Non-Gaussianity in multidetector component separation

Paper in preparation

M. Citran,[a,b,1] **H.V. Tran,**[a] **B. van Tent**[b] **G. Patanchon**[a]

[a]Université Paris Cité, CNRS, Astroparticule et Cosmologie, 75013 Paris, France
[b]Université Paris-Saclay, CNRS/IN2P3, IJCLab, 91405 Orsay, France

Michele Citran

13/10/2025

Presentation Outline

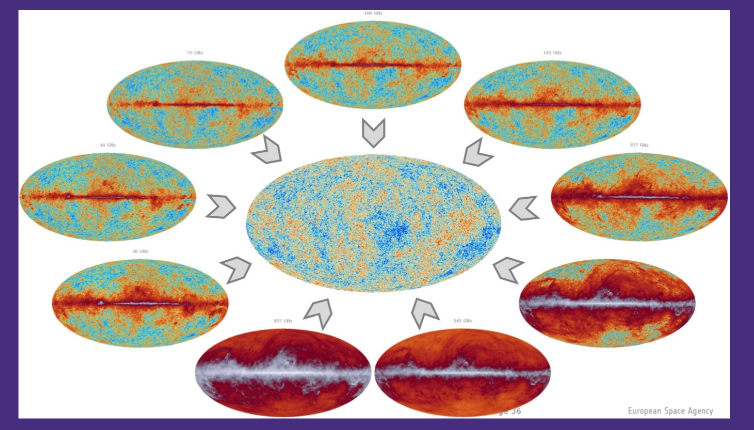
FIRST PART: non-Gaussianity in component separation

- The **definition of Non-Gaussianity (and anisotropy): generic and bispectral** non-Gaussianity
- The process of **component separation: SMICA**
- How to **include non-Gaussianity in harmonic space based component separation methods**
- Results

SECOND PART: multi-detector multi-component bispectrum estimator

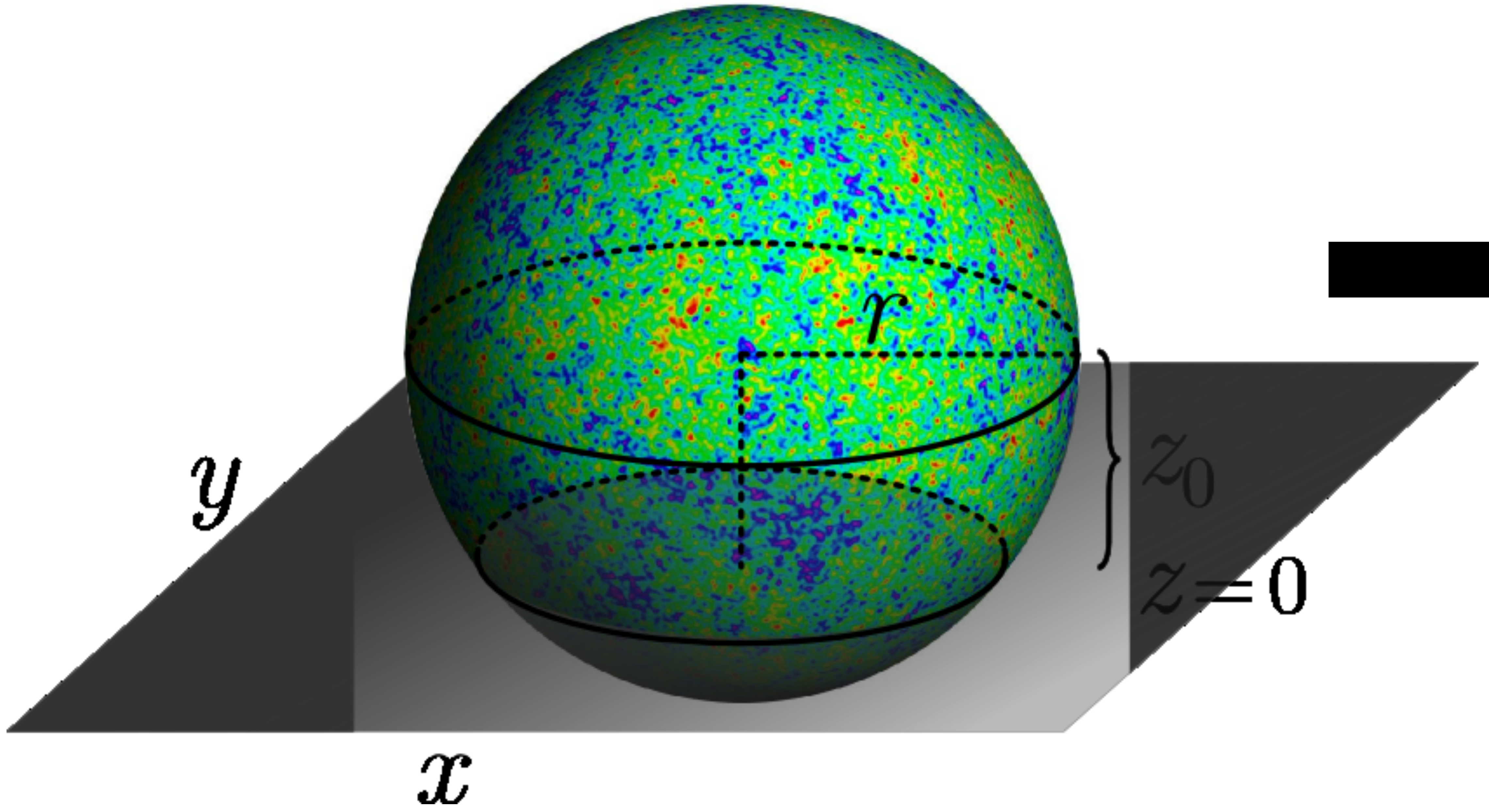
- **Bispectrum likelihood**
- **Binning and bispectrum model**
- Results

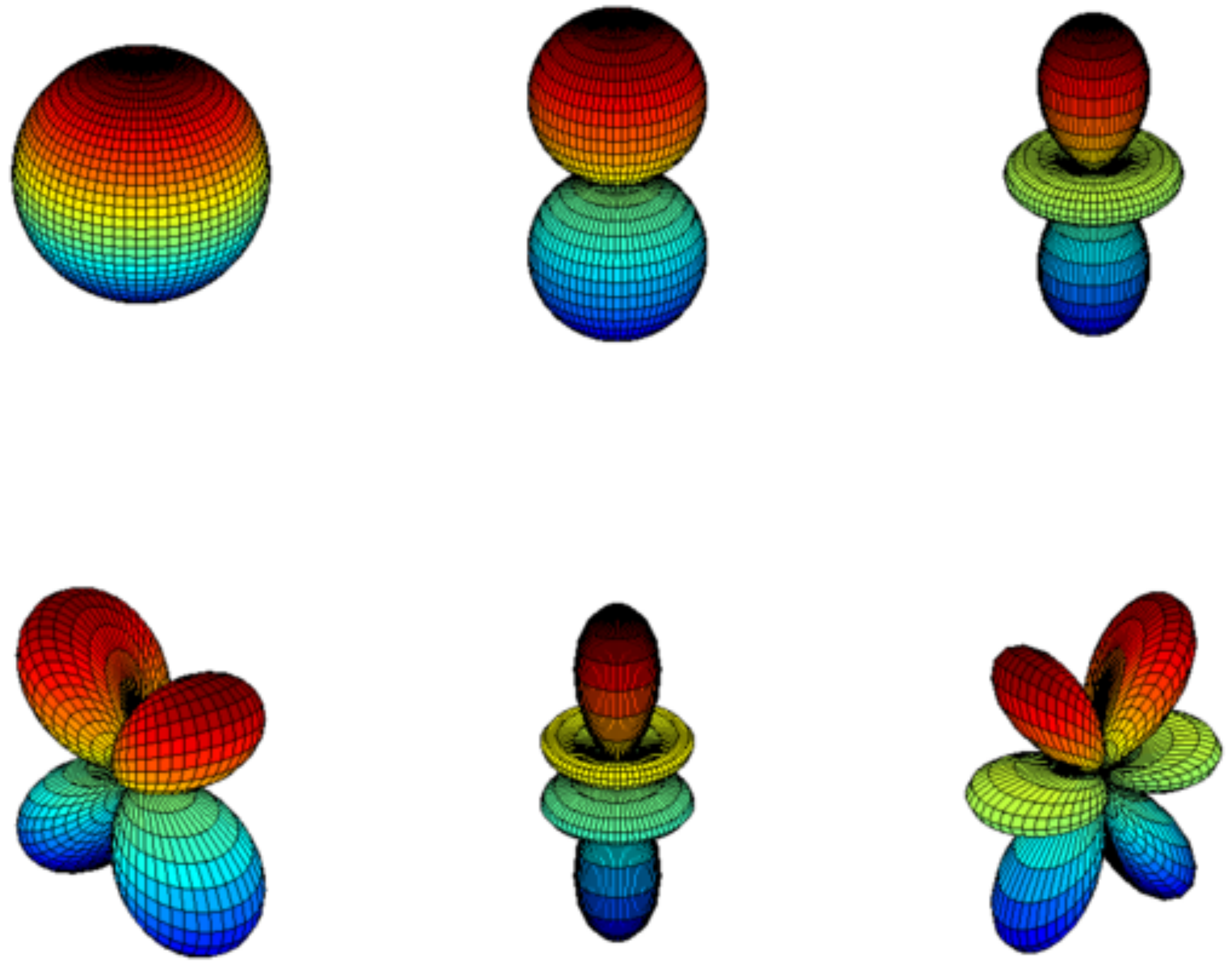
Non Gaussianity and Anistropy: Definition



Isotropic and gaussian

$$P(a_{lm}) = G(a_{lm} \,|\, C_l) \propto \exp(-\frac{1}{2} a^*_{lm} C_l^{-1} a_{lm})$$

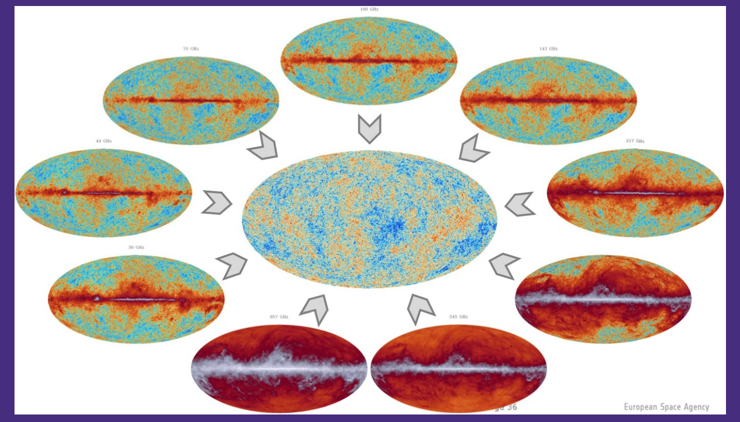




Spherical harmonics decomposition

$\phi(\hat{n}) \rightarrow a_{lm}$

Non Gaussianity and Anistropy: Definition



Non Gaussianity: anything that deviates from a gaussian:

$$P(a_{lm}) \propto \exp(-\frac{1}{2} a^*_{lm} C_l^{-1} a_{lm}) \quad \langle a^*_{lm} a_{l'm'} \rangle = \delta_{ll'} \delta_{mm'} C_l$$

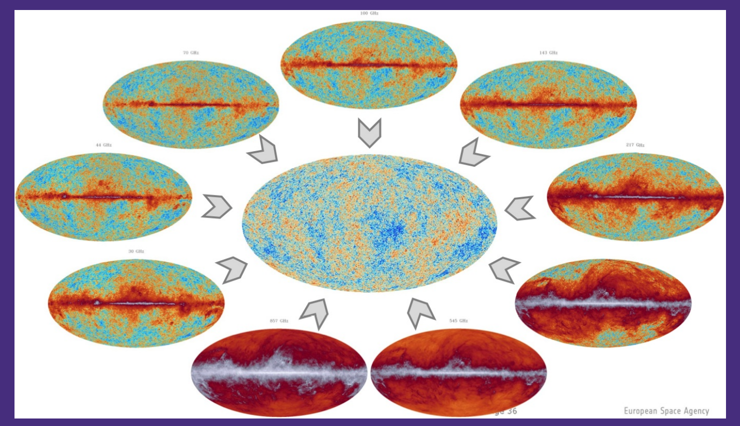
Non Gaussianity can be described by **higher order correlators**

$$\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle, \ \langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} a_{l_4 m_4} \rangle, \ \ldots$$

Anisotropy by **off-diagonal terms**

$$C_{l_1 l_2, m_1 m_2} = \langle a^*_{l_1 m_1} a_{l_2 m_2} \rangle$$

Non Gaussianity: Bispectrum



In the isotropic case, the **Bispectrum** is defined by the 3-point correlator

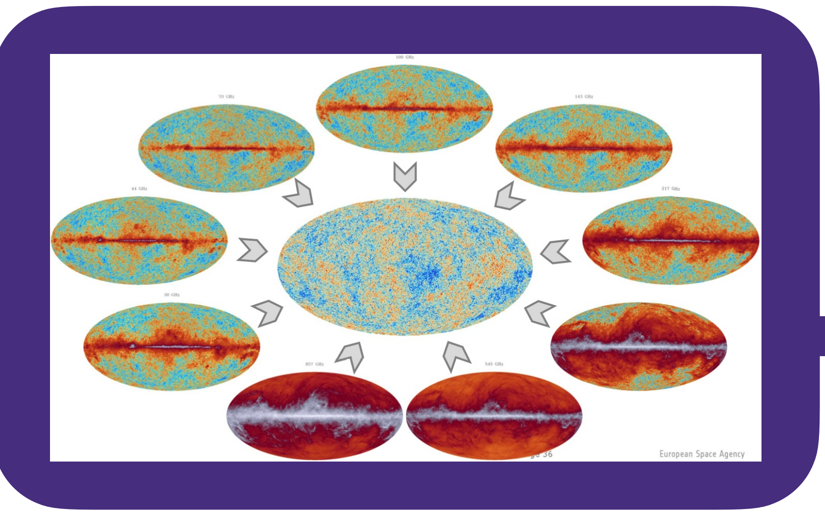
$$\langle a_{l_1m_1}a_{l_2m_2}a_{l_3m_3}\rangle = \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} B_{l_1l_2l_3}$$

