Factor analysis on multi-channel image data with Gaussian mixture priors

Simon Wilson[1] Ercan Kuruoğlu[2] Emanuele Salerno[2]
Alicia Quirós Carratero[3]

[1]Trinity College Dublin

[2]ISTI, CNR–Pisa

[3]Universidad Rey Juan Carlos, Madrid

Factor analysis

- For $\mathbf{d}_j \in \mathbb{R}^{n_f}$, $j = 1, \ldots, J$:

$$\mathbf{d}_j = A\mathbf{s}_j + \epsilon_j, \ \mathbf{s}_j \in \mathbb{R}^{n_s},$$

  where $\epsilon_j \sim N(0, \mathrm{diag}(\tau_1^{-1}, \ldots, \tau_{n_f}^{-1}))$, A is a $n_f \times n_s$ factor loading matrix;
- $\mathbf{S}_k = (s_{1k}, \ldots, s_{Jk})$ is kth factor;
- Goal: observe the $\mathbf{d}_j$ and infer the $\mathbf{s}_j$ and A;
- Known as source separation in signal processing literature where:
  - A known as the mixing matrix;
  - $\mathbf{S}_k$ known as the sources.

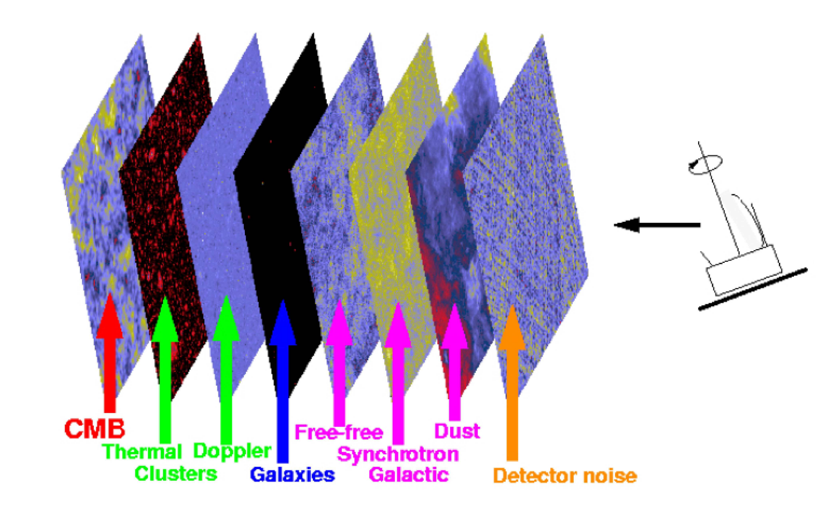
Source separation for multi-channel images

- n_f images, each of J pixels:
  - $\mathbf{d}_j$ is vector of intensities at pixel j across all n_f images;
- In remote sensing, each image is of same scene at different frequencies $\nu_1, \ldots, \nu_{n_f}$;
- $\mathbf{s}_j$ is then the vector of the n_s different sources at pixel j;
- A_{ik} is the contribution of the kth source at the ith frequency.

Cosmic microwave background (CMB)

- Discovered by accident in 1964;
- By 1970's agreed to be an image of the first scattering of EM radiation at recombination $\approx$ 300,000 years after Big Bang;
- Of great interest as an observation of the state of the early universe:
  - In particular it is remarkably uniform;
  - But accurate measurement of the small anisotropies place strong restrictions on theories of big bang, galaxy formation etc.;
- Cosmic expansion $\Rightarrow$ radiation has cooled to 2.7K (microwave);

Source separation problem for the CMB



Source separation problem for the CMB

- Upcoming data (Planck satellite) will have $J \approx 10^7$, $n_f = 9$ at frequencies from 30 to 857 GHz;
- $\mathbf{S}_k,\ k = 1, \ldots, n_s$ are the sources that make up the microwave signal received by the satellite:
  - One of these sources is the CMB (source 1);
  - Other important ones are synchrotron radiation and galactic dust;
  - There are many others.... is n_s known?
  - A lot is known from physics about the properties of these sources e.g. their spectrum, mean, variance etc;
- A is not known but the physics tell us a lot about it;
- *A lot of "prior" information $\Rightarrow$ a Bayesian approach looks promising*.

