Source Separation of Astrophysical Images Using Particle Filters

Mauro Costagli, Ercan E. Kuruoğlu, A. Ahmed
ISTI-CNR Via Moruzzi,1 - 56124 Pisa

November 2003

Abstract

The problem of separating a superposition of different, simultaneous signals from their mixture appears very frequently in various fields of engineering, such as speech processing, telecommunications, biomedical imaging and financial data analysis. In this work, we will confront the problem of source separation in the field of astrophysics, where the contributions of various Galactic and extra-Galactic components need to be separated from a set of observed noisy mixtures. Most of the previous work on the problem perform a blind separation, assume noiseless models, and in the few cases when noise is taken into account it is generally assumed to be Gaussian and space-invariant. Our objective is to study a novel technique named *particle filtering*, and implement it for the non-blind solution of the source separation problem. Particle filtering is an advanced Bayesian estimation method which can deal with non-Gaussian and nonlinear models, and additive space-varying noise, in the sense that it is a generalization of the Kalman Filter. In this work, particle filters are utilized with objectives of both noise filtering and separation of signals: this approach is extremely flexible, as it is possible to exploit the available a-priori information about the statistical properties of the sources through the Bayesian theory. Especially in case of low SNR, our simulations show that the output quality of the separated signals is better than that of ICA, which is one of the most widespread methods for source separation. On the other hand, since a wide set of parameters, which can take from a large range of values, has to be initialized, the use of this approach needs extensive experimentation and testing.

Keywords

Cosmic Microwave Background, Bayesian Source Separation, A-priori Information, Particle Filtering, Sequential Markov Chain Monte Carlo, Non-Gaussian Models, Non-Stationary Noise, Blind Source Separation, Independent Component Analysis.

1 Introduction

The signal separation problem is one of the most fundamental in engineering: in fact, it is of interest to many applications, such as crosstalk removal in multichannel communications and multipath channel identification. Also consider, for example, the need of separating the superposition of different, simultaneous audio and speech signals. Moreover, source separation techniques are also used in medical science for functional magnetic resonance imaging (fMRI), electromyograms (EMG), and magnetoencephalography (MEG), in financial science for the analysis of time series, and in the context of feature extraction.

In this work we will confront the problem of source separation in the field of astrophysics.

1.1 The Astrophysical Problem

The European Space Agency (ESA) will launch a satellite called Planck in the year 2007: the ultimate goal of this mission is to answer the most fundamental questions concerning the origins of our Universe. How did it form? How old is it? Will the Universe continue its expansion forever, or will it collapse into a *Big Crunch*?

Answers to these (and other) questions will be found via the analysis of the microwave radiation coming from the vault of heaven. This signal is the superposition of many independent astrophysical sources, and the first important step in our analysis will be their separation: the *Cosmic Microwave Background* (CMB) is surely one of the most important sources to be analyzed, as high-resolution and high-sensivity measurements of its anisotropies would enable cosmologists to assess the validity of the present competing cosmological theories.

The signal measured in CMB experiments is however contaminated by several sources (Figure 1): together with the intrinsic noises due to the satellite microwave detectors, astrophysical contaminants (the so-called *foregrounds*) are present because of different phenomena. The most relevant ones are due to several different processes that occur in our Galaxy (the dust emission, synchrotron and free-free radiation), while other contaminations come from extragalactic microwave sources and from the so-called Sunyaev-Ze'ldovich effect. Before achieving cosmological information from the statistical analysis of the CMB anisotropies, all these components must be separated from the intrinsic CMB signal. The Planck satellite will provide nine different sky maps, at different frequencies between 30 GHz and 857 GHz, in order to have the oppurtunity of analysing nine different mixtures, which will be our only available observations: each of them is a different mixture of the above named sources, that is a superposition obtained through one of the nine different unknown mixing matrices. As also any other astrophysical source signal carries information about relevant phenomena, the most proper approach is to try to extract each source from the mixture maps by a source separation strategy, instead of looking just for a CMB map as clean as possible, through a noise cancellation one.

1.2 PCA, ICA and IFA

A widely used *blind* technique in signal processing is *Principal Component Analysis* (PCA), which is able to separate mixed variables, providing uncorrelated output signals, that is separation up to second order statistics ([11], [23]).

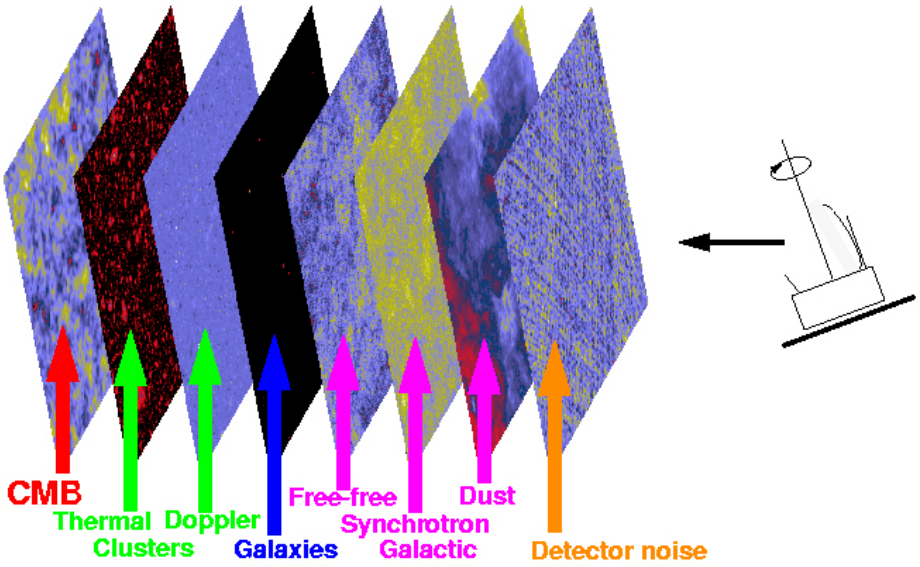

Figure 1: Introduction - The radiation observed by the satellite antenna is the result of the superposition of various astrophysical components ([24]).

Another classical approach, which performs a blind source separation, is *Independent Component Analysis* (ICA) [2], [9], [14], [39]. Assuming that all the source processes are statistically independent, ICA techniques operate in order to transform the set of observed data and obtain independent output signals, separated in all orders of statistics. Unfortunately this approach does not provide good results, because noise is not taken into account in the ICA model: in earlier experience on source separation in astrophysical images it has been observed that this algorithm has a significantly deteriorating performance when the noise level is increased [36].

Indeed, we possess some information about the statistical distributions of the sources and noise, so we are not forced to follow a completely blind approach, as this knowledge can be somehow exploited. A method that incorporates prior information about the sources is called *Independent Factor Analysis* (IFA), and has been introduced both in signal processing [37] and neural networks literature [7], and recently it has been studied in the context of simulated but realistic sky radiation maps, in order to identify its potentials and drawbacks [34].

1.3 Objective

In this work a different approach, named *Particle Filtering*, will be studied and implemented in order to solve the source separation problem in a different promising way. Our objective is to provide a new, robust, and elaborate solution to the astrophysical source separation problem, as up to now the rest of literature has considered almost only stationary models, in absence of noise. Particle Filters, also known as Sequential Monte Carlo Methods, were introduced for the first time in automatic control by Handschin and Mayne [22] at the end of the 60's, but these methods have been overlooked until the early 90's because of the low computational power available. The renewed interest in these methods was due also to the progress in methodology

described by Gordon et al. [21], Liu and Chen [35] and Pitt and Shephard [38].

Particle filtering is an extension of Kalman filtering, and it is general technique which can deal with non-Gaussian and nonlinear models: its theory is well covered in [1], but only basic and conventional particle filters have been presented so far, and none of them was intended to be used to separate astrophysical sources. The basic generic particle filter algorithm has been adapted and implemented in order to improve its performances to our purpose, using the a-priori information available about the mixing matrix and the statistical properties of the sources.

On the plus side, the particle filter is extremely flexible, and so very attractive, because one has the opportunity of using many a-priori information about the statistical properties of the sources as inputs; on the other hand, the use of this approach can be very tricky, as the user needs to set a lot of parameters which can take a large range of values, as it will be explained later. This is the reason why the use of this approach needs a lot of experimentation and testing.

The following section is about Particle Filters, from the description of the algorithm to its implementation, through the Bayesian theory. The third section contains a brief introduction to the astrophysical components that will be mentioned in this work, while the fourth one describes the simulations done to test the algorithm. The conclusions are presented in the last part of this work.

2 Particle Filtering

Filtering is the problem of estimating the hidden variables (called *states*) of a system, as a set of observations becomes available on-line: the introduction of a *state space* formulation is a fundamental step, because it allows to deal with non-stationarity, as it will be shown later.

In many real-world data analysis applications, prior knowledge about the unknown quantities to be estimated is available, and this information can be exploited to formulate Bayesian models: prior distributions for the unknown quantities and likelihood functions that relate these quantities to the observations. Then, all inference on the unknown quantities is based on the posterior distribution obtained from Bayes' theorem.

It is possible to express the model in terms of a *state equation* and an *observation equation*:

$$\boldsymbol{\alpha}_t = f_t(\boldsymbol{\alpha}_{t-1}, \mathbf{v}_t);$$

$$\mathbf{y}_t = h_t(\boldsymbol{\alpha}_t, \mathbf{w}_t).$$

The state equation evaluates the state sequence: $\boldsymbol{\alpha}_t$ is the state at current step t, f_t is a possibly nonlinear function, $\boldsymbol{\alpha}_{t-1}$ is the state at the previous step, and $\mathbf{v}_t$ is called dynamic noise process. The observation equation is characterized by a nonlinear function h_t, and both the current state $\boldsymbol{\alpha}_t$ and the observation noise realisation $\mathbf{w}_t$ at time step t are taken into account to generate the observation $\mathbf{y}_t$.

The *Kalman filter* (KF) is an extension of the Wiener filter, and it was presented by R. E. Kalman in 1961 [30]: this filter derives an exact analytical expression to compute the evolving sequence of the posterior distributions, when the data are modelled by a linear Gaussian state-space model. The obtained posterior density at every time step is Gaussian, hence parametrized by a mean and a covariance.

The best known algorithm that allows a non-Gaussian and nonlinear model is the *Extended Kalman filter* (EKF) [4], based upon the principle of linearising the measurements and evolution models using Taylor series expansions. Unfortunately, this procedure may lead to poor representations of both the non-linear functions and the probability distributions of interest, so the filter can diverge.

The more recent *Unscented Kalman Filter* (UKF) is founded on the intuition that it is better to approximate a Gaussian distribution, than approximating arbitrary non-linear functions [29]. Also this approach has, however, a limitation, that is it does not apply to general non-Gaussian distributions.

A new technique to solve the general filtering problem is introduced in this section: this approach, named *particle filtering*, uses sequential Monte Carlo methods, and it was introduced for the first time in automatic control by Handschin and Mayne [22] at the end of the 60's, but it has been overlooked until the early 90's because of the low computational power available. The renewed interest in these methods brought to success in tracking problems (see [6] for a general review), and very recently it has been applied also to perform source separation ([1], [5], [19]). Sequential Monte-Carlo particle filters are able to solve time or space varying mixing problems, and allow for a complete representation of the posterior distribution of the states, so that any statistical estimates (mean, variance, and so on...) can be computed.

2.1 The Bayesian Approach

This subsection is a brief overview of the Bayesian approach, which is the basis of Particle Filtering. Given a set of observations $\mathbf{y}$ and the set of unknown sources $\boldsymbol{\alpha}$, we consider the *posterior* distribution

$$p(\boldsymbol{\alpha}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})}{p(\mathbf{y})}$$

where

$$p(\mathbf{y}) = \int p(\mathbf{y}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})d\boldsymbol{\alpha}$$

and where $p(\mathbf{y}|\boldsymbol{\alpha})$ denotes the *likelihood* and $p(\boldsymbol{\alpha})$ denotes the *prior* distribution. In order to keep the the same notation used in literature [17], we use $\boldsymbol{\alpha}_t$ to denote both the random variable and its realisation. Consequently, we express continuous probability distributions using $p(d\boldsymbol{\alpha}_t)$ instead of $\Pr(\boldsymbol{\alpha}_t \in d\boldsymbol{\alpha}_t)$, and discrete distributions using $p(\boldsymbol{\alpha}_t)$ instead of $\Pr(\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_t)$.

Given the posterior distribution, optimum estimators can be obtained, most notably the *Minimum Mean Squared Error* (MMSE) and the *Maximum A Posteriori* (MAP) estimates of $\boldsymbol{\alpha}$:

$$\hat{\boldsymbol{\alpha}}_{MMSE} = \int \boldsymbol{\alpha} p(\boldsymbol{\alpha}|\mathbf{y})d\boldsymbol{\alpha};$$

$$\hat{\boldsymbol{\alpha}}_{MAP} = \arg\max_{\boldsymbol{\alpha}} p(\boldsymbol{\alpha}|\mathbf{y}).$$

The aforementioned filters (KF, EKF, UKF) rely on various assumptions to ensure mathematical tractability. Unfortunately, real data sets are often very complex, typically high dimensional, non-linear, nonstationary and non-Gaussian: except in some simple cases, the integration (MMSE) or the optimisation (MAP) of the posterior are not analitically tractable. Moreover, classical optimisation methods need good initialisations and are sensitive to local minima.

On-line simulation based *Sequential Monte Carlo* (SMC) methods are a set of simulation-based approaches which use variates from the posterior, and provide an attractive solution to compute the posterior distribution of interest: at each time step, the posterior distribution is approximated by a set of *particles* generated by an *importance distribution* $\pi(\boldsymbol{\alpha}|\mathbf{y})$, chosen such that it is easy to sample, and whose support is assumed to include that of $p(\boldsymbol{\alpha}|\mathbf{y})$, as shown in the next subsection.

