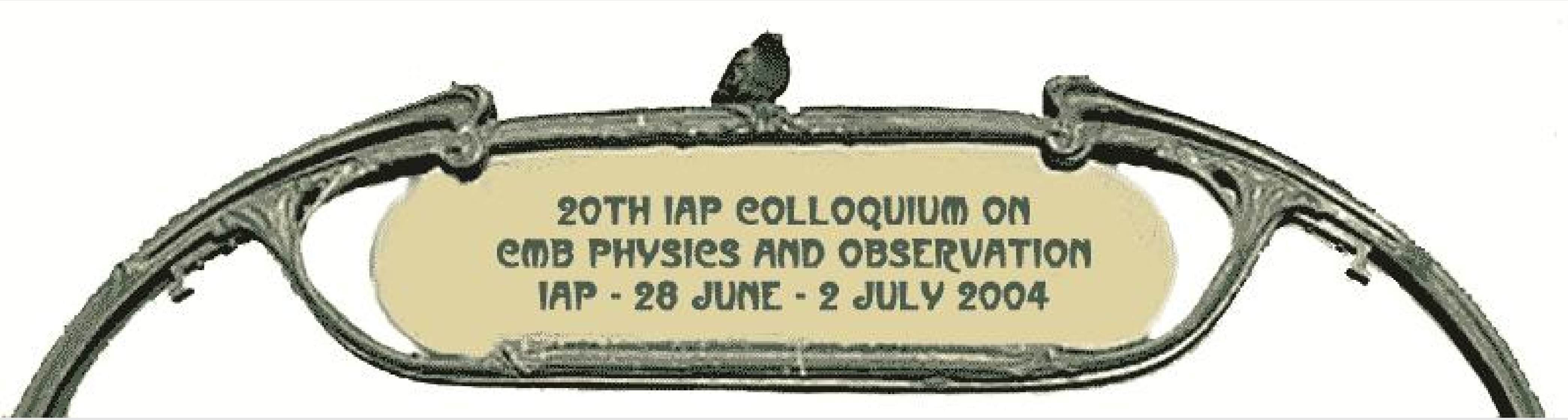


Component separation

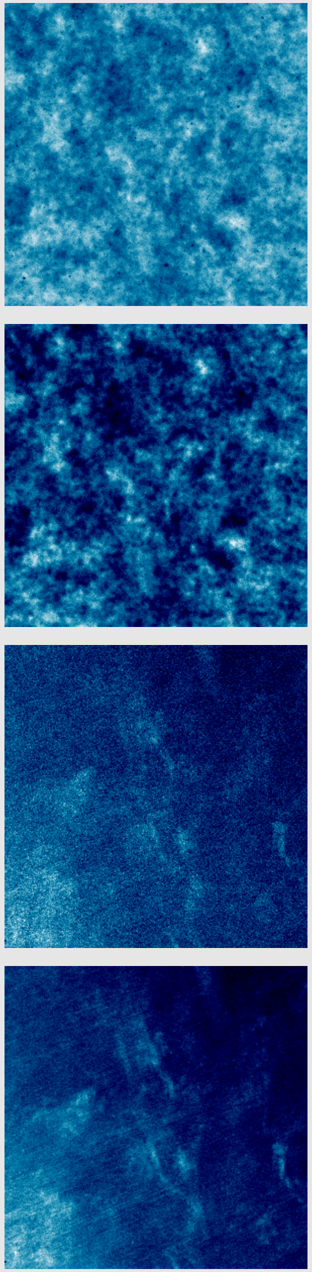
J.-F. Cardoso,
CNRS-STIC/ENST-LTCI
PCC Collège de France - IN2P3

Institut d'Astrophysique de Paris, juin 2004

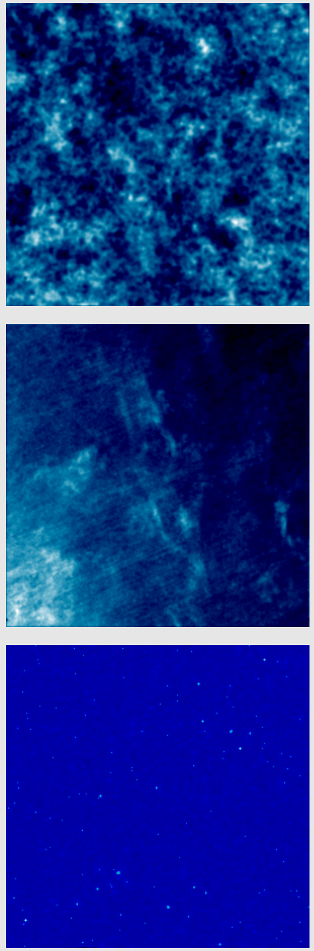
The idea of components

Noisy linear spatial mixture : $X(\vec{r}) = AS(\vec{r}) + N(\vec{r})$

$$\begin{bmatrix} X_1(\vec{r}) \\ X_2(\vec{r}) \\ X_3(\vec{r}) \\ X_4(\vec{r}) \end{bmatrix} =$$



$$= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix} \times$$



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

Related work

Blind and non blind

– Hobson 1998, Stolyarov et al 2001, Maximum entropy
– Bouchet et al 1999, Wiener filter
– Tegmark et al 2000
– Baccigalupi 2000
– Maino et al 2001
– Snoussi et al 2001
– Maino et al, 2003. astro-ph/0303657 (FastICA on COBE-DMR)

Outline

- Component separation :
  exploiting diversity to recover components (sources) from mixtures.

- Blind separation :
  exploiting statistical independence(s).

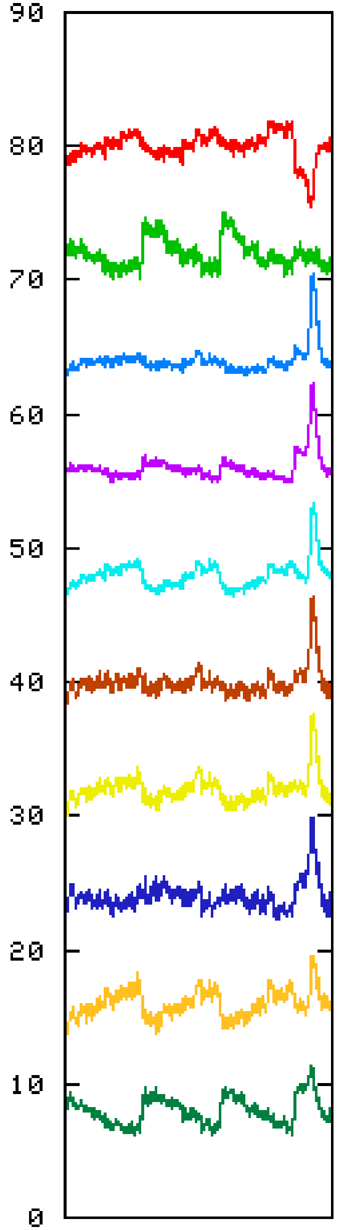
- Component separation versus spectrum separation : SMICA.

- What works when, and why.
  And why that which should not work may still work.

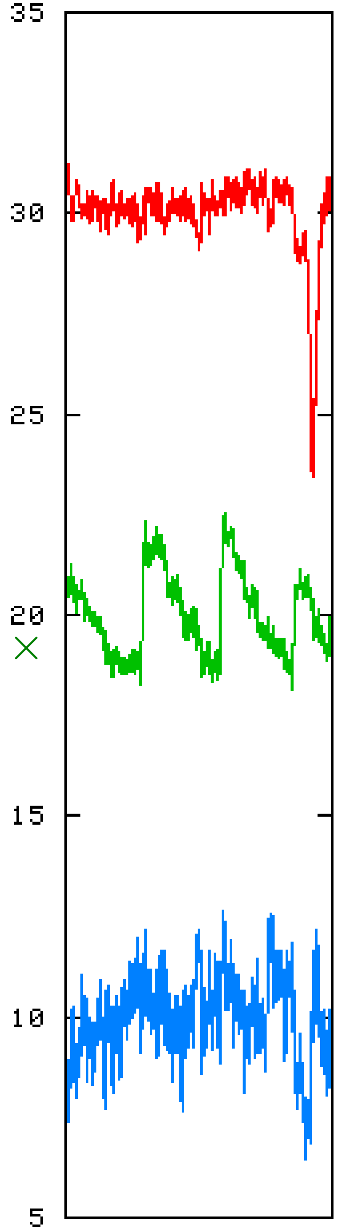
- Results on Archeops and W-MAP data (Guillaume Patanchon)

Blink'n'Scan (something non cosmic)

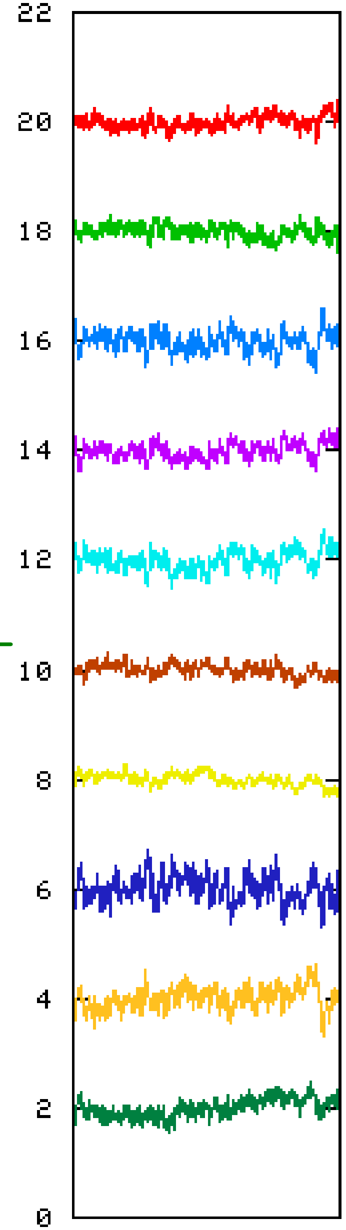
$$X = AS + N$$



$$= \begin{bmatrix} 58 & -45 & 66 \\ 4 & 94 & 29 \\ -97 & -13 & 6 \\ -90 & 37 & -12 \\ -78 & -56 & 18 \\ -98 & -11 & 11 \\ -88 & -40 & 21 \\ -90 & 36 & -4 \\ -57 & -71 & 33 \\ -45 & 84 & -21 \end{bmatrix} \times$$



$$+$$



Component separation : get S from X

Model $X = AS + N$: Linear mixtures AS of independent ($S \sim P_S = \prod_i P_{S_i}$) components, observed in Gaussian noise ($N \sim \mathcal{N}(0, R_N)$).

