

Sunyaev-Zel'dovich cluster reconstruction in multiband bolometer camera surveys

S. Pires¹, J. B. Juin¹, D. Yvon¹, Y. Moudden¹, S. Anthoine², and E. Pierpaoli³

¹ CEA – CE Saclay, DSM/DAPNIA, 91191 Gif-sur-Yvette Cedex, France
e-mail: yvon@hep.saclay.cea.fr

² PACM, Princeton University, Princeton, NJ 08544, USA

³ California Institute of Technology, Mail Code 130-33, Pasadena, CA 91125, USA

Received 12 July 2005 / Accepted 30 March 2006

ABSTRACT

We present a new method for the reconstruction of Sunyaev-Zel'dovich (SZ) galaxy clusters in future SZ-survey experiments using multiband bolometer cameras such as Olimpo, APEX, or Planck. Our goal is to optimise SZ-Cluster extraction from our observed noisy maps. None of the algorithms used in the detection chain is tuned using prior knowledge of the SZ-Cluster signal, or other astrophysical sources (Optical Spectrum, Noise Covariance Matrix, or covariance of SZ Cluster wavelet coefficients). First, a blind separation of the different astrophysical components that contribute to the observations is conducted using an Independent Component Analysis (ICA) method. This is a new application of ICA to multichannel astrophysical data analysis. Then, a recent non linear filtering technique in the wavelet domain, based on multiscale entropy and the False Discovery Rate (FDR) method, is used to detect and reconstruct the galaxy clusters. We use the Source Extractor software to identify the detected clusters. The proposed method was applied on realistic simulations of observations that we produced as mixtures of synthetic maps of the four brightest light sources in the range 143 GHz to 600 GHz namely the Sunyaev-Zel'dovich effect, the Cosmic Microwave Background (CMB) anisotropies, the extragalactic InfraRed point sources and the Galactic Dust Emission. We also implemented a simple model of optics and noise to account for instrumental effects. Assuming nominal performance for the near future SZ-survey Olimpo, our detection chain recovers 25% of the cluster of mass larger than $10^{14} M_{\odot}$, with 90% purity. Our results are compared with those obtained with published algorithms. This new method has a high global detection efficiency in the high-purity/low completeness region, being however a *blind* algorithm (i.e. without using any prior assumptions on the data to be extracted).

Key words. cosmology: large-scale structure of Universe – cosmology: observations – galaxies: clusters: general – methods: data analysis – methods: numerical

1. Introduction

Spectacular advances in cosmology have taken place through observation of the sky at millimeter and sub-millimeter wavelengths. The Boomerang (Netterfield et al. 2002), MAXIMA (Hanany et al. 2000), Archeops (Tristram et al. 2005), WMAP (Hinshaw et al. 2003), CBI (Contaldi et al. 2002) and DASI (Halverson et al. 2002) experiments measured the anisotropies of the Cosmic Microwave Background with high precision. Combined with other observations, such as distant supernovae of Riess et al. (1999); Knop et al. (2003) and/or Clusters (White et al. 1999; Pierpaoli et al. 2001) or the Large-Scale Structure Observation (Eisenstein et al. 2005), these experiments allowed one to place tight constraints on parameters of generic cosmological models. These results gave rise to new questions. Consensus cosmological models assume that the large-scale evolution of the Universe is driven by Dark Matter and Dark Energy. Dark Matter has been sought for 30 yr, and most candidates have been rejected in the light of experimental results. Dark Energy will be one of the toughest challenges of modern cosmology. One way to learn about the nature of dark energy is to study the evolution of the Universe at late times, during which dark energy is thought to have impacted on the formation of large-scale structures. Cluster surveys are an ideal tool for this purpose.

1.1. Cluster detection in the SZ survey

Large-scale structures can be studied using several probes: optical galaxy catalogs, X-ray cluster surveys such as XMM-LSS (Pierre et al. 2004), SZ cluster surveys such as Olimpo (Masi et al. 2003), APEX, AMI (Jones 2001), AMIBA (AMIBA 2005–2007) and weak shear such as CFHT-LS “wide”. The mass and redshift distribution of clusters of galaxies, $dN/dz dM d\Omega$ (where M , z and Ω are the cluster mass and redshift, and the solid angle), is predicted with fair precision using Halo models and N -body simulations. SZ cluster, X-ray and optical surveys measure quantities related to cluster mass through complex phenomena, such as interstellar gas heating processes, clustering and processes of stellar formation and evolution that are still not fully understood. The SZ signal from galaxy clusters is believed to be the simplest direct observable of cluster mass, though some uncertainties on extragalactic gas heating mechanisms remain. Measuring the cluster mass function will require these systematics to be understood. Pioneering SZ cluster surveys will obtain their first data this year. The Planck satellite survey will follow (Lamarre et al. 2004). In this paper we prepare for their analysis using simulated data.

Melin (2004) and Schafer (2004), following Herranz et al. (2002b,a), developed a cluster extraction tool based on an optimised filter. The filter is optimised to damp the noise as well

as the other astrophysical components. The SZ signal is small compared to the other astrophysical sources, rendering the task of filtering difficult. In addition a filter damps some of the signal at the same time as it rejects parasites. We developed the idea of a 3 step method using as a first step a source separation algorithm. The source separation algorithm sorts the noisy SZ signal from the other astrophysical components and the filtering step is left with the easier task of improving the SZ cluster signal versus noise. A detection algorithm extracts the SZ cluster candidates.

In the 100 to 600 GHz range, the brightest components of the sky are the Cosmic Microwave Background (CMB), the Infrared Point Sources, the Galactic dust emission and, swamped by the previous ones, the SZ clusters. It follows that the true sky map $X_v^{\text{true}}(\vartheta, \varphi)$, in a given optic band centered on ν , can be modeled as a sum of distinct astrophysical radiations as in

$$X_v^{\text{true}}(\vartheta, \varphi) = \text{CMB}_v(\vartheta, \varphi) + \text{IR}_v(\vartheta, \varphi) + \text{Gal}_v(\vartheta, \varphi) + \text{SZ}_v(\vartheta, \varphi) \quad (1)$$

where ϑ, φ denote spatial or angular indexes on 2D or spherical maps. The *observed* sky map $X_v(\vartheta, \varphi)$ is the result of the convolution of the true sky with the optical beam of the telescope plus the unavoidable contribution of instrumental noise $N_v(\vartheta, \varphi)$. Multiband signal processing in astrophysical applications deals with exploiting correlations between data maps at different ν , to enhance the estimation of specific objects of interest such as the power spectrum of the CMB spatial fluctuations, or the mass function of galaxy clusters. Considering the case where the radiative properties of the sources are completely isotropic, in the sense that they do not depend on the direction of observation, the above model can be rewritten in the following factored form:

$$X_v(\vartheta, \varphi) = \sum_i a_{v,i} S_i(\vartheta, \varphi) + N_v(\vartheta, \varphi) \quad (2)$$

where S_i is the spatial template (i.e. the convolution of the true component map with the instrumental beam at frequency ν) and $a_{v,i}$ the emission law of the i th astrophysical component. Although this is a coarse approximation in the case of Infrared Point Sources as described in Sect. 2.2, it is mostly valid for the other three components. With observations available in m channels, Eq. (2) can be written in matrix form:

$$X(\vartheta, \varphi) = A S(\vartheta, \varphi) + N(\vartheta, \varphi) \quad (3)$$

where $X(\vartheta, \varphi)$ is a vector in \mathbb{R}^m , A is an $m \times n$ matrix, n is the number of contributing astrophysical components, $S(\vartheta, \varphi)$ is now a vector in \mathbb{R}^n and $N(\vartheta, \varphi)$ in \mathbb{R}^m . An important simplifying assumption made here is that the instrumental beam does not vary with the optical frequency ν . This is acceptable with Olimpo but definitely not for other multichannel experiments such as Planck. Accounting for a varying instrumental beam requires considering a more complex convolutive mixture model for the observed multichannel data either with known or parameterized beams. Some methods have been developed to deal with such mixtures but most of these move to the Fourier domain and consider an instantaneous mixture model in this representation. This is not acceptable here as the component of interest (i.e. SZ clusters) would be below the noise level in all channels in this representation: this is due to the non-locality of the Fourier basis functions. Instead, we chose to move the data to the wavelet domain. Wavelets are known to provide sparse representations of images with point singularities in an otherwise smooth background. Thus enhancement of the SZ signal concentration is expected. Then, once computed, in order to reconstruct the

astrophysical component maps, the estimated separating filter is applied onto the original data.

Equation (3) shows that the observations consist of linear mixtures of astrophysical components with different weights and additive noise. Assuming no prior knowledge of A , the problem of recovering the particular component of interest i.e. the SZ effect map of galaxy clusters, can be seen as a blind source separation problem which we approach in this paper using Independent Component Analysis (ICA). Such methods have been previously used quite successfully in the analysis of astrophysical data from present or future multichannel experiments such as WMAP or Planck (Moudden et al. 2005; Maino et al. 2002; Kuruoglu et al. 2003; Bobin et al. 2005) where the focus is on estimating the map of CMB anisotropies. We concentrate here on separating the SZ component, which has very different statistical properties compared to CMB and thus requires the use of other ICA methods. Then, due to the additive instrumental noise, the source separation process has to be followed by a filtering technique.

We use an iterative filtering based on Bayesian methods to filter the noise (Starck et al. 2006). This method uses a Multiscale Entropy prior which is only defined for non-significant wavelet coefficients selected by the False Discovery Rate (FDR) Method (Benjamini & Hochberg 1995). Then, we identify SZ clusters in filtered maps using the Source Extractor software SExtractor (Bertin 2003) which conveniently provides source extraction along with adequate photometry and deblending.

1.2. Outline of the paper

In the following, we first describe our sky model. We use the Olimpo SZ-cluster survey (Masi et al. 2003) as an example of a future ambitious SZ survey. We model the Olimpo instrument and produce simulated observed maps. With minor parameter changes, these algorithms will also allow us to simulate observed local maps typical of other bolometer camera SZ surveys (i.e. Planck or APEX observations). Then we explain the data processing methods used to recover the map of SZ-cluster signal and the “observed” cluster catalog. We quantify the efficiency of our detection chain and compare it to a recently published method. This paper is at the interface of astrophysics and applied mathematics.

2. Simulations of contributing astrophysical components

The input of our simulation is a list of cosmological model parameter values. We use a Λ CDM cosmological model with $\Omega_t = 1$, $\Omega_m = 0.3$, $\Omega_b = 0.04$, $\Omega_{dm} = 0.26$, $\Omega_\Lambda = 0.7$, $h = 0.7$, $\sigma_8 = 0.85$, $n_s = 1$, $\tau = 0.1666$ and a seed for the random number generator. This simulation was fully written in C++, so that we can mass produce our simulations on a PC farm, when needed.

