Polarized CMB cleaning with non-parametric spectral matching

J.-F. Cardoso,
LTCI CNRS, Télécom ParisTech

APC, IAP & the Planck collaboration

Polarized Foreground for Cosmic Microwave Background

Workshop at Max-Planck-Institut für Astrophysik, Garching, Germany, November 2628, 2012.

Two important (to me) questions

- What am I doing here ?
  - Planck release of temperature only CMB maps is about to happen.
- Will I get stoned ?
  - For CMB cleaning, I will advocate a non-parametric approach.

Note: all figures from data simulated by the Planck Sky Model (see J. Delabrouille's talk).

CMB cleaning

[•] Tasks:

- Combine sky maps to disentangle astrophysical emissions: component separation proper.

  or

- Focus on CMB extraction/cleaning (this talk).

[•] A range of options for CMB cleaning:

- Very blind: template fitting, the ILC family,. . .
- Non-parametric: assumes some foreground coherence (this talk),
- Parametric: assumes SEDs, spectral indices, power laws. . .

Combining channels with the Best Linear Unbiased Estimator (BLUE)

• The BLUE

Given contaminated observations of s with known gains a_i, that is, $x_i = a_i s + n_i$, or

$$\mathbf{x} = \mathbf{a} s + \mathbf{n}$$

where contamination $\mathbf{n}$ is noise+foreground, the linear estimator

$$\widehat{s} = \sum_i w_i x_i = \mathbf{w}^\dagger \mathbf{x}$$

of s with zero bias ($\mathbf{w}^\dagger \mathbf{a} = 1$) and minimum variance has weights given by

$$\mathbf{w} = \frac{\mathbf{C}^{-1}\mathbf{a}}{\mathbf{a}^\dagger \mathbf{C}^{-1}\mathbf{a}} \qquad \mathbf{C} \stackrel{\text{def}}{=} \text{Cov}(\mathbf{x}) \qquad \text{[the BLUE]}$$

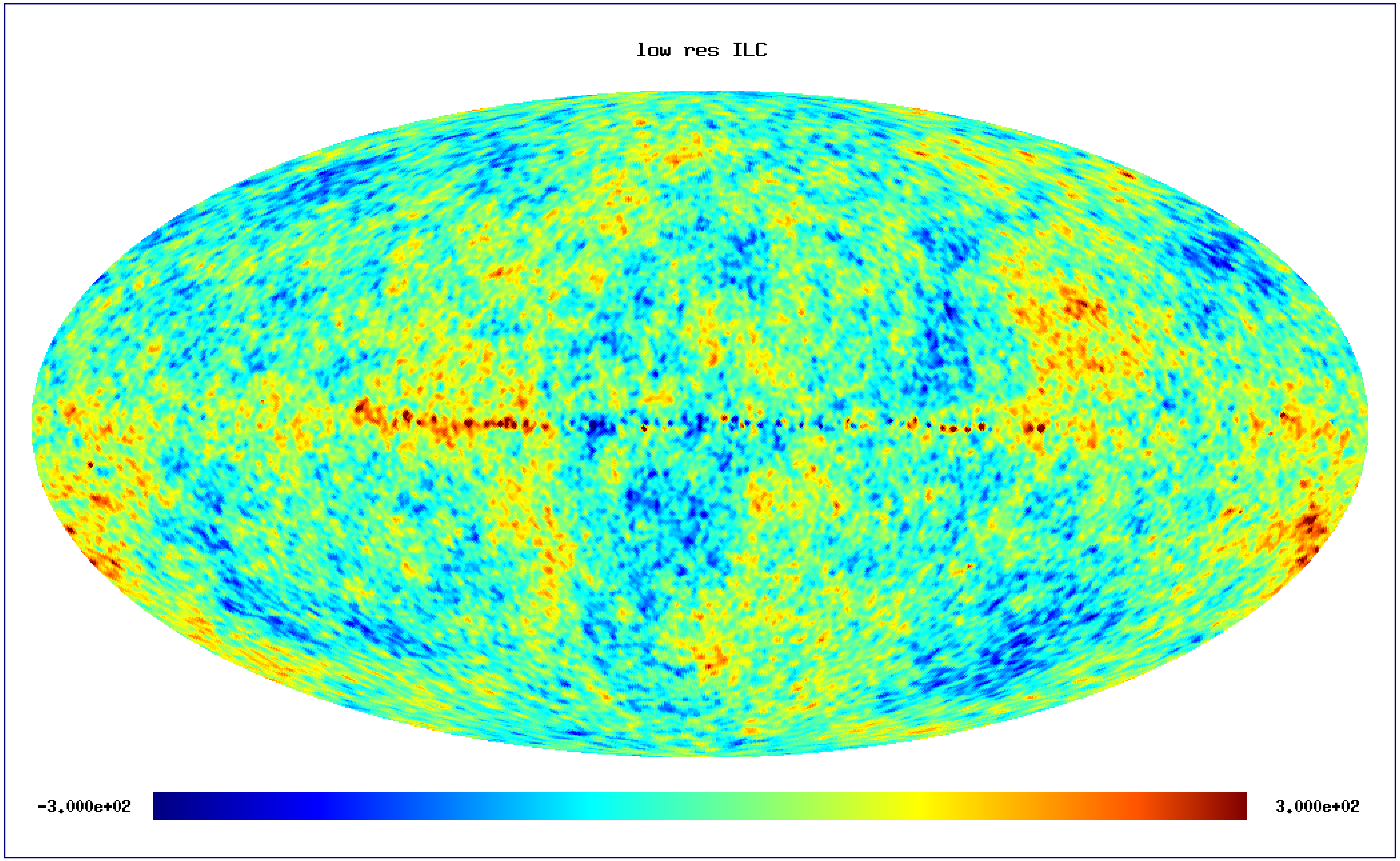
• Beauty of the BLUE: it only requires knowing:

1) the gain vector $\mathbf{a}$ *i.e.* a CMB-calibrated instrument

2) the covariance matrix of the data $\mathbf{C} = \text{Cov}(\mathbf{x})$

• Replacing the (unknown) covariance matrix $\mathbf{C}$ by its sample estimate $\widehat{\mathbf{C}} = 1/P \sum_p \mathbf{x}(p)\mathbf{x}(p)^\dagger$ yields the super simple ILC (Internal Linear Combination).

A plain, low-resolution (1 degree) ILC map from PSM simulations



Is it good enough ? Can we do better? Can we do better at high resolution ?

Beating (up) the BLUE

- Q: Given that

  - Linearity is a must,

  - The BLUE is MSE-optimal among linear filters,

can you beat it ? What could go wrong ?

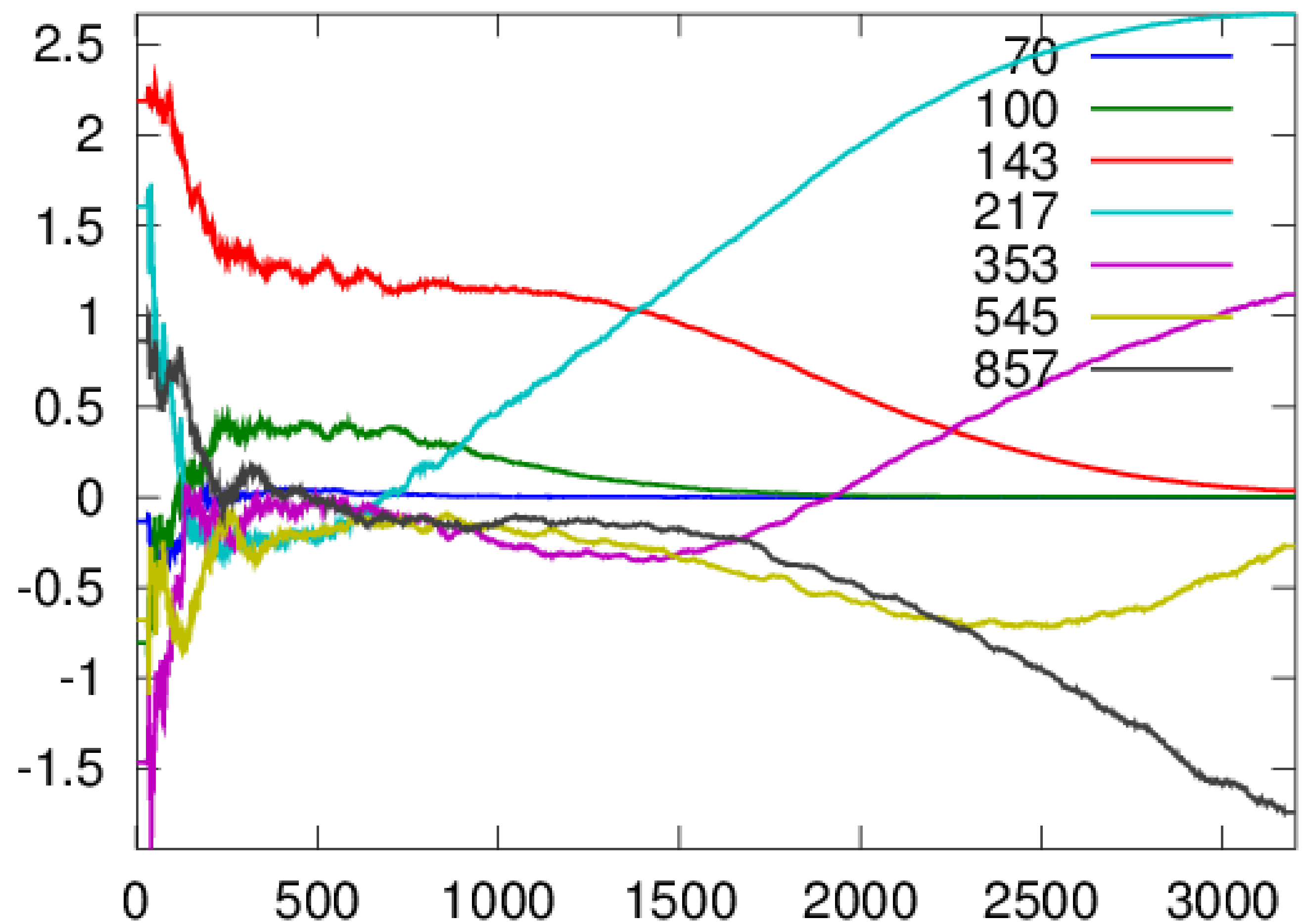
- A: Many things can go wrong in many ways !

  - Total mean-square error may not be the best criterion, after all.
    It lumps together foregrounds and noise. And also multipoles. And also sky regions.

  - Need to adapt to 'local conditions':
    We do not fight the same ennemy in various parts of the sky, in various multipole ranges.
    The case for harmonic filtering or even wavelet/needlet filtering.

