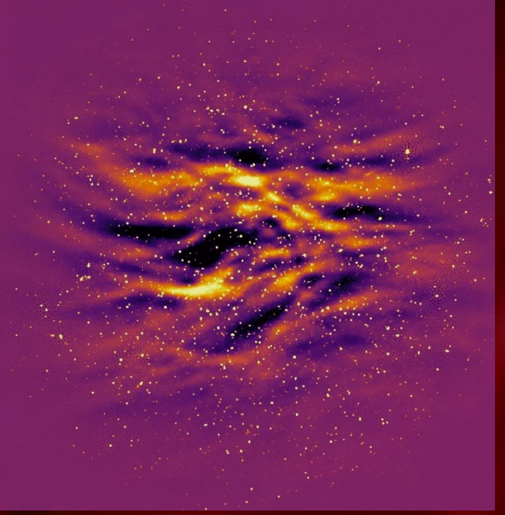
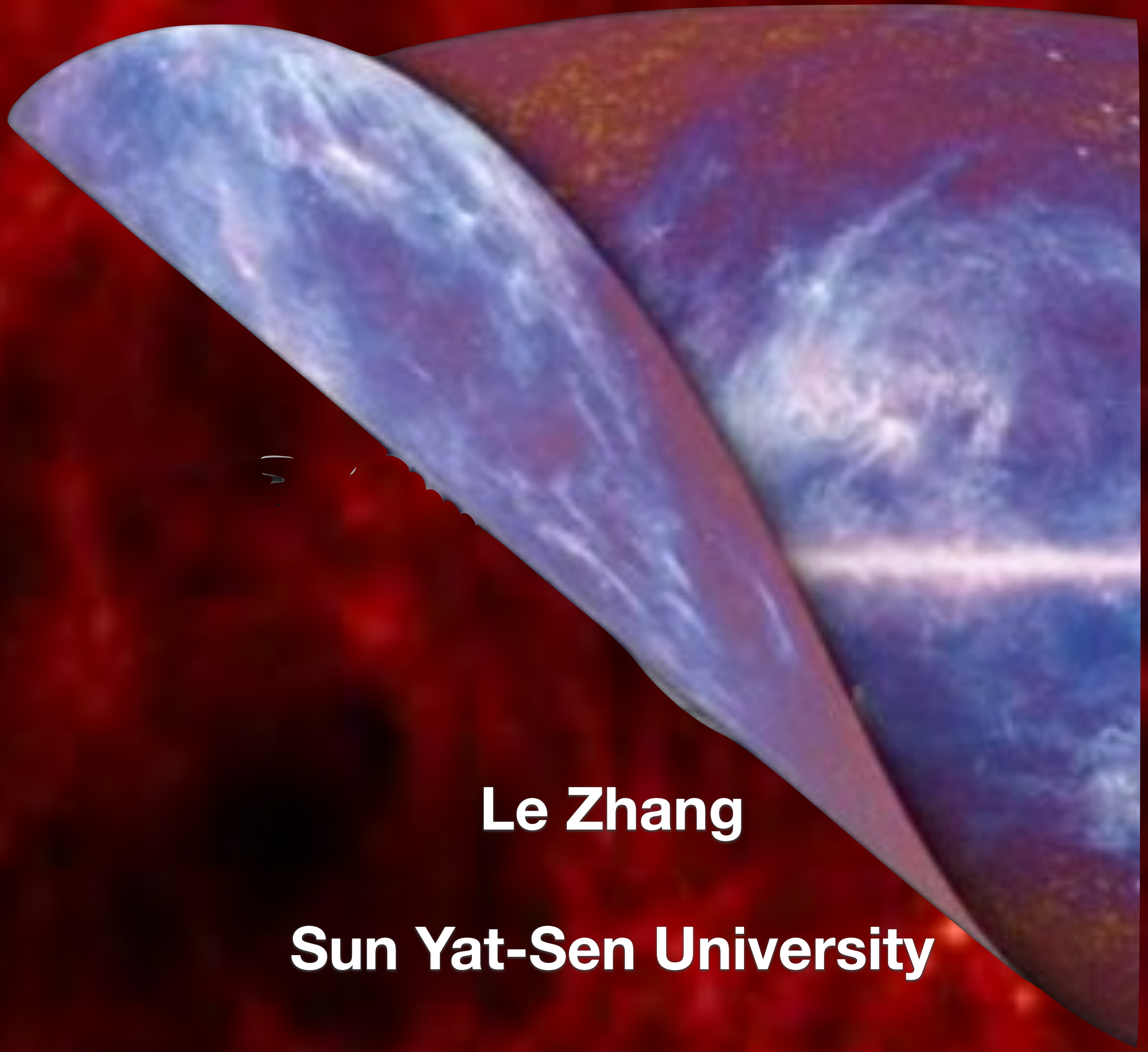


Foreground removal in 21 cm surveys

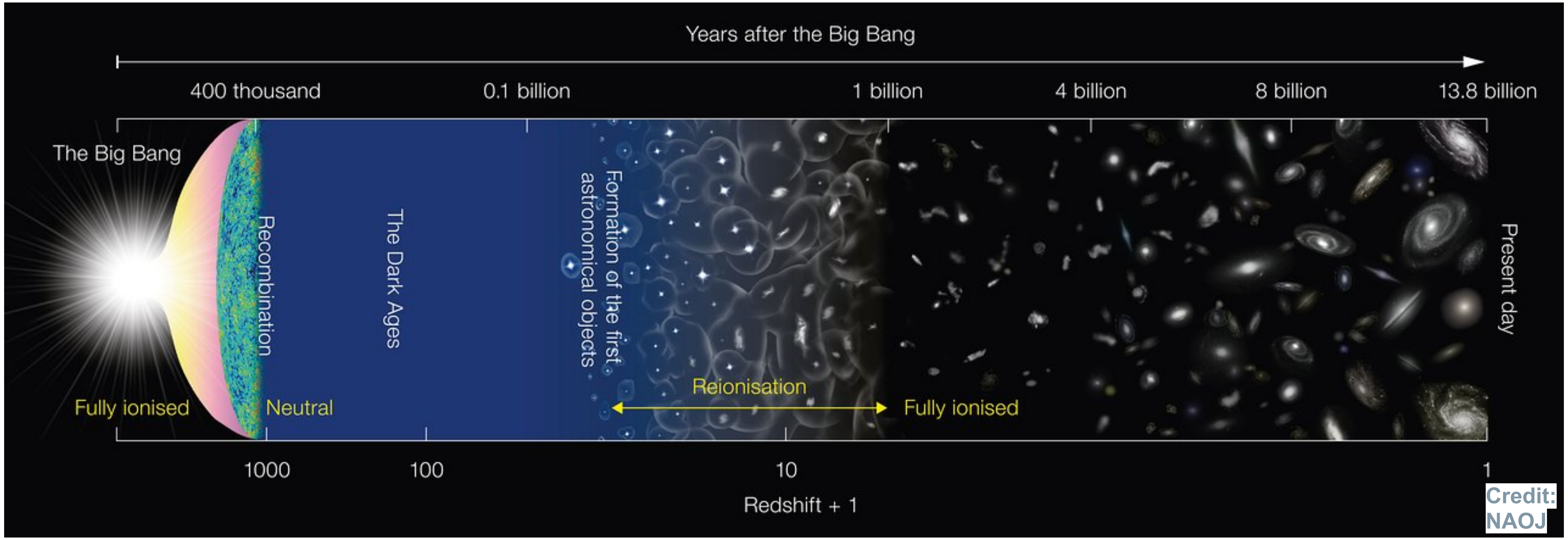


Le Zhang

Sun Yat-Sen University

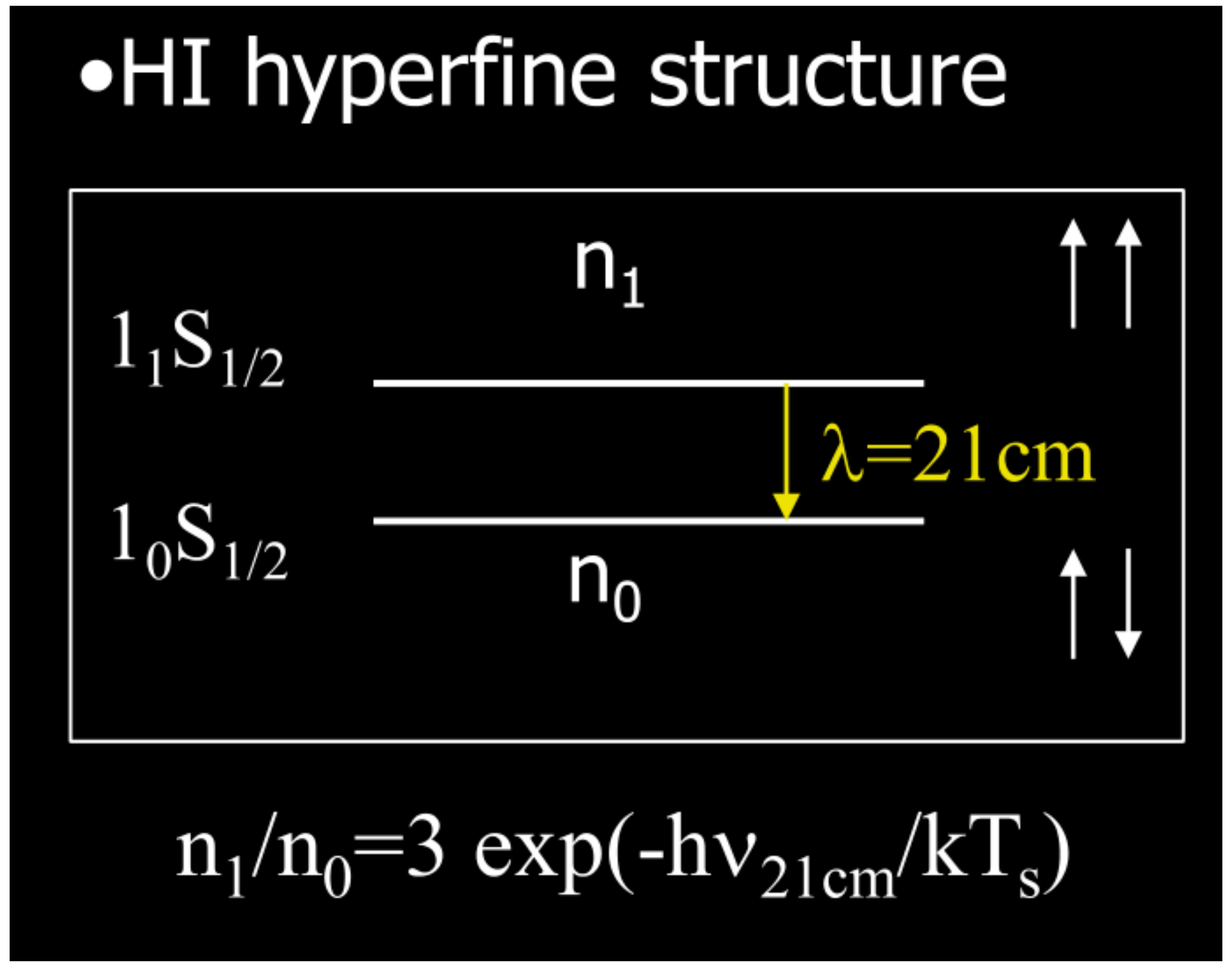
XI’An 2025.07.12

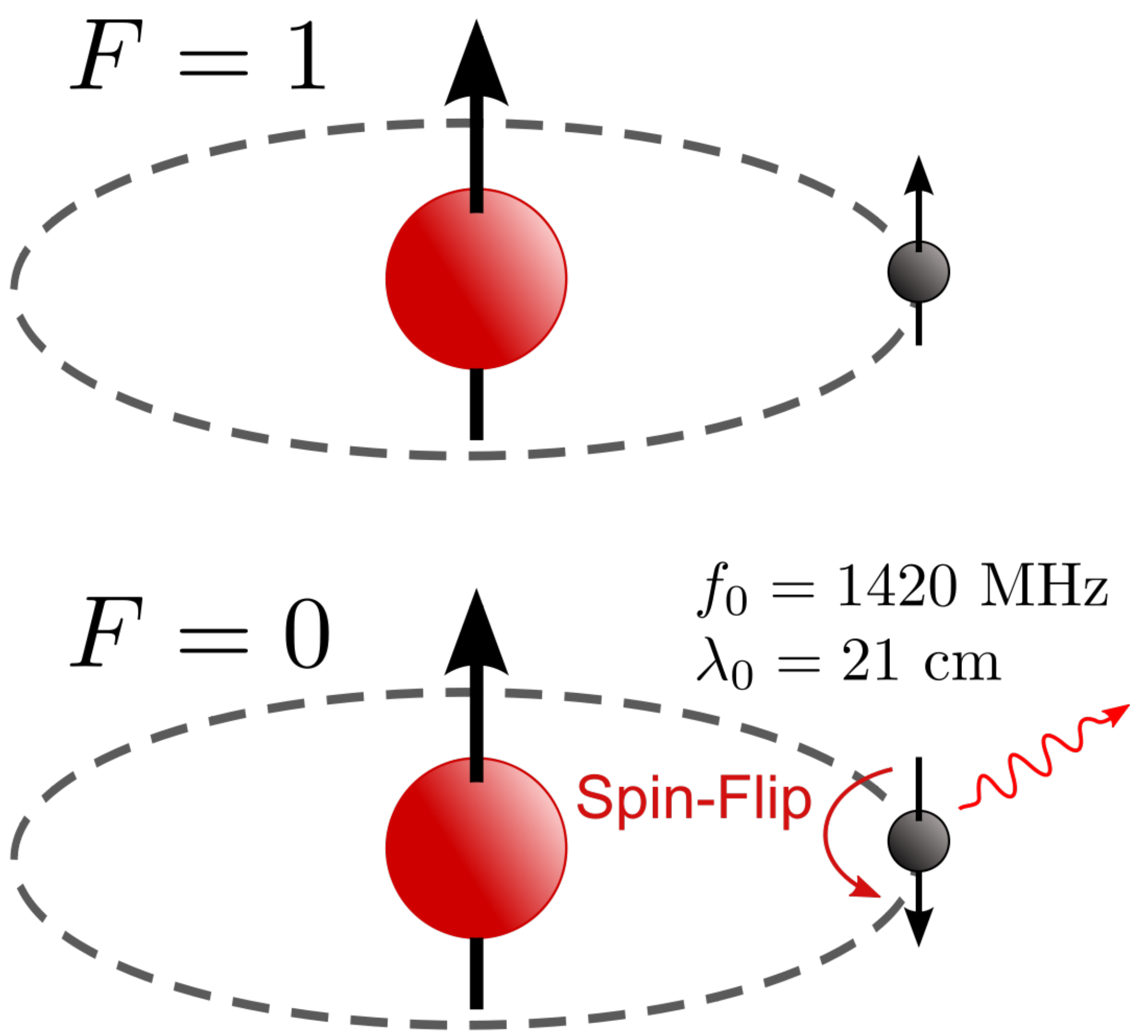
History of the Universe



- **Dark Ages: No luminous star exists (z > 30)**
- **Cosmic Dawn: Emergence of the initial stars and galaxies (z ~ 20-30)**
- **Epoch of Reionization (EoR): UV photons from luminous objects ionize HI in the IGM (z ~ 6-15)**

None of these have been detected !!!

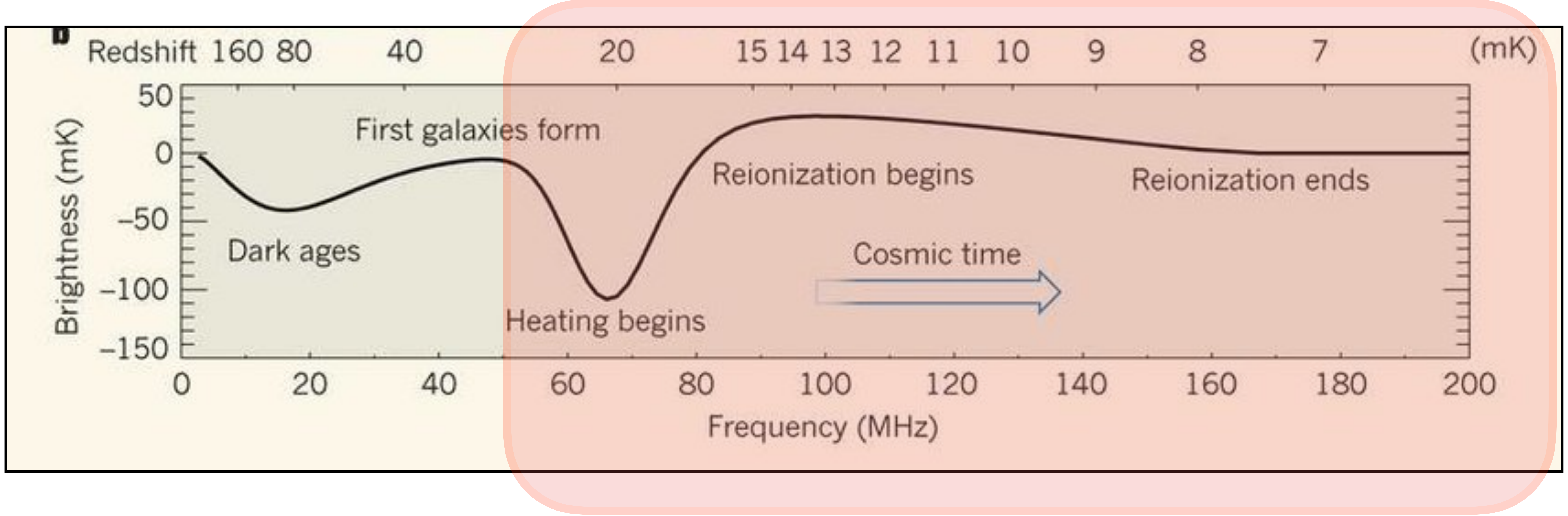




The only way to study this period:

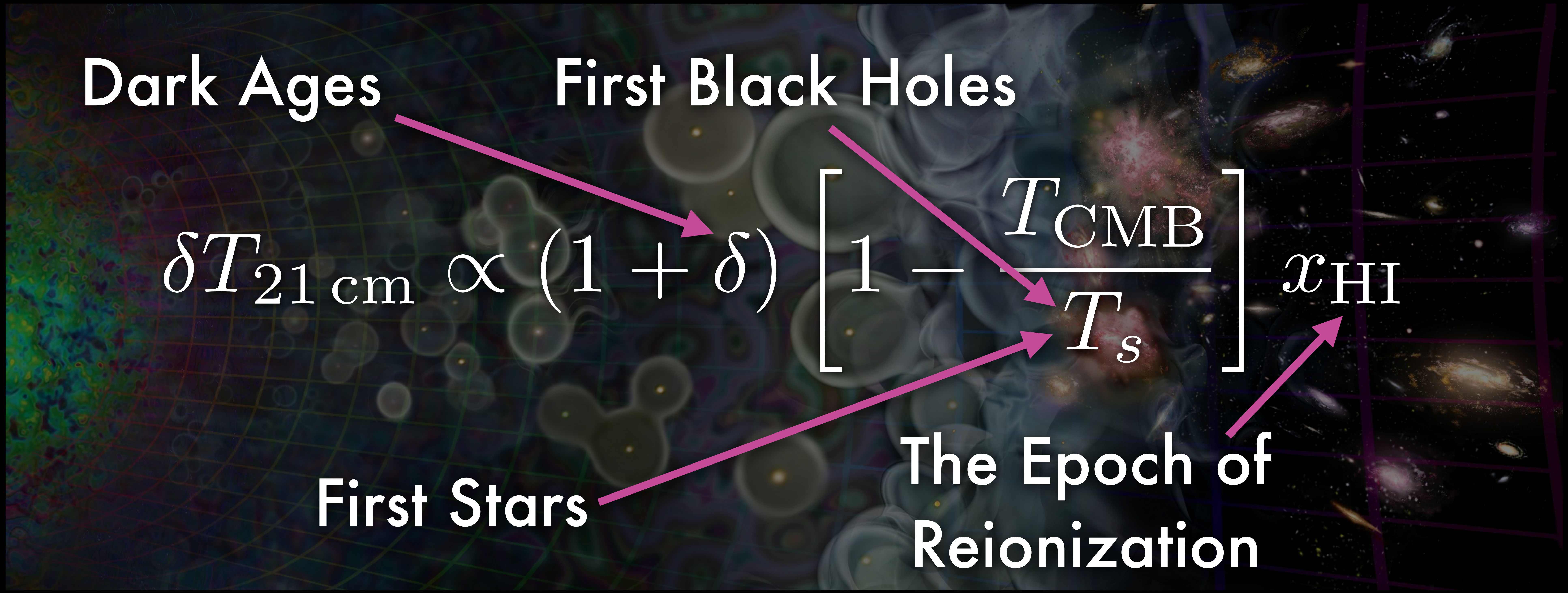
Observing Neutral Hydrogen 21cm Radiation

The SKA-low array:
Unveiling the Cosmic Dawn and reionization



- **Exploring the First Billion Years**
- **Mapping the Structure of the Early Universe**
- **Witnessing Births and Deaths of First Stars**
- **Unraveling the Formation of Earliest Galaxies**

Brightness temperature probes different physics at different times



z =1100 z ≈ 50 z ≈ 8 z < 6

credit: Josh Dillon

Global 21cm experiments

EDGES
50 – 100, 100 – 200 MHz
Murchison Radio Obs.