2.1. Cosmic Microwave Background anisotropies

Big Bang cosmology models assume that the Universe has been evolving from a hot and dense plasma. While it expands, its contents cools. At some time, *decoupling* occurred, when the temperature of the Universe dropped below $h\nu = k_B T \approx 0.2$ eV and the mean free path of photons *suddenly* became infinite so that the large majority of them have been propagating freely through the Universe ever since. These photons are nowadays observed to have a black-body spectrum at temperature 2.728 K

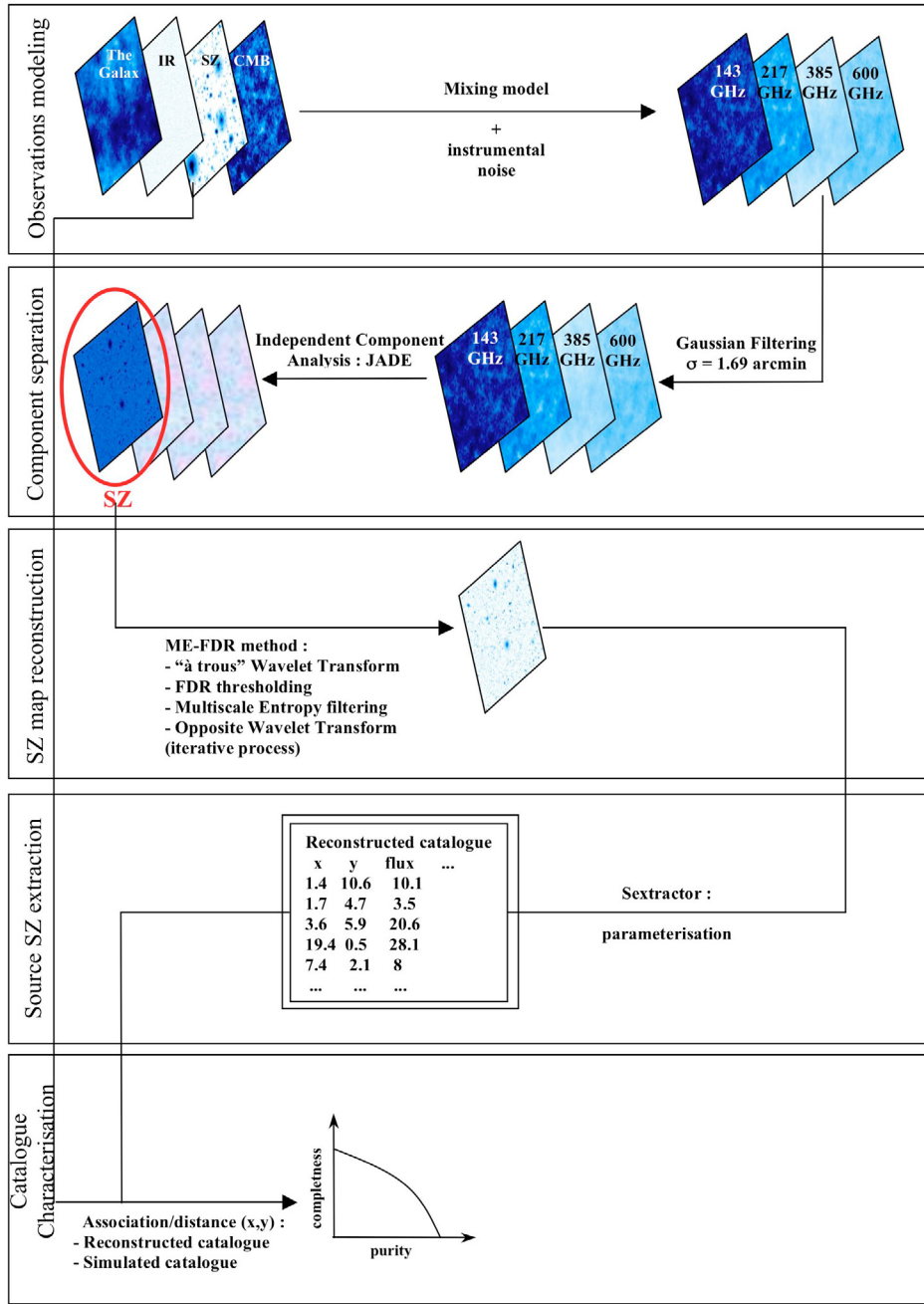


Fig. 1. Graphical summary of the simulation and detection chain presented in this paper.

(Hinshaw et al. 2003): the Cosmic Microwave Background (CMB). Small anisotropies in the CMB are observed and interpreted as resulting from temperature and density fluctuations in the primordial plasma at the decoupling time.

Bolometer cameras measure incoming power variations, while scanning the sky. Thus the main blackbody spectrum (monopole) will not be observed. In the following simulations we also assume that the CMB dipole (Doppler effect due to the Earth's movement) has been subtracted from the data.

In order to simulate a map of CMB anisotropies, we first enter the cosmological parameter list in the *cmbeasy* code (Doran & Mueller 2004) and compute a Power Spectrum in Spherical Harmonics up to multipole 5000 of the CMB fluctuations. Then using a small field approximation (White et al. 1999), we generate a random spatially correlated Gaussian field of primordial

anisotropies with the previous power spectrum. Figure 2 shows a typical simulated CMB anisotropy map in units of μK .

2.2. IR point sources

In our frequency range, infrared point sources are expected to be a significant contamination. Borys et al. (2003) from the SCUBA experiment published a list of infrared point sources at 350 GHz and a $\log N / \log S$ law from this data. We extend this law to lower brightness, and generate a catalog of IR point sources in our field of view. We assume that their optical spectra follow a grey body law parameterised as:

$$F(\nu) = \nu^\alpha * Planck(\nu, T_0). \quad (4)$$

Where *Planck* stands for Planck's black body spectrum, ν stands for frequency, T_0 is the temperature of the black body and α is

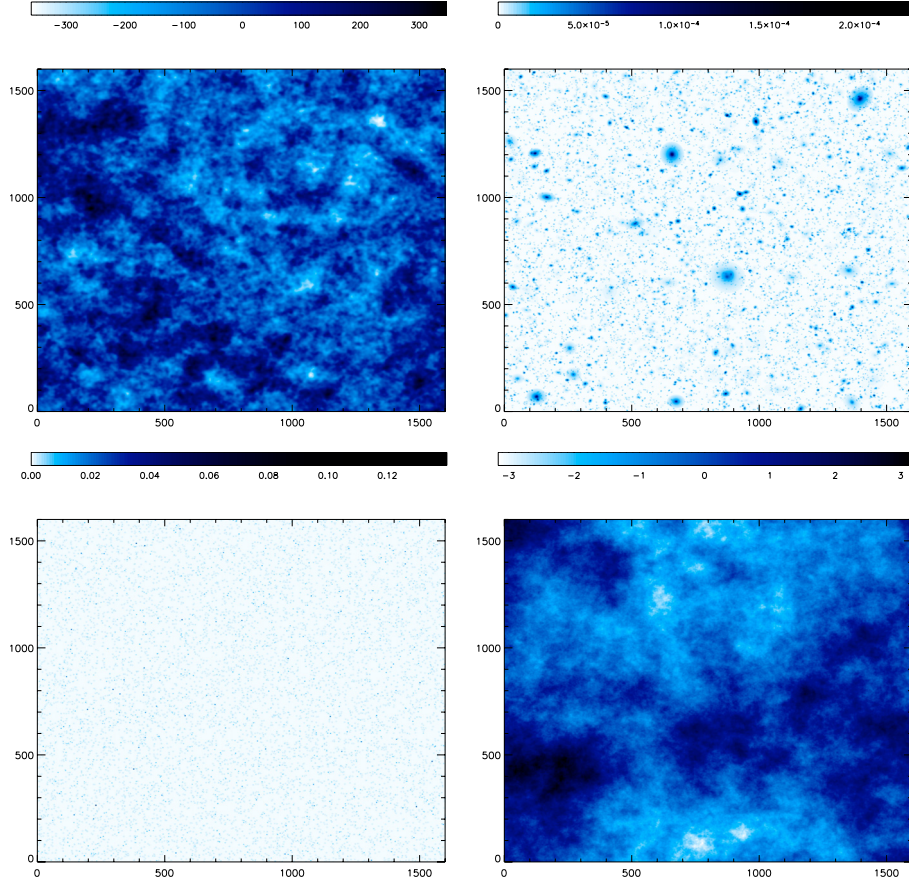


Fig. 2. The 4 physical components of the sky included in our simulation: *upper right* is a map of the CMB's anisotropies in unit of μK , *upper left* is the SZ Cluster map, in units of the Compton parameter on line of sight y , *lower right* is the IR point source map, convolved with a beam of 2 arcmin, in Jy at 350 GHz, *lower left* is the Galactic dust map in units of MJy/st at $100\ \mu\text{m}$.

by definition the spectral index. For each IR point source, the spectral index was randomly picked between 1.5 and 2, and T_0 was set to 30 K. Thus each IR point source has a different optical spectrum. Then the IR point source is placed on our map as a single pixel. IR sources are positioned randomly on the map but a user chosen fraction of them can be positioned within clusters of galaxies, and would make SZ-Cluster photometry more difficult. We obtain an IR point source map in units of Jy at 350 GHz as shown in Fig. 2.

2.3. Galactic dust emission

We know from FIRAS and IRAS that the galactic interstellar dust emits in the far infrared. For these simulations, we wanted to generate truly random galactic dust maps. So, instead of using the maps published by Schlegel et al., we followed Bouchet & Gispert (1999) and assumed that Galactic Dust was described by a power spectrum of spatial correlations in $C_l \propto 1/l^3$. We then randomly generated a map with the same algorithms as used for the CMB map simulations. We assume a homogeneous optical spectrum over our map given by Eq. (4) with $\alpha = 2$, and $T_0 = 20$ K. This is a reasonable assumption since SZ-Cluster observations will take place at high galactic latitudes, and we will be working on small size maps. Finally, we normalise the map rms at its observed values at high galactic latitude: at $\lambda = 100\ \mu\text{m}$, the rms flux is observed at 1 MJy/st. We obtain a galactic Dust Emission Map, shown in Fig. 2 in natural units of MJy/st at $100\ \mu\text{m}$.

2.4. Sunyaev-Zel'dovich clusters

Hot intracluster gas is observed as a plasma at a temperature of 2 to 20 keV. A small fraction of CMB photons traveling through clusters Compton-scatter on electron gas. Since the electron kinetic energy is much higher than the CMB photon energy, CMB photons statistically gain energy when diffused. This effect, known as the thermal Sunyaev-Zel'dovich effect (1980), results in a small distortion in the CMB Black-Body spectrum, in the direction of clusters of galaxies. The probability that a photon scatters while crossing a cluster is given by:

$$y = \int_{\text{los}} \frac{kT_e}{m_e c^2} n_e \sigma_T dl = \frac{k\sigma_T}{m_e c^2} \int_{\text{los}} T_e n_e dl. \quad (5)$$

The optical spectral dependance is given by:

$$j(\nu) = 2 \frac{(kT_0)^3}{(h\nu c)^2} \frac{x^4 e^x}{(e^x - 1)^2} \left(\frac{x}{\tan h(x/2) - 4} \right). \quad (6)$$

In these expressions, T_e , n_e , m_e refer to the electron temperature, density and mass, respectively; $\sigma_T = 6.65 \times 10^{-25}\ \text{cm}^2$ is the Thomson cross-section and $x = h\nu/kT_{\text{CMB}}$ is the dimensionless frequency of observation in terms of unperturbed CMB temperature $T = 2.728\ \text{K}$.

