Blind Component Separation in Wavelet Space. Application to CMB Data Analysis.

Yassir Moudden

CEA/DAPNIA/SEDI-SAP

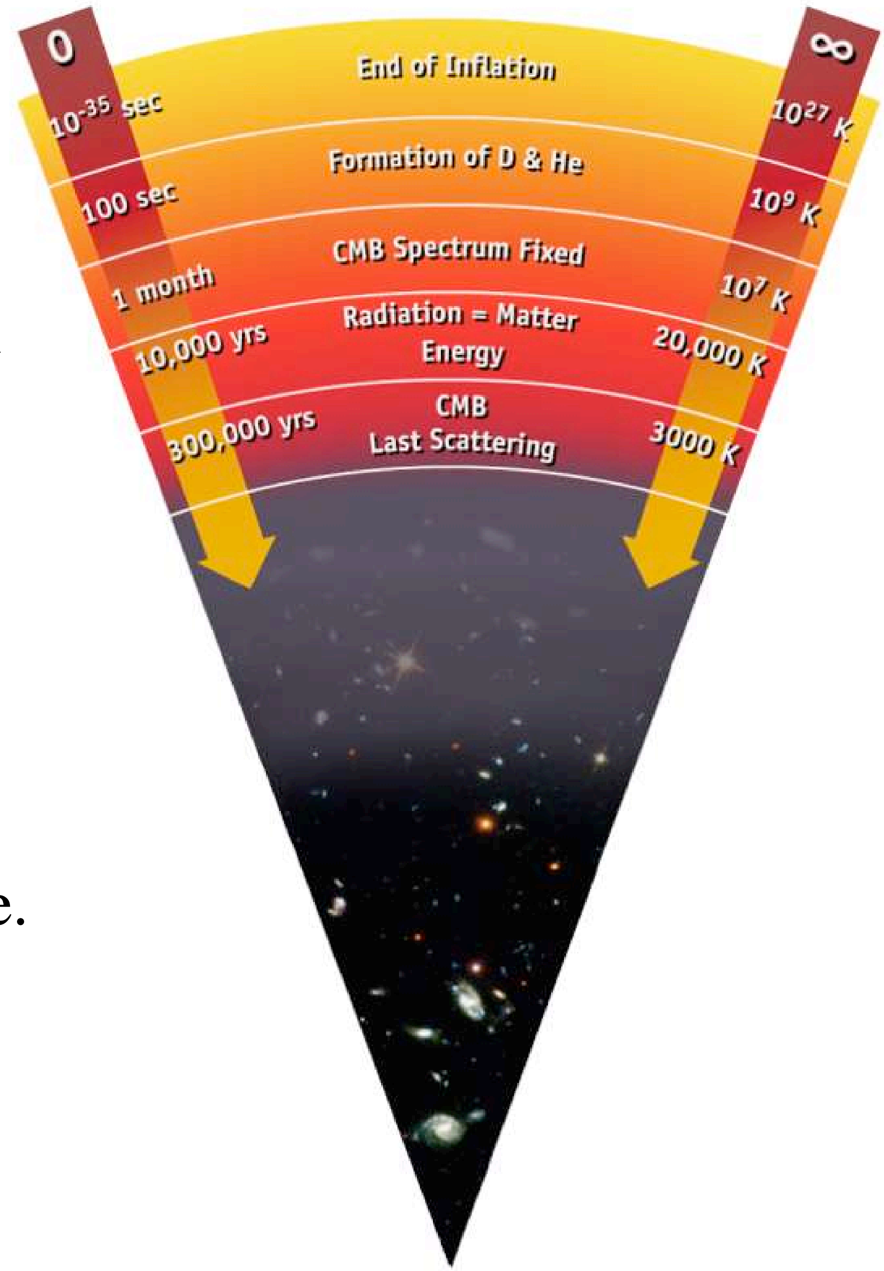
Joint work with :
- Jean-François Cardoso , CNRS-ENST
- Jacques Delabrouille, PCC/Collège de France
- Jean-Luc Starck, Pierrick Abrial, CEA/DAPNIA/SEDI-SAP

Outline

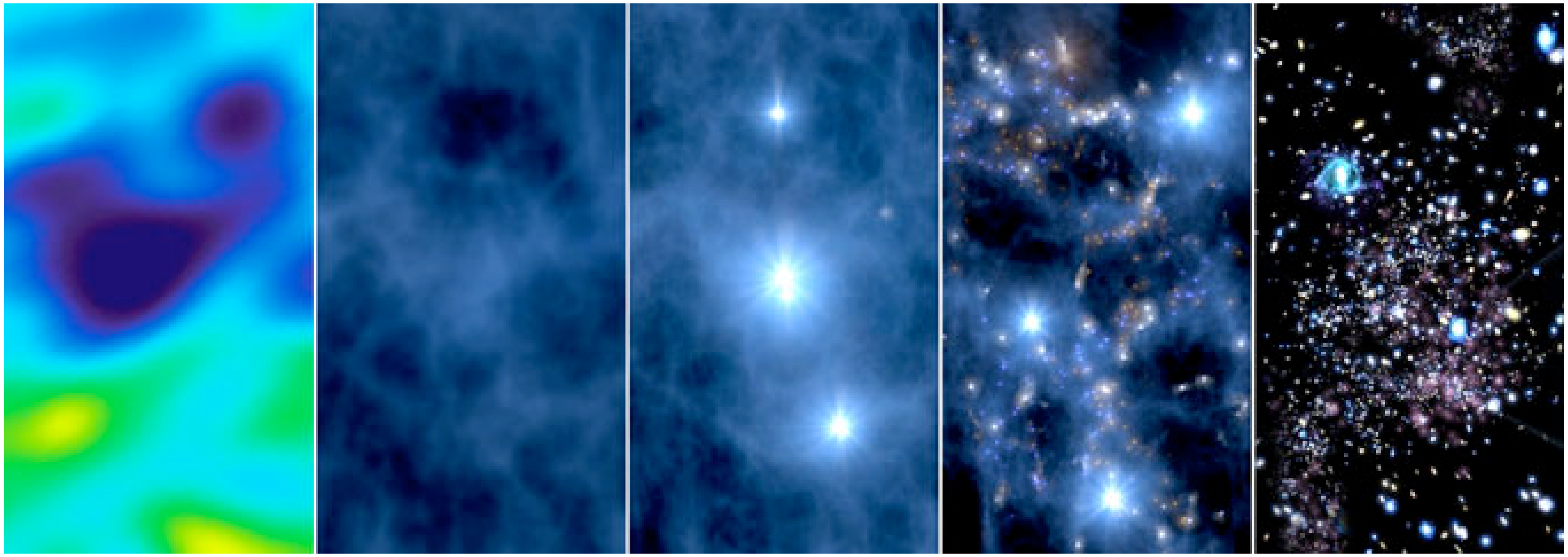
- What is Cosmic Microwave Background Radiation?
- Experimental issues
- A static linear mixture model
- ICA methods in Blind Source Separation
- Spectral Matching ICA
- A *variation* on SMICA using wavelets

The Big Bang , etc.

- At the beginning, it is all very hot.
- Fortunately, the Universe is expanding and cooling down.
- Around 300 000 yrs after BB, photons and baryons decouple.



