Foreground Cleaning on Filtered Data in Simons Observatory

Challenges at the CMB Frontier

Baptiste Jost (IPMU)
Benjamin Beringue (APC), Magdy Morshed (Ferrara U), Amalia Villarrubia-Aguilar (APC), Sherry Song (IPMU), Simon Biquard (Manchester U), Artem Basyrov (APC), Josquin Errard (APC)

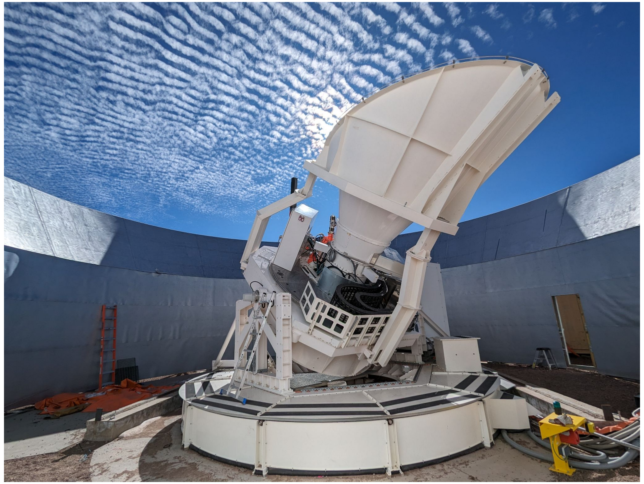
Looking For Primordial B-modes with Simons Observatory

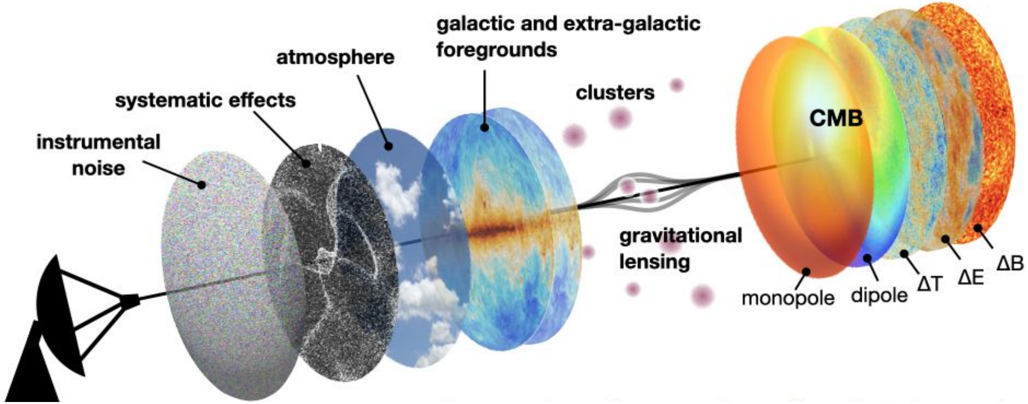
The SO **Small Aperture Telescopes** (SAT) are after large scales primordial B-modes **(see Yoshinori Sueno's talk)**

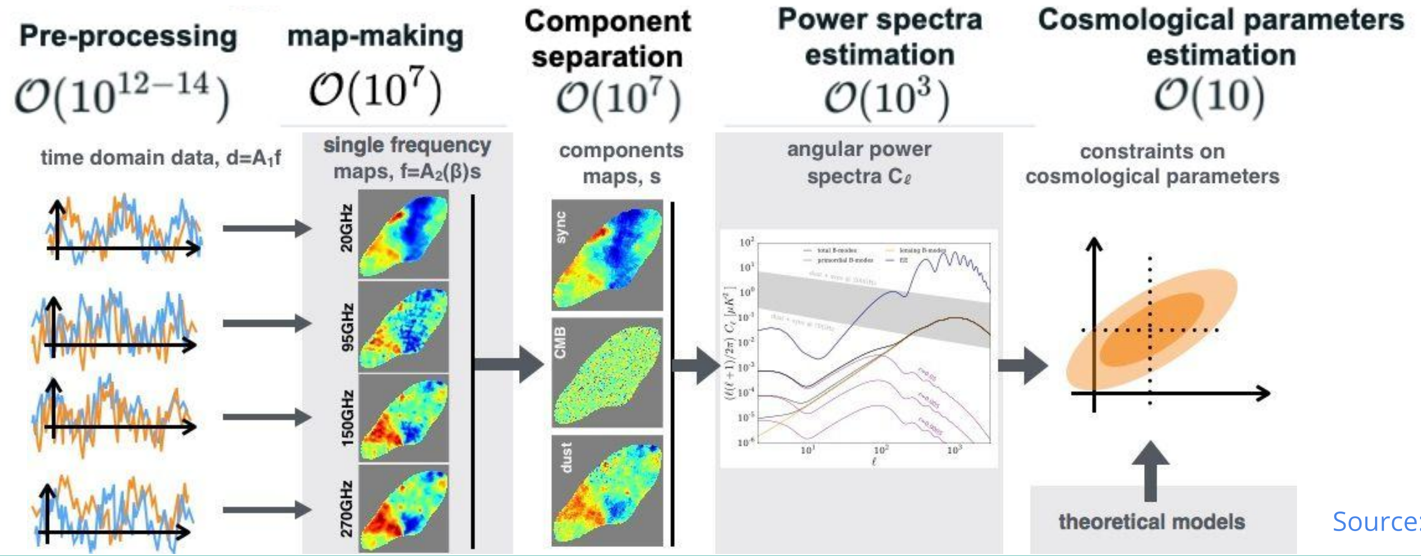
Three main pipelines are being developed **(see Alessandro Carones' talk)**:

- Cl-based **(parametric)**: Existing code: BBPower, Ongoing: SOOPERCOOL
- Map-based **(parametric)**: Existing code: FGBuster/XForecast, Ongoing: MEGATOP
- NILC **(non-parametric)**: BROOM.

