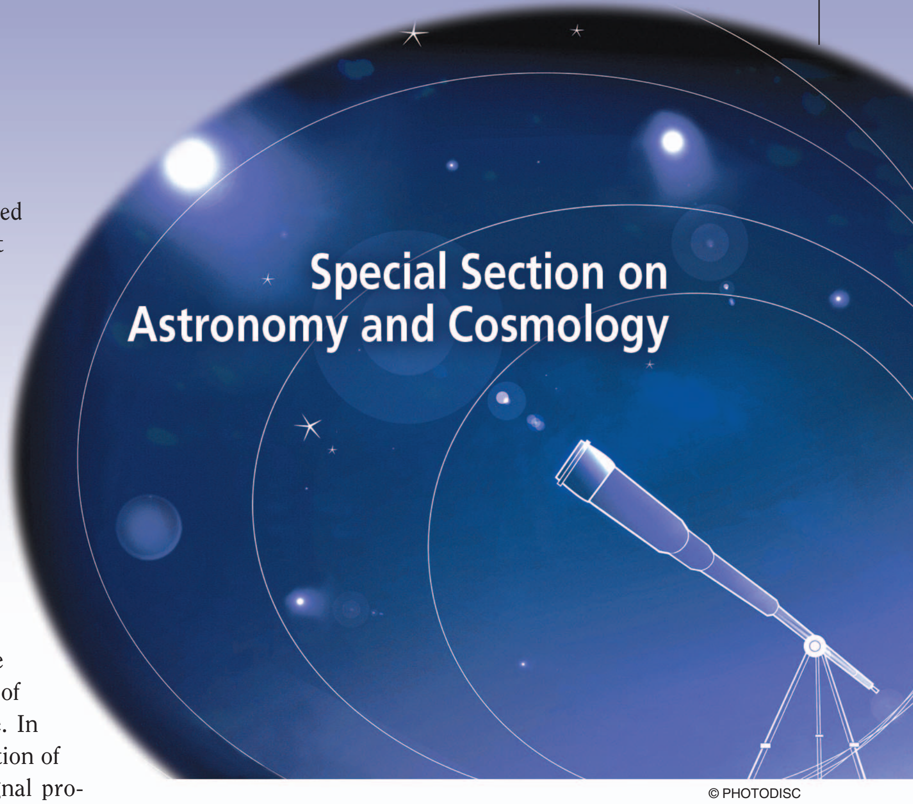
Ercan E. Kuruoglu



© PHOTODISC

Bayesian Source Separation for Cosmology

Estimating cosmological components from multichannel measurements

Recent satellite missions have provided and continue to provide us with vast amounts of data on radiation measurements that generally present themselves as superpositions of various cosmological sources, most importantly cosmic microwave background (CMB) radiation and other galactic and extragalactic sources. We would like to obtain the estimates of these sources separately since they carry vital information of cosmological significance about our Universe. Although initial attempts to obtain sources have utilized blind estimation techniques, the presence of important astrophysical prior information and the demanding nature of the problem makes the use of informed techniques possible and indispensable. In this article, our objective is to present a formulation of the problem in Bayesian framework for the signal processing community and to provide a panorama of Bayesian source separation techniques for the estimation of cosmological components from the observation mixtures.

INTRODUCTION

One of the most important discoveries of the past century was undoubtedly the 1964 observation of CMB radiation by Penzias and Wilson. Their accidental discovery gave the much-searched for proof of the hot big bang theory, which was predicted by Gamov in 1948.

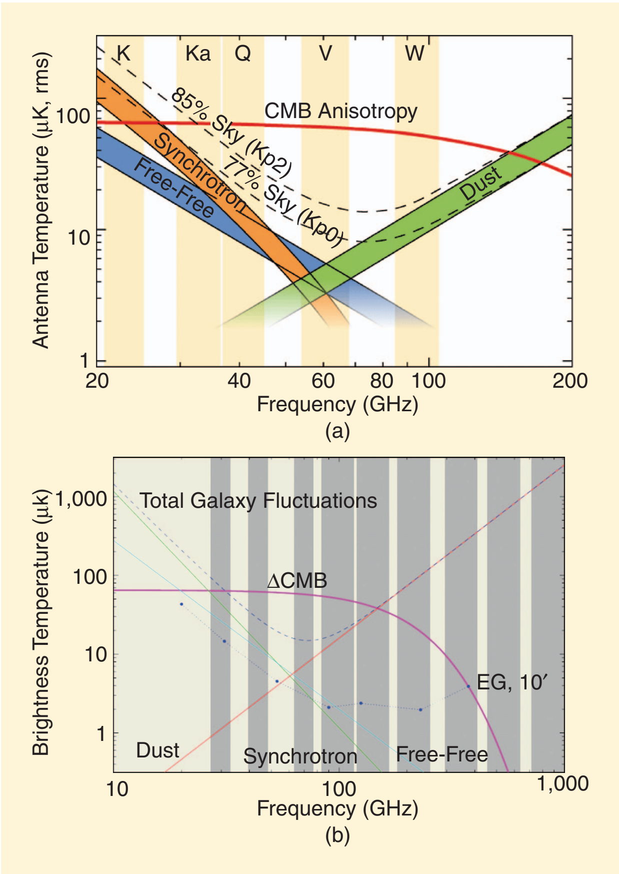
The hot big bang model, which aims to provide an explanation for the formation of our Universe, has the main thesis that the Universe evolved into its current state expanding and cooling from an initially much denser and hotter Universe. According to this theory, the very early Universe was characterized by a very high energy density and very high temperatures and pressures and was rapidly expanding and cooling. Approximately 10^{-37} s into the expansion, a phase transition caused a cosmic inflation, during which the Universe grew exponentially. Matter was dominating the Universe and radiation was trapped in matter. The Universe continued to grow in size and to cool down, hence the typical energy of each particle continued to decrease. A few minutes into the expansion, as the

Digital Object Identifier 10.1109/MSP.2009.934718

Universe cooled, the rest of the mass energy density of matter started to dominate that of the photon radiation. After about 379,000 years, the electrons and nuclei combined into atoms (mostly hydrogen); hence the radiation decoupled from matter and was freely released through space. CMB radiation is this radiation when light was freed for the first time.

ONE OF THE MOST IMPORTANT DISCOVERIES OF THE PAST CENTURY WAS UNDOUBTEDLY THE 1964 OBSERVATION OF CMB RADIATION BY PENZIAS AND WILSON.

Initially, this radiation had a temperature of about 3,000 K but as the Universe continued to expand, the temperature dropped. Today, about 14 billion years later, CMB is still detectable but at a much lower temperature, around 2.7 K. Like fossil remains giving us a chance to discover now extinct species, this relic radiation carries immensely important information about the Universe, and therefore it is of tremendous importance to make a full-sky measurement of CMB. First, it is the picture of the Universe shortly after it has started and houses vital information on the early Universe. Second, it has something to tell about the current Universe: the inhomogeneities in the CMB are the seeds of the galaxies that exist today, and it provides us information about the formation of the galaxies and the topology of the Universe. Third, CMB not only informs us about the past and present but also about the future of the Universe. The precise measurement of CMB and the calculation of angular spectrum and certain cosmological constants from it will enable us to choose between competing theories for the future evolution of our Universe. In particular, we will be able to tell whether the Universe will continue to expand, ending up in a zero-entropy state, whether its expansion will stop and the dimension will stabilize, or whether it will be followed by a phase of shrinking that will end with a "big crunch."



[FIG1] Frequency bands of (a) WMAP, (b) Planck satellites, and cosmological sources [4], [6]. Figures used with permission.

To be able to answer such important questions and others, many attempts have been made to measure the CMB accurately. These attempts range from ground-based measurements such as the degree angular scale interferometer (DASI) [1] to balloon missions such as balloon observations of millimetric extragalactic radiation and geophysics (BOOMERANG) [2] and to satellite missions such as cosmic background explorer (COBE) [3] and the Wilkinson microwave anisotropy probe (WMAP) [4]. The COBE satellite mission of NASA in 1992 has detected small-amplitude fluctuations or anisotropies in otherwise flat-looking CMB. Since the anisotropies in CMB are the main information carrying part, it is vital to recover them in high resolution. Considering COBE's limited angular resolution (7°), it became evident that more accurate measurements were needed. The WMAP [4] was launched in 2001 with the aim of obtaining much higher-resolution maps of the Universe (0.23–0.93° angular resolution) in five frequency bands between 22–100 GHz. Finally, another satellite mission, the Planck surveyor, [which will provide images with even higher resolution (between 0.083–0.55°) and higher sensitivity on a wider frequency span (nine frequency bands between 30–857 GHz)], was launched on 14 May 2009 by the European Space Agency [5]. The frequency bands in which WMAP and Planck operate can be seen in Figure 1.

An important problem is that the signals measured by these satellites do not contain radiation only from CMB but also contributions from a number of other sources, namely foreground radiations and extragalactic sources in addition to antenna receiver noise. Foreground sources emission from our galaxy includes synchrotron, dust, and free-free emission. Extragalactic sources include point sources and Sunyaev-Zeldovich (SZ) clusters.