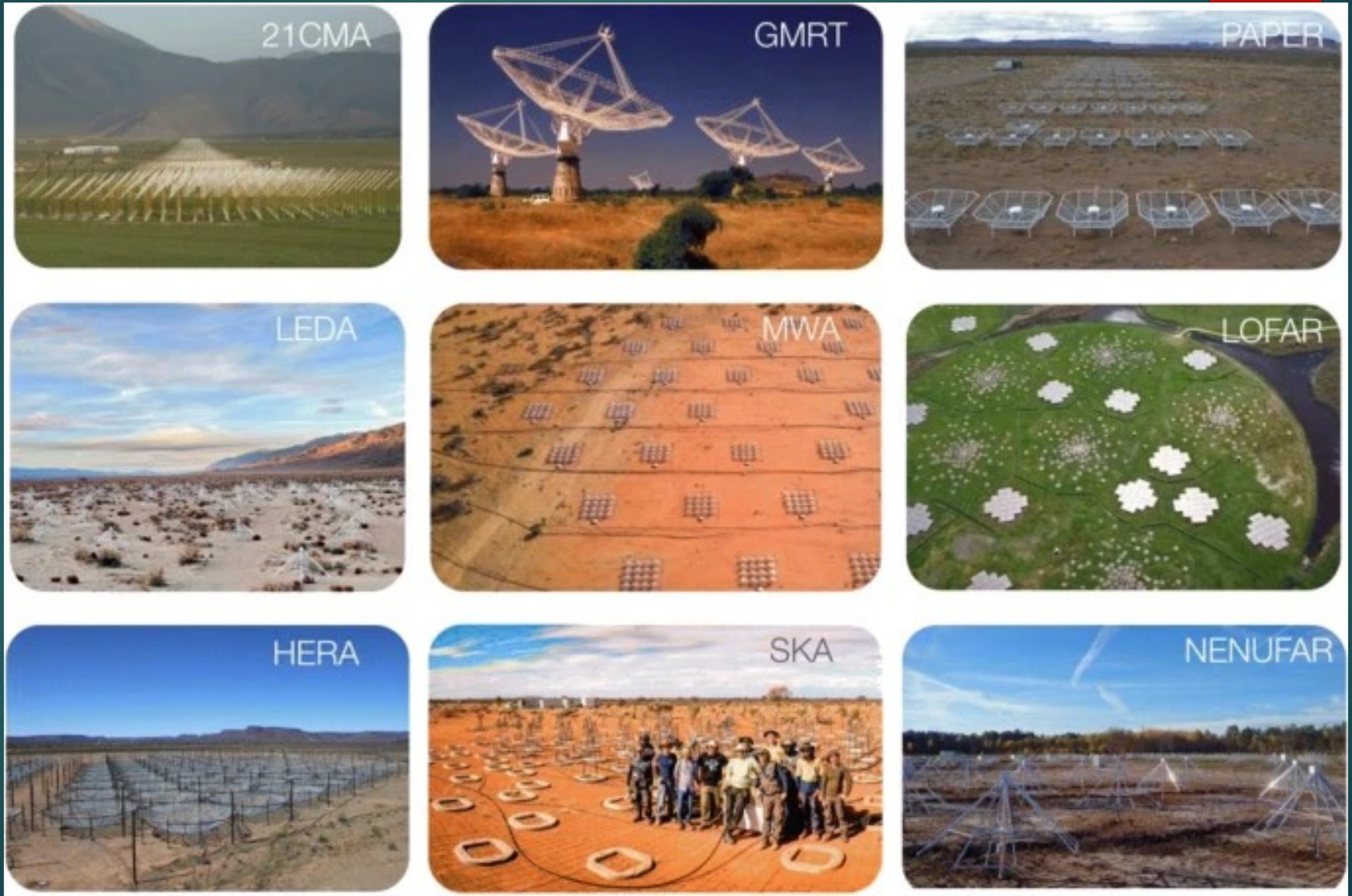
BIGHORNS
50 – 200 MHz
Western Australia

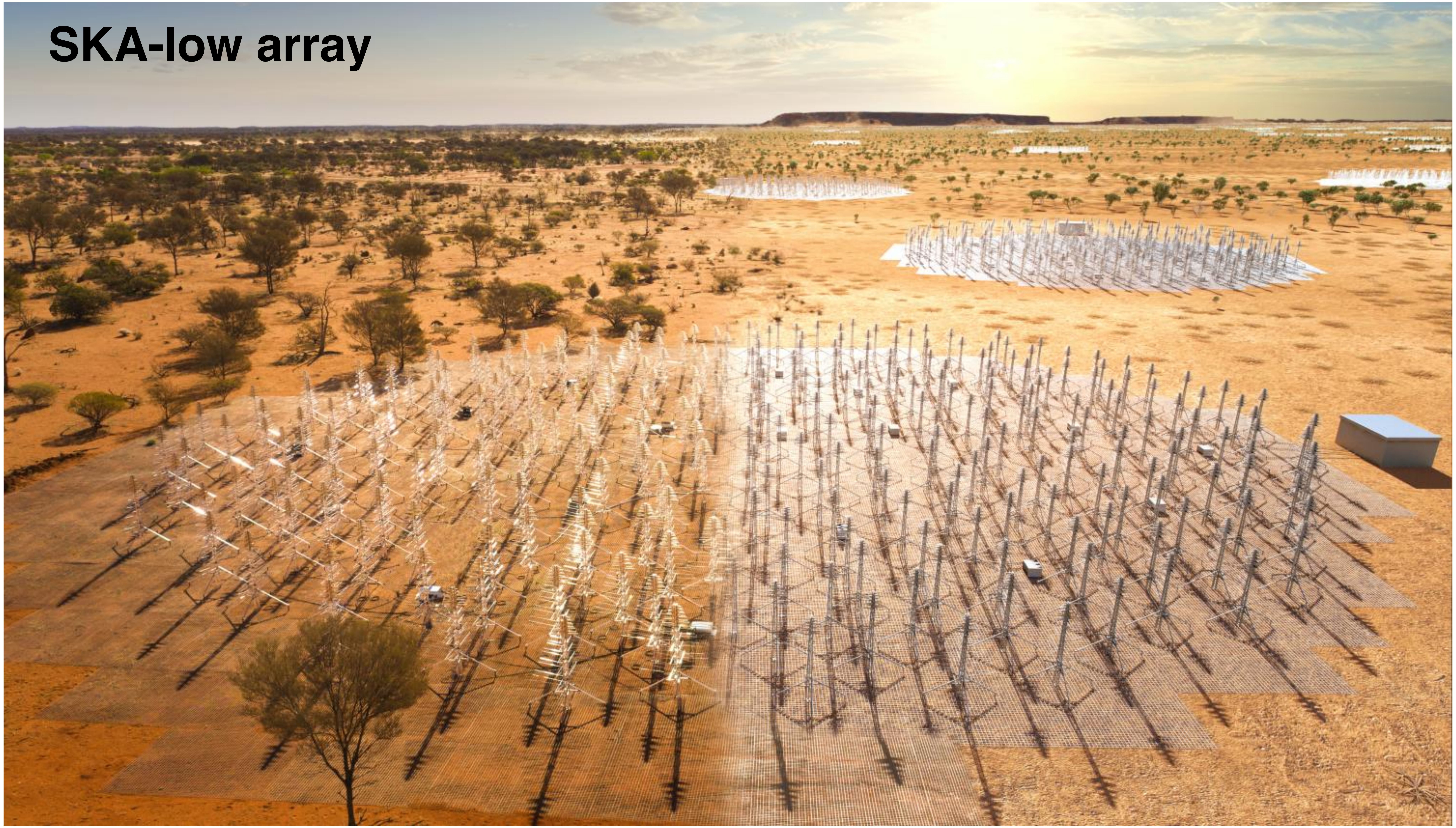
LEDA
30 – 88 MHz
Owens Valley

SARAS2
87.5 – 175 MHz
Gauribidanur Obs., India

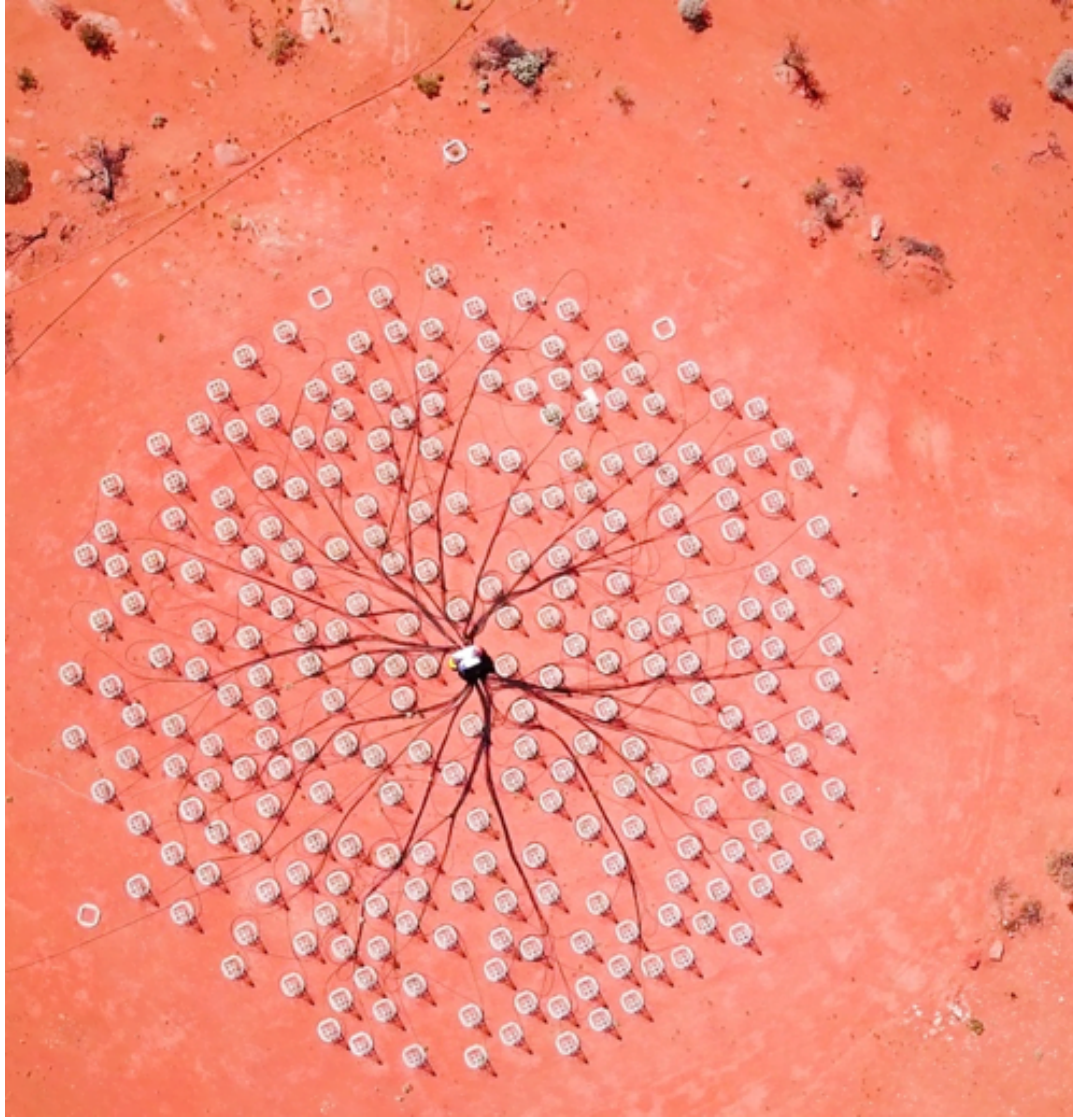
HYPERION
30 – 120 MHz
Owens Valley

21cm Power spectrum measurements





Compared to LOFAR:
25% better resolution, 8$\times$ the sensitivity, 135$\times$ faster

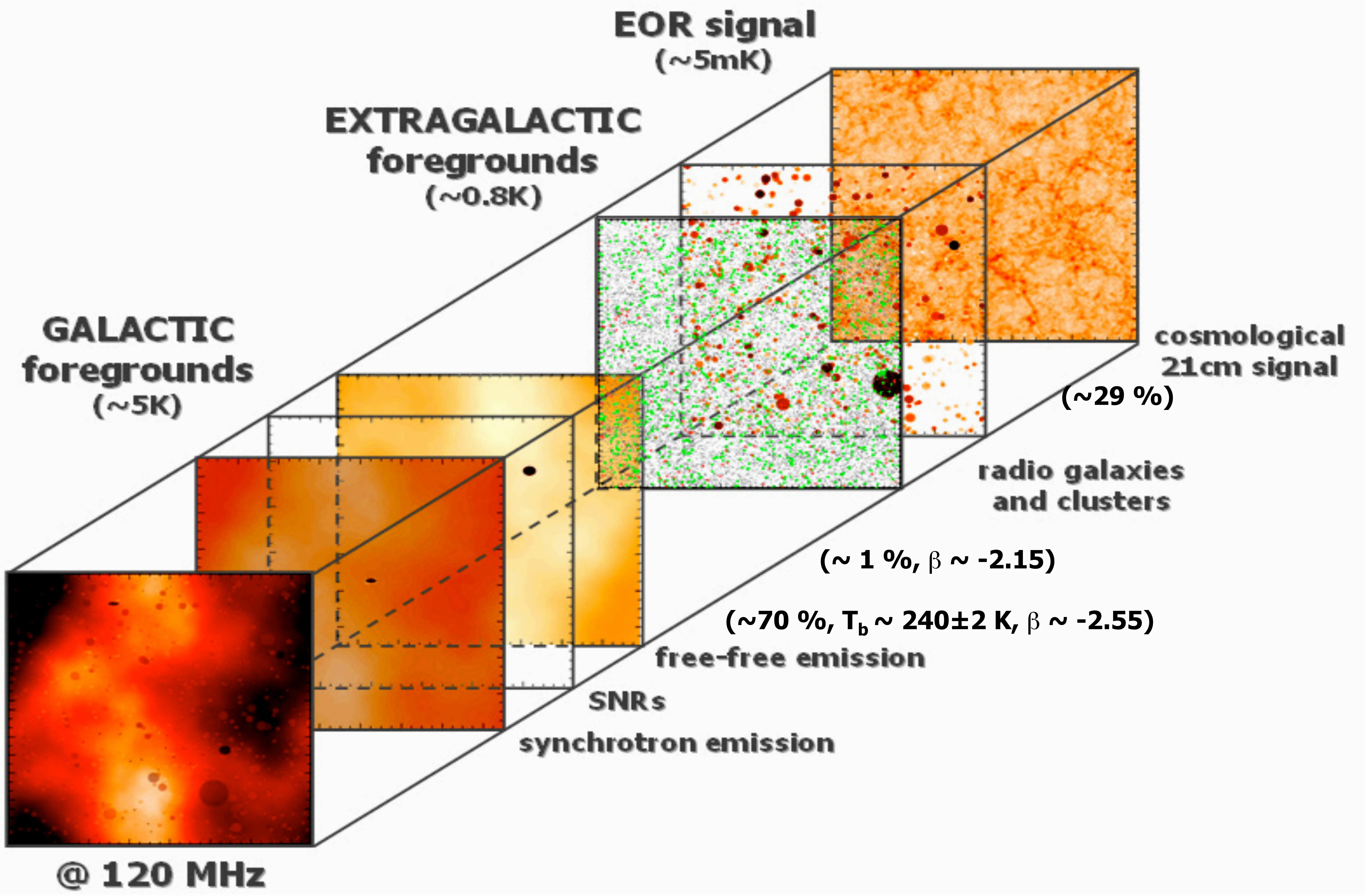


- 131,072 log-periodic antennas spread between 512 stations (~74km)
- Collecting area: 419,000 m^2
- Frequency range: 50 MHz - 350 MHz

Why is it so difficult to observe the 21 cm signal from the early universe?

Foreground Challenges

• featureless power law
• variation in spectral index with position on the sky and with frequency



SIMULATIONS: 5º x 5º filed of view, ~ 0.6 arcmin resolution and freq. range: 115-180 MHz LOFAR 50-370 SKA-low

Jelic et al., 2008, MNRAS

nature > letters > article

Letter | Published: 01 March 2018

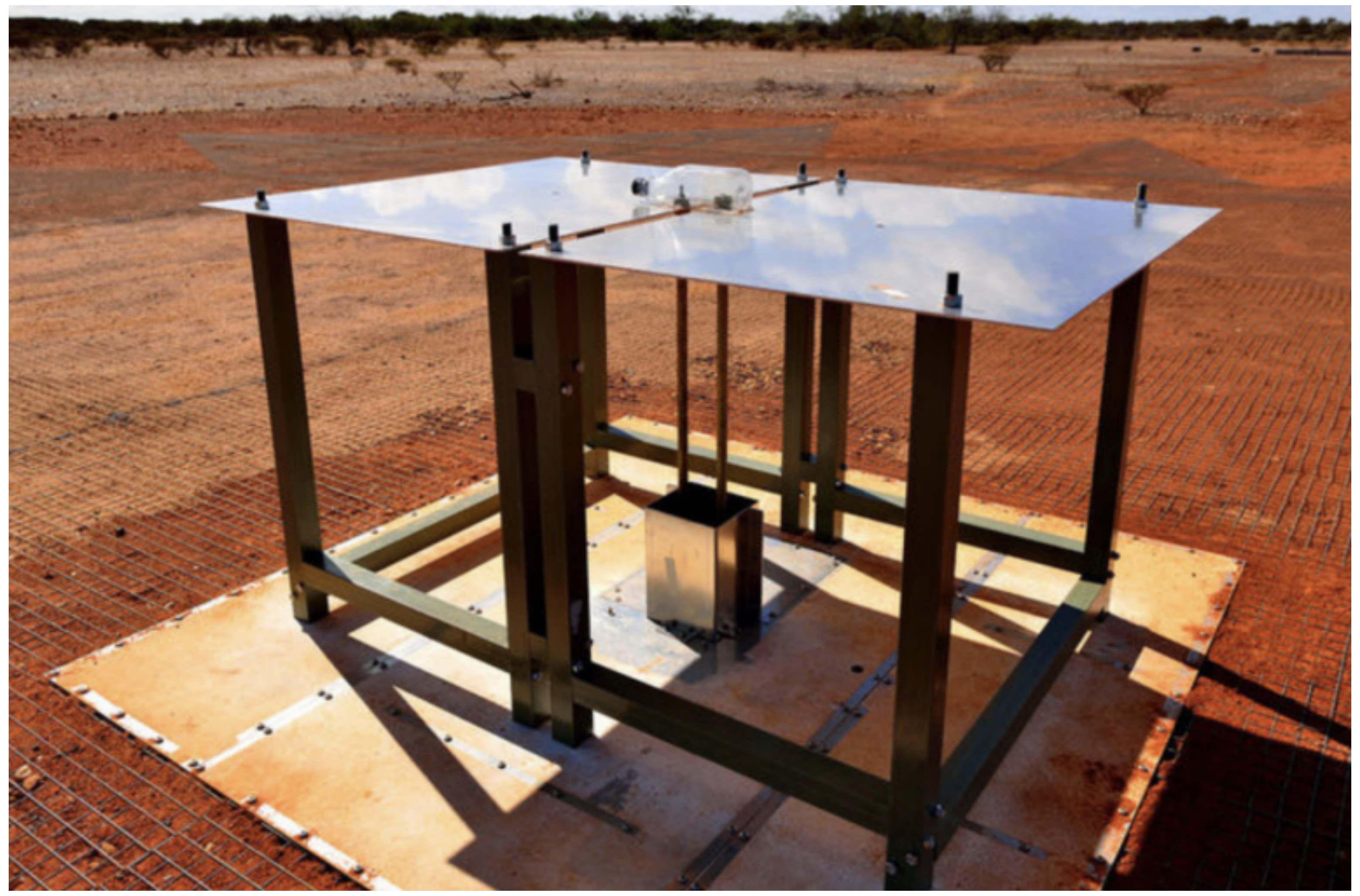
An absorption profile centred at 78 megahertz in the sky-averaged spectrum

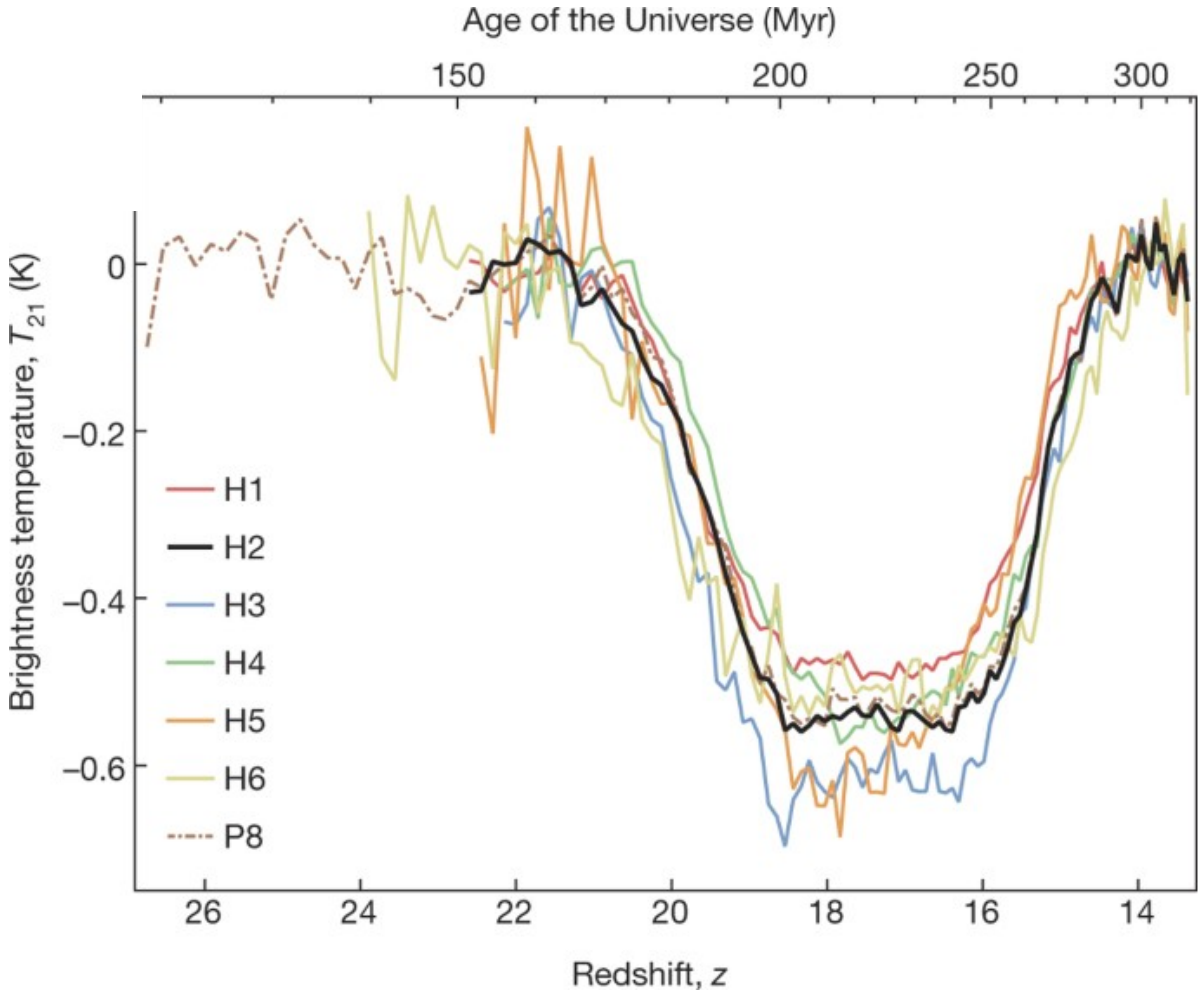
Judd D. Bowman, Alan E. E. Rogers, Raul A. Monsalve, Thomas J. Mozdzen & Nivedita Mahesh

Nature 555, 67–70 (2018) | Cite this article

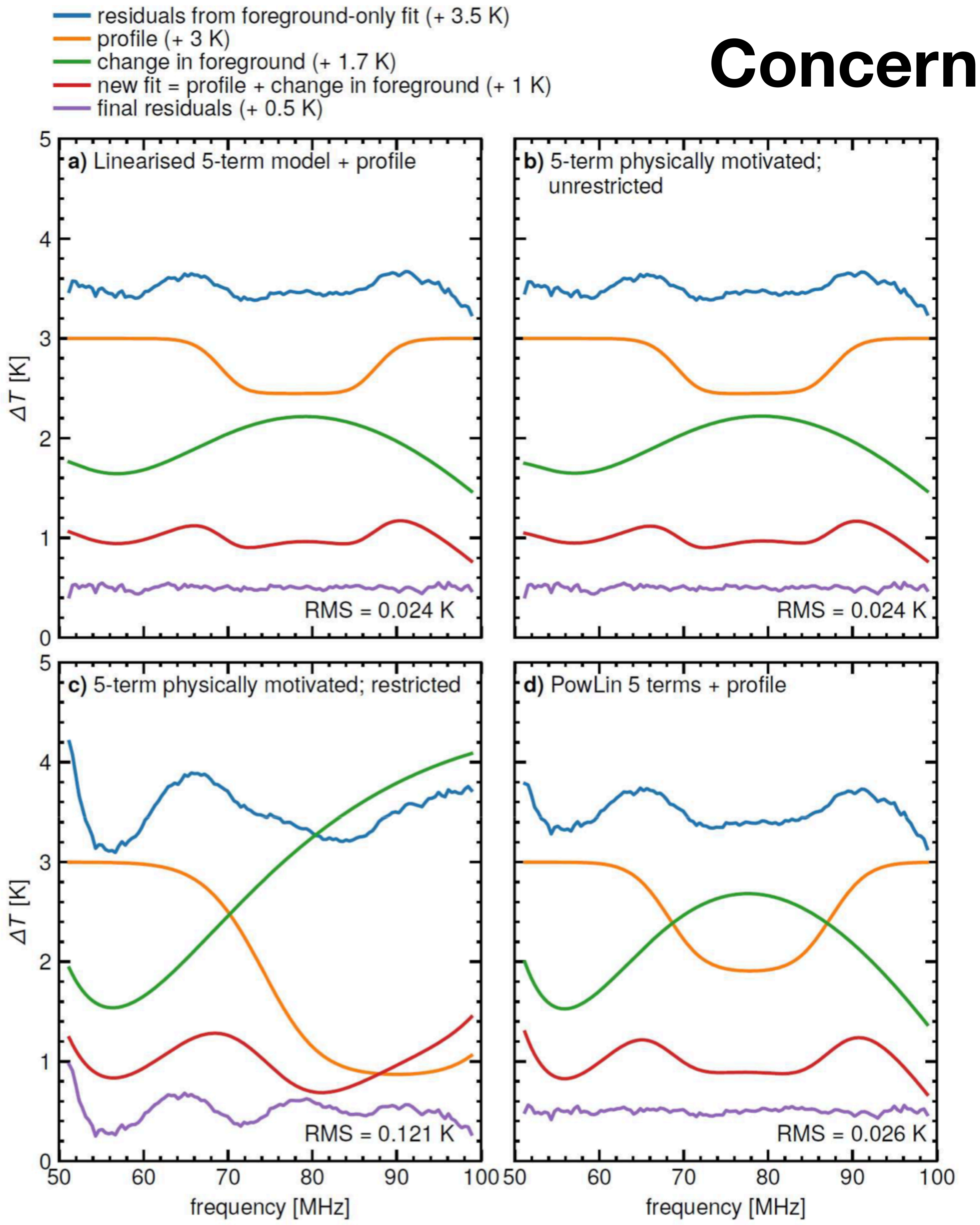
41k Accesses | 1005 Citations | 2104 Altmetric | Metrics

A Brief Communications Arising to this article was published on 19 December 2018





Concerns about Modelling of the EDGES Data



- **Fit foregrounds using low-order polynomials in log-log space**

$$T_{\mathrm{ant}} = T_{75}\left(\frac{\nu}{\nu_{75}}\right)^{\beta+\gamma\ln\left(\frac{\nu}{\nu_{75}}\right)+a_4\left[\ln\left(\frac{\nu}{\nu_{75}}\right)\right]^2+a_5\left[\ln\left(\frac{\nu}{\nu_{75}}\right)\right]^3} + T_{\mathrm{CMB}}$$