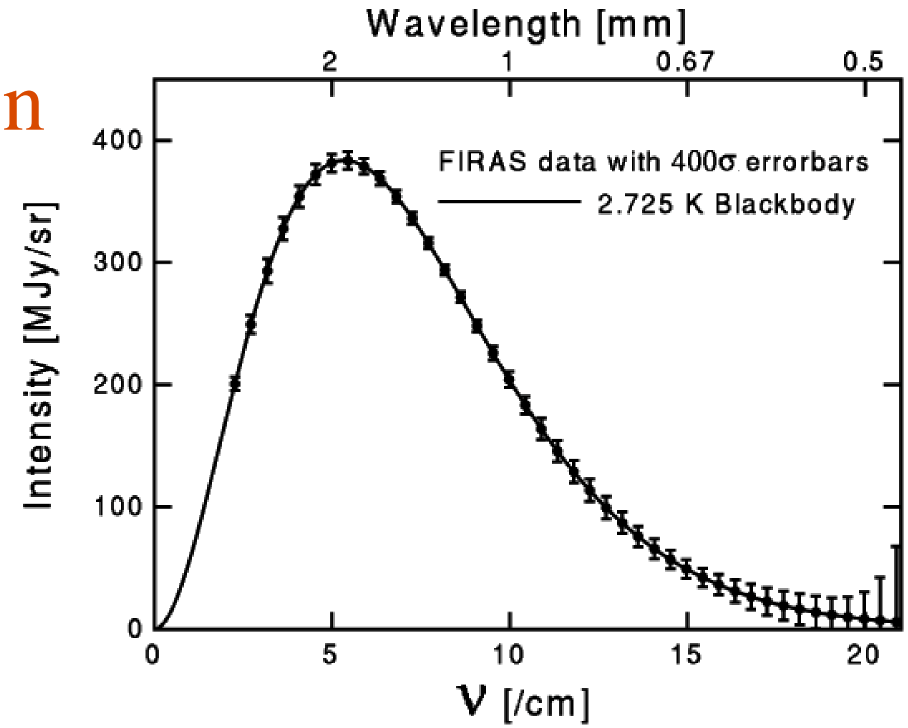
The destiny of matter

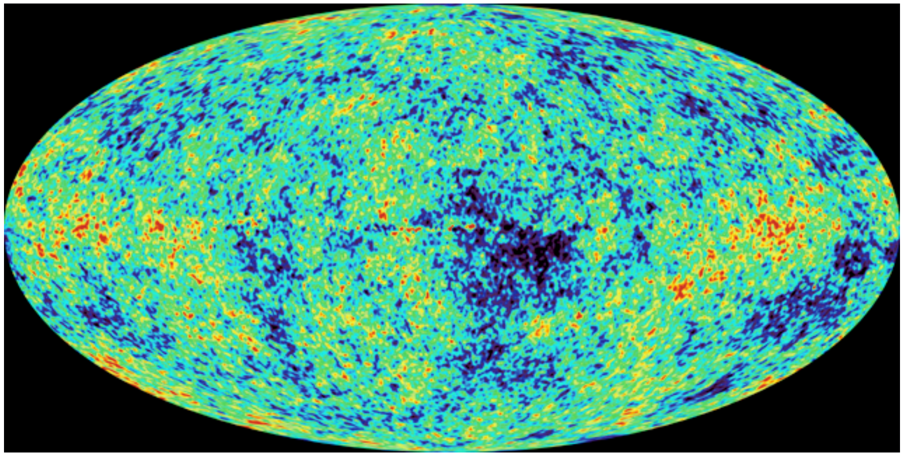


Small density fluctuations are the seeds of large scale structures.

The destiny of radiation

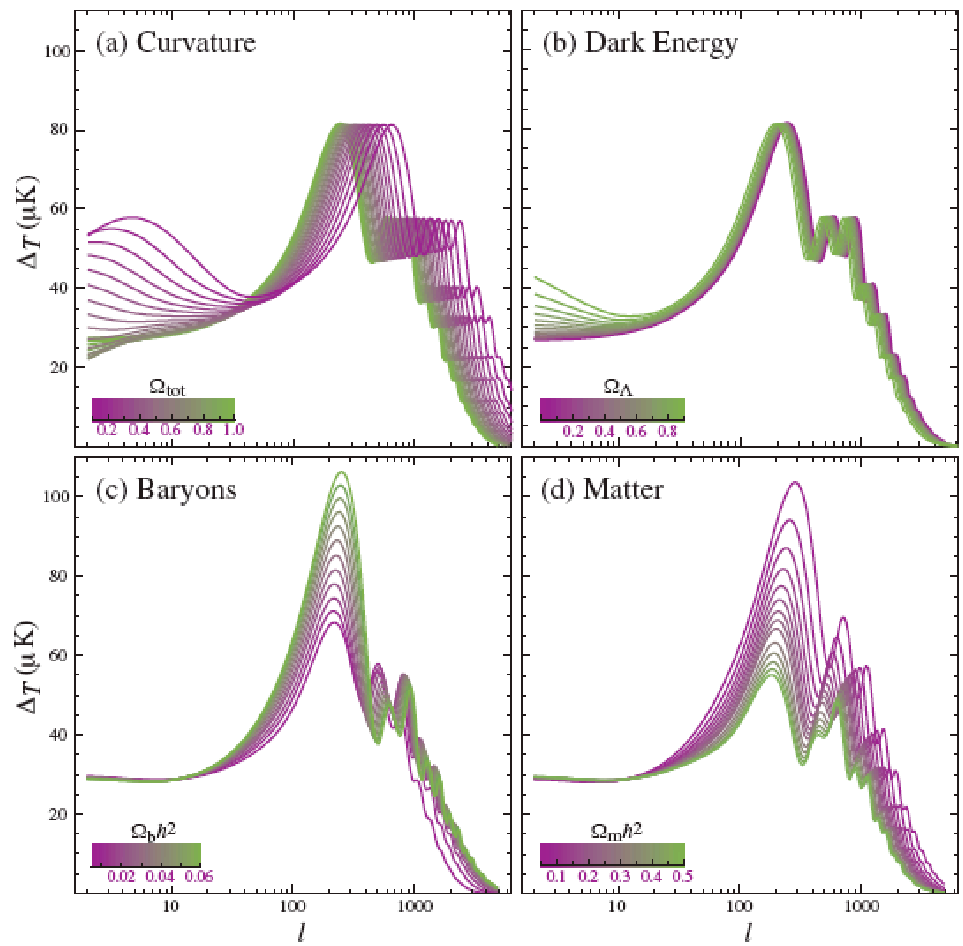
- Wander endlessly through the Universe …
- CMB photons have cooled down from 3000K to 3K.
- Planck’s blackbody spectrum depends only on T.
- $\Delta T / T \sim 10^{-5}$





Modeling the CMB anisotropies

- Linearized dynamics
- Parametric model
- Input: scale invariant gaussian fluctuations
- Output is also gaussian, completely defined by its power spectrum.



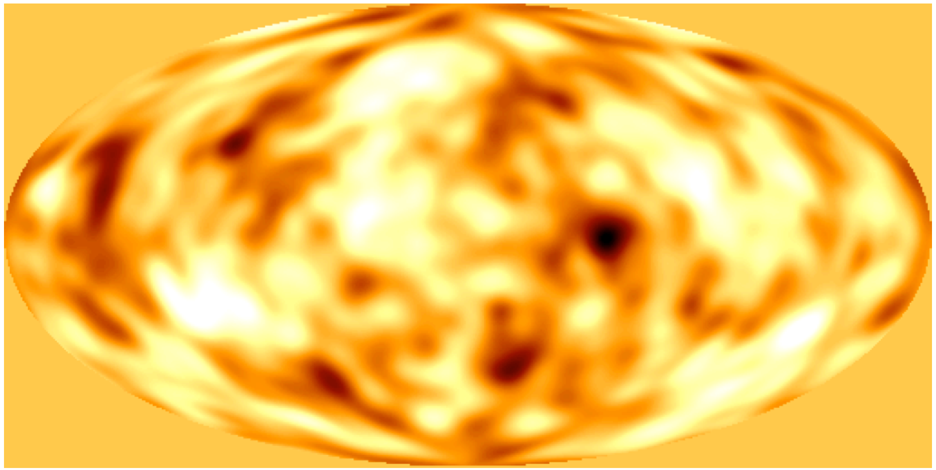
Measuring the CMB anisotropies

- Penzias & Wilson (1965)
- COBE (1990)
- Boomerang, Maxima, Archeops, etc.
- WMAP
- PLANCK (2007)

- Better angular resolution, sensitivity, sky coverage
- Reduce instrumental noise levels, foreground contamination

Measuring the CMB anisotropies

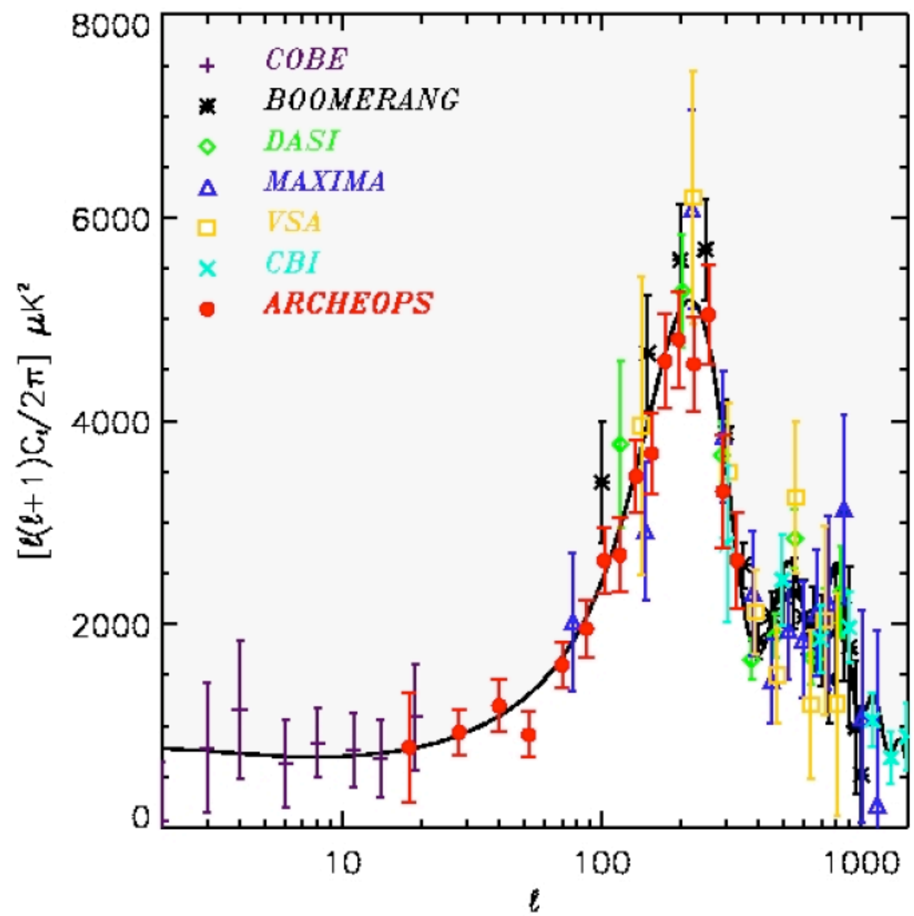
- Penzias & Wilson (1965)
- COBE (1990)
- Boomerang, Maxima, Archeops, etc.
- WMAP
- PLANCK (2007)