The most likely S once X is observed —the maximum *a posteriori* (MAP) estimate— is

$$\widehat{S}(X) = \arg\max_S \; P(S|X) = \arg\min_S \; (X - AS)^\dagger R_N^{-1}(X - AS) + \sum_i \phi_i(S_i)$$

where $\phi_i(\cdot) = -2 \log P_{S_i}(\cdot)$ depends on the (hypothetical) component pdf.

Component separation : get S from X

Model $X = AS + N$: Linear mixtures AS of independent ($S \sim P_S = \prod_i P_{S_i}$) components, observed in Gaussian noise ($N \sim \mathcal{N}(0, R_N)$).

The most likely S once X is observed —the maximum *a posteriori* (MAP) estimate— is

$$\widehat{S}(X) = \arg\max_S \; P(S|X) = \arg\min_S \; (X - AS)^\dagger R_N^{-1} (X - AS) + \sum_i \phi_i(S_i)$$

where $\phi_i(\cdot) = -2 \log P_{S_i}(\cdot)$ depends on the (hypothetical) component pdf.

- Gaussian components : $\phi_i(S_i) = S_i^2/\sigma_i^2$.

– The MAP is linear : $\widehat{S}(X) = (A^\dagger R_N^{-1} A + R_S^{-1})^{-1} A^\dagger R_N^{-1} X$

– It is 'biased', unlike $\widehat{S}_\star(X) = (A^\dagger R_N^{-1} A)^{-1} A^\dagger R_N^{-1} X = S + \text{noise}$.

– It is also the minimizer of $E(S - f(X))^2$: Wiener filter.

Component separation : get S from X

Model $X = AS + N$: Linear mixtures AS of independent ($S \sim P_S = \prod_i P_{S_i}$) components, observed in Gaussian noise ($N \sim \mathcal{N}(0, R_N)$).

The most likely S once X is observed —the maximum *a posteriori* (MAP) estimate— is

$$\widehat{S}(X) = \arg\max_S \; P(S|X) = \arg\min_S \; (X - AS)^\dagger R_N^{-1}(X - AS) + \sum_i \phi_i(S_i)$$

where $\phi_i(\cdot) = -2 \log P_{S_i}(\cdot)$ depends on the (hypothetical) component pdf.

- Gaussian components : $\phi_i(S_i) = S_i^2/\sigma_i^2$.

– The MAP is linear : $\widehat{S}(X) = (A^\dagger R_N^{-1} A + R_S^{-1})^{-1} A^\dagger R_N^{-1} X$

– It is 'biased', unlike $\widehat{S}_\star(X) = (A^\dagger R_N^{-1} A)^{-1} A^\dagger R_N^{-1} X = S + \text{noise}$.

– It is also the minimizer of $E(S - f(X))^2$: Wiener filter.

- Non Gaussian comp's : $\phi_i(S_i) = S_i^2/(S_i^2 + \sigma_i^2)$ for instance, for heavy tails.

– Other functions ϕ_i for other priors. MAP is MEM.

– The MAP estimate $\widehat{S}(X)$ is non linear for non Gaussian priors.

Two issues

In order to separate components (invert the noisy mixture $X = AS + N$), one may use, for instance, the Gaussian MAP (a.k.a. Gaussian Wiener filter)

$$\widehat{S}(X) = (A^\dagger R_N^{-1} A + R_S^{-1})^{-1} A^\dagger R_N^{-1} X$$

or even smarter filters : non Gaussian MAP, non Gaussian Wiener filters.

Two issues

In order to separate components (invert the noisy mixture $X = AS + N$), one may use, for instance, the Gaussian MAP (a.k.a. Gaussian Wiener filter)

$$\widehat{S}(X) = (A^\dagger R_N^{-1} A + R_S^{-1})^{-1} A^\dagger R_N^{-1} X$$

or even smarter filters : non Gaussian MAP, non Gaussian Wiener filters.

1 • But even the simple Wiener filter requires knowing the mixture coefficients (the A matrix) and the power R_S, R_N of components and noise ...

Two issues

In order to separate components (invert the noisy mixture $X = AS + N$), one may use, for instance, the Gaussian MAP (a.k.a. Gaussian Wiener filter)

$$\widehat{S}(X) = (A^\dagger R_N^{-1} A + R_S^{-1})^{-1} A^\dagger R_N^{-1} X$$

or even smarter filters : non Gaussian MAP, non Gaussian Wiener filters.

1 • But even the simple Wiener filter requires knowing the mixture coefficients (the A matrix) and the power R_S, R_N of components and noise ...

2 • For noisy mixtures, there can be no clean maps of separated components

(*Would you like some noise in your contaminants ?*)

so spectral estimation from separated maps still face the issue of removing 'stuff' (noise and/or signal) from the spectral estimates. $\widehat{S}(AS) \neq S$.

Two routes to component spectra

Component separation :

$$X_i(\theta, \phi) \xrightarrow{1} S_j(l, m) \xrightarrow{2} \widehat{C}_j(l) = \langle S_j(l, m)^2 \rangle_{l,m \in \mathrm{bin}(l)}$$

1 : Component separation and spherical harmonic transform.

2 : Non parametric (auto)-spectral estimation.

Two routes to component spectra

Component separation :

$$X_i(\theta,\phi) \xrightarrow{1} S_j(l,m) \xrightarrow{2} \widehat{C}_j(l) = \langle S_j(l,m)^2 \rangle_{l,m\in \mathrm{bin}(l)}$$

1 : Component separation and spherical harmonic transform.

2 : Non parametric (auto)-spectral estimation.

Spectrum estimation :

$$X_i(\theta,\phi) \xrightarrow{1} \widehat{R}_{ij}(l) = \langle X_i(l,m) X_j(l,m) \rangle_{l,m\in \mathrm{bin}(l)} \xrightarrow{2} \widehat{C}_j(l)$$

1 : Non parametric, cross- and auto-spectral estimation.

2 : Spectrum separation.

3 : Optionally, component separation (a by product, now easy).

Spectral matrices

The data-based auto- and cross-spectra $\widehat{R}_{ij}(l) = \langle X_i(l,m) X_j(l,m) \rangle_{l,m\in \mathrm{bin}(l)}$ are collected in a sample spectral matrix $\widehat{R}(l)$.

It is the natural estimate of the spectral matrix $R(l)$ at mode l.

According to the model $X = AS + N$, it is structured as

$$\begin{bmatrix} R_{11}(l) & R_{12}(l) & R_{13}(l) \\ R_{21}(l) & R_{22}(l) & R_{23}(l) \\ R_{31}(l) & R_{32}(l) & R_{33}(l) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \cdot \begin{bmatrix} C_1(l) & 0 \\ 0 & C_2(l) \end{bmatrix} \cdot \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}^\dagger + \begin{bmatrix} \sigma_1^2(l) & 0 & 0 \\ 0 & \sigma_2^2(l) & 0 \\ 0 & 0 & \sigma_3^2(l) \end{bmatrix}$$

for $m = 3$ detectors, $n = 2$ components and spatially uncorrelated noise. Here, $C_i(l)$ is the harmonic spectrum of the ith component.

Spectral matrices

The data-based auto- and cross-spectra $\widehat{R}_{ij}(l) = \langle X_i(l,m) X_j(l,m) \rangle_{l,m \in \mathrm{bin}(l)}$ are collected in a sample spectral matrix $\widehat{R}(l)$.

It is the natural estimate of the spectral matrix $R(l)$ at mode l.

According to the model $X = AS + N$, it is structured as

$$\begin{bmatrix} R_{11}(l) & R_{12}(l) & R_{13}(l) \\ R_{21}(l) & R_{22}(l) & R_{23}(l) \\ R_{31}(l) & R_{32}(l) & R_{33}(l) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \cdot \begin{bmatrix} C_1(l) & 0 \\ 0 & C_2(l) \end{bmatrix} \cdot \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}^{\dagger} + \begin{bmatrix} \sigma_1^2(l) & 0 & 0 \\ 0 & \sigma_2^2(l) & 0 \\ 0 & 0 & \sigma_3^2(l) \end{bmatrix}$$

for $m = 3$ detectors, $n = 2$ components and spatially uncorrelated noise. Here, $C_i(l)$ is the harmonic spectrum of the ith component.

The left hand side is estimated from the data by $\widehat{R}(l)$.
The right hand side can be **uniquely** fitted to it,
by adjusting **any chosen subset** of parameters $\left\{a_{ij}, C_i(l), \sigma_j^2(l)\right\}$.

SMICA : Spectral matching independent component analysis

A measure of spectral mismatch

Plan : Estimate all unknown parameters by spectral matching.

Specifically : The unknown parameters $\theta = (C_1(l), \ldots)$ are found by minimizing

$$\phi(\theta) = \sum_{l=1}^{L} n_l \; K\left(\widehat{R}_l | R_l(\theta)\right)$$

where $K(\cdot|\cdot)$ is the 'Kullback divergence' between positive matrices

$$K(R_1|R_2) = \mathrm{trace}(R_1 R_2^{-1}) - \log\det(R_1 R_2^{-1}) - m.$$

A measure of spectral mismatch

Plan : Estimate all unknown parameters by spectral matching.

Specifically : The unknown parameters $\theta = (C_1(l), \ldots)$ are found by minimizing

$$\phi(\theta) = \sum_{l=1}^{L} n_l \; K\left(\widehat{R}_l | R_l(\theta)\right)$$

where $K(\cdot|\cdot)$ is the 'Kullback divergence' between positive matrices

$$K(R_1|R_2) = \mathrm{trace}(R_1 R_2^{-1}) - \log\det(R_1 R_2^{-1}) - m.$$

Why ? Because $-\phi(\theta)$ is the Whittle approximation to the log-likelihood.