Simulation of SZ Cluster maps requires modelling large-scale structure formation. For this we translated in C++ the ICosmo semi-analytic model (Refregier & R. Massey 2004) which, starting from a cosmological parameter set, allows us to compute the cluster density versus mass and redshift: $dn/dz dM d\Omega(M, z)$. Several fits to data and physical assumptions

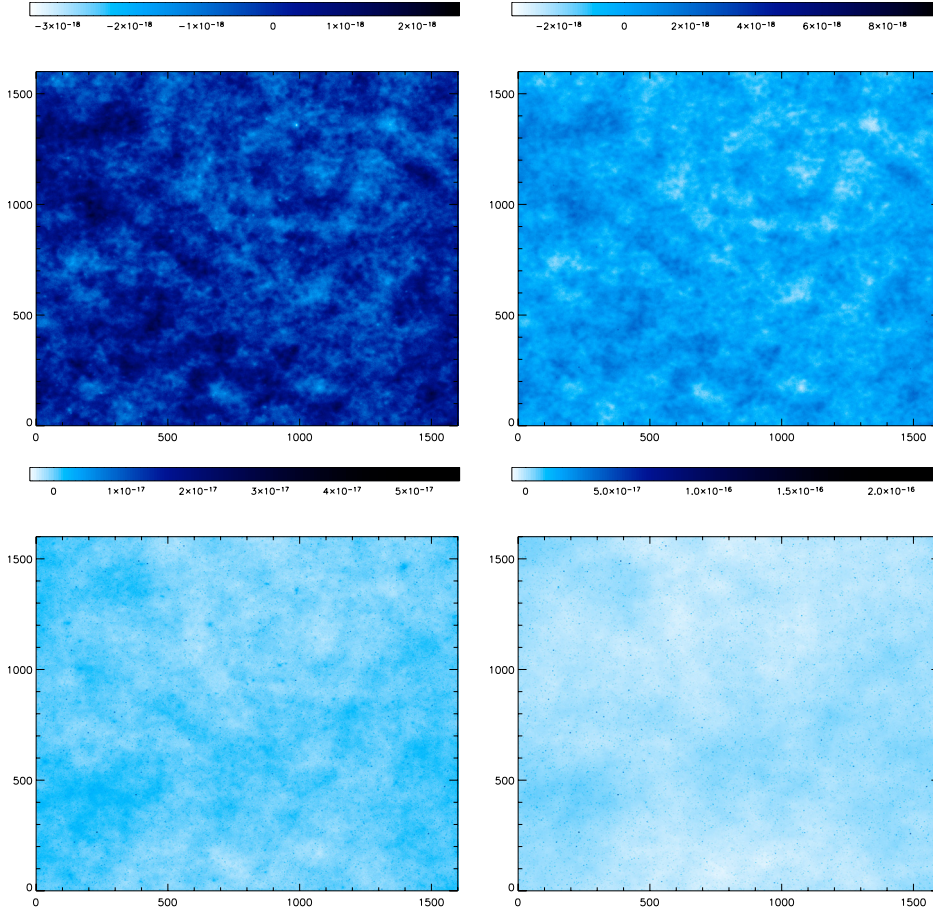


Fig. 3. Simulated maps in Olimpo's four frequency bands: *upper right* is the 147 GHz band, *upper left* is the 217 GHz band, *lower right* is the 385 GHz band: CMB anisotropies, IR point sources and Galactic Dust blend in this band. *Lower left* 500 GHz: IR point sources and Galactic Dust are the dominant features at high frequencies. The SZ cluster signal is dominated by other astrophysical sources at all frequencies.

are involved and implemented in this computation. We run C-ICosmo using the options of a concordance Λ CDM model and the mass function from Sheth & Thormen (Sheth et al. 2001). The virial radius of a cluster is computed assuming a spherical collapse (Peebles 1993; Peacock 1999). This is the radius where the average overdensity is $\Delta_c = 1 + \delta = 200$ times the critical density (chosen independent of redshift). Gas heating is modeled according to Pierpaoli et al. (2003), with the additional parameter T^* . Then we randomly generate a cluster catalog in our field of view, according to density profile keeping those clusters which have an integrated Compton parameter Y larger than 3×10^{-12} st (corresponding to a mass cut of 10^{14} Solar Mass).

An ellipticity is attributed to each cluster according to experimental observations (Cooray 2000). We assume that the cluster gas density follows an elliptical beta model. For the time being, clusters are positioned on the map randomly and the orientation of the main axis of each ellipse is also chosen randomly. We project to account for two point correlations in cluster positions in the near future. A typical SZ cluster map in natural units of y -Compton is shown in Fig. 2.

3. Modeling observations

3.1. Instrumental effects

A detailed model of observations can only be developed after data acquisition. In the following, we choose to use a very simple observation model, nevertheless representative of an ongoing

project, the Olimpo balloon experiment. The Olimpo camera will observe the sky at 4 frequencies, 143, 217, 385 and 600 GHz. The millimeter-wave filter bandwidths will be close to 30 GHz and are assumed to have a top-hat shape. The beams at the 4 frequencies are assumed to be symmetric, Gaussian and of 3, 2, 2 and 2 arcmin $FHWM$ respectively. The optical efficiency (mirrors, filters), the photon noise and the bolometer sensitivities and noise contributions are summarized by an equivalent observed noise temperature n_{eqT} on the Sky in unit of $\mu K_{CMB}/Hz^{-1/2}$. Based on the BOOMERanG experiment, typical values in the four channels of Olimpo are expected at 150, 200, 500, and 5000 $\mu K_{CMB}/Hz^{-1/2}$ respectively.

3.2. Noise model

The simplest and most optimistic model is to assume that the observed noise along bolometer timelines will be white, in which case, after projection (or *coaddition*, Yvon & Mayet 2005) on 2D maps, pixel noise can be very easily computed as:

$$noise_{pix} = n_{eqT} * \sqrt{N_{Bol} * t_{pix}} \quad (7)$$

where N_{Bol} stands for the number of bolometers working in a particular frequency band, and t_{pix} is the observation length on this pixel. High resolution SZ bolometer surveys will cover their large survey adding up small patches of the sky. Noise map patterns are likely to be complex, and will only be known after data has been acquired. We chose in this first paper to begin with a simple case of homogeneous white noise on the full field of view.

Frequency [GHz]	143	217	385	600
<i>FWHM</i> [arcmin]	3	2	2	2
White noise level [$\mu\text{K}/\sqrt{\text{Hz}}$]	150	200	500	5000

3.3. Mixing model

We now have all the tools to compute simulated observed sky maps. For each frequency band, we compute a conversion factor from the maps in their natural units into the observed unit at the bolometer level namely $\text{pW}/\text{m}^2/\text{st}$. The exception is the IR point source map where, because of the random spectral indexes of the point sources, a conversion factor is computed for each point source. We sum the four physical components into a “true” sky map which is then convolved with the experimental beam. Then we add the noise. We end up with one map per frequency band (Fig. 3). These maps would be what the analysis team would recover from the data, after pointing reconstruction, parasite removal, de-correlation of instrumental systematics in the data, and map-making: much analysis work. In the following, we explain a set of algorithms optimised for recovering SZ cluster signals in the observed maps.

4. Separation of astrophysical components

The simulated observations generated according to the method described in the previous section consist of linear mixtures of the four main sources of diffuse radiation as expected in the four channels of the Olimpo instrument. Focusing on separating the SZ map, we want to make the most of the data in the four frequency bands (centered on 143, 217, 385 and 600 GHz) and exploit the fact that each map is actually a different point of view on the same scene: multichannel data requires processing these maps in a coherent manner. We now provide an overview of the general principles of ICA for component separation from multichannel data. Then, we give a more detailed description of JADE, the specific ICA algorithm we used.

4.1. Independent component analysis

Blind Source Separation (BSS) is a problem that occurs in multi-dimensional data processing. The overall goal is to recover unobserved signals, images or *sources* S from mixtures of these sources X observed typically at the output of an array of m sensors. The simplest mixture model takes the form of Eq. (3):

$$X = AS + N \quad (8)$$

where X and S are random vectors of respective sizes $m \times 1$, $n \times 1$ and A is an $m \times n$ matrix. The entries of S are assumed to be independent random variables. Multiplying S by A linearly mixes the n sources into m observed processes.

Independent Component Analysis methods were developed to solve the BSS problem, i.e. given a batch of T observed samples of X , estimate the mixing matrix A or its inverse and reconstruct the corresponding T samples of the source vector S , relying mostly on the statistical independence of the source processes. Inverting the mixing process requires that the number of mixtures m be greater than or equal to the number of sources n . Solving the under-determined case where $m < n$ requires some more a priori information on the sources. This is an currently active area of research with close ties to the ideas of sparse representation (Lee et al. 1999; Lewicki & Sejnowski 2000; Georgiev et al. 2005). Note that with the above model, the independent

sources can only be recovered up to a multiplication by a *non-mixing* matrix i.e. up to a permutation and a scaling of the entries of S . Although independence is a strong assumption, it is in many cases physically plausible. This exceeds the simple second order decorrelation obtained for instance using Principal Component Analysis (PCA): decorrelation is not enough to recover the source processes since any rotation of a white random vector remains a white random vector.

Algorithms for blind component separation and mixing matrix estimation depend on the a priori model used for the probability distributions of the sources (Cardoso 2001) although rather coarse assumptions can be made (Cardoso 1998; Hyvärinen et al. 2001). In a first set of techniques, source separation is achieved in a noiseless setting, based on the non-Gaussian nature of all but possibly one of the components. Most mainstream ICA techniques belong to this category: JADE (Cardoso 1999), FastICA, Infomax (Hyvärinen et al. 2001). In a second set of blind techniques, the components are modeled as Gaussian processes and, in a given representation (Fourier, wavelet, etc.), separation requires that the sources have diverse, i.e. non proportional, variance profiles. For instance, the Spectral Matching ICA method (SMICA) (Delabrouille et al. 2003; Moudden et al. 2005) considers the case of mixed stationary Gaussian components in a noisy context: moving to a Fourier representation, colored components can be separated based on the diversity of their power spectra.

In the case where the main component of interest is well modeled by a peaked distribution with long tails (e.g. a Laplace distribution) as is the case with SZ maps, methods from the first set are expected to yield better results. Next we provide a description of JADE, the non-Gaussian ICA method we used.