- Better angular resolution, sensitivity, sky coverage
- Reduce instrumental noise levels, foreground contamination

Measuring the CMB anisotropies

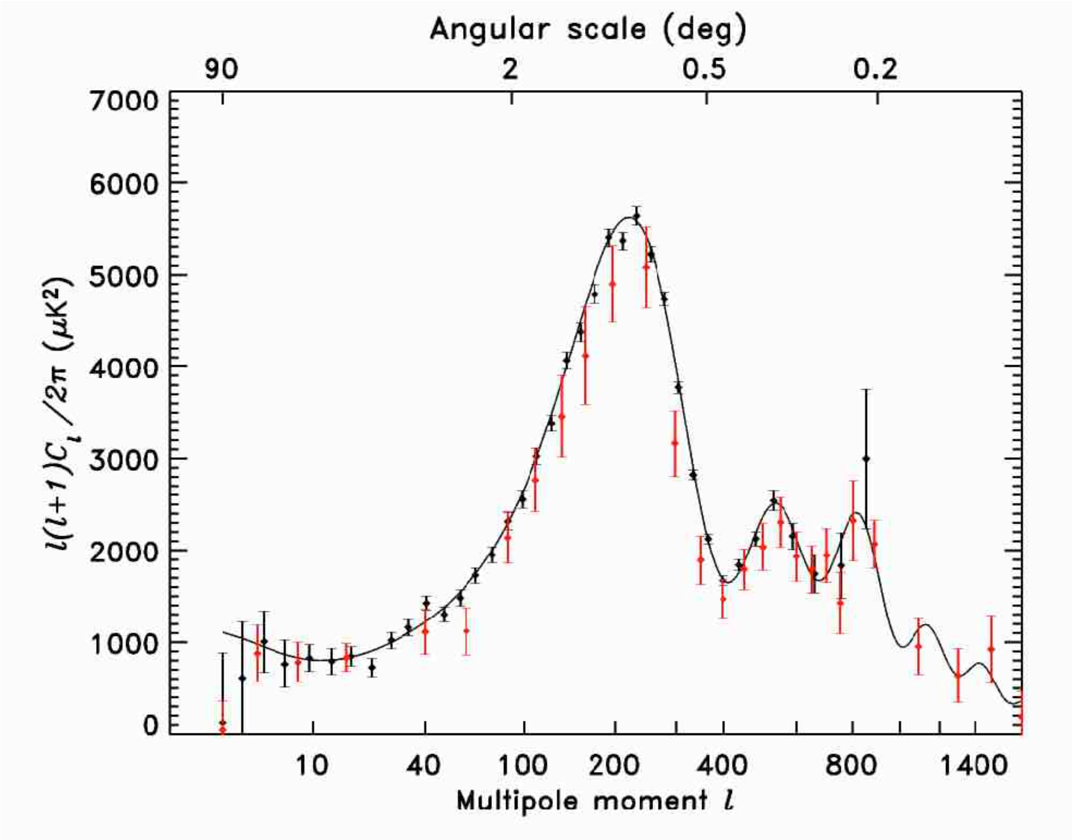
- Penzias & Wilson (1965)
- COBE (1990)
- Boomerang, Maxima, Archeops, etc.
- WMAP
- PLANCK (2007)



- Better angular resolution, sensitivity, sky coverage
- Reduce instrumental noise levels, foreground contamination

Measuring the CMB anisotropies

- Penzias & Wilson (1965)
- COBE (1990)
- Boomerang, Maxima, Archeops, etc.
- WMAP
- PLANCK (2007)

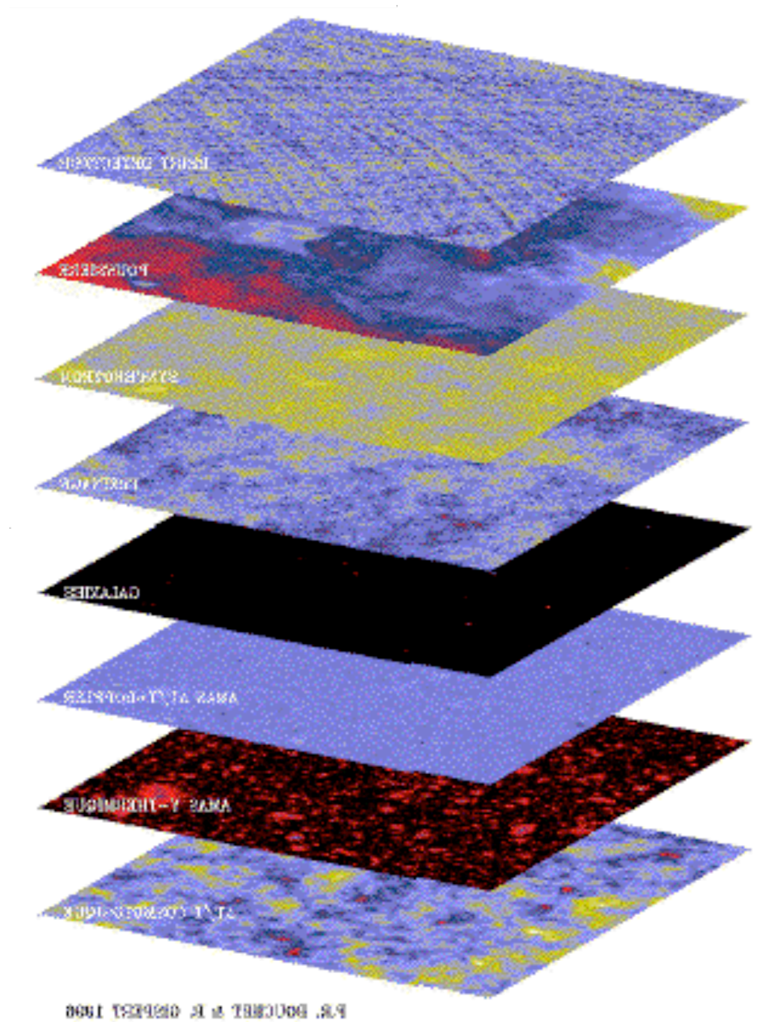


- Better angular resolution, sensitivity, sky coverage
- Reduce instrumental noise levels, foreground contaminations

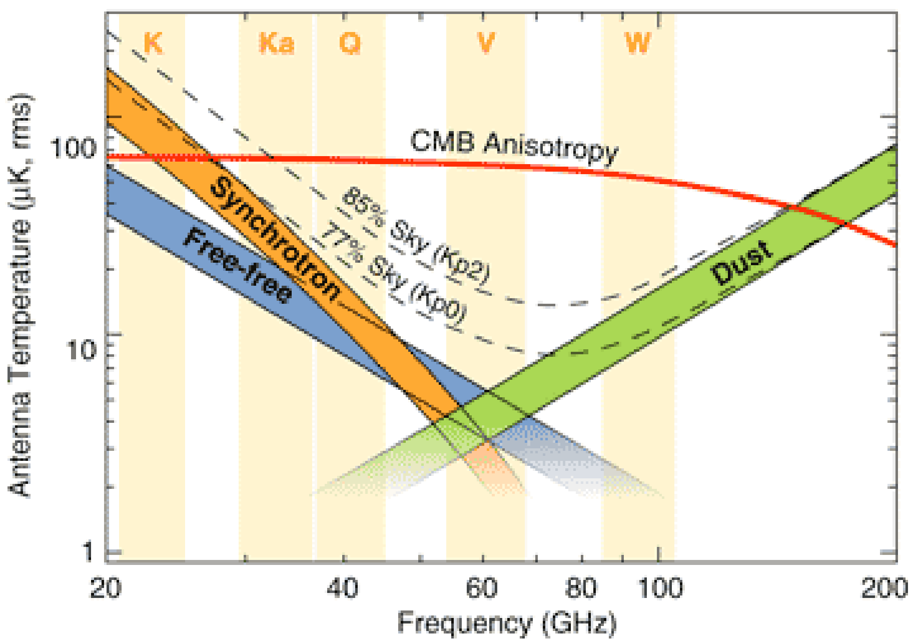
Foregrounds, foregrounds everywhere!

