E-team on Unsupervised Segmentation
Fully Bayesian Source Separation with Application to the Cosmic Microwave Background

Simon Wilson[1] Ercan Kuruoğlu[2] Emanuele Salerno[2]

[1]School of Computer Science and Statistics, Trinity College Dublin

[2]ISTI, CNR–Pisa

The Problem — Factor Analysis

- J observations of vectors $d_1, \ldots, d_J$, $d_j \in \mathbb{R}^{n_f}$;
- Assume:

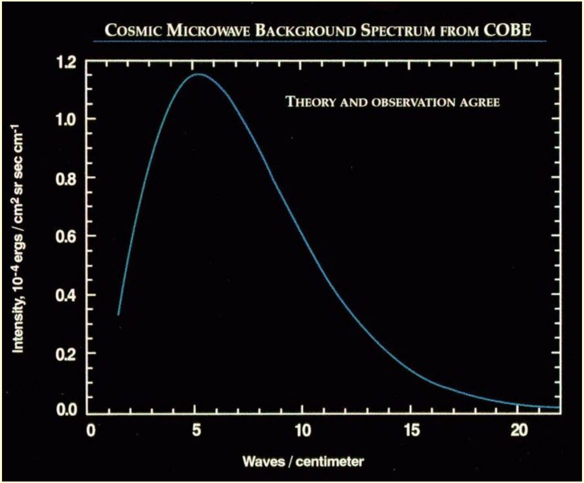
$$d_j = As_j + \epsilon_j,\ s_j \in \mathbb{R}^{n_s},$$

  where A is a $n_f \times n_s$ matrix (the mixing matrix);
- Assume $\epsilon_j \sim N(0, \mathrm{diag}(\tau_1^{-1}, \ldots, \tau_{n_f}^{-1}))$;
- Goal: observe the d_j and then infer A and recover ("separate") the s_j;

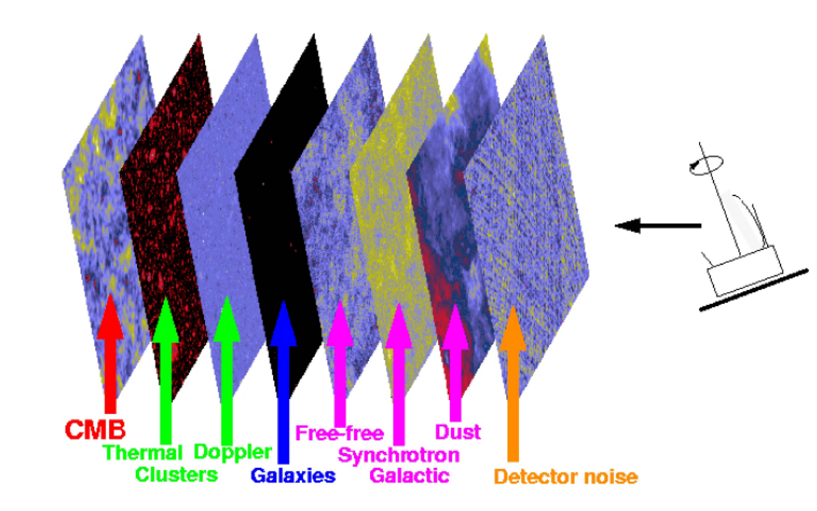
Cosmic Microwave Background (CMB)

- 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);

CMB Spectrum — Black Body



Inferring the CMB — Source Separation



Separating the Cosmic Microwave Background

- d_j, are observations at J pixels over the sky at n_f microwave frequencies $\nu_1, \ldots, \nu_{n_f}$;
  - Upcoming data (Planck satellite) will have $J \approx 10^7$, $n_f = 9$ at frequencies from 30 to 857 GHz;
- s_j are the sources that make up the microwave received by the satellite:
  - One of these sources is the CMB (source 1);
  - Other important ones are synchrotron radiation and galactic dust;
  - There are 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*.

Model

We can put all the d_j and s_j into matrices:

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

Model for Sources

- Each source $S_k = \{s_{kj} \mid j = 1, \ldots, J\}$ is an iid Gaussian mixture with an unknown number m_k of components.
- Define $\mu_k = (\mu_{k1}, \ldots, \mu_{km_k})$, $t_k = (t_{k1}, \ldots, t_{km_k})$ and $p_k = (p_{k1}, \ldots, p_{km_k})$ to be the mixture component means, precisions and weights for source k;
- So

$$\begin{aligned} & p(S_k \mid \mu_k, t_k, p_k) \\ & = \prod_{j=1}^{J} \sum_{a=1}^{m_k} p_{ka} \sqrt{\frac{t_{ka}}{2\pi}} \exp\left(-0.5 t_{ka}(s_{kj} - \mu_{ka})^2\right), s_{kj} \in \mathbb{R}. \end{aligned}$$

- Let $\underline{\mu} = (\mu_1, \ldots, \mu_{n_s})$, $\underline{t} = (t_1, \ldots, t_{n_s})$, $\underline{p} = (p_1, \ldots, p_{n_s})$ and $\underline{m} = (m_1, \ldots, m_{n_s})$ denote the vectors of all mixture means, precisions, weights and no. of components for all sources.