The recovery of the CMB signal from this mixture of various sources is a major task and is the subject of this article. One main approach is to consider all signals other than CMB as contaminants and try to recover only CMB. Another approach, which will be followed here, is to try to recover all sources separately and make a picture of other sources as well as CMB. The motivation of this approach is that the other sources also carry important information about the Universe, and in particular about our galaxy.

Various works over the last decade have addressed the problem in a source-separation framework. Many early publications on the problem adopted a blind approach and employed techniques such as independent component analysis. Such techniques assume no prior information about the sources, their distributions, spectral indices, or the noise other than that the mixing is linear and that the noise is Gaussian. Blind estimation also does not lend itself easily to error analysis. However, the cosmological problem under investigation has a wealth of prior information. Many researchers assume the Gaussianity of the CMB; we have very good knowledge about the antenna receiver noise and the beam pattern; the mixing parameters are dictated by the spectral indices, about the range of which we have a good idea from previous measurements, and so on.

CMB NOT ONLY INFORMS US ABOUT THE PAST AND PRESENT BUT ALSO ABOUT THE FUTURE OF THE UNIVERSE.

A mathematical framework is needed for incorporating this rich prior information into the formulation of the problem. This framework is provided successfully by Bayesian analysis that has the following advantages over blind methods:

1) The prior information is formulated with a probability density which conveniently modifies the likelihood so that the solution is "biased" to more probable areas of the solution space and improbable areas are avoided. This provides important savings in the search for a solution.

2) Bayesian estimation is an excellent learning model that mimics our learning mechanism. Bayesian estimation can be easily cascaded. With the arrival of new observations, we can simply update our previous analysis with the new likelihood.

3) Even when we do not have prior information, utilizing the Bayesian framework provides us with means to perform error analysis easily.

In contrast to blind source separation, the Bayesian framework provides us with means of informed source separation. The demanding nature of the physics of the problem has stimulated the incorporation of prior information in the mathematical formulation and the design of more and more elaborate techniques. In that sense, the cosmology problem have presented an ideal ground for the advance of source separation methodology. Current state-of-the-art techniques involve numerical Bayesian techniques and multilayer hierarchical statistical models. With the advance of the available methodology, issues that were not considered before such as the nonstationarity of the mixing, convolutive effects of the beam, and the statistical dependence between sources is also addressed, and further developments in the methodology are stimulated.

In this tutorial, we aim to present an easy-to-grasp Bayesian formulation and give a wide account of existing work in the literature. Due to the nature of the problem, many exciting new methodological directions are still under investigation and the article therefore is incomplete, however, we address some yet inconclusive areas of research and try to project new challenges in the data to excite more research in the field.

We hope that the signal processing audience will not only find the cosmological problem interesting to read but will also enjoy following the development of novel source separation methodologies provoked by a scientific problem.

PROBLEM STATEMENT

We start by giving a simple observation model adopted widely in the literature.

It is widely assumed [7], [8] that each radiation process $\tilde{\mathrm{s}}_n(\xi, \eta, \nu)$ from the microwave sky has a spatial pattern $\mathrm{s}_n(\xi, \eta)$ that is independent of its frequency spectrum $F_n(\nu)$

$$\tilde{\mathrm{s}}_n(\xi, \eta, \nu) = \mathrm{s}_n(\xi, \eta) F_n(\nu), \;\; n = 1, \ldots, N, \tag{1}$$

where ξ and η are angular coordinates on the celestial sphere and ν is frequency. The total radiation observed in a certain direction at a certain frequency is given by the sum of a number N of signals (processes, or components) described in (1), where subscript n is the process index. Assuming that the satellite observing beam in each measurement channel at the specified frequency is spatially invariant, we may write the beam-smoothing as a convolution and the observed signal at M different frequencies can be modeled as

$$\mathrm{X}(\xi, \eta) = \mathrm{P} \otimes \mathrm{A}\, \mathrm{S}(\xi, \eta) + \mathrm{E}(\xi, \eta), \tag{2}$$

where $\mathrm{X} = \{\mathrm{x}_m, m = 1, \ldots, M\}$ is the $M \times J$-matrix of the observations, with $\mathrm{x}_m = \{x_{m,j}, j = 1, \ldots, J\}$ being the observation in channel m (with central frequency ν_m) over J pixels. $\mathbf{A}$ is an $M \times N$ mixing matrix, $\mathrm{S} = \{\mathrm{s}_n, n = 1, \ldots, N\}$ is the $N \times J$-matrix of the individual source processes, and $\mathrm{E} = \{\mathrm{e}_m, m = 1, \ldots, M\}$ is the $M \times J$-matrix of instrumental noise. The elements of the matrix $\mathbf{P}$ give the beam profile at each channel frequency. The elements of $\mathbf{A}$ are related to the source spectra through the following formula:

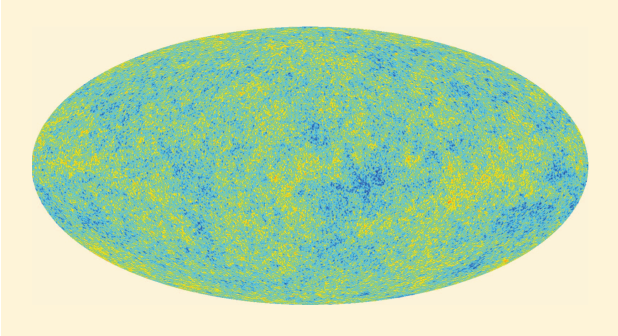
$$a_{mn} = \int F_n(\nu) b_m(\nu) d\nu, \tag{3}$$

where $b_m(\nu)$ is the instrumental frequency response in the mth measurement channel, which is known. If we assume that the source spectra are constant within the passbands of the different channels, (3) can be rewritten as

$$a_{mn} = F_n(\nu_m) \int b_m(\nu) d\nu. \tag{4}$$

The element a_{mn} is thus proportional to the spectrum of the nth source at the center frequency ν_m of the mth channel.

It has been preferred to work in the spatial frequency domain by several researchers [7], [9]–[11] due to a number of factors: the beam effects turn into a simple multiplication rather than convolution, it makes it easier to deal with varying resolution of each channel, and most importantly, it is very suitable to work with CMB, the angular spectrum of which is of central



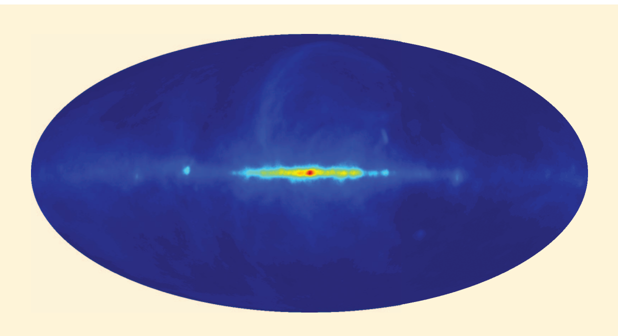
[FIG2] A simulation of CMB as will be seen by Planck [6]. (Courtesy of ESA.)

importance. In addition to these advantages, some techniques [12], [13] specifically exploit the spectral diversity of the sources.

However, there is an important counter argument against working in the frequency domain: although CMB is stationary (or homogeneous and isotropic in cosmology terminology), the foregrounds and the antenna receiver noise are highly nonstationary. The spectral indices of the foregrounds, in particular of the galactic dust, is spatially varying. This renders the power spectrum for these sources not well defined. This problem was underlined in [10], where the authors suggested making a Taylor expansion of the intensity and using the first few terms. However, the number of terms that one can use is limited such that the total number of fields to be reconstructed can not exceed the number of frequencies at which observations are made. Increased number of terms also increase the problem size and the computational complexity. Such complications rather diminish the advantages of working in the frequency domain and one can argue that one might as well work in the spatial domain. In the rest of this article, we will make the derivations in the spatial domain and will not consider the beam effects. Hence, we consider the following observation model:

$$\mathrm{X} = \mathrm{AS} + \mathrm{E}, \tag{5}$$

where X, A, S, and E are defined as above and the observation model consists of microwave amplitudes at M frequencies $(\nu_1, \ldots, \nu_M)$ over the sky at J pixels.



[FIG3] The Haslam 408 GHz map that is frequently used as a synchrotron template [4]. (Courtesy of NASA.)

COSMOLOGICAL COMPONENTS

CMB

The most important source among those that make up the celestial microwave radiation is the cosmic microwave radiation and is the principal objective of the WMAP and Planck satellite missions. It is widely accepted that CMB is Gaussian [14], [15] although recently there have been debates on the deviation of CMB from Gaussianity. It is also widely expected to be stationary, as validated on WMAP data in [16], and that CMB anisotropies can be represented as the multiplication of a spatial template with a nonlinear function of the frequency.

The CMB sources used in various work in the literature are generated synthetically using the CMBFAST software package, which assumes a spatially flat standard inflationary cold dark matter model with a Gaussian realization.