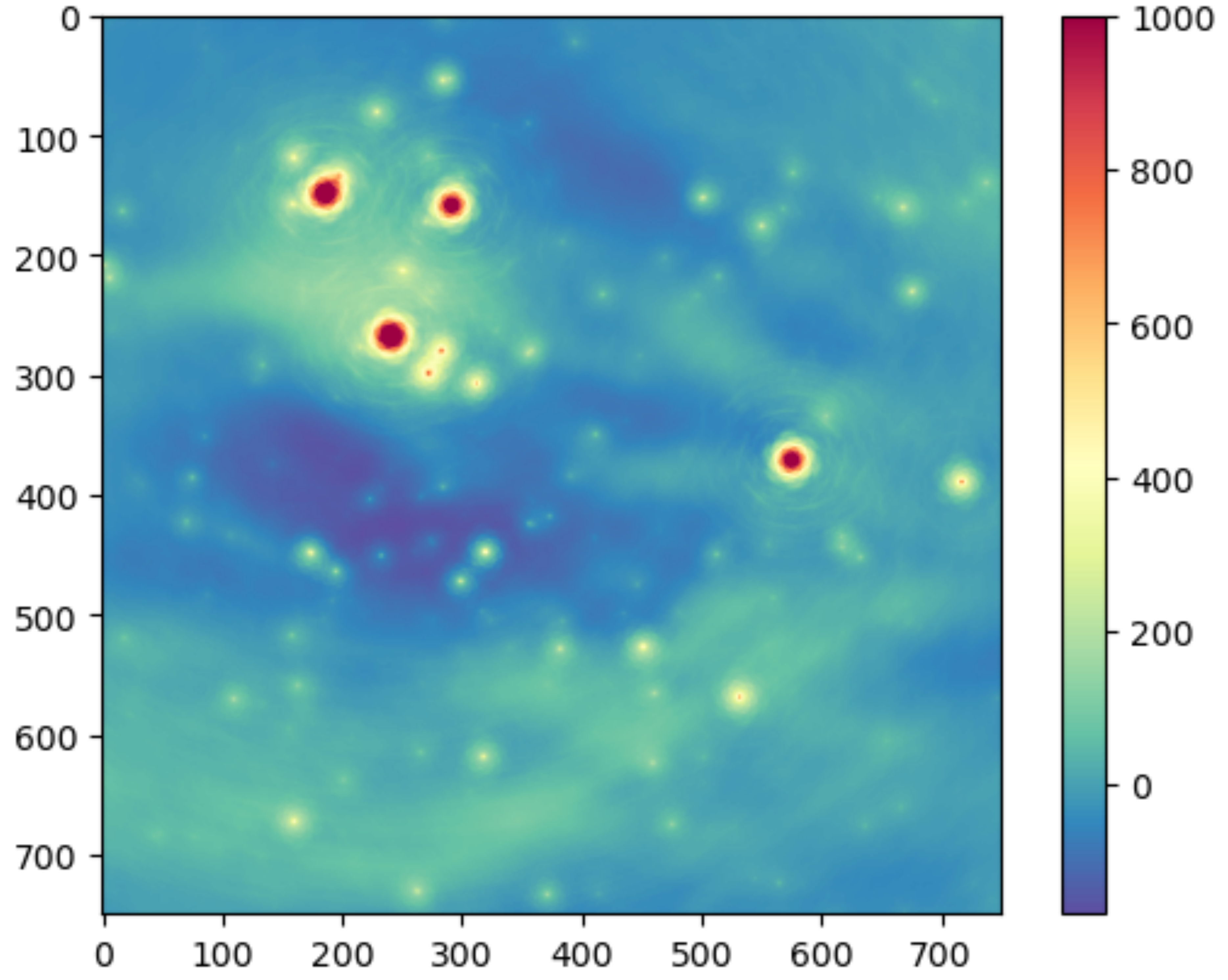
Four primary concerns:

- Physical foreground interpretation (Hills et al. 2019)
- Alternative models and goodness of model fits (Hills et al. 2019)
- Ground plane resonances (Bradley et al. 2019)
- Chromatic beam effects

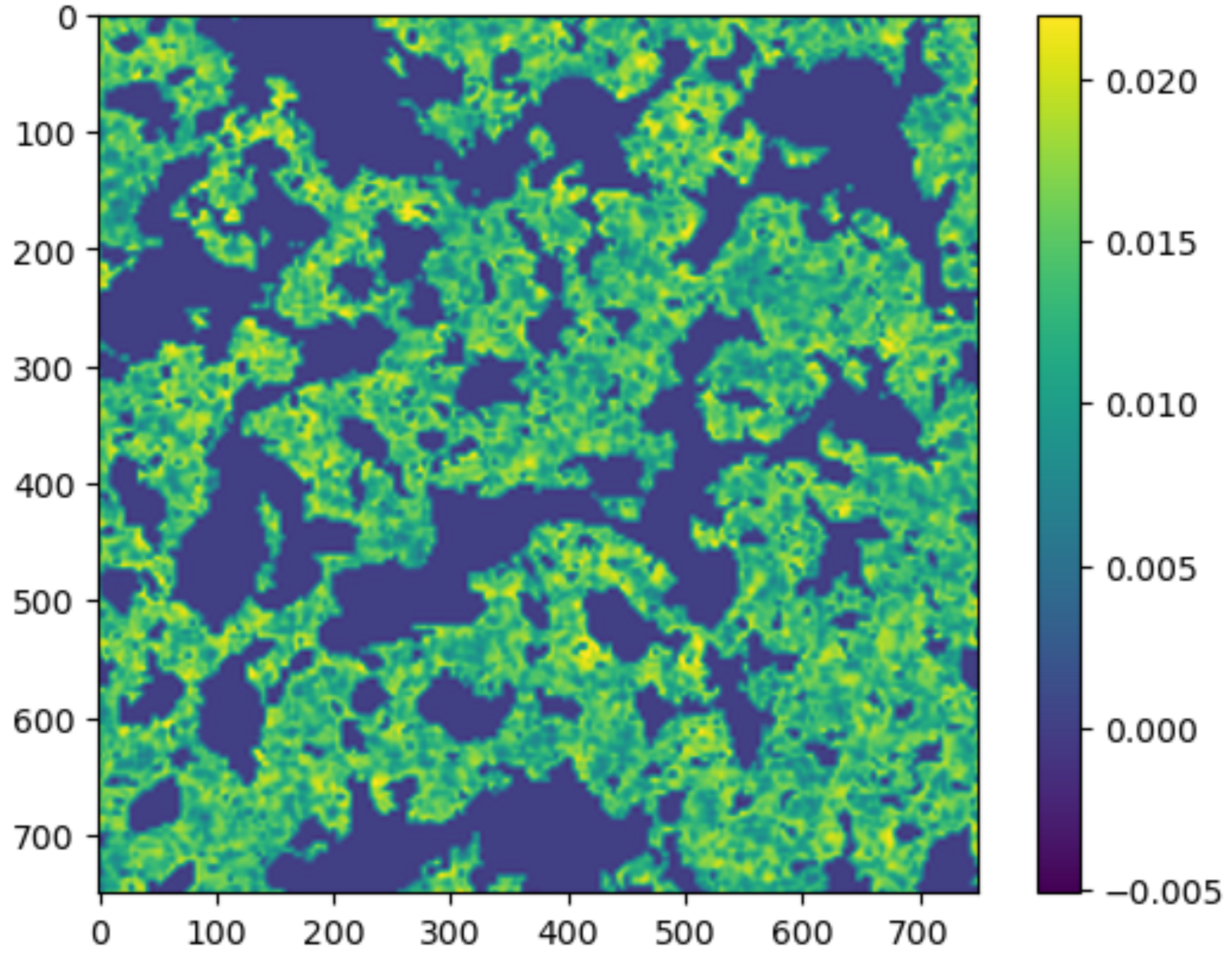
Challenge: FG contamination

- five orders of magnitude brighter
- FG removal accuracy of at least 1 in 10,000 required !!!

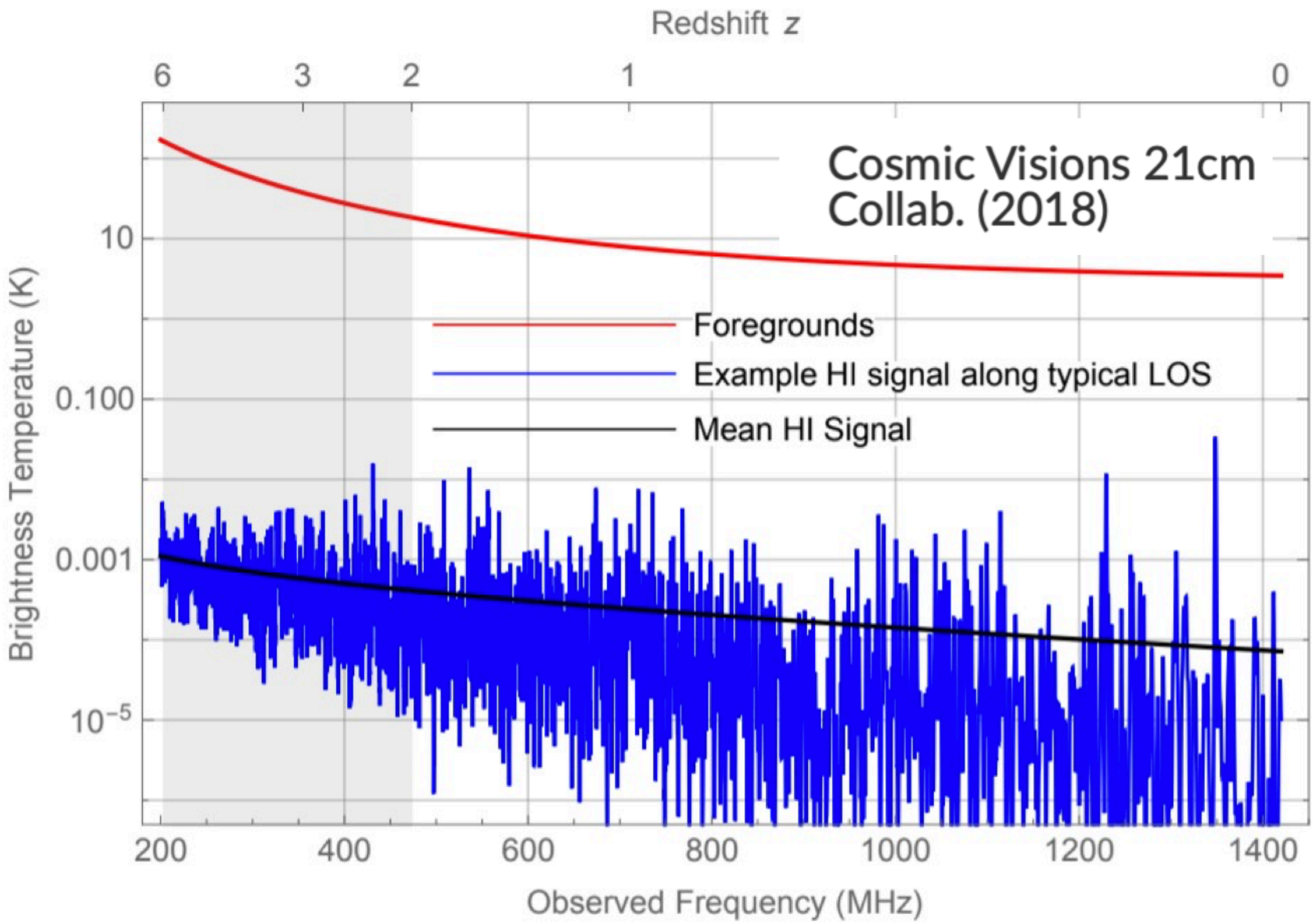
Raw SDC3a image
(natural weighting)

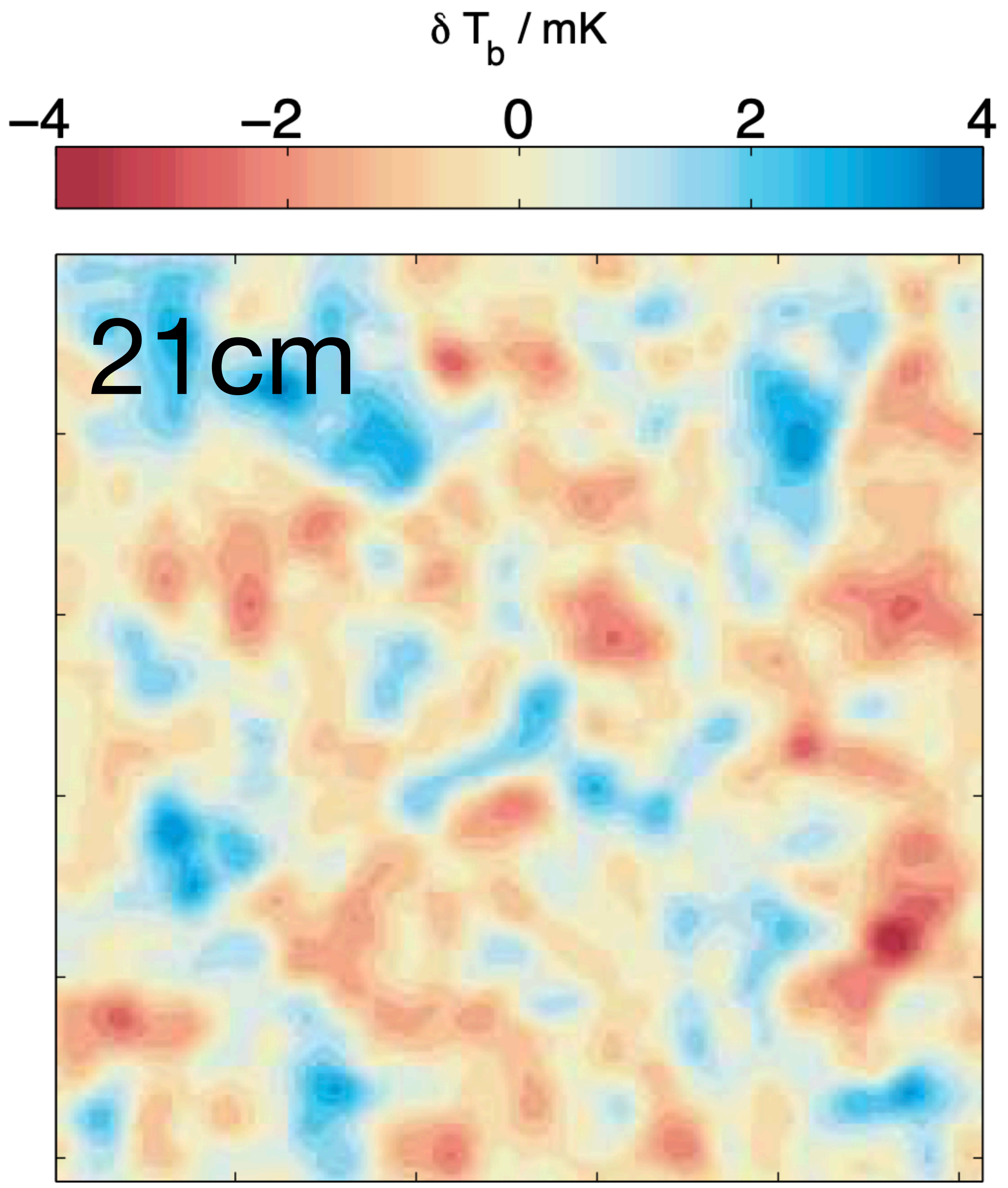


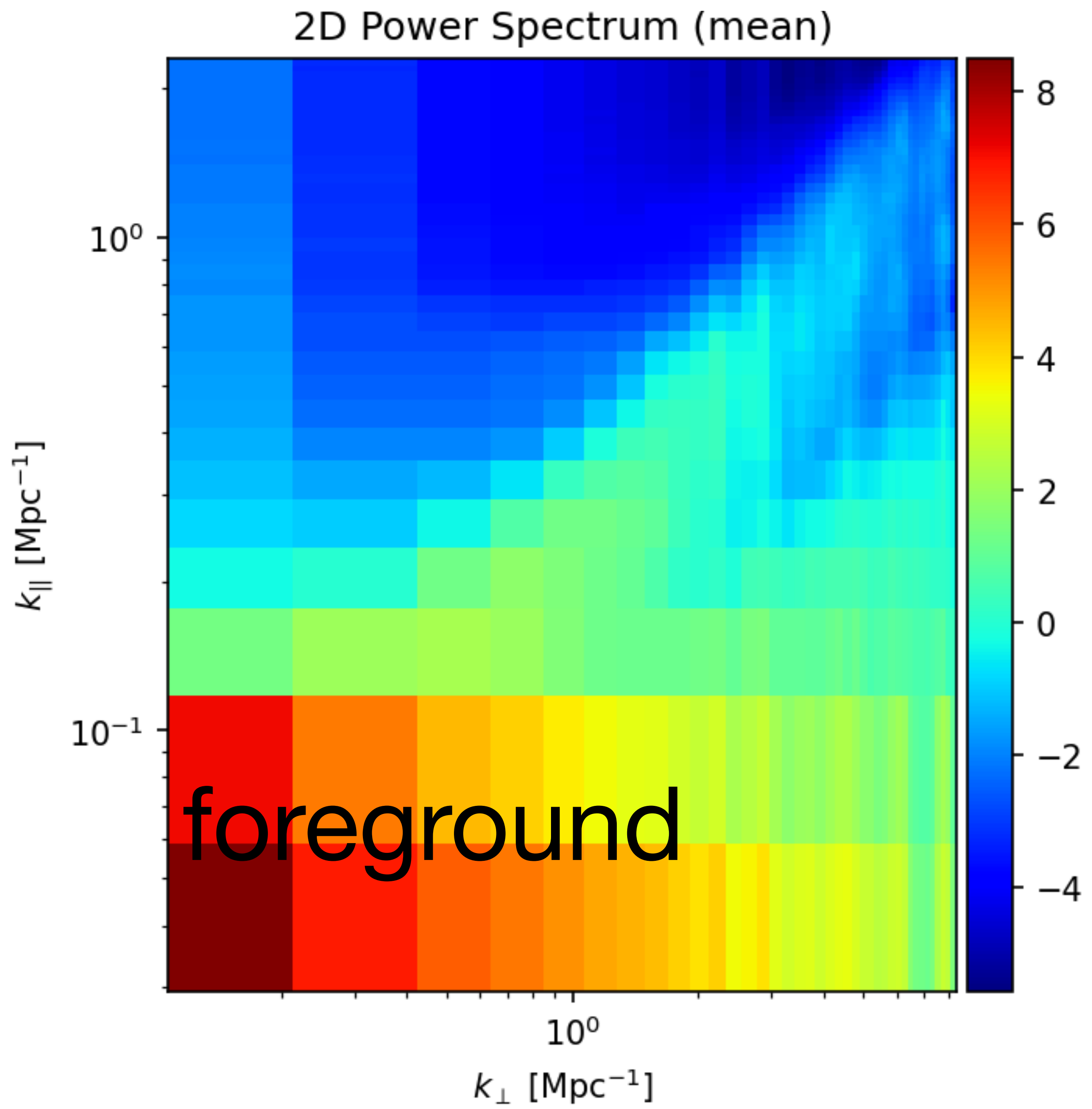
True EOR



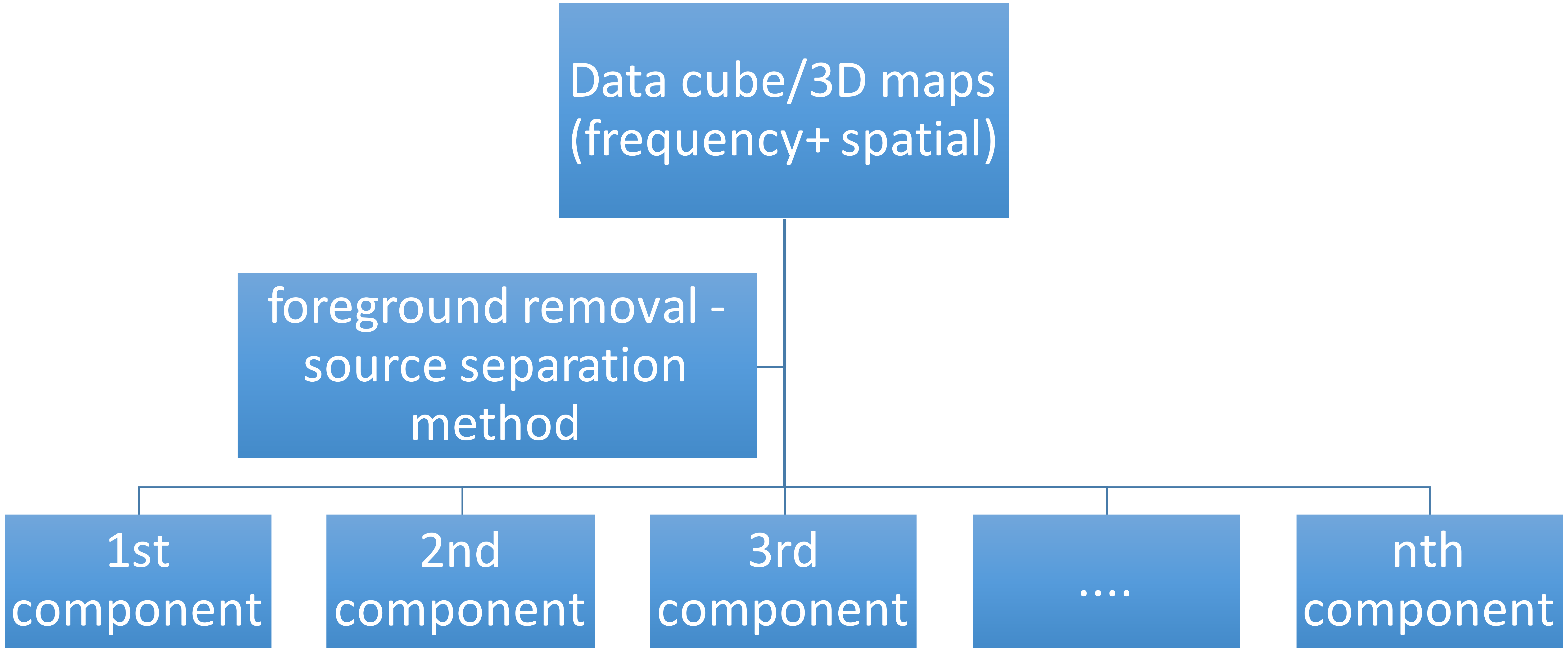
The key to separating out foregrounds: their spectral smoothness







Scientific goal



- Cosmological 21 cm signal

From Physics to Smooth Foreground Model

- Synchrotron radiation is emitted by relativistic electrons spiraling in magnetic fields.
- Brightness temperature approximately follows a power law in frequency: $T(\nu) \propto \nu^{-\alpha}$,where α is the spectral index related to the electron energy distribution.
- This power-law form arises because the electron energy spectrum is itself a power law: $N(E) \propto E^{-p}$

In practice, realistic spectra have small deviations. α varies slightly with frequency due to:

- Energy-dependent electron cooling,
- Superposition of multiple electron populations,
- Magnetic field variations,
- Propagation effects (e.g., absorption, scattering)

This means the spectrum is not exactly a power law, but rather a smoothly curved function of frequency

Mathematical modeling: frequency-dependent spectral index

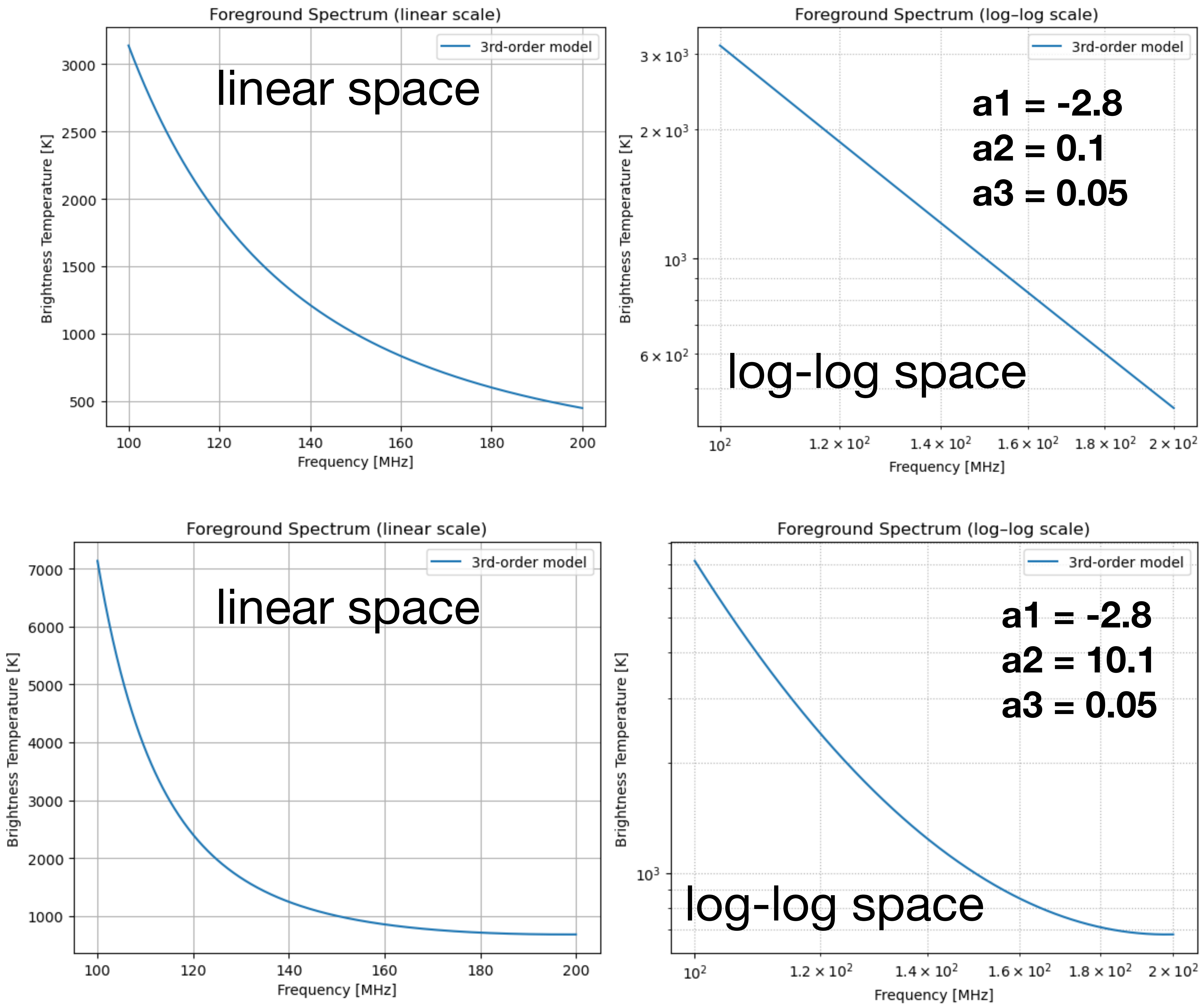
- The brightness temperature is: $T(\nu,\hat{\mathbf{n}}) \approx T_*(\hat{\mathbf{n}})\left(\frac{\nu}{\nu_*}\right)^{\alpha(\hat{\mathbf{n}},\nu)}$
- Real spectra deviate smoothly from pure power laws
- In reality, the spectral index α depends on frequency: $\alpha = \alpha(\nu,\hat{\mathbf{n}})$, with slow variation in ν
- Taylor (polynomial) expansion in log–frequency:

$$\ln T(\nu,\hat{\mathbf{n}}) = a_*(\hat{\mathbf{n}}) + a_1(\hat{\mathbf{n}})\ln(\nu/\nu_*) + \frac{1}{2}a_2(\hat{\mathbf{n}})\left[\ln(\nu/\nu_*)\right]^2 + \dots$$