Model for Mixing Matrix A

- Both A and s unknown $\Rightarrow$ solution up to a constant in each column (source) of A;
- Hence can arbitrarily fix one value in each column 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

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

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

Model for Mixing Matrix A

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

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

  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;

Priors

- We've parameterised the model in terms of:
  - Source mixture means, precisions, weights and no. of components: $\underline{\mu}$, $\underline{t}$, $\underline{p}$, $\underline{m}$;
  - Mixing matrix parameters θ;
  - Measurement noise precisions $\underline{\tau}$;
- We put the usual conjugate priors on these:
  - Normals on the mixture means;
  - Gammas on the mixture precisions;
  - Dirichlets on the mixture weights;
  - Geometrics on the no. of mixture components;
  - Gammas on the noise precisions;
  - For θ, physical arguments put quite tight bounds on their values — we put normal priors with high probability between these bounds;
- Our existing knowledge can put *very informative* priors on the source mixture parameters, and on the noise precisions;
- These should *greatly* help the inference.

Sampling from the Posterior Distribution

- Can be done by Gibbs sampling;
- Update parameters in blocks where possible:
  - Mixture means, precisions and weights updated jointly from their full conditional for each source by a Gibbs sampler;
  - No. of mixture components for each source sampled by the usual Richardson and Green (JRSS B, 1997) reversible jump move;
  - Components of θ updated jointly with their corresponding source by a Metropolis move (e.g. (θ_k, S_k));
  - Full conditional of each source at each pixel s_{kj} is a mixture of Gaussians;
    - Better: full conditional of vector of sources at each pixel $S_{\cdot j}$ is a multivariate mixture of Gaussians;
  - Full conditional of noise precisions are gamma;
- See pending paper for the details!!

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 5 COBE frequencies, giving:

$$A = \begin{pmatrix} 0.9770 & 32.8359 & 0.0990 \\ 0.9514 & 10.8140 & 0.2090 \\ 0.8823 & 2.8133 & 0.5107 \\ 0.7770 & 1.0000 & 1.0000 \\ 0.6044 & 0.3544 & 1.9256 \end{pmatrix}.$$

Example 1: simulated data

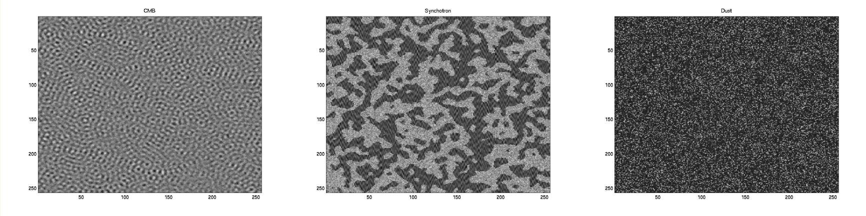


Figure: Simulated values of the 3 sources, from left to right, assigned to be CMB, synchrotron and dust.

Example 1: simulated data

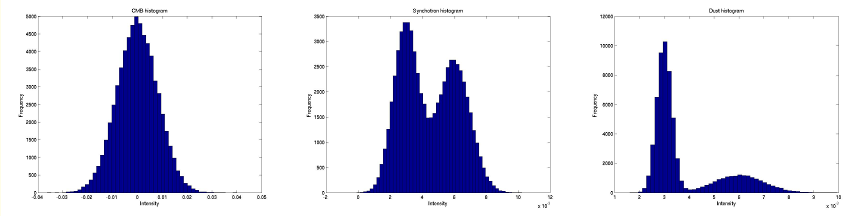


Figure: Histograms of the simulated values of the 3 sources, from left to right: CMB, synchrotron and dust.

