DEPENDENT GAUSSIAN MIXTURE MODELS FOR SOURCE SEPARATION

Alicia Quirós Carretero and Simon P. Wilson

Dep. de Estadística, Univ. Rey Juan Carlos
Móstoles (Madrid), Spain
phone: + (34) 91 4888236, fax: + (34) 91 4887626,
email: alicia.quiros@urjc.es

Dep. of Statistics, Trinity College Dublin
Dublin 2, Ireland
phone: + (353) 1896 1759, fax: + (353) 1677 0711,
email: swilson@tcd.ie

ABSTRACT

Source separation is a common task in signal processing and is often analogous to factor analysis. In this work we look at a factor analysis model for source separation of multi-spectral image data where prior information about the sources and their dependencies is quantified as a multivariate Gaussian mixture model with an unknown number of factors. Variational Bayes techniques for model parameter estimation are used. The development of this methodology is motivated by the need to bring an efficient solution to the separation of components in the microwave radiation maps to be obtained by the satellite mission Planck which has the objective of uncovering cosmic microwave background radiation. The proposed algorithm successfully incorporates a rich variety of prior information available to us in this problem in contrast to most of the previous work that assumes completely blind separation of the sources. Results on realistic simulations of Planck maps and on WMAP 5th year images are shown. The technique suggested is easily applicable to other source separation applications by modifying some of the priors.

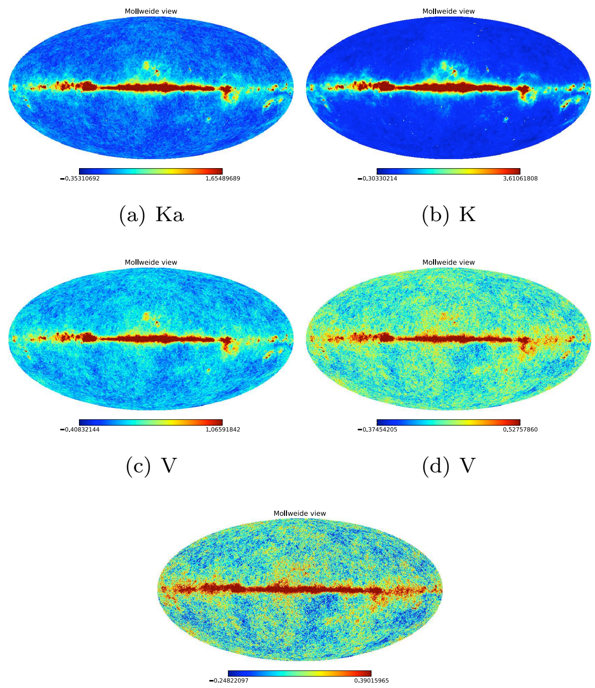
1. INTRODUCTION

The discovery of the cosmic microwave background (CMB) is strong evidence for the big bang theory of the formation and development of the universe. According to the theory, the early universe was smaller and hotter but cooled as it expanded. Once the temperature cooled to about 3000K, photons were free to propagate without being scattered off ionized matter; the CMB is an image of this event and is visible across the entire sky. Three satellites have been launched to measure the CMB: the cosmic background explorer (COBE), Wilkinson microwave anisotropy probe (WMAP) and most recently the Planck surveyor. Planck is the highest resolution data to date, of the order of 10^7 pixels across the sky measured at 9 channels.

Unfortunately, the signals measured by these satellites as in Fig. 1 contain radiation not 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 from our galaxy include synchrotron, dust and free-free emission. Therefore, the separation of the CMB signal from other sources is an important stage in the production of CMB maps [1].

To date, there have been several attempts to achieve it in a Bayesian framework: Gaussian mixture model (GMM) prior [2], and Markov Random Field (MRF) prior [3, 4]. Full sky maps at low resolution through MCMC, using masks to reduce the effect of the signal in the galactic plane, were described in [5]. Some of these are fully Bayesian source separation methods which are developed to separate the underlying CMB from the mixed observed signals of extraterrestrial microwaves made at several frequencies.

A common assumption among works in the literature is the independence 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, as stated in [1]. Recently, a small number of researchers have started addressing this problem [6, 7]. Various dependent component analysis approaches are compared in [8], demonstrating their superior performance with respect to classical ICA.