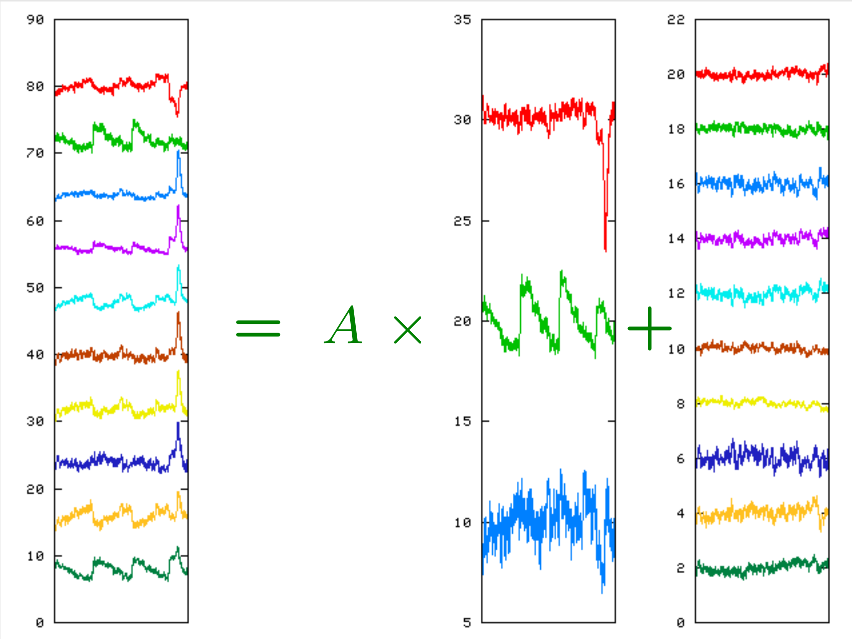
Meaning : Fisher optimality for Gaussian stationary components.

What else ?

What about these non Gaussian ICA methods ?

Independent component analysis $X = AS + N$

– Separation : Given A (plus component properties), build S from X.
– Blind analysis : find the mixing A in the first place.

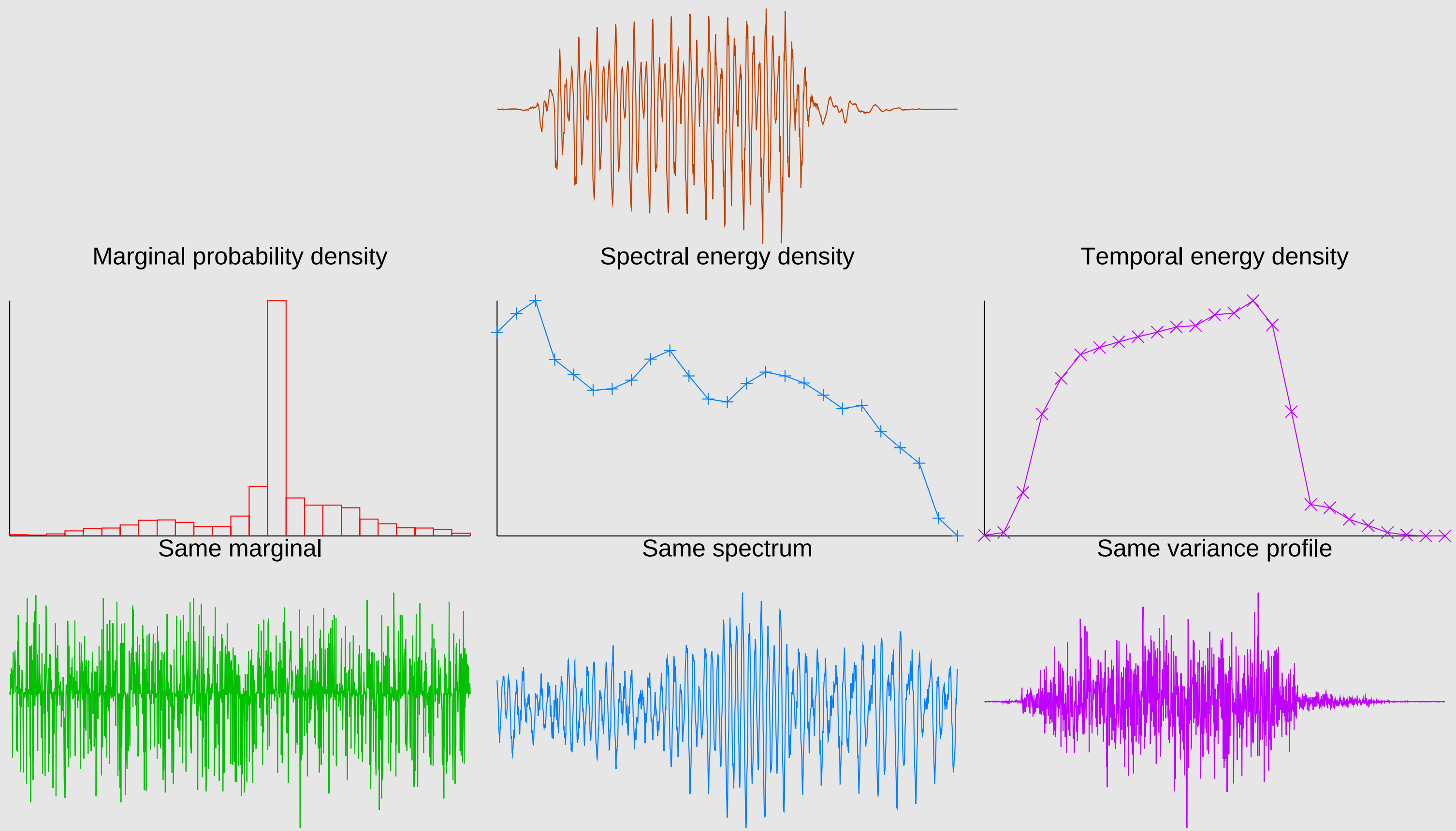


Ideas for blind analysis :

– Make rows of S mutually independent
– Make each row of S simple
– Make each row of S sparse
– Make each row of S neguentropic
– Do not constrain A at all. . .
– . . . or do it, if so inclined.
– . . .

The general idea : Select a *simple* statistical model for each component and look at the maximum likelihood solution.

Three points of view on a random (? ? ?) process



All models are wrong, but some are useful —George Box

Independences

The invertible linear transform making the entries of $Y(t) = BX(t)$ are 'as independent as possible' is such that, for any pair $i \neq j$

$$\frac{1}{T}\sum_{t=1}^{T} \psi_i(Y_i(t))Y_j(t) = 0$$

for i.i.d. sequences and $\psi_i = -r_i'/r_i$ with r_i the pdf of $Y_i(t)$ (for any t) *or*

Independences

The invertible linear transform making the entries of $Y(t) = BX(t)$ are 'as independent as possible' is such that, for any pair $i \neq j$

$$\frac{1}{T}\sum_{t=1}^{T} \psi_i(Y_i(t))Y_j(t) = 0$$

for i.i.d. sequences and $\psi_i = -r_i'/r_i$ with r_i the pdf of $Y_i(t)$ (for any t) *or*

$$\frac{1}{T}\sum_{t=1}^{T} \frac{Y_i(t)}{a_i^2(t)} Y_j(t) = 0$$

for modulated Gaussian i.i.d. sequences and $a_i(t)$ is the amplitude of $Y_i(t)$ *or*

Independences

The invertible linear transform making the entries of $Y(t) = BX(t)$ are 'as independent as possible' is such that, for any pair $i \neq j$

$$\frac{1}{T}\sum_{t=1}^{T} \psi_i(Y_i(t))Y_j(t) = 0$$

for i.i.d. sequences and $\psi_i = -r_i'/r_i$ with r_i the pdf of $Y_i(t)$ (for any t) *or*

$$\frac{1}{T}\sum_{t=1}^{T} \frac{Y_i(t)}{a_i^2(t)} Y_j(t) = 0$$

for modulated Gaussian i.i.d. sequences and $a_i(t)$ is the amplitude of $Y_i(t)$ *or*

$$\frac{1}{T}\sum_{t=1}^{T} \frac{\tilde{Y}_i(t)}{P_i(t/T)} \tilde{Y}_j(t) = 0$$

for Gaussian stationary sequences where $\tilde{Y}_i$ is the DFT of Y_i and $P_i(\nu)$ is its power spectrum.

Some experimental results

Mostly conducted by Guillaume Patanchon.

- First release of Archeops data.