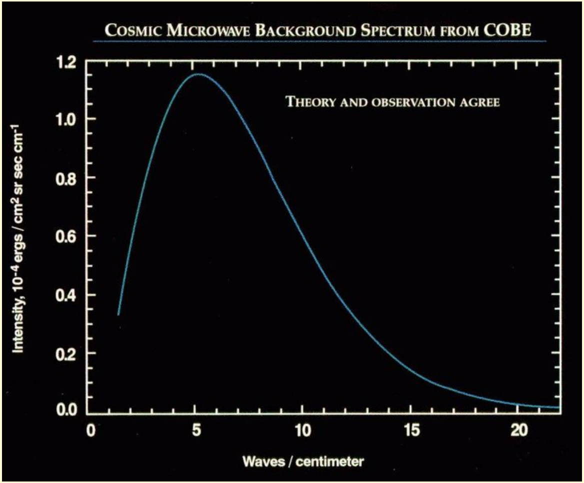
Likelihood

- Different number r_j of observations at pixel j. So

$$p(\mathbf{D} \mid \mathbf{S}, A, \underline{\tau}) \propto \left(\prod_{i=1}^{n_f} \tau_i^{\sum_j r_j} \right)^{0.5} \exp \left(-0.5 \sum_{i=1}^{n_f} \tau_i \sum_{j=1}^{J} r_j \left(d_{ij} - A_{i.}\mathbf{s}_j \right)^2 \right),$$

  where now d_{ij} is average over the r_j observations;
- $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_J), \mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_J)$.
- r_j are assumed known.

Prior information about A: CMB is black body



Prior for A

- Both A and any $\mathbf{s}_j$ unknown $\Rightarrow$ solution up to a constant in each column (source) of A;
- A_{ik} interpreted as the response of the detector at frequency ν_i to source k;
- The physics tells us a lot about what this should be for each source;
- The CMB is black body radiation at $T_0 = 2.725$K, so response at ν_i is *known*:

$$A_{i1} = \left(\frac{h\nu_i}{kT_0}\right)^2 \frac{e^{h\nu_i/kT_0}}{(e^{h\nu_i/kT_0} - 1)^2},$$

h is Planck constant, k is Boltzmann's constant.

Prior for A

- For other sources, physical argument to say that *approximately*:

$$A_{ik} = \left(\frac{\nu_i}{\nu_{0,k}} \right)^{\theta_k},$$

  for a reference frequency $\nu_{0,k}$ and parameter θ_k;
- So one free parameter θ_k per column of A;
- So A parameterised by $(n_s - 1)$ dimensional θ;
- Physical theories give tight bounds on these θ: $(-3.5, -2.0)$ for synchrotron, $(0.5, 1.5)$ for galactic dust, etc.

Prior model for sources

- Source distributions across sky show skewness, multimodality;
- Sources exhibit within and across (spatial) pixel correlations:
    - Galactic sources, extra-galactic sources and CMB should be independent;
    - Sources within galaxy show dependencies, similarly extra-galactic;
    - Most sources show spatial smoothness.

Prior model for sources

- This all suggests a mixture of GMRFs as a prior for **S**:

$$p(\mathbf{S} \mid m, \mathbf{p}, \boldsymbol{\mu}, \mathcal{Q}) \propto \sum_{a=1}^{m} p_a |Q_a|^{0.5} \exp(-0.5(\mathbf{S} - \mu_a)^T Q_a(\mathbf{S} - \mu_a)),$$

  where $\mu_a \in \mathbb{R}^{n_s J}$, Q_a is an $n_s J \times n_s J$ precision matrix.
- Lots of structure in the Q_a (but can it be made into a band matrix?);
- Two simplifications:
  1. Between source only: no spatial correlation $\Rightarrow$ each Q_a is block-diagonal of J identical $n_s \times n_s$ matrices;
  2. Independent: no within-pixel correlation either $\Rightarrow$ Q_a is diagonal (blocks of J identical $n_s \times n_s$ diagonal matrices);
- For last case, each source is an iid Gaussian mixture.