Example 1: simulated data

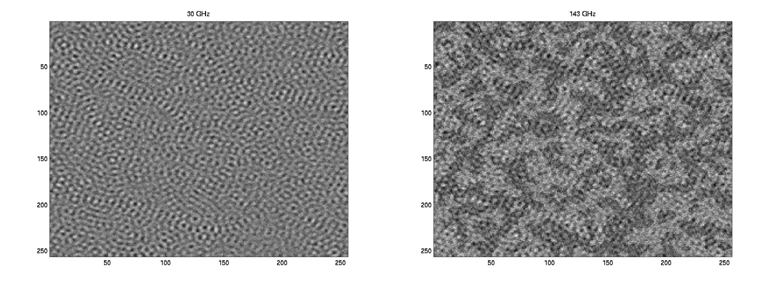


Figure: Resulting observed signal at two frequencies: 30 GHz (left) and 143 GHz (right).

Example 1: simulated data

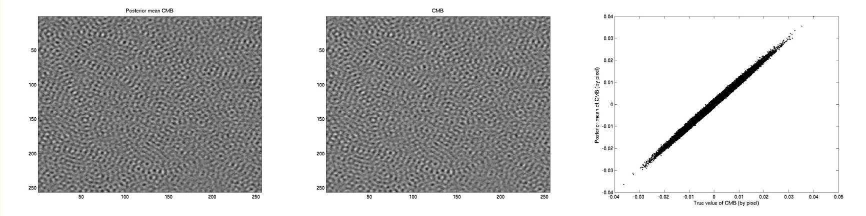


Figure: The posterior mean reconstruction of the CMB (left), the true (centre) with a scatter plot of true vs posterior mean (right).

Example 1: simulated data

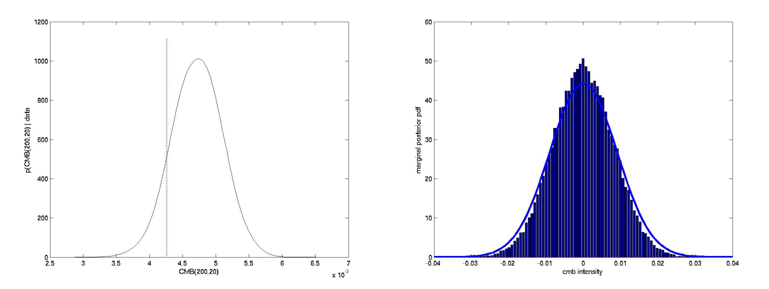


Figure: On the left, the posterior distribution of the CMB at pixel (200,20). The true value is indicated by the vertical line. On the right, the marginal posterior distribution of the CMB, with the histogram of true values for comparison.

Example 1: simulated data

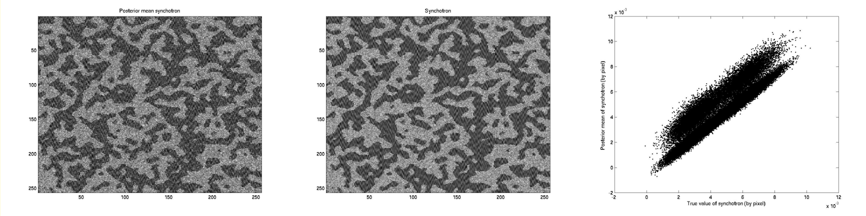


Figure: The posterior mean reconstruction of synchrotron (left), the true (centre) with a scatter plot of true vs posterior mean (right).

Example 1: simulated data

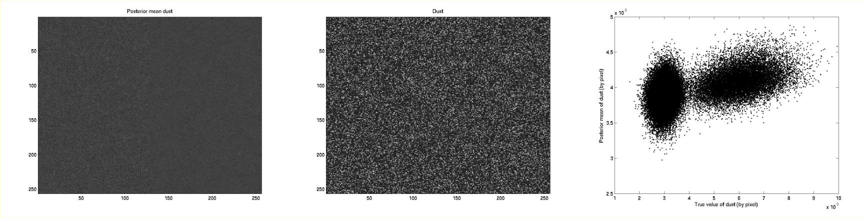


Figure: The posterior mean reconstruction of dust (left), the true (centre) with a scatter plot of true vs posterior mean (right).

Example 1: simulated data

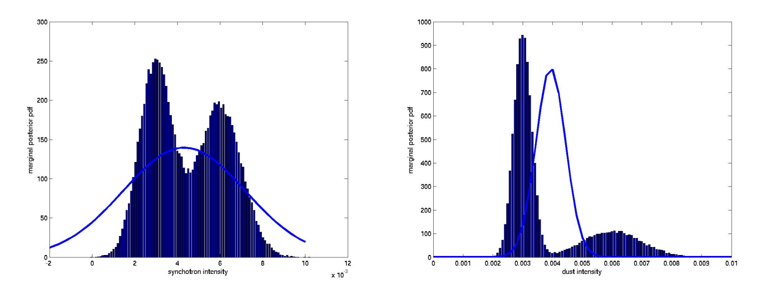


Figure: The fitted marginal posterior distribution of synchrotron (left) and dust (right), along with the histogram of their true values for comparison.