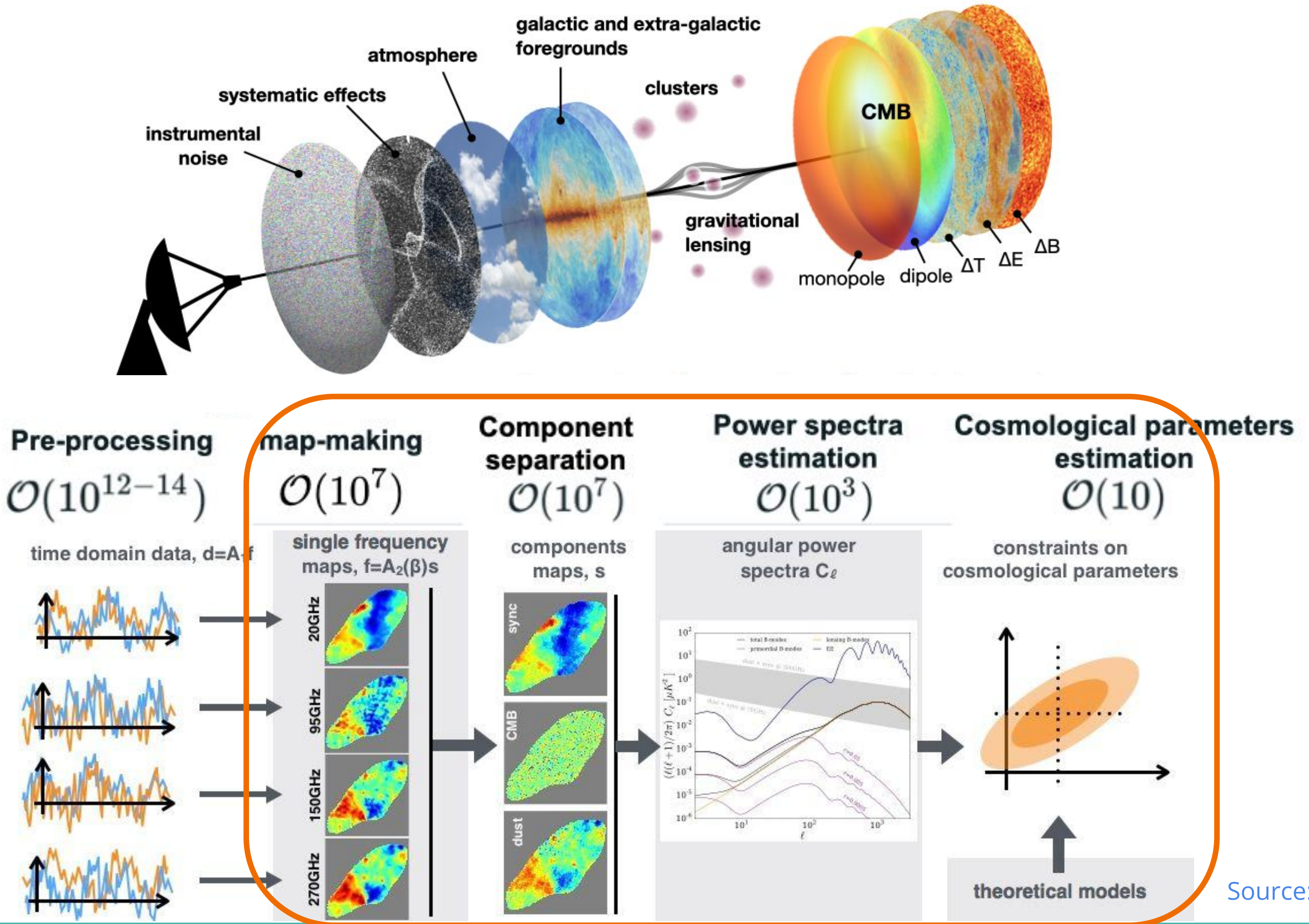
Allows for robustness against different systematic effects



