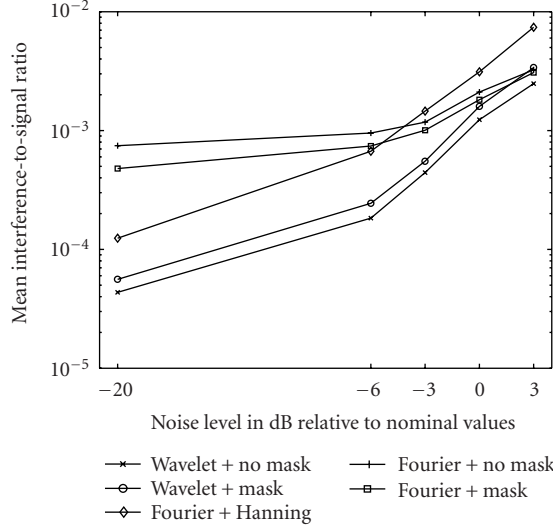
FIGURE 10: Comparison of the mean ISR for CMB as a function of noise in five different configurations, namely, wSMICA without mask, wSMICA with mask, fSMICA without mask, and fSMICA with mask, and fSMICA with Hanning apodizing window.
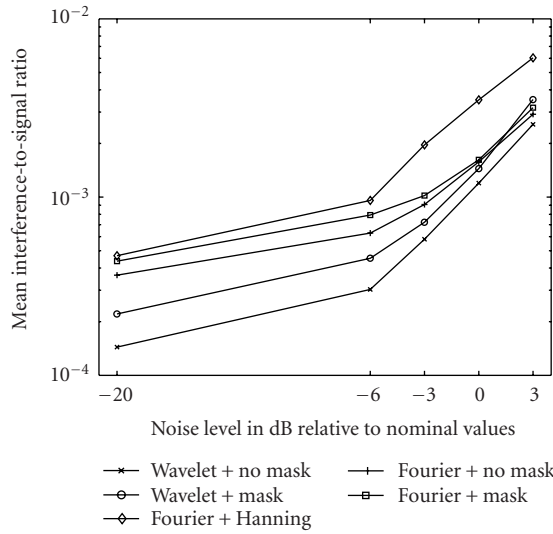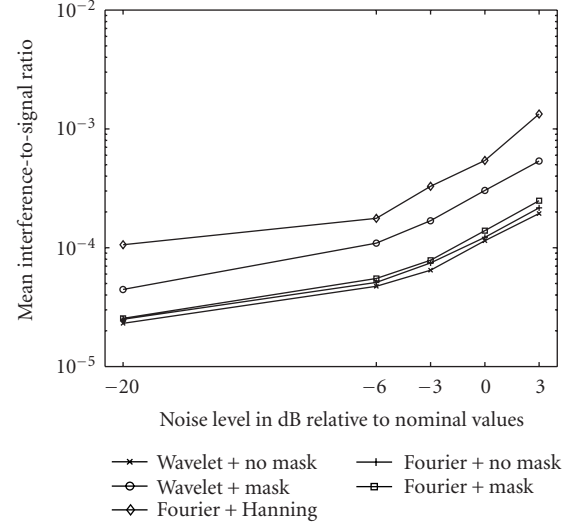

FIGURE 11: Comparison of the mean ISR for *dust* as a function of noise in five different configurations, namely, wSMICA without mask, wSMICA with mask, fSMICA without mask, fSMICA with mask, and fSMICA with Hanning apodizing window.

The EM functional is again expressed as

$$\Phi(\underline{\theta}, \theta) = -\frac{1}{2}(\underline{\mathcal{A}} - \mathcal{M})^{\dagger}\mathcal{Q}(\underline{\mathcal{A}} - \mathcal{M}) + \mathrm{const}_{\underline{A}}, \quad (A.10)$$

where in this case

$$\mathcal{Q} = \sum_q n_q R_{N,q}^{-1} \otimes R_q^{ss}, \qquad \mathcal{M} = \mathcal{Q}^{-1} \mathrm{vect}\left(\sum_q n_q R_{N,q}^{-1} R_q^{xs}\right). \quad (A.11)$$

FIGURE 12: Comparison of the mean ISR for SZ as a function of noise in five different configurations, namely, wSMICA without mask, wSMICA with mask, fSMICA without mask, fSMICA with mask, and fSMICA with Hanning apodizing window.

Then, the maximum of the EM functional with respect to $\underline{\theta}$ subject to the specified linear constraints is again reached for

$$\underline{\mathcal{A}} = \mathcal{M} - \mathcal{Q}^{-1}\mathcal{C}\left(\mathcal{C}^{\dagger}\mathcal{Q}^{-1}\mathcal{C}\right)^{-1}\mathcal{C}^{\dagger}(\mathcal{M} - \mathcal{A}_0),$$

$$\underline{R}_{S,q} = \mathrm{diag}\left(R_q^{ss}\right). \quad (A.12)$$

These expressions of the reestimates of the mixing matrix can become algorithmically very simple when, for instance, the linear constraints to be dealt with affect separate lines of A, or even simpler when the constraints are such that the entries of A are affected separately.