2.2 Monte Carlo Particle Filters

Problem Statement

As stated before, we usually cannot obtain an analytic expression for the posterior distribution: this is the reason why we have to resort to stochastic simulation. The unobserved signal (hidden state) $\boldsymbol{\alpha}_t$ is modelled as a Markov process of initial distribution $p(\boldsymbol{\alpha}_0)$ and transition equation $p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1})$:

$$\begin{array}{rl} p(\boldsymbol{\alpha}_0) & \text{for } t = 0, \\ p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}) & \text{for } t = 1, 2, 3, \cdots. \end{array}$$

We denote by $\boldsymbol{\alpha}_{0:t} \stackrel{\triangle}{=} \{\boldsymbol{\alpha}_0, \cdots, \boldsymbol{\alpha}_t\}$ and $\mathbf{y}_{1:t} \stackrel{\triangle}{=} \{\mathbf{y}_1, \cdots, \mathbf{y}_t\}$ the signals and the observations respectively, up to step t.

Our objective is to estimate recursively in time the posterior distribution $p(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})$, its associated features (including the marginal distribution $p(\boldsymbol{\alpha}_t|\mathbf{y}_{1:t})$, known as the *filtering distribution*), and the expectations

$$I(f_t) = E_{p(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})}\{f_t(\boldsymbol{\alpha}_{0:t})\} \stackrel{\triangle}{=} \int f_t(\boldsymbol{\alpha}_{0:t}) p(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t}) \mathrm{d}\boldsymbol{\alpha}_{0:t}$$

for some function of interest, like the mean of the sources, or their covariance.

At any time t, the posterior distribution is given by Bayes' theorem:

$$p(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_{1:t}|\boldsymbol{\alpha}_{0:t}) p(\boldsymbol{\alpha}_{0:t})}{\int p(\mathbf{y}_{1:t}|\boldsymbol{\alpha}_{0:t}) p(\boldsymbol{\alpha}_{0:t}) \mathrm{d}\boldsymbol{\alpha}_{0:t}}.$$

A recursive formula for this joint distribution can be obtained as follows:

$$p(\boldsymbol{\alpha}_{0:t+1}|\mathbf{y}_{1:t+1}) = p(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t}) \frac{p(\mathbf{y}_{t+1}|\boldsymbol{\alpha}_{t+1}) p(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})}.$$

The marginal distribution $p(\boldsymbol{\alpha}_t|\mathbf{y}_{1:t})$ also satisfies the following recursive equations (prediction and update respectively):

$$\begin{aligned} p(\boldsymbol{\alpha}_t|\mathbf{y}_{1:t-1}) &= \int p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}) p(\boldsymbol{\alpha}_{t-1}|\mathbf{y}_{1:t-1}) \mathrm{d}\boldsymbol{\alpha}_{t-1}; \\ p(\boldsymbol{\alpha}_t|\mathbf{y}_{1:t}) &= \frac{p(\mathbf{y}_t|\boldsymbol{\alpha}_t) p(\boldsymbol{\alpha}_t|\mathbf{y}_{1:t-1})}{\int p(\mathbf{y}_t|\boldsymbol{\alpha}_t) p(\boldsymbol{\alpha}_t|\mathbf{y}_{1:t-1}) \mathrm{d}\boldsymbol{\alpha}_t}. \end{aligned}$$

Monte Carlo integration methods have the great advantage of not being subject to any linearity or Gaussianity constraints on the model, and they also have appealing convergence properties. The basic idea is that a large number of samples drawn from the required posterior distribution is sufficient to approximate the posterior distribution itself, and to approximate the integrals appearing in the "prediction and update" equations mentioned before.

Importance Sampling

Assume that $N >> 1$ random samples $\{\boldsymbol{\alpha}_{0:t}^{(i)};\ i = 1, \cdots, N\}$, called *particles* (hence the term *particle filters*), have been generated from the posterior $p(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})$: a Monte Carlo approximation of this function is thus given by:

$$p_N(d\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\boldsymbol{\alpha}_{0:t}^{(i)}}(d\boldsymbol{\alpha}_{0:t}),$$

where $\delta_{\boldsymbol{\alpha}_{0:t}^{(i)}}(d\boldsymbol{\alpha}_{0:t})$ denotes the delta-Dirac mass located in $\boldsymbol{\alpha}_{0:t}^{(i)}$. The following estimate of the function of interest $I(f_t)$ can be obtained straightforwardly by:

$$I_N(f_t) = \int f_t(\boldsymbol{\alpha}_{0:t}) p_N(d\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t}) = \sum_{i=1}^{N} f_t\left(\boldsymbol{\alpha}_{0:t}^{(i)}\right).$$

Unfortunately, it is usually impossible to sample efficiently from the posterior distribution at any step t, since it is, in general, multivariate, non-standard, and only known up to a proportionality constant. A classical solution consists of using the *importance sampling* method [20], which introduces an arbitrary *importance function* (also referred to as the *proposal distribution* or the *importance sampling distribution*) $\pi(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})$. Provided that the support of $\pi(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})$ includes the support of $p(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})$, we get the identity

$$I(f_t) = \frac{\int f_t(\boldsymbol{\alpha}_{0:t}) w(\boldsymbol{\alpha}_{0:t}) \pi(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t}) d\boldsymbol{\alpha}_{0:t}}{\int w(\boldsymbol{\alpha}_{0:t}) \pi(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t}) d\boldsymbol{\alpha}_{0:t}},$$

where $w(\boldsymbol{\alpha}_{0:t})$ is known as the *importance weight*:

$$w(\boldsymbol{\alpha}_{0:t}) = \frac{p(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})}{\pi(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})}.$$

Consequently, it is possible to obtain a Monte Carlo extimate of $I(f_t)$ using N particles $\{\boldsymbol{\alpha}_{0:t}^{(i)};\ i = 1, \cdots, N\}$ sampled from $\pi(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})$:

$$\bar{I}_N(f_t) = \frac{\frac{1}{N}\sum_{i=1}^{N} f_t\left(\boldsymbol{\alpha}_{0:t}^{(i)}\right) w\left(\boldsymbol{\alpha}_{0:t}^{(i)}\right)}{\frac{1}{N}\sum_{j=1}^{N} w\left(\boldsymbol{\alpha}_{0:t}^{(i)}\right)} = \sum_{i=1}^{N} f_t\left(\boldsymbol{\alpha}_{0:t}^{(i)}\right) \tilde{w}_t^{(i)},$$

where the *normalised importance weights* $\tilde{w}_t^{(i)}$ are given by:

$$\tilde{w}_t^{(i)} = \frac{w\left(\boldsymbol{\alpha}_{0:t}^{(i)}\right)}{\sum_{j=1}^{N} w\left(\boldsymbol{\alpha}_{0:t}^{(i)}\right)}.$$

This integration method can be interpreted as a sampling method, where the posterior distribution is approximated by:

$$\bar{p}_N(d\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t}) = \sum_{i=1}^{N} \tilde{w}_t^{(i)} \delta_{\boldsymbol{\alpha}_{0:t}^{(i)}}(d\boldsymbol{\alpha}_{0:t}).$$

It is clear that importance sampling needs all the data set $\mathbf{y}_{1:t}$ before estimating $p(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})$. That makes this method not adequate for recursive estimation, because, whenever new data $\mathbf{y}_{t+1}$ become available, the importance weights over the entire state sequence need to be recomputed. As the complexity of this operation increases with long sequences, recursive techniques for overcoming this problem have been studied.

Sequential Importance Sampling

Our aim is to estimate the posterior density function $p(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})$ without modifying the past simulated trajectories $\{\boldsymbol{\alpha}_{0:t-1}^{(i)};\ i = 1, \cdots, N\}$. This means that the importance function $\pi(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})$ has to admit $\pi(\boldsymbol{\alpha}_{0:t-1}|\mathbf{y}_{1:t-1})$ as marginal distribution, which happens when the importance function is restricted to be of the general form:

$$\begin{aligned} \pi(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t}) &= \pi(\boldsymbol{\alpha}_{0:t-1}|\mathbf{y}_{1:t-1})\pi(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{0:t-1}, \mathbf{y}_{1:t}) \\ &= \pi(\boldsymbol{\alpha}_0) \prod_{k=1}^{t} \pi(\boldsymbol{\alpha}_k|\boldsymbol{\alpha}_{0:k-1}, \mathbf{y}_{1:k}). \end{aligned}$$

This importance distribution allows the importance weights to be avaluated recursively:

$$\tilde{w}_t \propto \tilde{w}_{t-1} \frac{p(\mathbf{y}_t|\boldsymbol{\alpha}_t^{(i)}) p(\boldsymbol{\alpha}_t^{(i)}|\boldsymbol{\alpha}_{t-1}^{(i)})}{\pi(\boldsymbol{\alpha}_t^{(i)}|\boldsymbol{\alpha}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})}.$$

The only constraints on the selection of the importance function are those that have been mentioned so far. It follows that a wide choice for $\pi(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})$ is allowed.

Selection

Unfortunately, for the importance distributions of the form specified before, a degeneracy phenomenon may occur: after a few iterations, all but one of the normalised importance weights are very close to zero. This happens because the variance of the importance weights can only increase (stochastically) over time, as demonstrated in [1]. As a result of the degeneracy phenomenon, it is indispensable to include one more step (called *selection*) in the particle filter algorithm. The purpose of this procedure is to discard the particles with low importance weights, and to multiply the particles having high importance weights: the idea is that of associating with each particle (say $\tilde{\boldsymbol{\alpha}}_{0:t}^{(i)} : i = 1, \cdots, N$) a number of "children" $N_t^{(i)}$, such that $\sum_{i=1}^{N} N_t^{(i)} = N$, in order to obtain N new particles $\{\boldsymbol{\alpha}_{0:t}^{(i)} : i = 1, \cdots, N\}$. For each particle, if $N_t^{(j)} = 0$, then $\tilde{\boldsymbol{\alpha}}_{0:t}^{(j)}$ is discarded, otherwise it has $N_t^{(j)}$ children at step $t+1$. More formally, the weighted empirical distribution $\bar{p}_N(d\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t}) = \sum_{i=1}^{N} \tilde{w}_t^{(i)} \delta_{\boldsymbol{\alpha}_{0:t}^{(i)}}(d\boldsymbol{\alpha}_{0:t})$ is replaced by the unweighted measure

$$p_N(d\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t}) = \frac{1}{N} \sum_{i=1}^{N} N_t^{(i)} \delta_{\boldsymbol{\alpha}_{0:t}^{(i)}}(d\boldsymbol{\alpha}_{0:t}),$$

where $N_t^{(i)}$ is the number of offsprings associated to the particle $\boldsymbol{\alpha}_{0:t}^{(i)}$. After the selection step, all the importance weights are divided by N; since they do not depend on any past values of the normalised importance weights, all information regarding the old importance weights is discarded.

There is a variety of selection schemes, including *Residual Resampling* [31], *Stratified Sampling* [31], and *Multinomial Sampling*, also known as SIR (*Sampling Importance Resampling*) [21]: all of them can be implemented in a number of operations which is proportional to the number of particles N, and their aim is to provide the coefficients $N_t^{(i)}$ such that $p_N(d\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})$ is close to $\bar{p}_N(d\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})$, in the sense that, for any function f_t,

$$\int f_t(\boldsymbol{\alpha}_{0:t}) p_N(d\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t}) \approx \int f_t(\boldsymbol{\alpha}_{0:t}) \bar{p}_N(d\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t}),$$

according to different criteria.

As we can see in figure 2, the filtering density is approximated by an adaptive stochastic grid. This is a direct consequence of the Monte Carlo approach, where the particles interact with each other randomly in time, and either give birth to children, or die out, depending on the magnitude of their weights.