(a) Ka (b) K (c) V (d) V (e) W

Figure 1: Observed WMAP 7 year data

In this work we present a dependent components model for source separation of multi-spectral image data, where prior information about the sources and between-source dependencies is quantified as a multivariate GMM, using Variational Bayes techniques for model parameter estimation. The work in this paper can thus be considered as an extension of [2], modelling dependencies between sources through generalizing the prior to multivariate GMM.

The rest of the document is structured as follows. Section 2 gives the model for the mixing problem and describes the hierarchical Bayesian model that we use, including the prior we assume for the sources. Section 3 describes the variational Bayes approach we use for the implementation of the separation. Section 4 provides results on both synthetic Plank and real WMAP images. Finally, we provide a discussion of the results in Section 5.

© EURASIP, 2011 - ISSN 2076-1465

2. MODEL

The model description is defined in terms of the microwave source separation problem, where there are n_f maps of the sky at frequencies $(\nu_1, \ldots, \nu_{n_f})$, each map consisting of J pixels. The data are denoted $\mathbf{d}_j \in \mathbb{R}^{n_f}, j = 1, \ldots, J$. The source model consists of n_s sources and is represented by the vectors $\mathbf{s}_j \in \mathbb{R}^{n_s}$, with each component representing the amplitude of a physical source of microwaves. We assume that the $\mathbf{d}_j$ can be represented as a linear combination of the $\mathbf{s}_j$:

$$\mathbf{d}_j = \mathbf{A}\mathbf{s}_j + \mathbf{e}_j, \tag{1}$$

where $\mathbf{A}$ is an $n_f \times n_s$ "mixing" matrix and $\mathbf{e}_j$ is a vector of n_f independent Gaussian error terms with precisions (inverse variances) $\tau = (\tau_1, \ldots, \tau_{n_f})$. For convenience, define

$$\begin{aligned} \mathbf{D} &= \{d_{ij} | i = 1, \ldots, n_f, j = 1, \ldots, J\}; \\ \mathbf{S} &= \{s_{kj} | k = 1, \ldots, n_s, j = 1, \ldots, J\} \end{aligned}$$

to represent all data and sources.

We assume dependence between the sources, defined by a prior distribution $p(\mathbf{S}|\psi)$ with parameters ψ. The goal is to estimate the $\mathbf{S}$ and the parameters ψ associated with the model for $\mathbf{S}$, given observation of $\mathbf{D}$. The noise variances τ and the mixing matrix $\mathbf{A}$ are assumed known. GMM are used to represent the non-Gaussian sources, in which case it is an example of a model known as a mixture of factor analysers [9]. Like in [9], we adopt a Bayesian approach to the data fitting, implemented by a variational Bayes approach.

Bayesian inference will be based on the posterior distribution, which following the above description can be factorized as:

$$p(\mathbf{S}, \psi | \mathbf{A}, \mathbf{D}, \tau) \propto p(\mathbf{D}|\mathbf{S}, \mathbf{A}, \tau) p(\mathbf{S}|\psi) p(\psi). \tag{2}$$

Each element of this distribution is defined next in turn.

i) Noise structure

Gaussian error, $\mathbf{e}_j$, is assumed independent within and between pixels j and frequency, which gives

$$p(\mathbf{D}|\mathbf{S}, \mathbf{A}, \tau) = \prod_{j=1}^{J} \prod_{i=1}^{n_f} \sqrt{\frac{\tau_i}{2\pi}} \exp\left(-\frac{\tau_i}{2}(d_{ij} - \mathbf{A}_{i\cdot}\mathbf{s}_j)^2\right) \tag{3}$$

where $\mathbf{A}_{i\cdot}$ is the ith row of $\mathbf{A}$.

ii) Mixing Matrix Structure

In this application, $\mathbf{A}$ is parameterized and denoted $\mathbf{A}(\theta)$. Each column of $\mathbf{A}(\theta)$ is the contribution to the observation of a source at different frequencies, which is written as a function of the frequencies and θ. These parameterizations are approximations that come from the current state of knowledge about how the sources are generated. Here, we merely state the parameterization that we are going to use, and refer to [10] for a more detailed exposition on the background to them. Some restrictions are usually placed on $\mathbf{A}(\theta)$ in order to force a unique solution; this is achieved here by setting one row of $\mathbf{A}(\theta)$ to be ones.