- Return to linear (up to the 3rd order) scale:

$$T(\nu,\hat{\mathbf{n}}) = T_*(\hat{\mathbf{n}})\left(\frac{\nu}{\nu_*}\right)^{a_1(\hat{\mathbf{n}})}\exp\left[\frac{1}{2}a_2(\hat{\mathbf{n}})\left(\ln\frac{\nu}{\nu_*}\right)^2 + \frac{1}{6}a_3(\hat{\mathbf{n}})\left(\ln\frac{\nu}{\nu_*}\right)^3\right]$$

some examples:



Frequency covariance:

$$T(\nu, \hat{\mathbf{n}}) = T_f(\hat{\mathbf{n}})\left(\frac{\nu}{\nu_f}\right)^{-\alpha(\hat{\mathbf{n}})}$$

Then the covariance of temperature is related to :

$$\langle T(\nu)T(\nu')\rangle = \left\langle A^2\left(\frac{\nu\nu'}{\nu_f^2}\right)^{-\alpha}\right\rangle = A^2\left(\frac{\nu\nu'}{\nu_f^2}\right)^{-\bar{\alpha}}\left\langle \exp\left[-\delta\alpha\cdot\ln\left(\frac{\nu\nu'}{\nu_f^2}\right)\right]\right\rangle$$

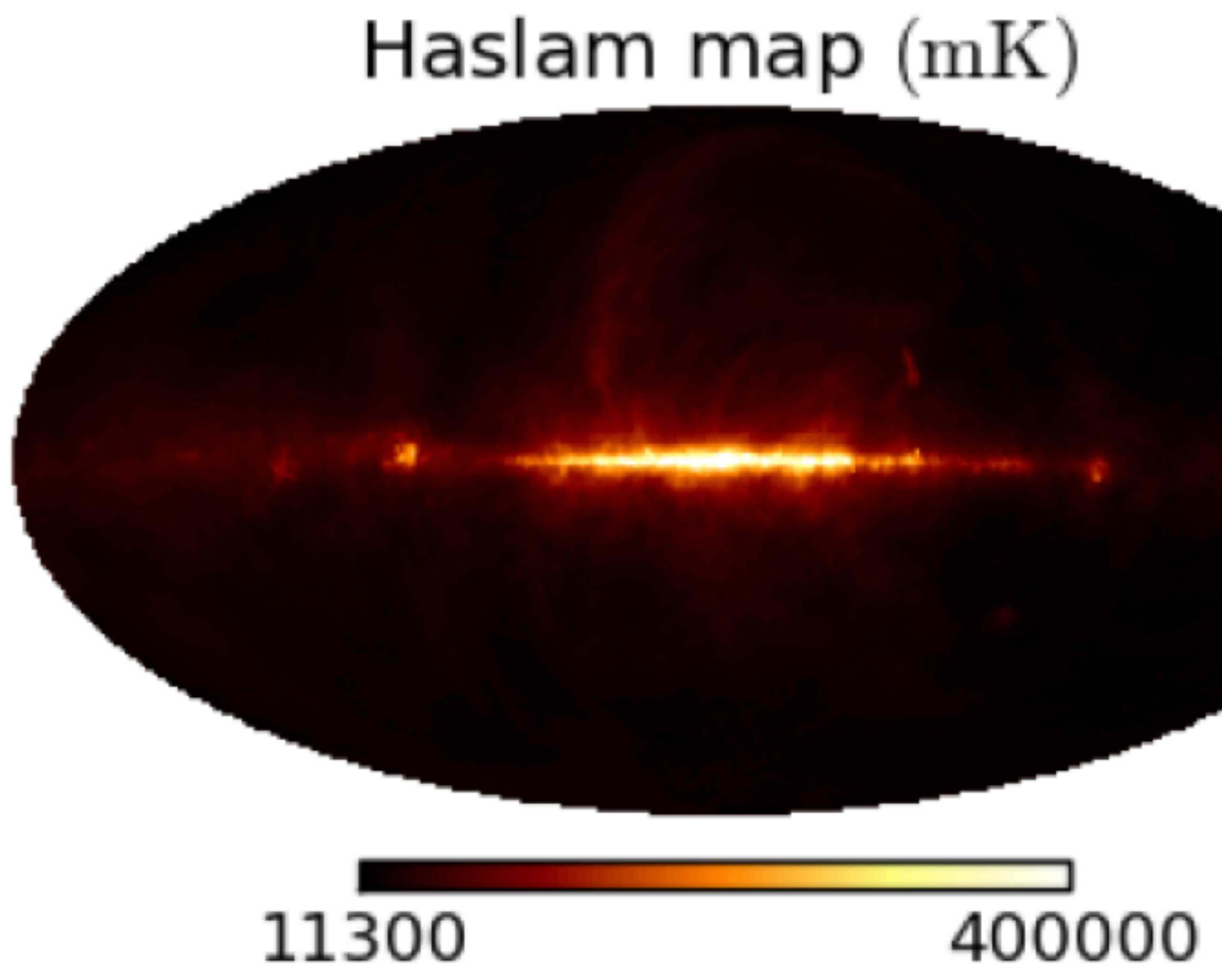
So the expectation becomes (for Gaussian pdf): $\left\langle e^{-\delta\alpha\cdot\ln\left(\nu\nu'/\nu_f^2\right)}\right\rangle = \exp\left(\frac{1}{2}\sigma_\alpha^2\left[\ln\left(\frac{\nu\nu'}{\nu_f^2}\right)\right]^2\right)$

with some approximations, one can find $\mathrm{Cov}(\nu, \nu') \propto \exp\left(-\frac{\ln^2(\nu/\nu')}{2\xi^2}\right)$

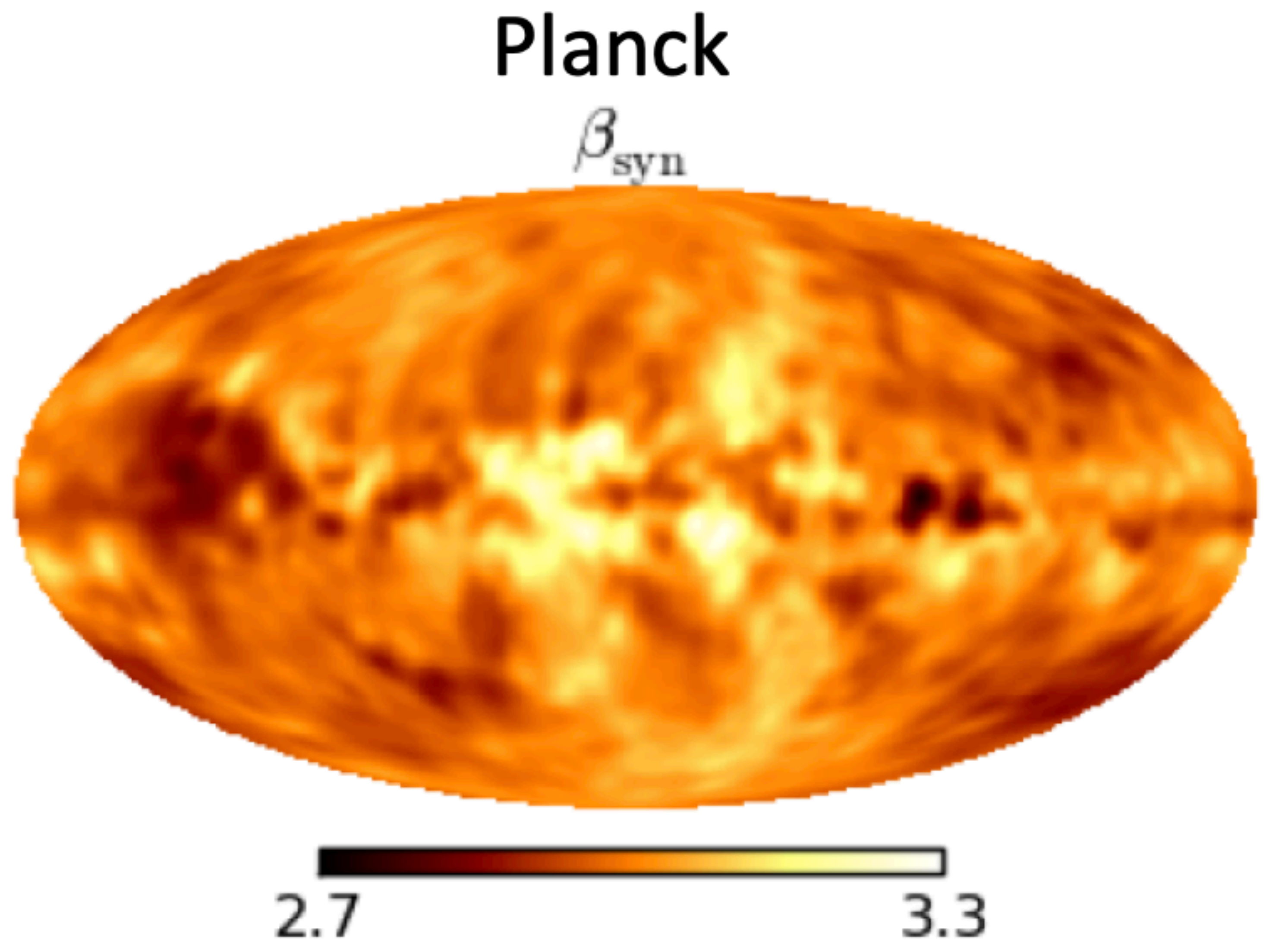
21cm Foreground removal methods

Type	**Reference**	**Methods**
Foreground Avoidance	Datta et al . 2010 Trott et al. 2012 Morales et al. 2012 Parsons et al. 2013 Pober et al. 2013 Liu et al. 2014a,b	EOR Window, delay spectrum
Foreground suppression	Zaldarriaga et al 2004; McQuinn et al. (2006); Morales et al. 2006; Gleser et al. 2008; Jelic et al. 2008; Bowman et al. 2006; Liu et al. 2009; Datta et al. 2009; Petrovic & Oh (2011), Datta et al. 2010, Trott et al. 2012, Morales et al. 2012, Liu et al. 2008, Harker et al. 2008, 2018, Chapman et al. 2013, Chapman et al. 2012, 2016, Zhang 2015, Bonaldi & Brown 2015, Shaw 2015, Mertens 2018, Makinen 2021, Tauscher 2018, 2020, Rapetti 2020, Kennedy 2023,…	Power-law fitting, PCA/SVD, FASTICA, WP smoothing, CCA, GMCA, HIEMICA, K-L transform, Gaussian Process, Deep learning, Gibbs sampling, Eigenanalysis,…

A Fast Simulation: Galactic synchrotron



+



Delabrouille et al. 2013 (PSM)

+ Small scale fluctuations

408 MHz

$$T_{\rm sync}(\nu, \hat{\mathbf{n}}) = T_{408}(\hat{\mathbf{n}}) \left(\frac{\nu_H}{\nu} \right)^{\alpha(\hat{\mathbf{n}})} + \Delta T_{\rm Gaussian}(\nu, \hat{\mathbf{n}})$$

Santos, Cooray, Knox 2003

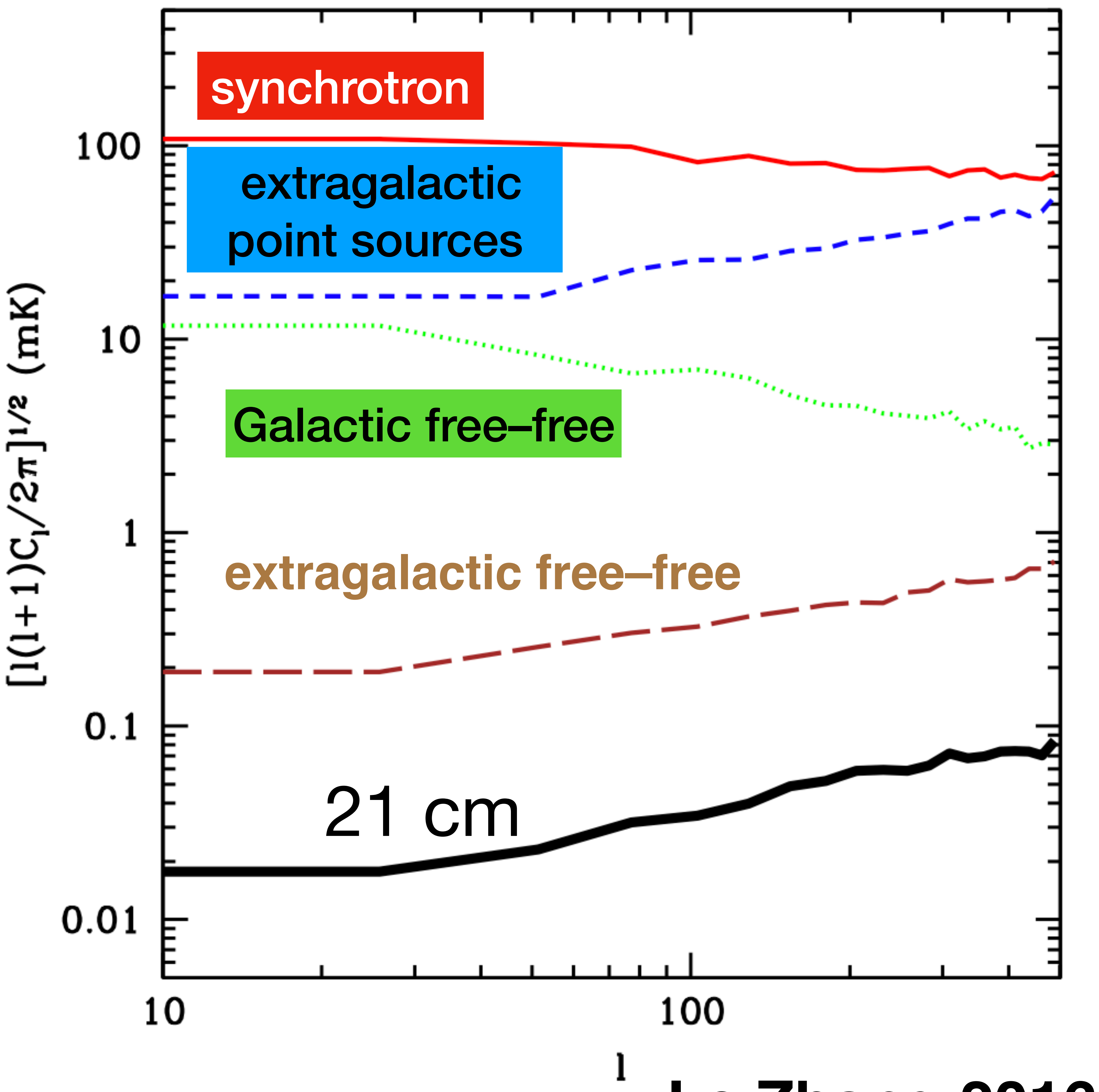
let’s simulate them step-by-step

Foreground simulations: isotropic components

$$C_f(\ell,\nu,\nu') = A\left(\frac{1000}{\ell}\right)^{\beta}\left(\frac{\nu_f^2}{\nu\nu'}\right)^{2\alpha}\exp\left(-\frac{\ln^2(\nu/\nu')}{2\xi^2}\right)$$

Santos et al. (2005)	$A\,(\mathrm{mK}^2)$	β	α	ξ	$\Delta\ \alpha$
Extragalactic point sources	57.0	1.1	2.07	1.0	0.2
Extragalactic free–free	0.014	1.0	2.10	35	0.03
Galactic synchrotron	700	2.4	2.80	4.0	0.15
Galactic free–free	0.088	3.0	2.15	35	0.03

830 MHz



Le Zhang 2016

Case I: a random fluctuations

$$T_{\text{sync}}(\nu, \hat{\mathbf{n}}) = T_{408}(\hat{\mathbf{n}}) \left(\frac{\nu_H}{\nu} \right)^{\alpha(\hat{\mathbf{n}})} + T_{\text{Gaussian}}(\nu, \hat{\mathbf{n}})$$

- **Frequency range: 100–200 MHz (relevant for Epoch of Reionization studies)**
- **Brightness model: power-law dependence with spatially varying amplitude and spectral index**
- **Output: a data cube Tb (v,x,y) representing the brightness temperature in Kelvin**

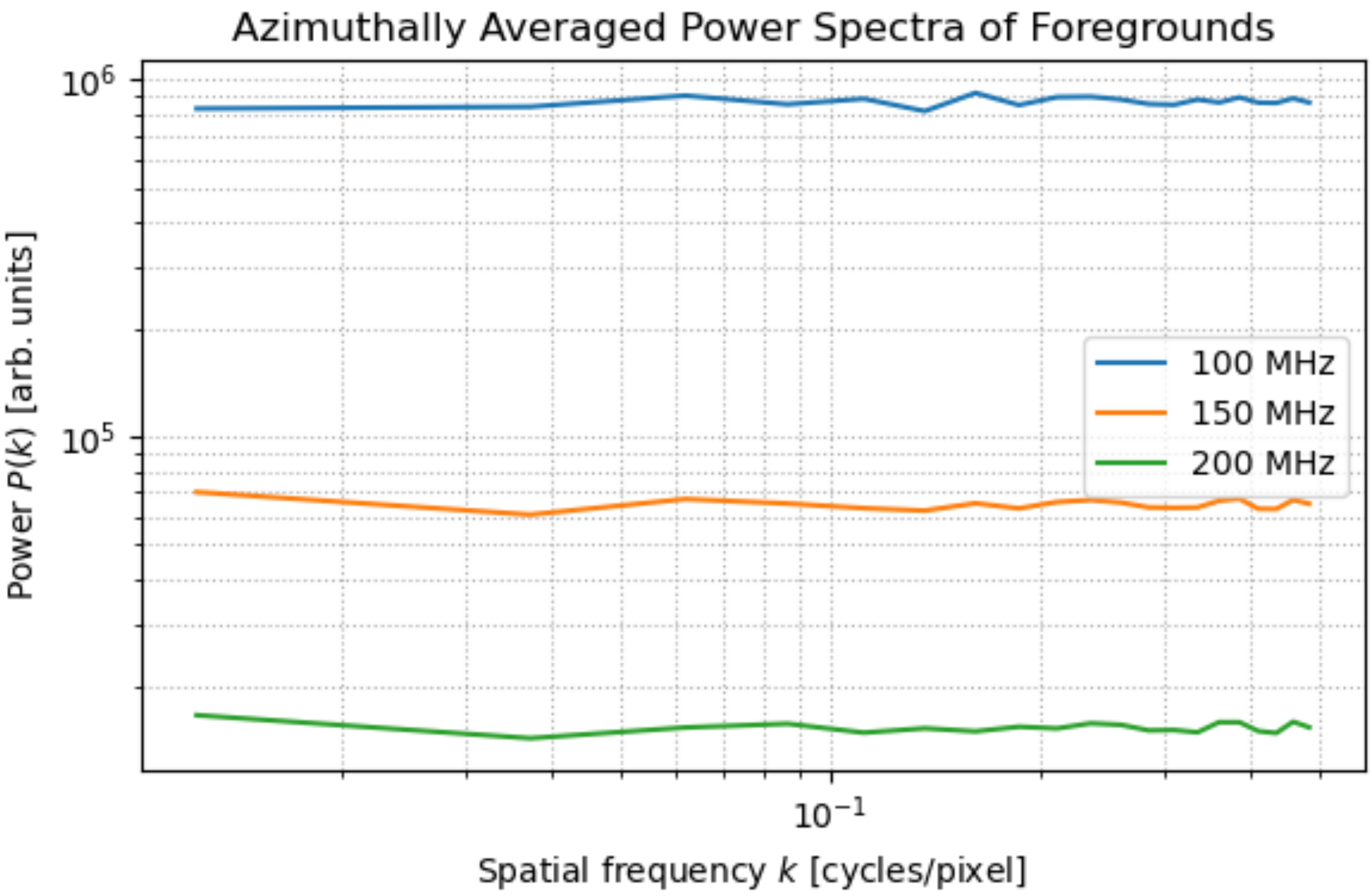
The emission spectrum follows a power law:

$$T_{\text{Gaussian}}(\nu, x, y) = A(x, y) \left(\frac{\nu}{150\text{MHz}} \right)^{\alpha(x,y)}$$

- **The amplitude map $A(x, y)$ follows a Gaussian distribution**
- **The spectral index $\alpha(x, y)$ varies spatially around a mean of -2.8 , as seen in real synchrotron dominated sky surveys**

Temperature maps

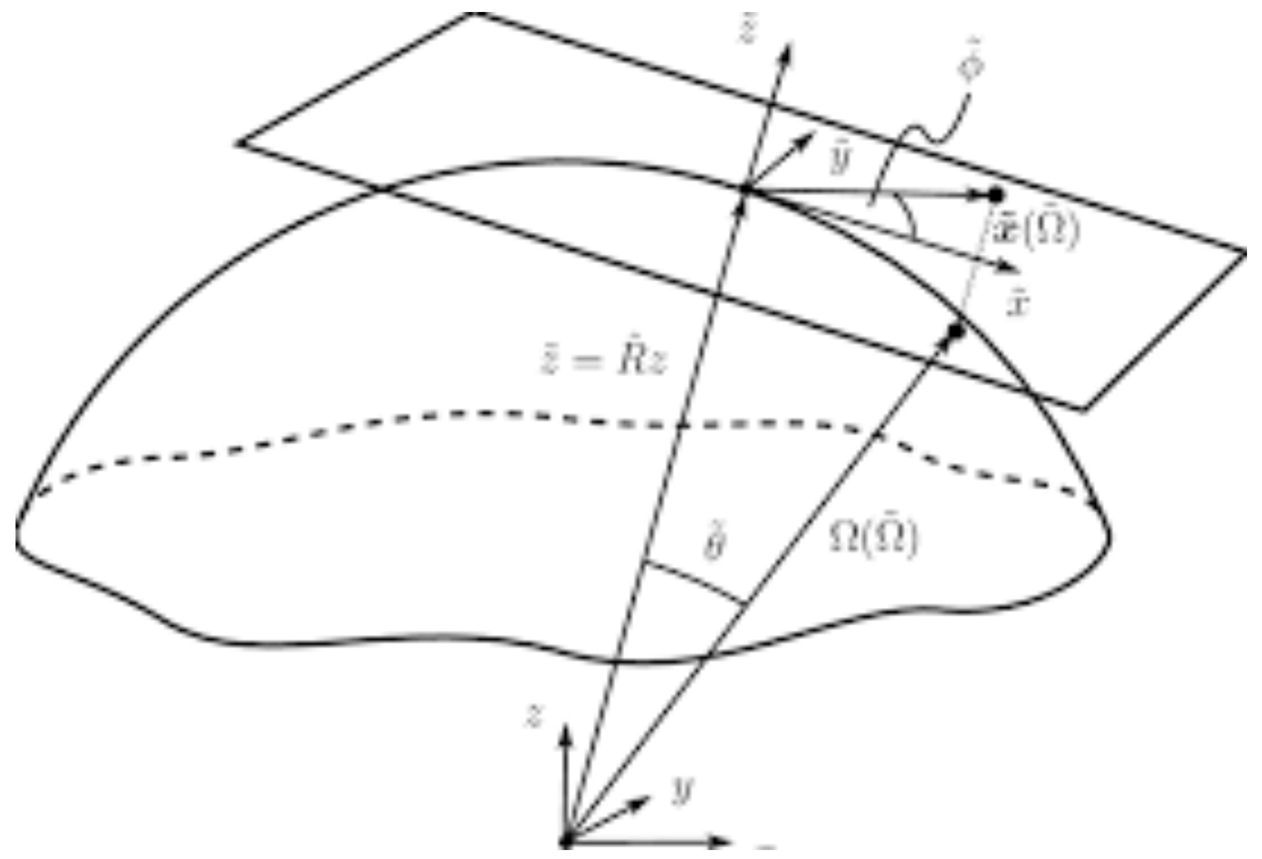
Flat angular power spectra (like shot noise)