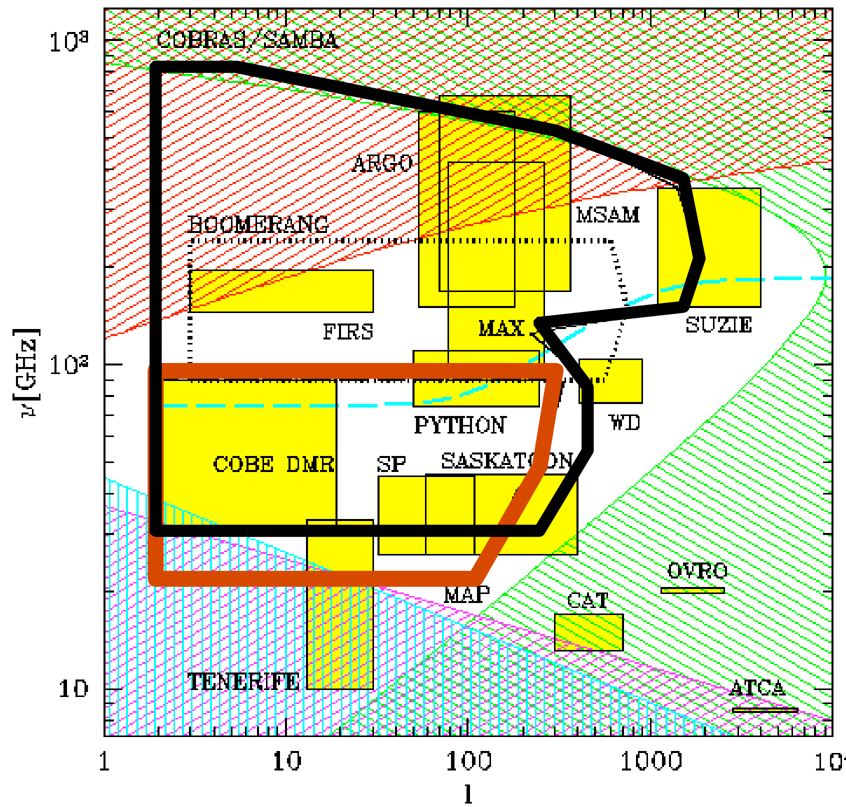
Other contributions in the microwave range :

- Detector noise
- Galactic dust
- Synchrotron
- Free - Free
- Point Sources
- Thermal SZ
- …



Foregrounds, foregrounds everywhere?





Multispectral imaging of the microwave sky.

A static linear mixture model

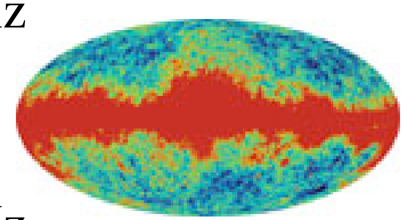
- map observed in a given frequency band sums the contributions of various astrophysical components
- the contributions of a component to two different detectors differ only in intensity.
- additive noise
- all maps are at the same resolution

$$x_d(\vartheta, \varphi) = \sum_{j=1}^{N_c} A_{dj} s_j(\vartheta, \varphi) + n_d(\vartheta, \varphi)$$

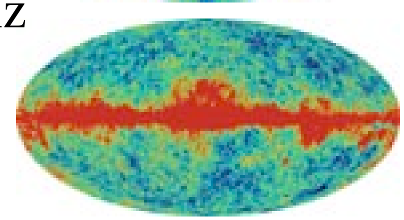
$$X(\vartheta, \varphi) = AS(\vartheta, \varphi) + N(\vartheta, \varphi)$$

A static linear mixture model

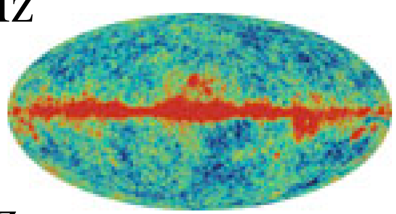
23 GHz



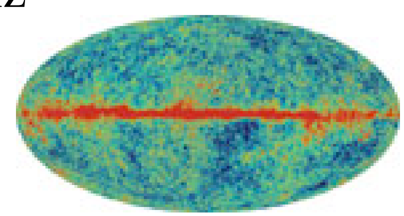
33 GHz



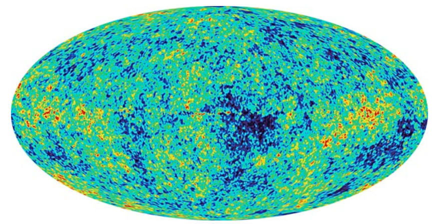
41 GHz

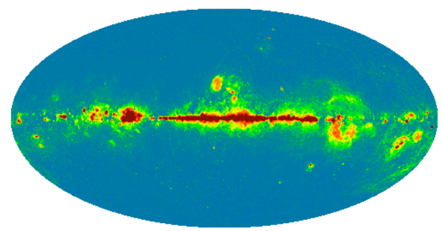


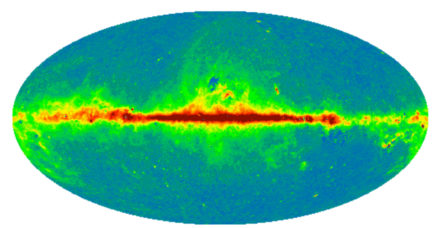
94 GHz



$$= \begin{array}{c} \begin{array}{ccc} \text{CMB} & \text{Free-free} & \text{Synchrotron} \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix} \end{array} \times$$







$$+ \begin{bmatrix} n_1(\vec{r}) \\ n_2(\vec{r}) \\ n_3(\vec{r}) \\ n_4(\vec{r}) \end{bmatrix}$$

Foreground removal as a source separation problem.

- Processing coherent observations.

$$X(\vartheta, \varphi) = AS(\vartheta, \varphi) + N(\vartheta, \varphi)$$

- All component maps are valuable !

- How much prior knowledge?
  - Mixing matrix, component and noise spatial covariances known [Tegmark96, Hobson98].
  - Subsets of parameters known.
  - Blind Source Separation assuming independent processes [Baccigalupi00, Maino92, Kuruoglu03, Delabrouille03].

Independent Component Analysis

- The noiseless case : $X = AS$
- Goal is to find a matrix B such that the entries of $Y = BX$ are independent.
- Existence/Uniqueness :
  - Most often, such a decomposition does not exist.
  - Theorem [Darmois, Linnik 1950] : Let S be a random vector with independent entries of which at most when is Gaussian, and C an invertible matrix. If the entries of $Y = CS$ are independent, then C is almost the identity.

➡ A mixture of non-Gaussian sources can be demixed by restoring independence (JADE, fastICA, InfoMax, etc.)

ICA for gaussian processes?

Two linearly mixed independent Gaussian processes do not separate uniquely into two independent components.

Three ways away from the Gaussian, stationary, white model :

- Non-Gaussian, stationary, white
- Gaussian, non-stationary, white
- Gaussian, stationary, non-white

ICA of a mixture of independent Gaussian, stationary, colored processes (1)

- Same mixture model in Fourier Space : X(l) = A S(l)
- Parameters : the n*n mixing matrix A and the n power spectra $D_{Si}(l)$, for l = 1 to T .
- Whittle approximation to the likelihood :

$$\mathbf{p}(\tilde{\mathbf{X}}/\mathbf{A}, \tilde{\mathbf{D}}_{\mathbf{S}}) = \mid \det \mathbf{A} \mid^{-1} \prod_{\mathbf{l}=1}^{\mathbf{T}} \prod_{\mathbf{i}=1}^{\mathbf{N}} \frac{\mathbf{1}}{\sqrt{2\pi \tilde{D}_{S_i}(l)}} \exp(-\frac{\mid \tilde{\mathbf{Y}}_{\mathbf{i}}(\mathbf{l}) \mid^{2}}{\tilde{\mathbf{D}}_{\mathbf{S}_{\mathbf{i}}}(\mathbf{l})})$$

$$\tilde{\mathbf{Y}}(\mathbf{l}) = \mathbf{A}^{-1}\tilde{\mathbf{X}}(\mathbf{l})$$

ICA of a mixture of independent Gaussian, stationary, colored processes (2)

- Assuming the spectra are constant over symmetric subintervals, maximizing the likelihood is the same as minimizing:

$$\phi(\theta) = \sum_{q=1}^{Q} \alpha_q \mathcal{D}\Big(\widehat{R}_X(\nu_q), AR_S(\nu_q)A^\dagger\Big)$$

$$\mathcal{D}(R_1, R_2) = \frac{1}{2}\Big(\mathrm{Tr}(R_1 R_2^{-1}) - \operatorname{logdet}(R_1 R_2^{-1}) - m\Big)$$

- D turns out to be the Kulback Leibler divergence between two gaussian distributions.

- After minimizing with respect to the source spectra, we are left with a joint diagonality criterion to be minimized wrt $B = A^{-1}$. There are fast algorithms.

A Gaussian stationary ICA approach to the foreground removal problem?

- CMB is modeled as a Gaussian Stationary process over the sky.
- Working with covariance matrices achieves massive data reduction, and the spatial power spectra are the main parameters of interest.
- Easy extension to the case of noisy mixtures.
- Connection with Maximum-Likelihood guarantees some kind of optimality.
- Suggests using EM algorithm.

- [Snoussi2004, Cardoso2002, Delabrouille2003]

Spectral Matching ICA

- Back to the case of noisy mixtures : $X(\vartheta,\varphi) = AS(\vartheta,\varphi) + N(\vartheta,\varphi)$
- Fit the model spectral covariances : $R^f_{X,q} = AR^f_{S,q}A^\dagger + R^f_{N,q}$
- to estimated spectral covariances : $\hat{R}^f_{X,q} = \dfrac{1}{n_q} \displaystyle\sum_{p=0,\frac{p}{T}\in F_q}^{T-1} \widetilde{X}(\frac{p}{T})\widetilde{X}(\frac{p}{T})^\dagger$
- by minimizing : $\phi(\theta) = \displaystyle\sum_{q=1}^{Q} n_q \mathcal{D}_{KL}\left(\widehat{R}^f_{X,q}, AR^f_{S,q}A^\dagger + R^f_{N,q}\right)$
- with respect to : $\theta = (A, R^f_{S,q}, R^f_{N,q})$.