The Algorithm

Now it is possible to describe in outline the general particle filter algorithm: as stated before, the recursive Monte Carlo filter operates on N particles $\{\boldsymbol{\alpha}_{0:t}^{(i)} : i = 1, \cdots, N\}$, given at step $t-1$, and distributed approximately according to $p(\boldsymbol{\alpha}_{0:t-1}|\mathbf{y}_{1:t-1})$. The algorithm has a structure that can be divided into two main blocks, and it proceeds as follows at step t:

- Sequential Importance Sampling Step:
  - For $i = 1, \cdots, N$, sample
    $$\tilde{\boldsymbol{\alpha}}_t^{(i)} \sim \pi(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})$$
    and set
    $$\tilde{\boldsymbol{\alpha}}_{0:t}^{(i)} = (\boldsymbol{\alpha}_{0:t-1}^{(i)}, \tilde{\boldsymbol{\alpha}}_t^{(i)});$$
  - For $i = 1, \cdots, N$, evaluate the importance weights, up to a normalising constant:
    $$w_t^{(i)} \propto \frac{p(\mathbf{y}_t|\tilde{\boldsymbol{\alpha}}_{0:t}^{(i)}, \mathbf{y}_{1:t-1}) p(\tilde{\boldsymbol{\alpha}}_t^{(i)}|\tilde{\boldsymbol{\alpha}}_{t-1}^{(i)})}{\pi(\tilde{\boldsymbol{\alpha}}_t^{(i)}|\tilde{\boldsymbol{\alpha}}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})};$$
  - For $i = 1, \cdots, N$, normalise the importance weights:
    $$\tilde{w}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{j=1}^N w_t^{(j)}}.$$

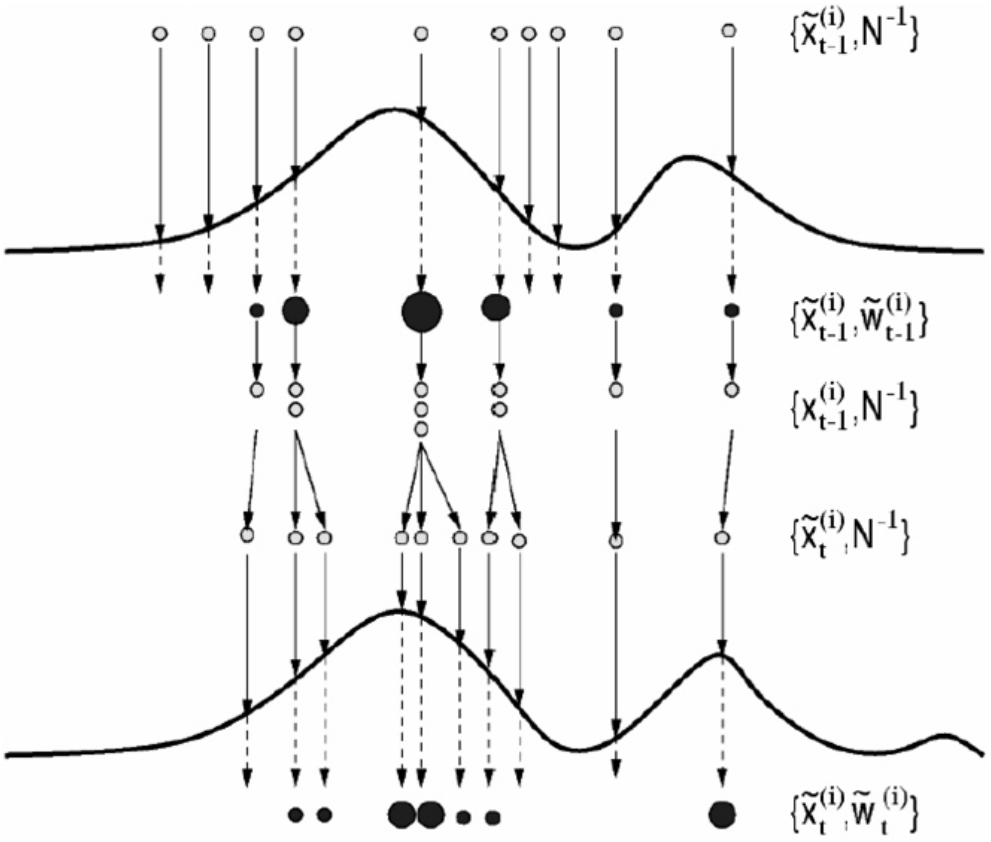

Figure 2: Particle Filtering - The adaptive stochastic grid and the selection step, when the density approximation is propagated from step $t-1$ to step t ([41]).

- Selection Step:
  - Discard / multiply particles $\{\tilde{\boldsymbol{\alpha}}_{0:t}^{(i)} : i = 1, \cdots, N\}$ with low / high normalised importance weights to obtain N particles:

$$\{\boldsymbol{\alpha}_{0:t}^{(i)} : i = 1, \cdots, N\}.$$

Although the computational complexity at each t step is proportional to N, the above algorithm is parallelisable, so that efficient implementation may be achieved by using parallel processors. It is worth mentioning that there is an unlimited number of choices for the implementation of this algorithm, as we have a lot of freedom both in the choice of the importance distribution and of the selection schemes.

2.3 Implementation

In this subsection, the particle filter is implemented considering the astrophysical context: our aim is to operate a Bayesian source separation of the different astrophysical components, given a set of observation, providing MMSE estimators of each source through the knowledge of the approximations of the posterior distributions computed by the particle filter. We allow the sources to have non-Gaussian distributions; the mixing-system is assumed to be non-stationary, and we also take space-varying noise into account.

Model Specification

Before illustrating the implementation of the particle filter algorithm, we introduce the model we will follow ([1]).

We consider instantaneous mixing of independent sources, each one modelled as a mixture of a known number of Gaussian components: Kuruoğlu et al. [34] demonstrated that it is possible to fit the curves of the astrophysical source distributions by a Gaussian mixture, using an Expectation-Maximization (EM) algorithm (figures 3 and 4). They tested the Gaussian mixture approximation on the main astrophysical components, for 15 different sky patches with a 15° aperture (256 × 256 pixels), and they observed that in all cases the Gaussian mixture model provides a good fit with three or five components only (see [34] for more details).

The mixing-system is assumed to be non-stationary, and we also take space-varying noise into account. Let n be the number of sensors: each of the n observations $\mathbf{y}$ will be represented as a row vector of t elements, where t is the number of the pixels we are taking into account. The number of independent sources $\boldsymbol{\alpha}$ is m (of course, each one is represented as a row vector of t elements).

The general model for the observations is thus, at time t:

$$\mathbf{y}_{1:n,t} = \mathbf{H}_t \boldsymbol{\alpha}_{1:m,t} + \mathbf{w}_{1:n,t}$$

where $\mathbf{y}_{1:n,t}$, $\boldsymbol{\alpha}_{1:m,t}$ and $\mathbf{w}_{1:n,t}$ are column vectors, representing the n observations, the m sources and the n additive noise samples at time t respectively. The $n \times m$ real valued mixing matrix $\mathbf{H}_t$ varies in t, and we can re-parametrise it into a vector $\mathbf{h}_t = vec\{\mathbf{H}_t\}$ so that $[\mathbf{h}_t]_{n(j-1)+1} = h_{i,j,t}$.

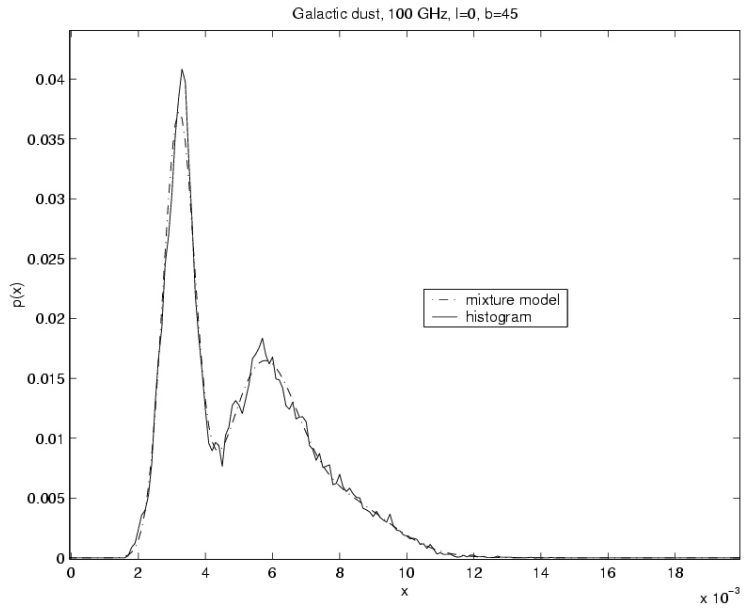

Figure 3: Source Separation - Galactic dust: histogram and Gaussian mixture model fit. ([34]).

Now we are able to express the model in terms of state equation and observation equation, in this way:

$$\mathbf{h}_t = \mathbf{A}_t \mathbf{h}_{t-1} + \mathbf{v}_t$$

$$\mathbf{y}_{1:n,t} = \mathbf{C}_t \mathbf{h}_t + \mathbf{w}_{1:n,t}$$

where $\mathbf{A}_t$ and $\mathbf{C}_t$ are $(nm \times nm)$ and $(n \times nm)$ real valued matrices respectively, and $\mathbf{w}_t$ is a $(n \times 1)$ real vector. $\mathbf{C}_t$ can be expressed in terms of the source signal vector $\boldsymbol{\alpha}_{1:m,t}$, as $\mathbf{C}_t = \boldsymbol{\alpha}_{1:m,t}^T \bigotimes \mathbf{I}_n$. In absence of further prior information, we assume $\mathbf{A}_t = \mathbf{I}_{nm}$, and of course $\mathbf{C}_t$ is unknown, as it consists of the source signals to be estimated. The distributions of the dynamic noise $\mathbf{v}_t$ and the observation noise $\mathbf{w}_t$ are assumed to be i.i.d. and mutually independent: $\mathbf{v}_t \sim \mathcal{N}(0, \boldsymbol{\sigma}_v)$ and $\mathbf{w}_t \sim \mathcal{N}(0, \boldsymbol{\sigma}_w)$, with obvious notation. The introduction of the state equation allows to deal with non-stationary mixing matrices, as the coefficients of $\mathbf{h}$ can be updated at every step.

In this formulation there is a scaling ambiguity, as we can multiply $\mathbf{H}_t$ by a non-zero constant c and divide the sources $\boldsymbol{\alpha}_{1:m,t}$ by c and obtain the same observations: in order to solve this ambiguity, we constrain $\mathbf{H}_t$ to have constant unity diagonal for $\mathbf{H}_t$ square $(m = n)$, or set the diagonal of sub-matrix $\mathbf{H}^a_{m \times m}$ to unity if $n > m$:

$$\mathbf{H}_t = \begin{bmatrix} \mathbf{H}^a_{m \times m} \\ \mathbf{H}^b_{(n-m) \times m} \end{bmatrix}$$

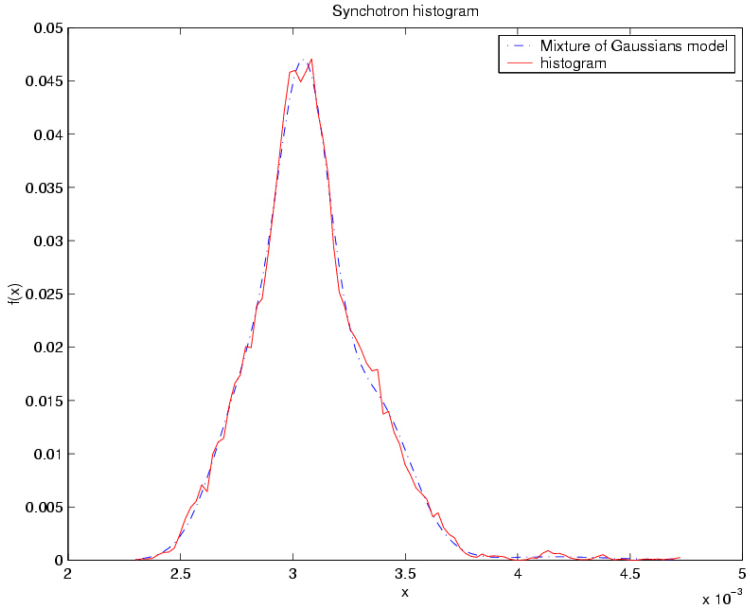

Figure 4: Source Separation - Synchrotron: histogram and Gaussian mixture model fit. ([34]).

Model of the Sources

As the m sources are statistically independent of one another:

$$p(\boldsymbol{\alpha}_{1:m,t}) = \prod_{i=1}^{m} p(\alpha_{i,t}).$$

Moreover, we can model each source by a finite mixture of Gaussians, so:

$$p(\alpha_{i,t}|\mu_{i,j,t}, \sigma^2_{i,j,t}) = \sum_{j=1}^{q_i} \rho_{i,j} \mathcal{N}(\alpha_{i,t}; \mu_{i,j,t}, \sigma^2_{i,j,t}); \sum_{j=1}^{q_i} \rho_{i,j} = 1,$$

where $\rho_{i,j}$ is the weight of the j^{th} Gaussian component of the i^{th} source, and q_i is the number of Gaussian components for the i^{th} source.

Now we will consider a hidden variable z_i which takes on a finite set of values $Z_i = \{1, \cdots, q_i\}$, so that we can denote the distribution of $\alpha_{i,t}$ as if at time t only the j^{th} Gaussian component is active, with probability $\rho_{i,j}$:

$$p(\alpha_{i,t}|z_{i,t} = j) = \mathcal{N}(\alpha_{i,t}; \mu_{i,j}, \sigma^2_{i,j})$$

At time t let $\mathbf{z}_{1:m,t} \triangleq [z_{1,t} \cdots z_{m,t}]^T$. Given that the sources are statistically independent of one another, $\alpha_{1:m,t}$ have distributions:

$$p(\boldsymbol{\alpha}_{1:m,t}|\mathbf{z}_{1:m,t}) = \mathcal{N}(\boldsymbol{\alpha}_{1:m,t}; \boldsymbol{\mu}(\mathbf{z}_{1:m,t}), \boldsymbol{\Gamma}(\mathbf{z}_{1:m,t})),$$

where

$$\boldsymbol{\mu}(\mathbf{z}_{1:m,t}) = [\mu_{1,z_{1,t}}, \cdots, \mu_{m,z_{m,t}}]^T$$

and

$$\boldsymbol{\Gamma}(\mathbf{z}_{1:m,t}) = diag\{\sigma^2_{1,z_{1,t}}, \cdots, \sigma^2_{m,z_{m,t}}\}.$$

It is possible to describe the discrete probability distribution of $\mathbf{z}_{1:m,t}$ using the i.i.d. model: in this case, the indicators of the states $z_{i,t}$ have identical and independent distributions. If we want to introduce temporal correlation beween the samples of a particular source, we have to consider the first-order Markov model case, where the vector of the states evolves as a homogeneous Markov chain for $t > 1$:

$$p(\mathbf{z}_{1:m,t} = \mathbf{z}_l | \mathbf{z}_{1:m,t-1} = \mathbf{z}_j) = \prod_{i=1}^{m} p(z_{i,t} = [\mathbf{z}_l]_i | z_{i,t-1} = [\mathbf{z}_j]_i) = \prod_{i=1}^{m} \pi^{(i)}_{j,l},$$

where $\pi^{(i)}_{j,l}$ is an element of the $q_i \times q_i$ real valued *transition matrix* for the states of the i^{th} source, denoted by $\boldsymbol{\pi}^{(i)}$. The state transition can be thus parametrised by a set of m transition matrices $\boldsymbol{\pi}^{(i)}$, $i \in \{1, \cdots, m\}$.