Example 2: simulated data with 9 channels

- Same 5 channels as example 1 with 4 more at higher frequencies;
- $$A = \begin{pmatrix} 0.9770 & 32.8359 & 0.0990 \\ 0.9514 & 10.8140 & 0.2090 \\ 0.8823 & 2.8133 & 0.5107 \\ 0.7770 & 1.0000 & 1.0000 \\ 0.6044 & 0.3544 & 1.9256 \\ 0.2194 & 0.1057 & 5.3763 \\ 0.0294 & 0.0258 & 11.3455 \\ 0.0019 & 0.0073 & 17.5890 \\ 0.0001 & 0.0020 & 21.9472 \end{pmatrix},$$
- Higher frequencies give much higher to dust.

Example 2: simulated data with 9 channels

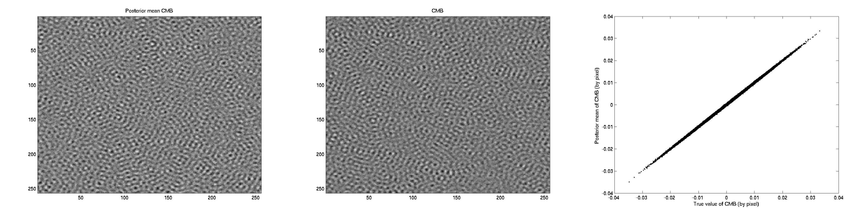


Figure: The posterior mean reconstruction of the CMB (left), the true (centre) with a scatter plot of true vs posterior mean (right).

Example 2: simulated data with 9 channels

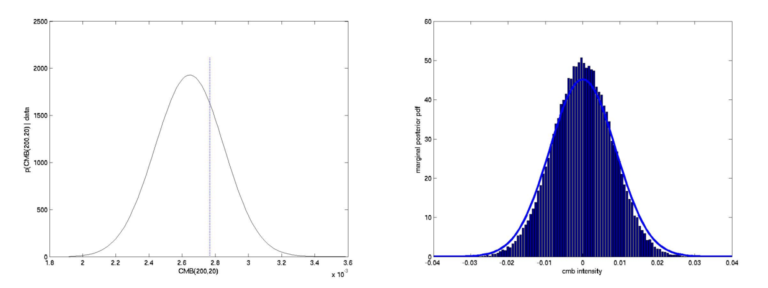


Figure: On the left, the posterior distribution of the CMB at pixel (200,20). The true value is indicated by the vertical line. On the right, the marginal posterior distribution of the CMB, with the histogram of true values for comparison.

Example 2: simulated data with 9 channels

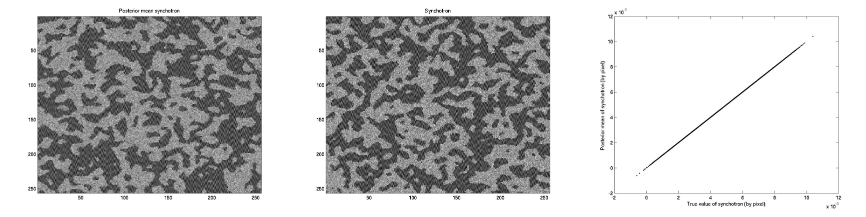


Figure: The posterior mean reconstruction of synchrotron (left), the true (centre) with a scatter plot of true vs posterior mean (right).

Example 2: simulated data with 9 channels

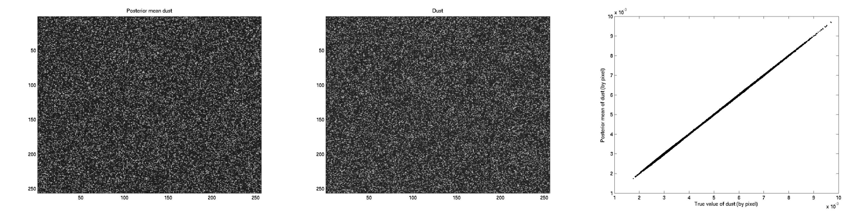


Figure: The posterior mean reconstruction of dust (left), the true (centre) with a scatter plot of true vs posterior mean (right).