The emission spectrum of the CMB is perfectly known, being a blackbody radiation [17]. In terms of antenna temperature, it is

$$F_{\mathrm{cmb}}(\nu) = \frac{\bar{\nu}^2 \exp \bar{\nu}}{[\exp(\bar{\nu}) - 1]^2}, \tag{6}$$

where $\bar{\nu} = h\nu / kT_{\mathrm{CMB}}$ is the normalized frequency, h is Planck constant, and k is Boltzmann's constant. Figure 2 shows a simulation of CMB as will be seen by Planck.

GALACTIC COMPONENTS

SYNCHROTRON

Synchrotron radiation is generated by the electrons spiraling (hence accelerating) along magnetic fields. Although synchrotron radiation originates in the galaxy, it extends also to outside the galactic plane and is less concentrated in the galactic plane when compared to other galactic foreground components. The synchrotron map commonly used for simulations in the literature is provided in the 408 MHz Haslam survey (see Figure 3) and extrapolated and scaled to Planck frequencies [18]. A large number of patches from this map were analyzed in [19], and it was shown that a Gaussian mixture with three to five components describes the histogram of the patch intensities with small error.

The synchrotron emission can be modeled by a power law over a wide range of frequencies [17], that is,

$$F_{\mathrm{syn}}(\nu) = B\left(\frac{\nu}{\nu_{0,\mathrm{syn}}}\right)^{\beta_s}, \tag{7}$$

where β_s is the synchrotron spectral index, B and $\nu_{0,\mathrm{syn}}$ are normalization factors. Various work on the surveys carried out on different frequency bands place the synchrotron spectral index in $-3.2 < \beta_s < -2.3$.

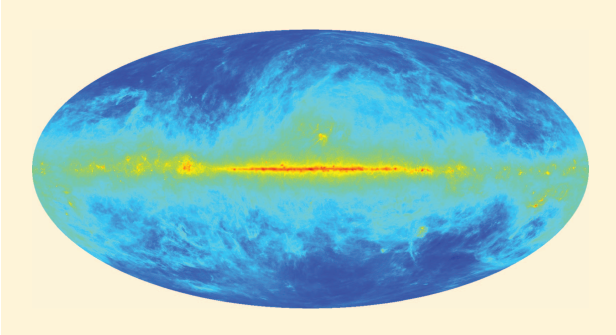
GALACTIC DUST

Galactic dust is made up of small particles that range from the order of nanometers to micrometers. They are made of various materials including silicate and carbon. The shapes vary and they can be of crystalline or amorphous structure. Their radiation is dominant especially in very high-frequency channels. Assuming a grey-body spectrum [17],

$$F_{\text{dust}}(\nu) = \frac{\overline{\nu}^{\beta_d+1}}{\exp(\overline{\nu}-1)}, \tag{8}$$

where $\overline{\nu} = h\nu/kT_d$ is the normalized frequency, h is Planck constant, k is Boltzmann's constant, and T_d is the physical dust temperature.

A frequently used galactic dust template is given in Figure 4.



[FIG4] **Predicted dust emission at 94 GHz by Finkbeiner et al. (1999) [4]. (Courtesy of NASA.)**

FREE-FREE EMISSION

Free-free emission or "Bremsstrahlung" emission is caused by the collision of free electrons with heavy ions in the ionized medium. Electrons lose energy in these collisions and emit photons. The emission can be described with [17]

$$F_{\text{free}}(\nu) = A_{\text{free}}\left(\frac{\nu}{\nu_{0,\text{free}}}\right)^{-2.14}. \tag{9}$$

The free-free emission as estimated by the WMAP consortium is given in Figure 5.

EXTRAGALACTIC SOURCES

SUNYAEV-ZELDOVICH CLUSTERS

The SZ effect is generated with the inverse Compton scattering of photons from CMB on electrons. There are two versions of the SZ effect: kinetic and thermal. They can be observed in the presence of a cluster of galaxies although they can also be caused by any large body with hot ionized gas. The SZ effect is of major cosmological importance since it can help determine the value of Hubble's constant. There are two different tendencies in the literature: to consider SZ custers either as diffuse sources or as compact sources.

POINT SOURCES

Point sources are caused by distant stars or galaxies that appear as localized, impulsive bursts of radiation. Unlike the sources discussed above, they are not diffuse sources: it is not possible to consider templates that scale in different frequencies and each channel needs to be considered separately. Due to these properties, the general approach in the literature is to detect and remove them from radiation maps before starting the component separation task. For more details on compact sources, the reader is referred to [20].

For more detailed information on cosmological sources, the reader is referred to [6].

BLIND SOURCE SEPARATION APPROACHES TO COSMIC SOURCE ESTIMATION

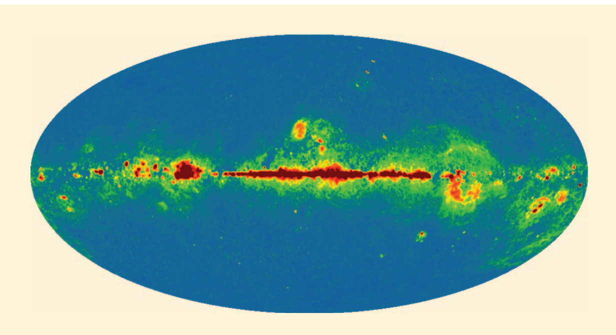
Despite the presence of earlier work on the estimation of cosmological components from multifrequency measurements such as [21] by Brandt et al. who utilized nonlinear least squares techniques for recovering components, some of the earliest work that utilize a full source separation approach were given in [22] and [8]. In these and other early work, the common followed approach was blind source separation. Blind source separation techniques do not assume any prior knowledge about the mixing or the sources and attempt to solve the problem using only observations. In particular, [8] uses independent component analysis (ICA) where (5) was solved using a gradient descent algorithm [23] for the recovery of CMB and galactic sources from synthetic Planck multispectral images; [22] made a comparative study of various ICA algorithms on Hubble telescope data; and in [24], a fixed-point algorithm, namely FastICA (FPICA) [25], was used on synthetic Planck images to obtain all-sky maps.

Another blind identification technique is proposed by Cardoso et al. [12], [26] and is called spectral matching ICA, which matches the sample spectral covariance matrices that depend on the data to their theoretical values, which in turn depend on the unknown parameters.

These techniques do not assume any prior knowledge about the mixing matrix or the cosmological sources other than the number and the independence of the sources, hence they are blind and ignore any available prior information.

BAYESIAN SOURCE SEPARATION

We start with a basic Bayesian formulation of the source separation problem as primarily suggested by Knuth [27] and Mohammad-Djafari [28].



[FIG5] **Free-free emission as estimated by the WMAP consortium [4]. (Courtesy of NASA.)**

The essence of the Bayesian methodology is the Bayes' theorem

$$p(\boldsymbol{\theta}|\mathrm{y}) = \frac{p(\mathrm{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathrm{y})}, \tag{10}$$

where $\boldsymbol{\theta}$ represents the vector containing model parameters and y the observations. The first term in the numerator is the likelihood that describes the probability of the observation given the model. Classical estimation schemes such as maximum likelihood utilize only this part that provides us with means of predicting based on our assumed model and comparing with observations. What is more interesting, however is the probability of the model given the observations, which is given by $p(\boldsymbol{\theta}|\mathrm{y})$ and is called the posterior probability. The second term in the numerator $p(\boldsymbol{\theta})$ gives the probability of the model and is the prior distribution. The denominator $p(\mathrm{y})$ is the probability of the observation and is called the evidence. Evidence does not depend on the model and can be ignored.

The prior encodes, in probability terms, our prior knowledge or "belief" about the model before we observe the data. It can be seen as our probability model of the quantity or process we are studying before observation. Once the observation is made, the likelihood modifies or updates our prior model to the posterior model. The posterior then is the model that incorporates all information we have originating from our belief or prior knowledge and from the observations we make. Faced with new observations, we can use our existing posterior model as our new prior and update it with the likelihood coming from new data and so on. The Bayes theory is an excellent formulation of our learning process.

Source separation is yet another estimation problem and can be modeled in the Bayesian framework. There are three important steps for Bayesian estimation: 1) adopting a model for the problem, i.e., construction of the likelihood; 2) formulating the prior information in the prior probabilities; and 3) evaluating the posterior and search in the hypothesis space for the probable solutions.

The second step is the critical design step where the researcher can make a change in the performance of technique by his choice of the prior. The most difficult computational part, however, is the final step. A careful analysis of the problem may suggest a hierarchical model with many nuisance and hyperparameters that need to be integrated out to obtain the estimates over parameters that are interest to us. In most cases other than special cases such as uniform priors or in the case of Gaussian likelihood, Gaussian priors this is an analytically intractable problem and expensive computational integration schemes need to be used.

Considering the characterization of the spectral responses described in the section "Cosmological Components," it is reasonable to parameterize A with a vector of spectral indices $\boldsymbol{\theta} = \{\beta_s, \beta_d, \ldots\}$ of considerably smaller dimension. We write the mixing matrix as $\mathrm{A}(\boldsymbol{\theta})$ to emphasize this point. It is assumed throughout the literature on the estimation of cosmological sources from multichannel measurements that the cosmological sources are independent, each defined by a prior distribution $p(\mathrm{s}_k \mid \boldsymbol{\psi}_k)$ with parameters $\boldsymbol{\psi}_k$.