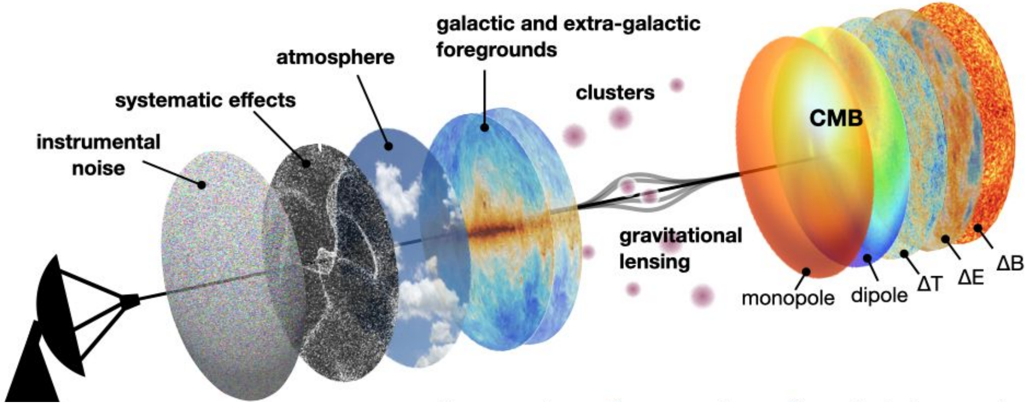


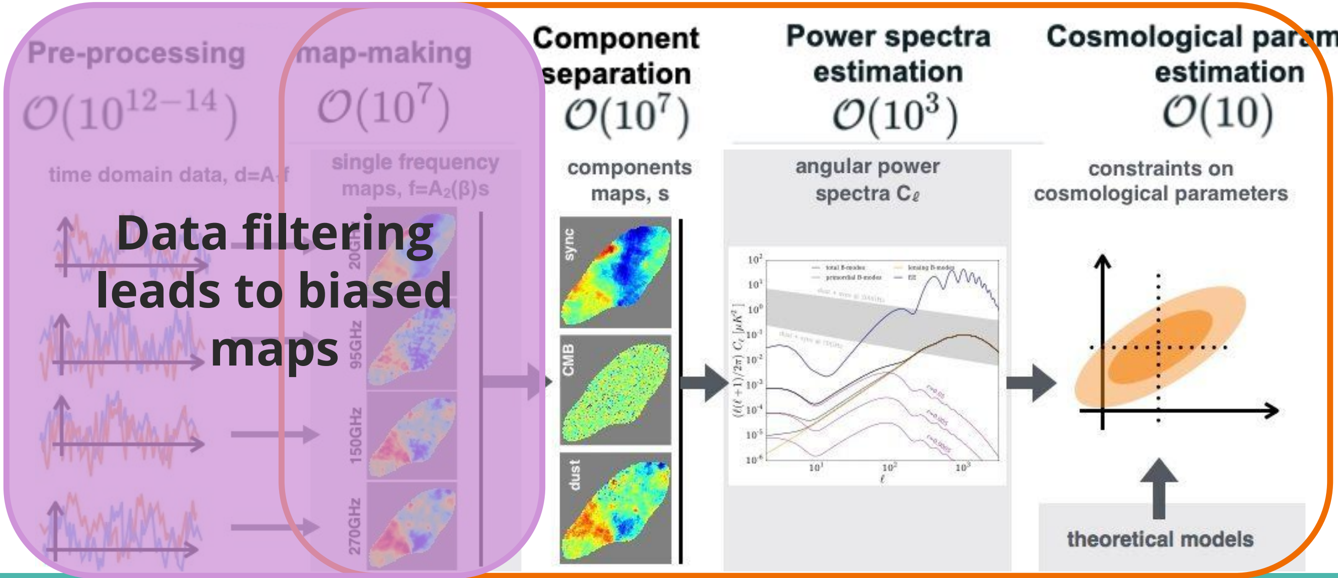


Source: J. Errard



Source: J. Errard





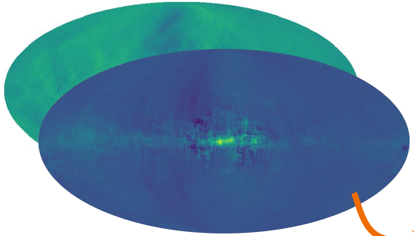
Data filtering leads to biased maps

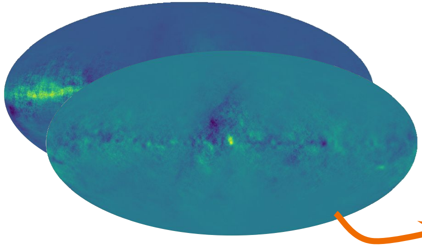
Source: J. Errard

The Standard Data Model

$$\mathbf{d}_p = \mathbf{A}(\beta)\mathbf{s}_p + \mathbf{n}_p$$

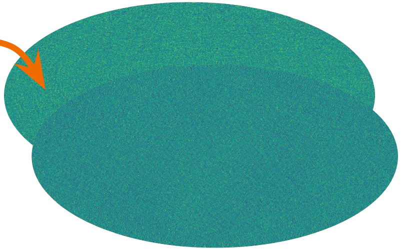
noise





$$\begin{pmatrix} Q_1 \\ U_1 \\ \vdots \\ Q_n \\ U_n \end{pmatrix}$$

$$\begin{pmatrix} Q^{CMB} \\ U^{CMB} \\ Q^d \\ U^d \\ Q^s \\ U^s \end{pmatrix}$$



Input frequency maps

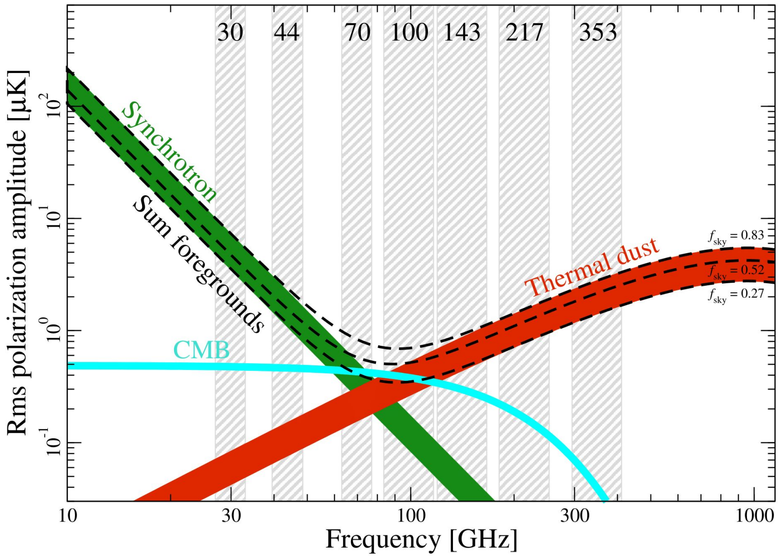
Component maps

The Mixing Matrix

$$\mathbf{d}_p = \mathbf{A}(\beta)\mathbf{s}_p + \mathbf{n}_p$$

$$\mathbf{A}(\{\beta_{\mathrm{fg}}\}) = \begin{pmatrix} 1 & A_1^d & A_1^s \\ \vdots & \vdots & \vdots \\ 1 & A_n^d & A_n^s \end{pmatrix}$$

CMB **Dust** **T_d , β_d** **Synchrotron** **β_s**



Source: Planck Collaboration 2018

Matrix encodes modelled foreground emission laws:

- **modified black-body** for dust
- **power law** for synchrotron

Variation across the sky is possible (Errard et al. 2019)

The Spectral Likelihood

We use the ***spectral likelihood*** (Stompor et al 2008) to estimate those parameters and remove foregrounds contamination:

fgbuster

$$\mathbf{d}_p = \mathbf{A}(\beta)\mathbf{s}_p + \mathbf{n}_p$$

$$-2\ln\mathcal{L}_{\rm spec}(\beta) = \mathrm{cst} - \left(\mathbf{A}^t\mathbf{N}^{-1}\mathbf{d}\right)^t \left(\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}\right)^{-1} \left(\mathbf{A}^t\mathbf{N}^{-1}\mathbf{d}\right)$$

$$\hat{\mathbf{s}} = \left(\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}\right)^{-1} \mathbf{A}^t\mathbf{N}^{-1}\mathbf{d}$$

Can deal with spatial variation, both in **noise properties** and **foreground parameters**

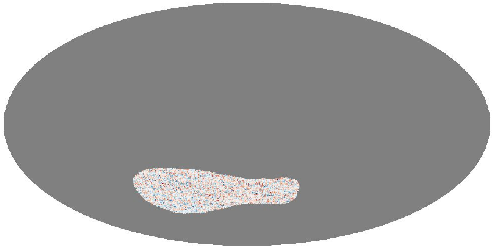
Outputs **CMB** and **Foregrounds maps**, allowing for residual foreground marginalization in the cosmological likelihood

From Component Maps to Cosmology

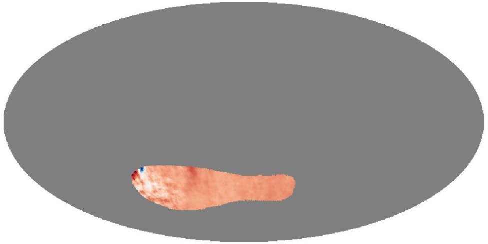
COMPONENT MAPS

$$\hat{\mathbf{s}} = \left(\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^t\mathbf{N}^{-1}\mathbf{d}$$