4.2. JADE

The Joint Approximate Diagonalization of Eigenmatrices method (JADE) assumes a linear mixture model as in (8) where the independent sources S are non-Gaussian iid¹ random processes with the additional assumption of a high signal to noise ratio (i.e. $N \approx 0$). The mixing matrix is assumed to be square and invertible so that (de)mixing is just a change of basis. Although the noiseless assumption may not be true in the problem at hand, the algorithm may still be applied and a proper change of representation can move us closer to such a setting, as discussed in the next Sect. 4.3.

As mentioned above, second order statistics do not retain enough information for source separation in this context: finding a change of basis in which the data covariance matrix is diagonal will not in general enable us to identify the independent sources properly. Nevertheless, decorrelation is *half the job* (Cardoso 1998) and one may seek the basis in which the data is represented by maximally independent processes among those bases in which the data is decorrelated. This leads to so-called orthogonal algorithms: after a proper whitening of the data by multiplication with the inverse of a square root of the covariance matrix of the data W , one seeks a rotation R (which leaves things white) so that \hat{S} defined by

$$\hat{S} = W^{-1} Y = W^{-1} R X_{\text{white}} = W^{-1} R W X \quad (9)$$

¹ The letters iid stand for independently and identically distributed meaning that each of the entries of X at a given position t are independent of X at any other position t' and that the distribution of X does not depend on position.

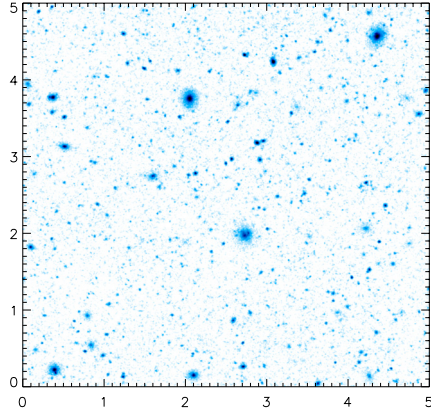


Fig. 4. SZ component map extracted by JADE from the four observed noisy maps. The SZ cluster signal, subdominant at all observed frequencies, now appears clearly. No obvious leftovers from other astrophysical sources are seen. Remaining noise is small, because we prefiltered data before JADE processing, and we simulated the nominal noise levels of an ambitious project: Olimpo.

and $\hat{B} = \widehat{A^{-1}} = W^{-1} R W$ are estimations of the sources and of the inverse of the mixing matrix.

JADE is such an orthogonal ICA method and, like most mainstream ICA techniques, it exploits higher order statistics so as to achieve a *non linear decorrelation*. In the case of JADE, statistical independence is assessed using fourth order cross cumulants defined by:

$$\begin{aligned} F_{ijkl} &= \text{cum}(y_i, y_j, y_k, y_l) \\ &= \mathcal{E}(y_i y_j y_k y_l) - \mathcal{E}(y_i y_j) \mathcal{E}(y_k y_l) \\ &\quad - \mathcal{E}(y_i y_l) \mathcal{E}(y_j y_k) - \mathcal{E}(y_i y_k) \mathcal{E}(y_j y_l) \end{aligned} \quad (10)$$

where \mathcal{E} stands for statistical expectation and the y_i are the entries of vector Y modeled as random variables. Then, the correct change of basis (i.e. rotation) is found by *diagonalizing* the fourth order cumulant tensor. If the y_i were independent, all the cumulants with at least two different indices would be zero. As a consequence of the independence assumption of the source processes S and of the *whiteness* of Y for all rotations R , the fourth order tensor F is well structured: JADE was precisely devised to take advantage of the algebraic properties of F . JADE objective function is given by

$$\mathcal{J}_{\text{JADE}}(R) = \sum_{ij} \sum_{k \neq l} \text{cum}(y_i, y_j, y_k, y_l)^2 \quad (11)$$

which can be interpreted as a joint diagonalization criterion. Fast and robust algorithms are available for the minimization of $\mathcal{J}_{\text{JADE}}(R)$ with respect to R based on Jacobi's method for matrix diagonalization (Pham 2001). More details on JADE can be found in Cardoso (1999, 1998); Hyvärinen et al. (2001).

4.3. JADE in wavelet space

We chose to use JADE after a Wavelet transform. Wavelets come into play as a sparsifying² transform. Moving the data to a wavelet representation does not affect its information content

² Data is sparse on a basis when this basis allows one to describe that signal with a small number of coefficients. This is a highly desirable property, since noise is not expected to be sparse at the same time in such a basis. Choosing a sparsifying basis thus allows one to enhance signal to noise ratio.

and applying a wavelet transform on both sides of (8) does not affect the mixing matrix and the model structure is preserved. However, the statistical distribution of the data coefficients in the new representation is different: wavelets are known to lead to sparse approximately iid representations of structured data. Further, the *local* (coefficient wise) signal to noise ratio depends on the choice of a representation. A wavelet transform tends to grab the informative coherence between pixels while averaging the noise contributions, thus enhancing structures in the data. Although the standard ICA model is for a noiseless setting, the derived methods can be applied to real data. Performance will depend on the detectability of significant coefficients i.e. on the sparsity of the statistical distribution of the coefficients. Moving to a wavelet representation will often lead to more robustness to noise. We used the bi-orthogonal wavelet transform based on the Antonini 7/9 filters (Antonini et al. 1992). Although this choice may not be optimal for the source separation objective, it does enhance separation. Finding the optimal representation of the data for component separation is not a trivial task given the diversity of structures encountered in the different component maps. Some very recent work in BSS considers exploiting the morphological diversity of the source processes by resorting to highly overcomplete dictionaries (i.e. unions of several bases) to enhance source separation in some specific cases (Zibulevsky & Pearlmutter 2001; Starck et al. 2004; Bobin et al. 2005; Starck et al. 2005). How this could be useful for astrophysical component separation in experiments such as Olimpo or Planck is currently being investigated.

Once the data has been transformed to a proper representation (e.g. wavelets but also ridgelets and curvelets (Starck et al. 2003), if the 2D or 3D data is strongly anisotropic), we apply the standard JADE method to the new multichannel coefficients. Once the mixing matrix is estimated, the initial source maps are obtained using the adequate inverse transform (Fig. 4).

5. Clusters restoration method – noise filtering

JADE has been used to separate the signals from four mixtures and four sources in the presence of Gaussian noise. We have added the expected experimental level of Gaussian noise to each observed mixture map. In order to prevent bias induced by pixelisation in our simulation and detection algorithms, we overpixelised our observed maps, and thus pixel noise is enhanced. But all algorithms for BSS that require whitening are sensitive to additive noise. In order to minimise the impact of this noise at the ICA step, we chose to convolve our observed map before JADE with a Gaussian of optical beams width. To optimise signal recovery, we perform a filtering to remove the remaining noise. We tested different methods to remove noise. Thanks to the original simulated SZ map (without noise), we can easily compare the results of those filterings in the next part.

5.1. Linear filtering

5.1.1. Gaussian filter

A rather common linear filtering technique uses a Gaussian filter, generally isotropic. The standard method consists of convolving the observed map S_{obs} with a Gaussian window G with standard deviation σ_G :

$$S_G = G * S_{\text{obs}}. \quad (12)$$

The filtering depends strongly on the value of σ_G . This value is adjusted arbitrarily or based on a priori information.

5.1.2. Wiener filter

An alternative to Gaussian filtering is Wiener filtering which is a method that attempts to minimize the mean squared error between the original and the restored signal. Wiener filtering convolves the observed map S_{obs} with a weight function i.e. by assigning the following weight to each mode in Fourier Space:

$$\hat{w}(u, v) = \frac{|\hat{S}(u, v)|^2}{|\hat{S}(u, v)|^2 + |\hat{N}(u, v)|^2} \quad (13)$$

where u, v are the spatial frequencies, $|\hat{S}(u, v)|^2$ is a model of the map power spectrum and is in practice derived from the data. The weight function makes it possible to attenuate or to remove part of the frequencies if the signal-to-noise ratio is low. The filtering depends on the model of the noise. The Wiener filter is the optimal filter if both the signal and the noise are well modeled as Gaussian Random Fields. This condition is not verified for SZ maps which display highly non-Gaussian features. Nevertheless, Wiener filtering generally outperforms the simple Gaussian filtering.

In the following paragraphs, we have tested a state of the art non-linear filtering method in order to improve signal recovery.

5.2. Multiscale entropy method using False Discovery Rate (FDR)

5.2.1. Maximum Entropy Method (MEM)

The Maximum Entropy Method (MEM) is commonly used in astronomy for image processing (see Starck et al. 2001; Marshall et al. 2002; Starck & Murtagh 2002, for a full description). It is based on entropy and Bayesian methods. The Bayesian approach provides the means to incorporate prior knowledge in data analysis. Choosing the prior is one of the most critical aspects of Bayesian analysis. Here an entropic prior is used. In data filtering, an entropy functional assesses the information content of a data set. Among several possible definitions of entropy, the most commonly used in image processing is the Gull & Skilling (1991) definition:

$$H_g(S) = \sum_{k,l} \left[S(k, l) - m(k, l) - S(k, l) \ln \left(\frac{S(k, l)}{m(k, l)} \right) \right] \quad (14)$$

where k and l are respectively the position and the scale index. m is a model, chosen typically to be the sky background. H_g has a global maximum at $S = m$. However, MEM does not allow negative values in the solution, which is a problem in experimental SZ data analysis, where we measure fluctuations about zero. To overcome this problem, it has been proposed (Maisinger et al. 2004) to replace H_g by:

$$H_{\pm}(S) = \sum_{k,l} \psi(k, l) - 2m - S(k, l) \ln \left(\frac{\psi(k, l) + S(k, l)}{2m} \right) \quad (15)$$

where $\psi(k, l) = \sqrt{S^2(k, l) + 4m^2}$. Here m does not play the same role as in (14). It is a constant fixed to the expected signal rms.

To overcome the difficulties encountered by the MEM to restore images containing both high and low frequencies, Pantin & Starck (1996) have suggested a definition of entropy in a multiscale framework which we describe in the next section. It has been shown that the main drawbacks of the MEM (i.e. model dependent solution, oversmoothing of compact objects, ...) disappear.

5.2.2. Multiscale Entropy

1. Multiscale Entropy definition: the Multiscale Entropy method is based on the standard MEM prior derived from the wavelet decomposition of a signal. In this section, we use the “a trous” algorithm, with the B^3 -Spline filter that matches well our cluster shapes and is fast to compute. The idea is to consider the entropy of a signal as the sum of the information at each scale of its wavelet transform. The information of a wavelet coefficient is related to the probability of its being due to noise. The Undecimated Isotropic Wavelet Transform (UIWT) decomposes an $n \times n$ image S as in:

$$S(k, l) = C_{J(k, l)} + \sum_{j=1}^J w_{j(k, l)}$$

where k, l are the position index and j is the scale index. C_J is a coarse or smooth version of the original image S and w_j represents the details in S at scale 2^{-j} (see Starck & Murtagh for details). Thus, the algorithm outputs $J+1$ sub-band arrays of size $n \times n$. We will use an indexing convention such that $j = 1$ corresponds to the finest scale (high frequencies). Denoting as $H(S)$ the information relative to the signal and $h(w_j(k, l))$ the information relative to a single wavelet coefficient, the entropy is now defined as:

$$H(S) = h(C_J(k, l)) + \sum_{j=1}^J \sum_{k,l=1}^n h(w_j(k, l)) \quad (16)$$

where l is the number of scales and n is the size of map.