The goal is to estimate the sources S, the parameters $\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{n_s})$ associated with the models for S, $\boldsymbol{\theta}$, and the noise variances $\boldsymbol{\sigma}^2$, given the observation of x.

For the cosmological component separation problem we can write the posterior as follows:

$$\begin{aligned} p(\mathrm{S}, \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\sigma}|\mathrm{x}) &\propto p(\mathrm{x}|\mathrm{S}, \mathrm{A}(\boldsymbol{\theta}), \boldsymbol{\sigma})P(\mathrm{S}|\boldsymbol{\psi})\, p(\boldsymbol{\psi})\, p(\boldsymbol{\sigma})\, p(\boldsymbol{\theta}) \\ &= \left[\prod_{j=1}^{J} p(\mathrm{x}_j|\mathrm{s}_j, \mathrm{A}(\boldsymbol{\theta}), \boldsymbol{\sigma})\right]\left[\prod_{n=1}^{N} p(\mathrm{s}_k|\boldsymbol{\psi}_k)\, p(\boldsymbol{\psi}_k)\right] p(\boldsymbol{\sigma})p(\boldsymbol{\theta}). \end{aligned} \tag{11}$$

In the next few subsections, each component of this distribution is defined in turn [see (12)–(15)].

LIKELIHOOD

The error e_j primarily arises from antenna receiver noise and can be reasonably assumed to be Gaussian distributed with zero mean. The errors are mutually independent within and between pixels j and frequency, and are assumed to be distributed at each frequency with standard deviation σ_i. Most source separation studies, where noise is modeled, have assumed that the noise is i.i.d. while the noise is actually nonstationary, since while scanning the sky the antenna makes multiple passes. For most CMB data, and certainly for Planck, the noise variances are known from extensive simulation.

In works such as [29], to take care of the nonstationarity as well, a simple form of stochastic spatial dependence for the noise was adopted, since the noise variance is dependent on the number of observations at each point in the sky. This number may differ, according to the scanning schedule adopted by the detector. The number of measurements at pixel j is assumed known to be r_j. Where $r_j \geq 2$, x_j is taken to be the mean of the measurements and since the mean of r_j i.i.d. Gaussian random variables is also Gaussian, the variance of the noise at pixel j is $r_j^2\sigma^2$. The term $p(\mathrm{x}_j|\mathrm{s}_j, \mathrm{A}(\boldsymbol{\theta}), \boldsymbol{\sigma})$ in (11) is therefore written as

$$\begin{aligned} p(\mathrm{x}_j|\mathrm{s}_j, \mathrm{A}(\boldsymbol{\theta}), \boldsymbol{\sigma}) &= \prod_{m=1}^{M} \frac{1}{(2\pi)^{1/2}\, r_j\sigma_m} \exp\left(-\frac{(x_{mj} - \mathrm{A}_{m\cdot}\mathrm{s}_j)^2}{2r_j^2\sigma_i^2}\right), \\ & x_{mj} \in \mathbb{R}, \end{aligned} \tag{12}$$

where A_m is the mth row of $\mathrm{A}(\boldsymbol{\theta})$.

PRIORS

One of the earliest Bayesian formulation for cosmological component estimation was given by Hobson et al. in [7], where the authors study a reduced problem assigning priors only to the sources. They proposed two different priors, the first one being a multivariate Gaussian prior for the sources. With Gaussian distribution as a conjugate prior, this leads to a very simple posterior that leads to a Wiener filter solution. Unfortunately, despite its analytical ease, this is not a valid choice since the sources other than CMB are clearly non-Gaussian. Second, they propose an entropy prior for a maximum entropy solution. Again this prior is not based on the distributions of the sources and the choice is far from optimal. Following on the same line, Jewell et al. [30] gave a general analytical framework of source separation and extended it to multiresolution source separation. They present the full Bayesian framework for the solution of the problem but do not specify priors.

In some of the work that uses a Bayesian formulation for the cosmological parameter estimation from CMB measurements, such as in [31] and [11], the priors are taken to be uniform for the sake of simplicity. Hence, the posterior reduces to the likelihood. This choice retains most of the advantages of the Bayesian framework such as error analysis, however the potential to incorporate prior information and hence reducing search space has not been exploited. Again, Gaussian prior was used for foregrounds in [32] for reasons of analytical simplicity.

A number of other works propose more generic distributions to model foregrounds, in particular [19], [33], and [29], Gaussian mixture distributions are suggested based on the observation that foregrounds follow multimodal distributions. On simulated data, it has been demonstrated that a small number of Gaussian components is enough for close approximation of foreground histograms [19].

Modeling the distribution of s_{kj}, source k at pixel j, as a Gaussian mixture model (GMM) with an unknown number of components m_k provides a very flexible but tractable class of models for the sources. Let $\boldsymbol{\mu}_k = (\mu_{k1}, \ldots, \mu_{km_k})$, $\mathrm{t}_k = (t_{k1}, \ldots, t_{km_k})$ and $\mathrm{p}_k = (p_{k1}, \ldots, p_{km_k})$ be the mixture component means, variances and weights for the kth source. Hence the parameters ψ_k of the kth source are $\boldsymbol{\psi}_k = (\boldsymbol{\mu}_k, \mathrm{t}_k, \mathrm{p}_k, m_k)$ and

$$p(\mathrm{s}_k|\boldsymbol{\psi}_k) = \prod_{j=1}^{J}\sum_{c=1}^{m_k}\frac{p_{kc}}{\sqrt{2\pi t_{kc}}}\exp\left(-\frac{(s_{kj}-\mu_{kc})^2}{2t_{kc}}\right),\ s_{kj}\in\mathbb{R}, \quad (13)$$

where $\mathrm{s}_k = \{s_{kj}|j=1,\ldots,J\}$. Let $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{n_s})$, $\mathrm{t} = (\mathrm{t}_1, \ldots, \mathrm{t}_{n_s})$, $\mathrm{p} = (\mathrm{p}_1, \ldots, \mathrm{p}_{n_s})$, and $\mathrm{m} = (m_1, \ldots, m_{n_s})$ denote the vectors of all mixture means, variances, weights, and number of components for all the sources, so that $\boldsymbol{\psi} = (\boldsymbol{\mu}, \mathrm{t}, \mathrm{p}, \mathrm{m})$.

The remaining terms in (11) are $p(\boldsymbol{\psi}_k)$, $p(\boldsymbol{\sigma})$, and $p(\boldsymbol{\theta})$. In [29], conjugate priors are suggested that facilitate the computation of the posterior and are flexible enough to incorporate useful prior information. Given m_k, the prior distributions for the means $\boldsymbol{\mu}_k = (\mu_{k1}, \ldots, \mu_{km_k})$ are assumed to be independent and identical Gaussian distributions with means ξ_k and variances κ_k, the prior distribution of the variances $\mathrm{t}_k = (t_{k1}, \ldots, t_{km_k})$ are assumed to be independent and identical gamma distributions with shapes α_k and scales β_k and the prior distribution of the weights p_k is assumed to be a Dirichlet distribution with equal parameters $(\delta_k, \ldots, \delta_k)$. The number of components m_k is assigned a geometric prior with mean λ_k, hence

$$\begin{aligned} p(\boldsymbol{\psi}_k) &= \frac{\Gamma(m_k\delta_k)}{\Gamma(\delta_k)^{m_k}}\, p_{kc}^{\delta_k-1}\,(1-\lambda_k^{-1})^{m_k-1}\lambda_k^{-1} \\ &\quad\times\left(\prod_{c=1}^{m_k}(\kappa_k/2\pi)^{1/2}\exp(-0.5\kappa_k(\mu_{kc}-\xi_k)^2)\right. \\ &\quad\left.\times\,\beta_k^{\alpha_k}t_{kc}^{\alpha_k-1}\frac{e^{-\beta_k t_{kc}}}{\Gamma(\alpha_k)}\right). \end{aligned} \quad (14)$$

The prior distribution parameters themselves—ξ_k, κ_k, α_k, β_k, δ_k, and λ_k, for $k = 1, \ldots, n_s$—are assigned values to reflect what is known currently about the values of the sources.

Since antenna characteristics and sky-scanning strategy is completely known, the noise variances $\boldsymbol{\sigma}^2$ are known, and so $p(\boldsymbol{\sigma}) = 1$ at the specified value. If it were to be assumed unknown, the gamma distribution could be used as it is the conjugate distribution in this case.

As mentioned above, rough bounds on the values of the spectral indices $\boldsymbol{\theta}$ are known. One can use independent normal distributions for each $\boldsymbol{\theta}$,

$$p(\boldsymbol{\theta}_k) = \frac{1}{\sqrt{2\pi}\sigma_{\theta,k}}\exp(-(\theta_k - m_{\theta,k})^2/2\sigma_{\theta,k}^2) \quad (15)$$

with mean $m_{\theta,k}$ and standard deviation $\sigma_{\theta,k}$ so that major part of the prior probability lies within the rough bounds. Alternatively, uniform distributions in this range can also be used.