CMB map



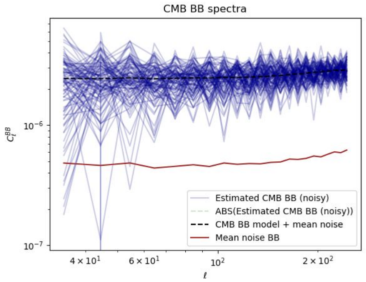
Foreground map



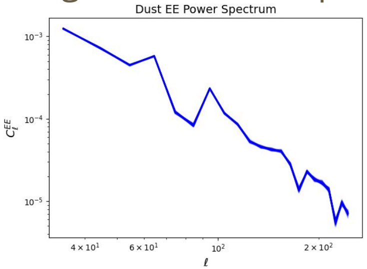
POWER SPECTRA ESTIMATION

Estimated using Namaster pseudo spectra approach

CMB BB Power Spectra



Foreground Power Spectra



COSMOLOGICAL ANALYSIS

➤ The **Cl s** of the observed CMB map are modeled as:

$$\tilde{C}_\ell^{\mathrm{CMB}}(r, A_{\mathrm{lens}}) \equiv C_\ell^{\mathrm{prim}}(r) + C_\ell^{\mathrm{lens}}(A_{\mathrm{lens}}) + \tilde{N}_\ell^{\mathrm{CMB}}$$

➤ Cosmological parameters are inferred via the **likelihood:**

$$-2\,\log\mathscr{L}^{\mathrm{cosmo}} = \sum_\ell (2\ell+1)\, f_{\mathrm{sky}} \left(\frac{\tilde{C}_\ell^{\mathrm{CMB}}}{C_\ell^{\mathrm{CMB}}} + \log\left(C_\ell^{\mathrm{CMB}}\right)\right)$$

➤ The foreground maps allow for **marginalisation** over the residual dust spectra by extending the model:

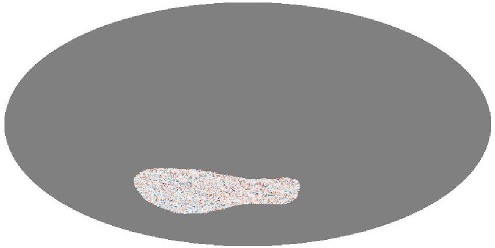
$$\tilde{C}_\ell^{\mathrm{CMB}} = \tilde{C}_\ell^{\mathrm{CMB}}(r, A_{\mathrm{lens}}) + A_{\mathrm{dust}}\tilde{C}_\ell^{\mathrm{dust}}$$

From Component Maps to Cosmology

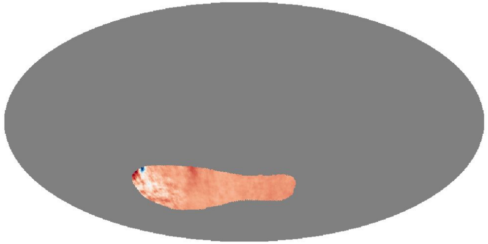
COMPONENT MAPS

$$\hat{\mathbf{s}} = \left(\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^t\mathbf{N}^{-1}\mathbf{d}$$

CMB map



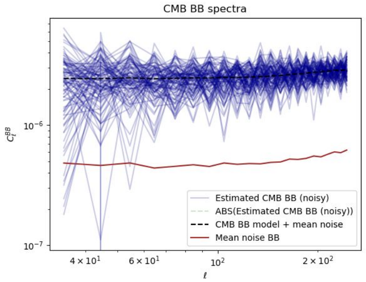
Foreground map



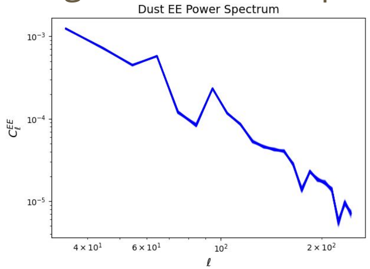
POWER SPECTRA ESTIMATION

Estimated using Namaster pseudo spectra approach

CMB BB Power Spectra

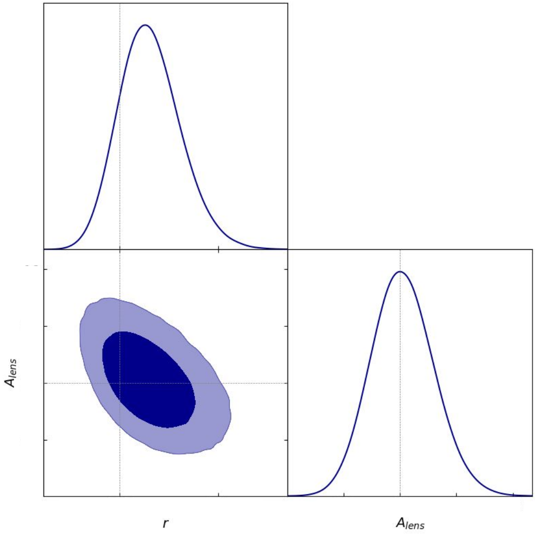


Foreground Power Spectra



COSMOLOGICAL ANALYSIS

Cosmological results for one particular simulation:

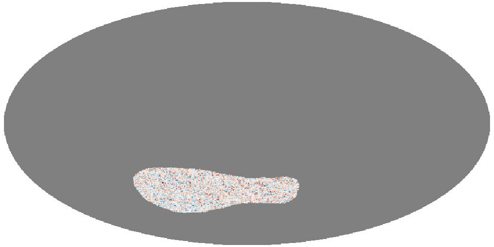


From Component Maps to Cosmology

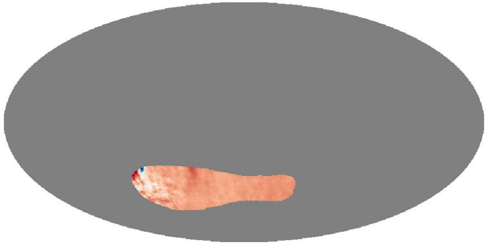
COMPONENT MAPS

$$\hat{\mathbf{s}} = \left(\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^t\mathbf{N}^{-1}\mathbf{d}$$