2. Entropy definition: the function h in (16) assesses the amount of information carried by a specific wavelet coefficient. Several functions have been proposed for h . A discussion and comparison between different entropy definitions can be found in Starck et al. (2006). We choose the NOISE-MSE entropy Starck et al. (2001) for the SZ reconstruction problem in which the entropy is derived using a model of the noise contained in the data:

$$h(w_j(k, l)) = \int_0^{|w_j(k, l)|} P_n(|w_j(k, l)| - u) \left(\frac{\partial h(x)}{\partial x} \right)_{x=u} du \quad (17)$$

where $P_n(w_j(k, l))$ is the probability that the coefficient $w_j(k, l)$ can be due to the noise: $P_n(w_j(k, l)) = \text{Prob}(W > |w_j(k, l)|)$. For Gaussian noise, we have:

$$\begin{aligned} P_n(w_j(k, l)) &= \frac{2}{\sqrt{2\pi}\sigma_j} \int_{|w_j(k, l)|}^{+\infty} \exp(-W^2/2\sigma_j^2) dW \\ &= \text{erfc} \left(\frac{|w_j(k, l)|}{\sqrt{2}\sigma_j} \right) \end{aligned} \quad (18)$$

and Eqs. (17) becomes (19)

$$h(w_j(k, l)) = \frac{1}{\sigma_j^2} \int_0^{|w_j(k, l)|} u \text{erfc} \left(\frac{|w_j(k, l)| - u}{\sqrt{2}\sigma_j} \right) du. \quad (19)$$

The NOISE-MSE is very close to the l_1 norm (i.e. absolute value of the wavelet coefficient) when the coefficient value is large, which is known to produce good results for the analysis of piecewise smooth images (Donoho & Elad 2003). In the case of non-Gaussian or spatially inhomogeneous noise, our implementation will take a noise map as input. Then we can easily derive its entropy using the above definition.

3. Signal and noise information: the SZ component obtained by blind separation is swamped by noise. The following algorithm assumes that the observed map can be decomposed as:

$$S_{\text{obs}} = S + N. \quad (20)$$

Then, we can decompose the information contained in our image in two components, the first one (H_s) corresponding to the non-corrupted part, and the other one (H_n) describing a component which contains no information for us:

$$H(S_{\text{obs}}(k, l)) = H_s(S_{\text{obs}}(k, l)) + H_n(S_{\text{obs}}(k, l)). \quad (21)$$

For each wavelet coefficient $w_j(k, l)$, we have to estimate the fractions h_n and h_s of h :

$$H(S_{\text{obs}}(k, l)) = \sum_{j=1}^J \sum_{k,l=1}^n h_s(w_j(k, l)) + \sum_{j=1}^J \sum_{k,l=1}^n h_n(w_j(k, l)). \quad (22)$$

5.2.3. Multiscale Entropy Filtering

1. Filtering: the problem of filtering the observed map S_{obs} can be expressed as follows. We look for a filtered map S_f such that the difference between S_f and S_{obs} minimizes the information due to the signal (to recover all the signal) and such that S_f minimizes the information due to the noise (we want no noise). These two requirements are competing. A tradeoff is necessary, because, on the one hand, we want to remove all the noise (heavy filtering) and on the other hand, we want to recover the signal with fidelity. In practice, we minimize for each wavelet coefficient $w_j(k, l)$:

$$l(\tilde{w}_j(k, l)) = h_s(w_j(k, l) - \tilde{w}_j(k, l)) + \beta h_n(\tilde{w}_j(k, l))$$

where $w_j(k, l)$ are the wavelet coefficients of the observed map S_{obs} , $\tilde{w}_j(k, l)$ the wavelet coefficients of the filtered map S_f and β is the so-called regularization (trade-off) parameter.

2. Selecting significant Wavelet coefficients: whatever the filtering, the signal is always substantially modified. We want to fully reconstruct significant structures, without imposing strong regularization while eliminating the noise efficiently. The introduction of the multiresolution support (Murtagh et al. 1995), helps to do so. It applies the previous regularization (i.e. filtering) only on the non-significant (noisy) wavelet coefficients (Pantin & Starck 1996). Thus the choice of the regularization parameter is not a critical point. The other components of the maps are left untouched. The new Multiscale Entropy becomes:

$$\tilde{h}(w_j(k, l)) = \bar{M}(j, k, l) h(w_j(k, l)) \quad (23)$$

where $\bar{M}(j, k, l) = 1 - M(j, k, l)$, and M is the “multiresolution support” defined as:

$$M(j, k, l) = \begin{cases} 1 & \text{if } w_j(k, l) \text{ is significant} \\ 0 & \text{if } w_j(k, l) \text{ is not significant} \end{cases} \quad (24)$$

M describes, in a Boolean way, whether the data contains information at a given scale j and at a given position (k, l) . $w_j(k, l)$ is said to be significant if the probability that the wavelet coefficient is due to noise is small. In the case of Gaussian noise, a coefficient $w_j(k, l)$ is significant if

$|w_j(k, l)| > k\sigma_j$, where σ_j is the noise standard deviation at scale j , and k is a constant. Without an objective method for selecting the threshold, it is adjusted arbitrarily, generally taken between 3 and 5 (Murtagh et al. 1995).

3. Selecting significant Wavelet coefficients using the FDR: the False Discovery Rate (FDR) is a new statistical procedure by Benjamini & Hochberg (1995) which offers an effective way to select an adaptative threshold to compute the multiresolution support. This technique has been described by Miller et al. (2001); Hopkins et al. (2002); Starck et al. (2006) with several examples of astrophysical applications. The FDR procedure provides the means to adaptively control the fraction of false discoveries over total discoveries. The FDR is given by the ratio (25), that is, the proportion of declared active pixels which are false positives:

$$\mathcal{FDR} = \frac{V_{\text{ia}}}{D_a}. \quad (25)$$

where V_{ia} is the number of pixels truly inactive declared active, and D_a is the number of pixels declared active. The FDR formalism ensures that, *on average*, the False Discovery Rate is no larger than α which lies between 0 and 1. This procedure guarantees control over the FDR in the sense that:

$$\mathcal{E}(\mathcal{FDR}) \leq \frac{T_i}{V} \alpha \leq \alpha. \quad (26)$$

The unknown factor $\frac{T_i}{V}$ is the proportion of truly inactive pixels. A complete description of the FDR method can be found in Miller et al. (2001). In Hopkins et al. (2002) and Starck et al. (2006), it has been shown that the FDR outperforms standard method for source detection. In this application, we use the FDR method in a multiresolution framework (see Starck et al. 2006). We select a detection threshold \mathcal{T}_j for each scale. The different values of α at each scale were set as in Starck et al. (2006) and left untouched during our work: $\alpha_j = \alpha_0 * 2^j$ where $\alpha_0 = 0.0125$.

A wavelet coefficient $w_j(k, l)$ is considered significant if its absolute value is larger than \mathcal{T}_j as seen below.

4. Multiscale Entropy Filtering algorithm: assuming Gaussian noise, the Multiscale Entropy restoration method reduces to finding the image S_f that minimizes $J(S_f)$, given the map S_{obs} output of source separation with:

$$J(S_f) = \frac{\|S_{\text{obs}} - S_f\|^2}{2\sigma_n^2} + \beta \sum_{j=1}^J \sum_{k,l} \tilde{h}_N((\mathcal{W}S_f)_{j,k,l}) \quad (27)$$

where σ_n is the noise standard deviation in S_{obs} , J is the number of Wavelet scales, \mathcal{W} is the Wavelet Transform operator and $\tilde{h}_N(w_{j,k,l})$ is the multiscale entropy only defined for non-significant coefficients (outside the support selected by the FDR thresholding). Full details of the minimization algorithm can be found in Starck et al. (2001).

The results presented in the next section are obtained on a SZ map with a uniform Gaussian white noise but the method still holds for a non-uniform Gaussian noise over the map.

6. Cluster detection

6.1. Extraction algorithm – SExtractor

We use a public source software for extraction of SZ-Cluster candidates from the filtered maps: SExtractor (Bertin 2003). SExtractor turned out to be fast, convenient, and easy to configure. It handles conveniently Cluster Extraction, with observed

sizes ranging from a few pixels up to angular diameters of a degree. We used the *Noiseless* configuration, assuming that our filters are efficient enough for this assumption to be valid. We mainly use the source identification capability and the *deblending* algorithms that be very useful when extracting large-mass clusters. In order to recover as many candidates as possible, we set `DEBLEND_MINCONT` = 0 and `DETECT_MINAREA` = 5 (the smallest cluster size after optical beam dilution is over 9 pixels). Photometry of Cluster candidates is done using SExtractor's `FLUX_AUTO` Mode. SExtractor photometry performs very well on our data: the recovery error on integrated Compton flux Y is smaller than 4 percent, for Y larger than 2.5×10^{-10} sr (id 3×10^{-3} arcmin²). Thus SExtractor outputs a catalog of Cluster candidates with their position, flux and size. This catalog is likely to be contaminated by false detections due to residual noise, or spurious point sources that have survived the Source Separation step. While processing observed data, we will have to live with this contamination. Using simulations, we quantify in the following the SZ-Cluster selection function (Completeness) and this contamination.

6.2. Completeness and Purity of the recovered catalog

In order to identify true SZ Clusters from contamination in our catalog, we used an association criterion: for each candidate cluster, we scan the generated cluster catalog (see Sect. 2.4) for the closest cluster and store the angular distance between the two. Doing so, we observed a distribution with a strong peak at a distance smaller than 4 pixels (3.5 arcmin), corresponding to the true detected clusters, and a long flat tail corresponding to random associations. We then decided to tag candidates as true when their distance to the closest cluster in the simulated catalog is less than 4 pixels, neglecting the small fraction of random associations passing this test. This criterion can and will be enhanced using an improved distance involving recovered cluster flux and spatial distance.

Purity is defined as the ratio of true detections to the total number of detections:

$$\text{purity} = \frac{\text{true detections}}{\text{total detections}}. \quad (28)$$

Completeness is defined as the ratio of true detections to the total number of clusters in the simulation:

$$\text{completeness} = \frac{\text{true detections}}{\text{Number of simulated clusters}}. \quad (29)$$

Completeness and purity are the main performance criteria of a detection chain and should be maximised. We will use them to compare detection chains in the results section.

7. Results

In the following, we will quantify the performance of our algorithm chain in two steps. We will focus on the SZ map reconstruction: first the source separation step with JADE and then the additional FDR-MultiScale Entropy Method (ME-FDR) filtering step. The ME-FDR filter will be compared first to the classical Gaussian and Wiener filters. Then we will discuss assumptions and performances of our method relative to the one published in Pierpaoli et al. (2005).