Algorithm

Start with sample covariance matrices $\langle \hat{S}_y \rangle_q$, and initial guesses for A, S_n and $\langle S_s \rangle_q$.

repeat

  {E-step ———— Compute conditional statistics}

  for $q = 1$ to Q **do**

    $C_q = (A^\dagger S_n^{-1} A + \langle S_s \rangle_q^{-1})^{-1}$

    $R_{yy}(q) = \langle \hat{S}_y \rangle_q$

    $R_{ys}(q) = \langle \hat{S}_y \rangle_q S_n^{-1} A \ C_q$

    $R_{ss}(q) = C_q \ A^\dagger S_n^{-1} \langle \hat{S}_y \rangle_q S_n^{-1} A \ C_q + C_q$

  end for

  $R_{ss} = \frac{1}{T} \sum_{q=1}^{Q} \ w_q R_{ss}(q)$

  $R_{ys} = \frac{1}{T} \sum_{q=1}^{Q} \ w_q R_{ys}(q)$

  $R_{yy} = \frac{1}{T} \sum_{q=1}^{Q} \ w_q R_{yy}(q)$

  {M-step ———— Update the parameters}

  $A = R_{ys} R_{ss}^{-1}$

  $S_n = \mathrm{diag}\left(R_{yy} - R_{ys} R_{ss}^{-1} R_{ys}^\dagger\right)$

  $\langle S_s \rangle_q = \mathrm{diag}\left(R_{ss}(q)\right)$ for $1 \le q \le Q$.

  Renormalize A and the $\langle S_s \rangle_q$

until convergence

- EM can become very slow, especially as noise parameters become small.

- Speed up convergence with a few BFGS steps.

- All /part of the parameters can be estimated.

Estimating component maps

- Wiener filtering in each frequency band :

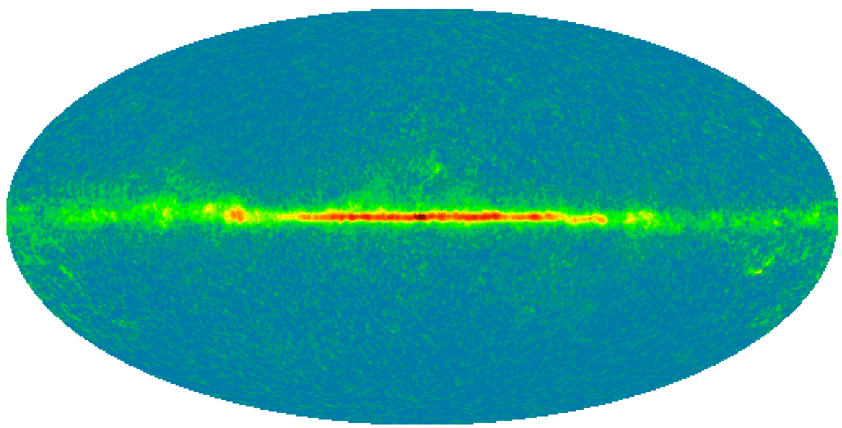
$$\widehat{S}(\nu) = (\widehat{A}^{\dagger} \widehat{R}_{N,q}^{f\,-1} \widehat{A} + \widehat{R}_{S,q}^{f\,-1})^{-1} \widehat{A}^{\dagger} \widehat{R}_{N,q}^{f\,-1} X(\nu)$$

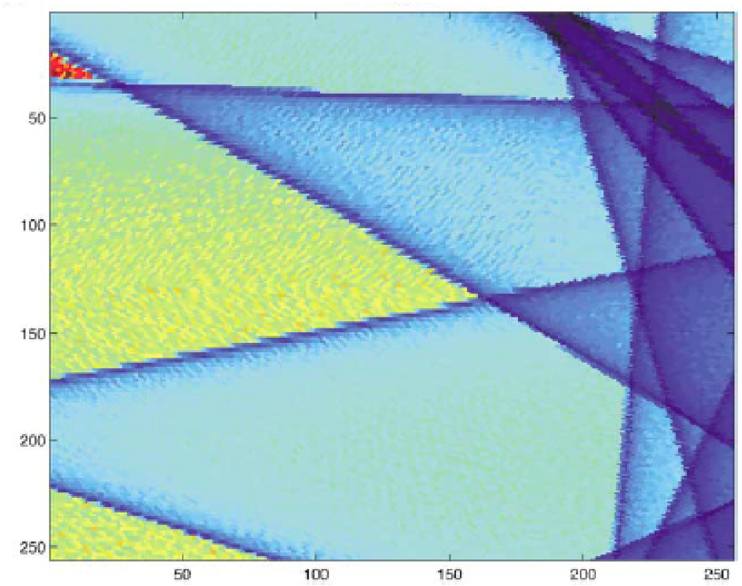
- In case of high SNR, use pseudo inverse :

$$\widehat{S}(\nu) = (\widehat{A}^{\dagger} \widehat{R}_{N,q}^{f\,-1} \widehat{A})^{-1} \widehat{A}^{\dagger} \widehat{R}_{N,q}^{f\,-1} X(\nu)$$

A *variation* on SMICA using wavelets

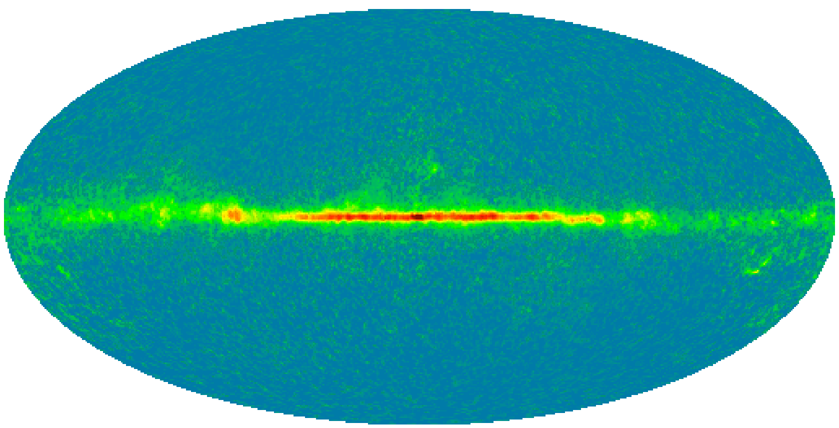
- Motivations :
  - Some components, and noise are a priori not stationary.
  - Galactic components appear correlated.
  - Emission laws may not be constant over the whole sky.
  - Incomplete sky coverage.

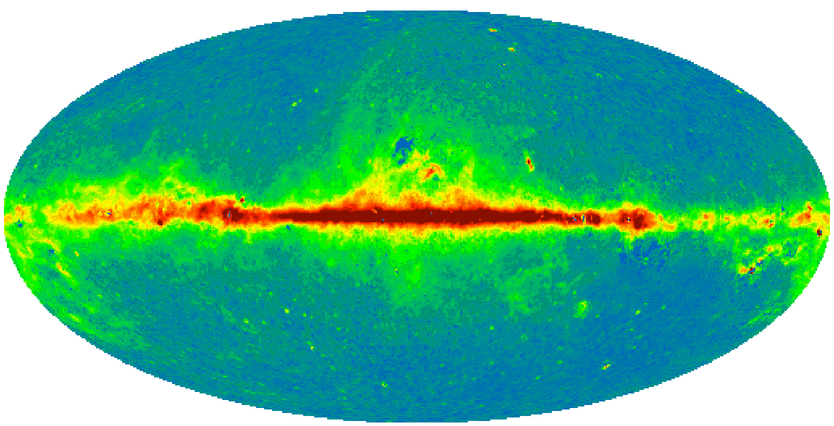




A *variation* on SMICA using wavelets

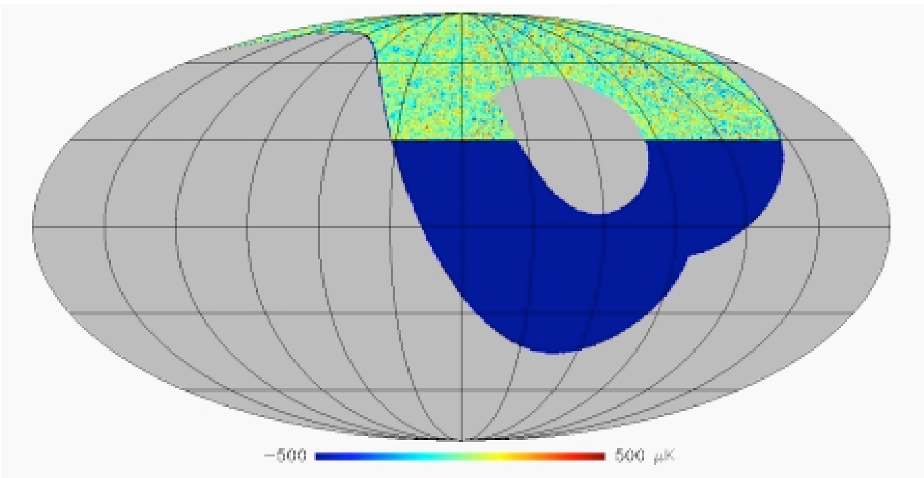
- Motivations :
  - Some components, and noise are a priori not stationary.
  - Galactic components are spatially correlated.
  - Emission laws may not be constant over the whole sky.
  - Incomplete sky coverage.