A very important addition to this formulation is suggested in [38] and [41], where the CMB angular power spectrum was included in the posterior in (11), too. In this way, the angular spectrum, which houses important information about cosmological parameters, is obtained directly rather than calculating from the CMB map, which would require the mapping of the full sky first.

In this section, we demonstrated how to develop the posterior for the cosmological component separation problem in a general setup. It should be clear to the reader that as more information is obtained, one can change the priors and can build even more elaborate posteriors. If a simpler posterior is desired, one can also use noninformative or conjugate priors to that effect.

IMPLEMENTING THE SOURCE SEPARATION: MARKOV CHAIN MONTE CARLO

The posterior developed in the previous section does not lend itself to an analytical solution when point estimates are desired. To obtain point estimates, we need to marginalize and integrate out the nuisance parameters and maximize the marginal posteriors depending on the choice of the estimator. The many dimensional integrations are prohibitive even when other priors used to obtain simpler posteriors.

The remedy to this technical problem is Monte Carlo integration methods, in particular, the Markov chain Monte Carlo (MCMC). The Monte Carlo techniques let us evaluate complicated integrals by sampling rather than by analytical or numerical methods. They can render otherwise impossible integrations feasible, however this is only part of their advantage. If sampling is done in an intelligent way, significant computational gains and robustness are obtained. MCMC sampling achieves this by enforcing that the subsequent samples follow a Markov chain with certain properties, i.e., irreducibility or aperiodicity [34]. Sampling in such a Markov chain ensures that we converge to the target distribution regardless of our initial guess of priors. This final point is very important: basically, even if we start with

> **IN CONTRAST TO BLIND SOURCE SEPARATION, THE BAYESIAN FRAMEWORK PROVIDES US WITH MEANS OF INFORMED SOURCE SEPARATION.**

"wrong" priors, MCMC sampling produces the correct posterior given enough observations. Then, the difference between choosing a good or bad prior reduces to the convergence speed.

The irreducibility and aperiodicity of the Markov chain is enforced by ensuring that the sampling leads to balanced transitions which is a sufficient condition [34], that is the probability of making a transition from a state i to j should be equal to the probability of the reverse transition.

The most popular and general MCMC algorithm is Metropolis-Hastings, a pseudocode of which is given in Table 1, where q represents the proposal function from which samples are generated, p is the target distribution, i.e., the posterior, and θ is the parameter vector.

The need for using MCMC in the cosmological component separation problem was first mentioned by Jewell et al. in [30] but no demonstrations were provided. In their later work [32], they propose the Metropolis-Hastings method for sampling and estimate CMB power spectrum in the presence of foreground sources.

In problems such as source separation, the posterior may become too complicated to design a good proposal distribution. In such cases, if the conditional distributions are available, one can use a simpler MCMC algorithm called Gibbs sampling. In Gibbs sampling, the basic idea is to sample consecutively from conditional distributions of all variables. There are no acceptance ratios since all transitions are accepted. It can be shown that Gibbs sampling guarantees convergence to the target distribution [37], [34]. A simple pseudocode of Gibbs sampling is given in Table 2, where $\boldsymbol{\theta}_{-k}^{(t)}$ represents the vector of all variables other than $\theta_k^{(t)}$.

Kuruoglu and Comparetti [33] and Wandelt et al. [38] independently suggest using Gibbs sampling and [33] reports results significantly better than obtained with blind techniques.

Eriksen et al. [11] adopt a previously suggested framework by Jewell et al. [30] and report results on WMAP data using Gibbs sampling. They used uniform priors all over, hence reducing the posterior to the likelihood. They later extended their method for joint component separation and CMB power spectrum estimation [41] and introduced some prior measurements as templates to the observation equation as in [38]. They use noninformative priors for most of the variables. Wilson et al. [29], unlike most previous work, adopt a generic prior for the foregrounds (Gaussian mixtures with unknown number of components) and construct a hierarchical framework and perform again Gibbs sampling.

In all of these works, the source separation is implemented in two stages: The first stage is Monte Carlo sampling of the posterior distribution of (11), specifically by an MCMC [34]. Once this is done, the second stage is to compute the point estimates (generally the sample mean of the samples of the sources); this average is taken to be the estimated source.

[TABLE 1] METROPOLIS-HASTINGS SAMPLING.

1) INITIALIZE $\theta^{(0)}$
2) FOR $t = 1$ TO T, T: METROPOLIS LOOP LENGTH
   - SAMPLE $u \sim \mathcal{U}[0, 1]$
   - SAMPLE $\tilde{\theta} \sim q(\tilde{\theta} | \theta^{(t)})$
   - IF $u < r(\theta^{(t)}, \tilde{\theta}) = \min\left(1, \dfrac{p(\tilde{\theta})q(\theta^{(t)}|\tilde{\theta})}{p(\theta^{(t)})q(\tilde{\theta}|\theta^{(t)})}\right)$
     $\theta^{(t+1)} = \tilde{\theta}$
     ELSE
     $\theta^{(t+1)} = \theta^{(t)}$

[TABLE 2] GIBBS SAMPLING.

1) INITIALIZE $\boldsymbol{\theta}_{1:K}^{(0)}$, K: DIMENSION OF THE PARAMETER VECTOR
2) FOR $t = 1$ TO T
   - FOR $k = 1$ TO K
     – SAMPLE $\theta_k^{(t+1)} \sim p(\theta_k | \boldsymbol{\theta}_{-k}^{(t)})$

TOWARDS MORE ELABORATE MODELS AND ALGORITHMS

FASTER ALGORITHMS

Perhaps the main criticism to Bayesian source separation with sampling methods, MCMC in particular, is their computational load and slow convergence. Regarding speed, they cannot compete with analytical methods or blind methods such as FastICA. This renders testing of the method, of the models, and the priors a time consuming task, which probably is the fact responsible for the delayed acceptance of these techniques in the astrophysics community. There are various ways, however, that the technique can be accelerated.

To speed up the algorithm, one can start with some good initial solutions, such as the results of the FastICA algorithm, and move on from there, hence avoiding long burn-in periods. Alternatively, one can use a pyramidal sampling scheme, starting from low-resolution images, and then move to increasingly higher resolutions with the low-resolution results as starting points.

Another way to speed up MCMC is to use intelligent random walks, or state transition proposals such that more relevant parts of the space is explored and rejection rate is decreased. Such a scheme is provided by so-called Langevin sampling. In statistical physics, the Langevin equation is used to describe the Brownian motion of the particles in a potential and has been used to obtain a smart MC algorithm in [42]. Another parallel sampling algorithm is the Hamiltonian Monte Carlo, which is the generalized version of the Langevin sampler. Taylor et al. utilized Hamiltonian sampling for fast estimation of CMB power spectrum from observations [43]. Kayabol et al. implemented the Langevin scheme [44] on the image source separation problem and report computational gains of two orders of magnitude when compared to regular Gibbs.

ALL OF THE POTENTIAL OF BAYESIAN MODELING SHOULD BE EXPLOITED TO DERIVE ANY INFORMATION PRESENT IN THIS OLDEST AND MOST IMPORTANT DATA SET OF OUR UNIVERSE.

Other strategies for improving MCMC can be found in [45].

FROM SIGNAL SEPARATION TO IMAGE SEPARATION

The vast majority of the techniques described up to this point perform one-dimensional (1-D) signal analysis, that is, convert the images first to 1-D signals before processing. On the other hand, the data under consideration are space radiation maps and contain important spatial information which is completely lost when converted to 1-D. This spatial information could be utilized to our profit for better separation. Few works attempt modeling the space dependence structure of the images (sources) and utilize this additional information to benefit the separation process. In [46], authors build a Markov random field (MRF)-based model for pixel interactions and use edge preserving priors for pixel gradients. The technique, which is called iterated conditional model, can be viewed as a naive Bayes method.

A MRF model is defined in Gibbs formulation that models the statistical relation between neighboring pixels. The probability density of $\mathbf{s}_n$ is expressed in a Gibbs formulation as

$$p(\mathbf{s}_n) = \frac{1}{Z(\beta)} e^{-U(\mathbf{s}_n)}, \tag{16}$$

where $Z(\beta)$ is the partition function to ensure that the total probabilit y $= 1$ and the clique potential or energy is defined as

$$U(s_{n,i}) = \frac{1}{2} \sum_{\{i,j\} \in \mathcal{C}} \beta\rho(s_{n,i} - s_{n,j}), \tag{17}$$

where $\rho(\cdot)$ is the potential function. The density given in (16) can be written in the vector form using the cliques in the eight compass directions as

$$p(\mathbf{s}_n) = \frac{1}{Z(\beta)} \exp\left\{ -\sum_{d=1}^{8} \beta\rho(\mathbf{s}_n - \mathbf{G}_d\mathbf{s}_n) \right\}, \tag{18}$$

where the clique differences in each direction are defined as $\mathbf{s}_n^{(d)} = \mathbf{s}_n - \mathbf{G}_d\mathbf{s}_n$. Here, $\mathbf{G}_d$ is the one-pixel shift operator in direction d. The clique potentials quantify the interaction or correlation between pixels.