The comparison between two methods for astrophysical image reconstruction can be performed in several ways. One may initially investigate the characteristics of residuals. This technique is typically used by mathematicians. An astrophysicist

may be more inclined to assess how well the objects of interest (SZ cluster of galaxies, in this case) are recovered: how many of them are found, and with what precision. Finally, a cosmologist would want to know to what extent the reconstruction procedure limits the ability to estimate cosmological parameters. This implies consideration of the purity and completeness as functions of cluster mass threshold, as well as precision in the reconstruction of the (central or integrated) Compton parameter for such masses. In what follows, we will compare the map's reconstruction error. Then we will focus on the SZ Cluster detection performance of the full detection chain, by comparing the recovered source catalogs to the Simulated SZ cluster catalog, in terms of purity and Completeness as defined in Sect. 6.2. This allows a parameter-free quantification of performances and trade-offs involved in cluster detection, being mostly independent of photometry details. Note also that the cluster detection efficiency does not convey information on the spread of the input/output y parameter. A more detailed description of the selection function, contamination, a discussion on photometry issues and methods and of cosmological implications will be given in a following paper (Juin et al. 2005).

7.1. JADE

Figure 4 shows the recovered SZ map, after source separation. JADE performs very well on our maps: the SZ signal that subdominant at all scales now appears clearly. No obvious residuals from other astrophysical sources are seen. JADE assumes and requires that input data have zero mean. This point is easy to meet in bolometer experiments since our bolometer cameras measure flux *variations* while the telescope scans the sky.

Second, JADE loses calibration information while processing data. Sky components are separated, but calibration of each output map has to be restored. It is of common observational procedure to calibrate a survey on the brightest sources in the field (which can be observed in follow up experiments). We chose in the following to calibrate our SZ maps using the 100 brightest clusters in the field. We average their recovered integrated flux (SExtractor `FLUX_AUTO` mode), and scale the map to match the average integrated flux of the 100 brightest simulated clusters. In the following, before computing statistical tests, all the maps have been normalised in this way. We will suppress this feature by taking into account prior knowledge (the optical spectral dependance of the SZ component) in future work.

Third, it is very important to minimise noise in the observed maps *before* source separation (JADE). JADE was designed to run on noiseless data. And it is quite sensitive to noise. We chose to apply to the four observed maps, before JADE, a simple Gaussian filter with the widths of the optical beams. If the noise level is not reduced before JADE, then the recovered mixing matrix will be inaccurate, and the SZ Cluster map will be polluted by the remains of the other astrophysical components, inducing efficiency losses. Once the mixing matrix has been carefully estimated, one can apply it to the unfiltered observed maps to extract the observed SZ cluster map, and then apply an optimised filter and cluster detection algorithms to extract the cluster catalog.

We characterize in our simulations the residual error on the estimated SZ map after component separation. The computed error map has a variance of 0.888, a skewness of 0.0019, a kurtosis of 0.0156 and is essentially white noise with no obvious visible structures. The foreground contribution to the noise map has a variance of 0.00273 (skewness of 0.0019, kurtosis of 0.0156) and shows that the component separation is *very* efficient on our simulated data, and strengthens the assumption of

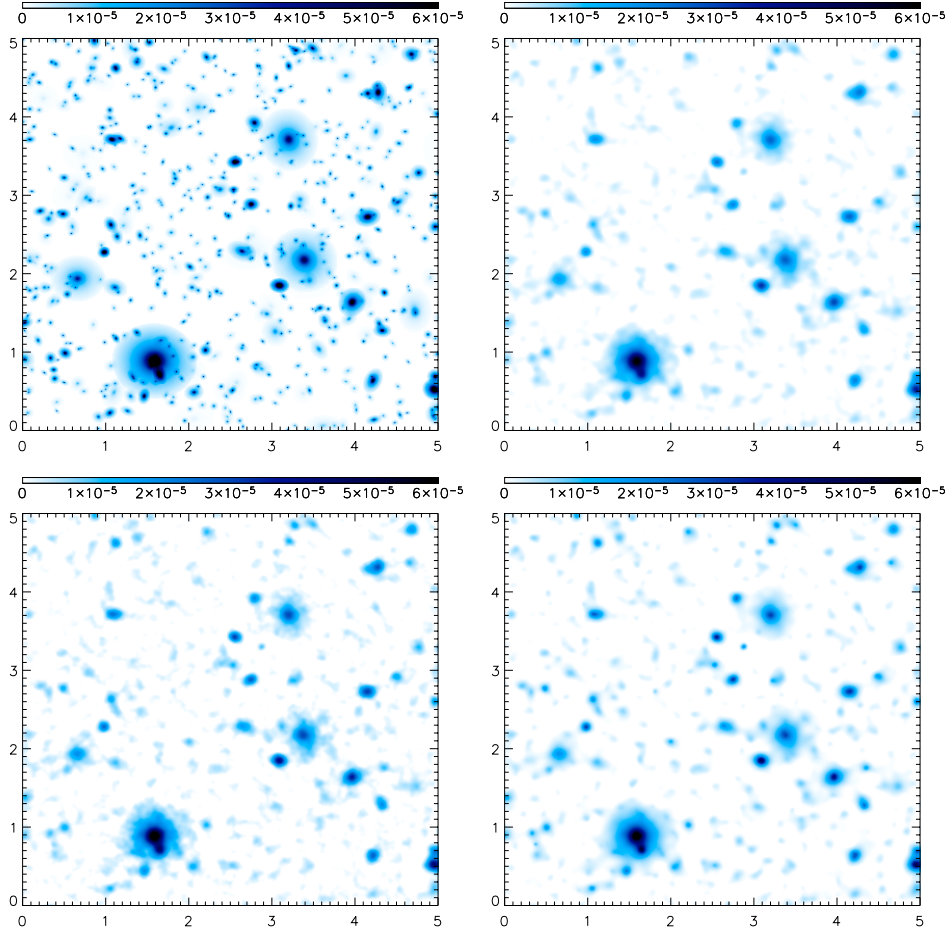


Fig. 5. Input SZ cluster map (*upper left*) map, and the 3 maps as recovered by Completeness and (*upper right*) Gaussian filtering $\sigma = 2.5 \text{ arcmin}^2$, *bottom left*, Wiener filtering and *bottom right* ME–FDR method. We chose to plot 25 deg^2 maps, to point out differences that do not appear on 400 deg^2 maps.

Gaussian noise made in the subsequent non-linear filtering and data processing steps. Should the experimental noise be significantly non-Gaussian, the presented detection chain could still be used. The price to pay will be a lack of statistical efficiency. The overall performance of the detection chain has to be assessed by means of simulations as shown in our paper.

7.2. Filtering method performances

Additional filtering is then applied to recover smaller clusters. We now quantify performances of the ME–FDR filter relative to the simpler filtering methods (Gaussian and Wiener). We chose to show results obtained with maps processed by JADE running in wavelet space, our best method. We will quantify the results by computing an error map, and its properties. Then we will compare the output of the extraction procedures using the three filters.

7.2.1. SZ map reconstruction

Figure 5 shows the maps after two classical filtering methods and our FDR multiscale entropy method. All maps look similar in nominal noise conditions of such ambitious experiments as future SZ Cluster survey, only statistical tests can discriminate between the 3 filtering methods. We computed the error map (filtered map, input map subtracted) for each filtered option. Computing the rms of the error maps divided by the rms

of the simulated SZ map (dimensionless rms) leads to $\sigma_{\text{Gauss}} = 0.617$, $\sigma_{\text{Wiener}} = 0.602$ and $\sigma_{\text{FDR}} = 0.570$. The CPU time on a 1024×1024 pixels map was $t_{\text{Gauss}} = 4 \text{ s}$, $t_{\text{Wiener}} = 26 \text{ s}$, and $t_{\text{FDR}} = 240 \text{ s}$. Thus our filtering algorithm is reasonably fast.

A more accurate test is to plot (see Fig. 6) the dimensionless standard deviation of error maps (filtered map, input map subtracted) at each scale of the wavelet decomposition:

$$\sigma_L = \frac{\sigma_{\text{Err},L}}{\sigma_{\text{Sim},L}} \quad (30)$$

where σ_L , $\sigma_{\text{Err},L}$ and $\sigma_{\text{Sim},L}$ are the standard deviation of the error map and the input SZ map, selecting the scale L of our wavelet transform. ME–FDR method performs better than simpler filters at all scales. Running our algorithms while changing the noise level showed that the higher the noise level, the larger the advantage in using our ME–FDR filtering compared to simple Gaussian, or Wiener filters.

7.2.2. Full extraction chain

Then considering our goal of detecting clusters, the relevant test is to compare the recovered catalog to the simulated cluster catalog, the input of the simulated map. Observed cluster catalogs are extracted and processed as explained in Sect. 6.2. In a future paper, we will present the selection function, contamination, a discussion on photometry issues and constraints on cosmology. In this paper where we discuss detection procedures the

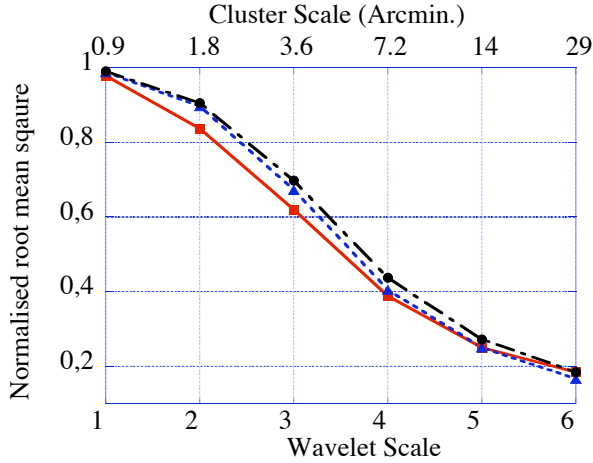


Fig. 6. Dimensionless standard deviation of error maps (filtered map, input SZ map subtracted) at each scale of the wavelet decomposition, for the three studied filters options. ME-FDR filtering method (red squares) is a significant improvement compared to simpler filtering methods (Gaussian, black dots and Wiener, blue triangles).

relevant information is the curve of purity versus completeness while raising the threshold on the cluster observed flux (Fig. 7). As expected, the higher the threshold, the higher the purity, and the lower the completeness. We see once more that ME-FDR outperforms the other filter options, allowing greater completeness at any required level of purity.

7.3. Comparison with another Wavelet-based methodology

In what follows, we will compare the method described above with the one presented in Pierpaoli et al. (2005). We will compare the general performance of the two methods in reconstructing the SZ images and then assess the global purity and completeness, as defined in Sect. 6.2. As we are not assessing here the goodness of the Y parameter reconstruction, a full comparison with the Pierpaoli et al. (2005) results (i.e. purity and completeness as a function of cluster mass threshold, spread of the input/output Y parameter relation) is not possible at this time.