Given the observations $\mathbf{y}_t$ (assuming that the number of sources m, the number of Gaussian components q_i for the i^{th} source, and the number of sensors n are known), we would like to estimate all the following unknown parameters of interest, grouped together:

$$\boldsymbol{\theta}_{0,t} = [\boldsymbol{\alpha}_{1:m,0:t}, \mathbf{z}_{1:m,0:t}, \{\mu_{i,j,0:t}\}, \{\sigma^2_{i,j,0:t}\}, \{\boldsymbol{\pi}^{(i)}_{0:t}\}, \{\boldsymbol{\sigma}^2_{\mathbf{w}_{1:n,0:t}}\}],$$

where we recall that $\boldsymbol{\alpha}_{1:m,0:t}$ are the sources, $\mathbf{z}_{1:m,0:t}$ is the matrix of the indicator variables which determines which Gaussian component is active at a particular time for each source, $\{\mu_{i,j,0:t}\}$ and $\{\sigma^2_{i,j,0:t}\}$ are the means and the variances of the j^{th} Gaussian component of the i^{th} source, $\{\boldsymbol{\pi}^{(i)}_{0:t}\}$ is the transition matrix for the evolution of $z_{i,0:t}$, and $\{\boldsymbol{\sigma}_{\mathbf{w}_{1:n,0:t}}\}$ represents the standard deviation of the observation noise.

Rao-Blackwellisation

In our case, referring to the model of the sources defined before, we want to estimate the wide set of unknown parameters grouped together in

$$\boldsymbol{\theta}_{0,t} = [\boldsymbol{\alpha}_{1:m,0:t}, \mathbf{z}_{1:m,0:t}, \{\mu_{i,j,0:t}\}, \{\sigma^2_{i,j,0:t}\}, \{\boldsymbol{\pi}^{(i)}_{0:t}\}, \{\boldsymbol{\sigma}^2_{\mathbf{w}_{1:n,0:t}}\}],$$

and we have also to consider that the mixing matrix is both space-varying and not precisely known. The particles we should deal with will be thus $\{(\mathbf{h}^{(i)}_{0:t}, \boldsymbol{\theta}^{(i)}_{0:t}) : i = 1, \cdots, N\}$, generated according to $p(\mathbf{h}_{0:t}, \boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$. An empirical estimate of this distribution is given by

$$\bar{p_N}(d\mathbf{h}_{0:t}, \boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\boldsymbol{h_{0:t}}^{(i)}}, \delta_{\boldsymbol{\theta_{0:t}}^{(i)}}(d\mathbf{h}_{0:t}, d\boldsymbol{\theta}_{0:t}),$$

and, as a corollary, an estimate of $p(\mathbf{h}_t, \theta_t | \mathbf{y}_{1:t})$ is

$$\bar{p_N}(d\mathbf{h}_t, \boldsymbol{\theta}_t | \mathbf{y}_{1:t}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\boldsymbol{h_t}^{(i)}}, \delta_{\boldsymbol{\theta_t}^{(i)}}(d\mathbf{h}_t, \boldsymbol{\theta}_t).$$

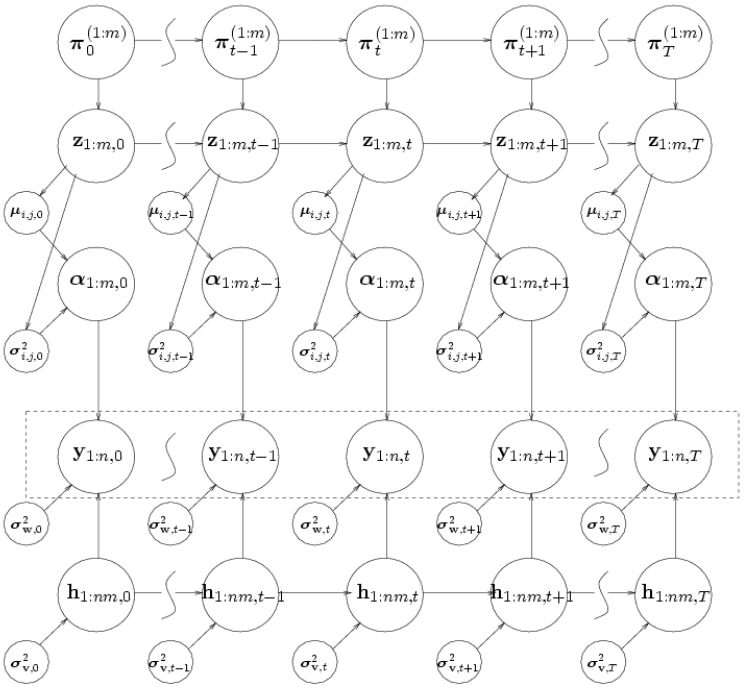

Figure 5: Particle Filtering - Graphical representation of the model ([1]).

It is possible to reduce the problem of estimating $p(\mathbf{h}_t, \boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$ to a simpler one of sampling from $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$. In fact,

$$p(\mathbf{h}_t, \boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) = p(\mathbf{h}_t|\boldsymbol{\theta}_{0:t}, \mathbf{y}_{1:t})p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}).$$

Given an approximation of $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$, an approximation of $p(\mathbf{h}_t|\boldsymbol{\theta}_{0:t}, \mathbf{y}_{1:t})$ may straightforwardly be obtained considering the following state space model for each particle:

$$\begin{aligned} \mathbf{h}_t^{(i)} &= \mathbf{A}_t \mathbf{h}_{t-1}^{(i)} + \mathbf{v}_t^{(i)}; \\ \mathbf{y}_{1:n,t} &= \mathbf{C}_t \mathbf{h}_t^{(i)} + \mathbf{w}_{1:n,t}^{(i)}; \end{aligned}$$

where we recall that $\mathbf{C}_t$ can be expressed in terms of the source signal vector $\boldsymbol{\alpha}_{1:m,t}$, as $\mathbf{C}_t = \boldsymbol{\alpha}_{1:m,t}^T \bigotimes \mathbf{I}_n$. The posterior distribution of the state $\mathbf{h}_t$ given the observations $\mathbf{y}_{1:n,t}$ can be recursively estimated in closed form using the Kalman filter [30].

This technique, called *Rao-Blackwellisation* [13], leads to better results, as we are reducing the size of the parameter set to be estimated by marginalising out the mixing coefficients $\mathbf{h}_t$ using the Kalman filter, so that the only distribution we have to estimate by particle filtering is $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$.

Prior Distribution as Importance Function

Referring to the approach defined before, the samples used to estimate the posterior density functions of the parameters of interest have to be drawn from an importance distribution of the general form

$$\pi(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) = \pi(\boldsymbol{\theta}_{0:t-1}|\mathbf{y}_{1:t-1})\pi(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t})$$

The best strategy is to choose, at step t, the importance distribution that minimises the variance of the importance weights, given $\boldsymbol{\theta}_{0:t-1}$ and $\mathbf{y}_{1:t}$. In [16] we find the proof that the importance distribution we are looking for is:

$$\pi(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t}) = p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t}).$$

From Bayes' rule, the optimal importance distribution may be expressed as

$$p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\boldsymbol{\theta}_{0:t}, \mathbf{y}_{1:t-1})p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})}{p(\mathbf{y}_t|\boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t-1})},$$

being

$$p(\mathbf{y}_t|\boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t-1}) = \int p(\mathbf{y}_t|\boldsymbol{\theta}_{0:t}, \mathbf{y}_{1:t-1})p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{0:t-1})d\boldsymbol{\theta}_t.$$

Unfortunately it is not easy to sample directly from the optimal importance distribution, and the above integral cannot be evaluated analitically, since $p(\mathbf{y}_t|\boldsymbol{\theta}_{0:t}, \mathbf{y}_{1:t-1})$ is a complex possibly non-linear function of $\boldsymbol{\theta}_t$. This is the reason why the following sub-optimal method will be employed throughout, taking the importance distribution at step t to be the *prior distribution*:

$$\pi(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) = p(\boldsymbol{\theta}_{0:t}) = p(\boldsymbol{\theta}_0)\prod_{k=1}^{t} p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{0:k-1}).$$

In this case, the importance weights can be computed recursively by

$$\tilde{w}_t^{(i)} \propto \tilde{w}_{t-1}^{(i)} p(\mathbf{y}_t|\boldsymbol{\theta}_t^{(i)})$$

whose evaluation requires only one step of the Kalman filter for each particle. Now it is convenient to factorize the prior importance function:

$$\begin{aligned} p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) &= p(\boldsymbol{\alpha}_{1:m,t}, \mathbf{z}_{1:m,t}, \boldsymbol{\pi}_t, \{\mu_{i,j,t}\}, \{\sigma^2_{i,j,t}\}, \{\sigma^2_{\mathbf{w},K,t}\}| \\ &\quad \boldsymbol{\alpha}_{1:m,t-1}, \mathbf{z}_{1:m,t-1}, \boldsymbol{\pi}_{t-1}, \{\mu_{i,j,t-1}\}, \{\sigma^2_{i,j,t-1}\}, \{\sigma^2_{\mathbf{w},K,t-1}\}) \\ &= p(\boldsymbol{\alpha}_{1:m,t}, \mathbf{z}_{1:m,t}, \boldsymbol{\pi}_t, \{\mu_{i,j,t}\}, \{\sigma^2_{i,j,t}\}| \\ &\quad \boldsymbol{\alpha}_{1:m,t-1}, \mathbf{z}_{1:m,t-1}, \boldsymbol{\pi}_{t-1}, \{\mu_{i,j,t-1}\}, \{\sigma^2_{i,j,t-1}\}) \times \\ &\quad p(\{\sigma^2_{\mathbf{w},K,t}\}|\{\sigma^2_{\mathbf{w},K,t-1}\}) \end{aligned}$$

If now we consider a new variable, $\tilde{\boldsymbol{\theta}}_t$, which excludes the observation noise variance, we obtain

$$\begin{aligned} p(\tilde{\boldsymbol{\theta}}_t|\tilde{\boldsymbol{\theta}}_{t-1}) &= p(\boldsymbol{\alpha}_{1:m,t}, \mathbf{z}_{1:m,t}, \boldsymbol{\pi}_t, \{\mu_{i,j,t}\}, \{\sigma^2_{i,j,t}\}|\tilde{\boldsymbol{\theta}}_{t-1}) \\ &= p(\boldsymbol{\alpha}_{1:m,t}|\mathbf{z}_{1:m,t}, \{\mu_{i,j,t}\}, \{\sigma^2_{i,j,t}\}) \times \\ &\quad p(\{\mu_{i,j,t}\}|\{\mu_{i,j,t-1}\}, \mathbf{z}_{i,t}) \times \\ &\quad p(\{\sigma^2_{i,j,t}\}|\{\sigma^2_{i,j,t-1}\}, \mathbf{z}_{i,t}) \times \\ &\quad p(\mathbf{z}_{1:m,t}|\mathbf{z}_{1:m,t-1}, \boldsymbol{\pi}_t) \times \\ &\quad p(\boldsymbol{\pi}_t|\boldsymbol{\pi}_{t-1}). \end{aligned}$$

This hierarchical structure allows us to obtain an approximation of the distribution of the sources exploiting the particles generated from the distributions of the other parameters, sampling subsequently from $p(\boldsymbol{\pi}_t|\boldsymbol{\pi}_{t-1})$, $p(\mathbf{z}_{1:m,t}|\mathbf{z}_{1:m,t-1}, \boldsymbol{\pi}_t)$, $p(\{\sigma^2_{i,j,t}\}|\{\sigma^2_{i,j,t-1}\}, \mathbf{z}_{i,t})$, $p(\{\mu_{i,j,t}\}|\{\mu_{i,j,t-1}\}, \mathbf{z}_{i,t})$, and finally obtain the particles of the distribution $p(\boldsymbol{\alpha}_{1:m,t}|\mathbf{z}_{1:m,t}, \{\mu_{i,j,t}\}, \{\sigma^2_{i,j,t}\})$.

3 Astrophysical Background

Cosmology is the scientific study of the large scale properties of the Universe as a whole. It endeavours to use the scientific method to understand the origin, evolution and ultimate fate of the entire Universe. Like any field of science, cosmology involves the formation of theories or hypotheses about the Universe which make specific predictions for phenomena that can be tested with observations. Depending on the outcome of the observations, theories need to be abandoned, revised or extended to accommodate the data. This section is an introduction to the cosmological context; for a detailed description, more information can be found in [24], [42], [12].

3.1 The Big Bang

One of the best known theories in Cosmology is the Big Bang, which is based on the idea that our Universe started out much hotter and denser than it is now, and it has been expanding since then. Extrapolating the history of the universe backwards using current physical models leads to a gravitational singularity, at which all distances become zero and temperatures and pressures become infinite. What this means is unclear, and most physicists believe that this result is due to our limited understanding of the laws of physics with regard to this type of situation.

This theory is based on observations of our universe, among which are two main evidences:

- the first one is the fact that external galaxies are receding in such a way that their recessional speed is proportional to the distance they are away from us (this is called *Hubble's Law* after Edwin Hubble who first noticed it in 1929). This observation is explained well by a uniform expansion of the universe: if the universe is expanding, it must have started out very small some time far in the past;

- the second evidence consists in the fact that, when we observe the night sky, we see an excess of radiation which is called the Cosmic Microwave Background (CMB) radiation.