Positivity constraints on the entries of A

Suppose a subset of entries of A are constrained to be positive. The maximization step of the EM algorithm on A alone, again has to be modified. We suggest dealing with such constraints in a combinatorial way rephrasing the problem in terms of equality constraints. If the unconstrained maximum of the EM functional is not in the specified domain, then one has to look for a maximum on the borders of that domain: on a hyperplane, on the intersection of two, three, or more hyperplanes. One important point is that the maximum of the EM functional with respect to A subject to a set of equality constraints will necessarily be lower than the maximum of the same functional considering any subset of these equality constraints. Hence, not all combinations need be explored, and a branch-and-bound-type algorithm is well suited [28]. A straightforward extension allows to deal with the case where a set of entries of the mixing matrix is constrained by upper and lower bounds.
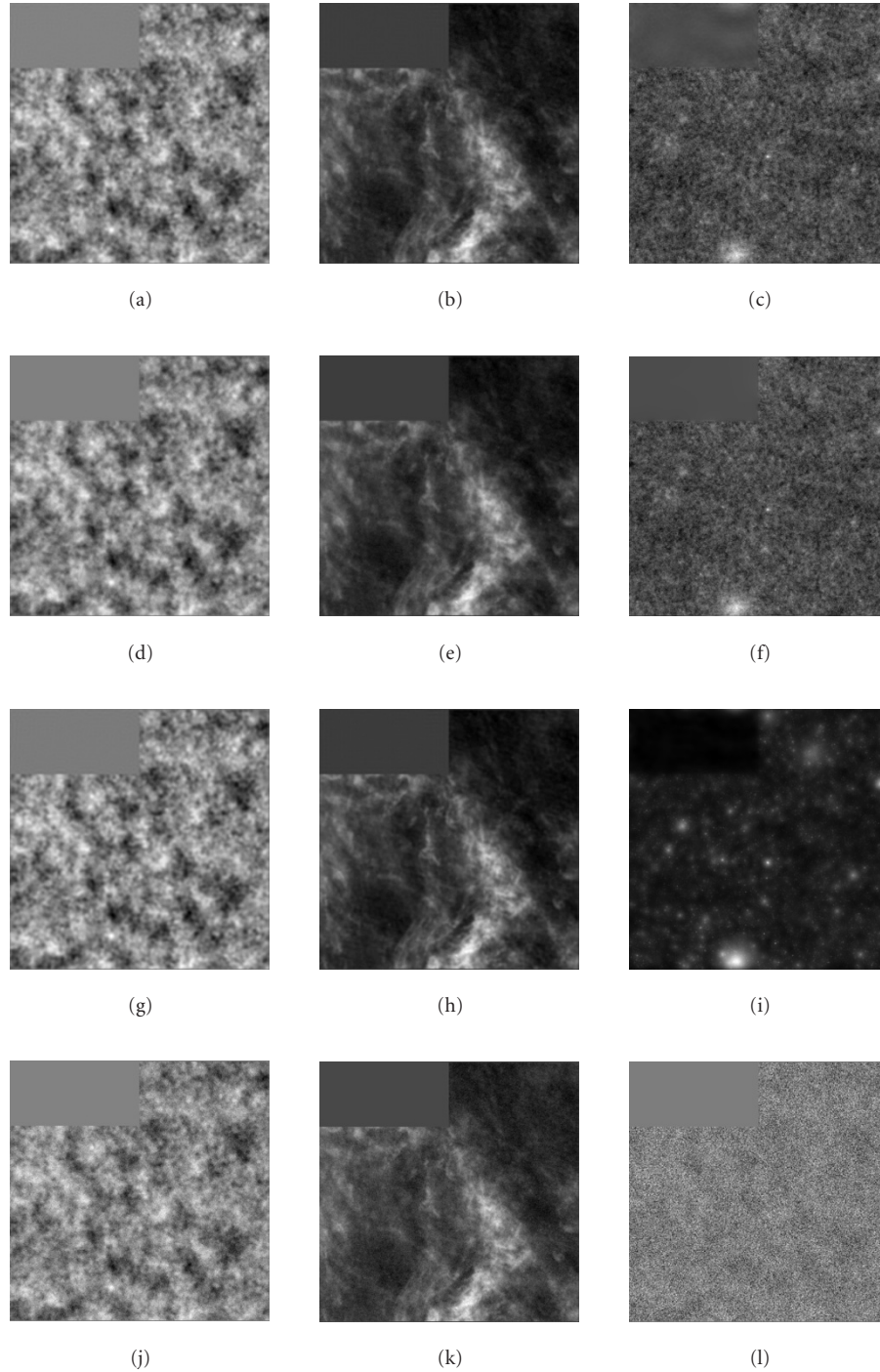
FIGURE 13: (a)–(f) Estimated component maps obtained with SMICA and wSMICA, respectively. These estimates result from applying a Wiener filter in each frequency band or wavelet scale based on the optimized model parameters (see Section 2.3). (g)–(i) The initial source templates after applying the optimal Wiener filter obtained with SMICA, that is, the same as (a)–(c) but leaving out noise and residual contaminations. (j)–(l) Maps estimated using JADE [27]. The initial source maps are shown in Figure 2.