  - Need to estimate the data covariance matrix.
    – Which covariance matrix ? (Pixel space, harmonic space, wavelet/needlet space ?)
    – Direct estimation from the data ? Beware chance correlations !
    – Maybe some modelling of the covariance matrix could help. . .

The ILC in harmonic space



ILC coefficients in harmonic space (for maps rebeamed at a 5').

ILC, template fitting, and chance correlation

Template fitting cleans map x_1 using the x_2 template according to $x_1 - \frac{\langle x_1 x_2 \rangle}{\langle x_2^2 \rangle} \cdot x_2$.

That does a perfect job with perfect templates, perfectly uncorrelated with the CMB.

Otherwise. . . let's look at a toy example:
a contaminated channel $x_1 = s + \alpha f$ and an (approximate) template $x_2 = f'$. Then:

$$\hat{s} = x_1 - \frac{\langle x_1 x_2 \rangle}{\langle x_2^2 \rangle} \cdot x_2 = s - \underbrace{\frac{\langle s f' \rangle}{\langle f'^2 \rangle} \cdot f'}_{\text{Chance corr.}} + \alpha \underbrace{\left(f - \frac{\langle f f' \rangle}{\langle f'^2 \rangle} \cdot f' \right)}_{\text{Non-rigid scaling}}$$