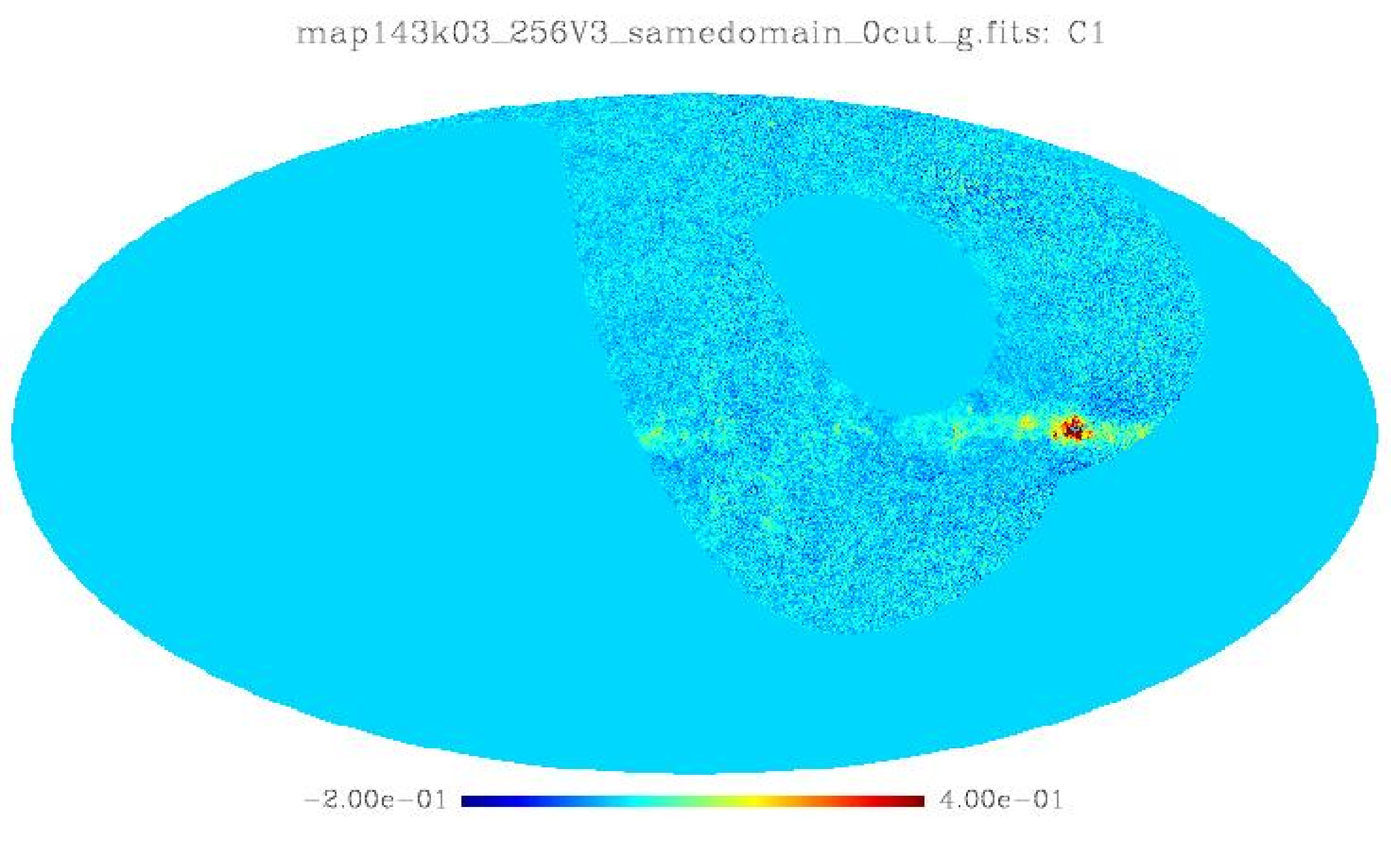
Several components.

- Second release of Archeops data.

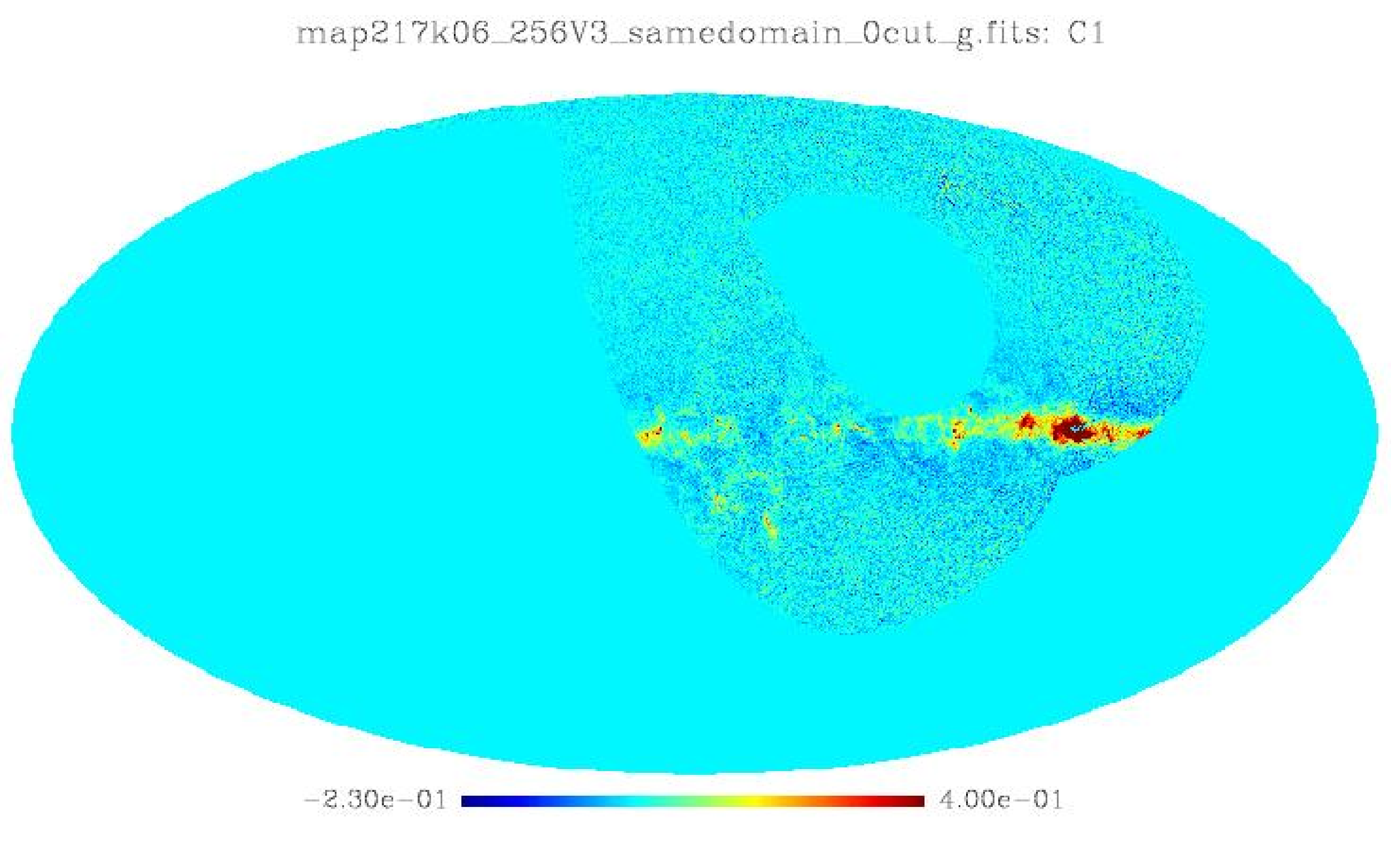
Checking for inter-calibration, selecting bolometers, goodness of fit, general feelgood feeling.

- Preliminary W-MAP processing.