CMB map



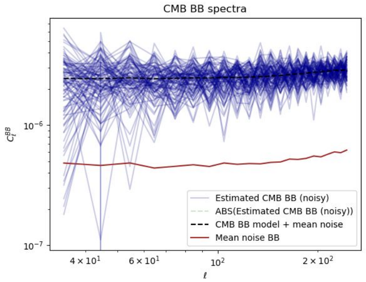
Foreground map



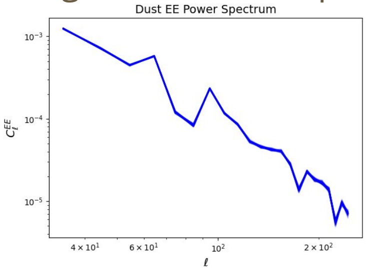
POWER SPECTRA ESTIMATION

Estimated using Namaster pseudo spectra approach

CMB BB Power Spectra

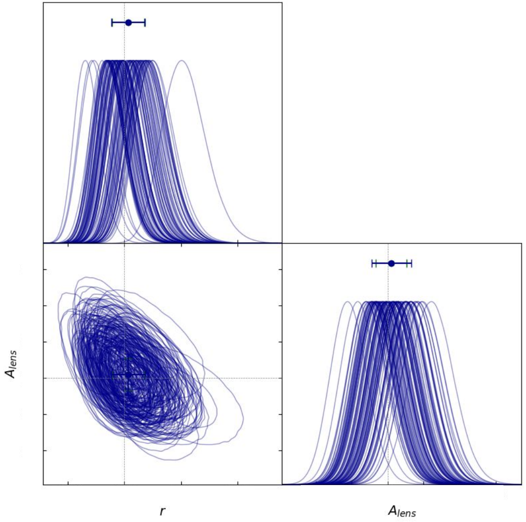


Foreground Power Spectra



COSMOLOGICAL ANALYSIS

We can run the full pipeline on many simulations efficiently

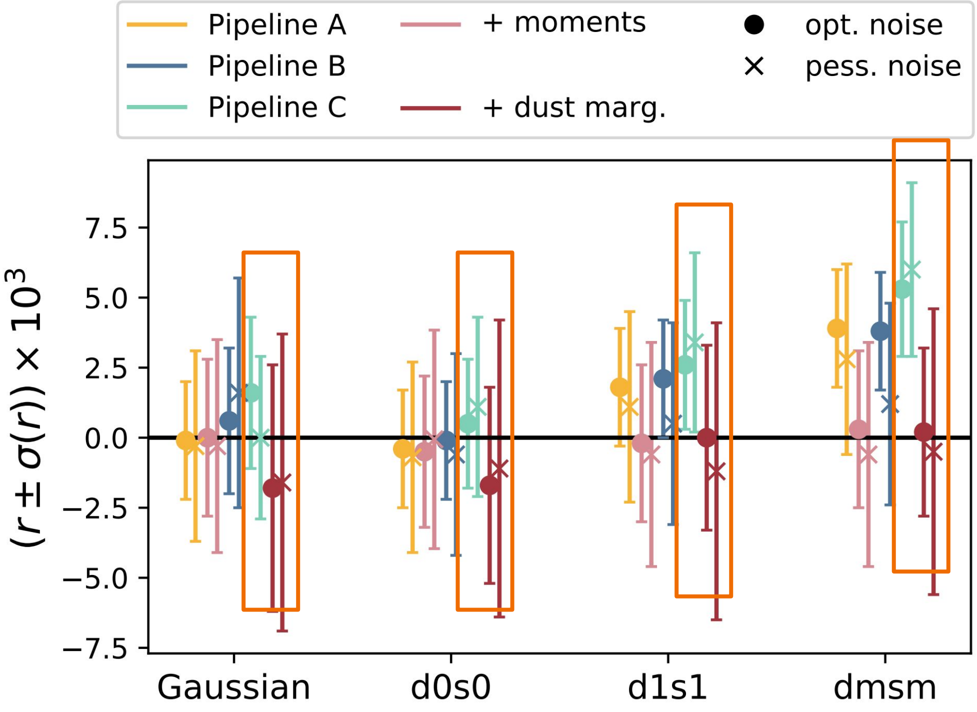


11

One of the Main Component Separation Pipeline in SO

Map-based method has been tested on Simons Observatory simulations (Wolz et al. 2024):

- comparable to other methods: ILC, Cross-spectra parametric method
- Robust to complex foregrounds thanks to dust marginalization

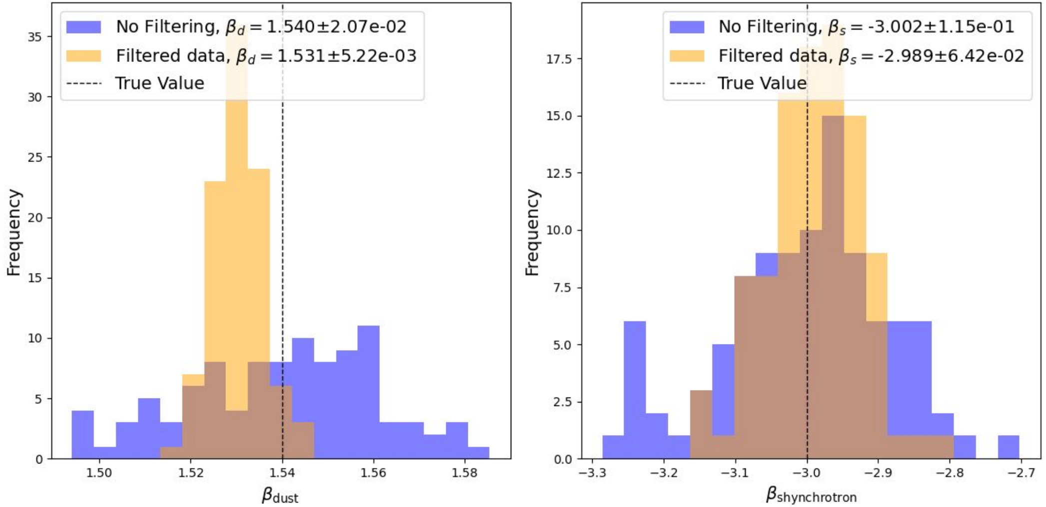


Source: Wolz et al 2024

Map-Based Parametric Component Separation On Filtered Data

Simulations with CMB + low complexity foregrounds (d0s0) for SO SAT survey using filtering operator (credit: SO simulations team).

The filtering is frequency dependent.