Its existence was predicted by George Gamow, Ralph Alpher, and Robert Hermann in the 1940s, and was accidentally observed for the first time in 1965 by Penzias and Wilson, who received a Nobel Prize for this discovery. The CMB is a very special light that fills the Universe: its formation occurred when the Universe was about only 300000 years old (a very early time compared to 10-20 thousand million years - the extimated age of Universe), when the galaxies had not formed yet, as it did not originate from one object in particular, but from the whole Universe. Being a sort of "shockwave" of the Big Bang, it can be detected today as coming from everywhere in the sky. This is the reason why, in literature, the CMB is also called *the oldest data set* in the Universe [12]. Observing this "first light" today is like seeing the Universe as it was only 300000 years after the Big Bang: for this reason, the Planck satellite is a sort of "time machine" astronomers will use to travel back in time, towards the beginning of space and time, in order to bring back precious data.

The Big Bang model says that the Universe started with a very dense and hot phase that expanded and cooled itself; for hundreds of thousands of years the temperature was so high that neutral atoms could not form, and thus matter consisted mostly of neutrons and charged particles (protons and electrons). Electrons interacted closely with the light particles, and therefore light and matter were tightly coupled at that time. As a consequence, light could not propagate and the Universe was opaque. It took some 300000 years for the Universe to cool down to a temperature at which atoms can form, that is about 3000 degrees. It happened what astronomers call the *recombination* process: matter became neutral, and the Universe became transparent. This allowed the light to travel freely, and that first light is what now we call the Cosmic Microwave Background radiation.

Since the time when the CMB was released, the Universe has expanded becoming at the same time cooler and cooler. The cosmic microwave background has been affected by the same process: it has expanded and cooled down. When we say expansion we actually mean that space has 'stretched' itself, and, with it, all length scales. Light is, after all, a wave, and when a wave is stretched its characteristic length scale and its frequency change. Today, the cosmic background can be detected at microwave frequencies, and the properly designed detectors of the Planck satellite will examine it with an accuracy never achieved before: the Planck instruments will detect the Cosmic Microwave Background by translating it into a temperature.

The waves of light are a form of energy, and temperature is a measure of this energy : the higher the frequency is, the more energy it has, and the hotter it is. So the detected microwaves can be also felt as a temperature: that is the same as saying that Planck will measure the temperature of each part of the whole sky. The CMB consists of a perfect black body, and now its temperature is already known to be 2.726 K all over the sky to three decimal figures: this degree of accuracy in the measurement may seem good enough, but it's not. In fact, thanks to previous observations, scientists know that slightly hotter or colder "patches" appear in the sky when measuring more precisely: one part in 100000, or 0.00001 degrees of difference. These differences in temperature are nothing less than the imprints left in the CMB by the primeval "seeds" of today's huge concentrations of matter (galaxies, galaxy clusters, ...), and this is the reason why the analysis of the CMB is currently one of the most promising ways to understand the birth and evolution of the Universe in which we live.

3.2 CMB Missions

Probably still the most famous mission dedicated to this subject is the Cosmic Background Explorer COBE satellite (launched by NASA in 1989) which made the first detection of the CMB *intrinsic* anisotropies - that is not due to other astrophysical source disturbs. More detailed measurements have been subsequently made using instruments placed on high flying balloons (the *Boomerang* experiment, in 1998, was one of them) that observed only small sections of the sky. In June 2001, NASA launched a second space mission, called MAP, to make detailed measurements of the anisotropies over the full sky.

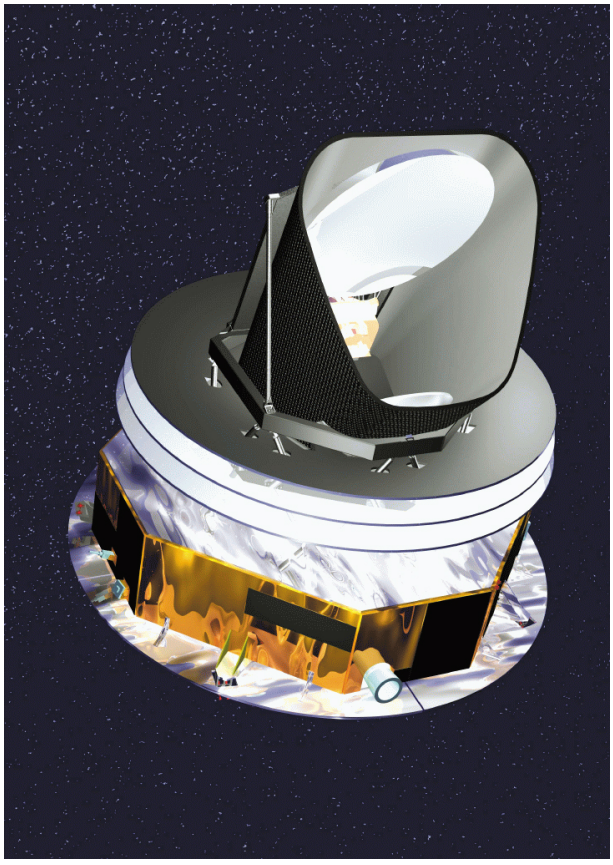

Figure 6: Astrophysical background - The Planck satellite ([24]).

Planck is the first European mission to study the birth of Universe, and will be launched in January 2007. Its primary objectives are the measurements of the CMB anisotropies and of the amplitude of its structures, and test inflationary models of the early Universe. Its instruments will analyze the microwave portion of the electro-magnetic spectrum between 30 and 857 GHz, with an angular resolution up to 30". The information Planck has to gather lies precisely in the pattern formed by these slightly hotter and colder regions. As a consequence, the Planck detectors will have to be highly sensitive: engineers will have to push current technology to its limits if they want to get useful scientific results (just to illustrate the challenge, Planck

detectors will have to work at temperatures very close to the absolute zero, otherwise their own heat emission will spoil the measurements).

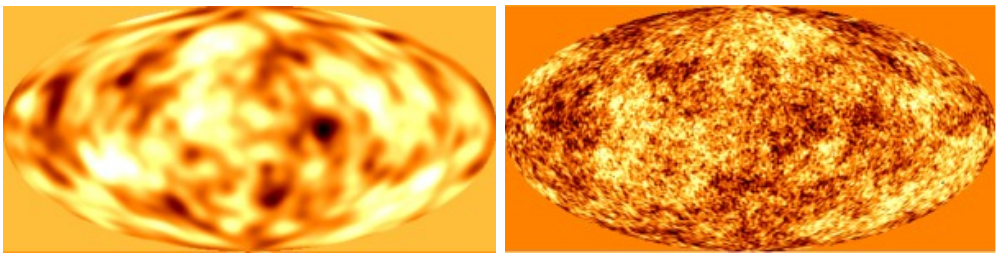

Figure 7: Astrophysical background - Cosmic Microwave Background radiation as measured by COBE (left), and simulated to Planck resolution (right) ([32]).

3.3 The Dipole Anisotropy

As introduced before, tiny fluctuations of the order of a few micro-Kelvins are present in the intensity of the CMB radiation, and their analysis will reveal how matter and energy were spatially distributed at the time of atomic recombination. However, the most influent anisotropy present in CMB maps is not generated by primitive inhomogenities or astrophysical contaminants, but to the so-called *dipole anisotropy*, which is due to the motions of the Earth (around the Sun), the Solar system (around our galaxy), and the Milky Way (which moves towards the Virgo Cluster). Nevertheless, the distortion caused by the dipole anisotriopy can be removed easily, exploiting the knowledge about direction and velocity of these motions.

3.4 The Foregrounds

The aim of this subsection is to introduce the main CMB contaminants that are present in sky maps. We divide these components into *Galactic* and *Extragalactic foregrounds*, depending on their origin.

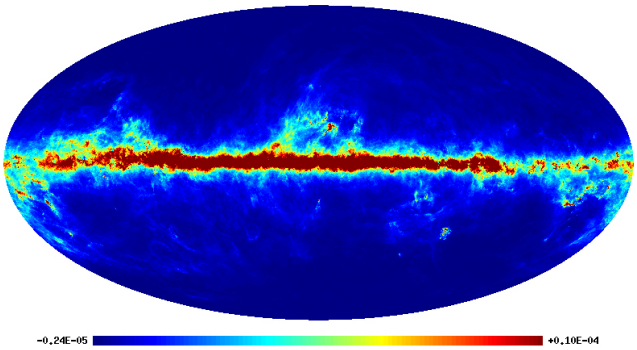

Figure 8: Astrophysical background - The contribution of Galactic dust ([42]).

Galactic Components

- *Thermal Dust*: The thermal dust emission is the strongest Galaxy contribution at high microwave frequencies, in particular at the high Planck ones. It is produced by small dust grains (the size is of a few μm) that absorb the ultra-violet light, which is re-emitted in the far-infrared.

- *Spinning Dust*: The spinning dust emission has been proposed recently as a mechanism to account for the anomalous dust-correlated emission at low frequencies. However, strong uncertainties are still present, and the Planck mission could help to resolve them.

- *Synchrotron*: The synchrotron emission is due to charged relativistic particles, accelerated by the magnetic field of the Galaxy. It carries information about the structure of the magnetic field, the spatial and energy distribution of relativistic electrons, and the variations of electron density, electron energy, and magnetic field introduced by supernova shocks.

- *Free-Free*: The free-free emission, or *bremsstrahlung*, is the most unknown of all Galactic emissions, since it is the dominant one in a short interval, where the Galactic signal is, in any case, quite weak by itself. This emission is due to the interaction of un-bound very hot electrons ($T_{electron} \sim 10^4 K$) with ions.

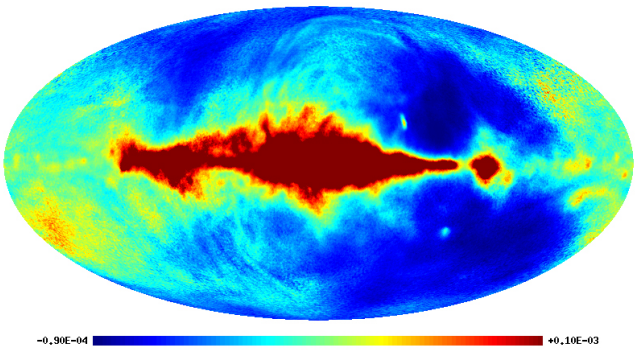

Figure 9: Astrophysical background - The contribution of synchrotron ([42]).

Extragalactic Components

- *Point Sources*: The effect produced by Point Sources (PS) appears in the small scales of the CMB fluctuations, and it is a very important contaminant generated by extragalactic microwave sources like radio galaxies, infrared galaxies and quasars.

- *The Sunyaev-Ze'ldovich Effect*: When the CMB photons cross the hot electron gas inside a galaxy cluster, they suffer the so-called *inverse Compton scattering* that modifies the CMB spectrum. This phenomenom is known as the *thermal* Sunyaev-Ze'ldovich (SZt) effect. This means that the SZt is not a foreground like the Galactic ones or the PS emission, but it could be considered as a *secondary* CMB anisotropy, because the involved photons are

the CMB ones. However, since we are focussing on the component separation problem in CMB images, we consider this effect like another CMB contaminant. Moreover, there exist a Doppler effect due to the relative velocity of the Galaxy clusters related to the CMB, too: this is called the *kinematic* Sunyaev-Ze'ldovich (SZk) effect, whose detection is a very difficult issue, being up to 100 times lower than the SZt. The study of this effect is of high interest, as it leads to put fundamental constraints on the cosmological parameters.

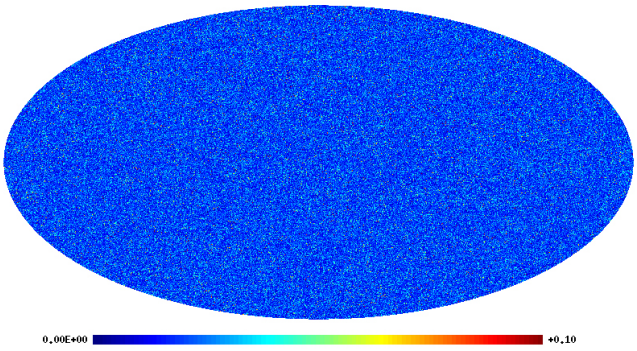

Figure 10: Astrophysical background - The contribution of extragalactic point sources ([42]).

3.5 CMB Intrinsic Anisotropies

After eliminating the dipole distortion and the foreground contributions, all the other anisotropies are due to the CMB itself, that is to matter density primitive fluctuations. These intrinsic anisotropies are essentially originated by three principal mechanisms [3]:

- The *adiabatic effect* is due to the fact that the primitive matter and radiation possess their own fluctuations, generated before the recombination epoch: these fluctuations are the original seeds of the structure of the Universe. This effect occurs at all angular scales, but it is predominant only at $\theta < 1'$.
- The *Sachs-Wolfe effect* was proposed in 1966, and it is related with a change in the early photons frequency, due to fluctuations of the gravitational potential, because of density matter perturbations. Its effects are present mainly at angular scales $\theta > 1°$.
- The *Doppler effect* is due to the fact that the particles which emitted the early photons had a non-zero velocity. The anisotropies generated by this effect are present at angular scales $0.1° < \theta < 2°$.

The magnitude of these anisotropies at different angular scales is studied through the analysis of the CMB power spectrum, which is almost constant, except for a series of peaks (named *Doppler peaks*) in correspondence of angles $\theta \sim 1°$, where the acoustic effect is dominant. The crucial point for Cosmology is that the position and height of these peaks depend on the value of the fundamental cosmological parameters, such as the Hubble Constant, the cosmological density parameters (related both to baryonic matter and to dark matter), and the cosmological constant, whose values are known with great uncertainty at the moment.

4 Numerical Experiments

4.1 Our Data

In order to test our algorithm [15] and its performance in the context of the Planck mission, we use synthetic but realistic maps of the CMB and of some foregrounds, each of 256x256 pixels, which have been generated according to the available a-priori information, at Planck's resolution. The expected noise RMS maps at different frequencies have been exploited to generate the additive noise samples.