Sampling from the Posterior Distribution

- Gibbs sampling;
- Update parameters in blocks where possible:
  - Full conditional samples of mixture means, precisions and weights;
  - No. of mixture components for each source sampled by the usual Richardson and Green (JRSS B, 1997) reversible jump move;
  - Full conditional of each source at each pixel s_{kj} is a univariate mixture of Gaussians;
    - Better: full conditional of $\mathbf{s}_j$ is a mixture of multivariate Gaussians;
  - Components of θ updated jointly with their corresponding source by a Metropolis move (e.g. $(\theta_k, \mathbf{S}_k)$);

Full conditional of $\mathbf{s}_j$ in the independent case

- s_{jk} is Gaussian mixture with m_k components;
- There are $\prod_k m_k$ components in $\mathbf{s}_j$, indexed $(l_1, \ldots, l_{n_s})$;
- Component $(l_1, \ldots, l_{n_s})$ has:
  - weight $\prod_k p_{kl_k}$;
  - precision matrix $\mathcal{T}_{(l_1 \cdots l_{n_s})}$ with elements:

$$(\mathcal{T}_{(l_1 \cdots l_{n_s})})_{ab} = I(a = b) t_{al_a} + \sum_j \tau_j A_{ja} A_{jb};$$

  - mean:

$$\mu_{(l_1 \cdots l_{n_s})} = \mathcal{T}^{-1}_{(l_1 \cdots l_{n_s})} \left(\begin{pmatrix} t_{1l_1} \mu_{1l_1} \\ \vdots \\ t_{n_s l_{n_s}} \end{pmatrix} + A^T \begin{pmatrix} \tau_1 d_{j1} \\ \vdots \\ \tau_{n_f} d_{jn_f} \end{pmatrix} \right).$$

Full conditional of $\mathbf{s}_j$ in the between-source dependence case

- $\mathbf{s}_j$ still a mixture of MVNs;
- Somewhat more complicated form for means and precision matrices (but still OK to compute quickly).

Metropolis step for $(\theta_k, \mathbf{S}_k)$

- Independent prior case:
  - Propose θ_k^* from $N(\theta_k, s_\theta^2) \Rightarrow$ new loading matrix A^*;
  - $\mathbf{S}_k^*$ proposed from $P(\mathbf{S}_k \mid A^*, \mathbf{S}_{-k}, \mathbf{D})$;
  - $(\theta_k^*, \mathbf{S}_k^*)$ accepted with probability

$$\frac{\int p(\mathbf{D} \mid \mathbf{S}, A^*, \boldsymbol{\tau})\, p(\mathbf{S}_k \mid \mathbf{p}_k, \boldsymbol{\mu}_k, \mathbf{t}_k)\, d\mathbf{S}_k}{\int p(\mathbf{D} \mid \mathbf{S}, A, \boldsymbol{\tau})\, p(\mathbf{S}_k \mid \mathbf{p}_k, \boldsymbol{\mu}_k, \mathbf{t}_k)\, d\mathbf{S}_k}$$

- Similarly for between-source dependence case:
  - But computationally more difficult.

Metropolis step for $(\theta_k, \mathbf{S}_k)$ (indpt. case)

$$\int p(\mathbf{D} \,|\, \mathbf{S}, A, \boldsymbol{\tau})\, p(\mathbf{S}_k \,|\, \mathbf{p}_k, \boldsymbol{\mu}_k, \mathbf{t}_k)\, d\mathbf{S}_k$$

$$\propto \prod_{j=1}^{J} \sum_{a=1}^{m_k} p_{ka} \frac{1}{\sqrt{1 + r_j \sum_{i=1}^{n_f} \tau_i A_{ik}^2 / t_{ka}}}$$

$$\times \exp\left\{ \frac{\left(\sum_{i=1}^{n_f} \tau_i A_{ik} \left(d_{ij} - \sum_{\substack{l=1 \\ l \neq k}}^{n_s} A_{il} s_{lj} \right) \right)^2}{2 \sum_{i=1}^{n_f} \tau_i A_{ik}^2} \right.$$

$$\left. - \frac{\left(\sum_{i=1}^{n_f} \tau_i A_{ik} \left(d_{ij} - A_{ik}\mu_{ka} - \sum_{\substack{l=1 \\ l \neq k}}^{n_s} A_{il} s_{lj} \right) \right)^2}{2(t_{ka}^{-1} + (r_j \sum_{i=1}^{n_f} \tau_i A_{ik}^2)^{-1}) \left(\sum_{i=1}^{n_f} \tau_i A_{ik}^2 \right)^2} \right\}.$$