Example 2: simulated data with 9 channels

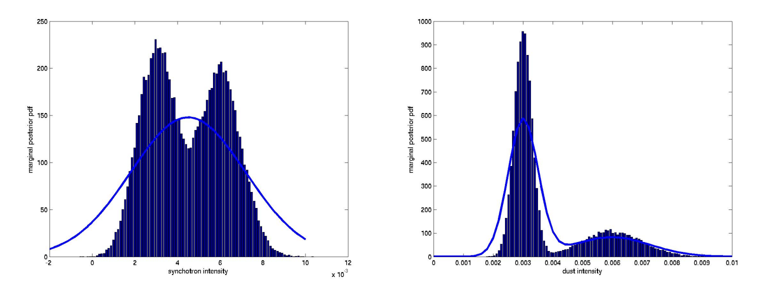
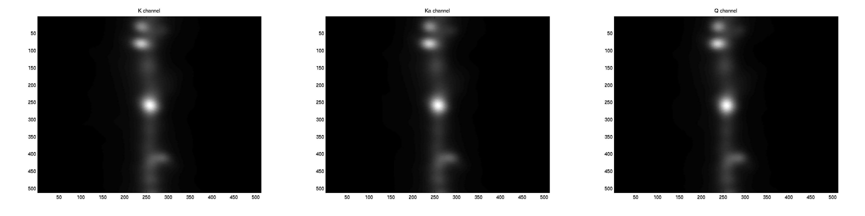


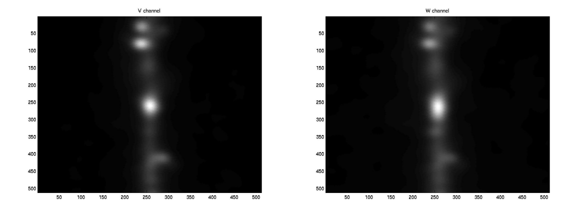
Figure: The fitted marginal posterior distribution of synchrotron (left) and dust (right), along with the histogram of their true values for comparison.

Real WMAP Data

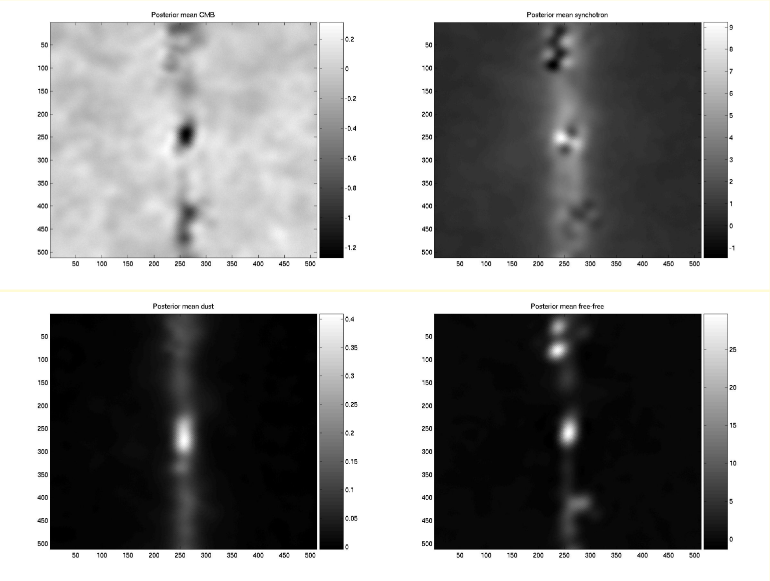
- Three patches of 512×512 pixels at 5 channels;
- Fit 4 sources: CMB, synchrotron, dust and free-free emission;
- The spectral index of free-free emission is assumed to be -2.19.