It is assumed that the CMB is the first source and therefore, it corresponds to the first column of $\mathbf{A}(\theta)$. It is modelled as a black body at a temperature, and its contribution is a known constant at each frequency. The parametrization of the mixing matrix is given as

$$\begin{aligned} \mathbf{A}_{i1}(\theta) &= \frac{g(v_i)}{g(v_1)}, \\ \mathbf{A}_{i2}(\theta) &= \left(\frac{v_i}{v_1}\right)^{\kappa_s} \text{ and} \\ \mathbf{A}_{i3}(\theta) &= \frac{\exp(\eta v_1/k_B T_1) - 1}{\exp(\eta v_i/k_B T_1) - 1} \left(\frac{v_i}{v_1}\right)^{1+\kappa_d} \text{ and} \\ \mathbf{A}_{i4}(\theta) &= \left(\frac{v_i}{v_1}\right)^{\kappa_f} \text{ and} \end{aligned}$$

where

$$g(v_i) = \left(\frac{\eta v_i}{k_B T_0}\right)^2 \frac{\exp(\eta v_i/k_B T_0)}{(\exp(\eta v_i/k_B T_0) - 1)^2},$$

$T_0 = 2.725$ is the average CMB temperature in Kelvin, $T_1 = 18.1$, η is the Plank constant and k_B is Boltzmann's constant. The ratio $g(v_i)/g(v_1)$ is designed to ensure that $\mathbf{A}_{11}(\theta) = 1$ as we constraint the first row of $\mathbf{A}(\theta)$ to be ones. There are three unknown model parameters for $\mathbf{A}$, for synchrotron $\kappa_s \in \{\kappa_s : -3.0 \leq \kappa_s \leq -2.3\}$, the spectral indices for dust $\kappa_d \in \{\kappa_d : 1 \leq \kappa_d \leq 2\}$ and for free-free emission $\kappa_f \in \{\kappa_f : -2.3 \leq \kappa_f \leq -3.0\}$.

iii) The sources

The distribution of $\mathbf{s}_j$ is modeled as a GMM with m factors. The model proposed allows for between source dependence; the vector of sources at a pixel is a mixture of multivariate Gaussians:

$$p(\mathbf{S}|\psi) = \prod_{j=1}^{J} \sum_{a=1}^{m} w_a p(\mathbf{s}_j | \mu_a, Q_a) \tag{4}$$

where

$$p(\mathbf{s}_j | \mu_a, Q_a) = \sqrt{\frac{|Q_a|}{(2\pi)^{n_s}}} \exp\left(-\frac{1}{2}(\mathbf{s}_j - \mu_a)^T Q_a (\mathbf{s}_j - \mu_a)\right),$$

for mixture component weights w_a, mean vectors μ_a and precision matrices Q_a, so that ψ is all the w_a, μ_a and Q_a, with $a = 1, \ldots, m$.

iv) Priors

The remaining term in equation (2) is $p(\psi)$. We use the conjugate prior distributions [11] that facilitate the computation of the posterior and yet flexible enough to incorporate good prior information: Gaussians for the component means, Dirichlet for the component weights and Wishart for precision matrices. In the microwave source application, background knowledge about the magnitude of the sources can be incorporated through specifying values of the parameters of these prior distributions. This prior specification follows [12], who discuss how to specify these values in more detail.

3. IMPLEMENTING THE SOURCE SEPARATION

The posterior developed in the previous section does not lend itself to an analytical solution. The MCMC techniques let us evaluate complicated integrals by sampling rather than by analytical or numerical methods. The main criticism to Bayesian source separation with sampling methods, MCMC in particular, is their computational load and slow convergence. Regarding the speed, they cannot compete with methods such as FastICA.

There are several approaches to speed up the algorithm, such as the strategies suggested in [13]. In the image source separation problem framework, the Langevin sampling scheme has been implemented [3], as a way to obtain a faster MC algorithm.

In this work, the source separation model presented in Section 2 is implemented by a variational Bayesian approach

[9, 14, 15], that allows for more efficient inference when dealing with large data when compared with MCMC techniques. In essence, given the data $\mathbf{D}$ and a model with parameters θ and latent variables $\mathbf{Z}$, the variational Bayes method is based on approximating the posterior distribution $p(\mathbf{Z},\theta|\mathbf{D})$ with a factorial approximation $q(\mathbf{Z},\theta|\phi) = q(\mathbf{Z}|\phi_{\mathbf{Z}})q(\theta|\phi_\theta)$, where ϕ are the variational parameters. The approximation is fitted by minimising the Kullback-Leibler divergence between q and p, or equivalently maximising a lower bound on marginal log-likelihood of the data.