Example 1: simulated data

- Three sources (simulated Gaussian mixtures and Gaussian MRFs) at five channels on a 256×256 grid;
- Mixing matrix A generated using reasonable values from CMB, synchrotron and dust at the 9 Planck frequencies, giving:

$$A = \begin{pmatrix} 1.0000 & 1.0000 & 1.0000 \\ 0.9737 & 0.3293 & 2.1109 \\ 0.9031 & 0.0857 & 5.1571 \\ 0.7952 & 0.0305 & 10.0990 \\ 0.6186 & 0.0108 & 19.4469 \\ 0.3417 & 0.0032 & 40.2804 \\ 0.0791 & 0.0008 & 87.0180 \\ 0.0064 & 0.0002 & 153.0462 \\ 0.0001 & 0.0001 & 221.6442 \end{pmatrix}.$$

Example 1: simulated data

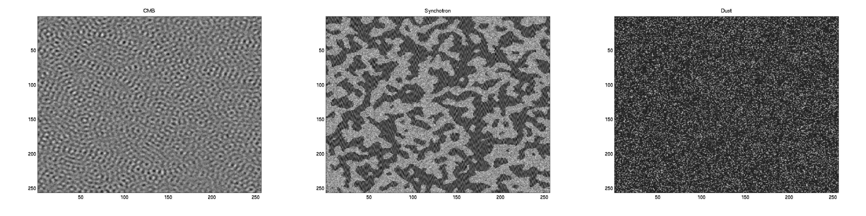


Figure: Simulated values of the 3 sources, from left to right, assigned to be CMB, synchrotron and dust.

Example 1: simulated data

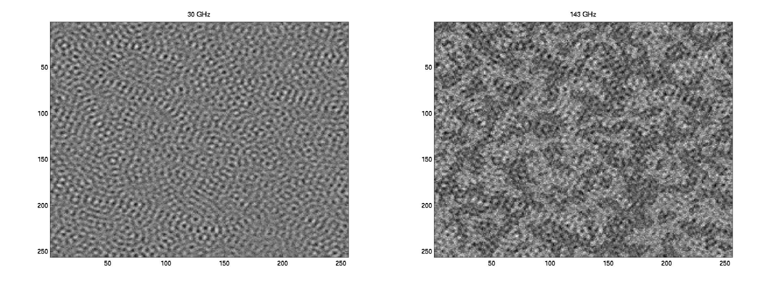


Figure: Resulting observed signal at two frequencies: 30 GHz (left) and 143 GHz (right).

Example 1: simulated data with independent prior

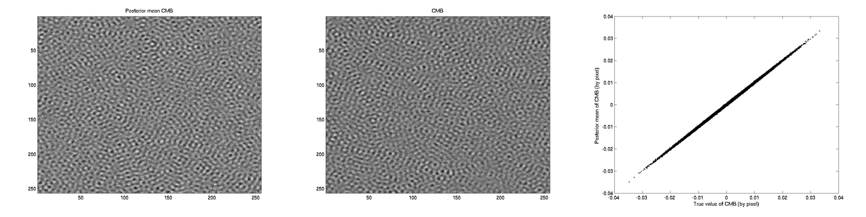


Figure: The posterior mean reconstruction of the CMB (left), the true (centre) with a scatter plot of true vs posterior mean (right).

Example 1: simulated data with independent prior

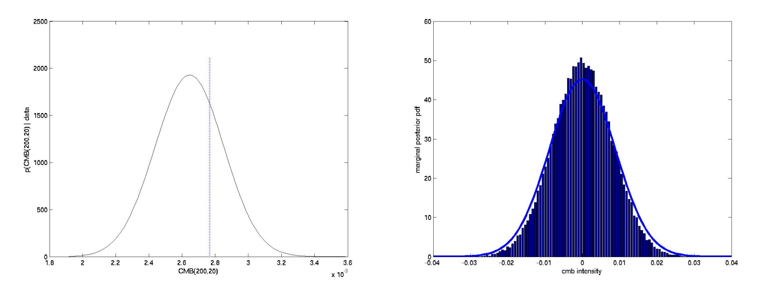
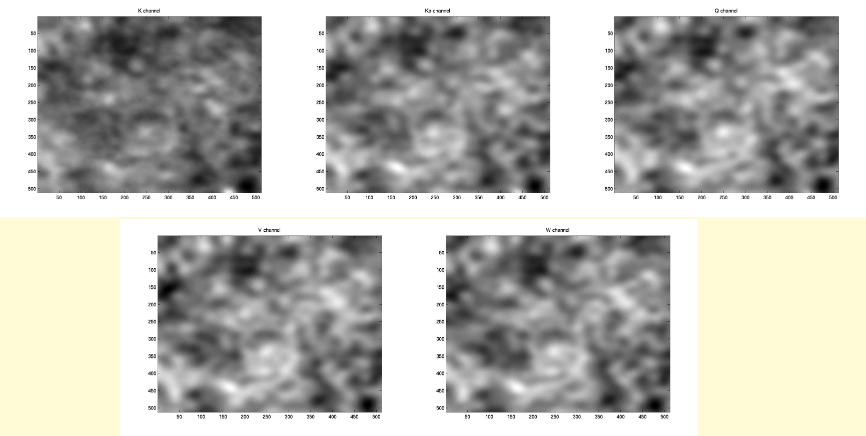