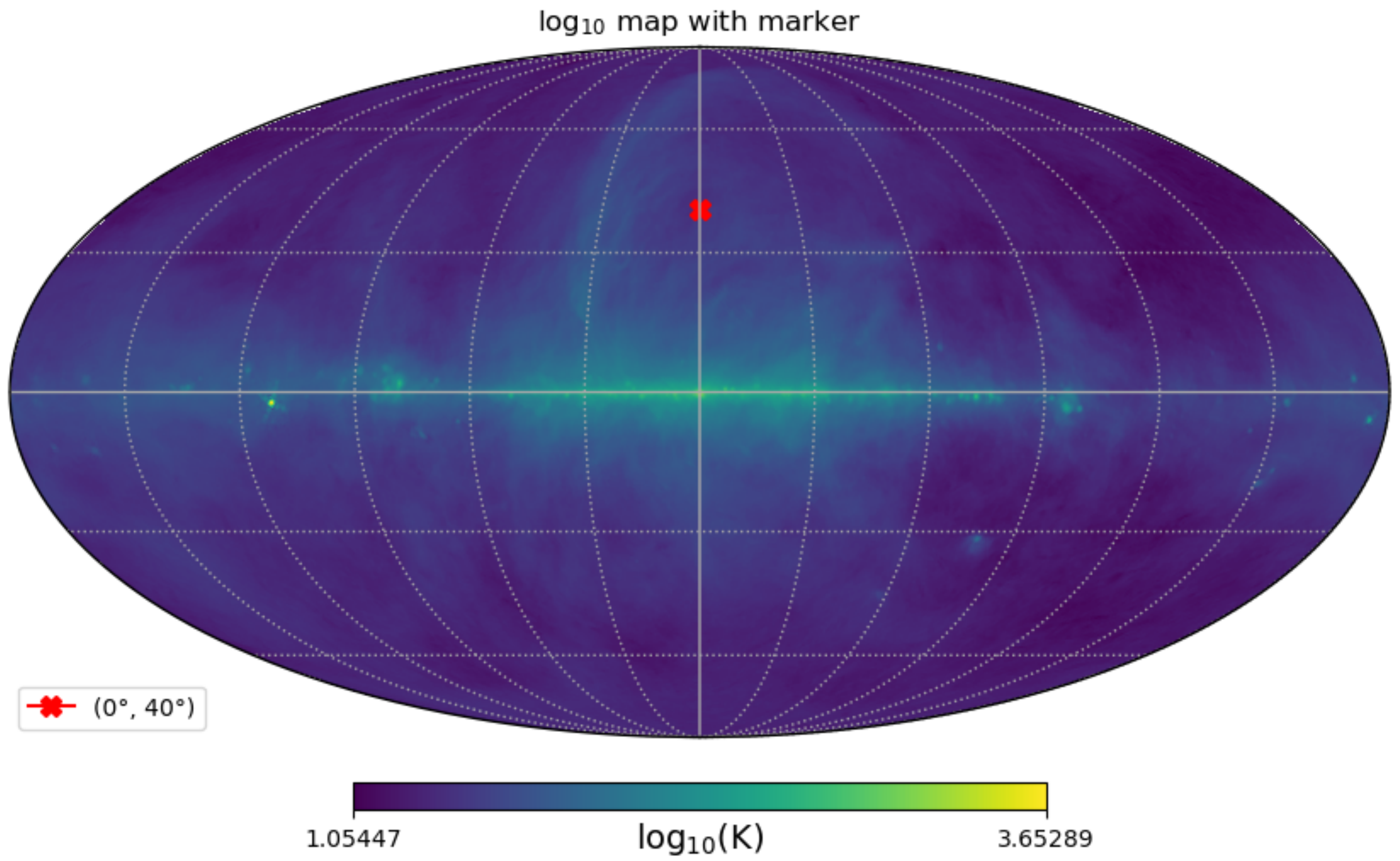


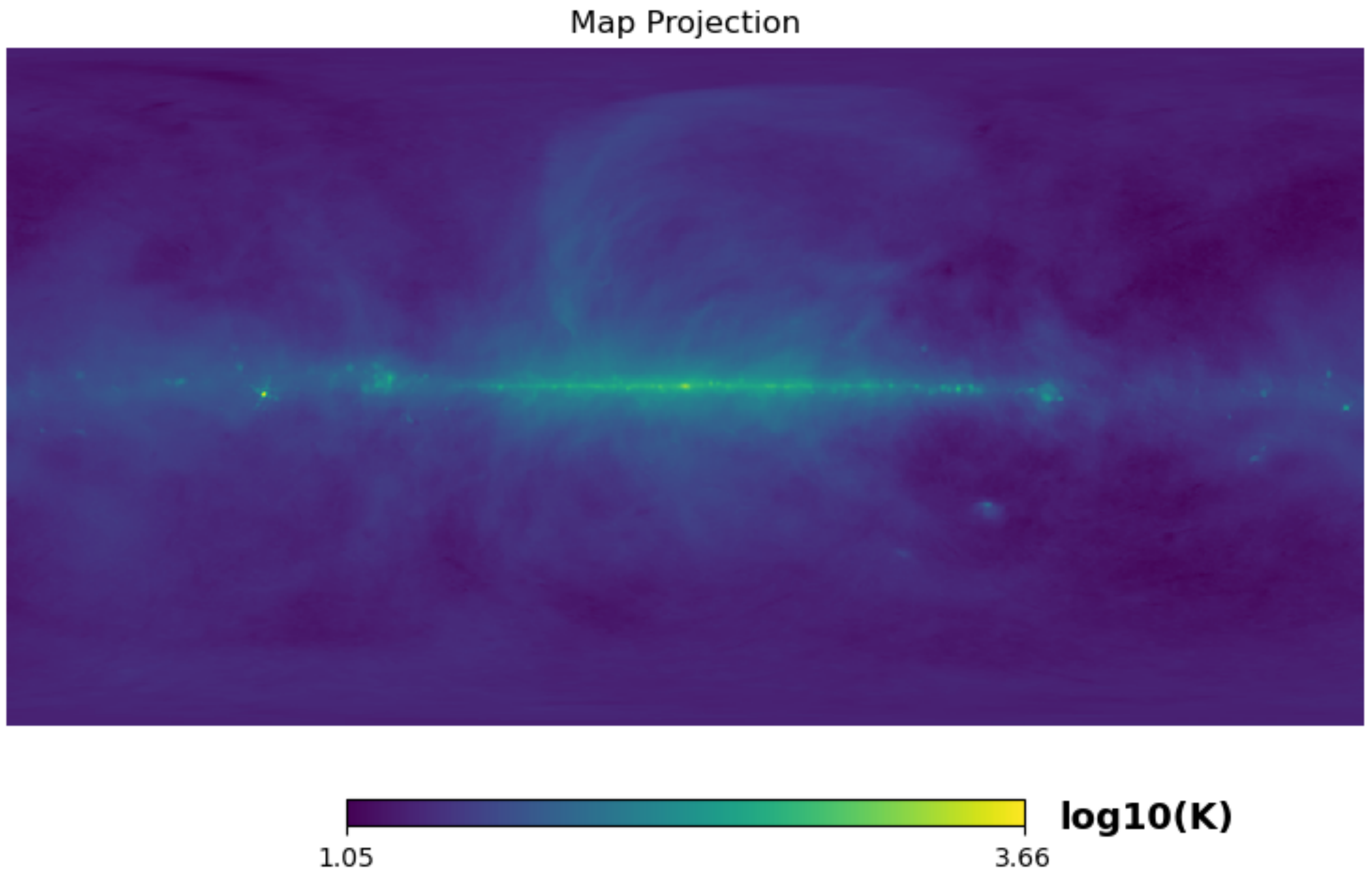
Extrapolate Haslam 408 MHz to …

$$T_{\rm sync}(\nu, \hat{\mathbf{n}}) = T_{408}(\hat{\mathbf{n}}) \left(\frac{\nu_H}{\nu}\right)^{\alpha(\hat{\mathbf{n}})} + T_{\rm Gaussian}(\nu, \hat{\mathbf{n}})$$

- project Haslam map to a flat sky



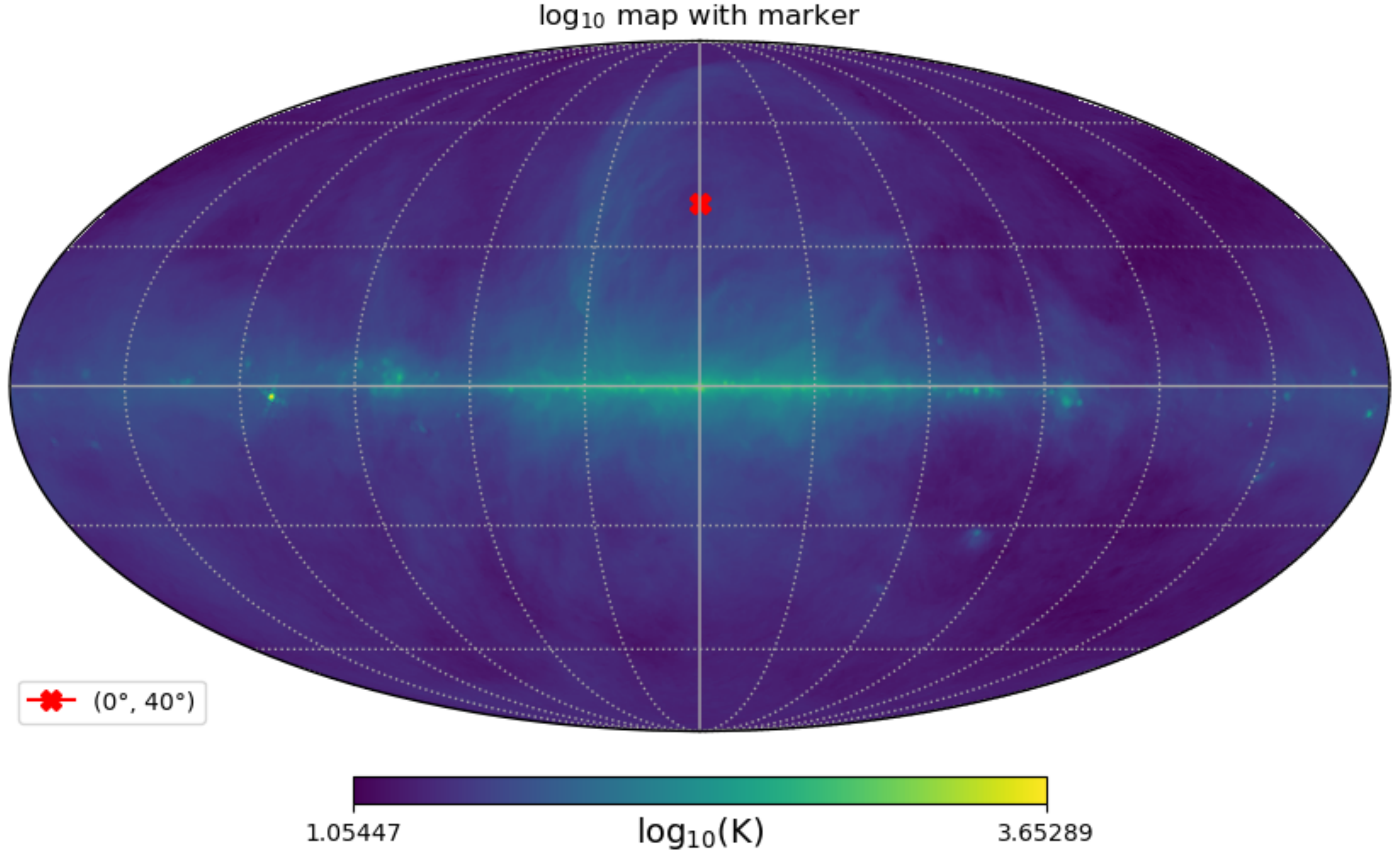


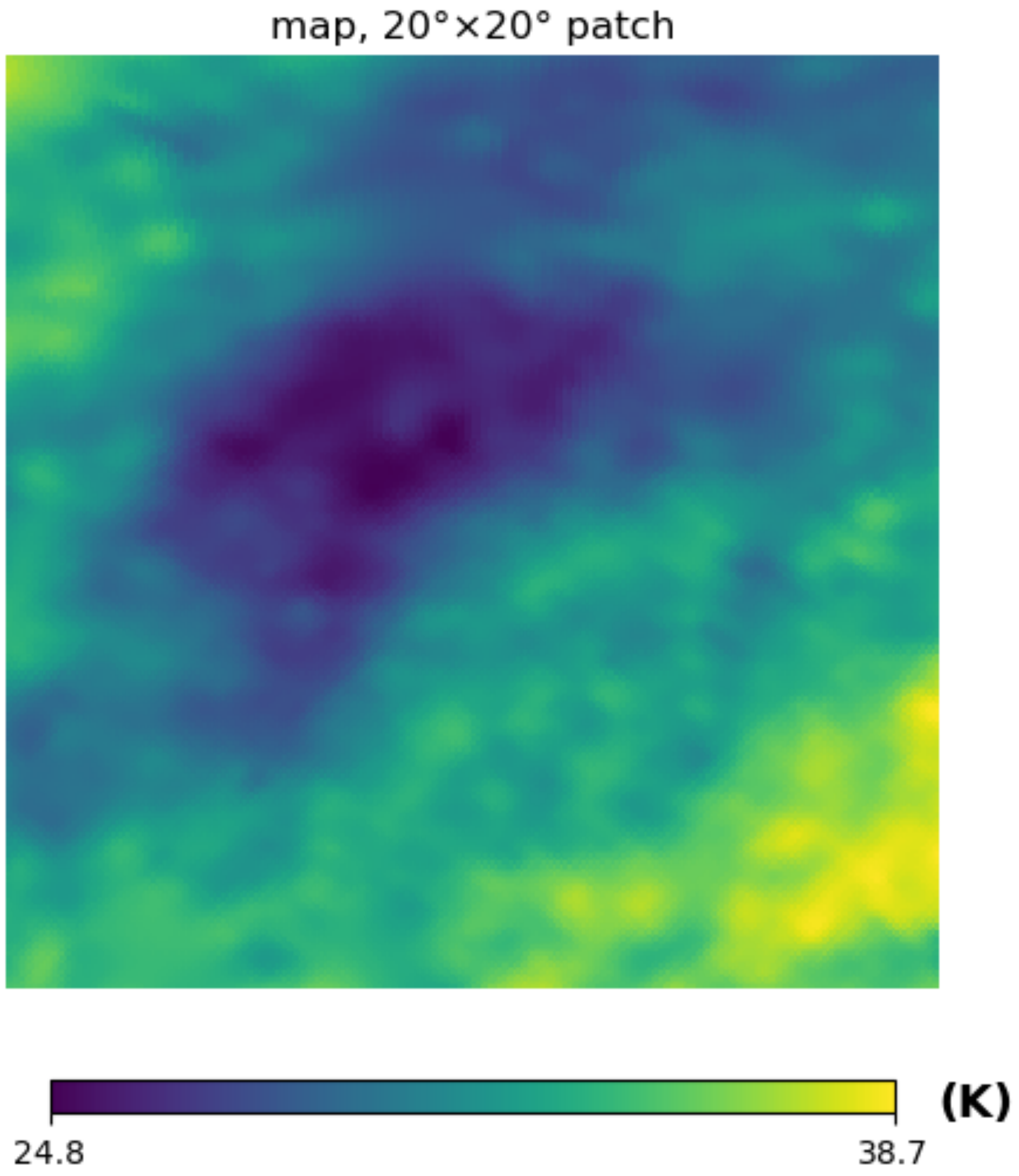


Extrapolate Haslam 408 to …

$$T_{\rm sync}(\nu, \hat{\mathbf{n}}) = T_{408}(\hat{\mathbf{n}})\left(\frac{\nu_H}{\nu}\right)^{\alpha(\hat{\mathbf{n}})} + T_{\rm Gaussian}(\nu, \hat{\mathbf{n}})$$

- project a small region to a flat patch (e.g., $20^\circ \times 20^\circ$)

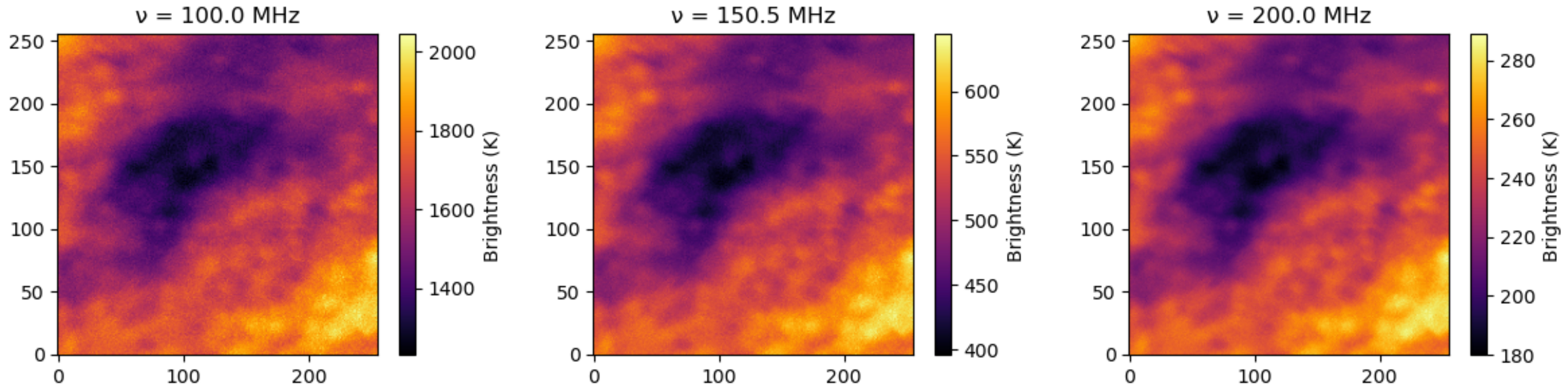




$$T_{\mathrm{sync}}(\nu, \hat{\mathbf{n}}) = T_{408}(\hat{\mathbf{n}}) \left(\frac{\nu_H}{\nu} \right)^{\alpha(\hat{\mathbf{n}})}$$

$\alpha = -2.8 + G(0,\ \sigma_\alpha)$ we assume $\sigma_\alpha = 0.01$

Alternatively, using a “spatially-varying” α-map (e.g. from Planck or simulations)



Blind foreground subtraction

- **LOS fitting: choose ad-hoc smooth functions. Usually polynomial fitting in log-log space.**
- **PCA: uncorrelated sources maximizing the variance. Diagonalize $\nu - \nu'$ covariance and subtract principal eigenvectors.**
- **ICA: independent sources maximizing the variance. Find independent sources by maximizing non-Gaussianity.**
- **GPR: Bayesian regression with smoothness priors; Learn smooth foreground functions per line-of-sight via a kernel**

(See also: Gleser et al. 2008, Liu et al. 2009, Ricciardi et al., 2010, Harker et al. 2009, Hyvärinen et al. 1999, Chapman et al. 2012, Wolz et al. 2013, Chapman et al. 2013)

- **Issues:**
- Mode mixing of angular and frequency fluctuations by frequency-dependent baselines (esp. interferometers); gain fluctuations…
- Robustness biasing introduced if foreground model poorly understood (esp. non-Gaussianities) although excellent results.
- Statistical optimality need to keep track of transformations on statistics, for optimal estimation

Polynomial Fitting in Log-Log Space

Fit and subtract smooth astrophysical foregrounds from frequency spectra using a polynomial model in log-log space, taking advantage of the fact that:

$$\ln T(\nu, \hat{\mathbf{n}}) = a_*(\hat{\mathbf{n}}) + a_1(\hat{\mathbf{n}})\ln(\nu/\nu_*) + \frac{1}{2}a_2(\hat{\mathbf{n}})\left[\ln(\nu/\nu_*)\right]^2 + \ldots$$

Given data:

Brightness temperature: $T(\nu)$ **in Kelvin**

Frequencies: $\nu_i \in \left[\nu_{\mathbf{min}}, \nu_{\mathbf{max}}\right]$ **- Choose a reference frequency** ν_0

Define: $x_i = \ln(\nu_i/\nu_0)$, $y_i = \ln T\left(\nu_i\right)$

a power law (up to the 1st order): $T(\nu) = T_0\left(\nu/\nu_0\right)^\alpha \Rightarrow y_i = \ln T_0 + \alpha \cdot x_i$ **; a straight line in log-log space**

Generalize to a Polynomial

To model spectral curvature, generalize the model to a polynomial:

$$y_i = \ln T\left(\nu_i\right) \approx \sum_{k=0}^{n} c_k \cdot x_i^k$$

This models:

- **Smooth foregrounds**
- **Pixel-by-pixel spectral index variations**
- **Higher-order spectral curvature**

Least-Squares Fit Using Vandermonde Matrix

$x_i = \ln(\nu_i/\nu_0)$ n for the n-th coefficient

$$\mathbf{V} = \begin{bmatrix} x_1^{n-1} & x_1^{n-2} & \cdots & x_1^1 & 1 \\ x_2^{n-1} & x_2^{n-2} & \cdots & x_2^1 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ x_{N_\nu}^{n-1} & x_{N_\nu}^{n-2} & \cdots & x_{N_\nu}^1 & 1 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} c_{n-1,1} & c_{n-1,2} & \cdots & c_{n-1,N} \\ c_{n-2,1} & c_{n-2,2} & \cdots & c_{n-2,N} \\ \vdots & \vdots & & \vdots \\ c_{1,1} & c_{1,2} & \cdots & c_{1,N} \\ c_{0,1} & c_{0,2} & \cdots & c_{0,N} \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,N} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,N} \\ \vdots & \vdots & & \vdots \\ y_{N_\nu,1} & y_{N_\nu,2} & \cdots & y_{N_\nu,N} \end{bmatrix}$$

- **Solve for the linear equation** $\mathbf{V}^{N_\nu \times n}\mathbf{C}^{n \times N} = \mathbf{Y}^{n_\nu \times N}$
- **best-fit coefficients (solve for the the linear system):**

$$\mathbf{C} = \left(V^\top V\right)^{-1} V^\top \cdot \mathbf{Y} = V^\dagger \mathbf{Y}$$

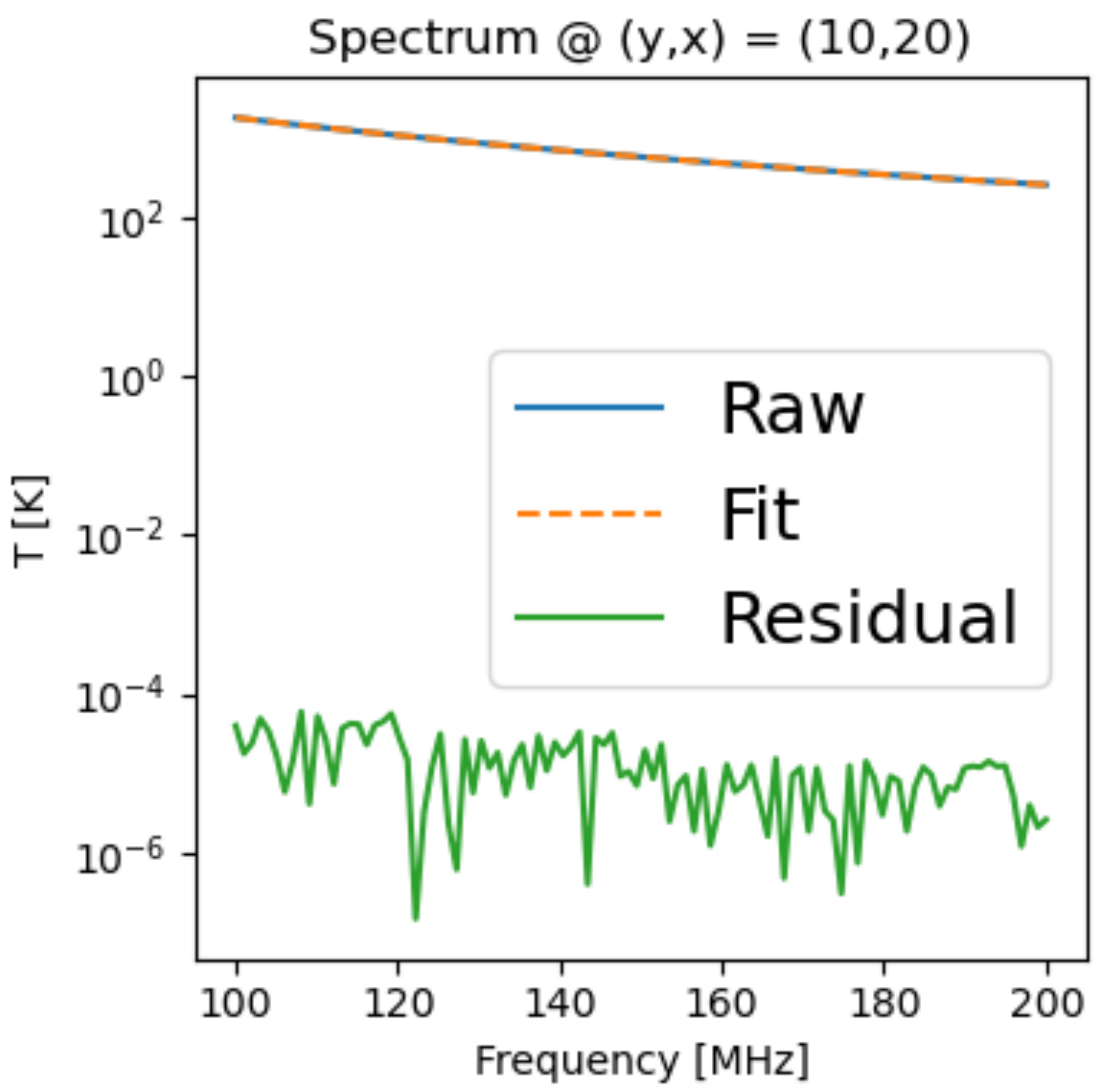
Reconstruct and Subtract

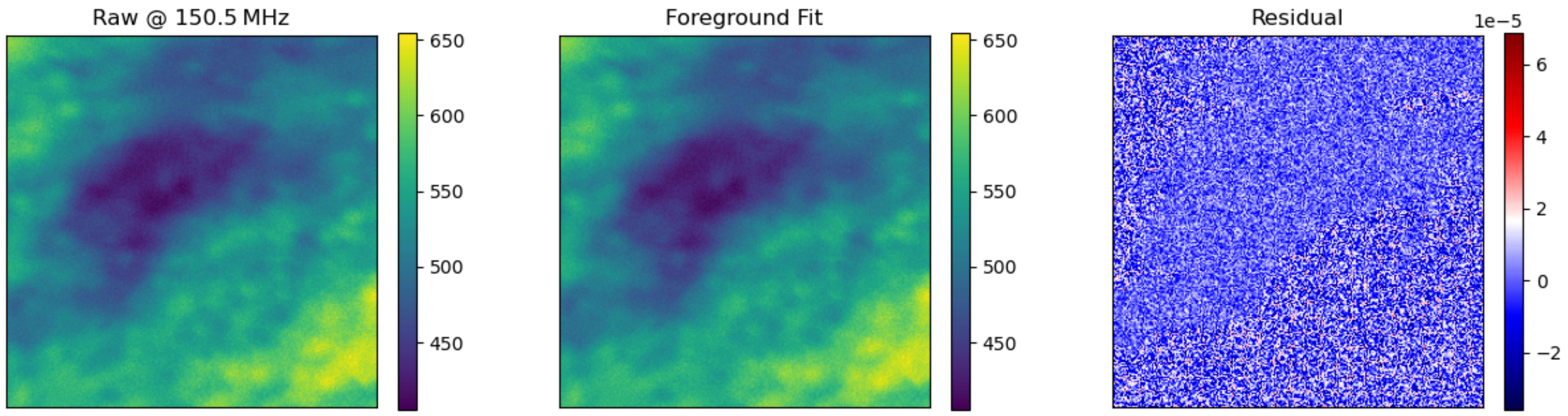
1. Evaluate the fitted model:

$$\widehat{\mathbf{Y}} = \mathbf{V} \cdot \mathbf{C} \quad \Rightarrow \quad \widehat{T}(\nu) = \exp(\widehat{y})$$