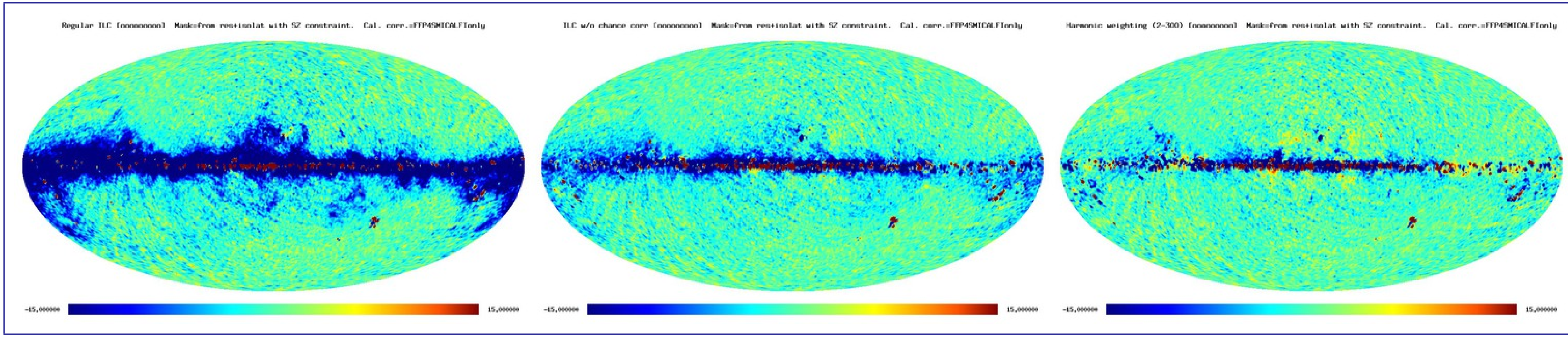
- The error due to chance correlation is independent of the level of α of contamination.

  One pays the price for any template thrown at the data, whether or not it's in there.

- If one assumes rigid scaling $f = f'$, chance correlation dominates the error.

What is hitting us harder: non-rigid scaling or chance chance correlation ?
<u>You tell me about the former, I tell you about the latter.</u>

ILC, chance correlation and harmonic weighting



Residuals ($\widehat{\mathrm{CMB}} - \mathrm{CMB}$) for 3 ILC's at low-resolution (1 degree) on a $\pm 15\mu K$ scale.

- Left: Plain pixel-based ILC.
- Center: Same with chance correlation CMB/fgd articially removed.
- Right: Covariance matrix estimated from weighted spherical harmonic coefficients.

From the BLUE to SMICA

We saw the 'optimality' of the BLUE but

- It must be made multipole dependent. That's easy:

$$\widehat{s}_{\ell m} = \mathbf{w}_\ell^\dagger \mathbf{x}_{\ell m}, \qquad \mathbf{w}_\ell = \frac{\mathbf{C}_\ell^{-1}\mathbf{a}}{\mathbf{a}^\dagger \mathbf{C}_\ell^{-1}\mathbf{a}}$$

  where the $N_{\text{chan}} \times N_{\text{chan}}$ matrix $\mathbf{C}_\ell$ contains all the auto- and cross-spectra.

- The spectra $\mathbf{C}_\ell$ are unknown and using the empirical covariance matrices:

$$\widehat{\mathbf{C}}_\ell \stackrel{\text{def}}{=} \frac{1}{2\ell+1} \sum_m \mathbf{x}_{\ell m} \mathbf{x}_{\ell m}^\dagger$$

  as a plugin replacement is not enough to tame chance correlation at large scales.

- So we set up a spectral model $\mathbf{C}_\ell(\theta)$:

$$\mathbf{C}_\ell(\theta) = \underbrace{\mathbf{a}\mathbf{a}^\dagger C_\ell}_{\text{CMB}} + \underbrace{\mathbf{C}_\ell^{\text{gal}}(\theta^{\text{gal}})}_{\text{galactic fgd}} + \underbrace{\mathbf{C}_\ell^{\text{efg}}(\theta^{\text{efg}})}_{\text{extra galactic}} + \underbrace{\text{diag}(\sigma_{i\ell}^2)}_{\text{noise}}, \qquad \theta = \{C_\ell, \theta^{\text{gal}}, \theta^{\text{efg}}, \sigma_{i\ell}^2\},$$

  and fit it (in the maximum likelihood sense) to $\widehat{\mathbf{C}}_\ell$, and use the result $\mathbf{C}_\ell(\widehat{\theta})$ in the BLUE.