With **Variance**, in the **weakly non-gaussian** case

$$V_{l_1l_2l_3} = \langle B^2_{l_1l_2l_3}\rangle - \langle B_{l_1l_2l_3}\rangle^2 = 6C_{l_1}C_{l_2}C_{l_3}$$

$$= 0$$

Non Gaussianity: Bispectrum



In the isotropic case, the **Bispectrum** is defined by the 3-point correlator

$$\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle = \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} B_{l_1 l_2 l_3}$$

With **Variance**, in the **weakly non-gaussian** case

$$V_{l_1 l_2 l_3} = \langle B^2_{l_1 l_2 l_3} \rangle - \langle B_{l_1 l_2 l_3} \rangle^2 = 6 C_{l_1} C_{l_2} C_{l_3}$$

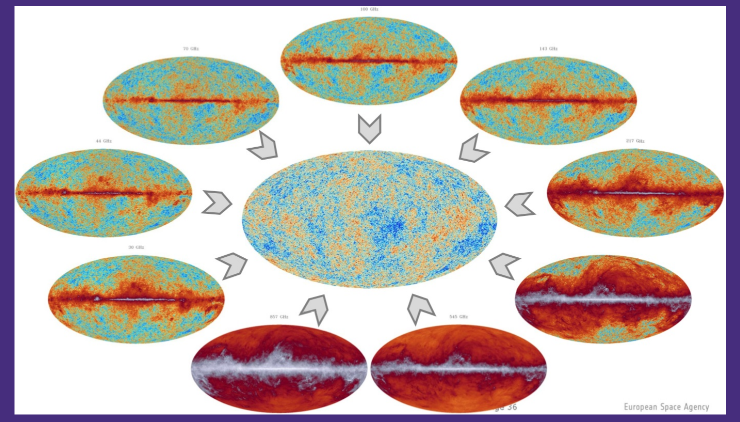
$$\langle B_{l_1 l_2 l_3} \rangle = 0$$

1) A **single realization** will always have a **non-zero bispectrum**

2) **Signal to noise** ratio is important

$$\propto \sqrt{\frac{B^2_{l_1 l_2 l_3}}{C_{l_1} C_{l_2} C_{l_3}}}$$

Component separation: Definition



Objective:

Combining **data from several detectors** to **separate** several **foreground components from the CMB**

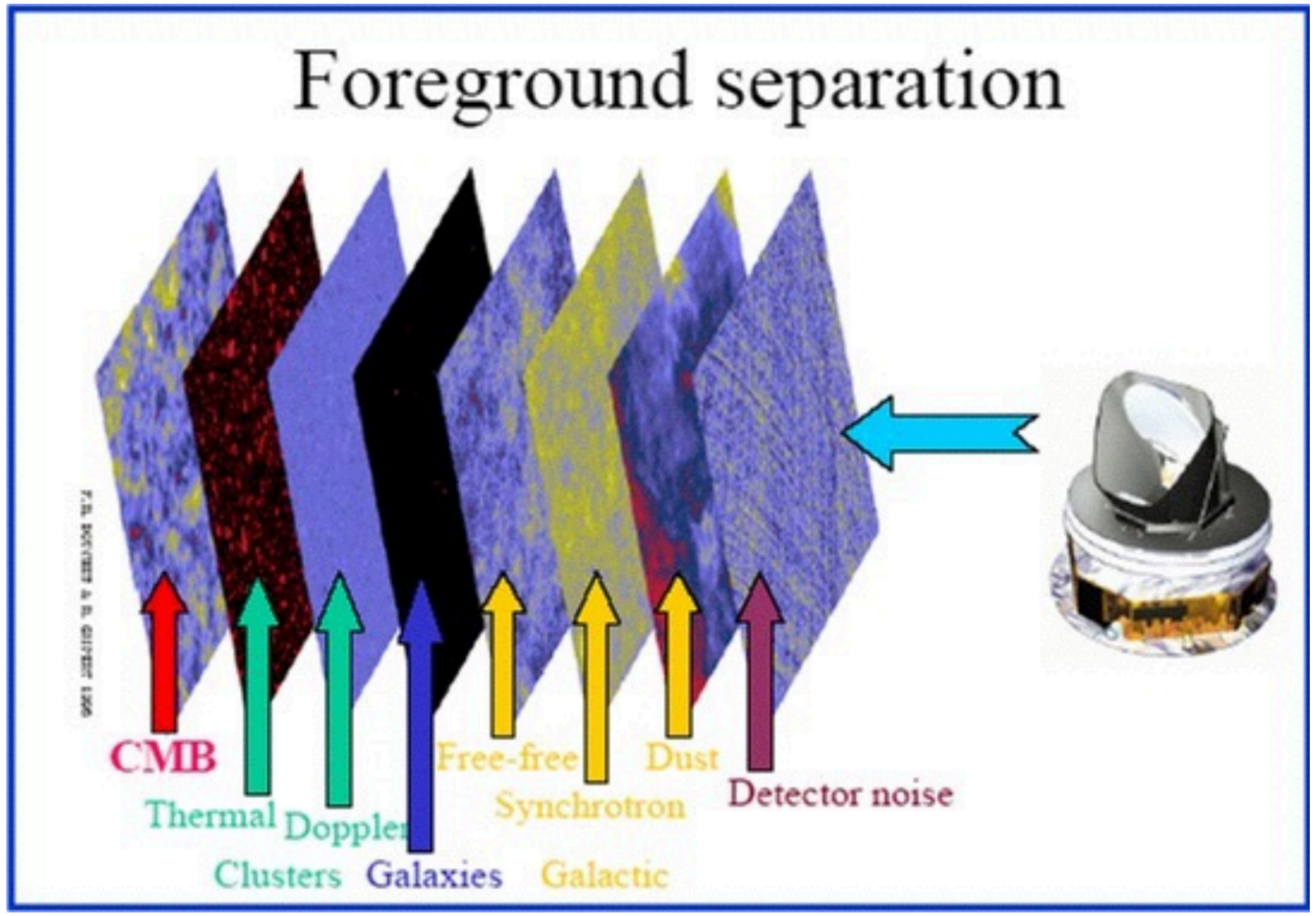
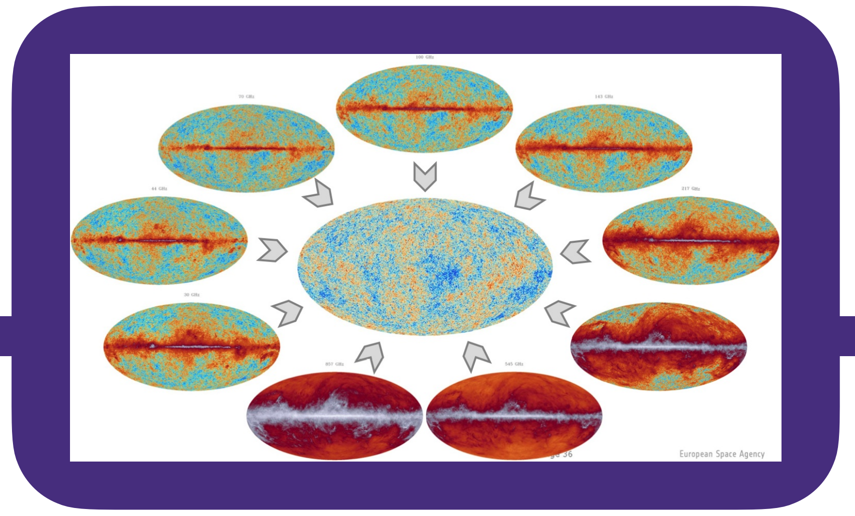


Image courtesy of F. Bouchet

Component separation: SMICA



SMICA (Spectral Matching Independent Component Analysis) is a **non-parametric** method [1],[2] that works in the **spherical harmonic domain**

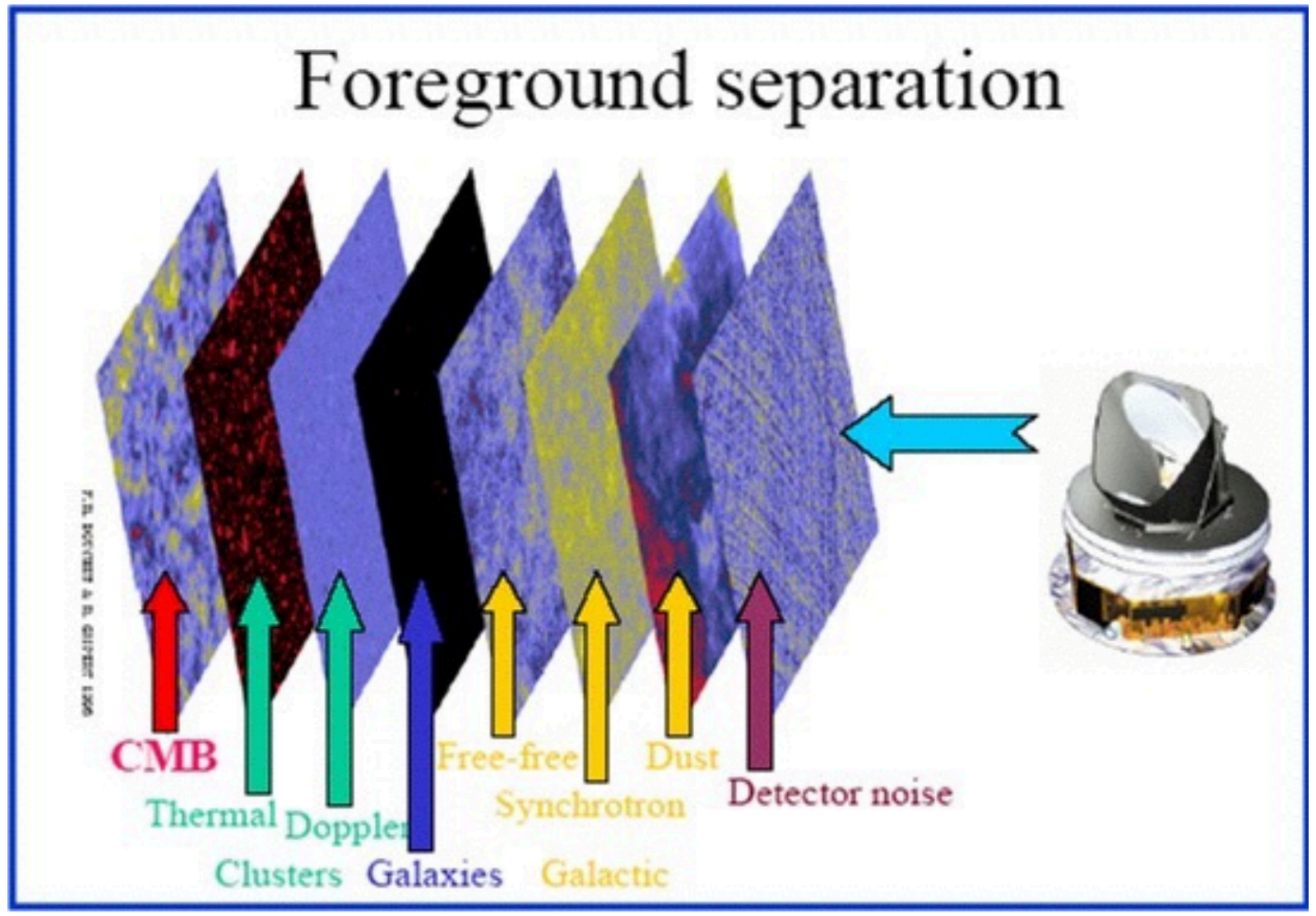
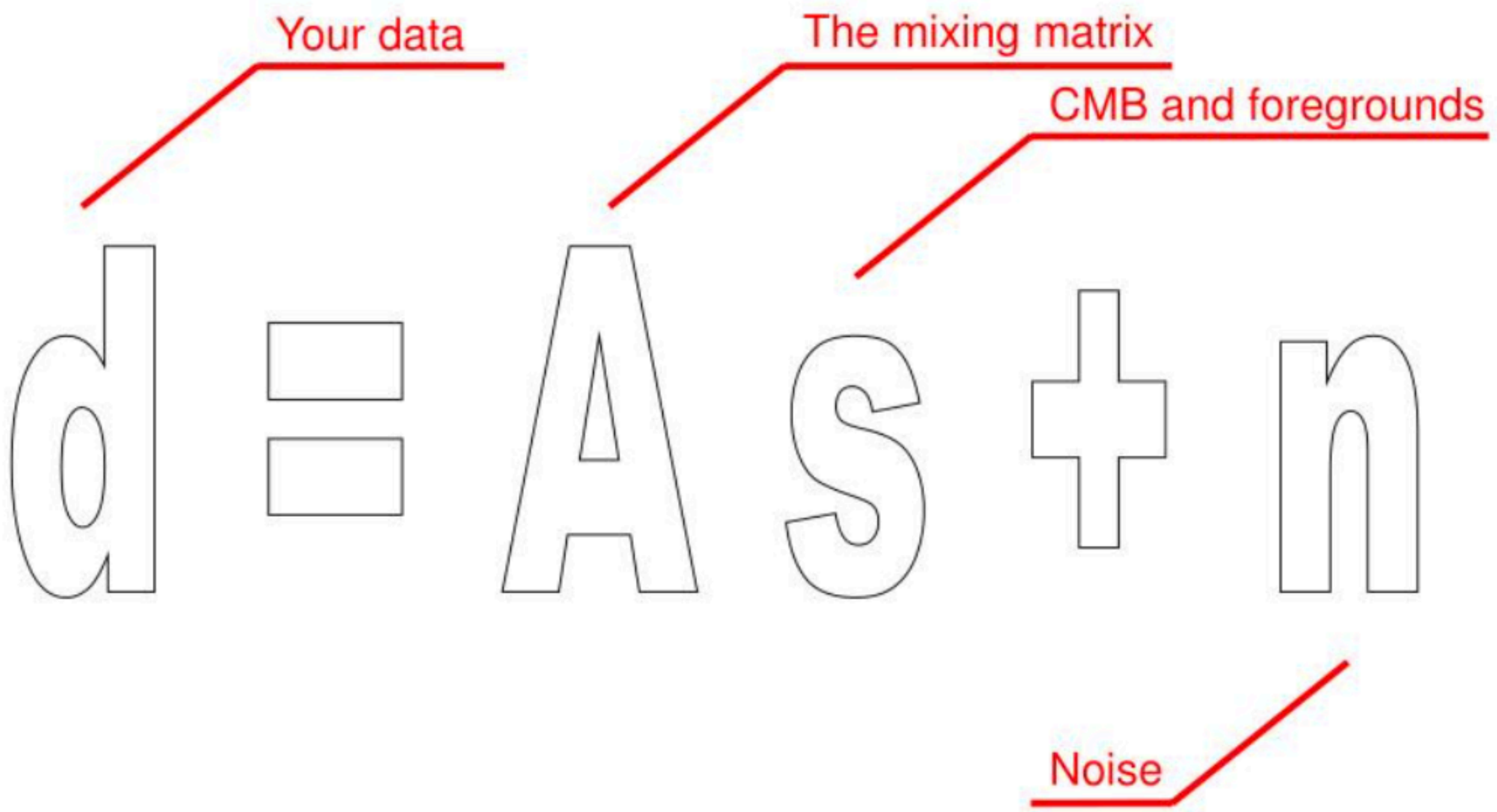