A full Bayesian treatment was given in [47], where the authors utilized Gibbs sampling with embedded Metropolis sampling, the pseudocode of which is given in Table 3.

After sampling all image pixels with Metropolis method described in Table 4, samples are drawn from the mixing matrix and from the noise variance. When all of the unknowns are sampled, one iteration of the Gibbs sampling algorithm is completed. Kayabol et al. provide simulation results which demonstrate significant gain over 1-D Gibbs sampling when MRF-Gibbs sampling is used [47].

FROM STATIONARY SOURCE SEPARATION TO NONSTATIONARY SOURCE SEPARATION

Most of the works in the literature consider separation of stationary mixing (constant mixing matrix) of stationary sources. However, the mixing matrix that is formed of the spectral indices of the cosmological sources at measurement channels is not necessarily constant. Most cosmological sources and the antenna receiver noise due to nonuniform scanning of the sky also demonstrate characteristics that cannot be described by stationary stochastic models. A limited number of work address this problem by extending the MCMC approaches to sequential Monte Carlo or particle filtering techniques. In particular, Costagli et al. in [48] propose a 1-D particle filter while in [49] they suggest the fusion of multiple 1-D particle filters that scan the image in different directions to approximate two-dimensional (2-D) particle filters.

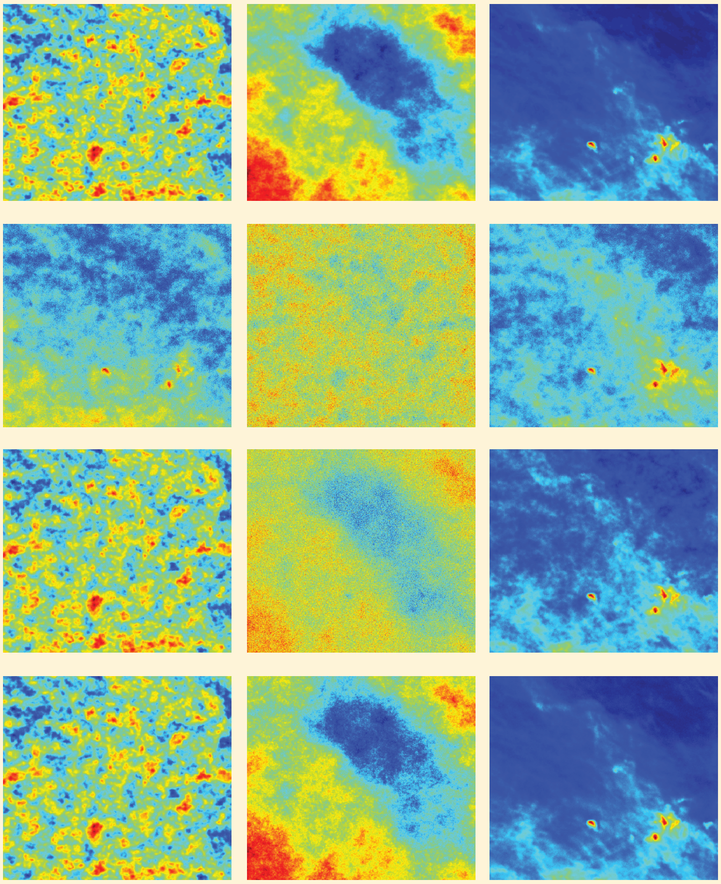
Sequential Monte Carlo or particle filtering [51] is an extension of Kalman filtering [50] for the solution of dynamic systems to possibly nonlinear and non-Gaussian systems and signals. It has gained wide acceptance and success in tracking problems. A nonstationary source mixing system can be visualized as a dynamic system and the mixing matrix and sources can be seen as hidden state variables of the system. Sequential Monte Carlo can potentially track space-varying spectral indices and nonlinear beam effects as well as space varying parameters of the stochastic models for the sources.

[TABLE 3] ONE CYCLE OF GIBBS SAMPLING WITH EMBEDDED METROPOLIS.

1) FOR ALL SOURCE IMAGES, $n = 1 : \mathrm{N}$
    FOR ALL PIXELS, $i = 1 : J$
        USING METROPOLIS METHOD IN TABLE 4
        $s_{n,i}^{t+1} \leftarrow \{p(s_{n,i}|\mathbf{x}_{1:M}, \theta_{-s_{n,i}}^{t})\}$
2) FOR ALL ELEMENTS OF THE MIXING MATRIX, $(m, n) = (1, 1) : (M, N)$
    $a_{m,n}^{t+1} \leftarrow \{p(a_{m,n}|\mathbf{x}_{1:M}, \theta_{-a_{m,n}}^{t})\}$
3) FOR THE NOISE VARIANCE, $m = 1 : M$
    $(\sigma_m^2)^{t+1} \leftarrow \{p(\sigma_m^2|\mathbf{x}_{1:M}, \theta_{-\sigma_m^2}^{t})\}$

[TABLE 4] METROPOLIS ALGORITHM FOR A SOURCE IMAGE. q_i: PROPOSAL DENSITY FOR THE iTH PIXEL OF SOURCE n; u: UNIFORM POSITIVE RANDOM NUMBER IN THE UNIT INTERVAL; w: NEW SAMPLE TO BE TRIED; r: ACCEPTANCE RATIO OF THE GENERATED SAMPLE.

1) PRODUCE w FROM $q_i(s_{n,i}, w)$.
2) CALCULATE $r = \min\left(1, \frac{p(\tilde{\theta})}{p(\theta^{(t)})}\right)$
3) IF $r \geq 1$ THEN $s_{n,i}^{t+1} = w$
    ELSE PRODUCE $u \sim \mathcal{U}[0, 1]$.
        IF $u < r$ THEN $s_{n,i}^{t+1} = w$,
        ELSE $s_{n,i}^{t+1} = s_{n,i}^{t}$
4) $i + 1 \leftarrow$ NEXT PIXEL AND GO TO 1ST STEP.



[FIG6] Comparison of performances of separation algorithms. (a) CMB, (b) synchrotron, and (c) dust. Row 1: original patches, row 2: FastICA results, row 3: SMICA results, and row 4: MRF+Gibbs results.

Another recent approach for dealing with nonstationary data is the incorporation of time and frequency information to the separation process. With this approach, potentially superior results can be obtained when compared to separation only in spatial domain or in frequency domain since diversity in both domains are exploited [52].

A related method is wavelet analysis, which provides the potential of modeling local features and deals with the problem in multiple scales. In [53], Moudden et al. extend the spectral matching ICA (SMICA) technique [12] to the wavelet domain. It is important to note that these techniques give the potential to deal with compact sources such as point sources in addition to the diffuse sources discussed in this article. Currently, time-frequency and wavelet-based techniques are mostly "blind" techniques while the Bayesian versions are in development. One such method is described in [54] where Bobin et al. propose to decompose sources into overcomplete orthonormal bases (hence obtaining morphological components) as in

$$s_j = \sum_{k}^{D} \varphi_{jk} = \sum_{k}^{D} \alpha_{jk} \phi_k, \qquad (19)$$

where ϕ_k are the ortonormal bases. Assuming the sparsity of sources in these bases, they adopt Laplace priors for α_{jk}. They note that cosmological sources such as CMB, galactic dust, and SZ clusters can be represented rather well with orthonormal wavelet bases. They propose estimating the morphological components φ_{jk} and the mixing matrix A by maximizing the joint posterior.

FROM INDEPENDENT SOURCES TO DEPENDENT SOURCES

A common assumption among works in the literature is the independency of the cosmological sources. Although it is well known that CMB is independent from the rest of the sources, the galactic sources demonstrate significant statistical dependence among themselves. Recently, a small number of researchers have started addressing this problem. Outside the Bayesian context, Bedini et al. [55], [17] proposed a second-order statistics-based method, namely correlated component analysis that also models the correlation between galactic components. The technique, despite being non-Bayesian, exploits the fact that the parametrization with spectral indices reduces the number of unknowns in the mixing matrix, which compensates for the increase in the number of parameters in the covariance matrix due to the correlation between some of the sources.

A more informed approach is presented in [56] and [57] that proposes a modified version of tree-dependent component analysis for the separation of dependent galactic sources. The technique provides a tree-based decomposition of hidden components in the observations using the Kullback-Leibler divergence. Simulations with WMAP data was successful in indicating dependence between galactic sources.

One of the challenges in this line of research is that almost no prior work exists on the dependence between galactic components, therefore one can come up with only very generic priors. It should be noted that dependence models based on covariances have only limited use since we do not have any indications of linear dependence between galactic sources. A parametric separation approach can be very useful in modeling the dependence between galactic sources.

A SIMULATION STUDY

We provide a simple simulation study on patches of dimension 512×512. We mix patches of CMB, synchrotron, and dust emission artificially using mixing coefficients reported in [8], and then try to recover original sources using some of the methods mentioned above. Figure 6 shows the original sources and the separation results. In Figure 7, we calculated the angular power spectrum of estimated CMB for each technique. We see that Gibbs sampling with MRF image model and SMICA obtain the angular power spectrum rather well, while FastICA cannot. Despite the success of SMICA in obtaining CMB, in Figure 6, we see that it fails recovering galactic sources that is the manifestation of the unsuitability of the stationarity assumption for these sources.