Astrophysical Sources

The simulated sky templates used in this work have been provided by the Planck Technical Working Group. The CMB fluctuations have been generated using the flat Cold Dark Matter (CDM) model, whose anisotropies have a Gaussian distribution, in accordance with the standard cosmological theories. Much less is known about the statistical distributions of the other astrophysical components: in our experiments, only Galactic dust and synchrotron emission have been taken into account. For Galactic dust, existing sky maps obtained from different frequency channel measurements of the COBE satellite [25] have been used as spatial templates, from which the specific emission values have been generated according to the hypothesized dust emission process. As for Galactic synchrotron emission, maps have been obtained by extrapolating existing COBE data, both for spatial resolution and for spectral emission.

The Mixing Matrix

In order to create realistic mixtures, we used the mixing coefficients reported in [8], to obtain the observations at 70 GHz and 100 GHz.

Noise Distribution

As it was stated before, previous approaches assumed noiseless models (for example, in [8] the ICA method is used), and in the few cases when noise is taken into account it is generally assumed to be Gaussian and space-invariant [36].

However, in our work noise is assumed to be Gaussian, space-varying, in accordance to the fact that the value given to each pixel of the data maps is obtained from multiple measurements by the Planck antenna in the same direction, and the number of multiple measurements is not the same for pixels with different locations (figure 11).

4.2 Choice of the Filter Parameters

The particle filter algorithm needs to be initialized, and a set of parameters and a-priori information must be given as an input of the filter itself. The following choices have been adopted to test our algorithm:

- *Number of Gaussian components*:
  The algorithm assumes that the CMB source has a Gaussian distribution, according to

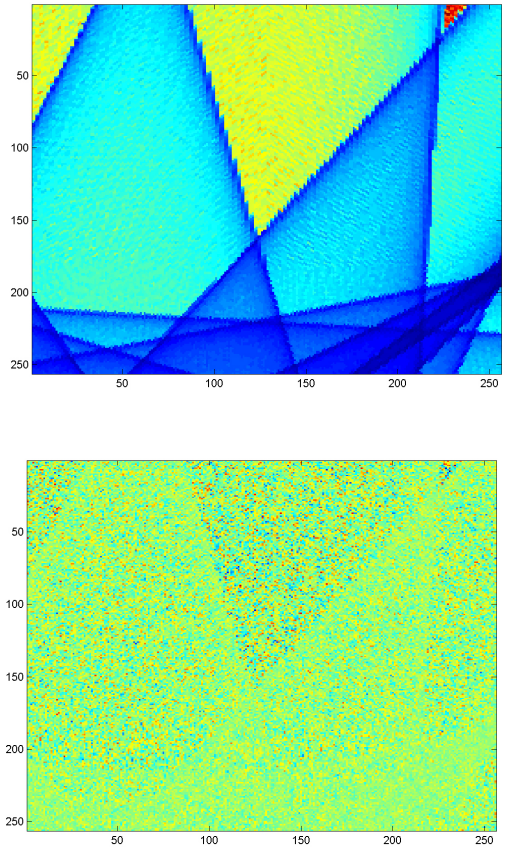

Figure 11: Simulations - RMS noise map for a sky patch of 256×256 pixels (top) and one possible noise realization (bottom).

the theoretical astrophysical knowledge which implies a flat Cold Dark Matter model for the CMB, whose anisotropies must have a Gaussian distribution. The foreground sources are separated assuming that their actual distributions can be approximated by a mixture of 3 or 5 Gaussians. Better approximations could be obtained by increasing the number of Gaussian components for each source, although in this case much more computational time would be needed. The active Gaussian component is selected through the indicator matrix $\mathbf{z}_{1:m,t}$, whose distribution is in accordance to the model presented in section 2.3.

- *Number of particles*:
  Simulations are performed by generating 70-100 particles at every *sequential importance sampling step*. Also in this case, better results could be obtained by increasing the number of particles.

- *RMS noise matrices*:
  A-priori information about the space-varying noise variance has been exploited: a realistic

256×256 matrix containing the noise variance of each pixel is given as an input to the algorithm for each observation.

- *Mixing matrix initialization*:
  The diagonal elements of the mixing matrix are constrained to unity. The other elements have been initialized with a 20 % error with respect to the values used to generate the mixtures.

- *Starting values of the parameters*:
  The starting values of the means $\{\mu_{i,j,0:t}\}$ and the variances $\{\sigma^2_{i,j,0:t}\}$ of the j^{th} Gaussian component of the i^{th} source are decided after a quick inspection of the histogram of each source: the starting values of the means of the q Gaussian components of each source are those which produce the q most significant local maxima of each histogram. The same qualitative approach is followed to give the starting values of the variances of each Gaussian component. The indicator matrix $\mathbf{z}_{1:m,0:t}$, which determines the active Gaussian component at a particular time for each source, is initialized with a discrete uniform distribution, while the transition matrix $\{\boldsymbol{\pi}^{(i)}_{0:t}\}$ for the evolution of $z_{i,0:t}$ is initialized with a continuous uniform distribution. No starting values for the sources $\boldsymbol{\alpha}_{1:m,0:t}$ are needed, as the particles that approximate the source distributions are generated using the particles of the other parameters, following the hierarchical structure of figure 5.

- *Drift parameters*:
  The drift parameters are used to generate the particles that describe the distributions of interest, so a correct choice is fundamental to obtain satisfactory results. In our simulations, the following values of the drift parameters have been used: for the particle of the matrix $\{\boldsymbol{\pi}^{(i)}_{0:t}\}$, a 3 % drift has been used, while small drifts of 0.0025%, or even smaller, have been chosen for the means and variances of the Gaussian components.

Possible alternative choices for future research are discussed in the last section.

4.3 Choice of Priors

The particles related to $\mathbf{z}_{1:m,0:t}$ are sampled from a Dirichlet prior, while those of $\{\sigma^2_{i,j,0:t}\}$ are generated by an inverted-Gamma distribution, according to [1]. The particles for the estimation of the other parameters of interest are sampled from Gaussian distributions, centered on the values of the particles themselves at the previous step, with a variance which depends on the value of the drift parameters defined above.

4.4 Simulations

In this subsection, we illustrate the results of the simulations obtained using the particle filtering technique. The standard *residual resampling* procedure has been employed to implement the *selection step*, and we used the prior distribution as importance function, because of its appealing properties, which have been described in the previous section.

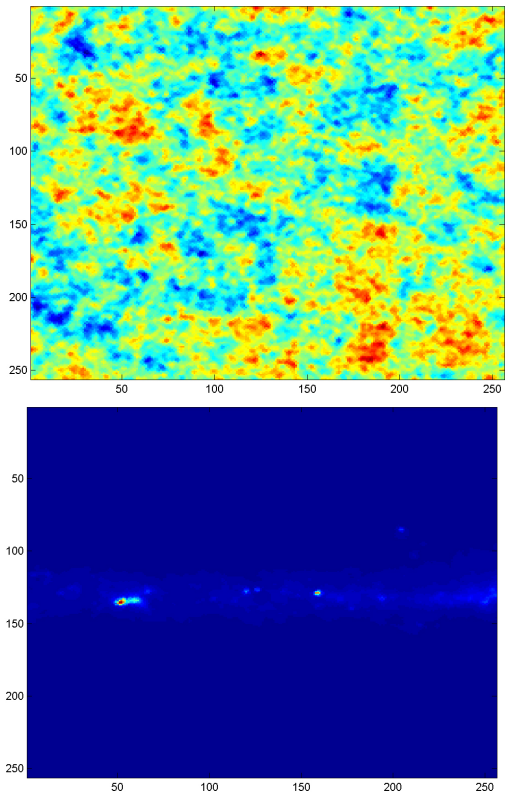

Figure 12: Simulations - Our simulated sources: CMB (top) and Galactic dust (bottom) on the Galactic plane.

Effect of the Choice of Drift Parameters

In this first simulation we will try to separate the CMB and the dust radiation on the Galactic plane, exploiting the information contained in the observed mixtures at 70 GHz and 100 GHz (figures 12 and 13). In this example, the Signal to Noise Ratio (SNR) varies between -4dB and +6dB: the dust signal is stronger on the Galactic plane, while the CMB radiation is predominant elsewhere, and we recall that noise is space-varying.

Two-dimensional data matrices are reduced to vectors by scanning each image by columns. 70 particles are used to generate each approximated probability density function of the parameters of interest, following the hierarchical scheme illustrated in figure 5.

First, we want to emphasize the importance of a correct choice of the drift parameters, as too small ones imply an unproper exploration of the distribution of interest, while a high drift implies the fact that also the variance of the estimates of the sources will be high. In this example we choose small values for the drift parameters: figure 14 represents the plot of the dust signal and its estimate. A sudden change of the statistical properties of the source signal implies a change of the estimated statistical distributions, too, and a set of small drift parameters is unable to explore a previously insignificant area of the statistical distribution itself which suddenly becomes very significant. In this case, the estimate was not able to fit the source.

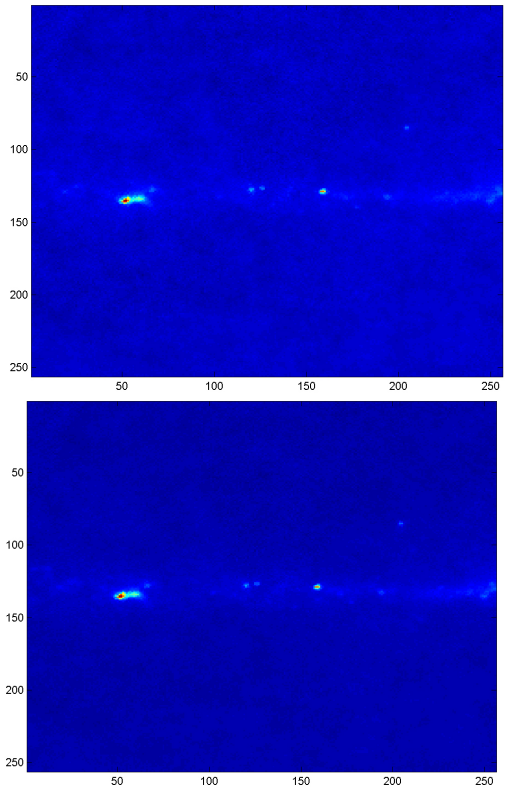

Figure 13: Simulations - The two observed signals. Notice that the Galactic dust component is much more stronger than the CMB radiation, on the Galactic plane, both at 70 GHz (top) and at 100 GHz (bottom).

ICA and Particle Filtering on the Galactic Plane

After some experimentation and testing, a satisfactory set of values for the drift parameters has been found. This example shows the comparison between an ICA-based estimation and the output of the particle filter. The sources to be separated and the mixtures are those which have been described in the previous example; the particle filter has been forced to approximate the distribution of the dust signal by means of a combination of 3 Gaussian components, sampling 70 particles for each parameter of interest, for each pixel.

In figure 15 we show that the ICA approach is unsuitable for this separation problem: in fact, in case of low SNR, the estimate of the Cosmic Microwave Background radiation consists of a noise realization only, as it can be seen easily by comparing this ICA estimate with the noise pattern shown in figure 11.

On the other hand, in the particle filter output it is possible to see the signal, which is also certified by the SIR (*Signal to Interference Ratio*) values (figure 15):

$$\mathrm{SIR}(\mathrm{CMB})_{\text{part. filt.}} = -0.32\mathrm{dB} \quad > \quad \mathrm{SIR}(\mathrm{CMB})_{\mathrm{ICA}} = -6.02\mathrm{dB}$$

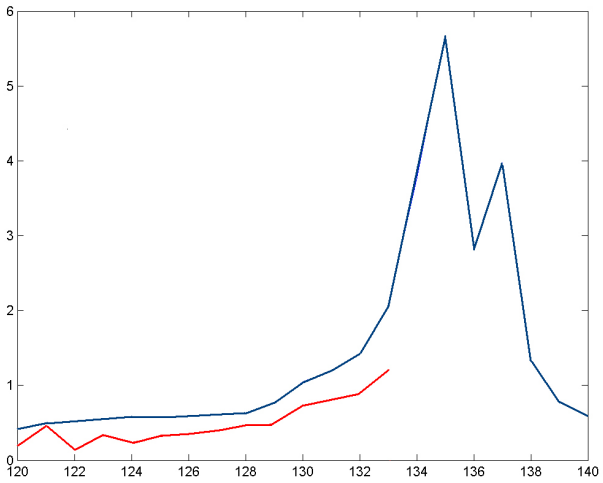

Figure 14: Simulations - Plot of the one-dimensional signal representation of one column of the Galactic dust image (blue), in correspondence of the Galactic plane. In case of an uncorrect choice of drift parameters and initial source distribution, the algorithm (red) is unable to recognize a sudden variation of the statistical properties of one of the source signals.

However, the output is very noisy and the separation is not completely satisfactory.

Scanning the Images

We experienced detrimental sudden variations in the source vectors, due to a non-optimal conversion of the data matrices to one-dimensional vectorial representations. In fact, while scanning an image column by column, from top to bottom, it is possible that the first pixel of a column has a completely different value from the first pixel of the previous column, so that the particle filter loses the track of the sources to be estimated. In the following simulation, matrices are converted into vectors in order to keep the correlation between adjacent pixels, by scanning the images column by column, one column from the top to the bottom, and the following one from the bottom to the top, and so on.

CMB and Synchrotron Radiation on the Galactic Plane

Another simulation has been done with the aim of separating the CMB and the synchrotron radiation on the Galactic plane (figure 17), exploiting two mixtures observed at 70 GHz and 100 GHz (figure 18), with a SNR = 10 dB. The particle filter algorithm has been forced to approximate the distribution of the synchrotron radiation with 4 Gaussian components.