Image courtesy of F. Bouchet

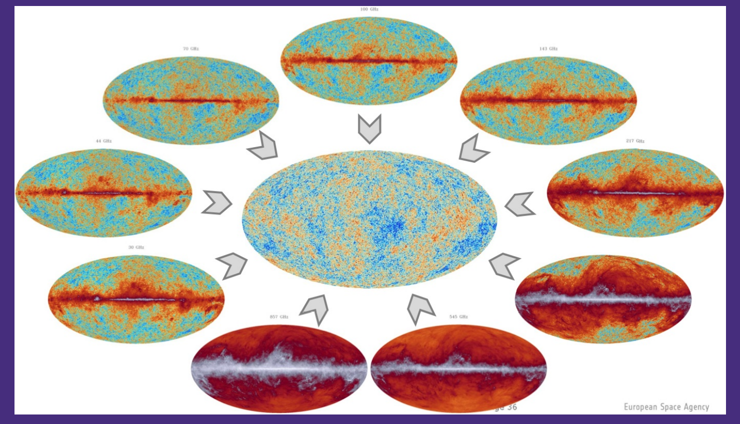
Linear superposition



[1] J. Delabrouille, J.-F. Cardoso, G. Patanchon, arXiv:astro-ph/0211504

[2]J.-F. Cardoso, M. Martin, J. Delabrouille, M. Betoule, G. Patanchon, arXiv:0803.1814

Component separation: SMICA



Under the assumptions of [3a-3b]

1) **Isotropy**
2) **Gaussianity**

Detector/Frequency index

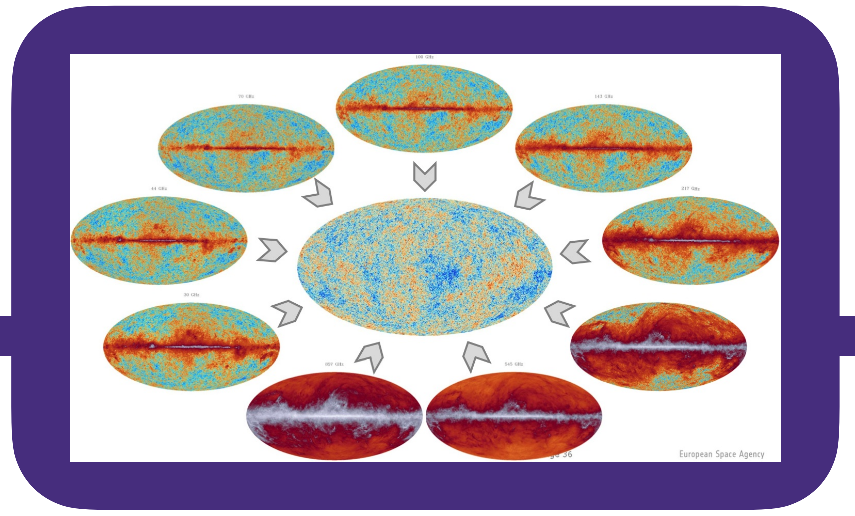
$$PDF(a^{d}_{lm}) = G(a^{d}_{lm}\,|\,R) \qquad \langle (a^{d}_{lm})^{*} a^{d'}_{l'm'} \rangle = \delta_{ll'}\,\delta_{mm'}\,R^{dd'}_{l}$$

SMICA also assumes:

3) **Statistical independence** between components
4) **Uncorrelated noise** across frequencies and in space

[3a] Marinucci, D. and Peccati, G. (2011). Cambridge Univ. Press
[3b] A. Lang, C. Schwab, arXiv:1305.1170v3

Component separation: SMICA

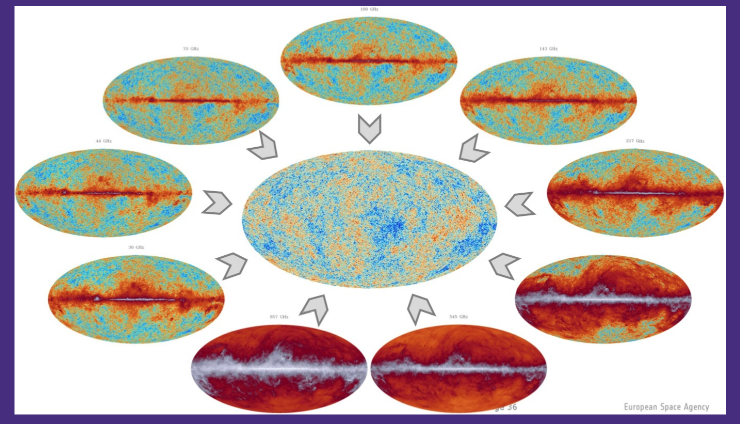


Under this set of assumptions

1) **Isotropy**
2) **Gaussianity**
3) **Statistical independence** between components
4) **Uncorrelated noise** across frequencies and in space

Our sky model is completely defined by $$R = R(A^{d,c}, C_l^c, n_l^d)$$

Component separation: SMICA



Under the assumptions this set of assumptions

1) **Isotropy**
2) **Gaussianity**
3) **Statistical independence** between components
4) **Uncorrelated noise** across frequencies and in space