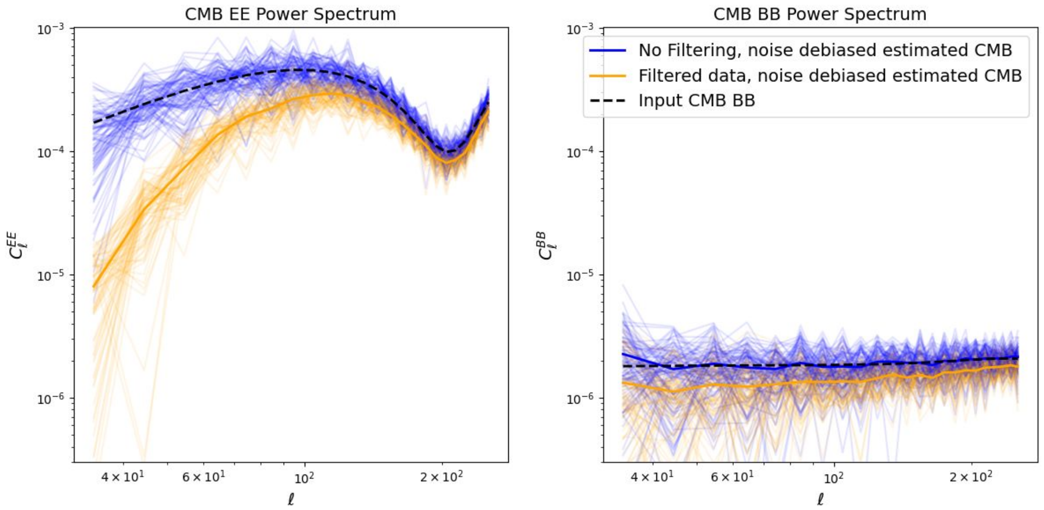
Ignoring the filtering in foreground cleaning and subsequent steps

⇒ Correction of filtering is necessary at the component separation level.

Map-Based Parametric Component Separation On Filtered Data

Simulations with CMB + low complexity foregrounds (d0s0) for SO SAT survey using filtering operator (credit: SO simulations team).

The filtering is frequency dependent.



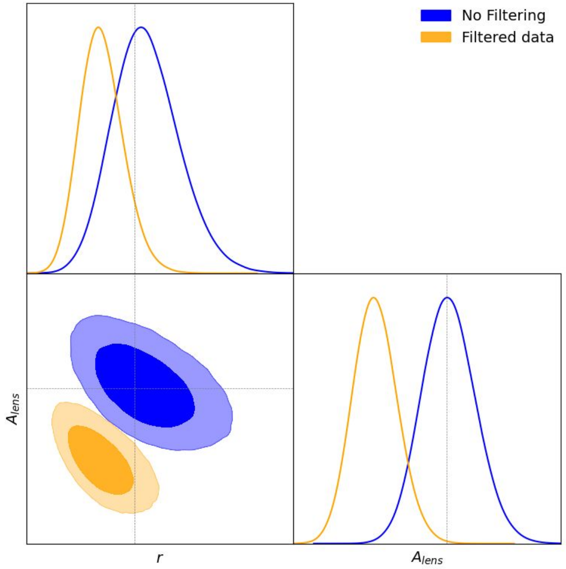
Ignoring the filtering in foreground cleaning and subsequent steps

⇒ Correction of filtering is necessary at the component separation level.

Map-Based Parametric Component Separation On Filtered Data

Simulations with CMB + low complexity foregrounds (d0s0) for SO SAT survey using filtering operator (credit: SO simulations team).

The filtering is frequency dependent.



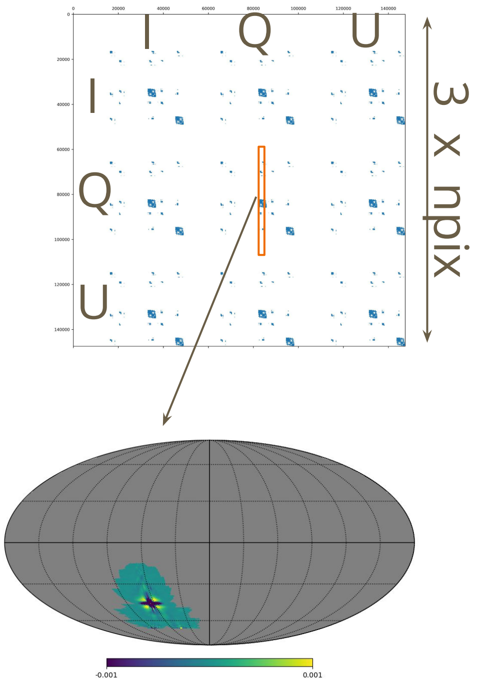
Ignoring the filtering in foreground cleaning and subsequent steps

⇒ Correction of filtering is necessary at the component separation level.

Accounting for Filtering in Component Separation

Data filtering expressed as linear operator on map space: the **Observation matrix** (Keck Array, BICEP2 Collaborations 2016)

$$\mathbf{m}_p^{\mathrm{out}} = \mathbf{O}_{pp'}\mathbf{m}_{p'}^{\mathrm{in}}$$



Constructed using instrument model and data preprocessing scheme.

Very large matrix, here for reduced resolution (nside=128):

- 600 000 x 600 000 elements
- >11 GB per map / frequency channel

Accounting for Filtering in Component Separation

We update the data model to include the observation matrix:

$$\mathbf{d}_p = \mathbf{O}_{pp'}(\mathbf{A}(\beta)\mathbf{s}_p + \mathbf{n}_p)$$

The spectral likelihood is modified accordingly:

$$-2\ln\mathcal{L}_{\rm spec}(\beta) \sim (\mathbf{A}^t\mathbf{O}^t\mathbf{N}^{-1}\mathbf{d})^t(\mathbf{A}^t\mathbf{O}^t\mathbf{N}^{-1}\mathbf{O}\mathbf{A})^{-1}(\mathbf{A}^t\mathbf{O}^t\mathbf{N}^{-1}\mathbf{d})$$

PROS

- No need for **simulations**
- Better suited for **pixel-based** analysis
- Describes filtering **spatial variability** and **correlations**
- Optimal noise weighting
- No need to filter external data
- No assumption of gaussianity of the foregrounds

CONS

- High **computational cost**
- Observations matrices need to be available
- Non-linear filtering are trickier to implement

Accounting for Filtering in Component Separation

We update the data model to include the observation matrix:

$$\mathbf{d}_p = \mathbf{O}_{pp'}(\mathbf{A}(\beta)\mathbf{s}_p + \mathbf{n}_p)$$

The spectral likelihood is modified accordingly:

$$-2\ln\mathcal{L}_{\rm spec}(\beta) \sim (\mathbf{A}^t\mathbf{O}^t\mathbf{N}^{-1}\mathbf{d})^t(\mathbf{A}^t\mathbf{O}^t\mathbf{N}^{-1}\mathbf{O}\mathbf{A})^{-1}(\mathbf{A}^t\mathbf{O}^t\mathbf{N}^{-1}\mathbf{d})$$

PROS

- No need for **simulations**
- Better suited for **pixel-based** analysis
- Describes filtering **spatial variability** and **correlations**
- Optimal noise weighting
- No need to filter external data*
- No assumption of gaussianity of the foregrounds