Figure: On the left, the posterior distribution of the CMB at pixel (200,20). The true value is indicated by the vertical line. On the right, the marginal posterior distribution of the CMB, with the histogram of true values for comparison.

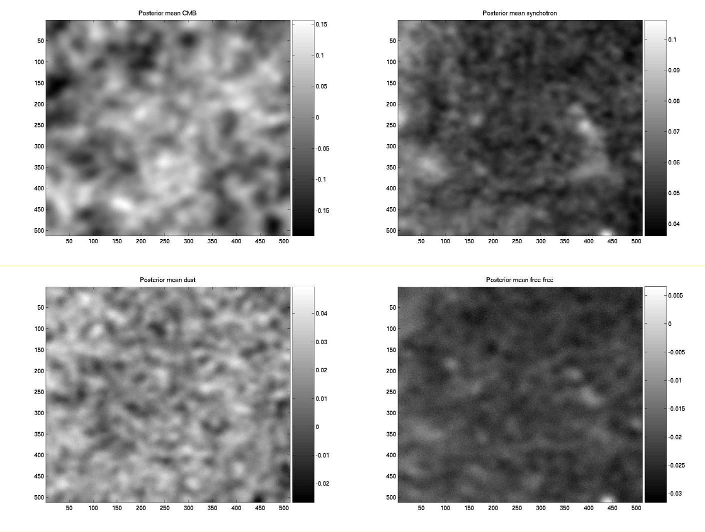
Example 2: WMAP Data

- 512×512 pixel patch from WMAP satellite at 5 channels;
- Fit 4 sources: CMB, synchrotron, dust and free-free emission;
- The spectral index of free-free emission is assumed to be -2.19.

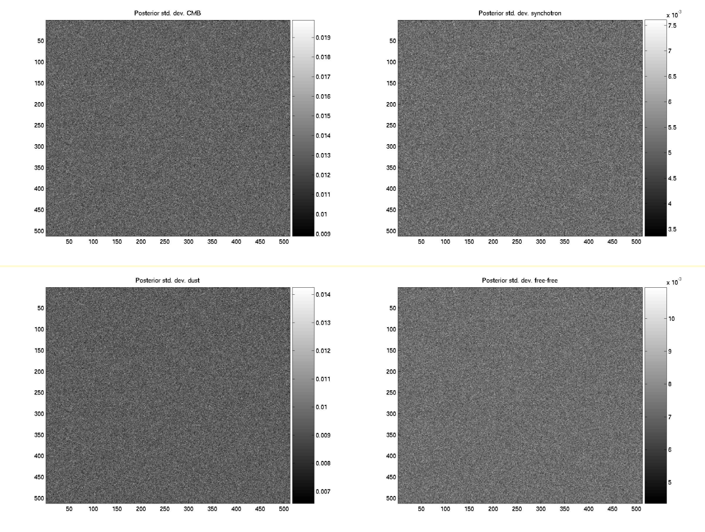
Example 2: WMAP Data



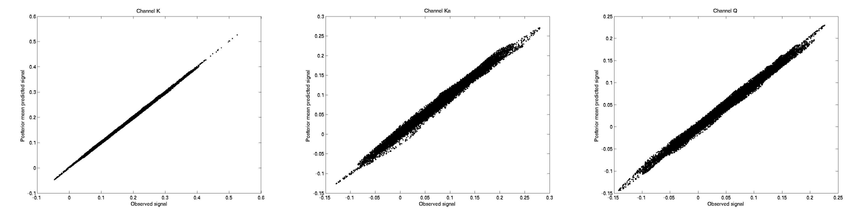
Example 2: Posterior mean of sources with independent priors