The FastICA algorithm estimates the CMB more satisfactorly than the particle filter, but is not able to recover the synchrotron signal (figure 19). The particle filter recovers both the signals, even if the estimates are still noisy (figure 20).

$$\begin{aligned} \mathrm{SIR}(\mathrm{CMB})_{\text{part. filt.}} = 8\mathrm{dB} &< \mathrm{SIR}(\mathrm{CMB})_{\mathrm{ICA}} = 10\mathrm{dB} \\ \mathrm{SIR}(\mathrm{SYNCHR})_{\text{part. filt.}} = 1.64\mathrm{dB} &> \mathrm{SIR}(\mathrm{SYNCHR})_{\mathrm{ICA}} = -7.74\mathrm{dB} \end{aligned}$$

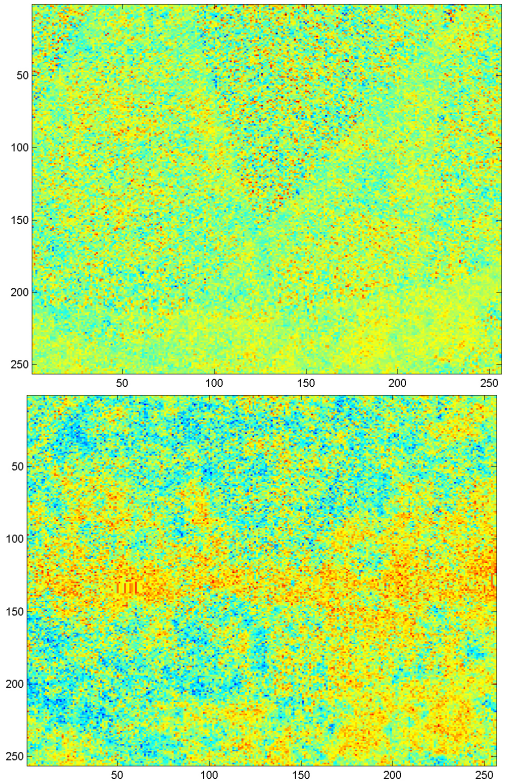

Figure 15: Simulations - CMB estimates: ICA (top) and particle filter (bottom).

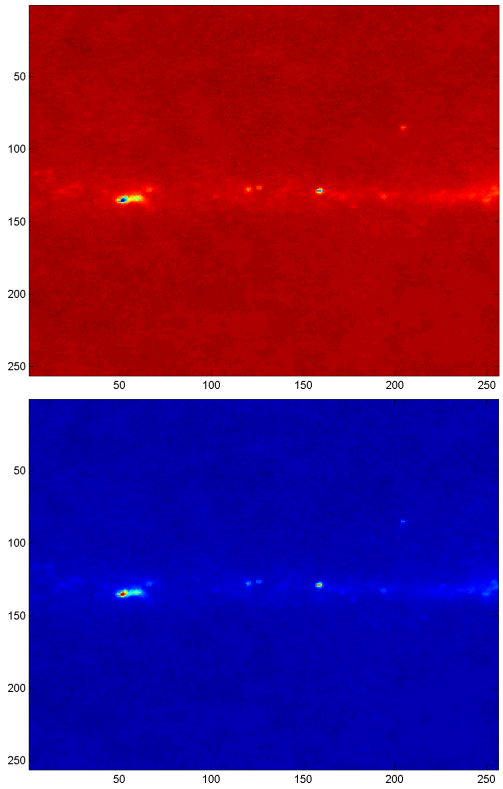

Figure 16: Simulations - Galactic dust estimates: FastICA (top) and particle filter (bottom).

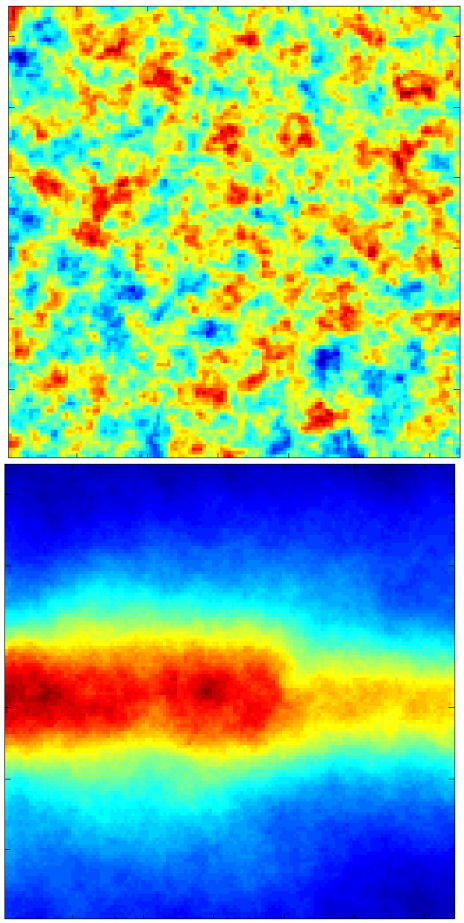

Figure 17: Simulations - The Cosmic Microwave Background (top) and the synchrotron radiation (bottom) on the Galactic plane.

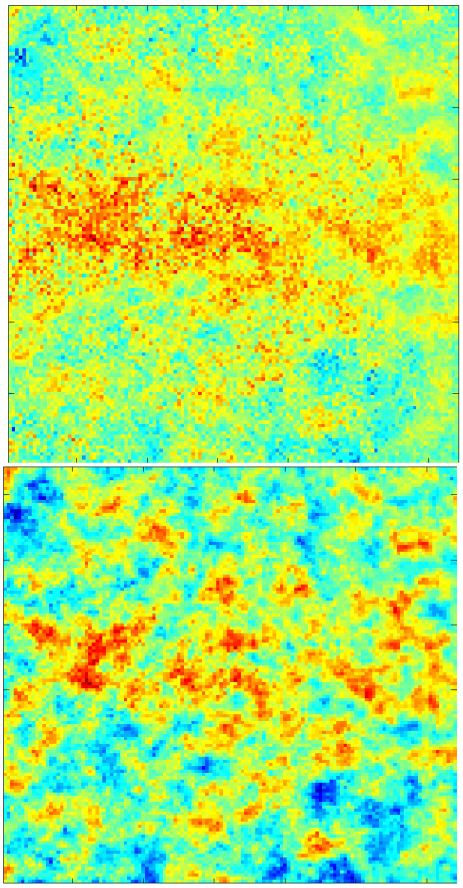

Figure 18: Simulations - The observed images at 70 GHz (top) and 100 GHz (bottom).

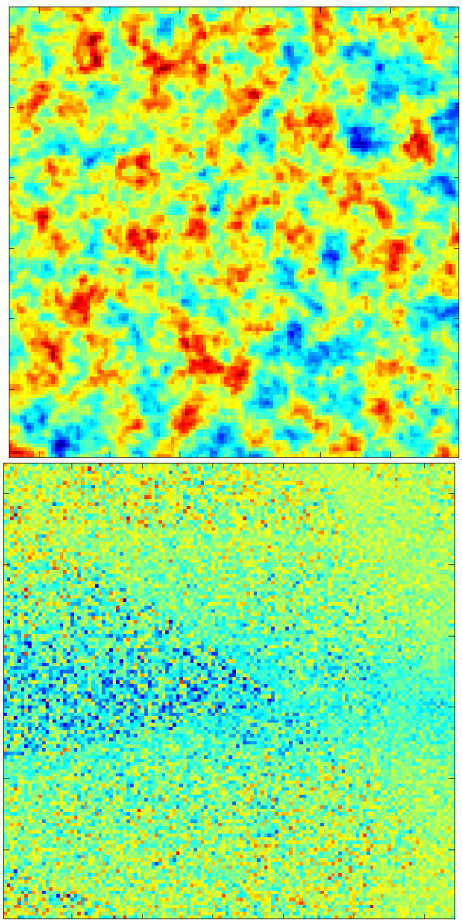

Figure 19: Simulations - ICA estimates of CMB (top) and the synchrotron radiation (bottom) on the Galactic plane.

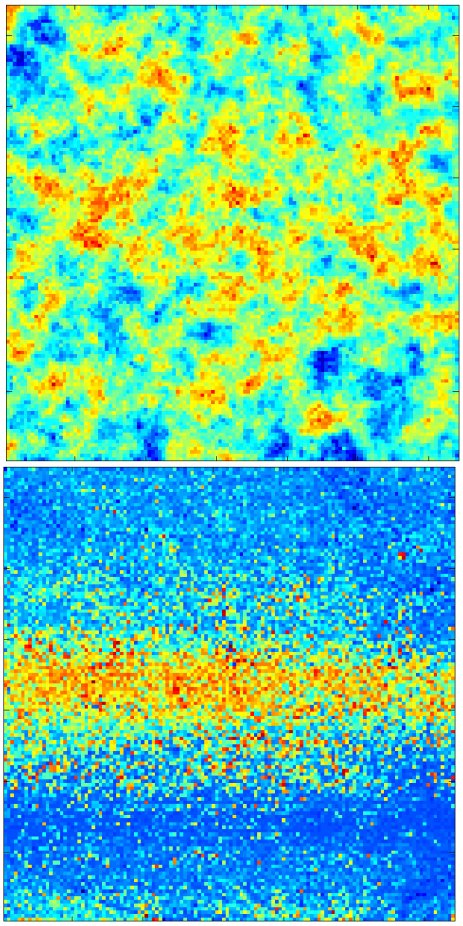

Figure 20: Simulations - Output images of the particle filter for CMB (top) and synchrotron radiation (bottom).

CMB and Synchrotron Radiation outside the Galactic Plane

The algorithm has been tested also with mixtures of CMB and synchrotron radiation, far from the Galactic plane (figure 21). Five Gaussian components are used to approximate the synchrotron posterior distribution, and 70 particles are generated at each step, for each parameter of interest.

In this case, the CMB radiation is predominant, as it can be seen in figure 22. When the

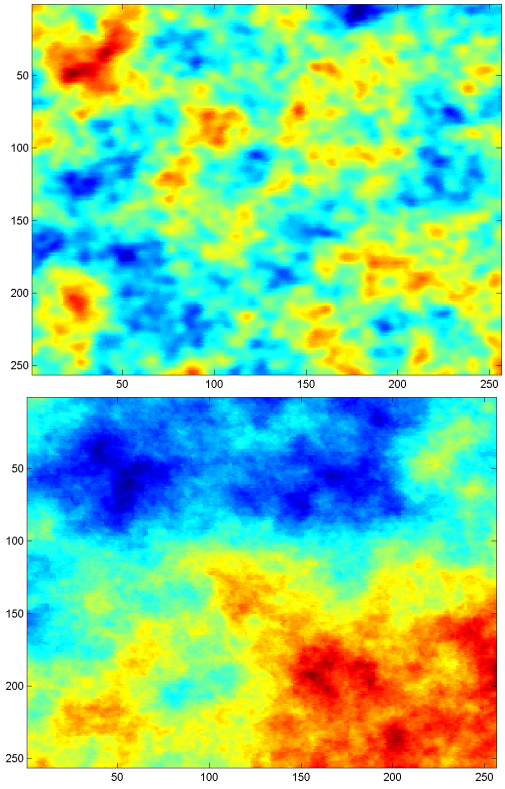

Figure 21: Simulations - CMB (top) and synchrotron radiation (bottom) outside the Galactic plane.

SNR is high, ICA and particle filtering have the same performance. In case of low SNR, the FastICA algorithm cannot recognize the contribution of the synchrotron radiation, but it can reconstruct the CMB signal with good accuracy, as it is shown in figure 23. The particle filter provides the two output signals by exploiting the a-priori information concerning the distribution of the two sources, that is the starting values of the means and the variances of each Gaussian component, while initial random distributions are chosen for the indicator matrix $\mathbf{z}_{1:m,0:t}$ and for the transition matrix $\{\boldsymbol{\pi}_{0:t}^{(i)}\}$.

The signal-to-interference ratios for both the CMB and the synchrotron radiation are reported below:

$$\begin{aligned} \text{SIR}(\text{CMB})_{\text{part. filt.}} = 10.88\text{dB} &> \text{SIR}(\text{CMB})_{\text{ICA}} = 10.55\text{dB} \\ \text{SIR}(\text{SYNCHR})_{\text{part. filt.}} = 6.9\text{dB} &> \text{SIR}(\text{SYNCHR})_{\text{ICA}} = -18\text{dB} \end{aligned}$$

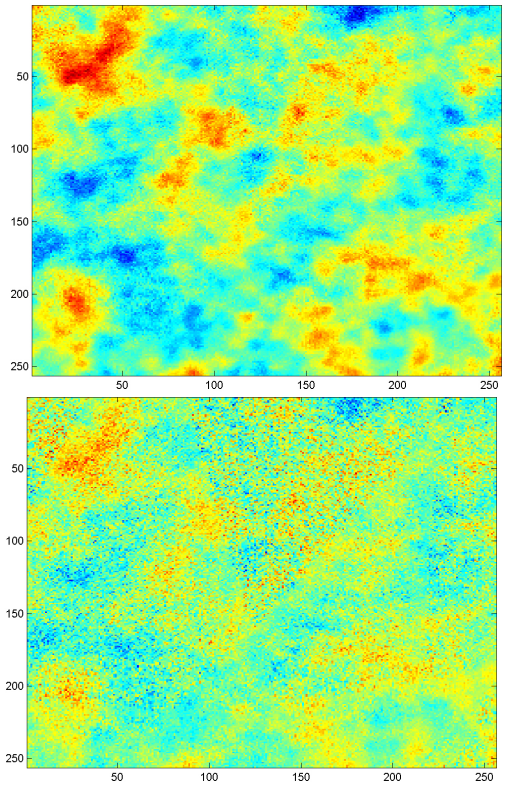

Figure 22: Simulations - Observed mixtures of CMB and synchrotron radiation in as sky patch outside the Galactic plane, at 100 GHz (top) and 70 GHz (bottom).