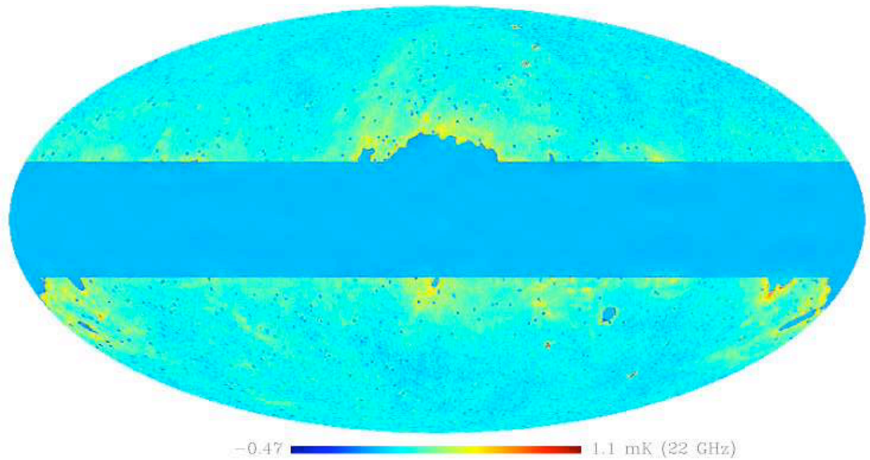




A *variation* on SMICA using wavelets

- Motivations :
  - Some components, and noise are a priori not stationary.
  - Galactic components appear correlated.
  - Emission laws may not be constant over the whole sky.
  - Incomplete sky coverage.





A *variation* on SMICA using wavelets

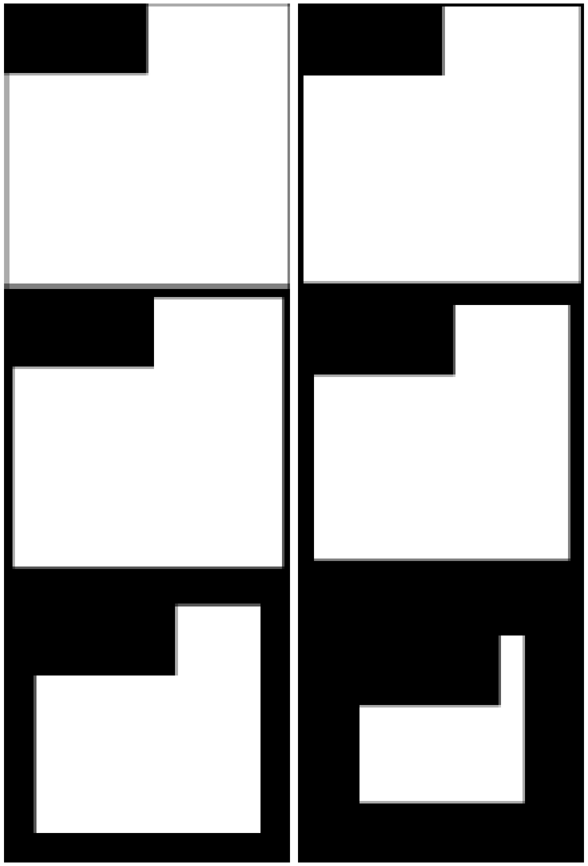
- Motivations :
  - Some components, and noise are a priori not stationary.
  - Galactic components appear correlated.
  - Emission laws may not be constant over the whole sky.
  - Incomplete sky coverage.

⇒ Use wavelets to preserve space-scale information.

The *à trous* wavelet transform

- shift invariant transform
- coefficient maps are the same size as the original map
- isotropic
- small compact supports of wavelet and scaling functions (B3 spline)
- Fast algorithms

$$c_0(k,l) = c_J(k,l) + \sum_{i=1}^{J} w_i(k,l)$$



wSMICA

Model structure is not changed :

$$R^w_{X,i} = AR^w_{S,i}A^\dagger + R^w_{N,i}$$

Objective function :

$$\phi(\theta) = \sum_{q=1}^{Q} \alpha_q \mathcal{D}\left(\widehat{R}^w_{X,q}, AR^w_{S,q}A^\dagger + R^w_{N,q}\right)$$

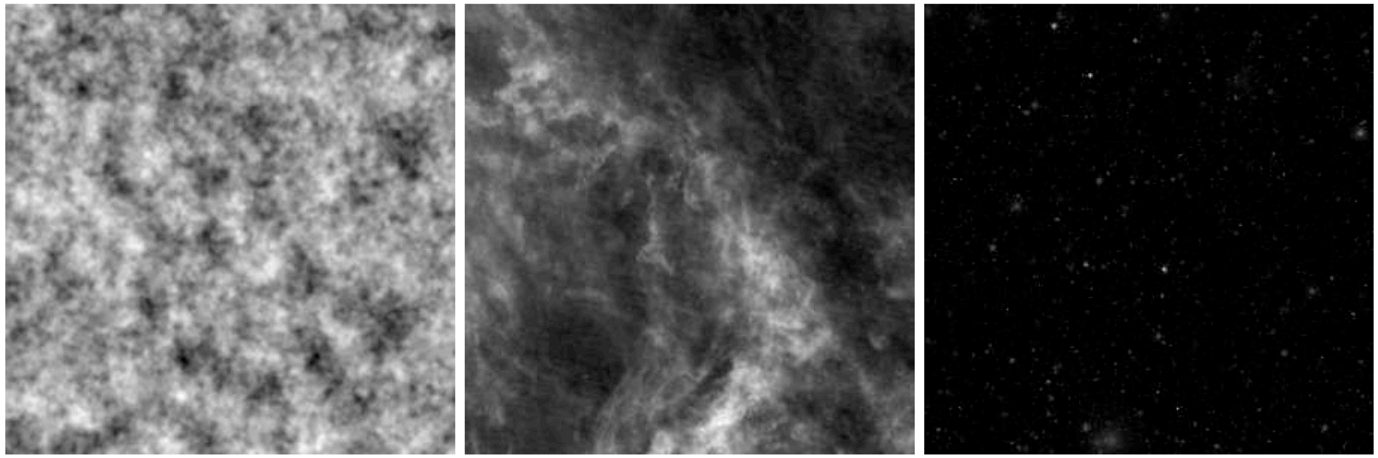
$$\{\alpha_1, \alpha_2, ..., \alpha_J, \alpha_{J+1}\} = \{\frac{3l_1}{4}, \frac{3l_2}{16}, ..., \frac{3l_J}{4^J}, \frac{l_{J+1}}{4^J}\}$$

Same minimization using EM algorithm and BFGS.

Experiments

Simulated Planck HFI observations at high galactic latitudes :