The reader is referred to [58] for a detailed simulation study of some blind or semiblind methods with a Bayesian technique [41] by the Planck consortium. Another simulation study can be found in [47].

CONCLUSIONS

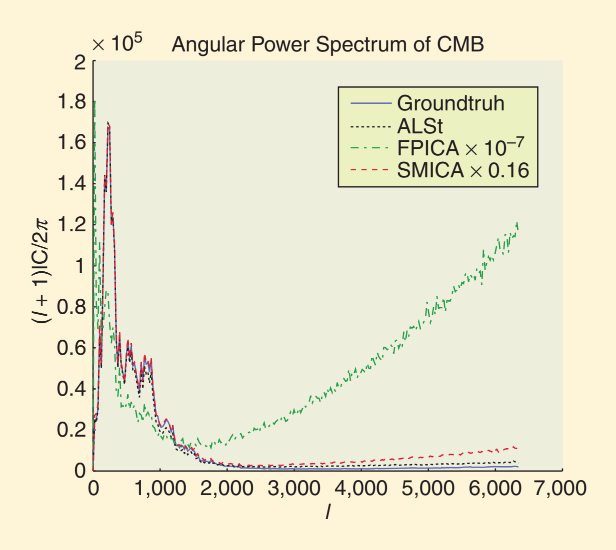
In this article, we defined the problem of cosmological source separation problem from multispectral measurements in a Bayesian framework and presented a panorama of Bayesian source separation methods that were proposed in the literature. We presented some new research frontiers where work is under progress, particularly nonstationary source separation, image separation, and dependent component analysis.

Despite the speed of blind source separation methods, they lack the flexibility of Bayesian framework in providing increasingly complex models for cosmological variables and for error analysis. With the arrival of WMAP and Planck satellite data, we will have the possibility of looking into increasingly detailed models that will also create the need to make error analysis for new models. All of these and more are handled seamlessly by the Bayesian framework, which provides us with the means to model our uncertainty about cosmological sources and propagate it to cosmological parameters.

We believe that the computational complexity of Bayesian methods should not be discouraging. All of the potential of Bayesian modeling should be exploited to derive any information present in this oldest and most important data set of our Universe.

ACKNOWLEDGMENTS

This work has been supported partially by ASI under contract Planck LFI Phase E2 activity. The simulation results in the section "A Simulation Study" were provided by Koray Kayabol. The source images used in the simulations were provided by Diego Herranz and the Planck data analysis team. The WMAP estimates of the source images in the section "Cosmological Components" were provided by NASA and the WMAP Science team [4]. Some of the ideas presented in this article have benefited from and matured during collaborations with a number of researchers over the years. The author would like to thank Simon Wilson, Koray Kayabol, Bulent Sankur, Diego Herranz, Emanuele Salerno, Mauro Costagli, Luigi Bedini, Anna Tonazzini, and Jose Luis Sanz.



[FIG7] CMB angular power spectrum estimates. (FPICA, SMICA, and adaptive Langevin sampling (ALSt) for MRF-Gibbs).

AUTHOR

Ercan E. Kuruoglu (ercan.kuruoglu@isti.cnr.it) received his B.Sc. and M.Sc. degrees in electrical and electronics engineering at Bilkent University in 1991 and 1993 and the M.Phil and Ph.D degrees in information engineering at Cambridge University in 1995 and 1998, respectively. In 1998, he joined Xerox Research Center in Cambridge. In 2000, he was an ERCIM fellow at INRIA-Sophia Antipolis. In 2002, he joined ISTI-CNR, Pisa where he is currently a senior researcher. He was a visiting professor at Georgia Institute of Technology, Shanghai in 2007. He has been on the editorial boards of four journals on signal processing. He takes part in Planck activities in Working Group 2 (diffuse component separation). He is a member of the IEEE Technical Committee on Signal Processing Theory and Methods. His research interests are in statistical signal processing and information theory with applications in image processing, astronomy, telecommunications, and bioinformatics. He is a Senior Member of the IEEE.

REFERENCES

[1] Degree Angular Scale Interferometer home page, 2004. [Online]. Available: http://astro.uchicago.edu/dasi/

[2] BOOMERang home page, 2008. [Online]. Available: http://www.astro.caltech.edu/lgg/boomerang_front.htm

[3] NASA. Cosmic Background Explorer, 1993. [Online]. Available: http://lambda.gsfc.nasa.gov/product/cobe/

[4] NASA. Legacy archive for microwave background data analysis (LAMBDA), 2009. [Online]. Available: http://lambda.gsfc.nasa.gov

[5] Planck, ESA. Planck surveyor home page, 2009. [Online]. Available: http://www.rssd.esa.int/index.php?project=PLANCK

[6] ESA, Planck Science Team. Planck Scientific Programme, 2005. [Online]. Available: http://www.rssd.esa.int/SA/PLANCK/docs/Bluebook-ESA-SCI(2005)1_V2.pdf

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

[8] C. Baccigalupi, L. Bedini, C. Burigana, G. D. Zotti, A. Farusi, D. Maino, M. Maris, F. Perotta, E. Salerno, L. Toffolatti, and A. Tonazzini, "Neural networks and the separation of cosmic microwave background and astrophysical signals in sky maps," *Monthly Notices Royal Astron. Soc.*, vol. 318, no. 3, pp. 769–780, 2000.

[9] H. Snoussi, G. Patanchon, J. Macias-Peres, A. Mohammad-Djafari, and J. Delabrouille, "Bayesian blind component separation for cosmic microwave background observations," in *AIP Proc. Maximum Entropy and Bayesian Inference Conf.*, 2001, pp. 125–140.

[10] V. Stolyarov, M. P. Hobson, A. N. Lasenby, and R. B. Barreiro, "All-sky component separation in the presence of anisotropic noise and dust temperature variations," *Monthly Notices Royal Astron. Soc.*, vol. 336, no. 1, pp. 97–111, 2005.

[11] H. K. Eriksen, C. Dickinson, C. R. Lawrence, C. Baccigalupi, A. J. Banday, K. M. Gorski, F. K. Hansen, P. B. Lilje, E. Pierpaoli, K. M. Smith, and K. Vanderlinde, "CMB component separation by parameter estimation," *Astrophys. J.*, vol. 641, no. 2, pp. 665–682, 2006.

[12] J.-F. Cardoso, H. Snoussi, J. Delabrouille, and G. Patanchon, "Blind separation of noisy Gaussian stationary sources: Application to cosmic microwave background imaging," in *Proc. EUSIPCO 2002*, Toulouse, France, pp. 561–564.

[13] G. Patanchon, J.-F. Cardoso, J. Delabrouille, and P. Vielva, "Cosmic microwave background and foregrounds in WMAP first year data," *Monthly Notices Royal Astron. Soc*, vol. 364, no. 4, pp. 1185–1194, 2005.

[14] M. White and J. D. Cohn, "The theory of anisotropies in the cosmic microwave background," *Amer. J. Phys.*, vol. 70, no. 2, pp. 106–118, 2002.

[15] W. Hu and S. Dodelson, "Cosmic microwave background anisotropies," *Ann. Rev. Astron. Astrophys.*, vol. 40, pp. 171–216, Sept. 2002.

[16] E. Komatsu, A. Kogut, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, L. Verde, E. Wollack, and E. L. Wright., "First year Wilkinson microwave anisotropy probe (WMAP) observations: Tests of Gaussianity," *Astrophys. J. Suppl. Ser.*, vol. 148, no. 1, pp. 119, 2003.

[17] A. Bonaldi, L. Bedini, E. Salerno, C. Baccigalupi, and G. De Zotti, "Estimating the spectral indices of correlated astrophysical foregrounds by a second-order statistical approach," *Monthly Notices Royal Astron. Soc.*, vol. 373, no. 1, pp. 271–279, 2006.

[18] C. G. T. Haslam, C. J. Salter, H. Stoffel, and W. E. Wilson, "A 408 MHz all sky continuum survey. II. The atlas of contour maps," *Astron. Astrophys. Suppl. Ser.*, vol. 47, pp. 4–142, Jan. 1982.

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

[20] D. Herranz and P. Vielva, "Cosmic microwave background images," *IEEE Signal Processing Mag.*, vol. 27, no. 1, pp. 67–75, 2010.

[21] W. N. Brandt, C. R. Lawrence, A. C. S. Readhead, J. N. Pakianathan, and T. M. Fiola, "Separation of foreground radiation from cosmic microwave background anisotropy using multifrequency measurements," *Astrophys. J.*, vol. 424, no. 1, pp. 1–21, 1994.

[22] D. Nuzillard and A. Bijaoui, "Blind source separation and analysis of multispectral astronomical images," *Astron. Astrophys. Suppl. Ser.*, vol. 147, pp. 129–138, Nov. 2000.

[23] S.-I. Amari and A. Cichocki, "Adaptive blind signal processing-neural network approaches," *Proc. IEEE*, vol. 86, no. 10, pp. 2026–2048, 1998.