Attias has recently developed a fully Bayesian approach to GMM [16] which variational approximation to the posterior, when choosing conjugate priors, leads to the following components: Wishart densities for the precisions, Q_a; Normal densities for the means, μ_a; and a Dirichlet for the mixing coefficients, $\mathbf{p}$; and a discrete distribution for the indicator posteriors, z_j, which indicates the component that explains information in pixel j. We further derived the variational approximation to the marginal posterior of sources, $\mathbf{s}_j$, turning out to be a multivariate Gaussian distribution. In brief:

$$
\begin{aligned}
q(\mathbf{s}_j) &\sim MVN(A^\star_j, B^\star_j)\\
q(\mathbf{p}) &\sim D(\lambda)\\
q(\mu_a|Q_a) &\sim N(\xi_a, \beta_a Q_a)\\
q(Q_a) &\sim W(\eta_a, V_a)
\end{aligned}
$$

and

$$
\begin{aligned}
q(z_j = a) \propto \exp&\left(\Psi(\lambda_a) - \Psi\left(\sum_{a'}\lambda_{a'}\right)\right)|V_a|^{\frac{1}{2}}2^{\frac{n_s}{2}}\\
&\exp\left(\frac{1}{2}\sum_{i=1}^{n_s}\Psi\left(\frac{\eta_a+1-i}{2}\right)\right)\exp\left(-\frac{n_s}{2\beta_a}\right)\\
&\exp\left(-\frac{\eta_a}{2}\left((A^\star_j-\xi_a)^T V_a(A^\star_j-\xi_a) + tr(V_a(B^\star_j)^{-1})\right)\right)
\end{aligned}
$$

where Ψ denotes the digamma function. Note that $q(z_j = a)$ is the probability that component a is responsible for information in pixel j in sources, $\mathbf{s}_j$.

The quantities of interest, i.e. the hyper-parameters to be computed, $A^\star_j$, $B^\star_j$, λ, ξ_a, β_a, η_a and V_a for $j = 1,\ldots,J$ and $a = 1,\ldots,m$, have the following values:

$$
\begin{aligned}
(B^\star_j)_{kl} &= \sum_{i=1}^{n_f}\tau_i A_{ik}A_{il} + \sum_{a=1}^{m} q(z_j=a)\eta_a(V_a)_{kl}\\
(A^\star_j)_k &= (B^\star_j)^{-1}\mathbf{v}(k), \text{ with}\\
\mathbf{v}(k) &= \sum_{i=1}^{n_f}\tau_i d_{ij}A_{ik} + \sum_{a=1}^{m} q(z_j=a)\eta_a\sum_{l=1}^{n_s}(V_a)_{kl}(\xi_a)_l\\
\lambda_a &= \sum_j q(z_j=a) + \lambda_a^{\text{prior}}\\
\xi_a &= \frac{\sum_j[q(z_j=a)A^\star_j] + \beta_a^{\text{prior}}\xi_a^{\text{prior}}}{\sum_j q(z_j=a) + \beta_a^{\text{prior}}}\\
\beta_a &= \sum_j q(z_j=a) + \beta_a^{\text{prior}}\\
\eta_a &= \sum_j q(z_j=a) + \eta_a^{\text{prior}}
\end{aligned}
$$

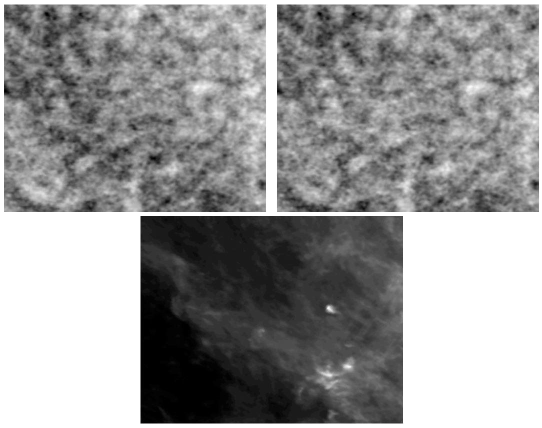


Figure 2: Observations on the 3 of the 9 channels (lowest, middle and highest frequencies are shown) of the data generated from the source separation model with realistic simulations of CMB, synchrotron and galactic dust.