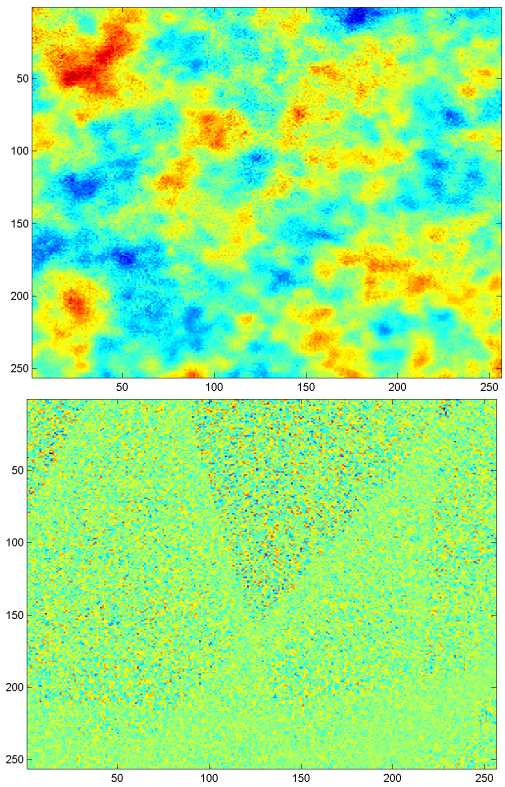

Figure 23: Simulations - ICA estimates of CMB (top) and synchrotron (bottom).

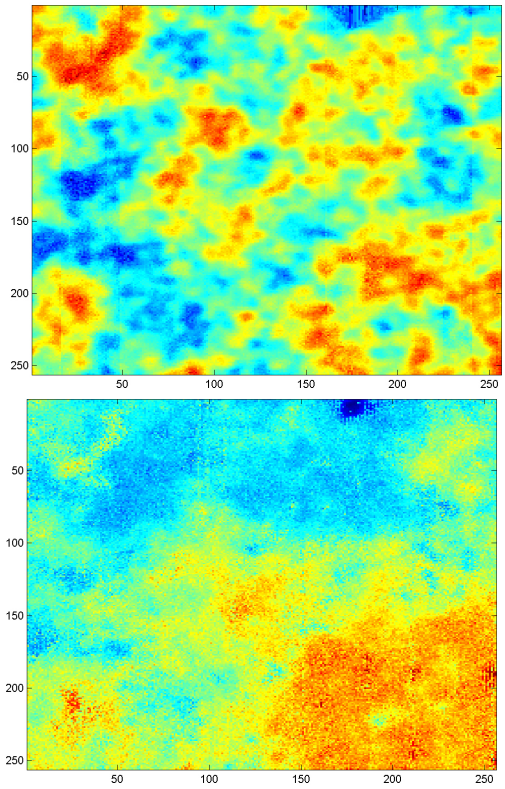

Figure 24: Simulations - Output signals of the particle filter, related to CMB (top) and synchrotron (bottom).

5 Results and Future Research

In this work we have introduced and implemented a new, general approach, named *particle filtering*, to solve the source separation problem in the astrophysical context. This method is able to deal with non-Gaussian and non-linear models, non-stationary sources and space-varying noise. The extreme flexibility of particle filters allows to include all the available a-priori information about the statistical properties of the sources to be separated, together with the available a-priori information about noise, and, in case, about the mixing matrix.

In the previous section we demonstrated that the particle filter provides better results in comparison with one of the most widespread methods to perform source separation (ICA), especially in case of low SNR. We also showed that it is possible to improve the filtering performance by using a greater number of particles, and allowing more Gaussian components to approximate each source distribution. Moreover, it must be considered the fact that, as we have already stressed, particle filtering is a highly flexible method: it will be possible to obtain better results with more sperimentation, by changing importance function, trying different values for the drift paramenters, or using different selection algorithms.

5.1 Importance Function

A fundamental step in the implementation of the particle filter algorithm is the choice of the importance function, as it has been explained before: the use of the prior distribution as the important fuction has been already discussed, and presented as an appealing solution to overcome the analytic untractability of the optimal importance function. Unfortunately this choice gives results which are not fully satisfactory, because this importance function does not allow us to exploit any information about the observed data. Better results may be obtained adopting different choices of the importance function, in order to include the information derived from the statistical distribution of the observations.

5.2 Drift Parameters

A set of constant drift parameters has been used to perform the simulations shown in the previous section. However, finding optimal values for these parameters is a very difficult task: a trade-off has to be considered when choosing the drifts, as too small ones imply an unproper exploration of the distribution of interest, while a high drift implies the fact that also the variance of the MMSE estimates of the sources will be high. An alternate strategy would be to initialise the drift parameters to high values, and let these values decrease gradually in magnitude. However, the drift cannot become too small, because a sudden change of the statistical properties of one of the sources could not be recognized by the algorithm, when it occurs. A good solution could be the use of drift parameters which are able to adapt their magnitude step by step, exploiting the information coming from the observations.

5.3 Selection Step

The results of our simulations have been obtained by employing the standard residual resampling procedure to implement the selection step. While debugging the code, we experienced that only a small percentage of the generated particles have significant importance weights, which means that this procedure does not avoid completely the degeneracy phenomenon. Future research should focus on alternative implementations of the selection step by means of different resampling procedures.

5.4 Two-Dimensional Data

Another key-point to work on is the exploitation of the spatial correlation between pixels: in the algorithm presented here, each image has been converted into a vector and then treated as one-dimensional. To exploit the spatial information completely, we should extend our data model to two dimensions taking into account both vertical and horizontal neighbours, which can be achieved utilizing Markov Random Fields (MRF) [40].

References

[1] Ahmed A.: "Signal Separation",
PhD Thesis, Signal Processing Group, Department Of Engineering, University Of Cambridge, U.K. (2000)
http://www-sigproc.eng.cam.ac.uk/publications/archive/AlijahAhmedThesis.zip

[2] Amari S. I., Cichoki A.: "Adaptive Blind Signal Processing - Neural Network Approaches" (1998). *Proceedings of IEEE, 86*, 2026-2048.

[3] Amendola L., Corasaniti P. S.: "Il codice genetico dell'Universo" (2001). *L'Astronomia, 216*, 18-29.

[4] Anderson B. D., Moore J. M.: "Optimal Filtering" (1979). *Prentice-Hall, New Jersey.*

[5] Andrieu C., Godsill S.J.: "A Particle Filter for Model Based Audio Source Separation (2000)". *International Workshop on Independent Component Analysis and Blind Signal Separation, ICA 2000, Helsinki, Finland.*

[6] Arulampalam M. S., Maskell S., Gordon N., Clapp T.: "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking" (2002). *IEEE Transactions on Signal Processing, Vol. 50, No.2,*, 174-188.

[7] Attias H.: "Independent Factor Analysis" (1999). *Neural Computation, 11*, 803-851.

[8] Baccigalupi C, Bedini L., Burigana C., De Zotti G., Farusi A., Maino D., Maris M., Perrotta F., Salerno E.: "Neural Networks and the Separation of Cosmic Microwave Background and Astrophysical Signals in Sky Maps" (2000). *Monthly Notices of the Royal Astronomical Society, 318*, 769-780.

[9] Bell A. J., Sejnowski T. J.: "An Information-Maximization Approach to Blind Separation and Deconvolution" (1995). *Neural Computation, 7*, 1129-1159.

[10] Bernardo J. M., Smith A. F. M.: "Bayesian Theory" (1995). *John Wiley and Sons.*

[11] Boersma P., Weenink D.: "Principal Component Analysis"
http://www.fon.hum.uva.nl/praat/manual/
Principal_component_analysis.html

[12] Cardoso J. F.: "Looking for Components in the Universe's Oldest Data Set" (2002). *Proc. EUSIPCO 2002*, http://tsi.enst.fr/∼cardoso/Papers.PDF/eusipco02.oldest.pdf.

[13] Casella G., Robert C. P.: "Monte Carlo Statistical Methods" (1999). *Springer Texts in Statistics.*

[14] Comon P.: "Independent Component Analysis, a New Concept?" (1994). *Signal Processing, 36*, 287-314.

[15] Costagli M.: "Source Separation of Astrophysical Images Using Particle Filters: Matlab Codes - Version 1.0" (2003). *ISTI-CNR, Pisa, Internal Note B4-26 2003.*

[16] Doucet A.: "Monte Carlo Algorithms for Bayesian Estimation of Hidden Markov Models" (1997). *PhD Thesis, University Paris-Sud, Orsay.*

[17] Doucet A., De Freitas J. F. G., Gordon N. J.: "Sequential Monte Carlo Methods in Practice" (2001). *Springer-Verlag.*

[18] Doucet A., Godsill S., Andrieu C.: "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering" (2000). *Statistics and Computing, 10*, 197-208;

[19] Everson R. M., Roberts S. J.: "Particle Filters for Non-stationary ICA" (2000). *Advances in Independent Components Analysis, M. Girolami (Ed.) 23-41, Springer.*

[20] Geweke J.: "Bayesian Inference in Econometric Models Using Monte Carlo Integration" (1989). *Econometrica, 24*, 1317-1399.

[21] Gordon N. J., Salmond D. J., Smith A. F. M.: "Novel Approach to Non-linear / Non-Gaussian Bayesian State Estimation" (1993). *IEE-Procedings-F, 140*, 107-113.

[22] Handschin J. E., Mayne D. Q.: "Monte Carlo Techniques to Estimate the Conditional Expectation in Multi-Stage Non-linear Filtering" (1969). *International Journal of Control, 9*, 547-559.

[23] Hollmen J.: "Process Modelling Unsing the Self-organizing Map" (1996) *Master Thesis, Department of Computer Science, Helsinki University of Technology, Finland*, http://www.cis.hut.fi/∼jhollmen/dippa/node30.html.

[24] http://astro.estec.esa.nl/planck/: "The Home Page of Planck".

[25] http://lambda.gsfc.nasa.gov/product/cobe/: "Lambda - COBE Mission Overview".

[26] Hyvärinen A.: "Independent Component Analysis: a Tutorial" (1999). hhtp://www.cis.hut.fi/projects/ica/.

[27] Hyvärinen A.: "Survey on Independent Component Analysis" (1999). http://www.cis.hut.fi/aapo/papers/NCS99web/NCS99web.html.

[28] Hyvärinen A., Oja E.: "A Fast Fixed-point Algorithm for Independent Component Analysis" (1997). *Neural Computation, 9 (7)*, 1483-1492 .

[29] Julier S. J., Uhlmann J. K.: "A New Extension of the Kalman Filter to Nonlinear Systems" (1997). *Proceedings of AeroSence: the 11th International Symposium on Aerospace/Defence Sensing, Simulation and Controls, Vol. Multi Sensor Fusion, Tracking and Resource Management II.*

[30] Kalman R. E.: "A New Approach to Linear Filtering and Prediction Problems" (1961). *Transactions of the ASME, Journal of Basic Engineering, 83*, 95-107.

[31] Kitagawa G.: "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models" (1996). *Journal of Computational and Graphical Statistics, 5*, 1-25.

[32] kurp-www.hut.fi/quasar/ planck/yksi.htm: "Planck".

[33] Kuruoğlu E. E., Bedini L., Paratore M. T., Salerno E., Tonazzini A.: "Numerical Experiments on Blind Separation of Astrophysical Maps by Independent Factor Analysis" (2002). *IEI-CNR, Pisa, Technical Report, 2002-TR-03.*

[34] Kuruoğlu E. E., Bedini L., Paratore M. T., Salerno E., Tonazzini A.: "Source Separation in Astrophysical Maps Using Independent Factor Analysis" (2003). *Neural Networks, 16*, 479-491.

[35] Liu J. S., Chen R.: "Sequential Monte Carlo Methods for Dynamic Systems" (1998). *Journal of the American Statistical Association, 93*, 1032-1044.

[36] Maino D., Farusi A., Baccigalupi C., Perrotta F., Banday A. J., Bedini L., Burigana C., De Zotti G., Grski K. M., Salerno E.: "All-Sky Astrophysical Component Separation with Fast Independent Component Analysis (FastICA)" (2002). *Monthly Notices of the Royal Astronomical Society, 334*, 53-68.

[37] Moulines E., Cardoso J. F., Gassiat E.: "Maximum Likelihood for Blind Separation and Deconvolution of Noisy Signals Using Mixture Models" (1997). *Proceedings of ICASSP '97, 5*, 3617-3620.

[38] Pitt M., Shephard N.: "Filtering via Simulation: Auxiliary Particle Filter" (1999). *Journal of the American Statistical Association, 94*, 590-599.

[39] Pope K. J., Bogner R. E.: "Blind Signal Separation. 1: Linear, Instantaneous Combinations" (1996). *Digital Signal Processing, 6*, 5-16.

[40] Tonazzini A., Bedini L., Kuruoğlu E. E., Salerno E.: "Blind Separation of Auto-Correlated Images from Noisy Mixtures Using MRF Models" (2003). *4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003) - Nara, Japan, April 2003. Proceedings*, 675-680.

[41] Van der Merwe R., Doucet A., De Freitas N., Wan E.: "The Unscented Particle Filter" (2000). *Advances in Neural Information Processing Systems, 13*, 584-590.

[42] Vielva Martinez P.: "Detection of Point Sources on Cosmic Microwave Background Maps" (2002).. *PhD Thesis, Departamento de Fisìca Moderna / Instituto de Fisìca de Cantabria - Consejo Superior de Investigaciones Cientìficas, Universidad de Cantabria, Spain.*