We will first describe the Pierpaoli et al. (2005) method, and then perform the comparison according to the mentioned criteria.

7.3.1. Description of the alternative image reconstruction method

The method presented in Pierpaoli et al. (2005) is formally a *non-blind* component separation that has been optimized to recover galaxy clusters. In this method component separation and the deconvolution of the beam effect are done in one step by computing the Bayes Least Square estimate under a Gaussian scale mixture model of “neighborhoods” of wavelet coefficients. These “neighborhoods” are sets of wavelet coefficients associated with the same location and that behave in a coherent manner. The optical frequency dependency of each component, the beam size and the noise level for each observation are assumed to be known. The procedure relies on the possibility to discriminate between the different components (CMB, clusters, infra-red point sources and galaxy dust) by modeling the joint probability of these neighborhoods by Gaussian scale mixtures (see Pierpaoli et al. 2005 for details). A Gaussian scale mixture is a random vector $x = \sqrt{z}u$, where u is a Gaussian

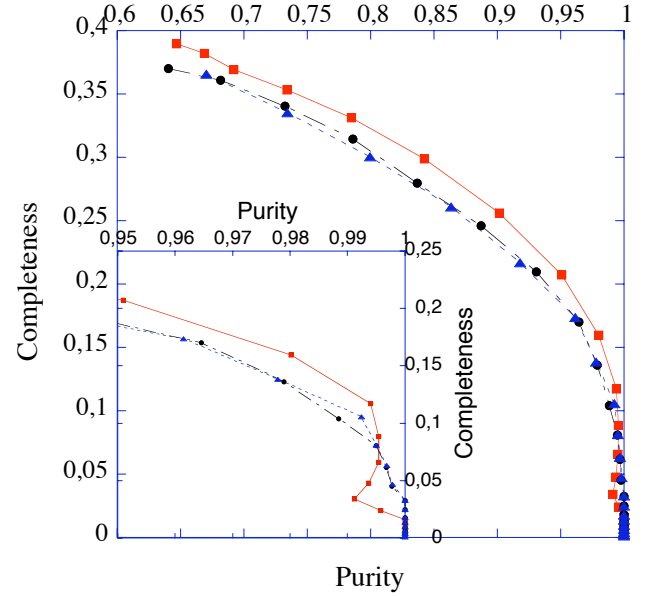


Fig. 7. Purity versus completeness for the three filtering options presented in this paper. Completeness + ME-FDR (red squares) outperforms the other filter options Gauss (black dots) and Wiener (blue triangles), allowing more completeness at any required level of purity. In the example presented here we simulated 9942 clusters according to Λ CDM distribution, with a threshold value on Y_{Comp} of 2×10^{-12} st; note that the completeness is normalized to this total number of simulated clusters. The insert is a zoom on the high purity region of the graph.

vector independent of the scalar random variable z – allowed to be non-Gaussian.

The model is determined by the probability distribution p_z and the covariance matrices of u . The covariance matrices have to be adjusted at each scale of the wavelet decomposition as a function of the resolution of the observed maps. They are computed prior to running the algorithm on simulated maps of each component at the correct resolution. The probability distribution p_z is also component-dependent. Pierpaoli et al. (2005) investigate several different possibilities for p_z , showing that different distributions lead to very different results in the mass reconstruction. For example, the Gaussian assumption (*Pier1* in the following) is most able to recover the maximum number of clusters but is inefficient in reconstructing the appropriate central intensity of these clusters; a non-Gaussian distribution modeled on the simulated SZ map (*Pier2*) is most able to recover the input Comptonisation parameter for bright clusters while missing the reconstruction of very low intensity ones.

In Pierpaoli et al. (2005) the focus is on reconstructing the signal of the most massive clusters, since those are the ones which are likely to lead to the most precise constraints on cosmological parameters. To this aim, the distribution *Pier2* is preferable to *Pier1*. As this paper is more focused on cluster detection efficiency (also including low-intensity clusters) and disregards the precision of the reconstruction, we will consider here the Gaussian prior for p_z (*Pier1* in what follows). A Gaussian prior for p_z reduces the estimator to a local Wiener filtering in Wavelet space, as all information about the signal non-Gaussianity is lost. The result of Pierpaoli et al. (2005) is a set of beam-deconvolved maps (one for each physical component considered) which can be directly compared with the input ones, or with other method’s. The cluster’s y map is used here for comparison. As for cluster detection, we will present results obtained with the code used in

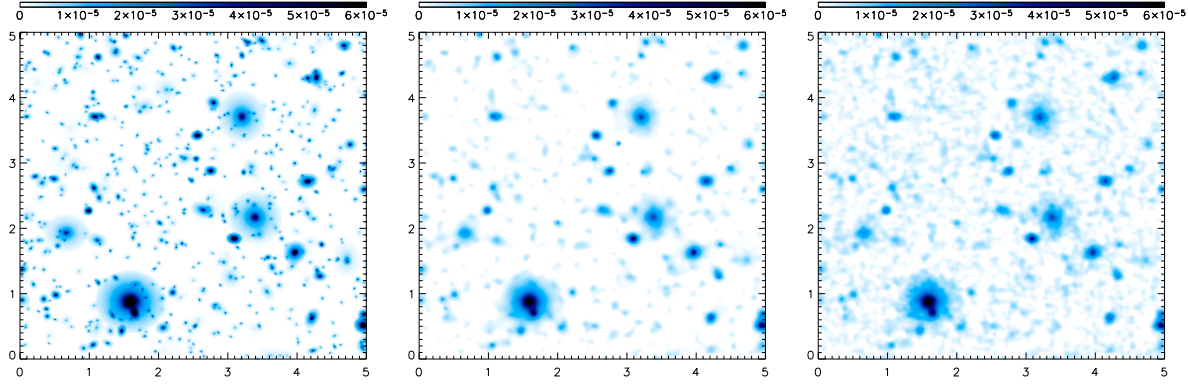


Fig. 8. Maps computed by two of the algorithms (25 deg² size). Left is the input SZ Cluster map. Centered is the map output of completeness after ME-FDR filtering. Right is the map output of the Pierpaoli et al. method, using the Gaussian assumption for the probability distribution.

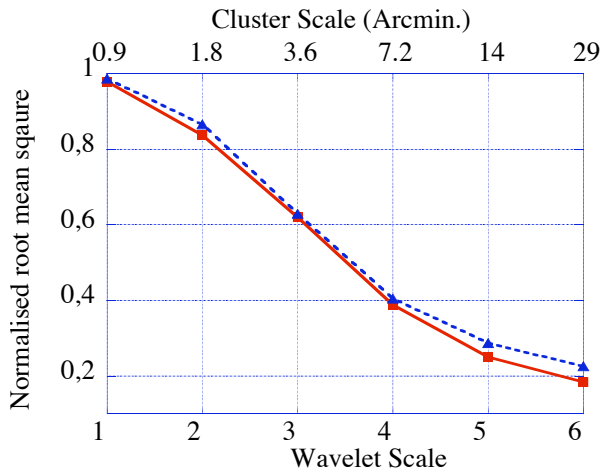


Fig. 9. Standard deviation versus scale for SZ cluster maps recovered by the completeness ME-FDR method (red squares) and Pierpaoli et al. *Pier1* method (blue triangles). The top axis gives the wavelet's typical scale (in arcmin) corresponding to the index on the bottom axis.

Pierpaoli et al. (2005), as well as the SExtractor code adopted here.

7.3.2. Map comparison

In Fig. 8 we present the maps that will be used for the comparison. While both maps recover well the intense clusters, the *Pier1* processed map shows more low-intensity structure than the ME-FDR map. This could be due to a better reconstruction of low-intensity clusters, as well as undesirable noise. Computing the *Pier1* error map rms, we find $\sigma_{Pier1} = 0.589$, which should be compared ME-FDR. An analysis of the residuals' normalised standard deviation for different wavelet scales (shown in Fig. 9) shows that the two methods mainly differ at angular scales equal and above 14 arcmin. The major contribution to the difference in the total rms is therefore not associated with the typical (low mass) cluster scale, but to much larger ones.

7.3.3. Detection efficiency

We compare here the detection efficiency defined in Sect. 6.2. For consistency with our detection chain (see Sect. 7.1), we recalibrate the Pierpaoli maps so that the average integrated Compton flux of the 100 brightest clusters matches the value observed in the input SZ cluster map (small correction). Figure 10

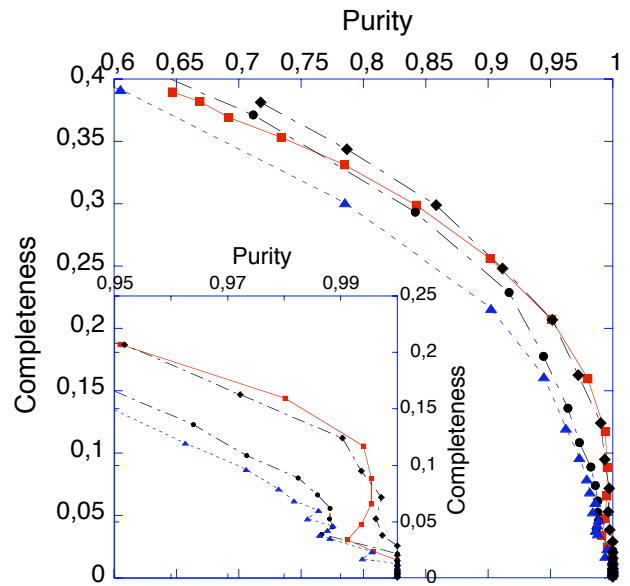


Fig. 10. Purity versus completeness curves for the four software detection chains studied in the following. We use red squares for Completeness-FDR+SExtractor, black dots for *Pier1* and blue triangles for *Pier2* with peak detection, black diamonds for *Pier1* using SExtractor for source detection. The inserted plot is a zoom on the high purity region of the graph.

presents the curves of completeness versus purity for completeness FDR-ME + SExtractor, *Pier1*, *Pier2* using the peak finding algorithm presented in Pierpaoli et al. (2005) to detect clusters and the *Pier1* method using SExtractor (PierNew). As expected, *Pier2* is not as good as the other methods at all purity, since the probability p_z used here is optimized for accurate recovery of the most massive cluster's central intensity and not for cluster detection. *Pier1* (the Gaussian distribution) improves *Pier2* results, especially when SExtractor is used to detect the clusters (PierNew). At low threshold (low purity) PierNew recovers more clusters than ME-FDR. The low-intensity structures visible in Fig. 8 contain a sizable number of clusters. As completeness FDR-ME is designed to filter out noise better, it also filters out some low-intensity clusters which then no longer contribute to the detection rate. At high purities, PierNew and completeness FDR-ME provide equivalent performance, given the statistical uncertainty of the small number of clusters involved in the high-threshold cut.