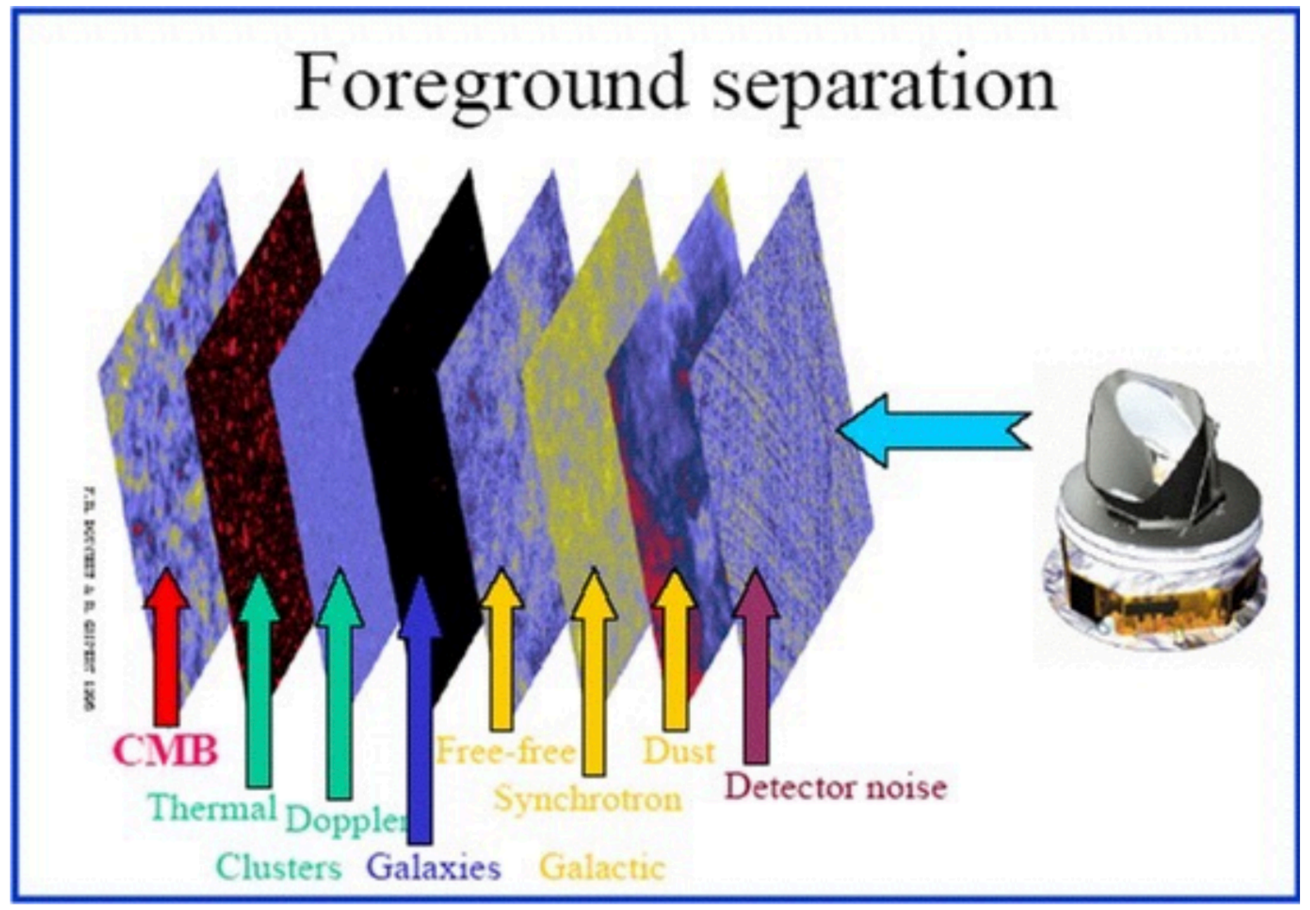


Image courtesy of F. Bouchet

Your data

The mixing matrix

CMB and foregrounds

$$d = A s + n$$

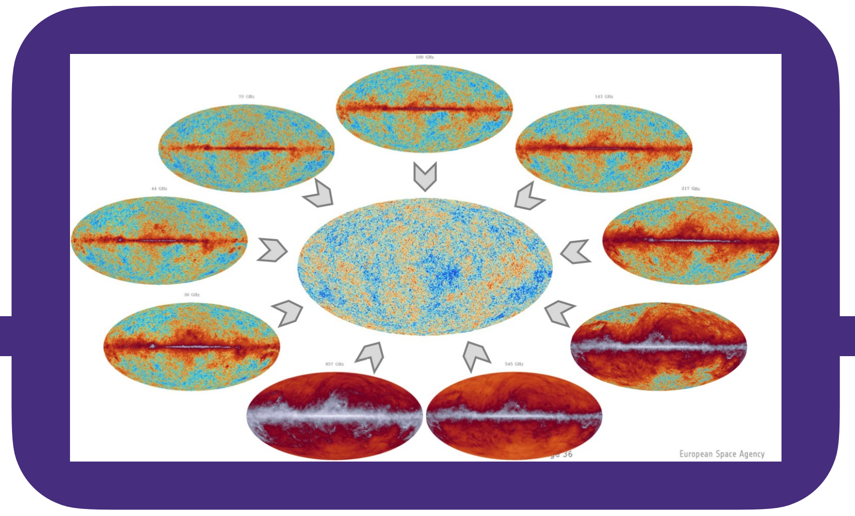
Noise

$$R = R(A^{d,c}, C_l^c, n_l^d)$$

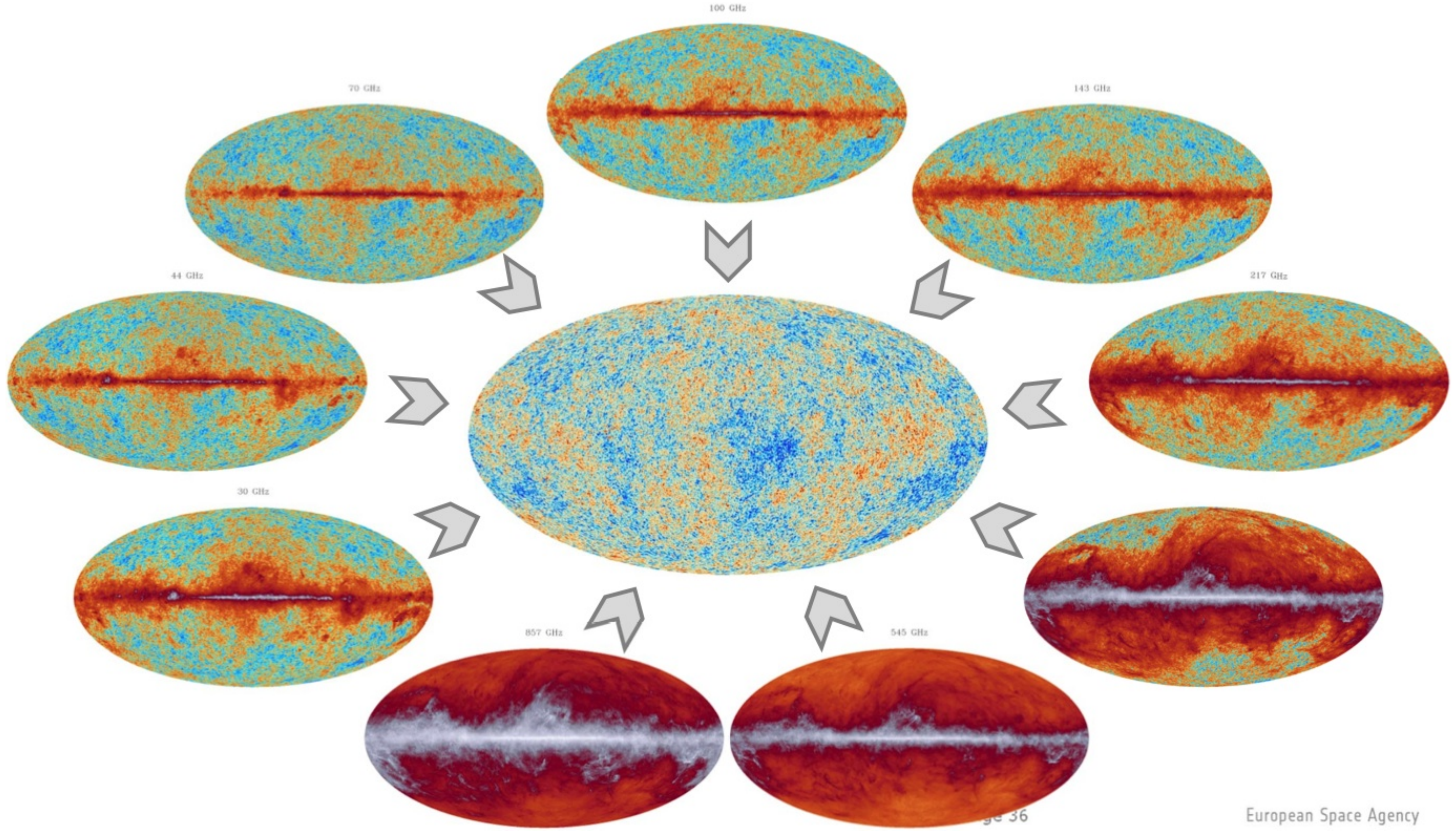
$$\vec{\theta} = \{A^{freq,comp}, C_l^{comp}, n^{freq}\}$$

the parameters describing the distribution

Component separation: SMICA



Observed sky at **different frequencies** → The **negative log-likelihood**

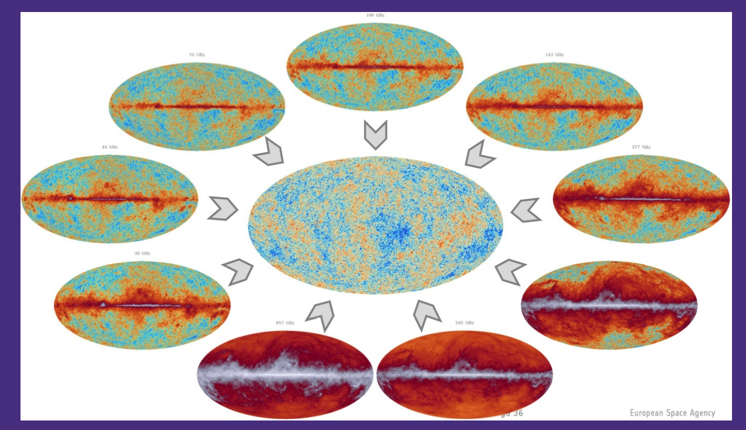


$$-2\log(\mathscr{L}(\vec{\theta})) = \sum_{l} (2l+1)\ D(\hat{R}_l, R_l(\vec{\theta})) + const$$

Where $D(R,\hat{R}) = tr(\hat{R}R^{-1}) + \log\det(\hat{R}R^{-1})$ is the so called **Kullback-Leibler divergence** between **the observed covariance matrix** and **the model covariance matrix**

$$D(R,\hat{R}) = tr(\hat{R}R^{-1}) + \log\det(\hat{R}R^{-1})$$

Component separation: SMICA

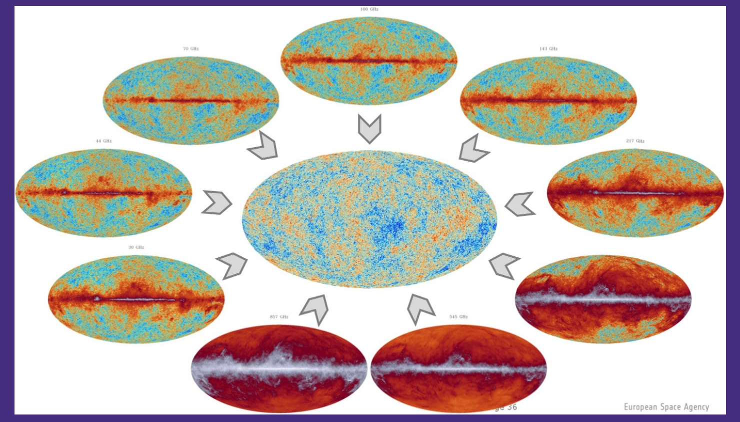


SMICA is a "**simple" conjugate gradient descent with** $\mathcal{O}(10^2)$ **parameters!**
We fit values for:

$$\vec{\theta} = \{A^{freq,comp}, C_l^{comp}, n^{freq}\}$$

But what if we wanted to **relax the assumption of Gaussianity**?

NG in component separation: $\mathscr{L}$

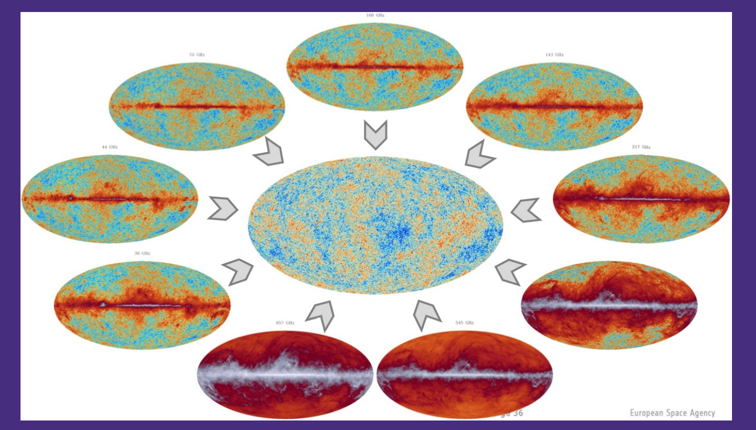


Gaussianity → **Weak non-Gaussianity**

New set of assumptions

1) **Isotropy**
2) **Weak non-Gaussianity**
3) **Statistical independence** between components
4) **Uncorrelated noise** across frequencies and in space

NG in component separation: $\mathscr{L}$



Gaussianity → **Weak non-Gaussianity**

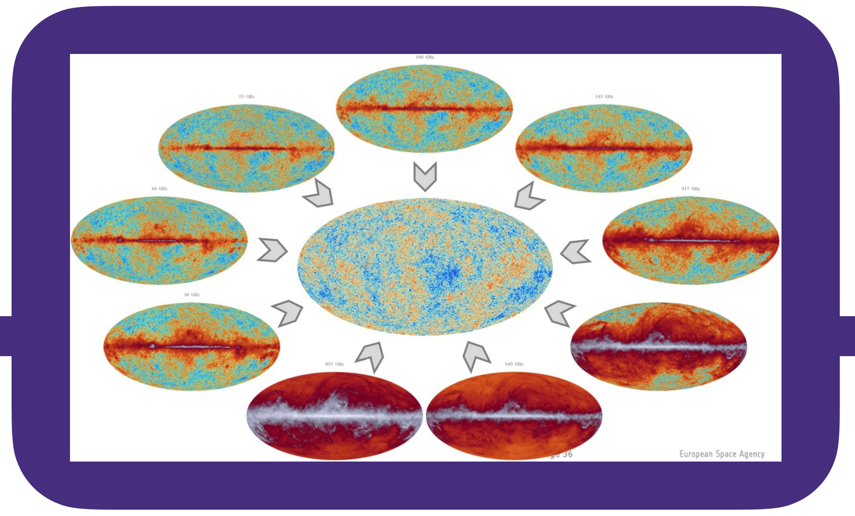
How does our distribution change from a Gaussian distribution?

Edgeworth multivariate expansion [4a-4d]

$$PDF(\mathrm{x}) = \frac{1}{\sqrt{2\pi \det(R)}} e^{-\frac{1}{2}x^T R^{-1} x} \longrightarrow \left(1 - \frac{1}{6}\langle x^i x^j x^k \rangle \frac{\partial^3}{\partial x^i \partial x^j \partial x^k}\right) \frac{1}{\sqrt{2\pi \det(R)}} e^{-\frac{1}{2}x^T R^{-1} x}$$

[4a] R. Juszkiewicz et al., arXiv:astro-ph/9308012
[4b] L. Amendola, https://doi.org/10.1093/mnras/283.3.983
[4c] A. Taylor et al., arXiv:astro-ph/0010014
[4d] N. Bartolo et al., arXiv:1107.4304

NG in component separation: $\mathscr{L}$



By computing the derivatives we obtain:

$$PDF(a^d_{lm}) = G(R) \longrightarrow PDF(a^d_{lm}) = G(R)\left(1 + \langle B, \hat{B} \rangle\right)$$

Mathematically well-defined **inner product** [5]

$$\langle B, \hat{B} \rangle = \sum_{l_i} B_{l_1 l_2 l_3} V^{-1}_{l_1 l_2 l_3} \hat{B}_{l_1 l_2 l_3}$$

$$B^{d_1 d_2 d_3}_{l_1 l_2 l_3} = A^{d_1,c} A^{d_2,c} A^{d_3,c} B^c_{l_1 l_2 l_3}$$

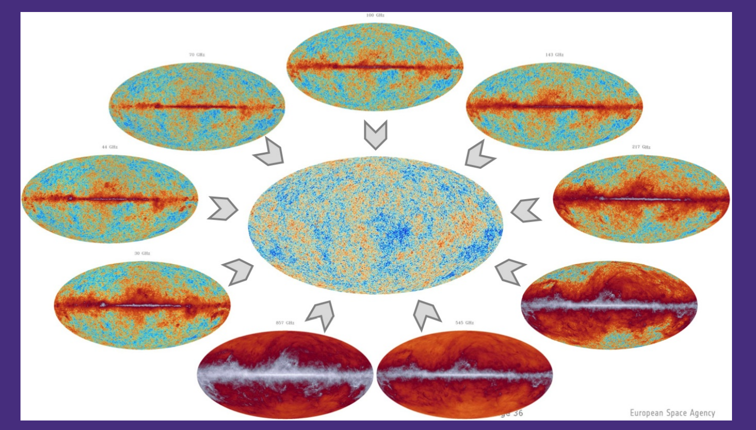
Model bispectrum

$$\hat{B}^{d_1 d_2 d_3}_{l_1 l_2 l_3}$$

Observed bispectrum

[5] M. Bucher, B. Racine and B. van Tent, arXiv:1509.08107

NG in component separation: $\mathscr{L}$



Now **the negative log-likelihood of the data given this model** is

$$-2\log(\mathscr{L}(\vec{\theta})) = -2\log(\mathscr{L}_G(R,\hat{R})) \quad -2\log(\mathscr{L}_{NG}(R,B,\hat{B}))$$

We could perform a conjugate gradient descent over a new set of parameters:

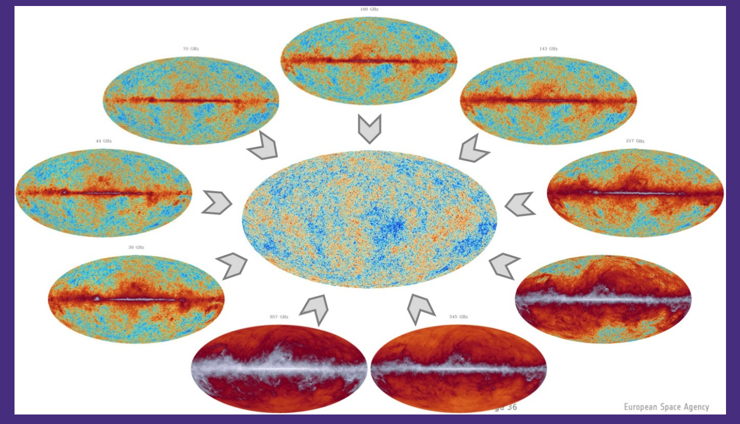
$$\vec{\theta} = \{A^{freq,comp}, C_l^{comp}, n^{freq}, B_{l_1 l_2 l_3}^{comp}\}$$