Patch 1 — Data

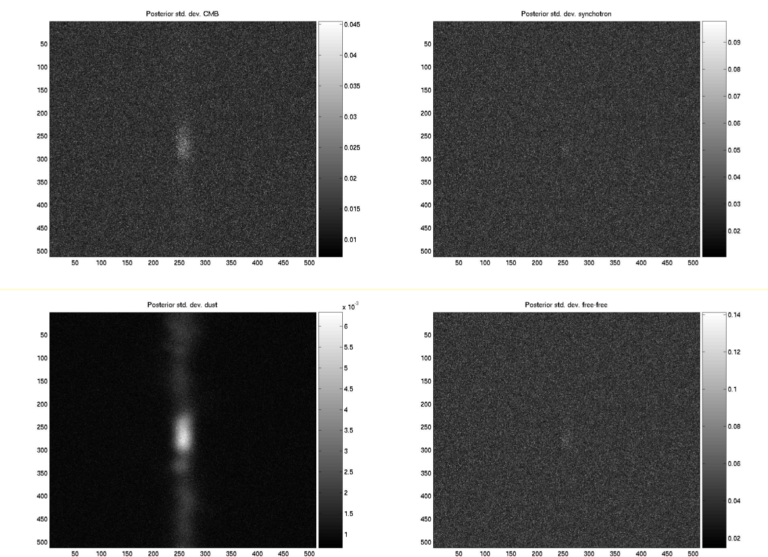

K channel
Ka channel
Q channel
50 100 150 200 250 300 350 400 450 500


V channel
W channel
50 100 150 200 250 300 350 400 450 500

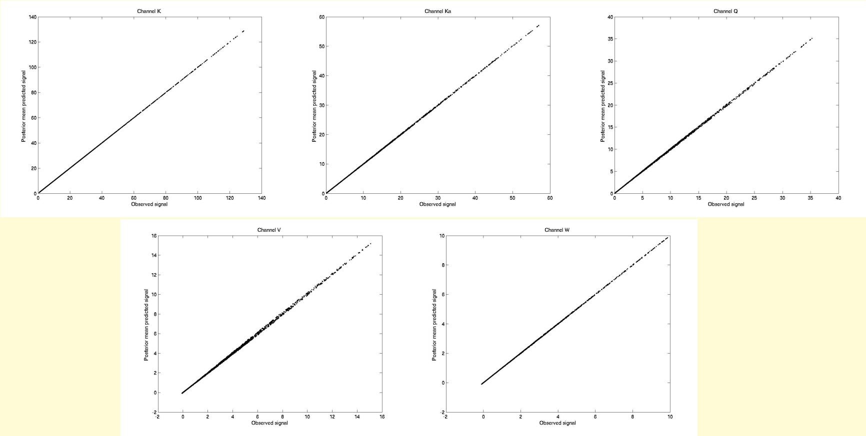
Patch 1 — Posterior mean of sources



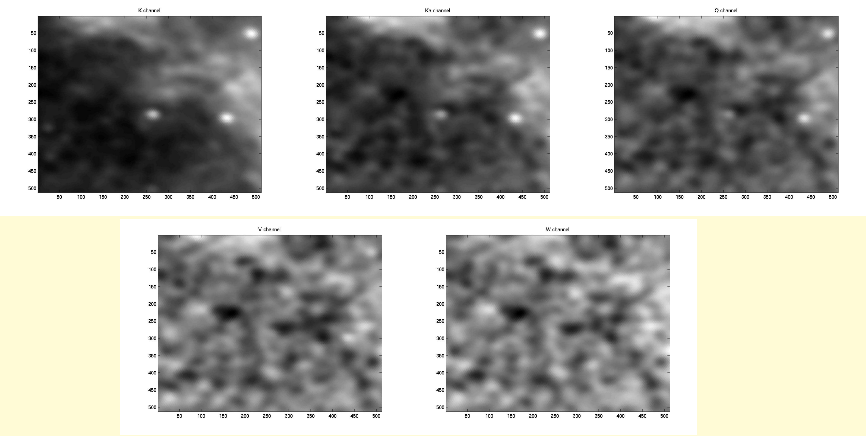
Patch 1 — Posterior standard deviation of sources



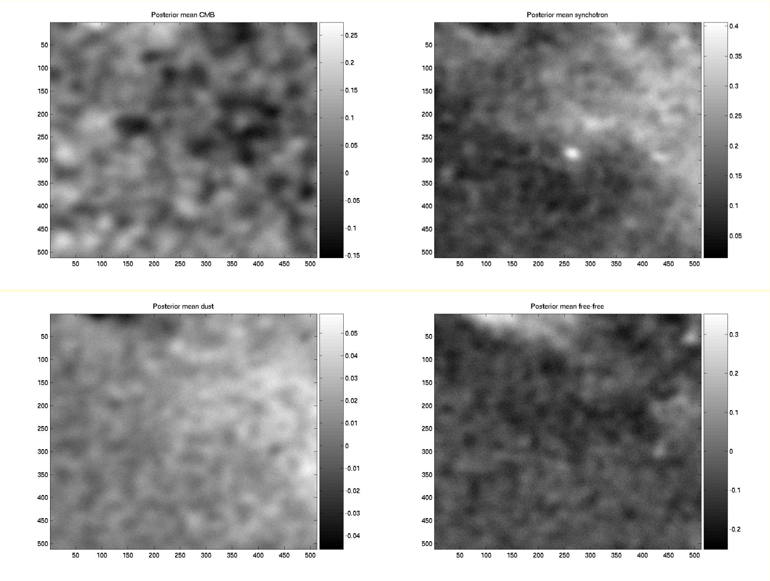
Patch 1 — Model fit: observed temperature vs. posterior mean temperature