$$
\begin{aligned}
V_a &= \sum_j\{q(z_j=a)[(1-\frac{2q(z_j=a)}{\sum_j q(z_j=a)})(B^\star_j)^{-1} + \Phi +\\
&+ (A^\star_j-\bar{\mu}_a)(A^\star_j-\bar{\mu}_a)^T]\} + V_a^{\text{prior}} +\\
&+ \frac{\beta_a^{\text{prior}}[\sum_j q(z_j=a)][\Phi + (\bar{\mu}_a-\xi_a^{\text{prior}})(\bar{\mu}_a-\xi_a^{\text{prior}})^T}{\beta_a}
\end{aligned}
$$

with

$$
\begin{aligned}
\bar{\mu}_a &= \frac{\sum_j[q(z_j=a)A^\star_j]}{\sum_j q(z_j=a)}\\
(\Phi)_{kl} &= \frac{\sum_j\{[q(z_j=a)]^2(B^\star_j)^{-1}_{kl}\}}{[\sum_j q(z_j=a)]^2}.
\end{aligned}
$$

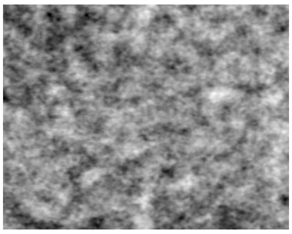
Computations were carried out using Matlab.

4. EXAMPLES

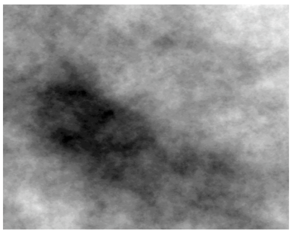
4.1 Analysis of simulated data

Figure 2 shows data obtained from realistic simulations of CMB, synchrotron and galactic dust on a 512×512 patch. The original sources are shown in Figure 3. The data were generated at the 9 frequencies that are observed by Planck from 30 to 857 GHz. The mixing matrix used was as defined in Section 2 with $\kappa_s = -2.9$ and $\kappa_d = 2.0$. Noise precisions were those published by the Planck research team. After exploring several values for m, the number of components in the GMM source model was fixed to be $m = 1$, as it provided the best fit, taking into account the compromise between fit and number of parameters in the model.

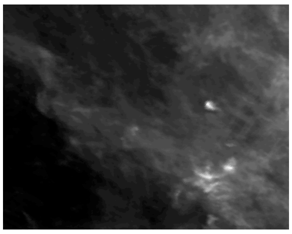
Figure 4 shows an estimate of CMB, along with a scatter plot of this estimate against the true value, as shown in Figure 3. Such a estimate is the average of the samples obtained for the first column of $A^\star$, which corresponds to CMB. We see from the scatter plot and from comparison with Figure 3 that the reconstruction of CMB is very accurate here. The same is true for the other two sources, as shown in Figure 5. Regarding the between-sources dependence structure, posterior estimates of $V^{-1}_{1k}, k = 2,3$ are approximately 0, suggesting independence between CMB and the other sources, as expected. On the other hand, posterior estimate of $V^{-1}_{23} \neq 0$, indicating dependence between synchrotron and galactic dust.



(a) CMB

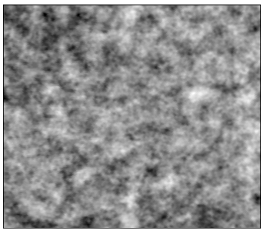


(b) Synchrotron



(c) Dust

Figure 3: The simulated sources used to generate the simulated data in Figure 2



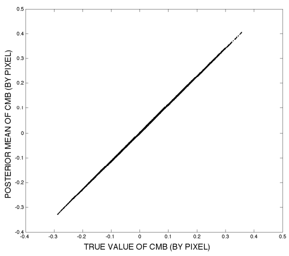
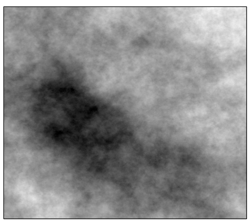


Figure 4: The posterior mean of the reconstruction of the CMB with a scatter plot of true versus posterior mean.



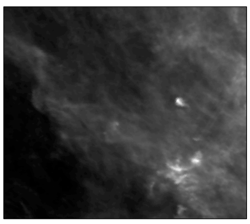


Figure 5: The posterior mean of the reconstruction of synchrotron and galactic dust.