REFERENCES

[1] G. F. Smoot, C. L. Bennett, A. Kogut, et al., "Structure in the COBE differential microwave radiometer first-year maps," *Astrophysical Journal Letters*, vol. 396, no. 1, pp. L1–L5, 1992.

[2] A. Benoît, P. Ade, A. Amblard, et al., "Archeops: a high resolution, large sky coverage balloon experiment for mapping cosmic microwave background anisotropies," *Astroparticle Physics*, vol. 17, no. 2, pp. 101–124, 2002.

[3] P. de Bernardis, P. Ade, J. J. Bock, et al., "A flat universe from

high-resolution maps of the cosmic microwave background radiation," *Nature*, vol. 404, no. 6781, pp. 955–959, 2000.

[4] S. Hanany, P. Ade, A. Balbi, et al., "MAXIMA-1: a measurement of the cosmic microwave background anisotropy on angular scales of 10 arcminutes to 5 degrees," *Astrophysical Journal Letters*, vol. 545, pp. L5–L9, 2000.

[5] C. L. Bennett, M. Halpern, G. Hinshaw, et al., "First year Wilkinson microwave anisotropy probe (WMAP) observations: preliminary maps and basic results," *Astrophysical Journal Supplement Series*, vol. 148, no. 1, pp. 1–27, 2003.

[6] G. Jungman, M. Kamionkowski, A. Kosowsky, and D. N. Spergel, "Cosmological-parameter determination with microwave background maps," *Physical Review D*, vol. 54, no. 2, pp. 1332–1344, 1996.

[7] F. R. Bouchet and R. Gispert, "Foregrounds and CMB experiments: I. Semi-analytical estimates of contamination," *New Astronomy*, vol. 4, no. 6, pp. 443–479, 1999.

[8] M. Tegmark and G. Efstathiou, "A method for subtracting foregrounds from multi-frequency CMB sky maps," *Monthly Notices of the Royal Astronomical Society*, vol. 281, no. 4, pp. 1297–1314, 1996.

[9] M. P. Hobson, A. W. Jones, A. N. Lasenby, and F. R. Bouchet, "Foreground separation methods for satellite observations of the cosmic microwave background," *Monthly Notices of the Royal Astronomical Society*, vol. 300, no. 1, pp. 1–29, 1998.

[10] M. Tegmark, D. J. Eisenstein, W. Hu, and A. de Oliveira-Costa, "Foregrounds and forecasts for the cosmic microwave background," *The Astrophysical Journal*, vol. 530, no. 1, pp. 133–165, 2000.

[11] C. Baccigalupi, L. Bedini, C. Burigana, et al., "Neural networks and the separation of cosmic microwave background and astrophysical signals in sky maps," *Monthly Notices of the Royal Astronomical Society*, vol. 318, no. 3, pp. 769–780, 2000.

[12] D. Maino, A. Farusi, C. Baccigalupi, et al., "All-sky astrophysical component separation with fast independent component analysis (FASTICA)," *Monthly Notices of the Royal Astronomical Society*, vol. 334, no. 1, pp. 53–68, 2002.

[13] E. E. Kuruoglu, L. Bedini, M. T. Paratore, E. Salerno, and A. Tonazzini, "Source separation in astrophysical maps using independent factor analysis," *Neural Networks*, vol. 16, no. 3-4, pp. 479–491, 2003.

[14] H. Snoussi, G. Patanchon, J. Macias-Perez, A. Mohammad-Djafari, and J. Delabrouille, "Bayesian blind component separation for cosmic microwave background observations," in *Bayesian Inference and Maximum Entropy Methods, MaxEnt Workshops*, pp. 125–140, American Institute of Physics, New York, NY, USA, August 2001.

[15] J.-F. Cardoso, H. Snoussi, J. Delabrouille, and G. Patanchon, "Blind separation of noisy Gaussian stationary sources. Application to cosmic microwave background imaging," in *Proc. 11th European Signal Processing Conference (EUSIPCO '02)*, vol. 1, pp. 561–564, Toulouse, France, September 2002.