CONS

- High **computational cost**
- Observations matrices need to be available
- Non-linear filtering are trickier to implement

***See Camille Gubeno's talk**

Accounting for Filtering in Component Separation

We update the data model to include the observation matrix:

$$\mathbf{d}_p = \mathbf{O}_{pp'}(\mathbf{A}(\beta)\mathbf{s}_p + \mathbf{n}_p)$$

The spectral likelihood is modified accordingly:

$$-2\ln\mathcal{L}_{\rm spec}(\beta) \sim (\mathbf{A}^t\mathbf{O}^t\mathbf{N}^{-1}\mathbf{d})^t(\mathbf{A}^t\mathbf{O}^t\mathbf{N}^{-1}\mathbf{O}\mathbf{A})^{-1}(\mathbf{A}^t\mathbf{O}^t\mathbf{N}^{-1}\mathbf{d})$$

With a precomputed preconditioner and JAX acceleration using the FURAX framework we are able to estimate foreground parameters! (credit: Magdy Morshed)

We can then estimate CMB and foreground maps using (BICEP/Keck Collaboration 2025 **see Dominic Beck's talk tomorrow!**):

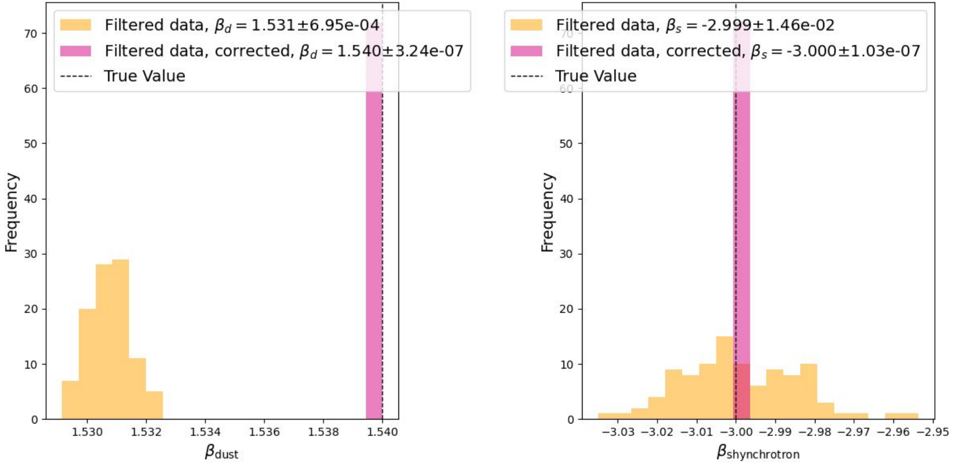
$$\hat{\mathbf{s}} = (\mathbf{A}^t\mathbf{O}^t\mathbf{N}^{-1}\mathbf{O}\mathbf{A})^{-1}\mathbf{A}^t\mathbf{N}^{-1}\mathbf{d}$$

Map-Based Parametric Component Separation On Filtered Data

Running on **noiseless** CMB + low complexity foregrounds for SO SAT survey.

Filtered using observation matrix for SO (credit: SO simulations team), one per frequency channel. Two cases:

- Ignoring the filtering in foreground cleaning and subsequent steps
- Including observation matrices

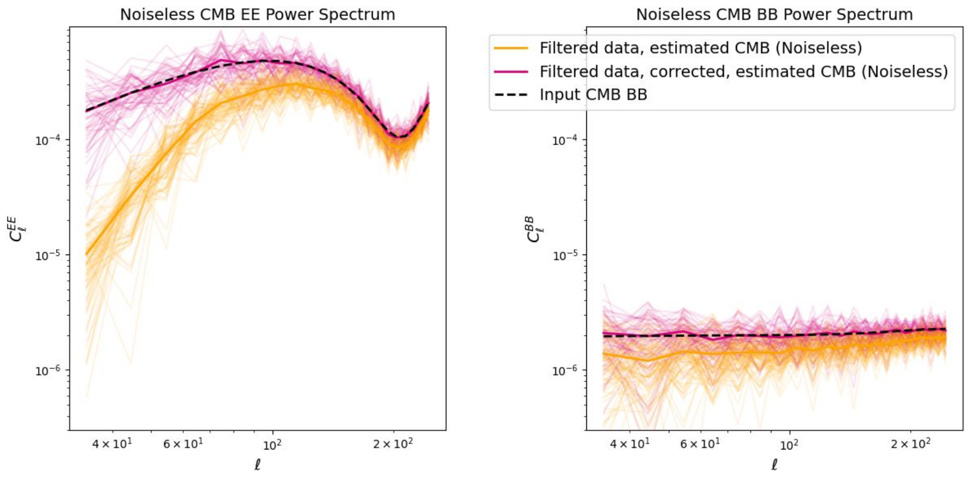


Map-Based Parametric Component Separation On Filtered Data

Running on **noiseless** CMB + low complexity foregrounds for SO SAT survey.

Filtered using observation matrix for SO (credit: SO simulations team), one per frequency channel. Two cases:

- Ignoring the filtering in foreground cleaning and subsequent steps
- Including observation matrices

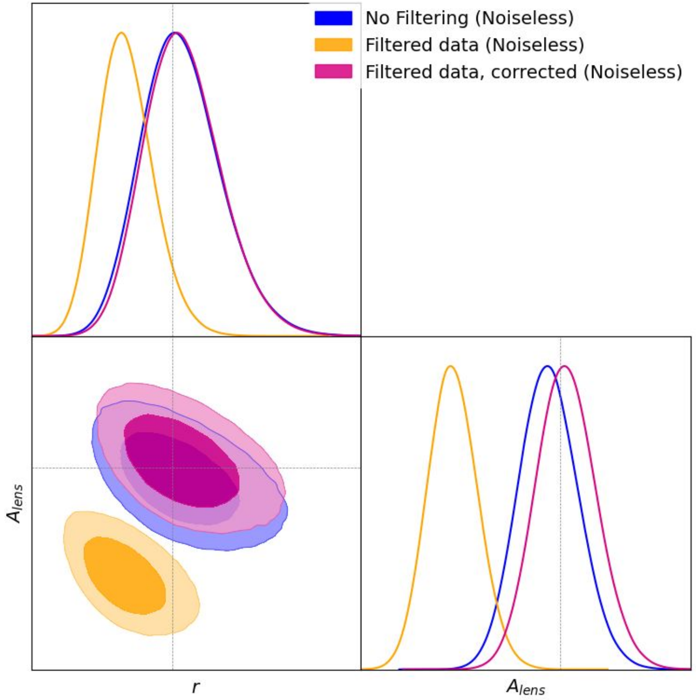


Map-Based Parametric Component Separation On Filtered Data

Running on **noiseless** CMB + low complexity foregrounds for SO SAT survey.