Due to obvious **computational limits** we can **choose a template for the bispectrum of the components:**

$$B_{l_1 l_2 l_3}^{comp}$$

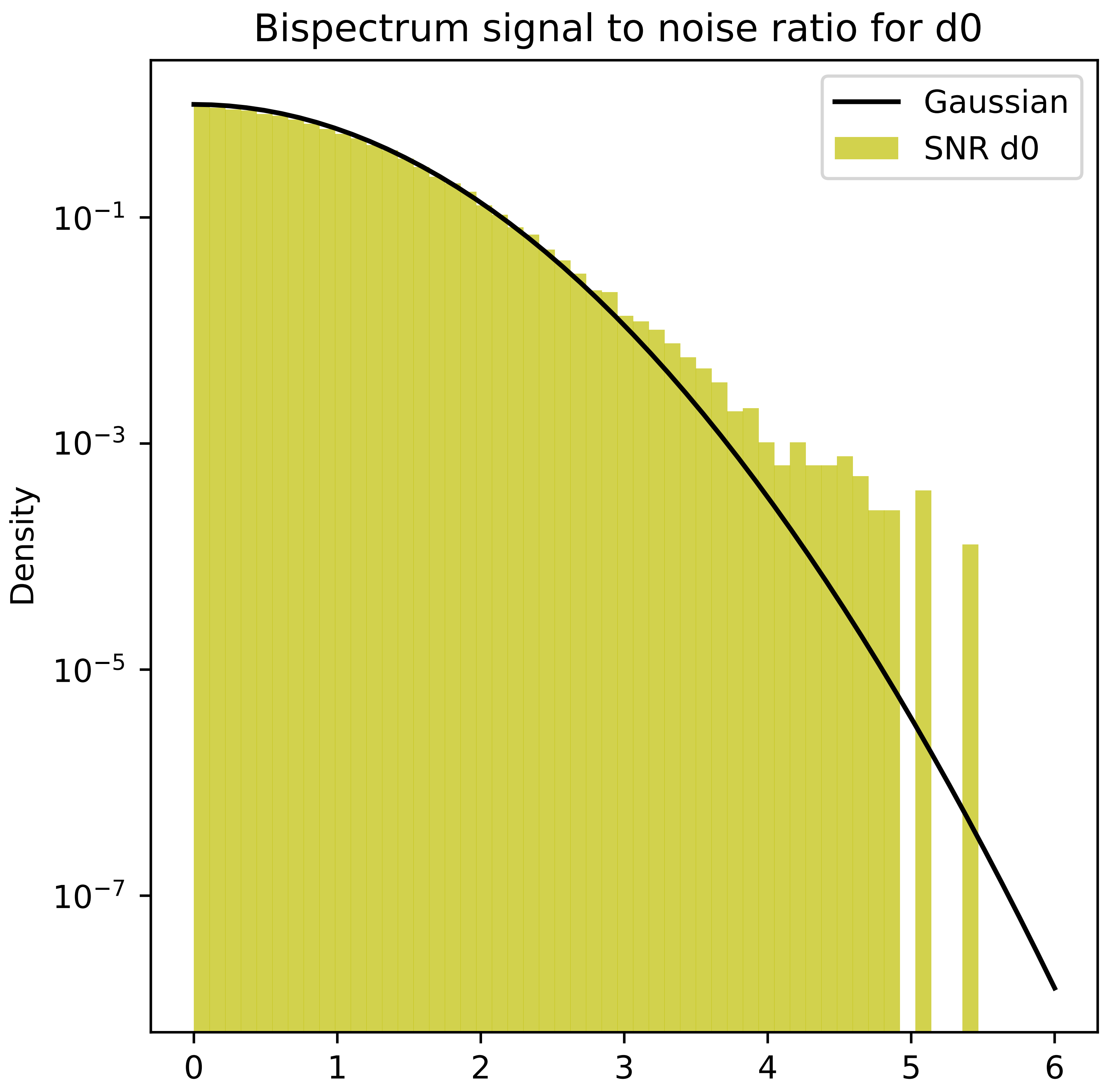
Which implies having **some information about the foregrounds present**.

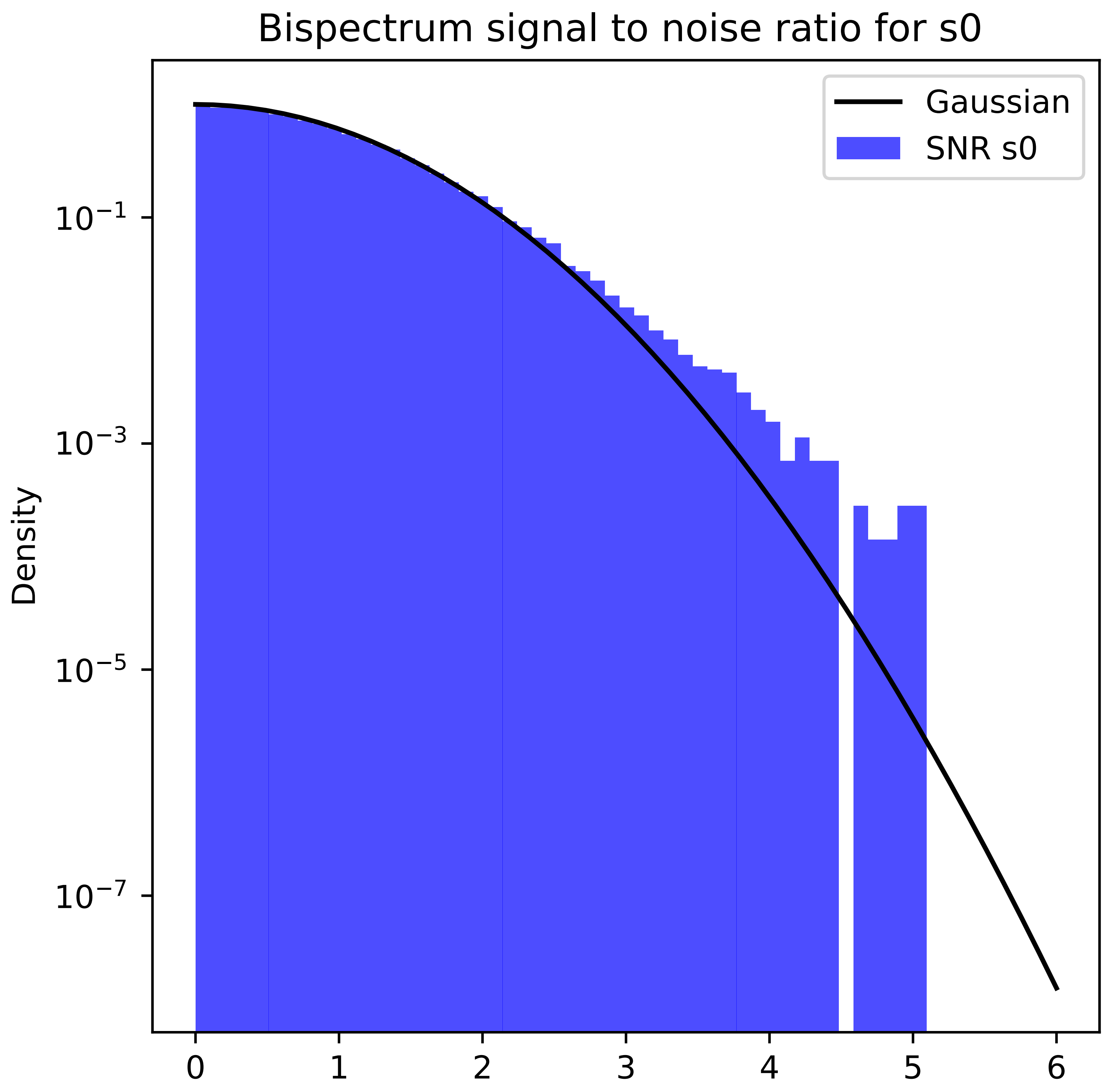
NG in component separation: Data



We focused on **B-mode polarization** for the LiteBIRD experiment based on [6]:

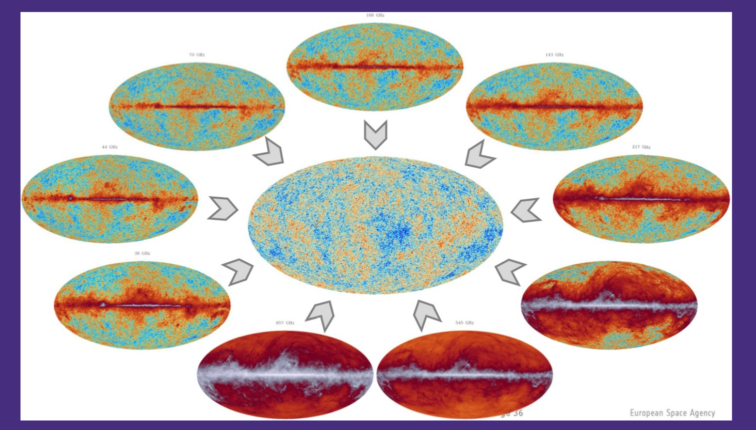
- **15 frequencies**
- **CMB with r=0**
- **Dust d0, Synchrotron s0** [starting model]
- **Gaussian instrumental noise**
- **Mask** $f_{\mathbf{sky}} \sim 60\,\%$
- $l_{\mathbf{max}} = 100$ [low nside to limit complexity]





[6] LiteBIRD collaboration, arXiv:2507.22618

NG in component separation: Data



Based on [7] we have realized a physically motivated **template for the bispectrum** of dust and synchrotron

$$B^{d}_{l_1 l_2 l_3} \qquad B^{s}_{l_1 l_2 l_3}$$

As a linear combination of **3 “basic” templates inspired by primordial non-gaussianity**

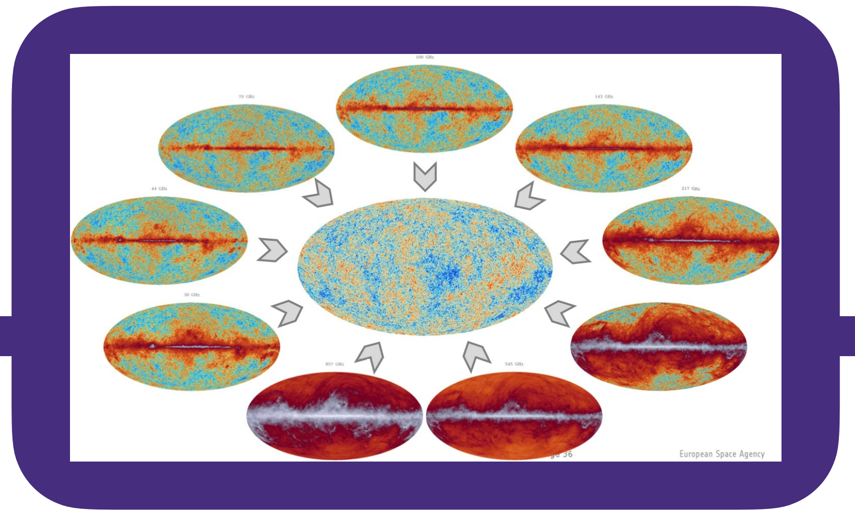
$$B^{\mathrm{loc}}_{l_1 l_2 l_3} = C_{l_1} C_{l_2} + 2 \text{ perm.} \qquad B^{\mathrm{eq}}_{l_1 l_2 l_3} = \left[C_{l_1} C_{l_2} + 2 \text{ perm.} \right] + 2 C_{l_1}^{2/3} C_{l_2}^{2/3} C_{l_2}^{2/3} - \left[C_{l_1} C_{l_2}^{2/3} C_{l_3}^{1/3} + 5 \text{ perm.} \right]$$

$$B^{\mathrm{ort}}_{l_1 l_2 l_3} = 3 \left[C_{l_1} C_{l_2} + 2 \text{ perm.} \right] + 8 C_{l_1}^{2/3} C_{l_2}^{2/3} C_{l_2}^{2/3} - 3 \left[C_{l_1} C_{l_2}^{2/3} C_{l_3}^{1/3} + 5 \text{ perm.} \right]$$

With computed **weights given by the inner product** $\langle B^{I}, B^{J} \rangle$

[7] G. Jung, B. Racine and B. van Tent, arXiv:1810.01727

NG in component separation: Results



Modifying the algorithm **has not shown any improvement** with respect to the standard gaussian smica

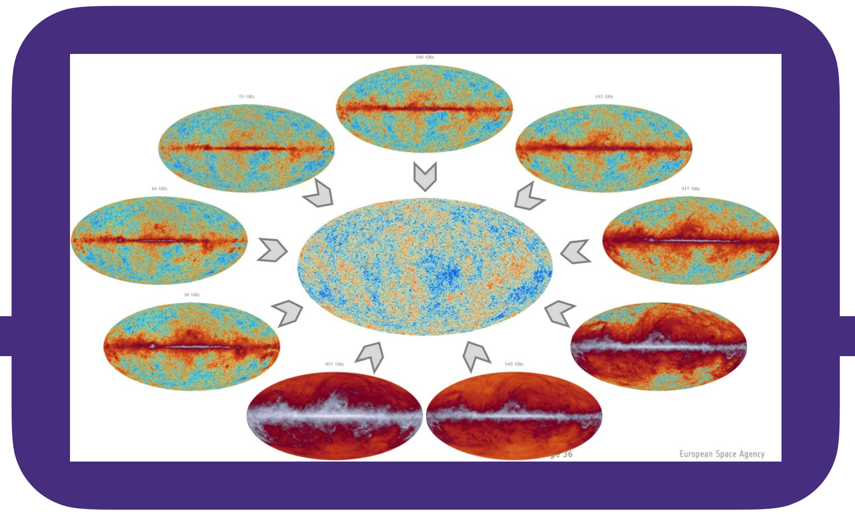
The explanation can be found in the term we have added to the log-likelihood:

$$-\log(\mathscr{L}_{NG}) = -\log(1 + \langle B, \hat{B} \rangle)$$

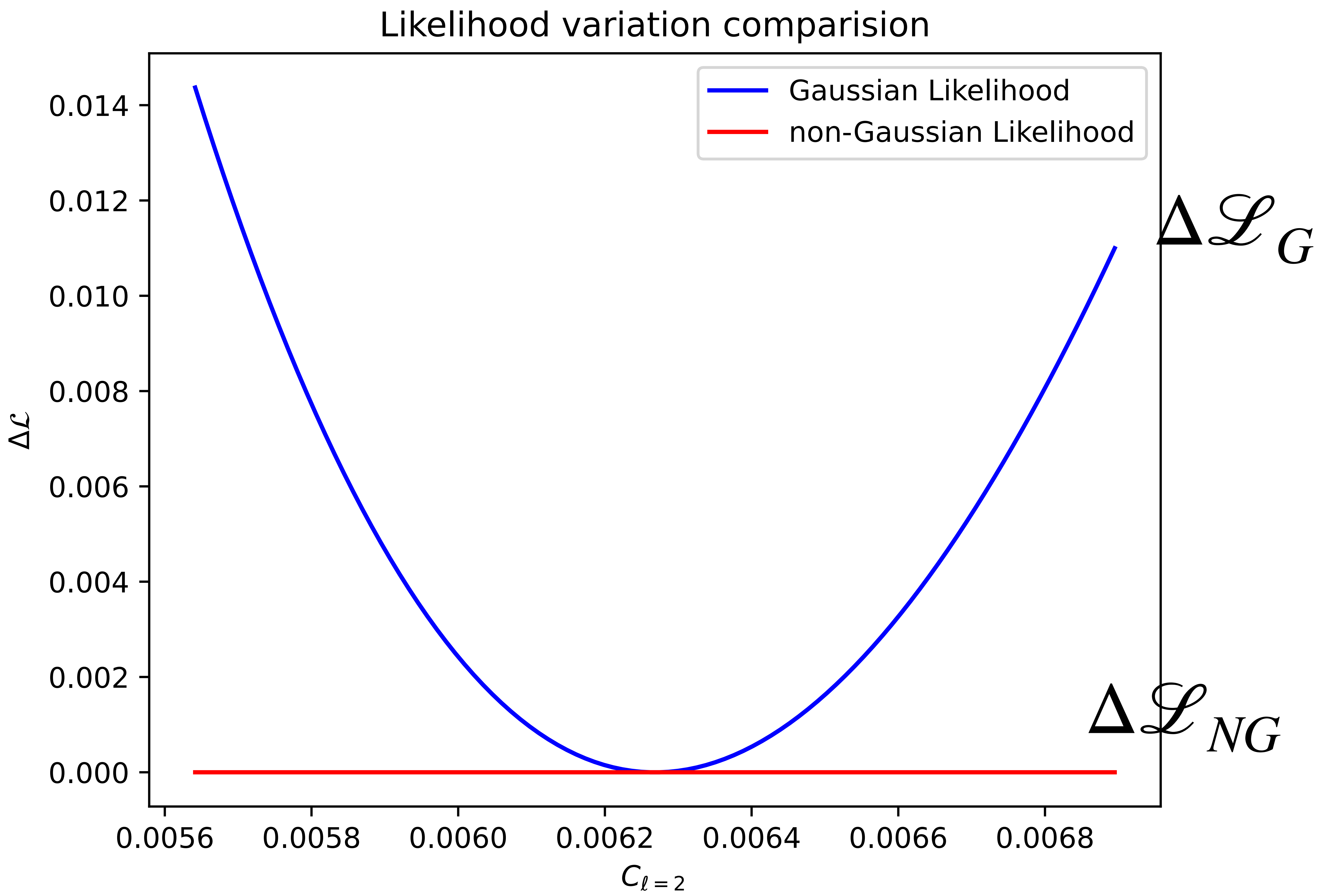
With respect to the original gaussian likelihood

$$-\log(\mathscr{L}_{G}) = \frac{1}{2} \sum_{l} (2l+1)\ D(\hat{R}_l, R_l(\vec{\theta})) + const$$

NG in component separation: Results

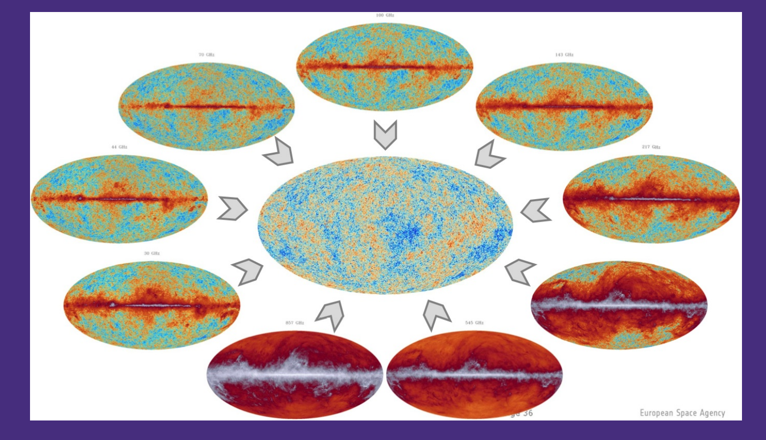


Likelihood variation **around the minimum** value found by Gaussian SMICA for **one of the parameters**



Variation of $\sim 1\%$

NG in component separation: Results



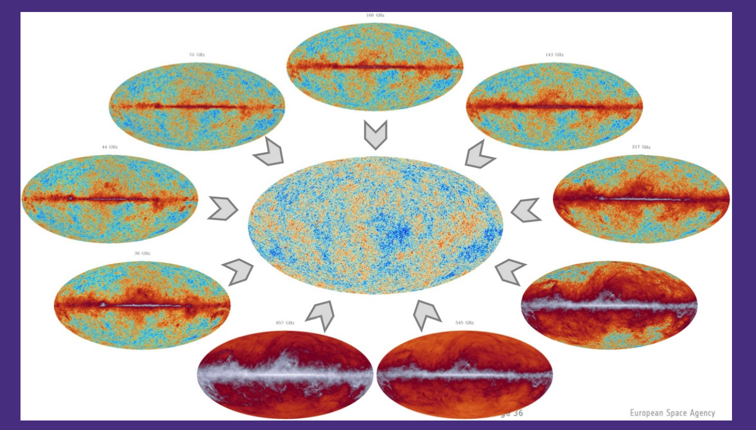
But this numerical cause has possible **physical reasons** behind it:

- The **amount of information** we are adding might be essentially **too small** with respect to the standard deviation of SMICA: possibly because **the mask** highly reduced the foregrounds' bispectrum signal-to-noise ratio, **the bispectrum model** was not precise enough, $\ell_{\max}$ was too low…

- **The gaussian information might be the only necessary** term needed in order to distinguish between the CMB and the other foregrounds

- The **bispectrum might not capture the complexity of the foregrounds** well enough with respect to **other higher-order correlators**, hence our likelihood expansion does not describe well the data

Multi-detector multi-component Bispectrum estimator



Let’s look at the glass half full!

Bispectrum does **not add information** → **independent powerspectrum and bispectrum estimation directly in frequency space!**

Gaussian SMICA output $\vec{\theta} = \{A^{freq,comp}, C_l^{comp}, n^{freq}\} \rightarrow R_l^{dd'}$

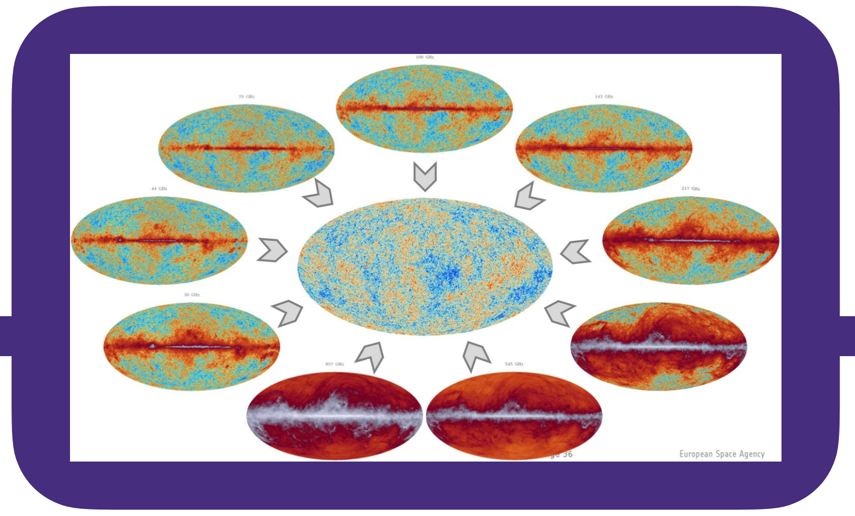
A likelihood describing the bispectrum, generalized to frequency space is [8]

$$\mathscr{L}(B\,|\,\hat{B}) \propto \exp\left(-0.5(\hat{B}-B)^{d_1d_2d_3}_{l_1l_2l_3}(Var(B)^{-1})^{d_1d_2d_3,d_1'd_2'd_3'}_{l_1l_2l_3}(\hat{B}-B)^{d_1'd_2'd_3'}_{l_1l_2l_3}\right)$$

Where $$Var(B)^{d_1d_2d_3,d_1'd_2'd_3'}_{l_1l_2l_3} = 6R^{d_1d_1'}_{l_1}R^{d_2d_2'}_{l_2}R^{d_3d_3'}_{l_3}$$

[8] W. Sohn et al., arXiv:2305.14646

Multi-detector multi-component Bispectrum estimator



Before we adopted a **model** to describe the bispectrum, now we want to **estimate it from frequency maps directly**:

We still have a problem with the amount of parameters we try to estimate!

It is pointless trying to estimate $\mathcal{O}(10^5)$ parameters! We **bin it** [5,9]!

12 bins $\sim$ 200 parameters

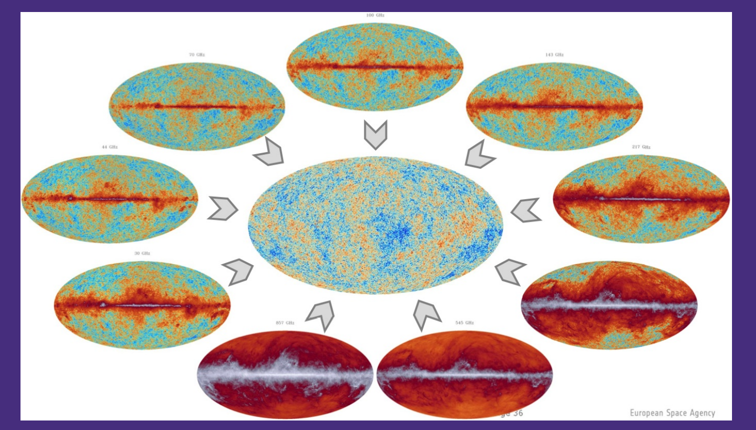
We divide the bispectrum into its **even** $l_1 + l_2 + l_3 = 2k, \quad k \in \mathbb{N}$ and odd $l_1 + l_2 + l_3 = 2k+1, \quad k \in \mathbb{N}$ parts

With **a binned bispectrum likelihood**

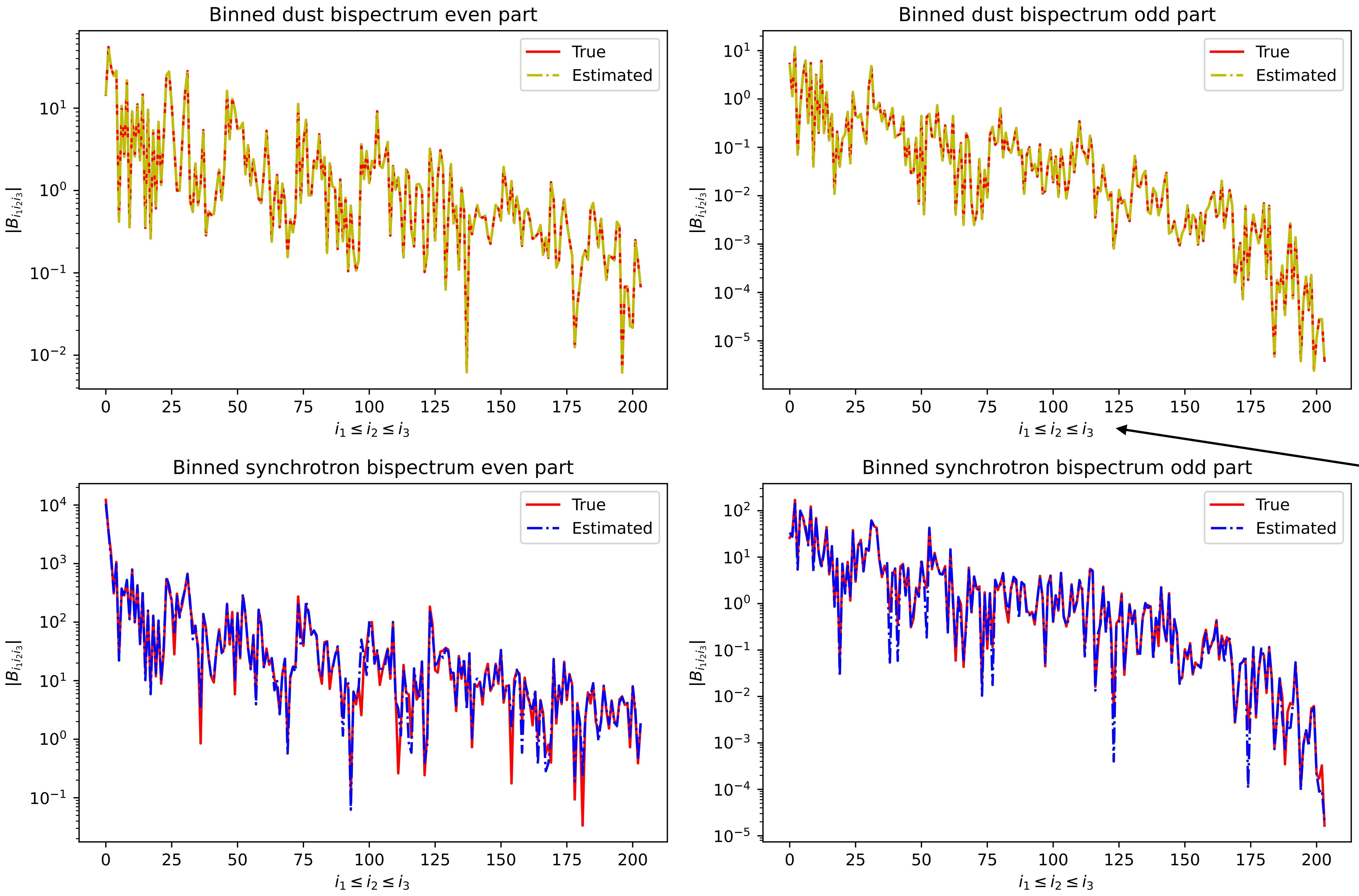
$$\mathcal{L}(B\,|\,\hat{B}) \propto \exp\Big(-0.5(\hat{B} - B)^{d_1d_2d_3}_{i_1i_2i_3} (Var(B)^{-1})^{d_1d_2d_3,d_1'd_2'd_3'}_{i_1i_2i_3} (\hat{B} - B)^{d_1'd_2'd_3'}_{i_1i_2i_3} \Big)$$