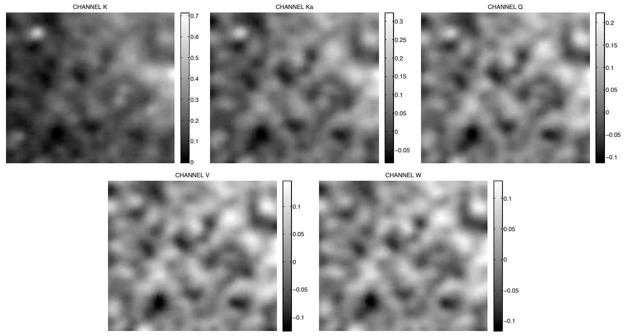


Figure 6: Temperatures (in mKs) at 20^o square patch of the sky from WMAP at 5 microwave frequencies.

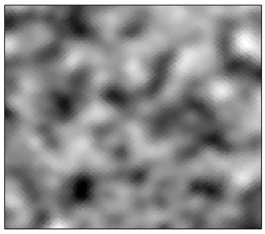


Figure 7: Estimated CMB.

4.2 Analysis of a WMAP Year 5 Patch

The WMAP was launched in 2001 and at the time of writing is still operational. It observes 5 frequencies from 22 to 90 GHz. Figure 6 shows a patch of 5-year WMAP data.

The algorithm of Section 3 was implemented with 4 sources (CMB, synchrotron, dust and free-free emission). The noise precisions were assumed to be the published values for WMAP detectors. The spectral density for free-free emission was fixed at -2.14 (following [10]) and the synchrotron and dust spectral indices were as in the first example. The number of components in the GMM source model were fixed to be $m = 2$, following the same reasoning as in the simulation study. Informative priors were placed on the GMM parameters, based on discussions on the expected marginal properties of the sources. Figure 7 shows the estimated CMB. The result obtained is in agreement with previous work [2]. In order to show the fit of the data to the model, Figure 8 is a scatter plot of the posterior expected value of the d_{jk} with the standardised residuals, with one figure for each frequency $k = 1, \ldots, 5$.

5. DISCUSSION

A fully Bayesian factor analysis algorithm has been presented and applied to a multi-channel image source separation problem, where dependencies between sources are modelled as a multivariate GMM. The algorithm performs very well on simulated Planck data and has been applied to data from WMAP.

Previous results [2] showed that source dependencies clearly exist in the posterior distribution, due to the stochastic linear constraint that $\mathbf{s}_j \approx \mathbf{A}\mathbf{d}_j$, thus prior modelling of them help to produce a more realistic separation. In this work, we extend that approach by allowing the source priors to be a mixture of multivariate Gaussian distributions for each pixel.

The development of this methodology is motivated by the need to bring an efficient solution to the separation of com-

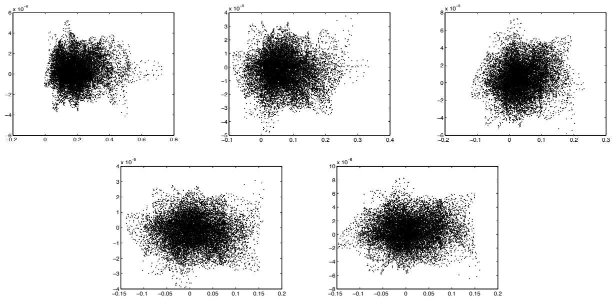


Figure 8: Assessment of model fit. Scatter plot of the posterior predicted values of d_j against the standardised residual over all pixels.

ponents in the microwave radiation maps to be obtained by the satellite mission Planck which has the objective of uncovering CMB radiation. The proposed algorithm successfully incorporates a rich variety of prior information available to us in this problem in contrast to most of the previous work that assumes completely blind separation of the sources. Further, the variational approach presented here overcomes the convergence problems of the MCMC stated in [17], when dealing with large data sets such as that will be provided by the satellite mission Plank.

In the analysis of simulated data, the number of components in the GMM source model turned out to be $m = 1$. This means that sources are multivariate Gaussian a priori. On the other hand, for real data, the number of components is $m = 2$. In blind source separation problem, identifiability rely on the independence of the sources. In this work, in spite of modelling the sources as Guassians when $m = 1$, identifiability is obtained because of the prior information which is incorporated to the model, given structure to the mixing matrix.