[16] J. Delabrouille, J.-F. Cardoso, and G. Patanchon, "Multidetector multicomponent spectral matching and applications for cosmic microwave background data analysis," *Monthly Notices of the Royal Astronomical Society*, vol. 346, no. 4, pp. 1089–1102, 2003.

[17] J.-F. Cardoso, "The three easy routes to independent component analysis; contrasts and geometry," in *Proc. 3rd International Conference on Independent Component Analysis (ICA '01)*, San Diego, Calif, USA, December 2001.

[18] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computation*, vol. 13, no. 4, pp. 863–882, 2001.

[19] M. M. Ichir and A. Mohammad-Djafari, "Wavelet domain blind image separation," in *Wavelets: Applications in Signal and Image Processing X*, vol. 5207 of *Proceedings of SPIE*, pp. 361–370, San Diego, Calif, USA, November 2003.

[20] G. Patanchon, J. Delabrouille, and J.-F. Cardoso, "Source separation on astrophysical data sets from the WMAP satellite," in *Proc. 5th International Conference on Independent Component Analysis (ICA '04)*, pp. 1221–1228, Granada, Spain, September 2004.

[21] J.-F. Cardoso and D.-T. Pham, "Optimization issues in noisy Gaussian ICA," in *Proc. 5th International Conference on Independent Component Analysis (ICA '04)*, pp. 41–48, Granada, Spain, September 2004.

[22] E. Hivon, K. M. Górski, C. B. Netterfield, B. P. Crill, S. Prunet, and F. Hansen, "MASTER of the cosmic microwave background anisotropy power spectrum: A fast method for statistical analysis of large and complex cosmic microwave background data sets," *The Astrophysical Journal*, vol. 567, no. 1, pp. 2–17, 2002.

[23] J.-L. Starck, F. Murtagh, and A. Bijaoui, *Image and Data Analysis: The Multiscale Approach*, Cambridge University Press, Cambridge, UK, 1998.

[24] M. Birkinshaw, "The Sunyaev-Zel'dovich effect," *Physics Reports*, vol. 310, no. 2-3, pp. 97–195, 1999.

[25] D.-T. Pham, "Blind separation of instantaneous mixture of sources via the Gaussian mutual information criterion," *Signal Processing*, vol. 81, no. 4, pp. 855–870, 2001.

[26] D. T. Pham, "Joint approximate diagonalization of positive definite hermitian matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 22, no. 4, pp. 1136–1152, 2001.

[27] J.-F. Cardoso, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, no. 1, pp. 157–192, 1999.

[28] P. Narendra and K. Fukunaga, "A branch and bound algorithm for feature subset selection," *IEEE Trans. Comput.*, vol. 26, no. 9, pp. 917–922, 1977.

Y. Moudden graduated in electrical engineering from SUPELEC, Gif-sur-Yvette, France, and obtained an M.S. degree in physics from the Université de Paris VII, France, in 1997. He received a Ph.D. degree in signal processing from the Université de Paris XI, Orsay, France. He was a Visitor at UCLA in 2004 and is currently a Postdocotoral Fellow with the CEA, Saclay, France, working on applications of signal processing to astronomy. His research interests include signal and image processing, data analysis, and statistics and information theory.
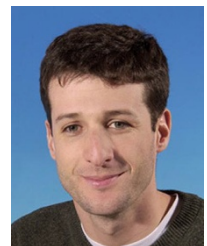
J.-F. Cardoso is with the French "Centre National de la Recherche Scientifique" (CNRS), in the Signal and Image Processing Department, École Nationale Supérieure des Télécommunications (ENST, Paris). His research area is statistical signal processing. Since 1989, he has been extensively working on all aspects of blind source separation and independent component analysis. In 2001, he started a collaboration with astronomers and cosmologists for the statistical analysis of astronomic data.

J.-L. Starck has a Ph.D. degree from University Nice-Sophia Antipolis and a Habilitation from University Paris XI. He was a Visitor at the European Southern Observatory (ESO) in 1993, at UCLA in 2004, and at Stanford's Statistics Department in 2000 and 2005. He has been a researcher at CEA since 1994. His research interests include image processing and statistical methods in astrophysics and cosmology. He is also the author of two books entitled *Image Processing and Data Analysis: The Multiscale Approach* (Cambridge University Press, 1998), and *Astronomical Image and Data Analysis* (Springer, 2002).

J. Delabrouille graduated in 1991 from the "École Nationale Supérieure des Télécommunications" with a major in aerospace telecommunication (1991). He then obtained in 1998 a Ph.D. degree in physics from the University of Chicago and a Ph.D. degree in astrophysics and space techniques from the University of Orsay. He is currently employed by CNRS in the Astroparticle and Cosmology (APC) Laboratory in Paris, where he is the Head of the Data Analysis and Simulation Group. His main research interests are in cosmology and gravitation, cosmic microwave background observations, and developing methods for complex data analysis.