[5] M. Bucher, B. Racine and B. van Tent, arXiv:1509.08107
[9] W. Coulton and D. N. Spergel, arXiv:1901.04515

Multi-detector multi-component Bispectrum estimator

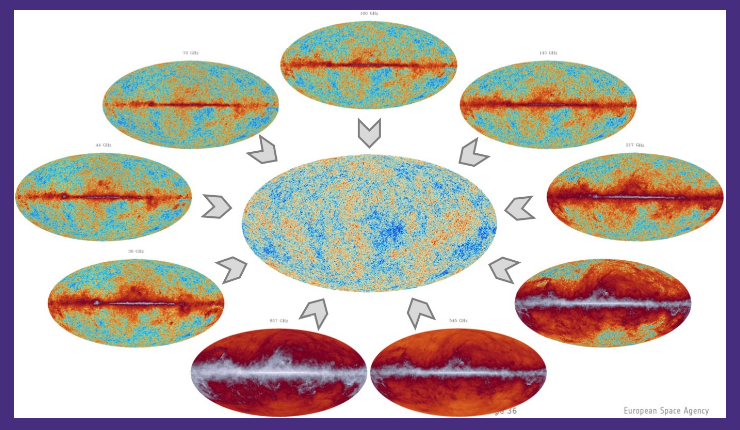


We are capable of recovering the correct bispectrum for the two foregrounds both in **B polarization**

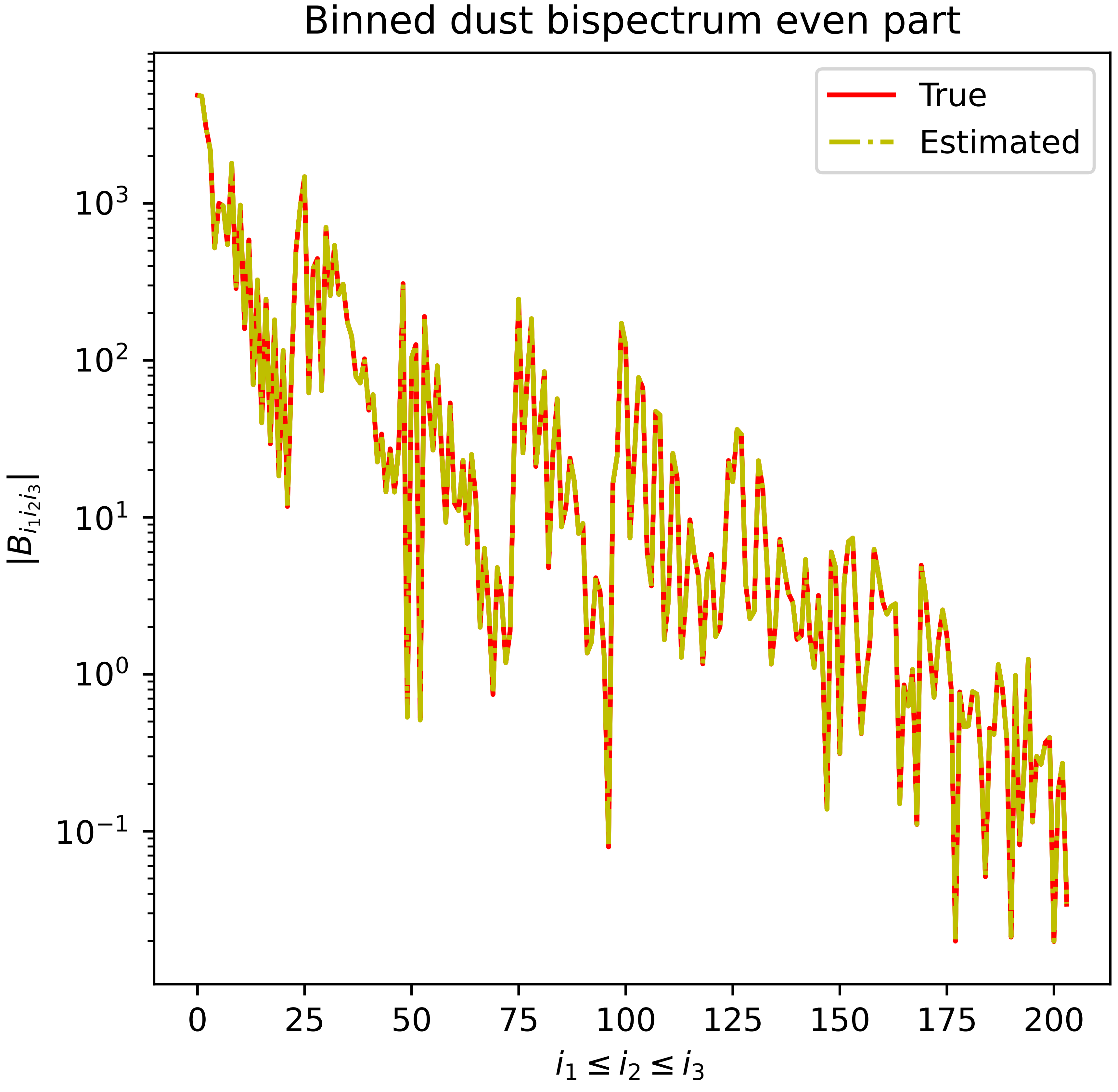


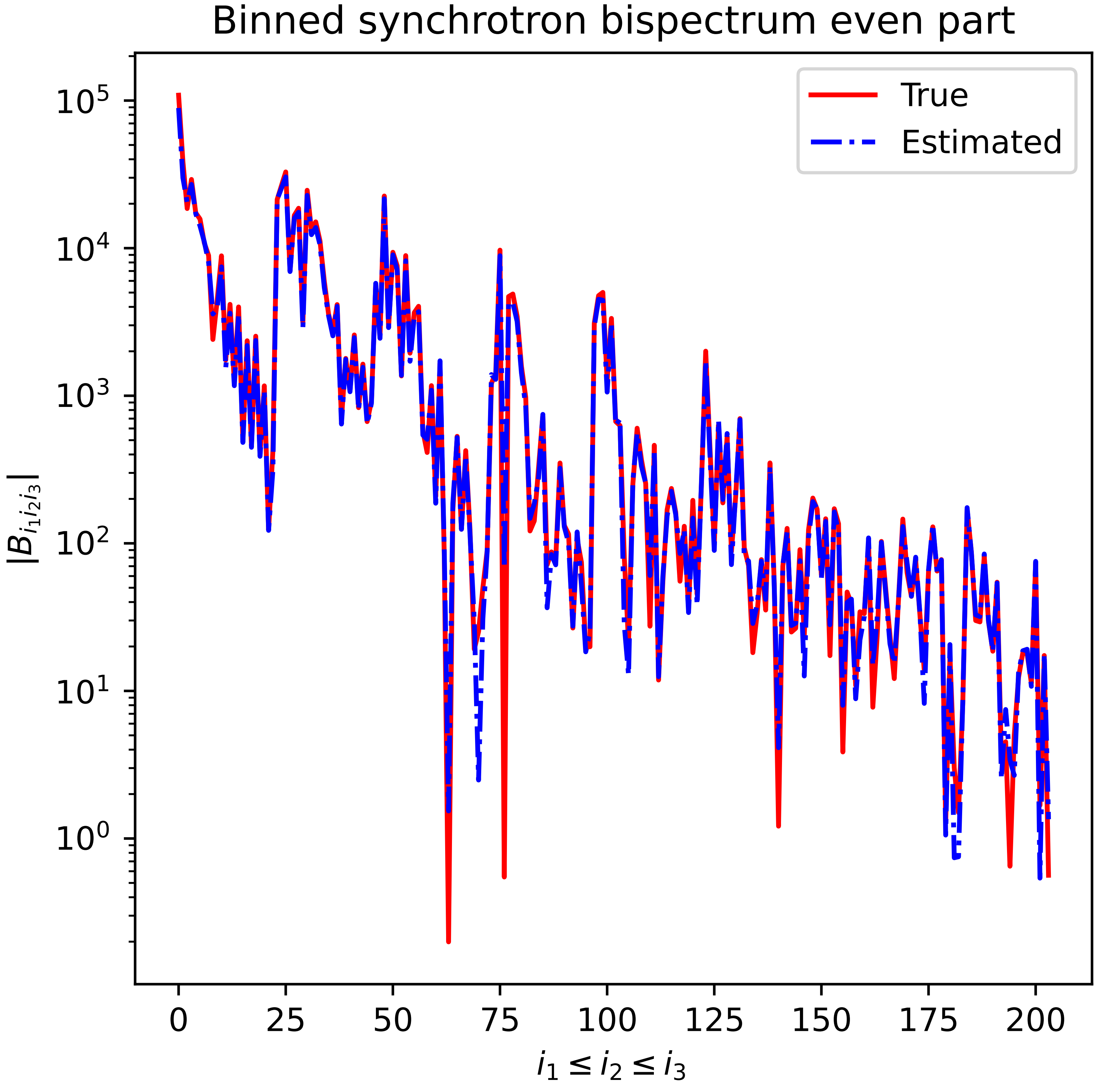
Combined index
$(i_1 \leq i_2 \leq i_3)$

Multi-detector multi-component Bispectrum estimator

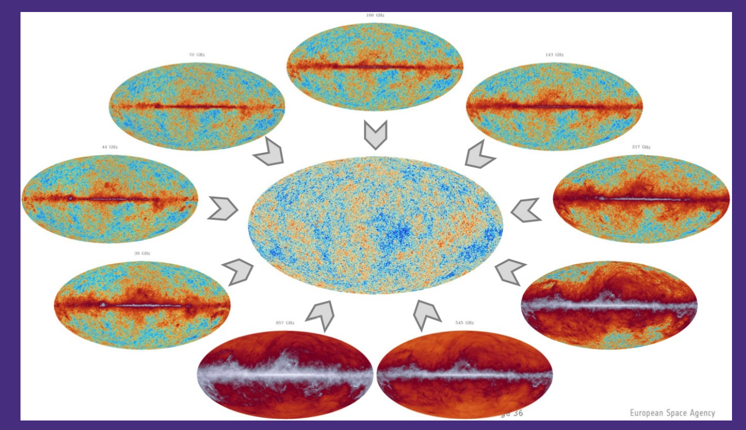


And in **E polarization,** where we have restricted our analysis to the even part, **in order to estimate The local shape** f_{NL}

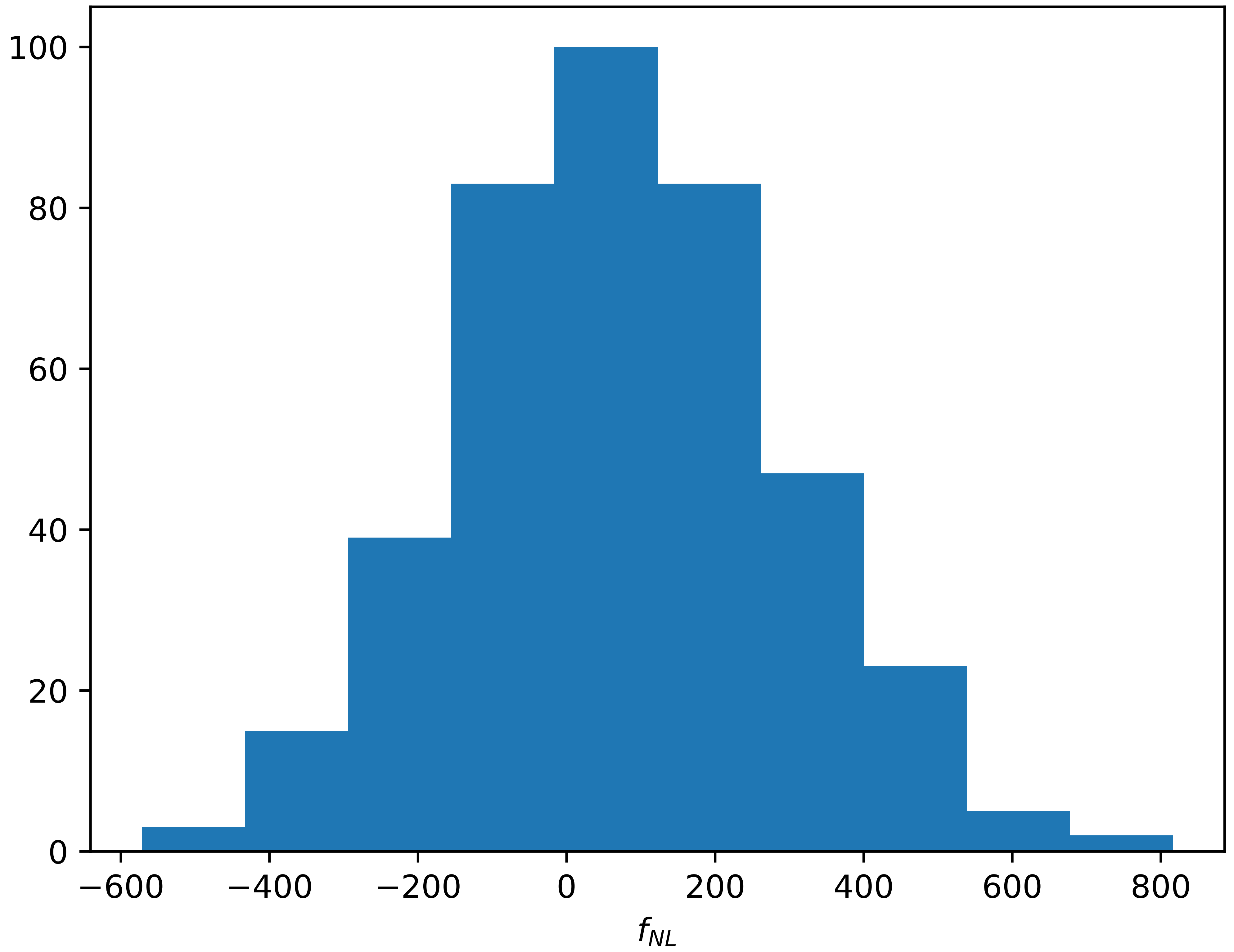




Multi-detector multi-component Bispectrum estimator



And we have obtained a value for $\sigma(f_{NL}^{\mathbf{loc}}) \sim 200$ **coherent with the value obtained with the usual binned bispectrum estimation after the component separation** step with **E polarization only and** $\ell_{\mathbf{max}} = 100$



Conclusions

Thank you for your attention!