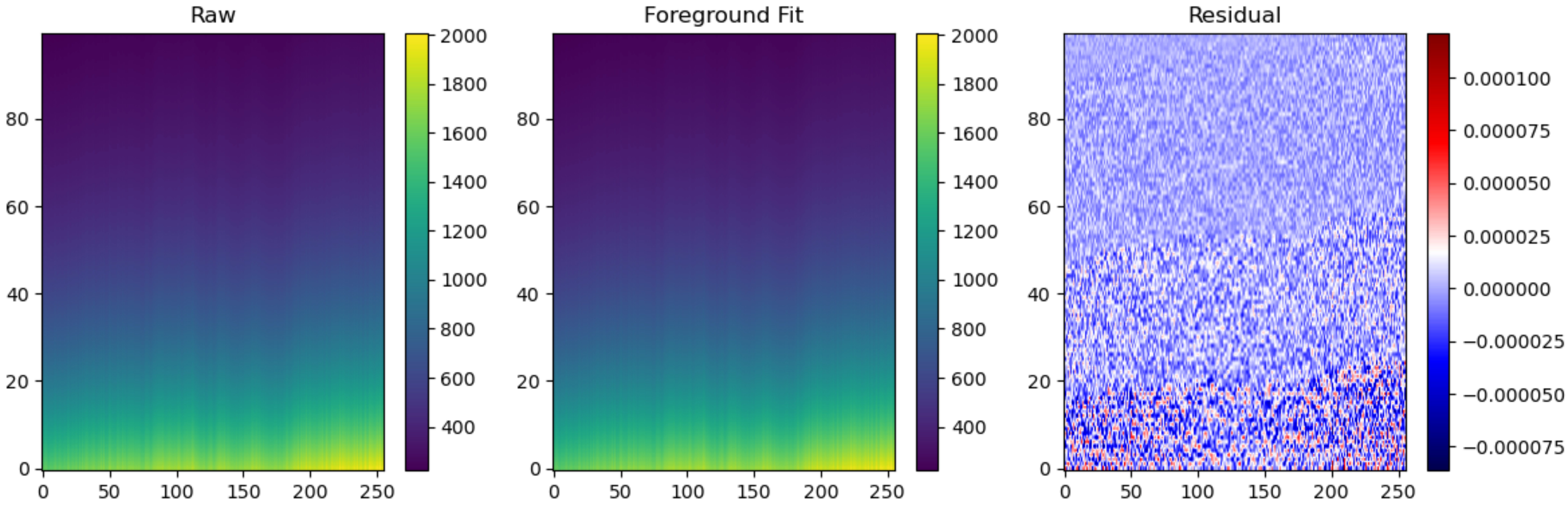
2. Subtract to get residuals (see residuals amplitude):

$$T_{\mathrm{resid}}(\nu) = T_{\mathrm{obs}}(\nu) - \widehat{T}(\nu)$$



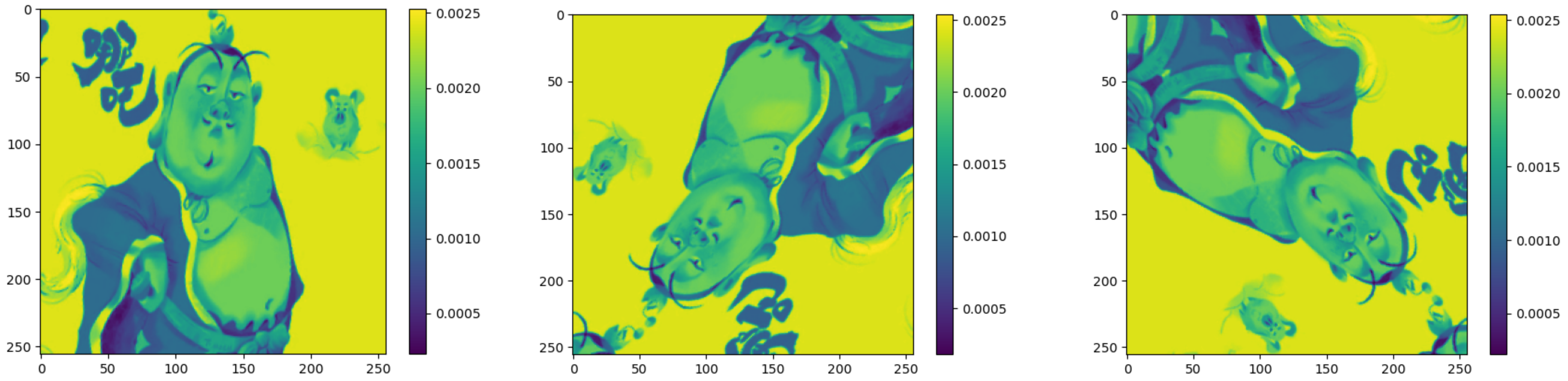


Residuals along LOS



Simulating a Toy 21cm Data Cube with Random Image Rotations

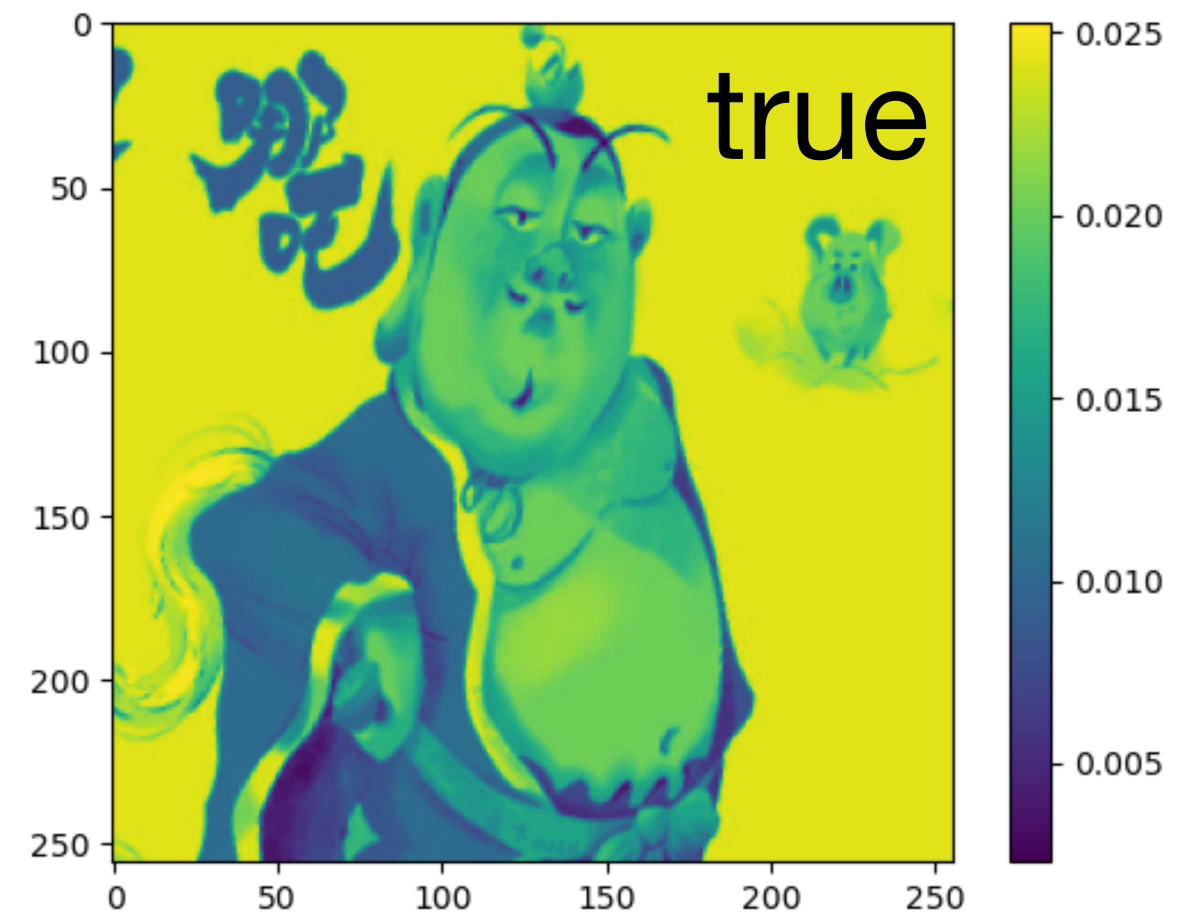
- Generate a mock 3D 21cm data cube by treating a 2D image as a spatial slice and simulating redshift evolution via random rotations along the frequency axis

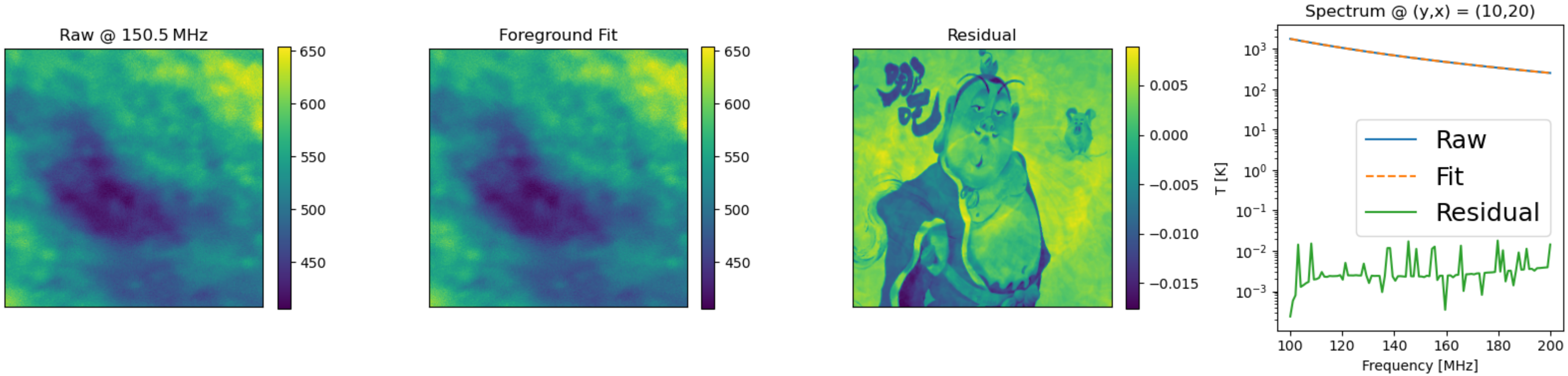


- More realistic case: run 21cmFAST (https://21cmfast.readthedocs.io)

Adding EoR to data cube…

Pearson correlation coefficient r=0.95





Foreground Removal via Singular Value Decomposition (SVD)

Why Use SVD for Foreground Removal?

- Foregrounds are spectrally smooth, but not strictly power-law

- Polynomial fits require a predefined functional form

- This smoothness means foreground spectra occupy a low-dimensional subspace in frequency space

- Foregrounds dominate the largest coherent structures in the data matrix

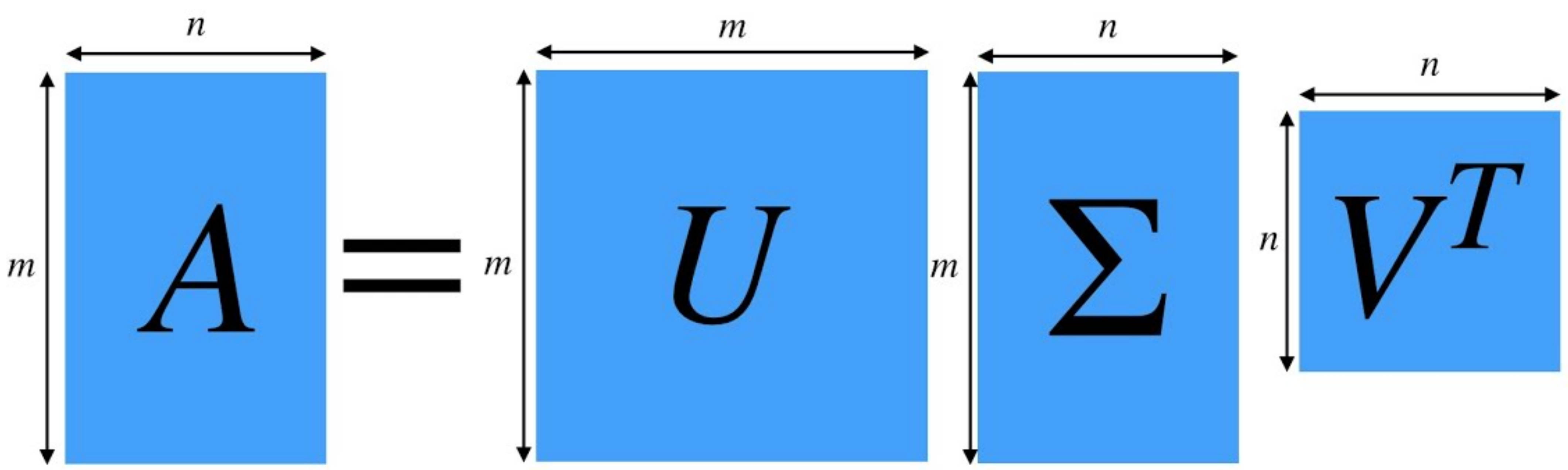
- SVD makes no assumptions on the functional basis

→ Let the data define its own foreground basis!

What’s Singular Value Decomposition

SVD is a fundamental matrix factorization technique that expresses any real matrix $\mathbf{A}$ as a product of three matrices:

$$A = U\Sigma V^T$$



- $\mathbf{A} \in \mathbb{R}^{m\times n}$ **general matrix**

- $\mathbf{U} \in \mathbb{R}^{m\times m}$ **— orthonormal columns ("left singular vectors")**

- $\mathbf{\Sigma} \in \mathbb{R}^{m\times n}$ **— diagonal matrix with non-negative singular values in descending order**

- $\mathbf{V} \in \mathbb{R}^{n\times n}$ **— orthonormal columns ("right singular vectors")**

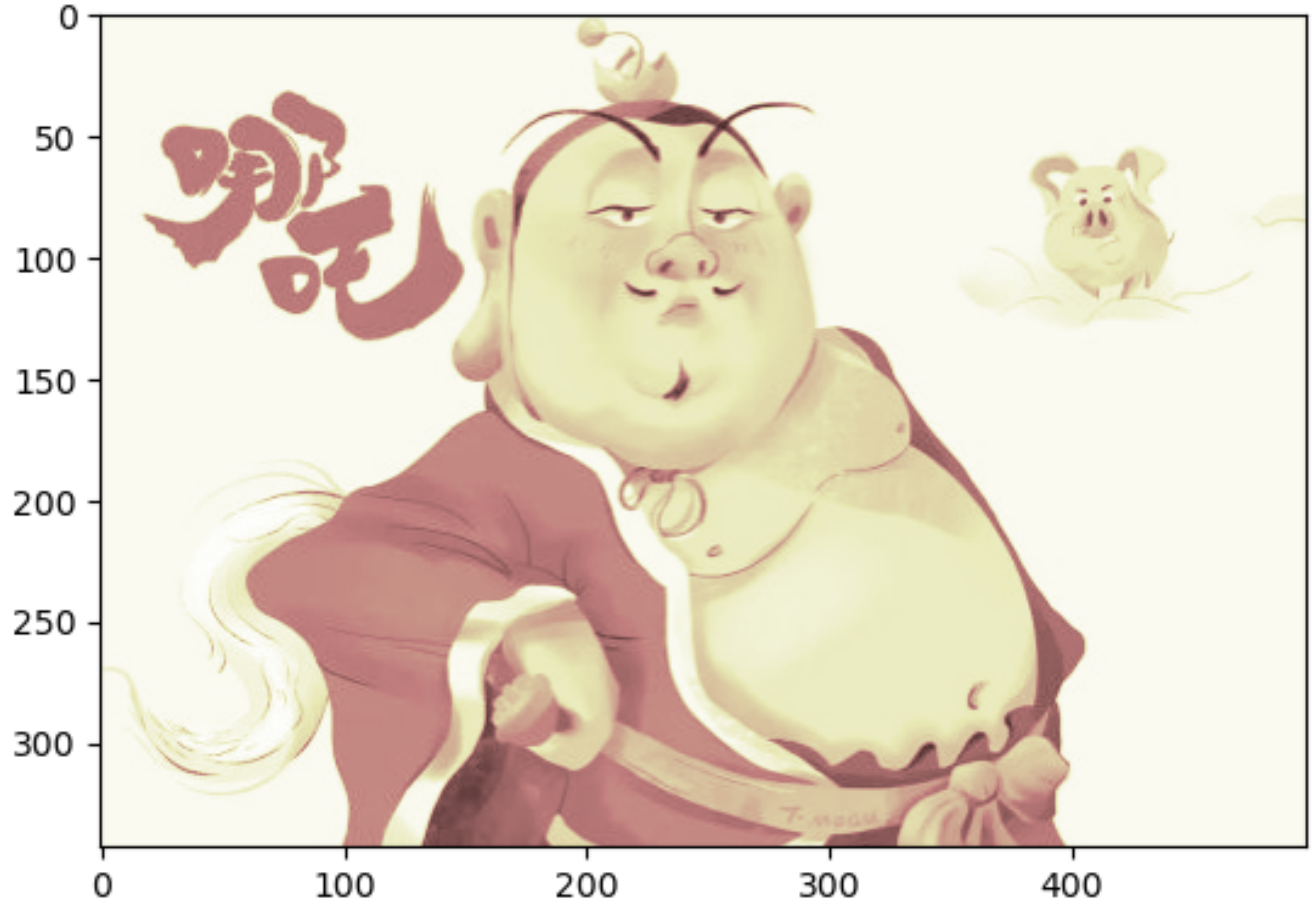
- **SVD decomposes data into orthogonal modes ranked by importance**
- **The singular values in Σ tell us how much variance each mode explains**

SVD for Beginners – Decomposing a 2D Map

Let's say we have a 2D data map: $\mathbf{A} \in \mathbb{R}^{N_y \times N_x}$

We can apply SVD:

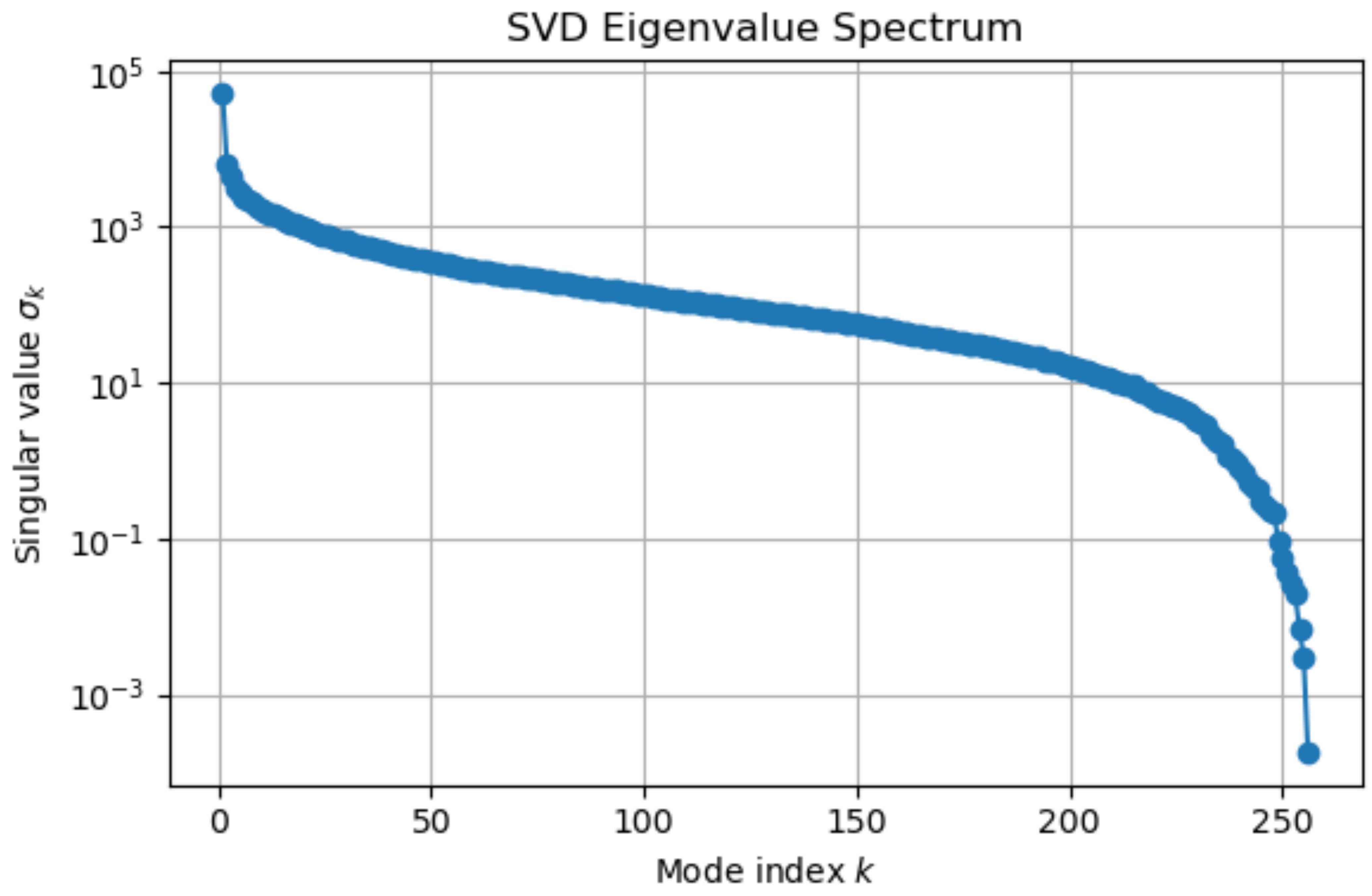
$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top = \sum_{k=1} \sigma_k \cdot \mathbf{u}_k \mathbf{v}_k^\top$$



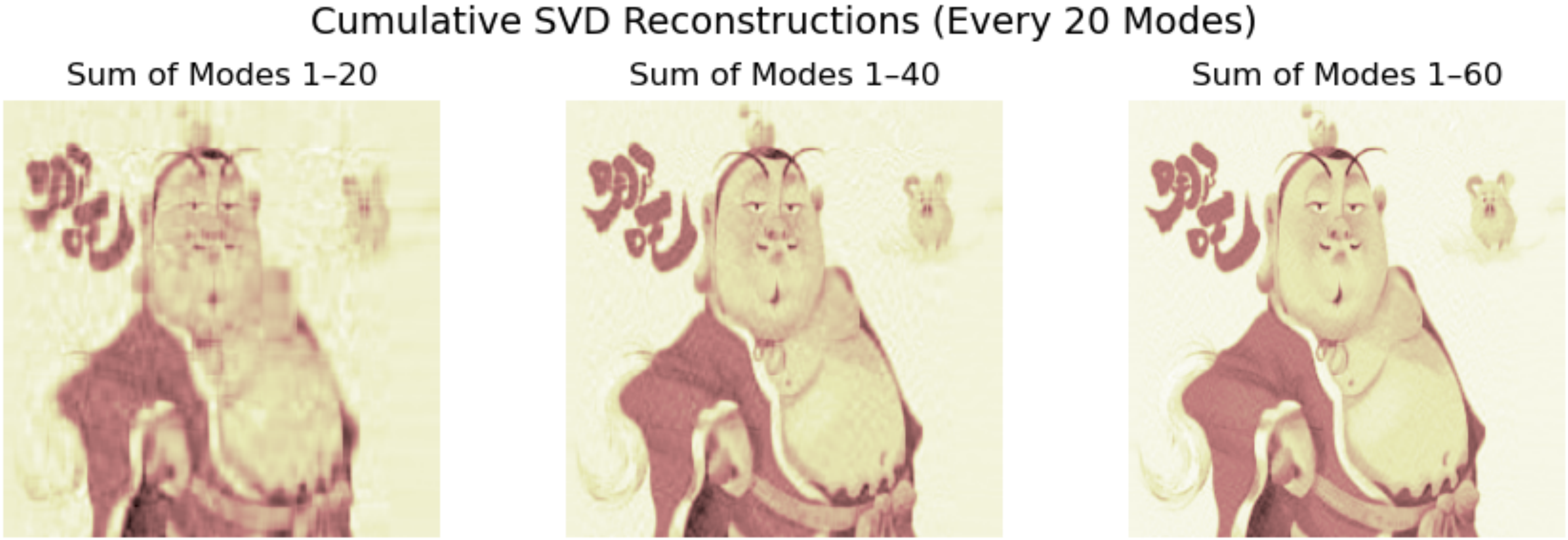
- $\mathbf{U}$ **contains vertical patterns (left singular vectors)**
- $\mathbf{V}^T$ **contains horizontal patterns (right singular vectors)**
- $\boldsymbol{\Sigma}$ **contains strengths (singular values) of each mode**

SVD of 2D Map — Modes & Eigenvalues

- **Each mode map is a rank-1 matrix:** $\mathbf{A}^{(k)} = \sigma_k \cdot \mathbf{u}_k \mathbf{v}_k^\top$
- **Reconstruct Map from SVD Modes:** $\mathbf{A} \approx \sum_{k=1}^{r} \sigma_k \cdot \mathbf{u}_k \mathbf{v}_k^\top$
- **Each mode map $\mathbf{A}^{(k)}$ captures one pattern in the data**



SVD of 2D Map — Modes & Eigenvalues



What Does SVD Do?