Another type of dependence is that a source is spatially correlated. Spatial dependence is most conveniently modelled by a Gaussian MRF and some preliminary work on this idea can be found in [4]. Combining with cross source correlations, one might ultimately consider a mixture of multivariate Gaussian MRF as a prior for the sources. Implementing the analysis with such a prior would be a significant challenge computationally; we hypothesize that it will be difficult to derive a well-behaved MCMC approach. Other functional approximations, such as that of [18], offer feasible alternative to computing the posterior distribution in this case.

Finally, although the technique was developed for the astrophysical source separation problem in mind, it is general and it is applicable to other source separation problems as well.

6. ACKNOWLEDGEMENTS

Simon Wilson's work is supported by the STATICA project, funded by the Principal Investigator programme of Science Foundation Ireland, contract number 08/IN.1/I1879.

REFERENCES

[1] E. E. Kuruoglu, "Bayesian source separation for cosmology," *IEEE Signal Processing Magazine*, vol. 27, pp. 43–54, Jan. 2010.

[2] S. P. Wilson, E. E. Kuruoglu, and E. Salerno, "Fully Bayesian source separation of astrophysical images modelled by a mixture of Gaussians," *IEEE Journal of selected topics in signal processing*, vol. 2(5), pp. 685–696, Oct. 2008.

[3] K. Kayabol, E. E. Kuruoglu, J. L. Sanz, B. Sankur, E. Salerno, and D. Herranz, "Adaptive Langevin sampler for separation of t-distribution modelled astrophysical maps," *IEEE Transactions on Image Processing*, vol. 19(9), pp. 2357–2368, Sep. 2010.

[4] K. Kayabol, E. E. Kuruoglu, and B. Sankur, "Bayesian separation of images modeled with MRFs using MCMC", *IEEE Transactions on Image Processing*, vol. 18 (5), pp. 982–994, 2009.

[5] C. Dickinson, H.K. Eriksen, A.J. Banday, J.B. Jewell, K.M. Gorski, G. Huey, C.R. Lawrence, I.J. O'Dwyer and B.D. Wandelt, "Bayesian component separation and cosmic microwave background estimation for the five-year WMAP temperature data", *Astrophysical Journal*, vol. 705, pp. 1607–1623.

[6] 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(15), pp. 2400–2412, 2005.

[7] 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(1), pp. 271–279, 2006.

[8] E.E. Kuruoglu, "Dependent Component Analysis for Cosmology," *Lecture Notes on Computer Science*, vol. 6365, pp. 538–545, 2010

[9] Z. Ghahramani and M. Beal, "Variational inference for Bayesian mixtures of factor analysers," in *Advances in Neural Information Processing Systems*, T. K. L. S. A. Solla and K.-R. Mller, Eds. MIT Press, 2000, vol. 12, pp. 449455.

[10] H. K. Eriksen, C. Dickinson, C. R. Lawrence, C. Baccigalupi, A. J. Banday, K. M. Górski, F. K. Hansen, P. B. Lilje, E. Pierpaoli, K. M. Smith, and K. Vanderlinde, "CMB component separation by parameter estimation," *Astrophysical Journal*, vol. 641, pp. 665–682, 2006.

[11] P. M. Lee, *Bayesian statistics: an introduction*, 3rd ed. London: Hodder Arnold H&S, 2004.

[12] S. Richardson and P. Green, "On Bayesian analysis of mixtures with an unknown number of components (with discussion)," *J. R. Statist. Soc. B*, vol. 59, pp. 731–792, 1997.

[13] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov chain Monte Carlo in practice*, Boca Raton, FL: CRC, 1996.

[14] D. J. C. McKay, *Information theory, inference and learning algorithms*, Cambridge University Press, Cambridge, 2003.

[15] C. M. Bishop, *Pattern recognition and machine learning*, Springer-Verlag, New York, 2006.

[16] H. Attias, *A variational Bayesian framework for graphical models*, in T. Leen et al., editor, NIPS 12, Cambridge, MA, 2000. MIT Press.

[17] S. P. Wilson, E. E. Kuruoglu, and A. Quirós, "Bayesian Factor Analysis using Gaussian Mixture Sources, with application to separation of the cosmic microwave background," in *2nd International Workshop on Cognitive Information Processing*, pp. 198–202.

[18] H. Rue, S. Martino, and N. Chopin, "Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations," *Journal of the Royal Statistical Society, Series B*, vol. 71, pp. 319–392, 2008.