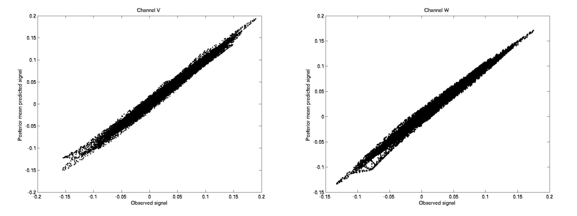


Example 2: Posterior standard deviation of sources with independent priors

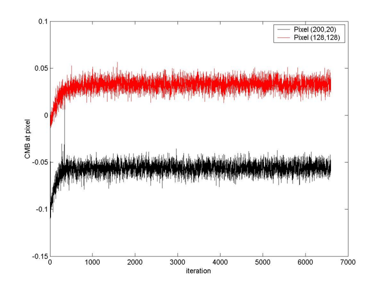


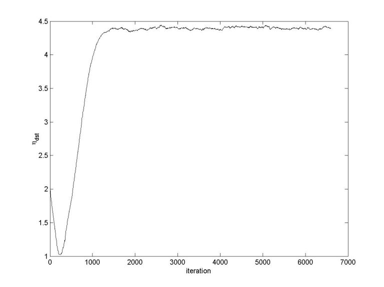
Example 2: Model fit: d_{ij} vs. $\mathbb{E}(A_{i\cdot}\mathbf{s}_j \mid \mathbf{D})$ for independent prior case



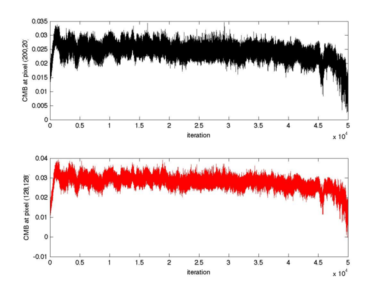


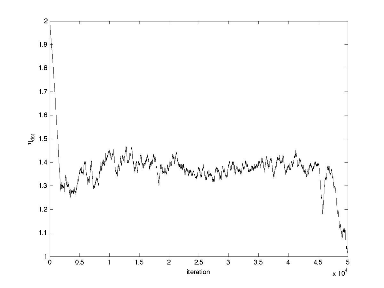
Example 2: How did the MCMC do? Independent prior case



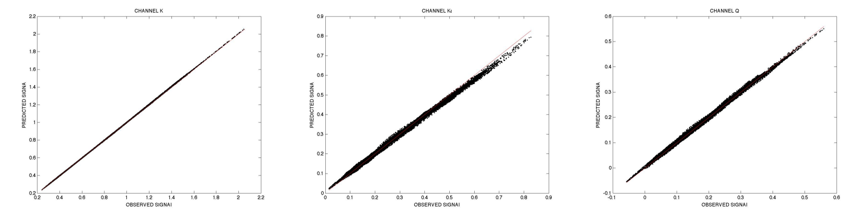


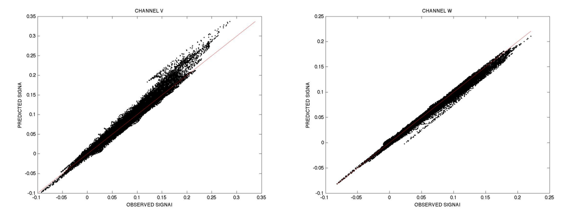
Example 2: How did the MCMC do? Between-source dependence case





Example 2: Model fit: d_{ij} vs. $\mathbb{E}(A_{i\cdot}\mathbf{s}_j \mid \mathbf{D})$ for between-source dependence case





Issues

- Currently solved problem for $J \times n_s \approx 10^6$ vs. 10^8 for whole sky reconstruction;
    - Whole sky needed to construct posterior for spectral density of CMB;
- Within pixel dependence MCMC run took 2 weeks with MATLAB!!
- How to practically implement the spatial dependence case;
- Is the prior important?
- Marginalise out the GMRF parameters (as in Nobile and Fearnside, Stats & Computing, 2007)?
- Separation at resolution of shortest wavelength;

Reference

Wilson, S.P., Kuruoğlu E. and Salerno, E. (2008). Fully Bayesian source separation of astrophysical images modelled by mixture of Gaussians. *IEEE Journal of Selected Topics in Signal Processing: Special Issue on Signal Processing for Astronomical and Space Research Applications*, **2**, 5: 685–696.