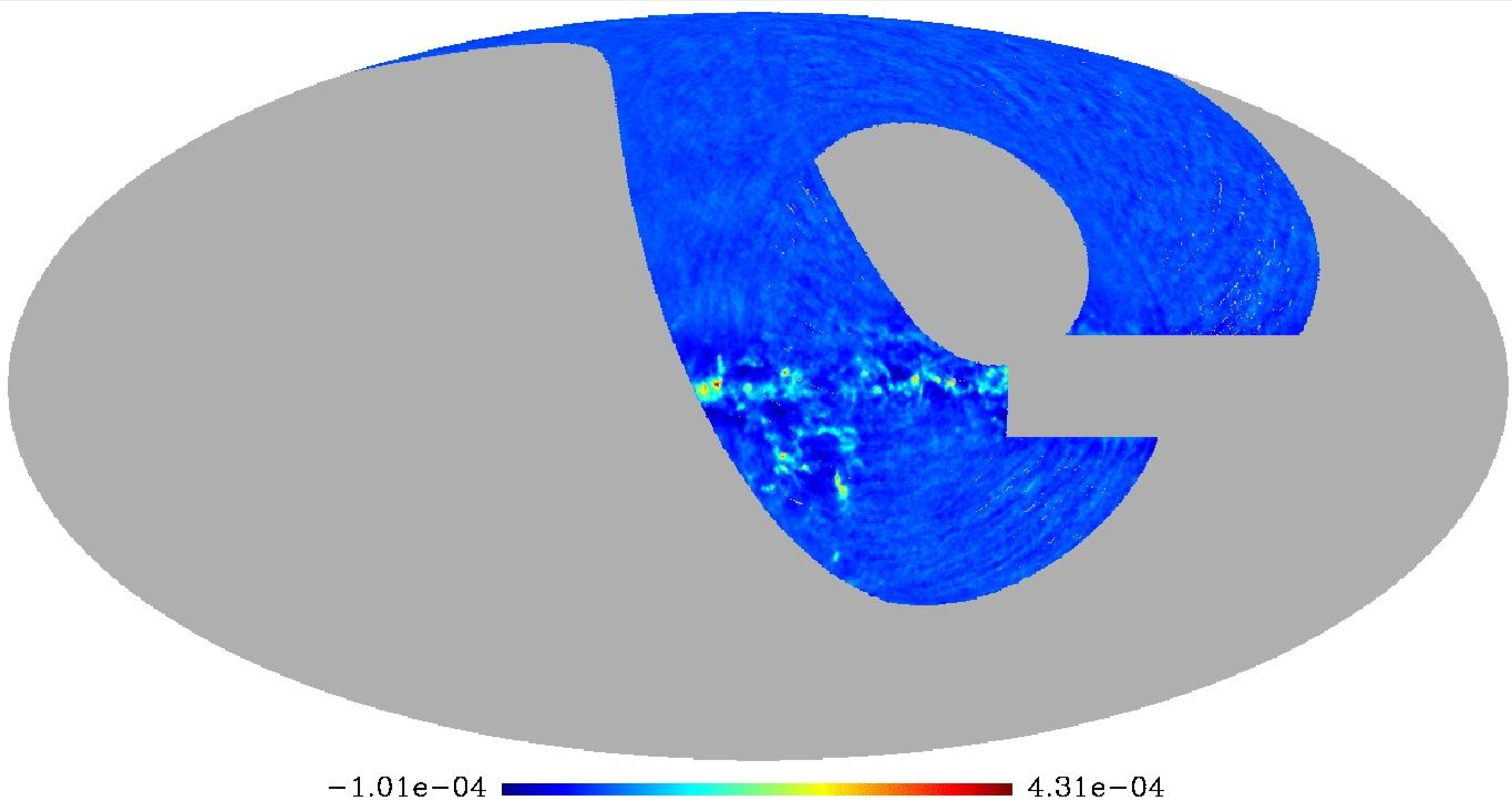
Archeops map in the 143 kHz channel



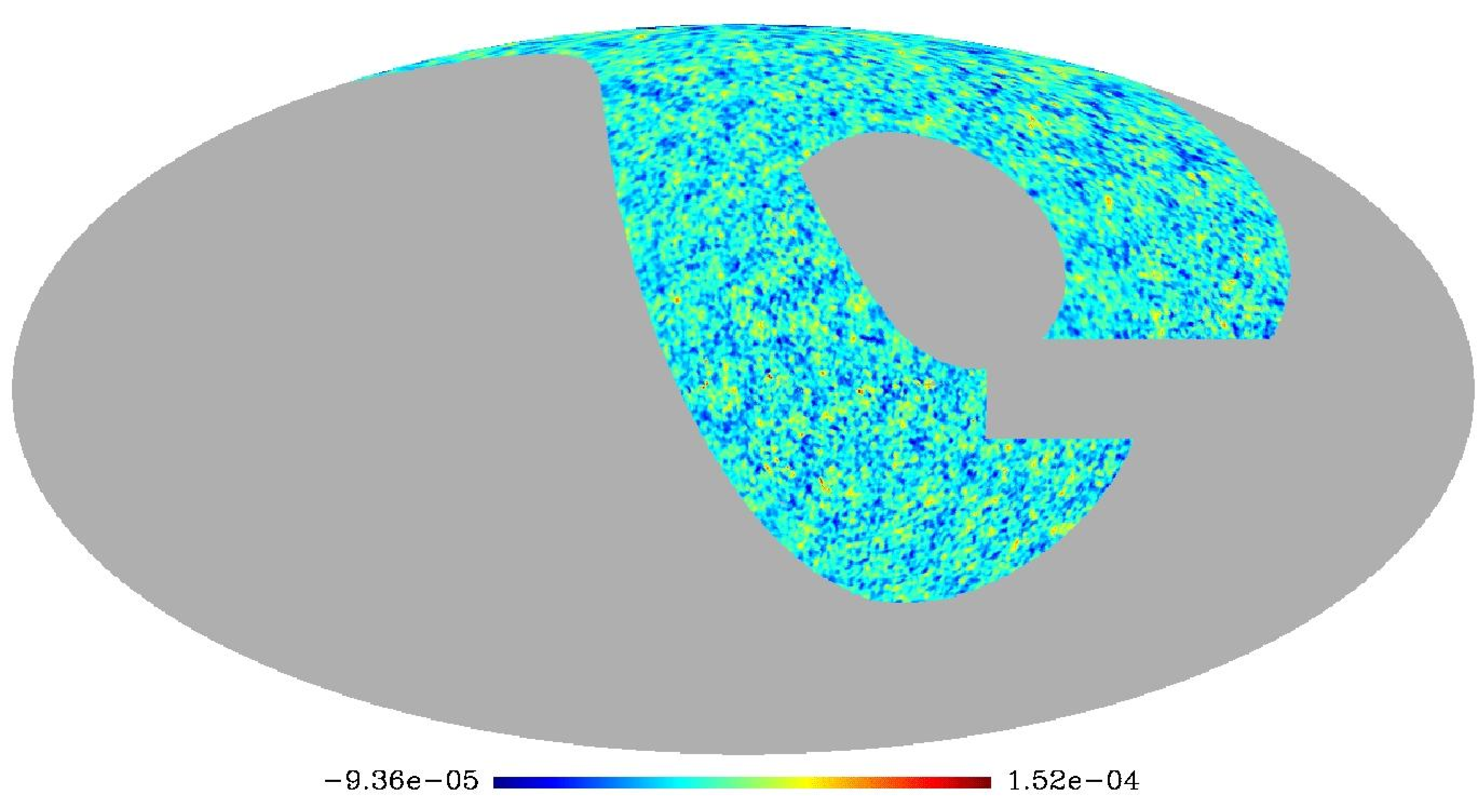
Archeops map in the 217 kHz channel



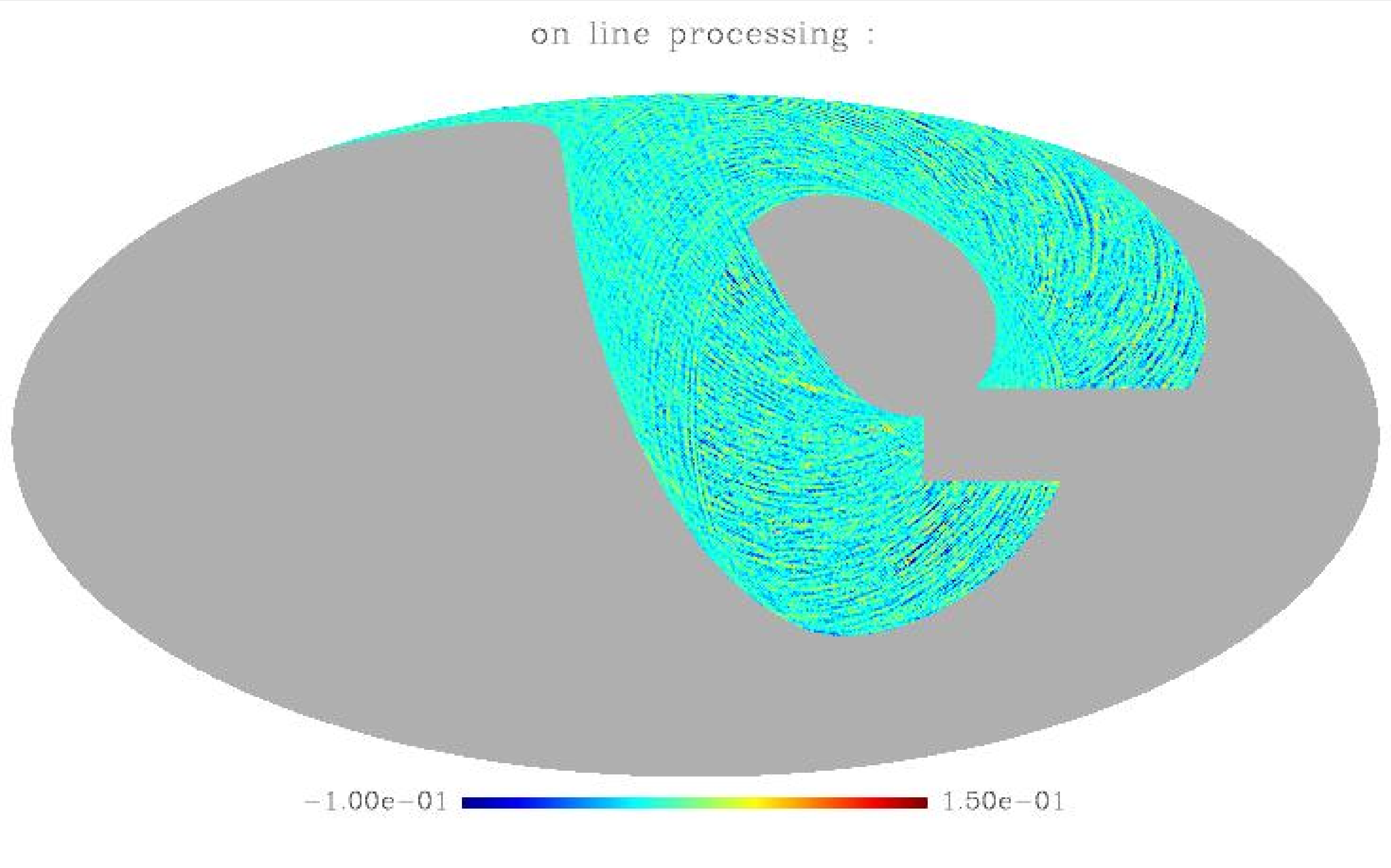
Separated dust component



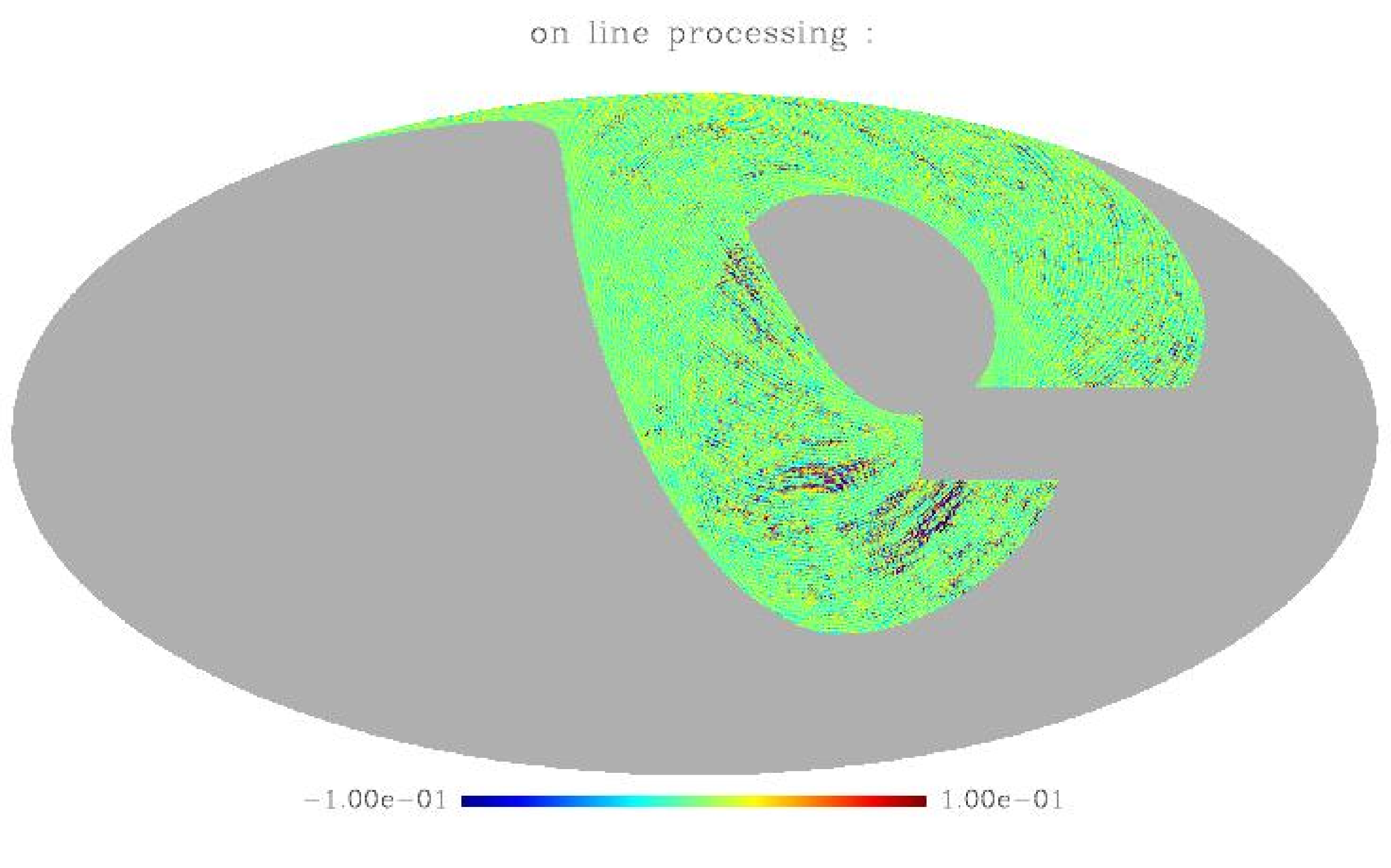
Separated CMB component



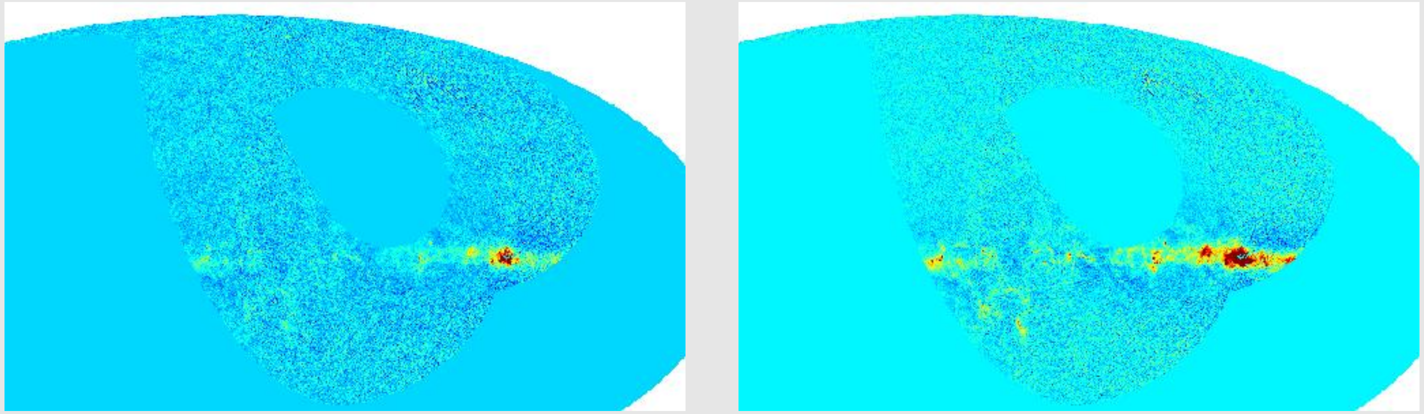