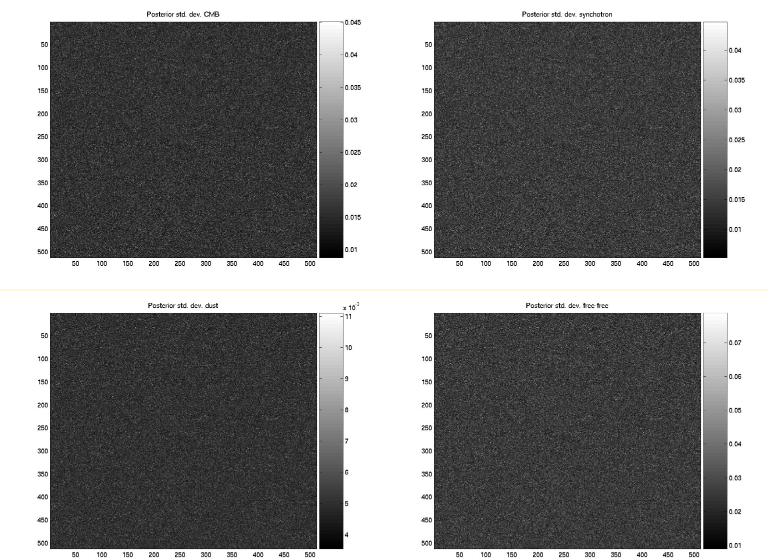
Patch 2 — Data


K channel
Ka channel
Q channel
V channel
W channel
50
100
150
200
250
300
350
400
450
500

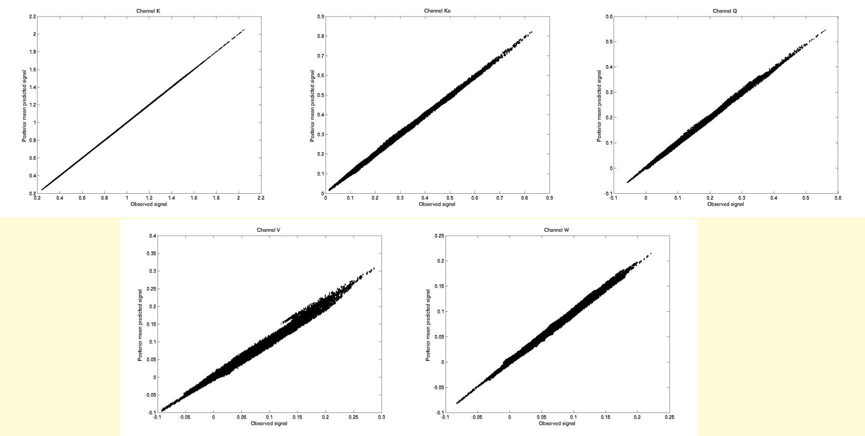
Patch 2 — Posterior mean of sources



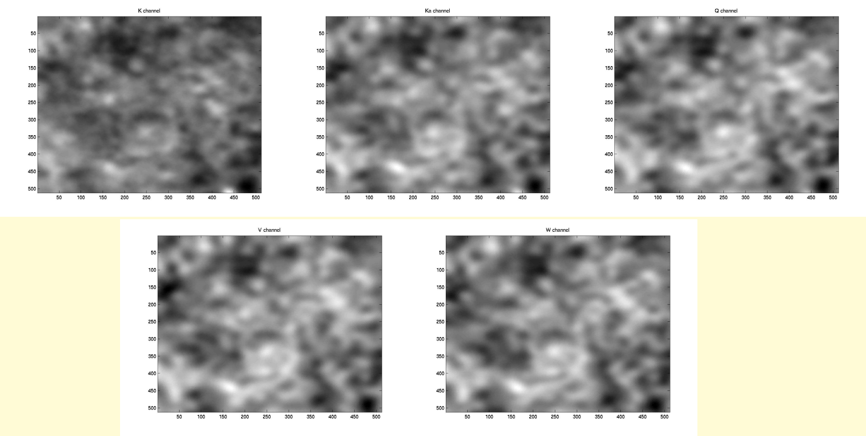
Patch 2 — Posterior standard deviation of sources