- $\rightarrow$ Spectral Matching Independent Component Analysis (SMICA).

Foreground models: parametric, or not.

- The global spectral model $\mathbf{C}_\ell(\theta) = \underbrace{\mathbf{a}\mathbf{a}^\dagger C_\ell}_{\text{CMB}} + \underbrace{\mathbf{F}\mathbf{P}_\ell\mathbf{F}^\dagger}_{\text{foreground}} + \underbrace{\text{diag}(\sigma_{i\ell}^2)}_{\text{noise}},$

Here, $\mathbf{F}$ is an $N_{\text{chan}} \times f$ matrix and $\mathbf{P}_\ell$ an $f \times f$ positive matrix depending on ℓ.

- A rigid model: $\mathbf{F}$ is made of known emission laws: $\mathbf{F} = [\mathbf{a}_{\text{dust}}\ \mathbf{a}_{\text{synch}}\ \mathbf{a}_{\text{CO}}\ \ldots]$.

Then matrix P_ℓ contains the auto- and cross-spectra of those f foregrounds.

- Sky-varying emissivity costs one column: $\mathbf{P} = [\mathbf{a}_{\text{dust}}\ \partial\mathbf{a}_{\text{dust}}/\partial T\ \mathbf{a}_{\text{synch}}\ \mathbf{a}_{\text{CO}}\ \ldots]$ at first order.

- A rigid but more flexible model, *e.g.* $\mathbf{P} = \mathbf{P}(T) = [\mathbf{a}_{\text{dust}}(T)\ \mathbf{a}_{\text{synch}}\ \mathbf{a}_{\text{CO}}\ \ldots]$.

•• The foreground emission matrix $\mathbf{P}$ can be controlled by many parameters.

•• Q: How many at most? A: as much as you want! (well, kind of).

Technically, spectral diversity guarantees the blind identifiability of the total foreground emission with f as large as $N_{\text{chan}} - 1$.

The underlying model is that $f < N_{\text{chan}}$ templates with arbitrary emissivities, arbitrary spectra and arbitrary correlations.

We consider here a 'catch-all' foreground component able to confine all the coherent contamination into a non-parametric 'foreground subspace' of dimension $N_{\text{chan}} - 1$ at most.

Models for polarization analysis

Now that we disposed of the painful need of parametric foreground modeling, we can serenely addreess polarization ;-)