First ozone component



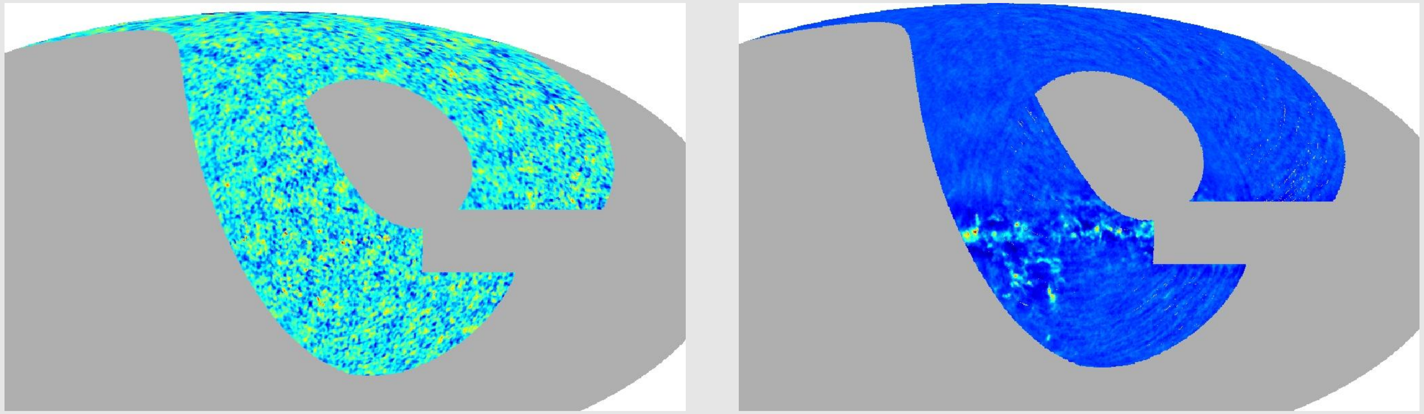
Second ozone component



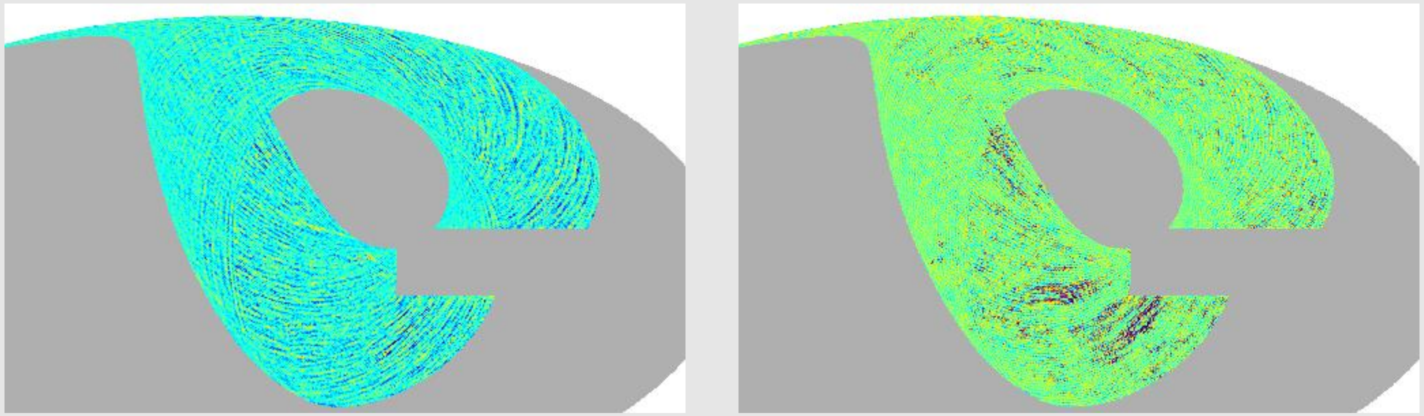
Archeops



Sky maps at 143 and 247 GHz. (2 out of 15)



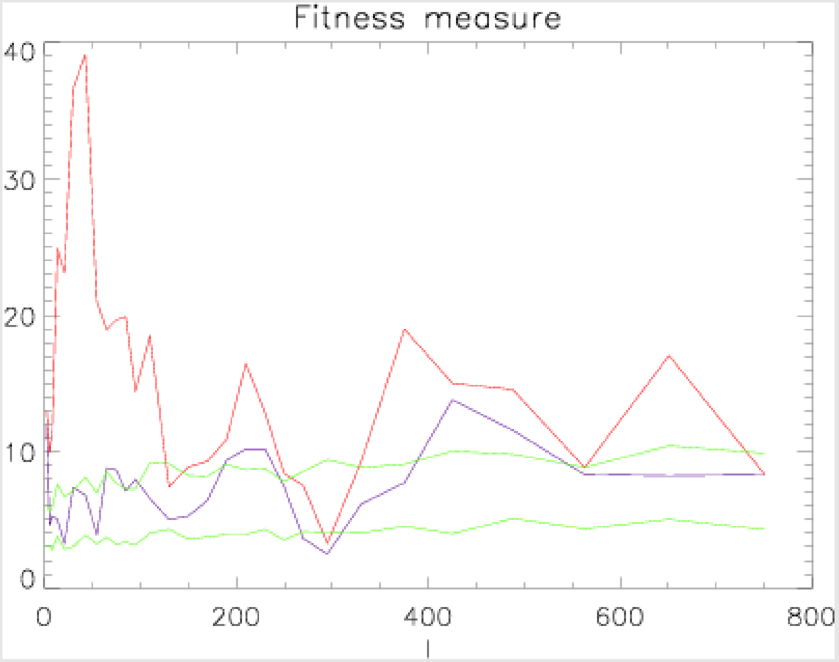
CMB (!) and dust.

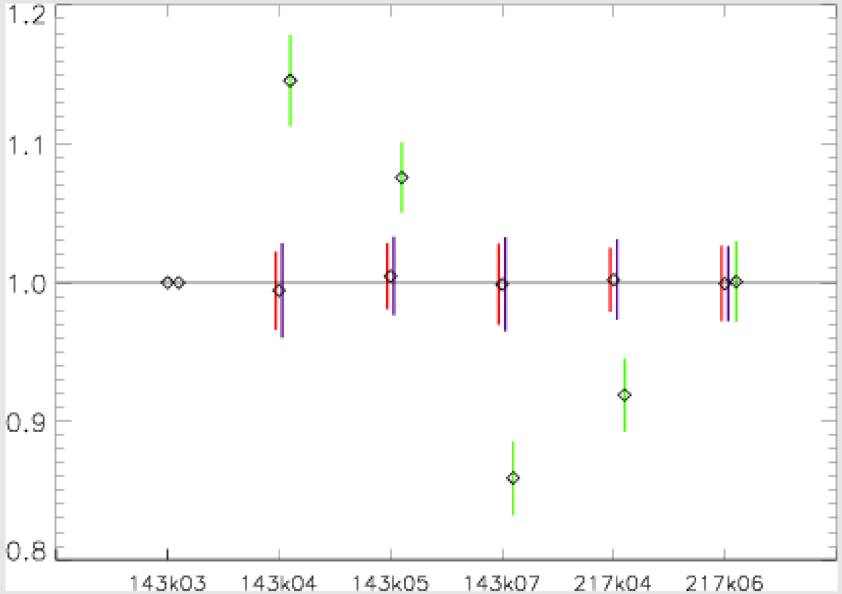


Two ozone components.