8. Conclusion

We simulated observed sky maps at frequencies of a typical large multiband bolometric SZ-Cluster survey. This simulation includes the main features of the “background”: point sources with varying spectral index and brightest sky components. We implemented a complete software detection chain, working in 3 steps. First we use a source separation algorithm that is based on a Wavelet transform and the JADE-ICA algorithm. Then, we filter the SZ-Cluster map using an FDR-Multiscale Entropy method. Finally, we detect the clusters on the filtered maps using the SExtractor software in a *noiseless* configuration. Despite several simplifying assumptions detailed in the paper, our detection chain proves to be very efficient, yielding 25% of clusters of mass larger than $10^{14} M_{\odot}$ detected with 90% purity. We compare our detection algorithm to a previously published method based on wavelets (Pierpaoli et al. 2005). We restrict our comparison criteria to the global detection efficiency, as defined in Sect. 6.2. We find a detection efficiency in the high-purity region comparable to the Gaussian probability case (*Pier1* method), which is the model that provides the best performance for this comparison among those presented in Pierpaoli et al. (2005). The ME-FDR detection efficiency slightly degrades at lower cluster intensity than the *Pier1* method, as ME-FDR filters out low-intensity clusters during the denoising procedure. These results, however, are impressive as ME-FDR, unlike Pierpaoli et al. (2005) methods, is a *blind* algorithm that makes no assumption on the physical properties of the signal to be recovered.

There remain few open issues concerning the use of the proposed detection chain to process upcoming real observations. Once configured, the detection chain was used without “close supervision”. Also, the wavelet transform used jointly with JADE was chosen a priori. JADE as well as the ME-FDR filter were run as black box algorithms, while the SExtractor parameters were set to fairly standard values, since the input data maps were pre-filtered. We also left out the smallest scales in the ME-FDR maps to reduce high frequency spatial noise. If the experimental data are not as clean as in our simulations, how will these simplifications affect the result?

If the noise level is increased or if the noise displays strong “striping” (as a result of low frequency noise in the bolometer timelines), JADE is no longer appropriate for the component separation step. Hence the experimental data processing will have to include a good “destriper” in the map-making calculations. The proposed ME-FDR filter proved to be more robust than the Gaussian or Wiener filter, when the noise level is increased. Our last concern is that the Galactic Dust Maps in our simulations are generated as spatially Gaussian, which is not true of observed “Galactic dust clouds”. But the optical spectrum of Galactic Dust being very different from SZ clusters, we foresee that ICA will separate these features efficiently.

We will focus on improving the source separation step. We will include optical spectrum knowledge in the source separation algorithms. This will allow us to overcome the intrinsic scaling indeterminacy of the blind linear mixture model and so prevent loosing track of map calibration as in Sect. 7.1. Additionally, JADE must run on noise-free data. We will design our source separation algorithm to handle noisy data. We studied the performance of our method in terms of detection efficiency. A further assessment in terms of reconstruction accuracy and photometry is necessary since these are valuable for cosmology. Of course detection efficiency doesn’t provide all the information we want to know in order to do cosmology, as the accuracy of the reconstruction and photometry issues are also important. In a

following paper (Juin et al. 2005), we will use these algorithms to compute selection functions, contaminations and constraints on cosmology foreseen for the upcoming bolometric SZ-cluster surveys.

Acknowledgements. The authors wish to thank A. Réfrégier, R. Teyssier, J. Rich, C. Magneville, J. P. Pansart, J. L. Starck (CEA/DAPNIA), J. B. Melin and J. Bartlett (Univ.P7, APC), E. Bertin (IAP), for many private discussions and the reviewer for helpful comments on an earlier version. E.P. is an ADVANCE fellow (NSF grant AST-0340648), also supported by NASA grant NAG5-11489.

References

- AMIBA 2005–2007, AMIBA experiment, see: <http://amiba.asiaa.sinica.edu.tw/>
- Antonini, M., Barlaud, M., Mathieu, P., & Daubechies, I. 1992, Vol. I Benjamini, Y., & Hochberg, Y. 1995, J. R. Stat. Soc. B, 57, 289
- Bertin, E. 2003, <http://terapix.iap.fr/>
- Bobin, J., Moudden, Y., Starck, J.-L., & Elad, M. 2005, in SPARS’05, Signal Processing with Adaptive Sparse Structured Representations, Vol. 5914, Rennes, France
- Borys, C., Chapman, S. C., Halpern, M., & Scott, D. 2003, MNRAS, 344, 385
- Bouchet, F. R., & Gispert, R. 1999, New Astron., 4, 443 [arXiv:astro-ph/9903176]
- Cardoso, J.-F. 1998, Proc. IEEE, Special issue on blind identification and estimation, 9, 2009
- Cardoso, J. F. 1999, Neural Computation, 11, 157
- Cardoso, J. F. 2001, in Proc. ICA 2001, San Diego
- Contaldi, C. R., Bond, J. R., Pogosyan, D., et al. 2002, Proc. XVIII IAP Coll. On the nature of dark energy, Paris, 1–5 July 2002 [arXiv:astro-ph/0210303]
- Cooray, A. R. 2000, MNRAS
- Delabrouille, J., Cardoso, J.-F., & Patanchon, G. 2003, MNRAS, 346, 1089
- Donoho, D. L., & Elad, M. 2003, Proc. Nat. Aca. Sci., 100, 2197
- Doran, M., & Mueller, C. M. 2004, Journal of Cosmology and Astroparticle Physics, 0409, 003 [arXiv:astro-ph/0311311]
- Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, ApJ, 633, 560
- Georgiev, P. G., Theis, F., & Cichocki, A. 2005, IEEE Transactions on Neural Networks, 16, 992
- Gull, S., & Skilling, J. 1991, MEMSYS5 Quantified Maximum Entropy User’s Manual, Royston, England
- Halverson, N. W., Leitch, E. M., Pryke, C., et al. 2002, ApJ, 568, 38
- Hanany, S., Ade, P., Balbi, A., et al. 2000, ApJ, 545, 15
- Herranz, D., Sanz, J., Hobson, M. P., et al. 2002a, MNRAS, 336, 1057
- Herranz, D., Sanz, J. L., Barreiro, R. B., & Martínez-González, E. 2002b, ApJ, 580, 610
- Hinshaw, G., Spergel, D. N., Bennett, C. L., et al. 2003, ApJS, 148, 97
- Hopkins, A. M., Miller, C. J., Connolly, A. J., et al. 2002, AJ, 123, 1086
- Hyvärinen, A., Karhunen, J., & Oja, E. 2001, Independent Component Analysis (New York: John Wiley), 481
- Jones, M. E. 2001, MNRAS, in 8th Taipei Astrophysics Workshop AMiBA 2001
- Juin, J. B., Yvon, D., & Réfrégier, A. 2005, A&A, submitted [arXiv:astro-ph/0512378]
- Knop, R. A., Aldering, G., Amanullah, R., et al. 2003, ApJ, 598, 102
- Kuruoglu, E., Bedini, L., Paratore, M., Salerno, E., & Tonazzini, A. 2003, Neural Networks, 16, 479
- Lamarre, J., Puget, J., Bouchet, F., et al. 2004, New Astronomy Reviews, Proc. of the workshop on The Cosmic Microwave Background and its Polarization, ed. S. Hanany, & R. A. Olive
- Lee, T.-W., Lewicki, M., Girolami, M., & Sejnowski, T. 1999, IEEE Signal Processing Letters, 6, 87
- Lewicki, M. S., & Sejnowski, T. J. 2000, Neural Computation, 12, 337
- Maino, D., Farusi, A., Baccigalupi, C., Perrotta, F., & Banday, A. 2002, MNRAS, 334, 53
- Maisinger, K., Hobson, M. P., & Lasenby, A. N. 2004, MNRAS, 347, 339
- Marshall, P. J., Hobson, M. P., Gull, S. F., & Bridle, S. L. 2002, MNRAS, 335, 1037
- Masi, S., Ade, P., de Bernardis, P., et al. 2003, Mem. Soc. Astron. It., 74, 96
- Melin, J.-B. 2004, Amas de galaxies et effet Sunyaev-Zel’dovich: observations et étude des effets de sélection des sondages, Univ. Denis Diderot, Paris VII
- Miller, C. J., Genovese, C., Nichol, R. C., et al. 2001, AJ, 112, 3492
- Moudden, Y., Cardoso, J.-F., Starck, J.-L., & Delabrouille, J. 2005, Eurasip Journal on Applied Signal Processing, 15, 2437
- Murtagh, F., Starck, J.-L., & Bijaoui, A. 1995, A&AS, 112, 179

- Netterfield, C., Ade, P., Bock, J., et al. 2002, *ApJ*, 571, 604
- Pantin, E., & Starck, J.-L. 1996, *A&AS*, 315, 575
- Peacock, J. 1999, *MNRAS*, 284, 885
- Peebles, P. 1993 (Princeton: Princeton Univ. Press)
- Pham, D.-T. 2001, *SIAM Journal on Matrix Analysis and Applications*, 22, 1136
- Pierpaoli, E., Anthoine, S., Huufenberger, K., & Daubechies, I. 2005, *MNRAS*, 359, 261
- Pierpaoli, E., Scott, D., & White, M. 2001, *MNRAS*, 325, 77
- Pierpaoli, E., Borgani, S., Scott, D., & White, M. 2003, *MNRAS*
- Pierre, M., Valtchanov, I., et al. 2004, *Journal of Cosmology and Astroparticle Physics*, 09, 11
- Refregier, A., & Massey, R. J. R. 2004, *AJ*, 127, 3102
- Riess, A. G., Filippenko, A. V., Li, W., & Schmidt, B. P. 1999, *AJ*, 118, 2668
- Schafer, B. 2004, *MNRAS*, submitted
- Sheth, R. K., Mo, H. J., & Tormen, G. 2001, *MNRAS*
- Starck, J.-L. & Murtagh, F. 2002, *Astronomical Image and Data Analysis* (Springer-Verlag)
- Starck, J.-L., Murtagh, F., Querre, P., & Bonnarel, F. 2001, *A&A*, 368, 730
- Starck, J.-L., Candes, E., & Donoho, D. 2003, *A&A*, 398, 785
- Starck, J.-L., Elad, M., & Donoho, D. 2004, *Advances in Imaging and Electron Physics*, 132
- Starck, J.-L., Moudden, Y., Bobin, J., Elad, M., & Donoho, D. 2005, in *SPIE Conference on Wavelets*, 5914, San Diego, CA, USA
- Starck, J.-L., Pires, S., & Refregier, A. 2006, *A&A*, 451, 1139
- Sunyaev, R., & Zeldovich, I. B. 1980, *ARA&A*, 18, 537
- Tristram, M., Patanchon, G., Macias-Perez, Y. F., et al. 2005, *A&A*, 436, 785
- White, M., Carlstrom, J. E., Dragovan, M., & Holzapfel, W. L. 1999, *ApJ*, 514, 12
- Yvon, D., & Mayet, F. 2005, *A&A*, 436, 729
- Zibulevsky, M., & Pearlmutter, B. 2001, *Neural-Computation*, 13, 863