Component maps



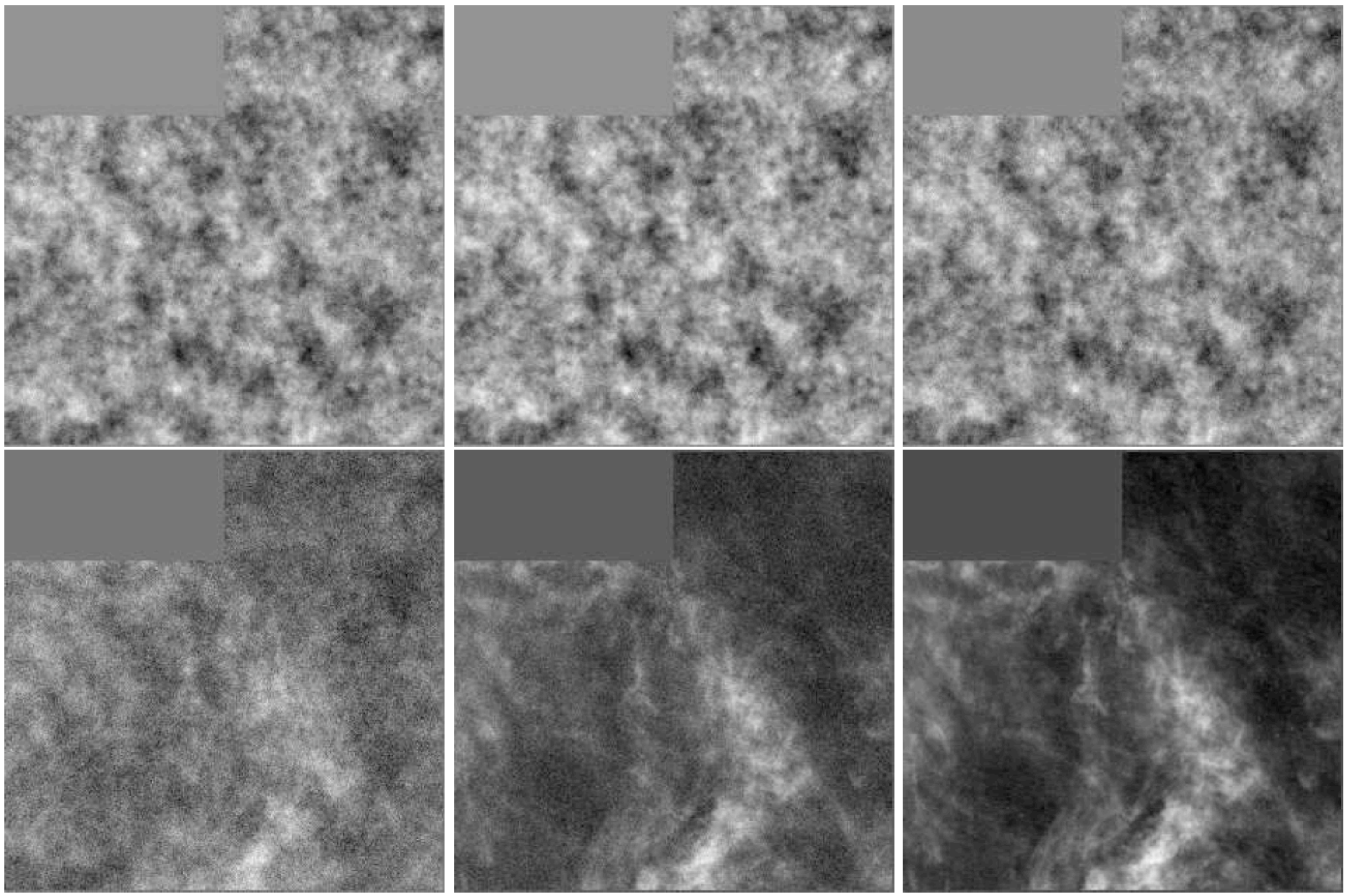
Mixing matrix

CMB	DUST	SZ	channel
7.452×10^{-1}	3.654×10^{-2}	-8.733×10^{-1}	100 GHz
5.799×10^{-1}	7.021×10^{-2}	-4.689×10^{-1}	143 GHz
3.206×10^{-1}	1.449×10^{-1}	-2.093×10^{-3}	217 GHz
7.435×10^{-2}	3.106×10^{-1}	1.294×10^{-1}	353 GHz
6.009×10^{-3}	5.398×10^{-1}	2.613×10^{-2}	545 GHz
6.115×10^{-5}	7.648×10^{-1}	5.268×10^{-4}	857 GHz

Planck HFI nominal noise levels

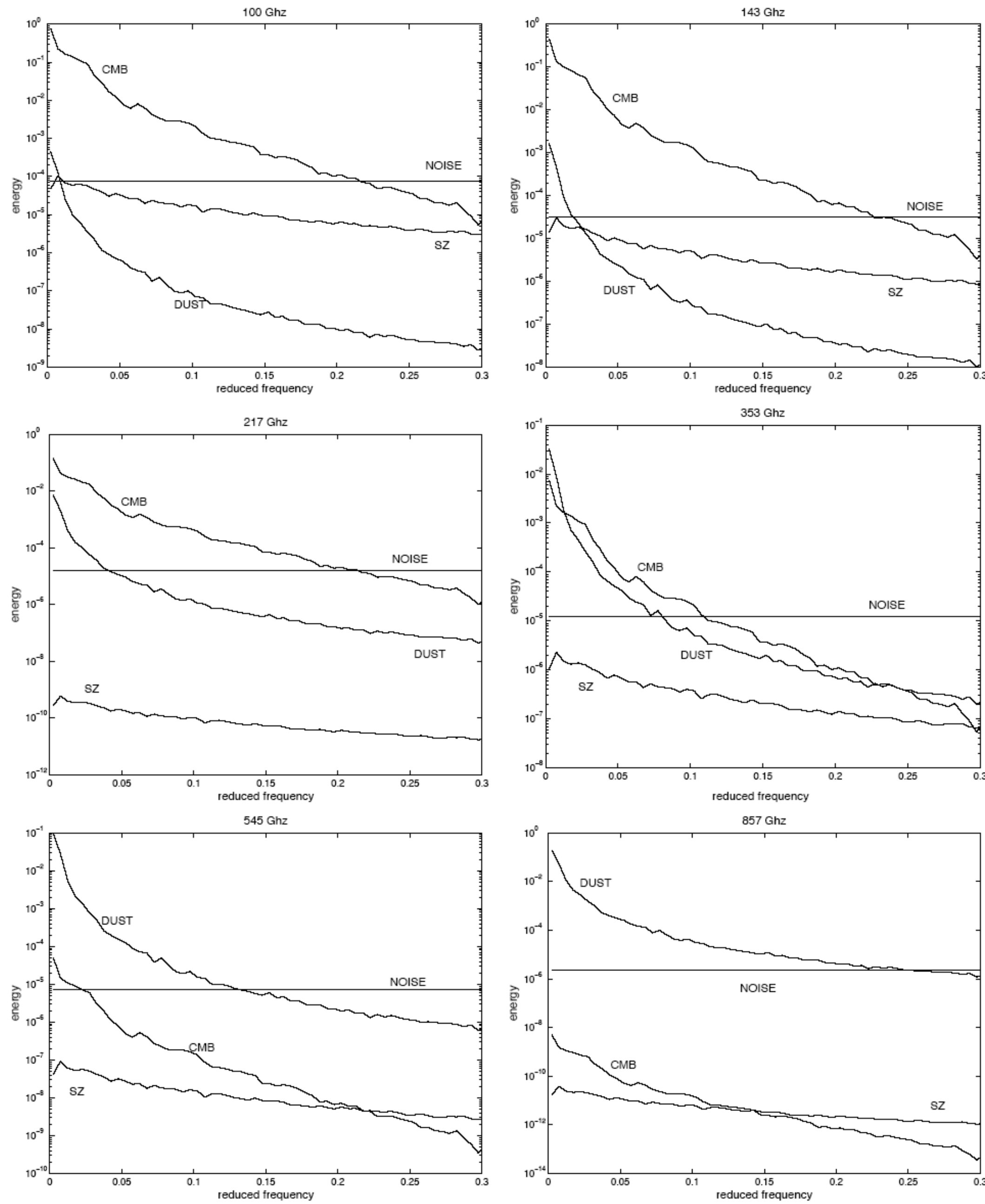
Experiments

Simulated Planck HFI observations at high galactic latitudes :

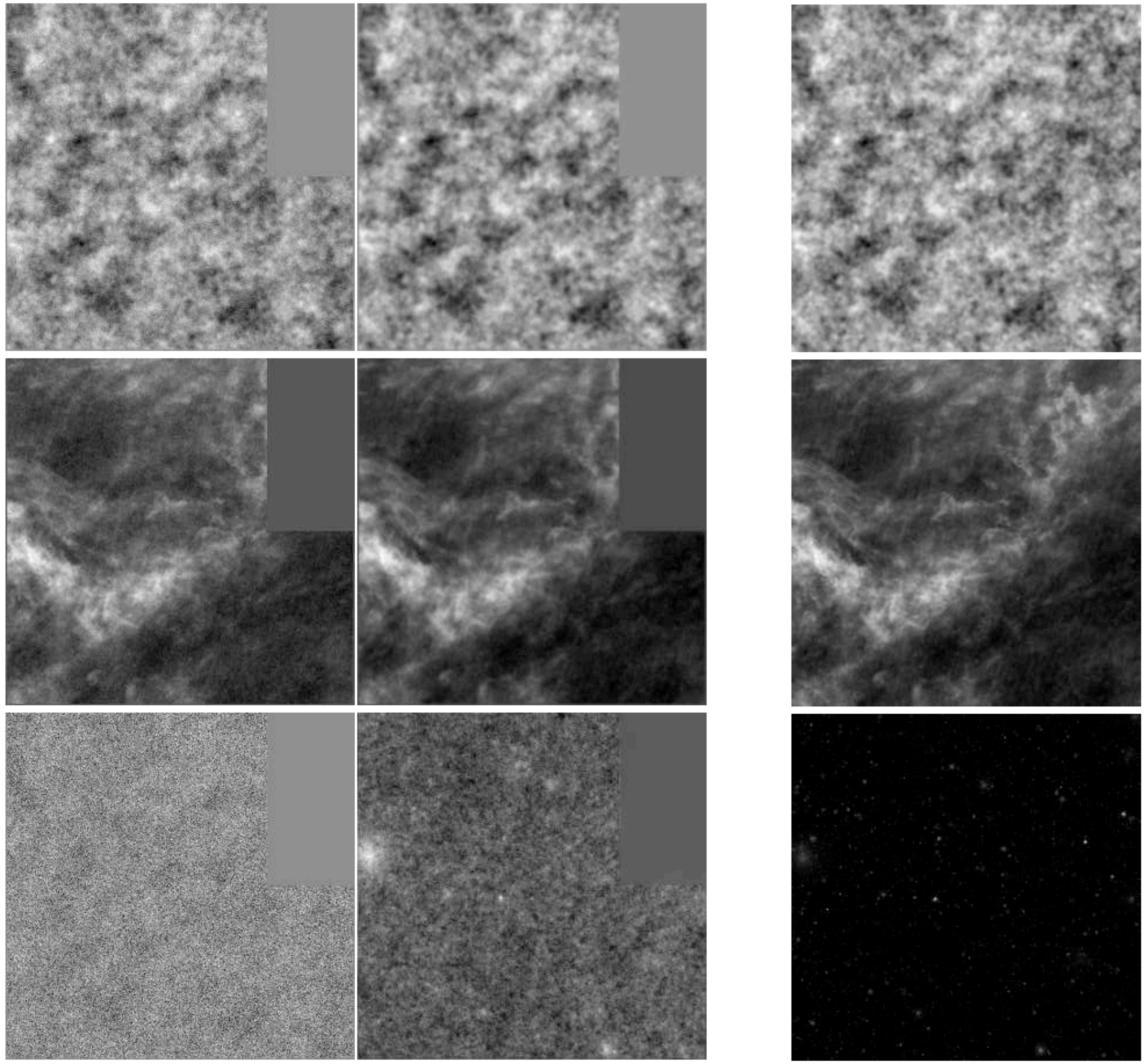


Experiments

Contribution of each component to each mixture as a function of spatial frequency :