Goodness of Archeops V2 fit

Left : Global spectral mismatch for 1 and 2 components.
Right : Best matching intercalibration.



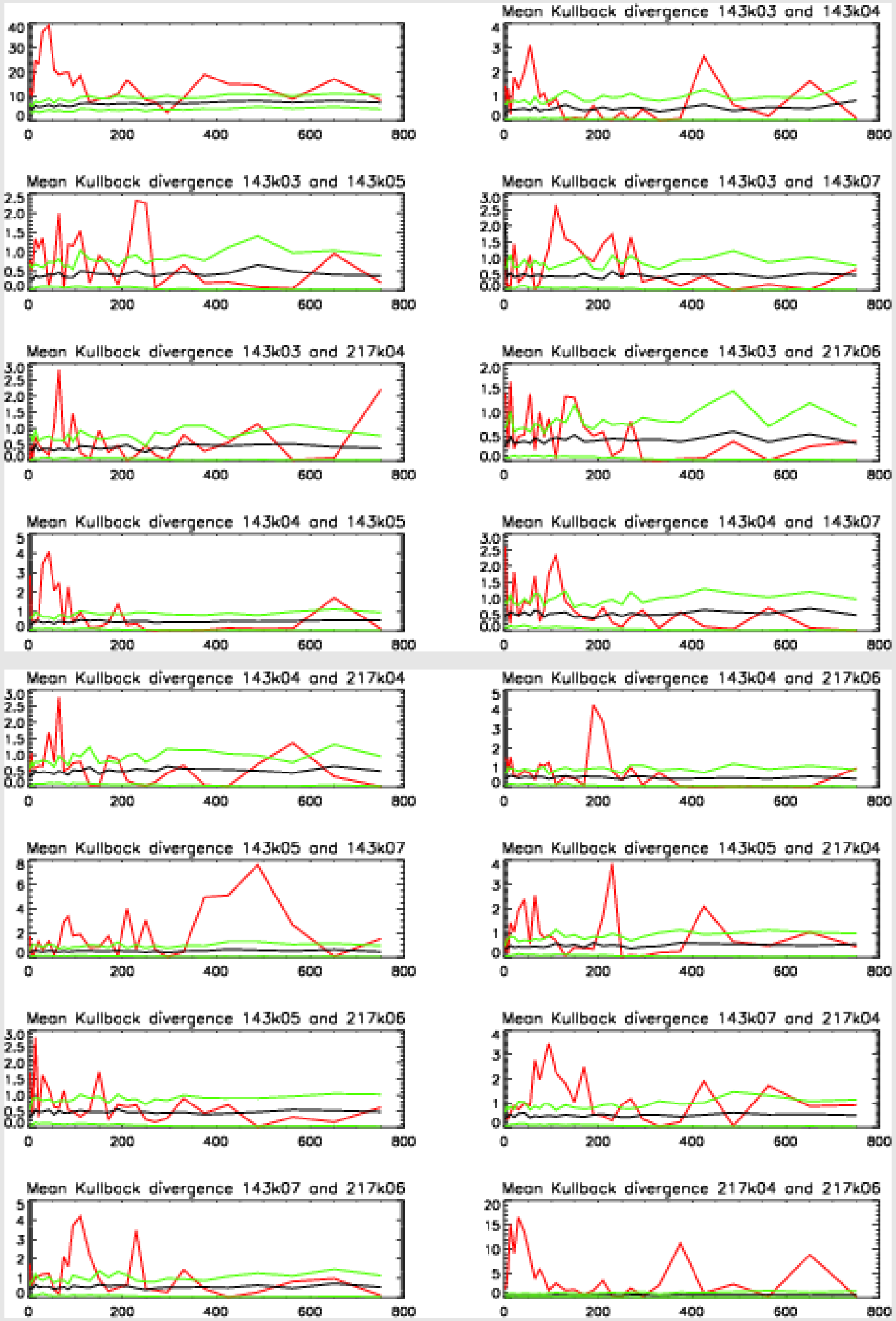


When the model holds :

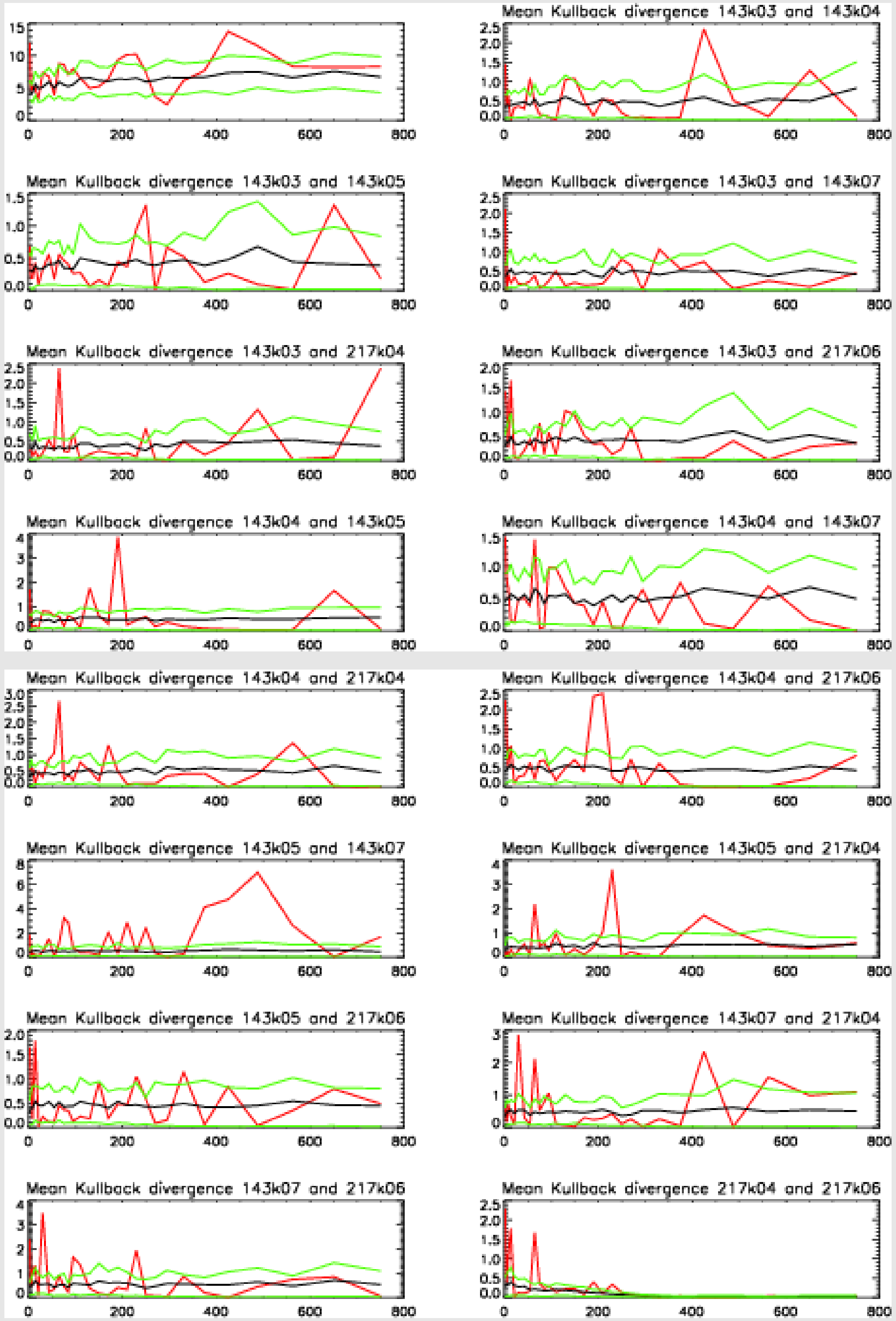
$$2n \left\langle \min_{\theta} K\left(\widehat{R},\ R(\theta)\right)\right\rangle \approx \frac{1}{2} N_{\text{bolo}}(N_{\text{bolo}}+1) - (N_{\text{bolo}} + N_{\text{comp}})$$

Pairwise mismatch for Archeops V2

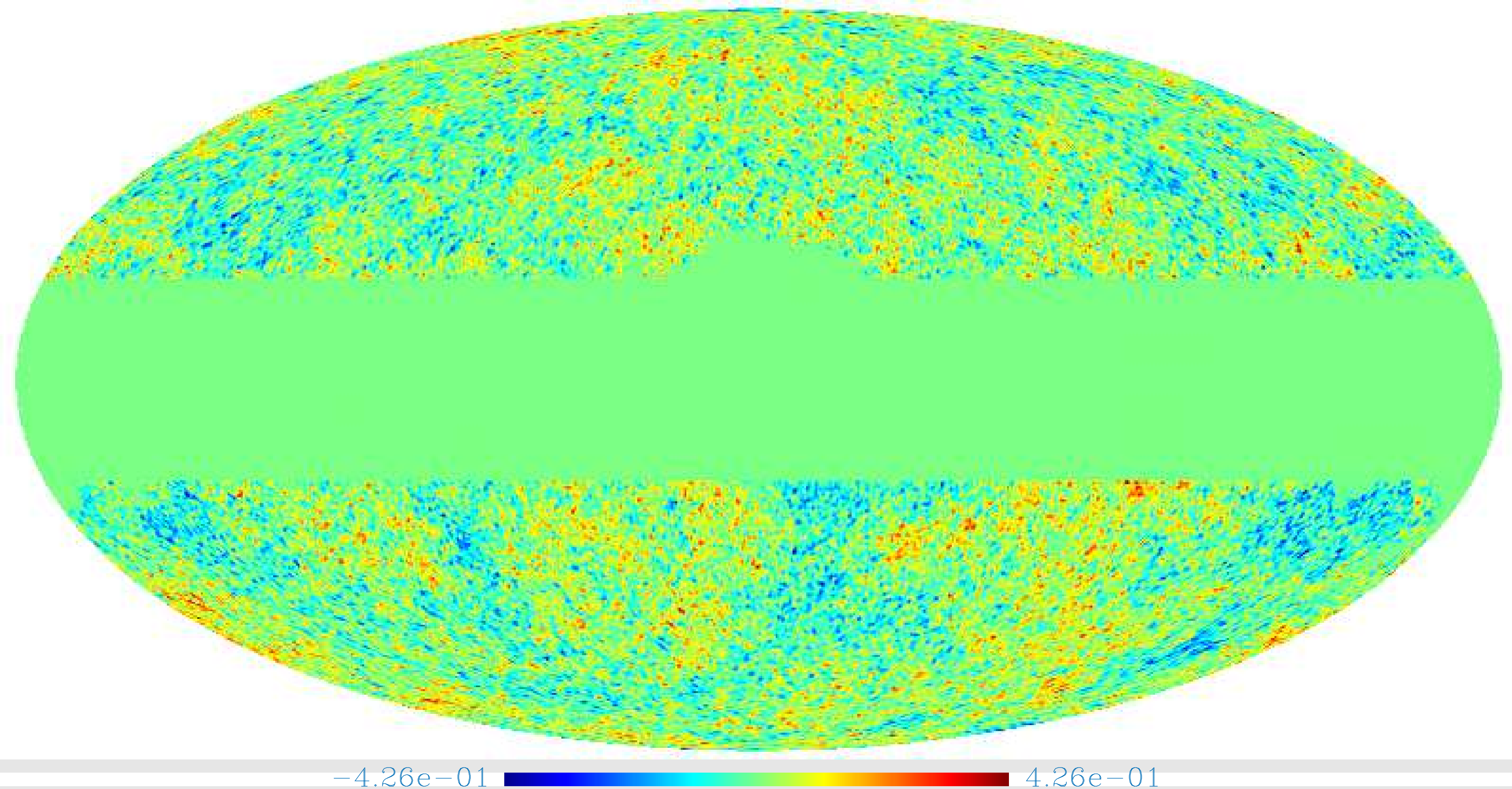
With 1 component (+ noise)