- The largest singular values correspond to modes that explain the most variance-dominated by foregrounds

- These dominant modes form a low-rank approximation of the foreground contamination

- By selecting the first k modes (largest singular values), we isolate the foreground subspace: $\mathbf{U}_{\mathrm{fg}} = \mathbf{U}[:\,,:k]$
- Projecting the data orthogonally to this subspace: $\mathbf{A}_{\mathrm{clean}} = \left(\mathbf{I} - \mathbf{U}_{\mathrm{fg}}\mathbf{U}_{\mathrm{fg}}^{\top}\right)\mathbf{A}$
- This removes the smooth foreground components, leaving behind the residuals, including the cosmological 21cm signal and noise

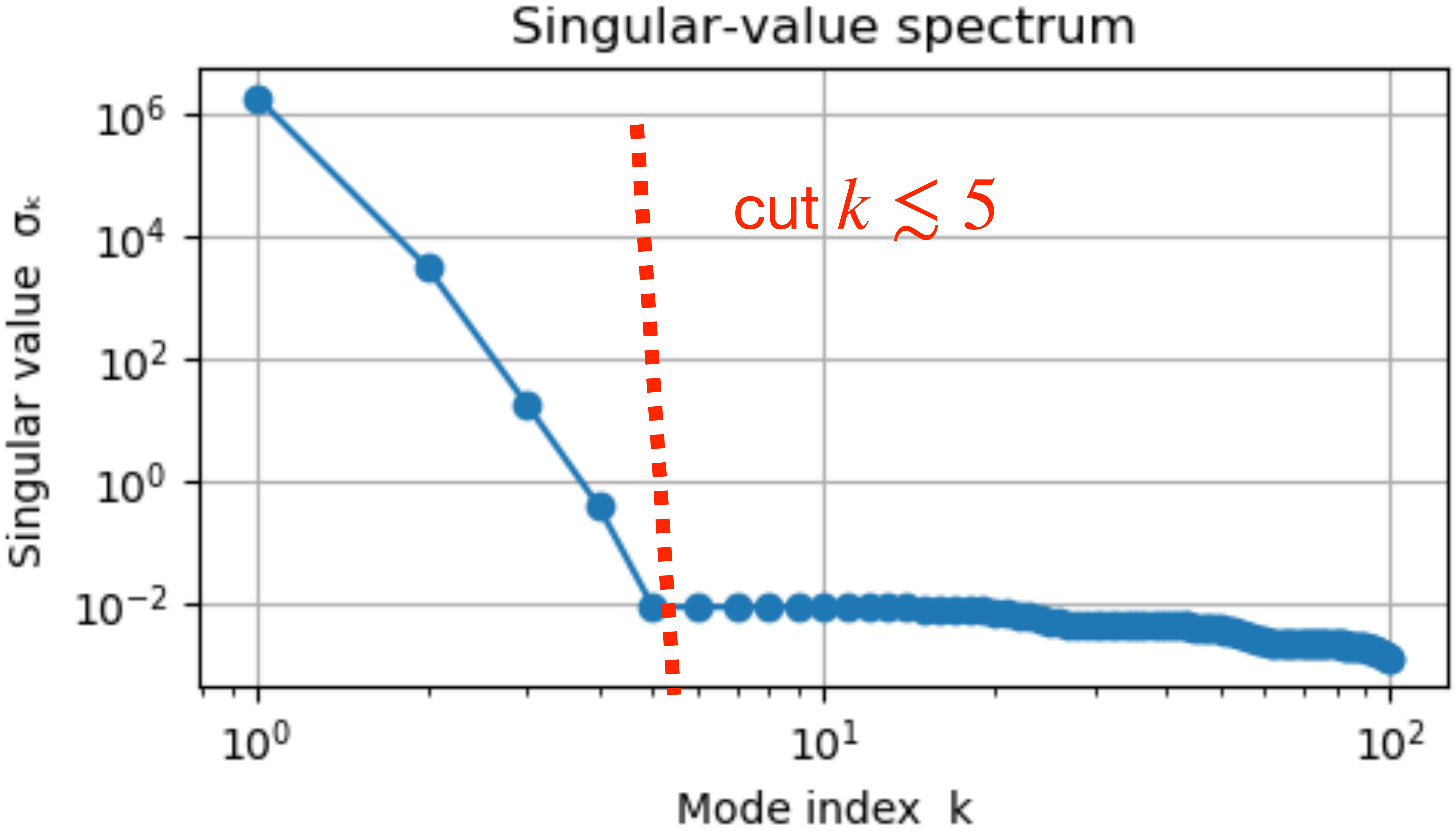
Data Cube Setup

Reshape the 3D cube to a 2D matrix: $\mathbf{A} \in \mathbb{R}^{N_\nu \times N_{\text{pix}}}$**, where** $N_{\text{pix}} = N_x \times N_y$

Apply Singular Value Decomposition: $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$$

$$(\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0)$$

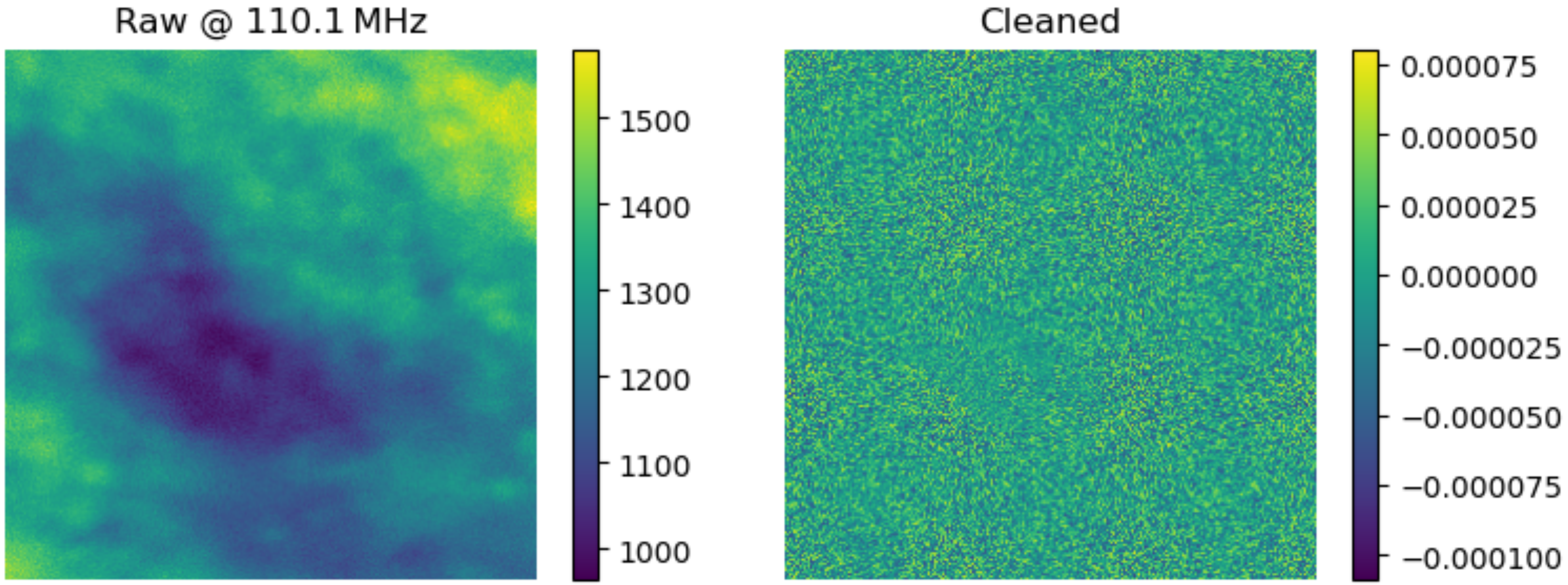


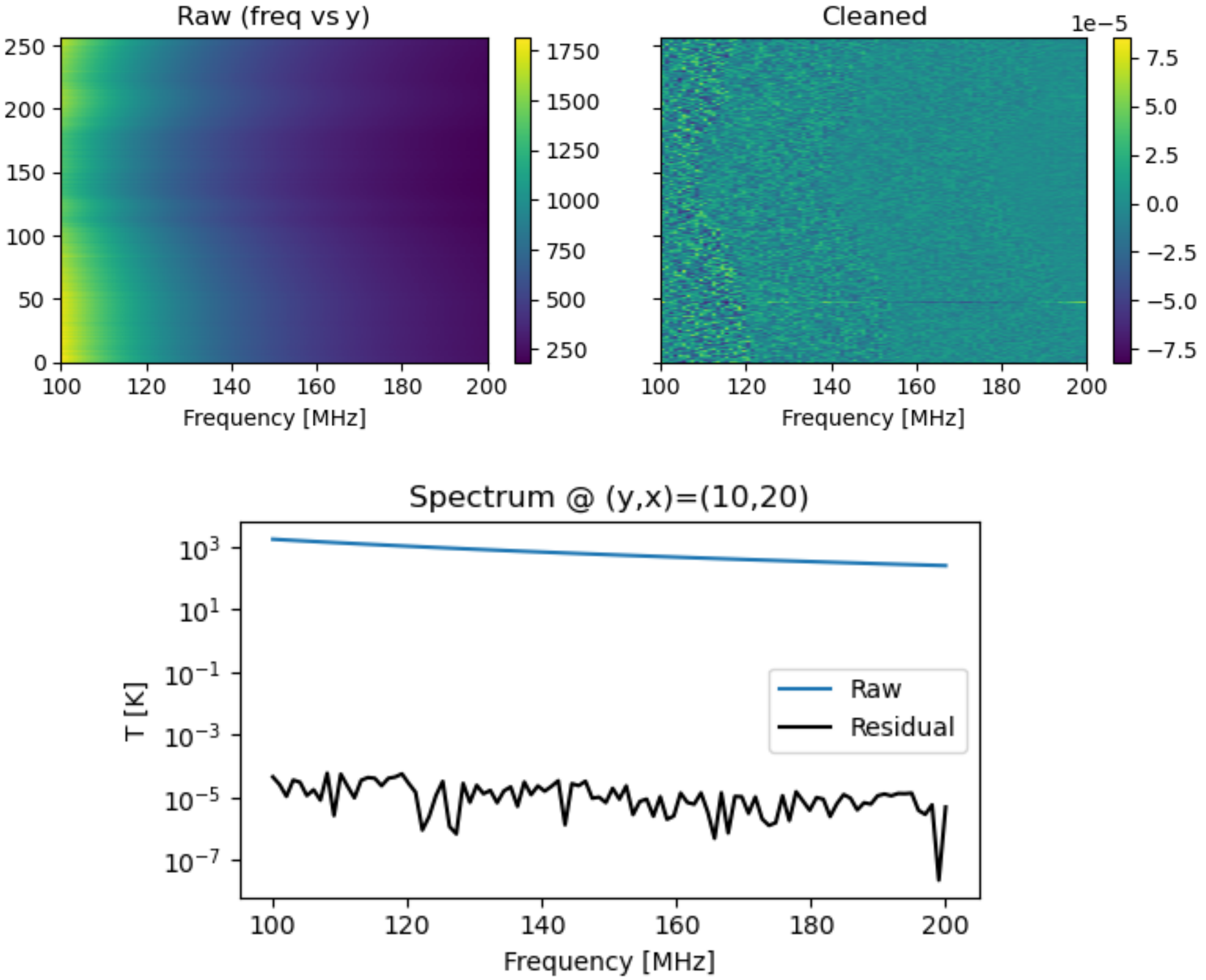
- **Foregrounds lie in a low-dimensional subspace:**

- **Keep only k dominant modes ⇒foreground modes:**

$$\mathbf{A}_{\mathrm{fg}} = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$$

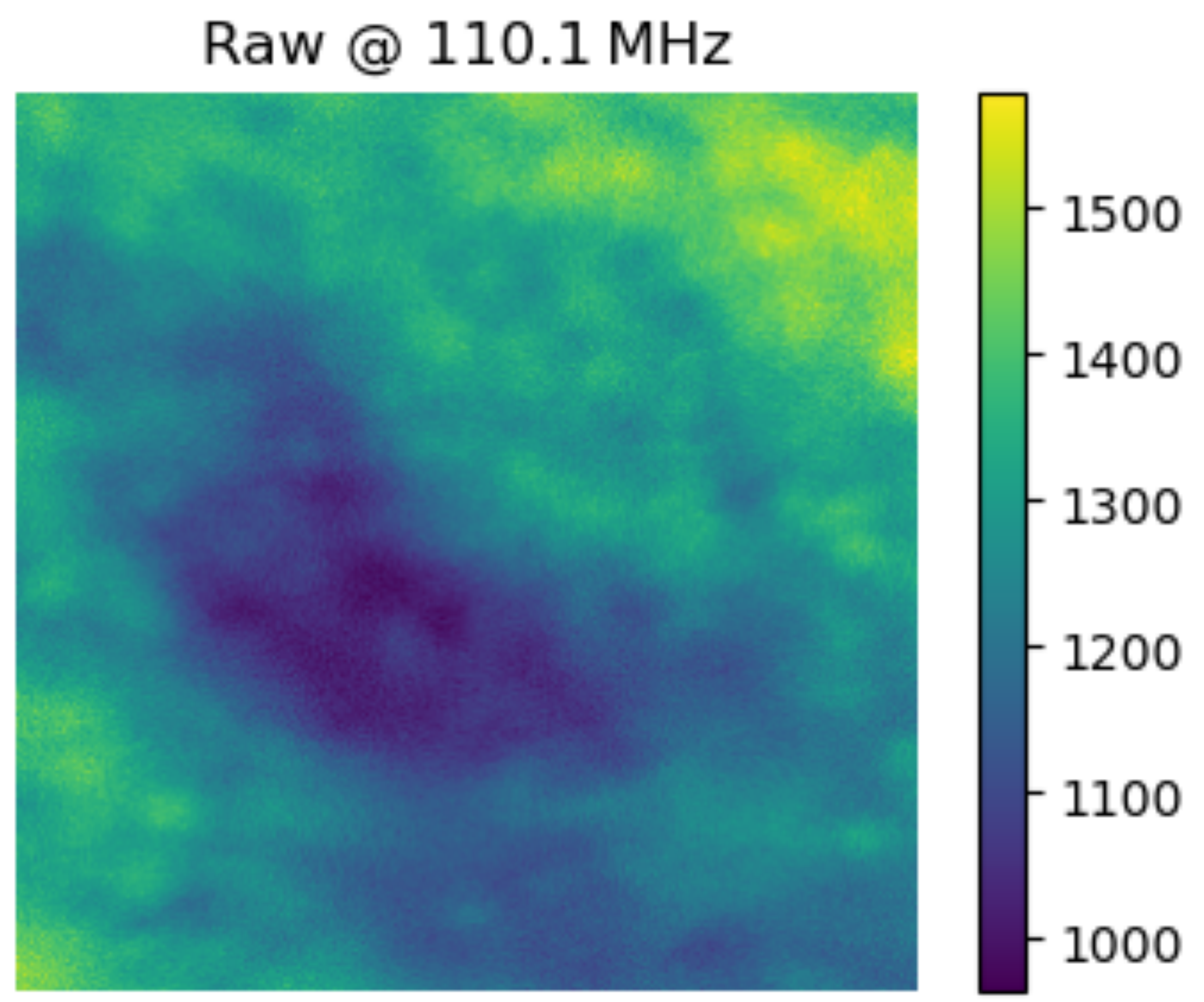
- **Cleaned data = projection orthogonal to foreground space:** $\mathbf{A}_{\mathrm{clean}} = (\mathbf{I} - \mathbf{U}_k \mathbf{U}_k^\top)\mathbf{A}$