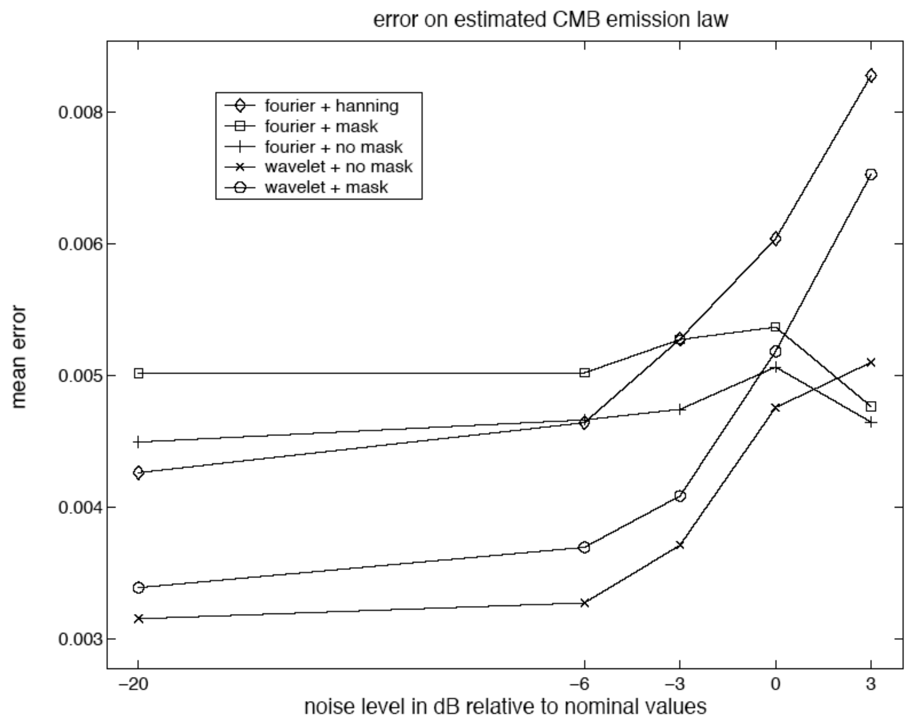
Results

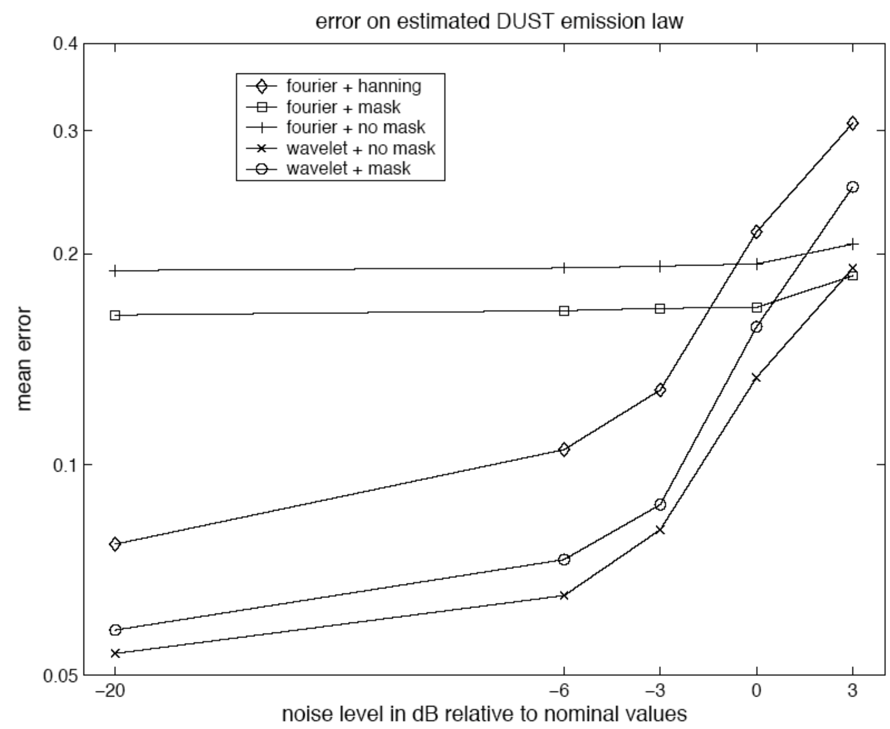


Jade wSMICA initial maps

Results

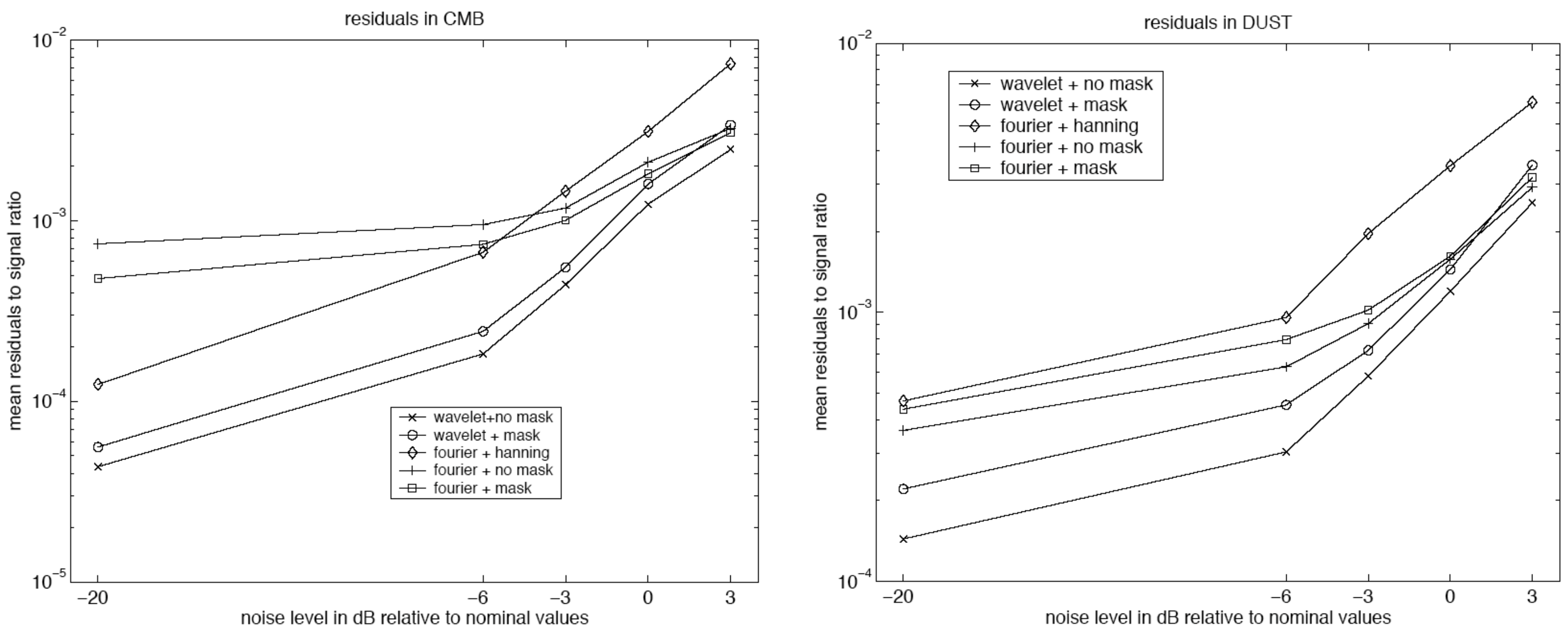
Errors on the estimated emission laws : $$QE_j = \left(\sum_{i=1}^{m} \left(A_{ij} - \widehat{A}_{ij} \right)^2 \right)^{\frac{1}{2}}$$





Results

Rejection rates : $ISR_j = \dfrac{\sum_{i \neq j} \mathcal{I}_{j,i}^2 \sigma_i^2}{\mathcal{I}_{j,j}^2 \sigma_j^2}$ where $\mathcal{I} = (\widehat{A}^\dagger \widehat{R}_N^{-1} \widehat{A})^{-1} \widehat{A}^\dagger \widehat{R}_N^{-1} A$



Conclusions

- Foreground removal using SMICA is remarkably good.
- Correctly handling non-stationarities, missing patches is important and wavelets help do that easily.
- Extending the results to full sky data using a wavelet transform on the sphere.
- The model can be extended to include spatial correlations between the components.
- Some components are clearly non Gaussian. Not using that prior is inefficient.
- Not clear yet how best to combine non-gaussianity, non-stationarity and non-whiteness for better source separation.