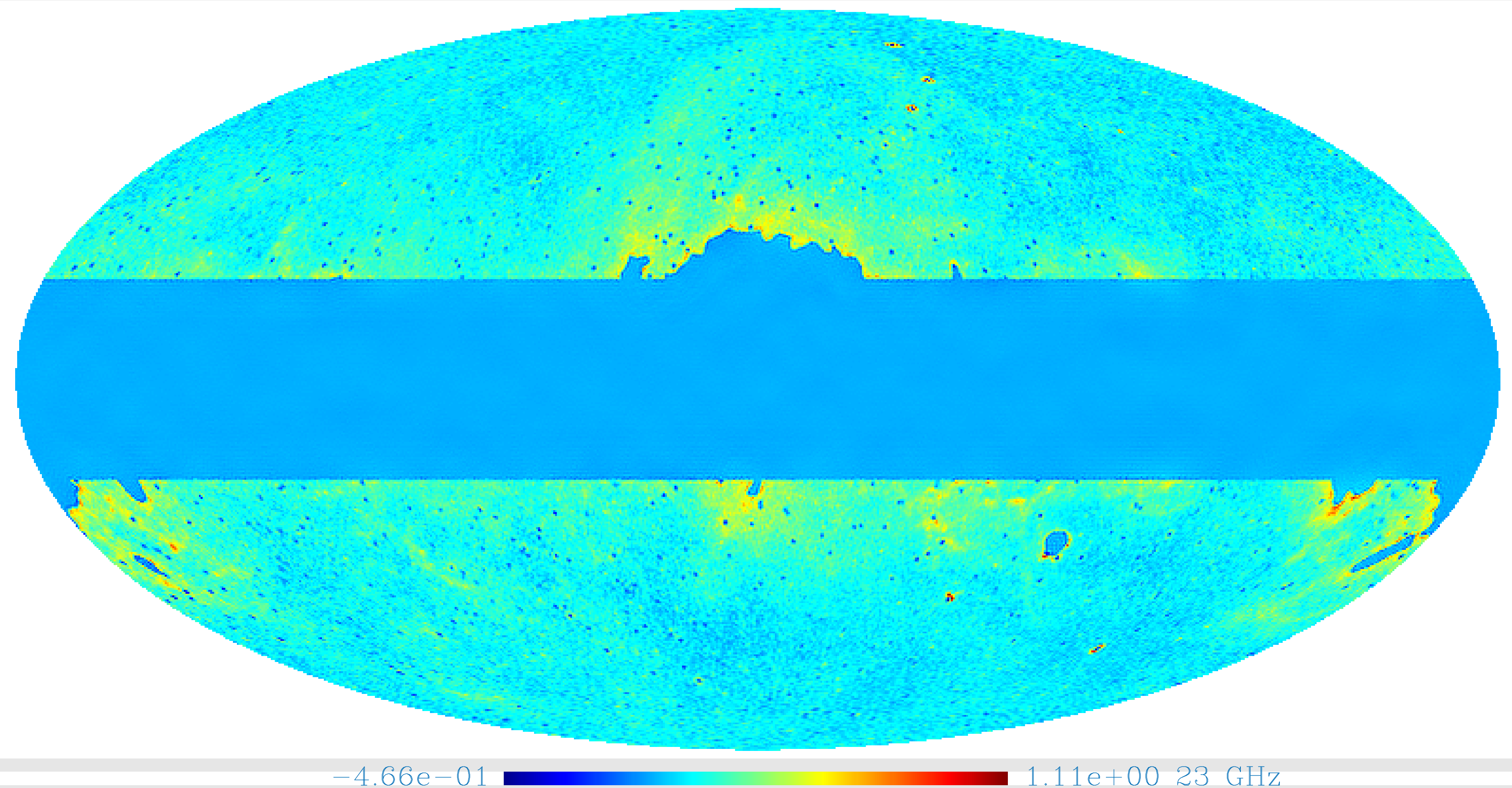
With 2 components (+ noise)



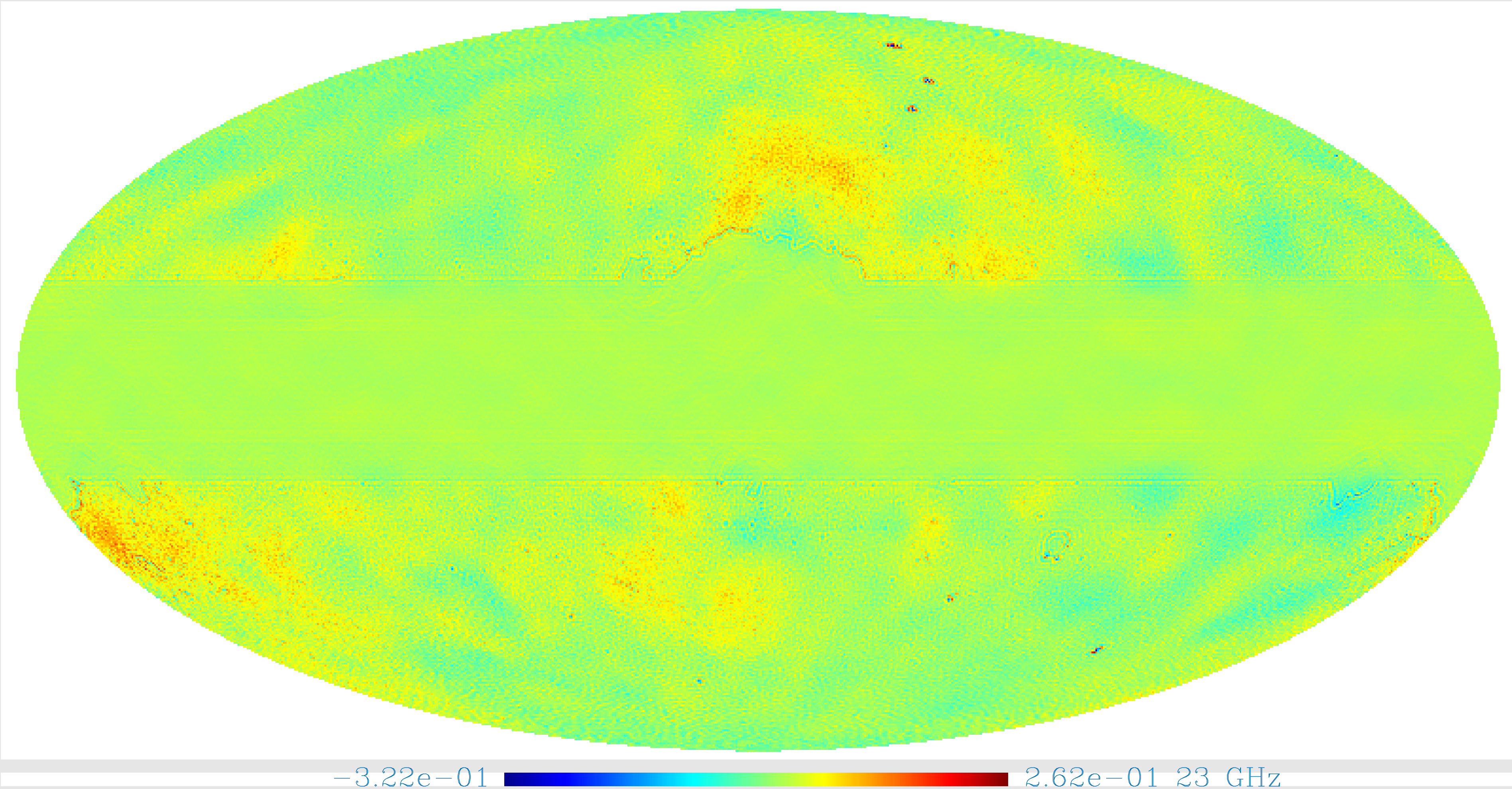
CMB



Galaxy



???



-3.22e-01
2.62e-01 23 GHz

Concluding comments

SMICA
– Flexible
– Maximum likelihood
– Built-in measure of fit
– Easy (manageable) model of correlated components
– Huge data compression
– Easy beam correction
– Optimal for Gaussian stationary, still consistent otherwise.
– *Poor* handling of highly non Gaussian comp's.

About MEM

- Maximum entropy method (a misnomer, IMHO)
- Uses simple non Gaussian models for (some) components.
- Reduces to the (linear) Wiener filter for Gaussian models
- Tremendous information in the non Gaussian part of the data...
- ...but poorly expressed in the frequency domain