1) Non-Gaussian statistics in component separation

↓

Developed a **formalism** able to include **higher order statistics into component separation**

↓

The **bispectrum** does **not** have **enough constraining power** to improve the estimation

Conclusions

Thank you for your attention!

1) Non-Gaussian statistics in component separation

↓

Developed a **formalism** able to include **higher order statistics into component separation**

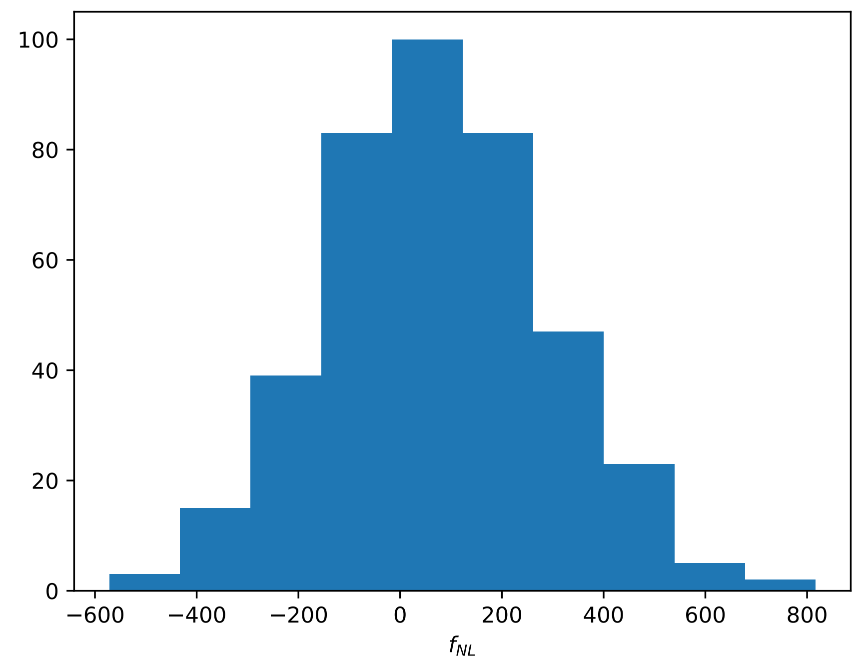
↓

The **bispectrum** does **not** have **enough constraining power** to improve the estimation

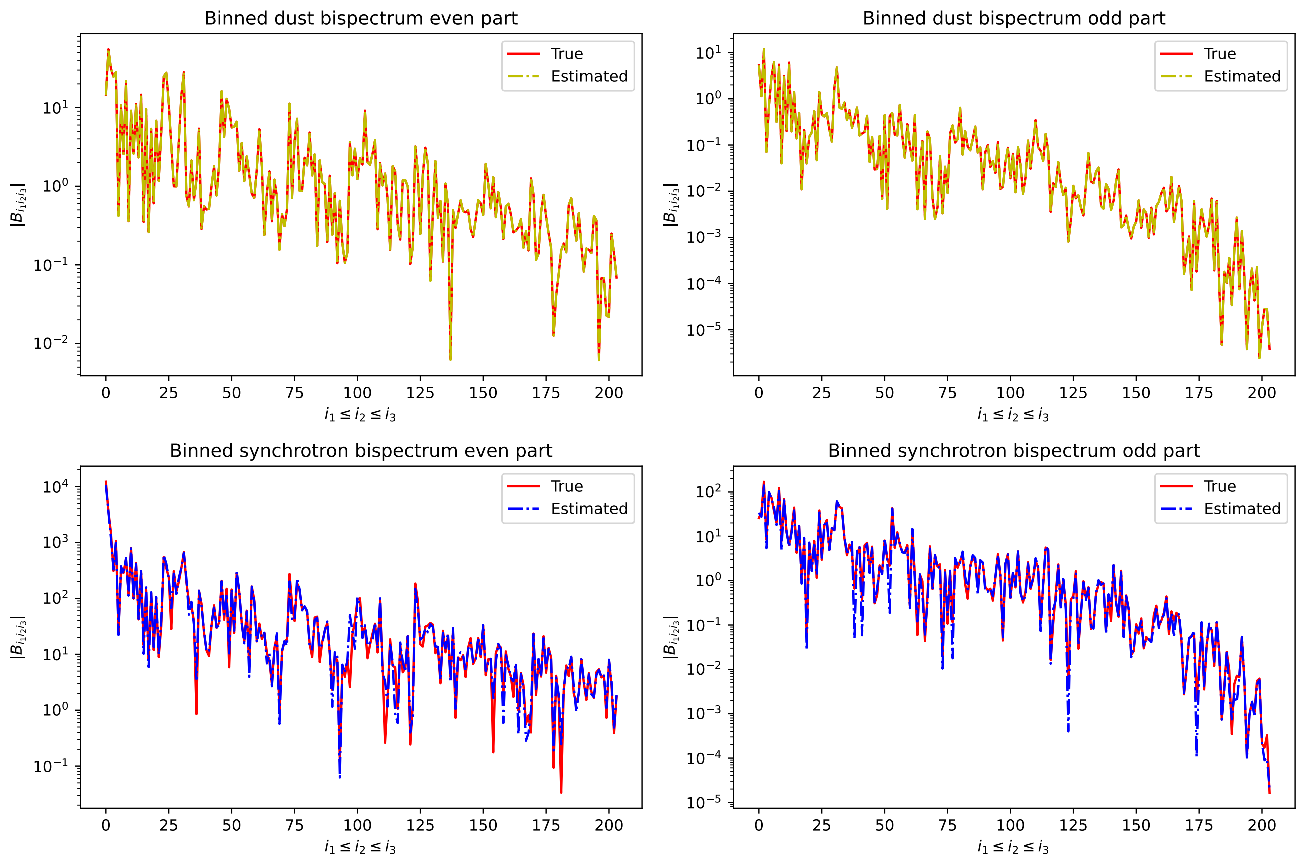
2) Multi-detector multi-component bispectrum estimation

↓

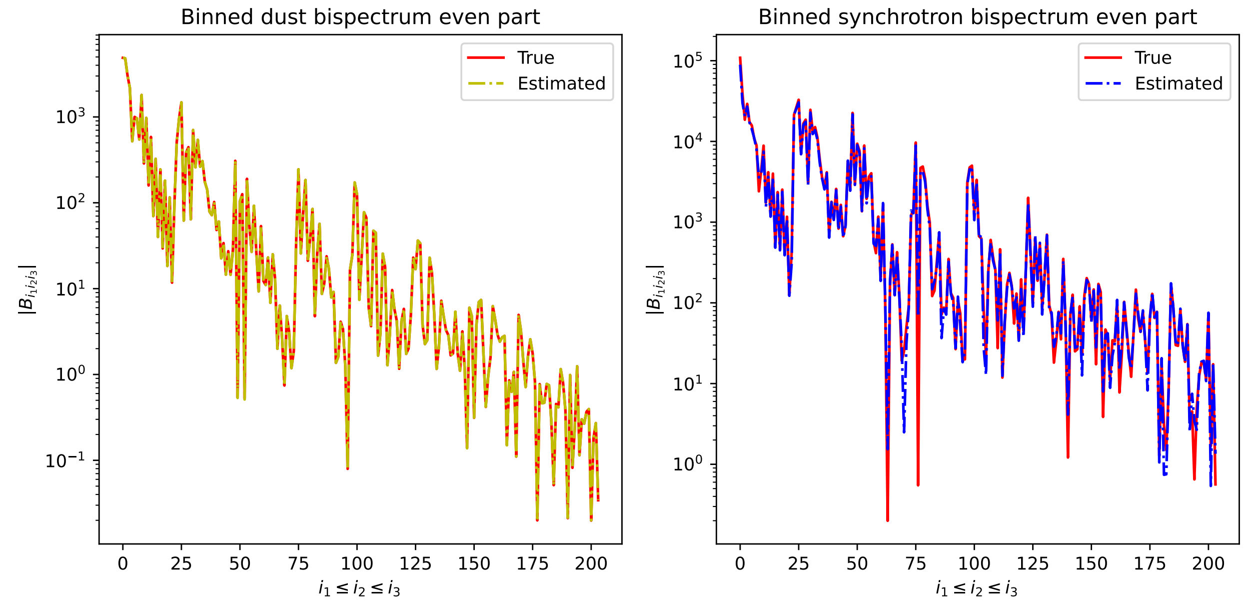
Built a **new estimator** in order to bring **non-Gaussian analysis a step before** usual bispectrum estimators



BBB



EEE



Michele Citran

Back up

In d0, one assumes **fixed spectral parameters** across the sky: in particular, a fixed dust spectral index (β) and a fixed dust temperature for the modified blackbody (MBB) emission law

Specifically, in PySM 3, d0 uses a fixed spectral index = 1.54 and fixed temperature = 20 K (i.e. no spatial variation) for the dust MBB scaling

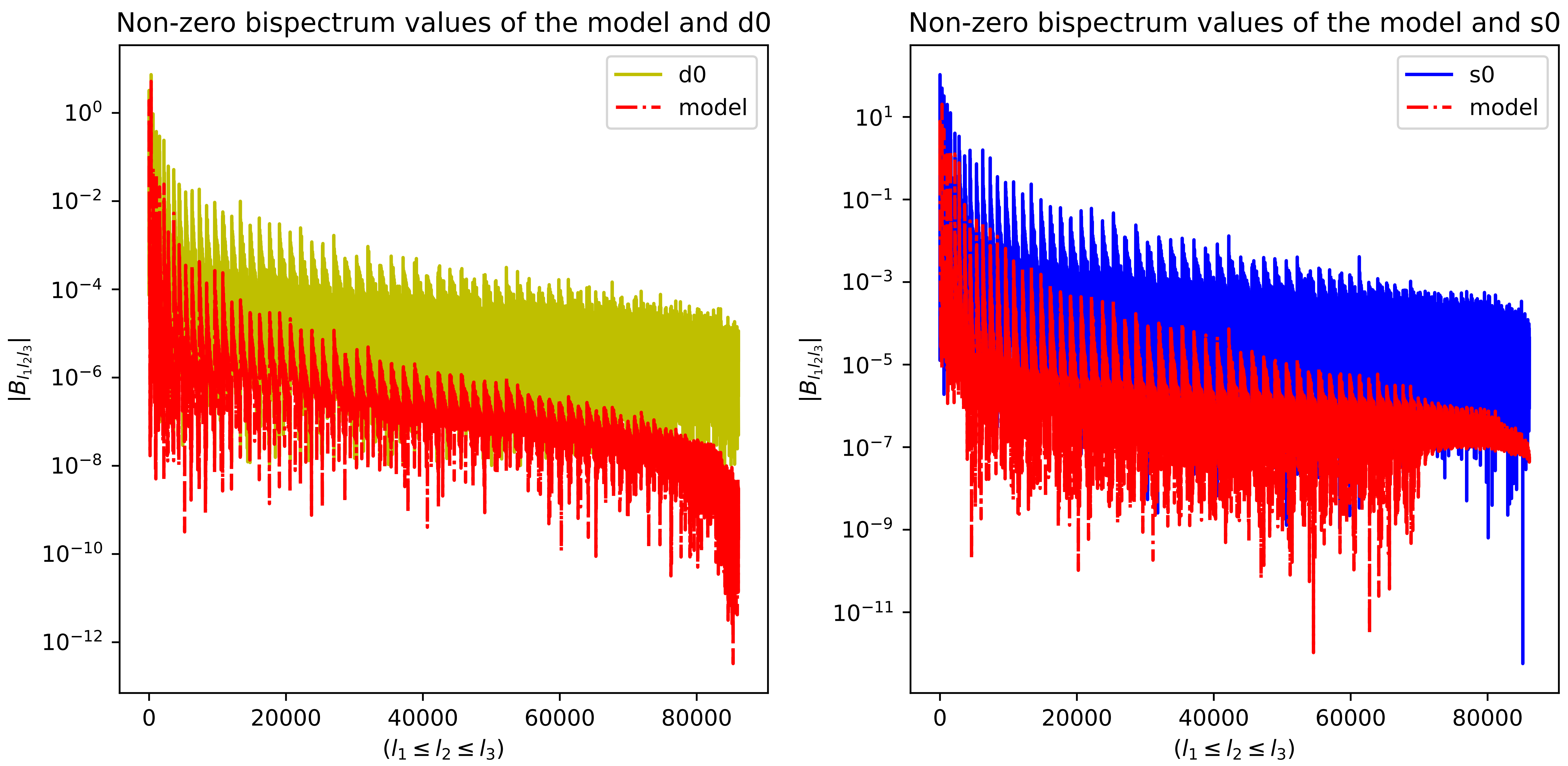
$$I(\nu, \hat{n}) = I(\nu_0, \hat{n}) \left(\frac{\nu}{\nu_0}\right)^{\beta_d} \frac{B(\nu, T_d)}{B(\nu_0, T_d)}$$

s0 uses a **constant spectral index** (i.e. no spatial variation) for the synchrotron power-law scaling

S0 -> Beta = -3

Haslam 408 MHz for intensity
WMAP/Planck for polarization)

$$I(\nu, \hat{n}) = I(\nu_0, \hat{n}) \left(\frac{\nu}{\nu_0}\right)^{\beta_s}$$


Non-zero bispectrum values of the model and d0
d0
model
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
$|B_{l_1 l_2 l_3}|$
0
20000
40000
60000
80000
$(l_1 \leq l_2 \leq l_3)$
Non-zero bispectrum values of the model and s0
s0
model
10^{1}
10^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-11}
$|B_{l_1 l_2 l_3}|$
0
20000
40000
60000
80000
$(l_1 \leq l_2 \leq l_3)$