Filtered using observation matrix for SO (credit: SO simulations team), one per frequency channel. Two cases:

- Ignoring the filtering in foreground cleaning and subsequent steps
- Including observation matrices

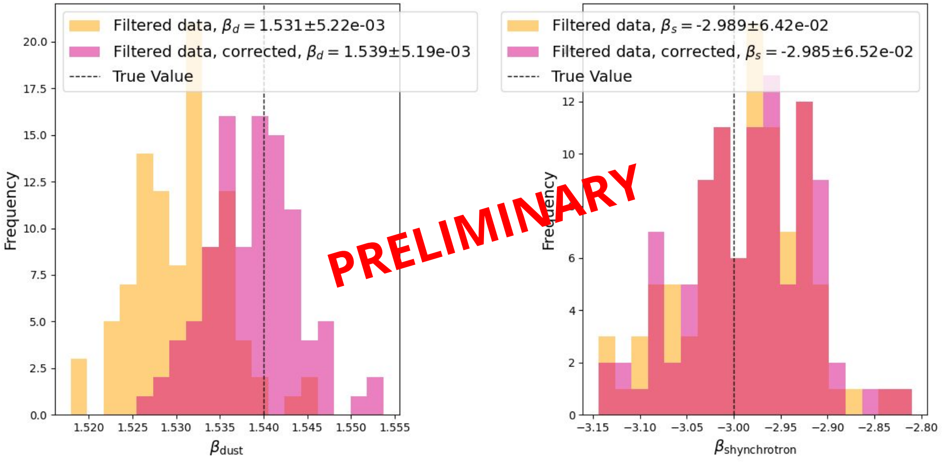


Map-Based Parametric Component Separation On Filtered Data

Preliminary results on **noisy** CMB + low complexity foregrounds for SO SAT survey.

Filtered using observation matrix for SO (credit: SO simulations team), one per frequency channel.

The noise is also filtered with the observation matrix.

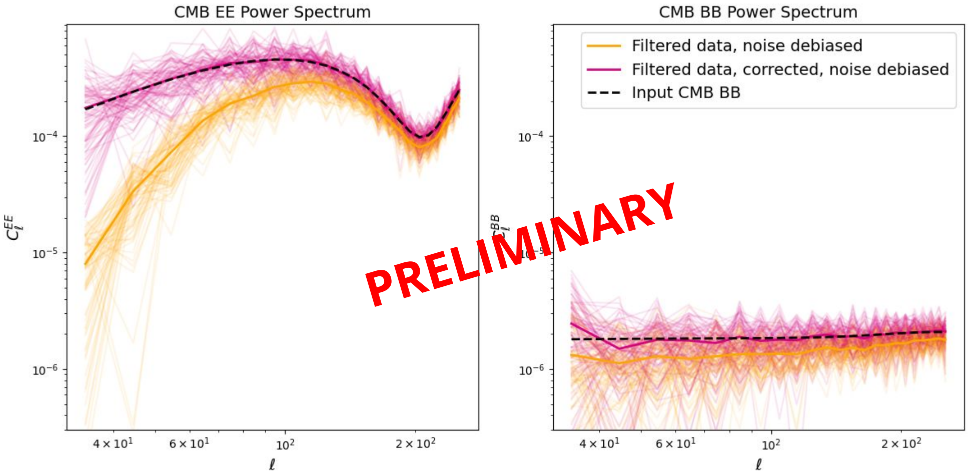


Map-Based Parametric Component Separation On Filtered Data

Preliminary results on **noisy** CMB + low complexity foregrounds for SO SAT survey.

Filtered using observation matrix for SO (credit: SO simulations team), one per frequency channel.

The noise is also filtered with the observation matrix.

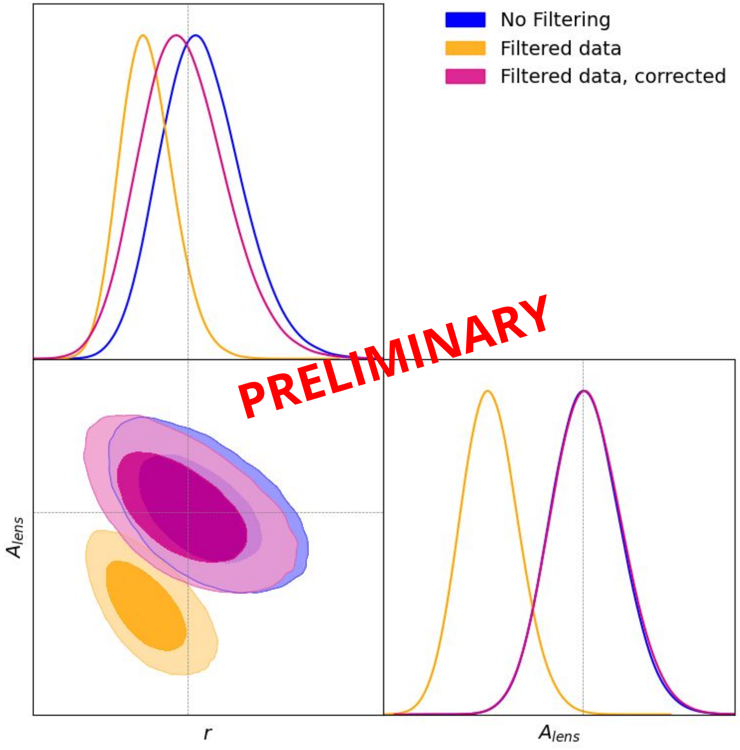


Map-Based Parametric Component Separation On Filtered Data

Preliminary results on **noisy** CMB + low complexity foregrounds for SO SAT survey.

Filtered using observation matrix for SO (credit: SO simulations team), one per frequency channel.

The noise is also filtered with the observation matrix.



Conclusion

MEGATOP is the **map-based** component separation for the SO SAT analysis

It is a versatile tool and provides cross-check for current and future CMB experiments.

Including the effect of filtering is necessary.

We are able to correct for it at the map level using observation matrices

Current/future prospects:

- application on external (**see Camille Gubeno's talk for preliminary study on NPIPE simulations**)
- internal SO data (work in progress)
- include beam operator in pixel based framework
- complex noise correlation
- polarization angles
- multi patch
- ...

THANK YOU!