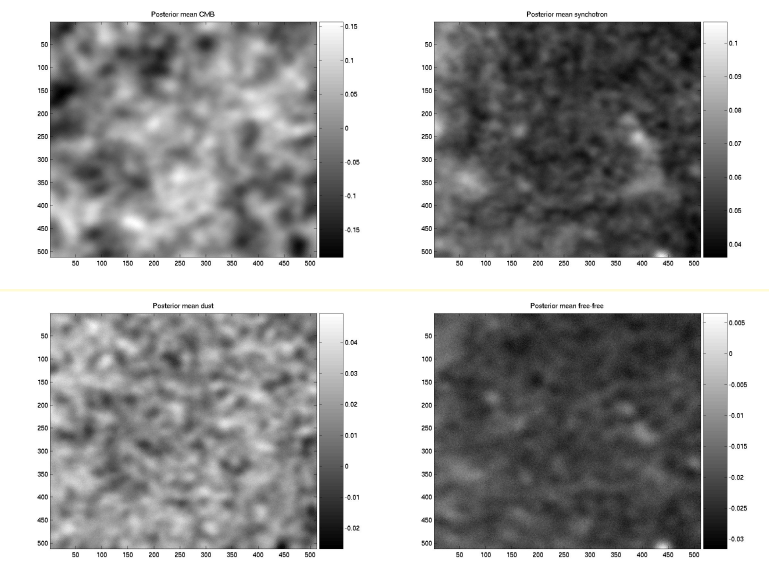
Patch 2 — Model fit: observed temperature vs. posterior mean temperature



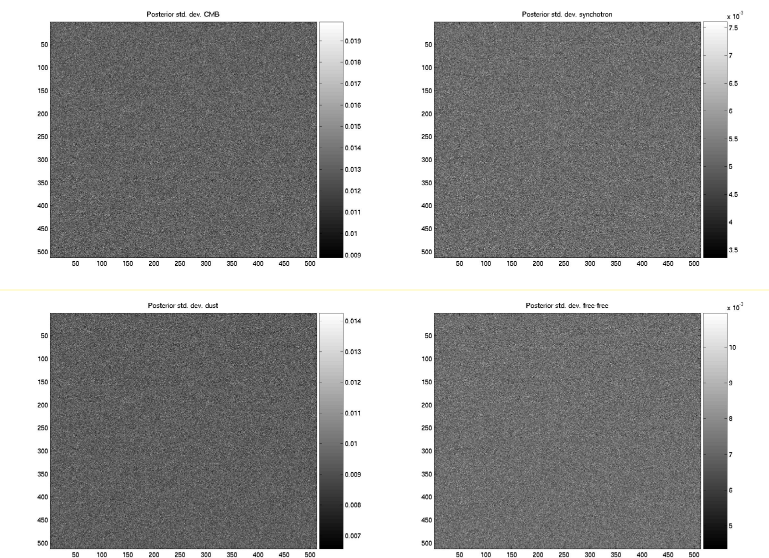
Patch 3 — Data



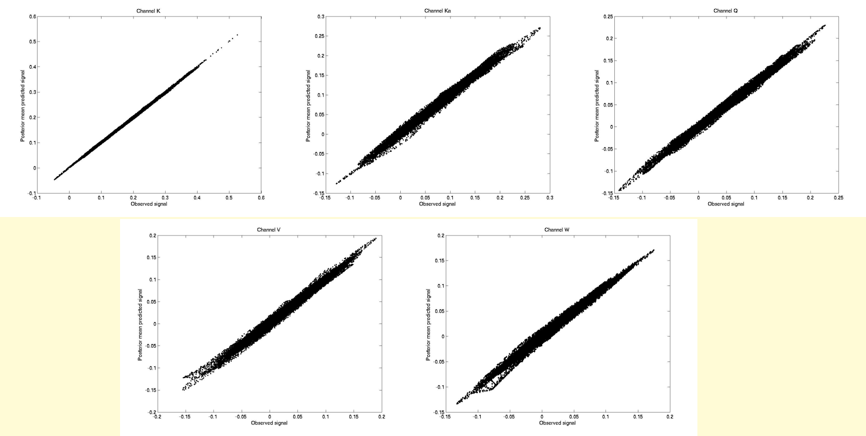
Patch 3 — Posterior mean of sources



Patch 3 — Posterior standard deviation of sources



Patch 3 — Model fit: observed temperature vs. posterior mean temperature



Posterior mean of spectral indices

	Synchrotron	Dust
Patch 1	-2.61	3.40
Patch 2	-2.84	1.34
Patch 3	-2.64	0.51

Dust spectral index varies considerably from patch to patch!!
Note: free-free spectral index assumed to be -2.19 (as in Eriksen et al., 2006).

Future Work

- Dependent sources (CMB vs. galactic vs. extra-galactic) by mixtures of multivariate Gaussians;
- Analysis of entire WMAP data;
- Separation at resolution of shortest wavelength;
- Planck data from 2008.