[24] D. Maino, A. Farusi, C. Baccigalupi, F. Perotta, A. J. Banday, L. Bedini, C. Burigana, G. D. Zotti, K. M. Gorski, and E. Salerno, "All-sky astrophysical component separation with fast independent component analysis," *Monthly Notices Royal Astron. Soc.*, vol. 334, no. 1, pp. 53–68, 2002.

[25] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, 1999.

[26] J. Delabrouille, J. F Cardoso, and G. Patachon, "Multi-detector multi-component spectral matching and applications for CMB data analysis," *Monthly Notices Royal Astron. Soc.*, vol. 346, no. 4, pp. 1089–1103, 2003.

[27] K. H. Knuth, "Informed source separation: A Bayesian tutorial," in *Proc. 13th European Signal Processing Conf. (EUSIPCO 2005)*, B. Sankur, A. E. Cetin, M. Tekalp, and E. Kuruoglu, Eds. Antalya, Turkey.

[28] A. Mohammed-Djafari, "A Bayesian approach to source separation," in *Proc. Maximum Entropy and Bayesian Inference*, 1999, pp. 221–224.

[29] S. P. Wilson, E. E. Kuruoglu, and E. Salerno, "Fully Bayesian source separation of astrophysical images modelled by mixture of Gaussians," *IEEE J. Select. Top. Signal Process.*, vol. 2, no. 5, pp. 685–696, 2008.

[30] J. Jewell, C. R. Lawrence, and S. Levin, "Bayesian approach to foreground removal," in *Proc. Microwave Foregrounds*, vol. 181, ASP, San Francisco, CA, 1999, pp. 357–366.

[31] N. Christensen, R. Meyer, L. Knox, and B. Luey, "Bayesian methods for cosmological parameter estimation from cosmic microwave background measurements," *Classic. Quantum Gravity*, vol. 18, no. 14, pp. 2677–2688, 2001.

[32] J. Jewell, S. Levin, and C. H. Anderson, "Application of Monte Carlo algorithms to the Bayesian analysis of the cosmic microwave background," *Astrophys. J.*, vol. 609, no. 1, p. 114, 2004.

[33] E. E. Kuruoglu and P. M. Comparetti, "Bayesian separation of independent sources in astrophysical radiation maps using MCMC," in *Proc. PHYSTAT (Statistical problems in Particle Physics, Astrophysics and Cosmology)*, L. Lyons, Ed., 2003, pp. 321–323.

[34] A. E. Gelfand and A. F. M. Smith, "Sampling based approaches to calculating marginal densities," *J. Amer. Stat. Assoc.*, vol. 85, no. 410, pp. 398–409, 1990.

[35] H. K. Eriksen, I. J. ODwyer, J. B. Jewell, B. D.Wandelt, D. L. Larson, K. M. Gorski, S. Levin, A. J. Banday, and P. B. Lilje, "Power spectrum estimation from high-resolution maps by Gibbs sampling," *Astrophys. J. Suppl. Ser.*, vol. 155, no. 2, p. 227, 2004.

[36] V. Stolyarov, M. P. Hobson, M. A. J. Ashdown, and A. N. Lasenby, "All-sky component separation for the Planck mission," *Monthly Notices Royal Astron.Soc.*, vol. 336, no. 1, pp. 97–111, 2002.

[37] C. Andrieu, N. de Freitas, A. Doucet, and M. Jordan, "An introduction to MCMC for Machine Learning," *Mach. Learn.*, vol. 50, no. 1–2, pp. 5–43, 2003.

[38] B. D. Wandelt, D. L. Larson, and A. Lakshminarayanan, "Global, exact cosmic microwave background data analysis using Gibbs sampling," *Phys. Rev. D*, vol. 70, no. 8, 083511, 2004.

[39] I. J. ODwyer, H. K. Eriksen, B. D.Wandelt, J. B. Jewell, D. L. Larson, K. M. Gorski, S. Levin, A. J. Banday, and P. B. Lilje, "Bayesian power spectrum analysis of the first-year WMAP data," *Astrophys. J. Suppl. Ser.*, vol. 612, 2004.

[40] J. B. Jewell, H. K. Eriksen, B. D. Wandelt, I. J. O'Dwyer, G. Huey, and K. M. Gorski, "A Markov chain Monte Carlo algorithm for analysis of low signal-to-noise CMB data" *Astrophys. J.,* vol. 697, no. 1, pp. 258–268, 2009.

[41] H. K. Eriksen, J. B. Jewell, C. Dickinson, A. J. Banday, K. M. Gorski, and C. R. Lawrence, "Joint Bayesian component separation and CMB power spectrum estimation," *Astrophys. J.*, vol. 676, no. 1, pp. 10–32, 2008.

[42] R. J. Rossky, and J. D. Doll, and H. L. Friedman, "Brownian dynamics as a smart Monte Carlo simulation," *J. Chem. Phys.*, vol. 69, no. 10, pp. 4628–4633, 1978.

[43] J. F. Taylor, M. A. J. Ashdown, and M. P. Hobson, "Fast optimal CMB power spectrum estimation with Hamiltonian sampling" *Monhtly Notices Royal Astron. Soc.,* vol. 389, no. 3, pp. 1284–1292, 2008.

[44] K. Kayabol, E. E. Kuruoglu, B. Sankur, E. Salerno, and L. Bedini, "Fast MCMC separation for MRF modelled astrophysical components," in *Proc. IEEE Int. Conf. Image Processing*, Egypt, Nov. 2009.

[45] W. R. Gilks and G. O. Roberts, "Strategies for improving MCMC," in *Monte Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds. Boca Raton, FL: CRC, 1996, pp. 89–114.

[46] E. E. Kuruoglu, A. Tonazzini, and L. Bianchi, "Source separation of astrophysical images modelled with Markov Random Field models," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Singapore, Oct. 24–27, 2004, pp. 2701–2704.

[47] K. Kayabol, E. E. Kuruoglu, and B. Sankur, "Bayesian separation of images modelled with MRFs using MCMC," *IEEE Trans. Image Process.*, vol. 18, no. 5, pp. 982–994, 2009.

[48] M. Costagli, E. E. Kuruoglu and A. Ahmed, "Astrophysical image separation using particle filters," in *Proc. 5th Int. Conf. Independent Component Analysis and Blind Signal Separation (ICA)* (Lect. Notes Comput. Sci. 3195), Granada, Spain, Sept. 22–24, 2004. New York: Springer-Verlag, 2004, pp. 930–937.

[49] M. Costagli and E. E. Kuruoglu, "Image separation using particle filters," *Digital Signal Process.*, vol. 17, no. 5, pp. 935–946, 2007.

[50] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, J. Basic Eng.*, vol. 82, no. 1, ser. D, pp. 35–45, 1960.

[51] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo methods for Bayesian filtering," *Stat. Comput.*, vol. 10, no. 3, pp. 197–208, 2000.

[52] T. Ozgen, E. E. Kuruoglu, and D. Herranz, "Separation of noisy astrophysical images by blind time-frequency source separation methods," *Digital Signal Process.*, vol. 19, no. 9, pp. 360–369, 2009.

[53] Y. Moudden, J. F. Cardoso, J. L. Starck, and J. Delabrouille, "Blind component separation in wavelet space: Application to CMB analysis," *EURASIP J. Appl. Signal Process.*, vol. 2005, no. 15, pp. 2437–2454, 2005.

[54] J. Bobin, Y. Moudden, J. L. Starck, J. Fadili, and N. Aghanim, "SZ and CMB reconstruction using generalized morphological component analysis," *Stat. Methodol.*, vol. 5, no. 3, pp. 307–317, 2008.

[55] L. Bedini, D. Herranz, E. Salerno, C. Baccigalupi, E. E. Kuruoglu, and A. Tonazzini, "Separation of correlated astrophysical sources using multiple-lag data covariance matrices," *EURASIP J. Appl. Signal Process.*, vol. 2005, no. 15, pp. 2400–2412, 2005.

[56] S. Fortunati, "Separazione di componenti dipendenti con applicazioni in astrofisica," Laurea thesis, Univ. Pisa, June 2008.

[57] E. E. Kuruoglu, "Separation of dependent components in astrophysical images using tree-dependent component analysis," Tech. Rep., ISTI-CNR, 2009.

[58] S. M. Leach, J.-F. Cardoso, C. Baccigalupi, R. B. Barreiro, M. Betoule, J. Bobin, A. Bonaldi, J. Delabrouille, G. de Zotti, C. Dickinson, H. K. Eriksen, J. González-Nuevo, F. K. Hansen, D. Herranz, M. LeJeune, M. López-Caniego, E. Martinez-González, M. Massardi, J. -B. Melin, M. -A. Miville-Deschênes, G. Patanchon, S. Prunet, S. Ricciardi, E. Salerno, J. L. Sanz, J. -L. Starck, F. Stivoli, V. Stolyarov, R. Stompor, and P. Vielva, "Component separation methods for the Planck mission," *Astron. Astrophys.*, vol. 491, no. 2, pp. 597–615, 2009.