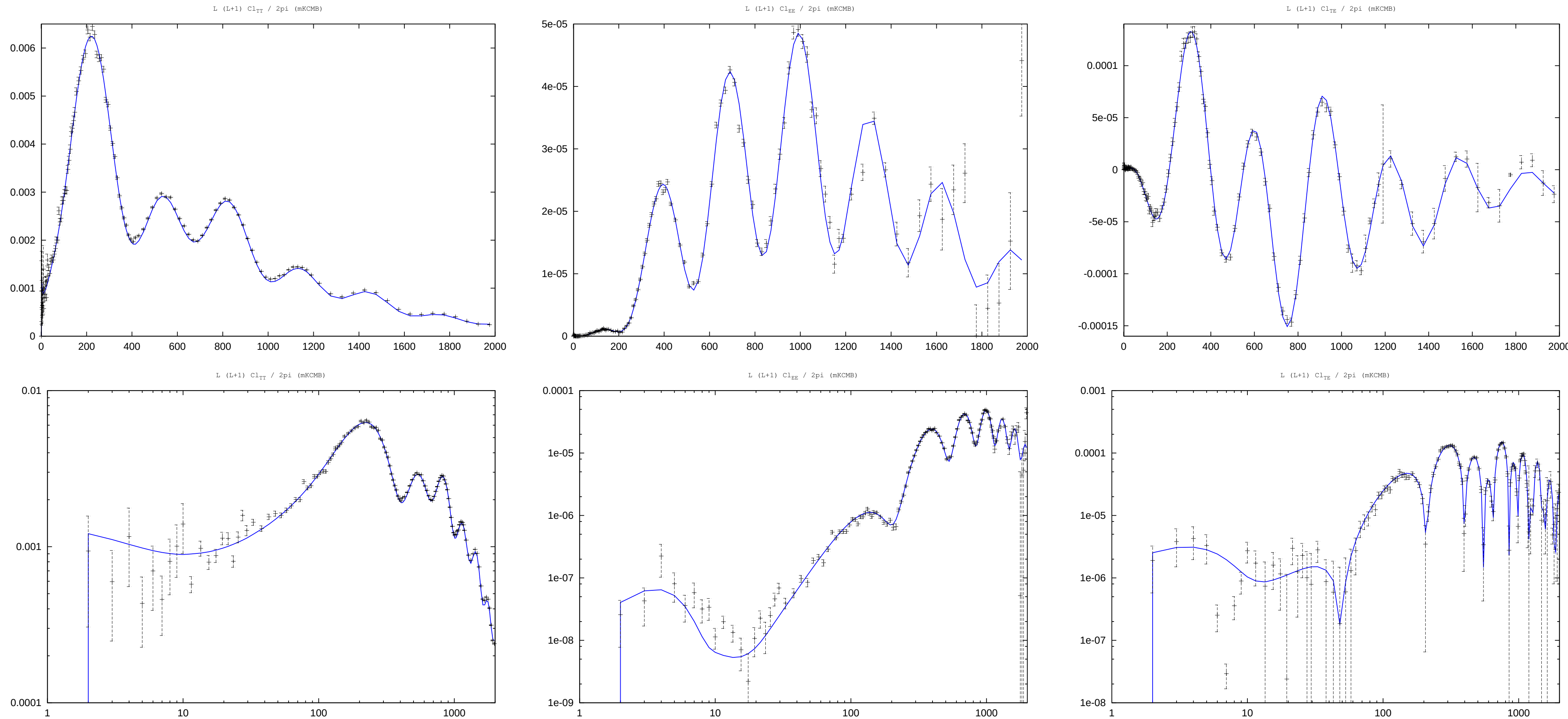
- $T + E$. For instance, for Planck, stacking the T modes of the 9 temperature channels and the E modes of the 7 polarized channels, we may use

$$\mathbf{C}_\ell = \mathsf{Cov}(\mathbf{x}_{\ell m}) = \mathsf{Cov}\begin{bmatrix} \mathbf{x}_{\ell m}^T \\ \mathbf{x}_{\ell m}^E \end{bmatrix} = \begin{bmatrix} \mathbf{a} & 0 \\ 0 & \mathbf{a} \end{bmatrix} \begin{bmatrix} C_{\mathsf{TT}}(\ell) & C_{\mathsf{TE}}(\ell) \\ C_{\mathsf{ET}}(\ell) & C_{\mathsf{EE}}(\ell) \end{bmatrix} \begin{bmatrix} \mathbf{a} & 0 \\ 0 & \mathbf{a} \end{bmatrix}^\dagger + \mathbf{F}\mathbf{P}_\ell\mathbf{F}^\dagger + \mathbf{N}_\ell$$

- B only. Measure of the tensor-to-scale ratio r in presence of foregrounds using SMICA. See paper by Betoule *et al.* 2009.

Some results from early (2008) Planck simulations

TT, TE, EE spectra using a 7-dimensional foreground component with a free (non-parametric) $(9+7) \times 7$ foreground emission matrix $\mathbf{P}$.



CMB power spectra, from top to bottom : TT, EE, and TE

Error bars $\pm 1\sigma$ from the Fisher information matrix.

Notes and conclusion

Notes:

- SMICA as a spectral estimator.
  Actually, it does component separation (at the map level) optionally after spectral separation.
- SMICA also is a likelihood (possiby parametric). See work on PLIK at IAP.
- SMICA as a calibrator.

Conclusions:

Some continuity: template fitting $\rightarrow$ ILC $\rightarrow$ non-parametric SMICA.

Non-parametric foreground modeling with SMICA.
All the more useful for CMB cleaning as long as polarized foreground models remain uncertain.

The parametric / non- parametric also is a tradeoff between statistical efficiency and robustness.
Need to learn from forthcoming Planck data and simulations.

Non parametric: Let your data talk and listen to them.