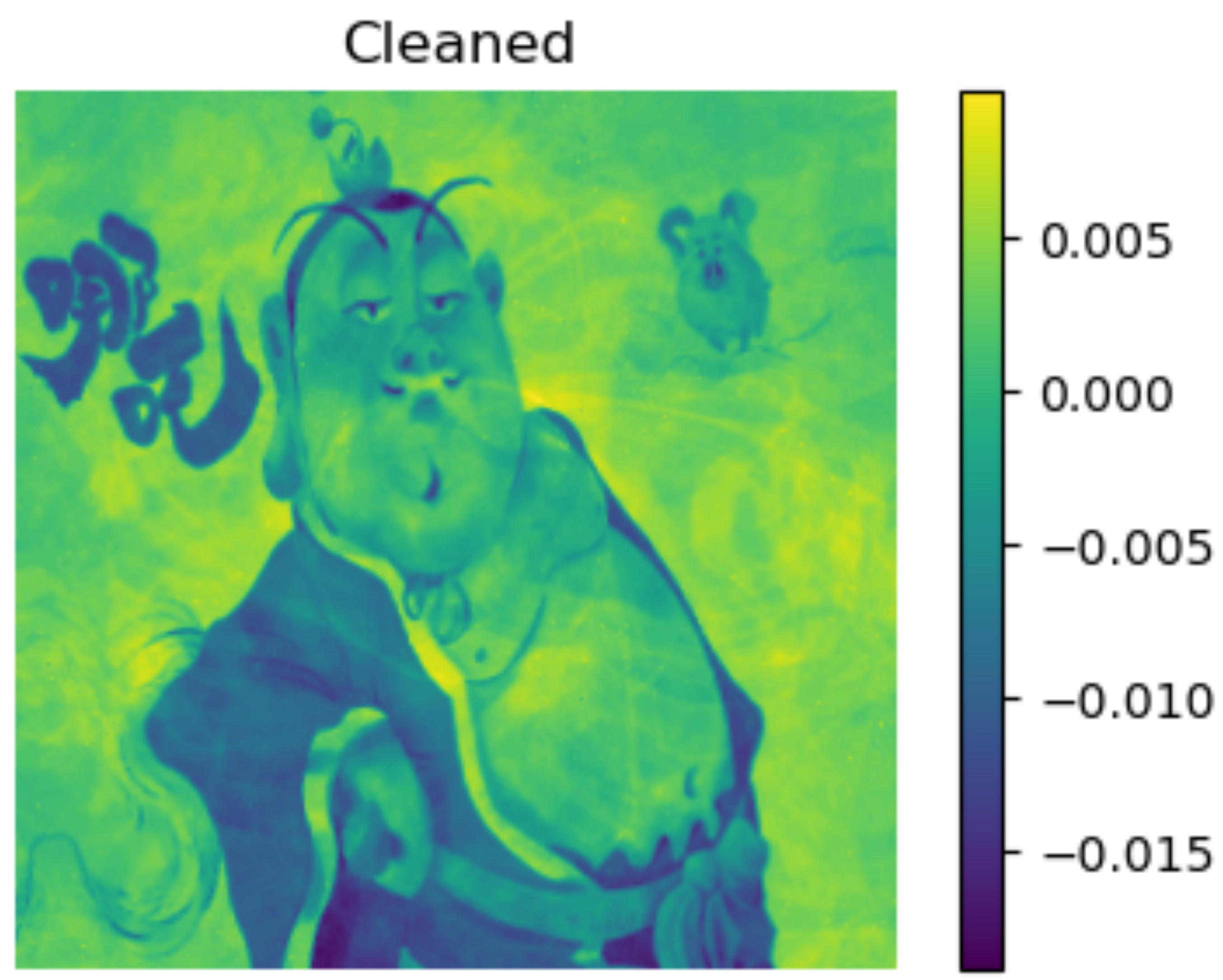


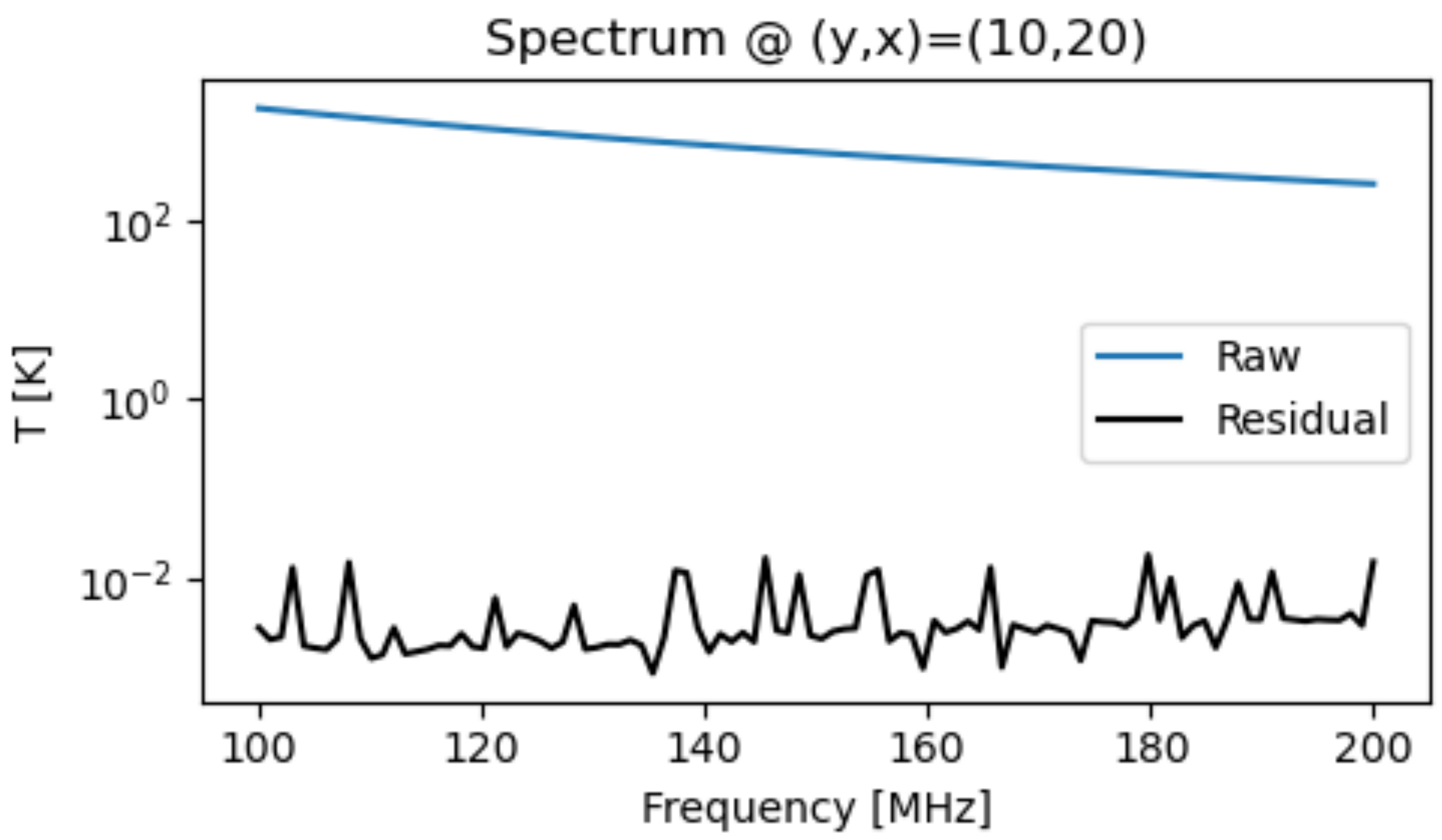


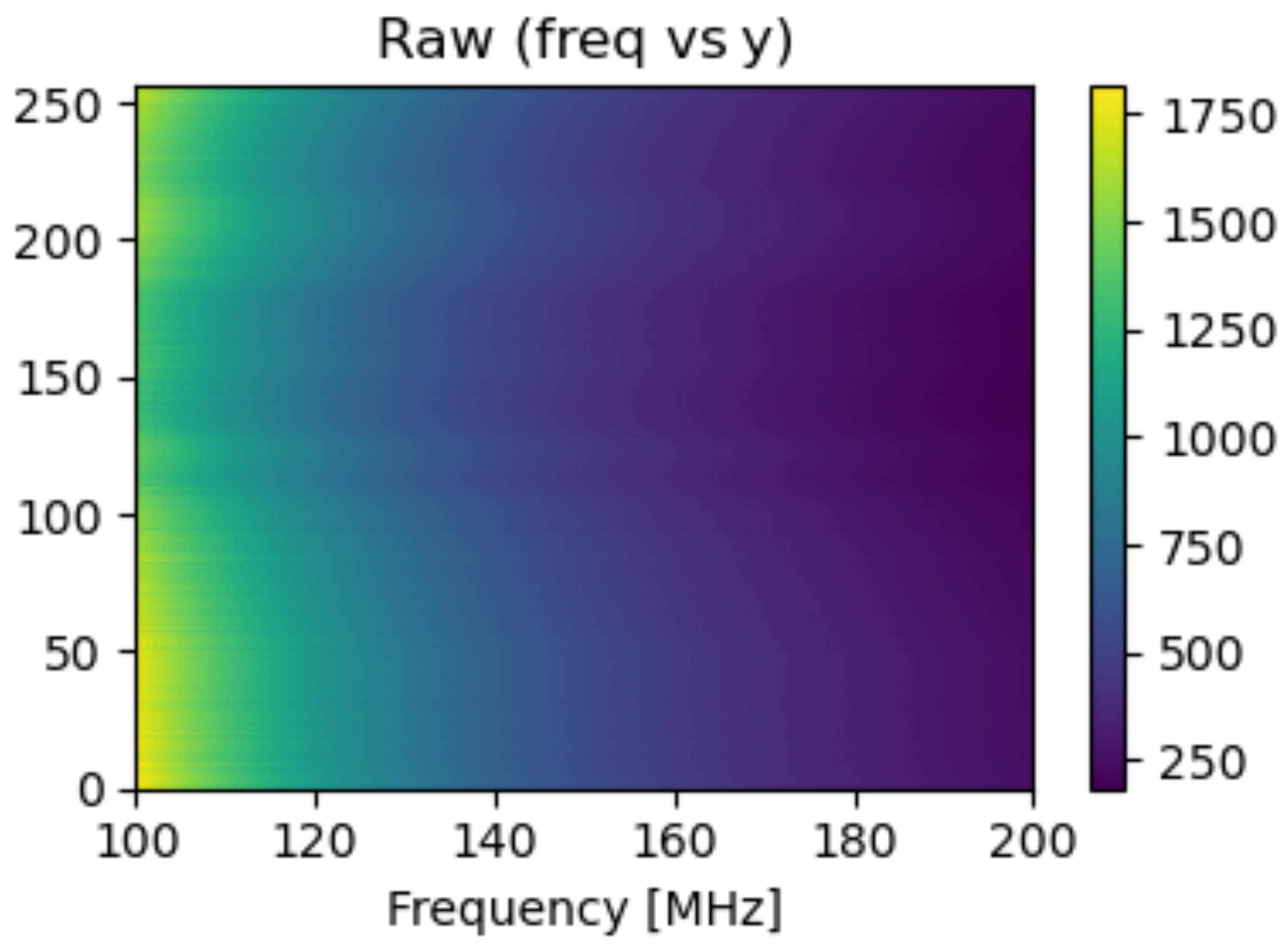
Raw (freq vs y)
Cleaned
1e−5
Frequency [MHz]
Frequency [MHz]
Spectrum @ (y,x)=(10,20)
T [K]
Raw
Residual
Frequency [MHz]

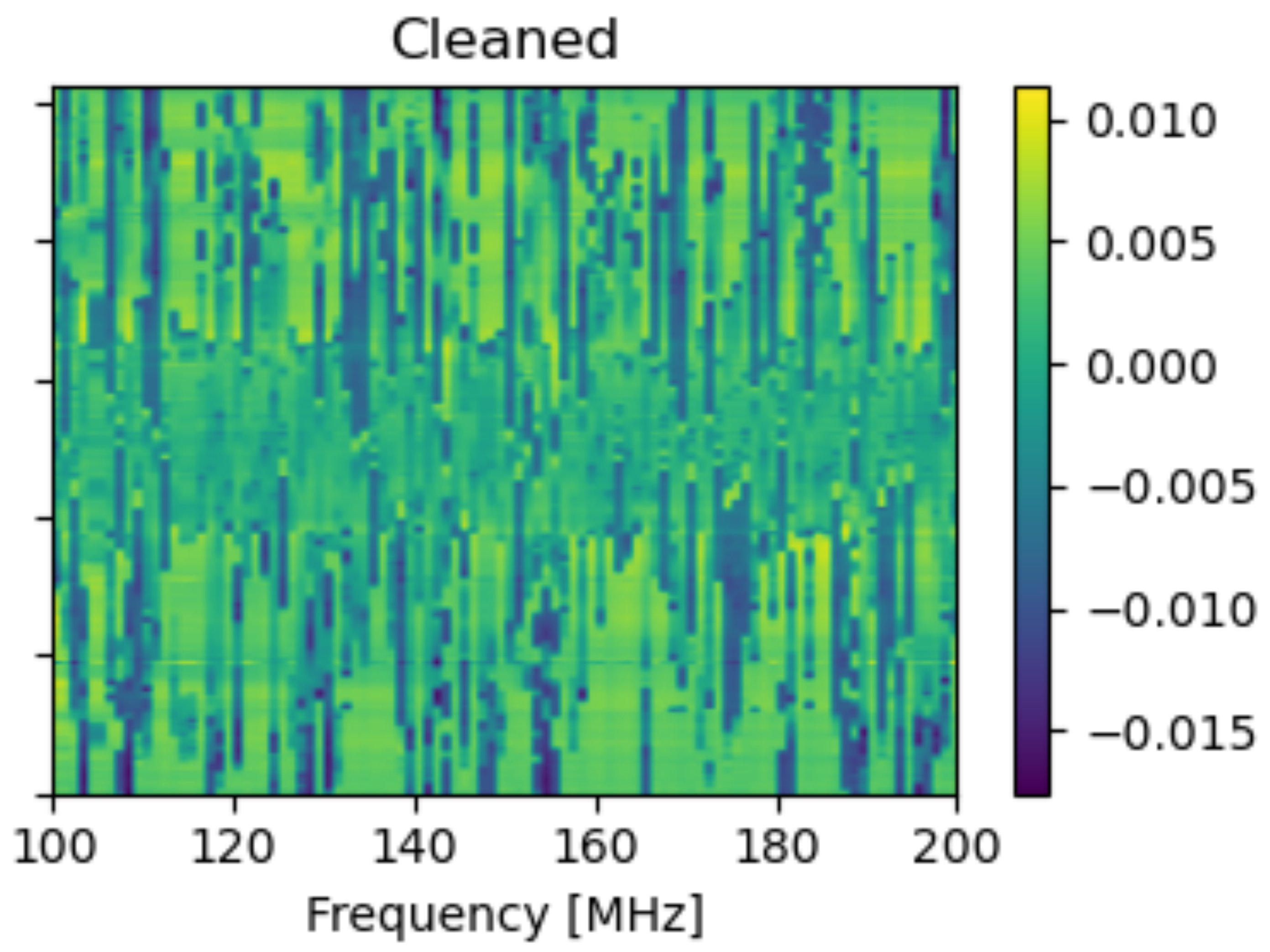
Adding EoR to data cube…





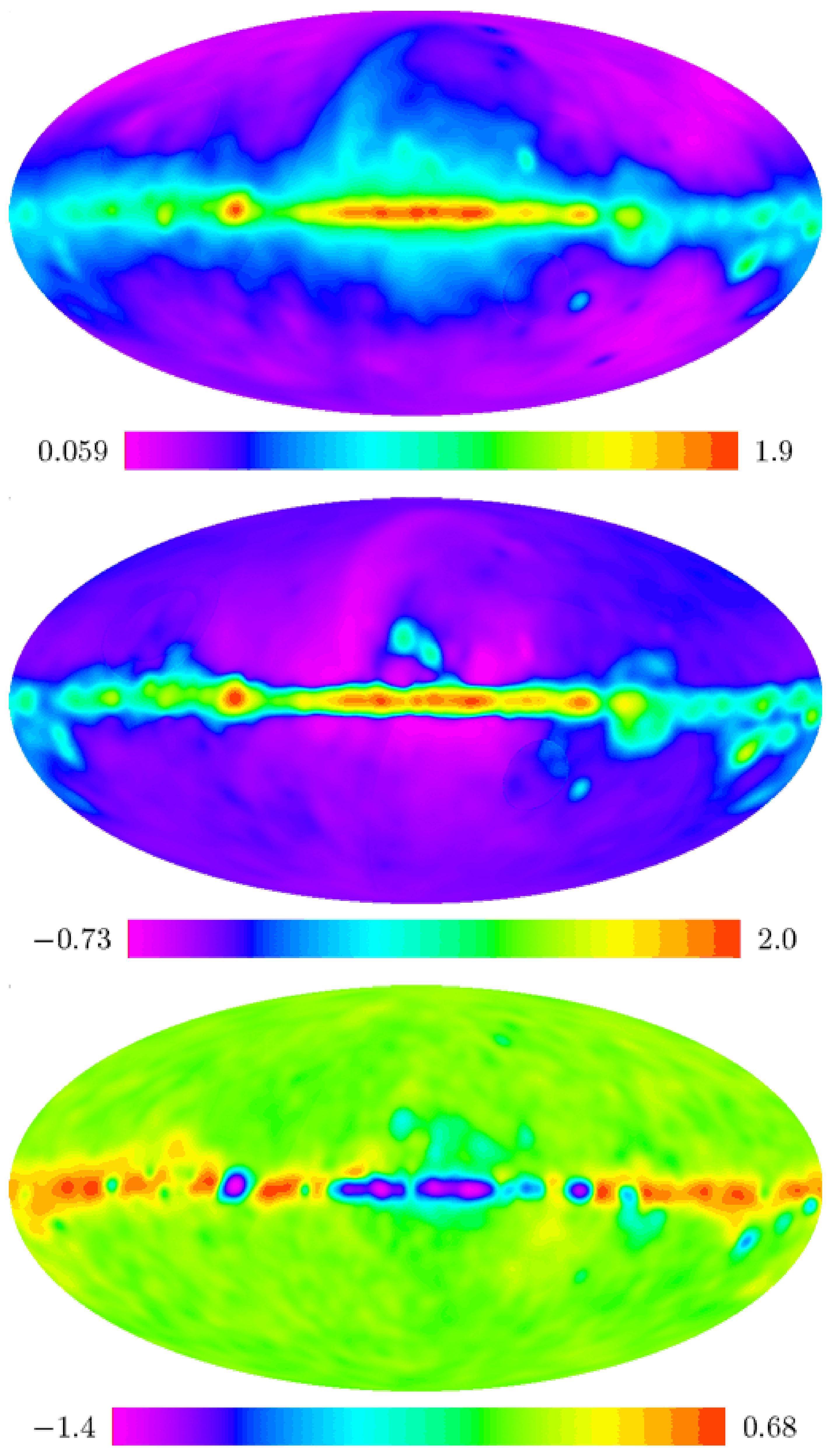




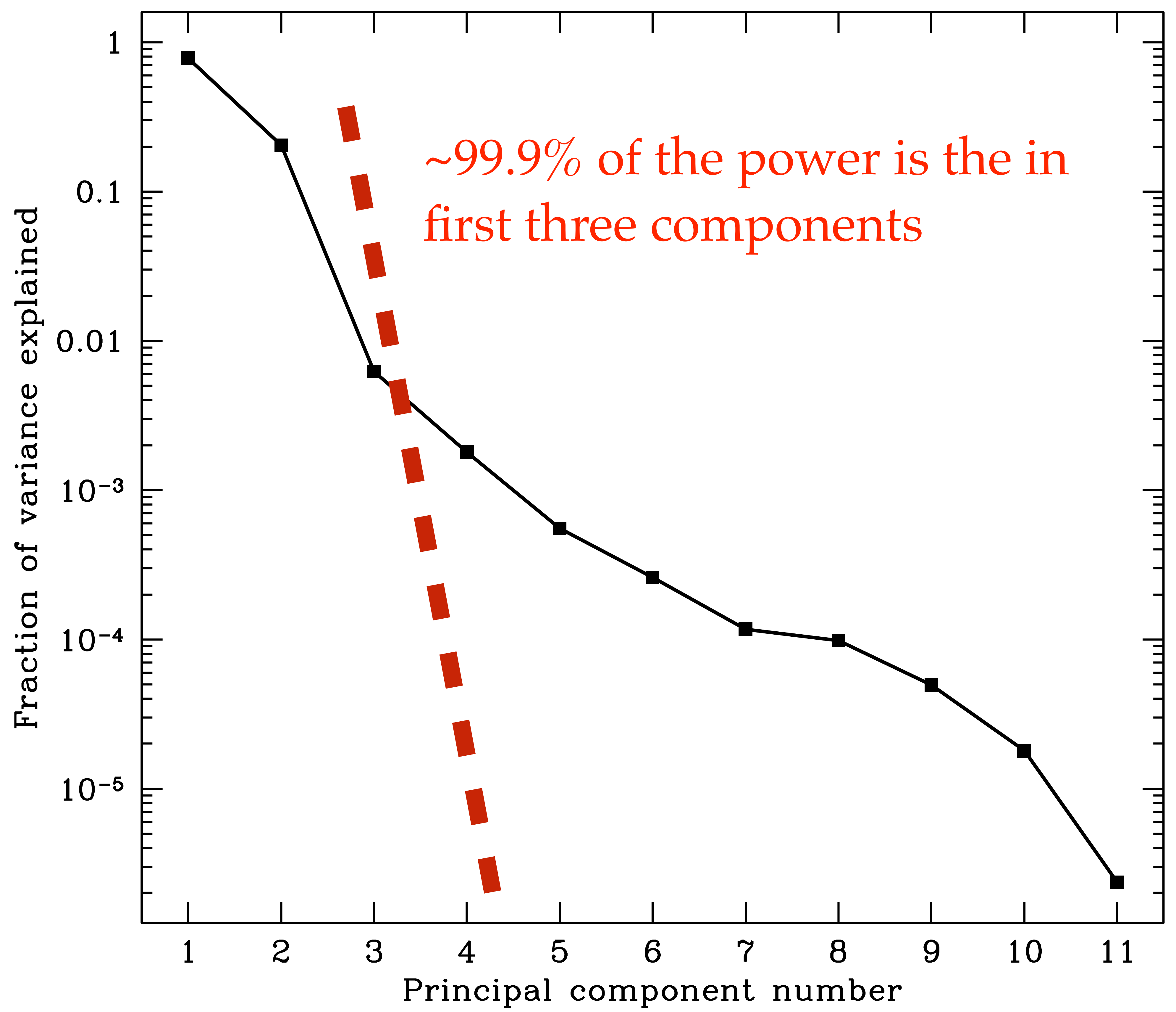


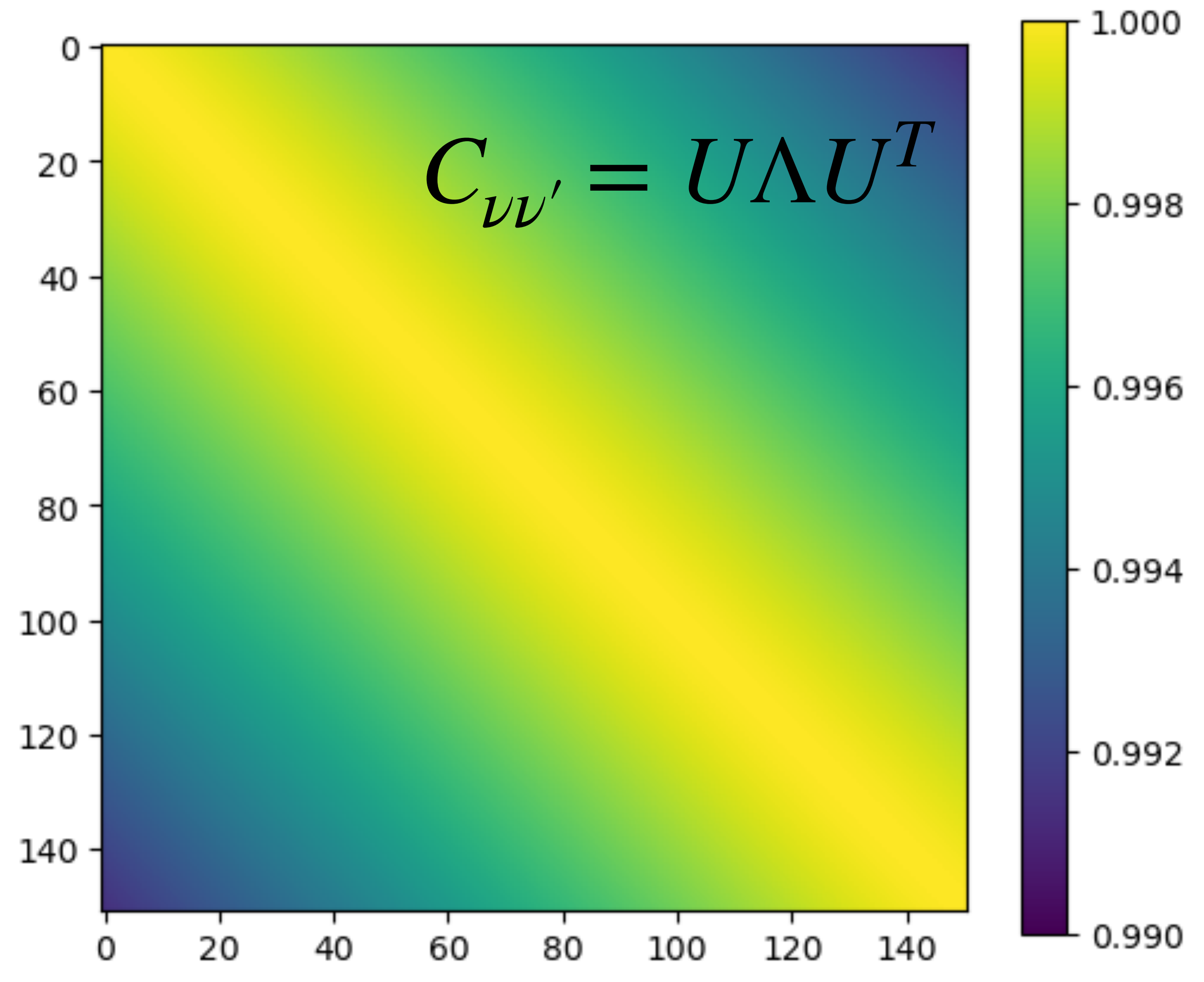
Pearson correlation coefficient r=0.96 (k=3)

PCA/SVD method



The first three principal components, which can be crudely interpreted as maps of total "stuff", synchrotron fraction and thermal dust fraction.





Blind Source Separation for Foreground Removal

In 21cm experiments, the observed brightness temperature cube T(v,x) contains:

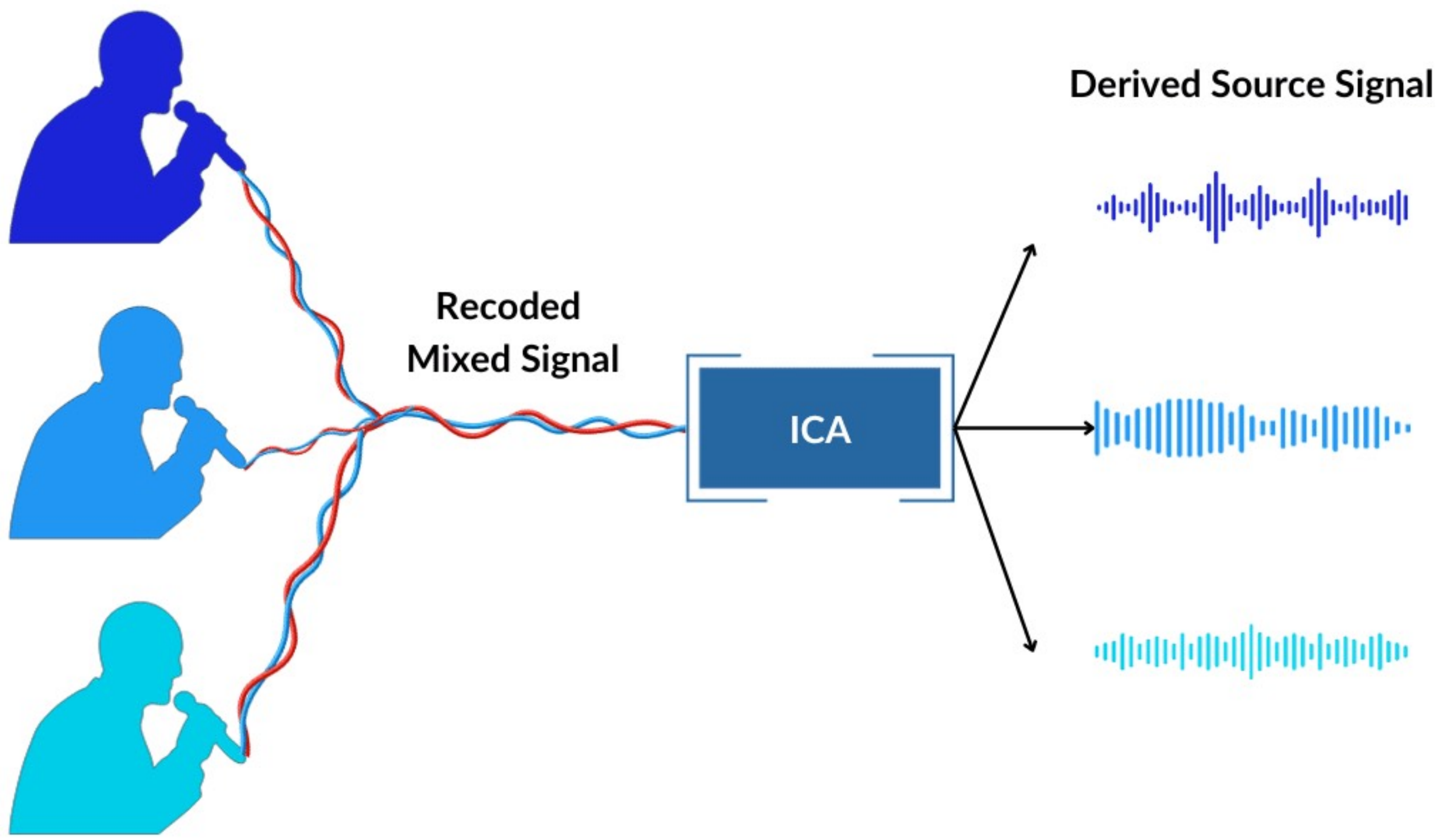
- **Bright foregrounds: galactic synchrotron, free-free emission,…**
- **Faint EoR signal: highly fluctuating in frequency, spatially random**
- **Instrumental noise**

These are mixed together in the observed data, with no labels.

A two-component foreground mixing model:

$$\mathbf{X}_{N_\nu \times N_{\mathrm{pix}}} = \underbrace{\begin{pmatrix} F_1(\nu_1) & F_2(\nu_1) \\ F_1(\nu_2) & F_2(\nu_2) \\ \vdots & \vdots \\ F_1(\nu_{N_\nu}) & F_2(\nu_{N_\nu}) \end{pmatrix}}_{\mathbf{A}_{N_\nu \times 2}} \underbrace{\begin{pmatrix} s_{F1,1} & s_{F1,2} & \cdots & s_{F1,N_{\mathrm{pix}}} \\ s_{F2,1} & s_{F2,2} & \cdots & s_{F2,N_{\mathrm{pix}}} \end{pmatrix}}_{\mathbf{S}_{2\times N_{\mathrm{pix}}}}$$

FastICA for Foreground Removal



$\mathbf{X} = \mathbf{AS} + \mathbf{N}$

Goal: Recover S without knowing A

- $\mathbf{X}$: observed data (e.g., frequency × pixel matrix)
- A: unknown mixing matrix; frequency dependence for each component
- S: unknown independent sources ($s_{F1,1}, s_{F2,2}$ foregrounds)
- N: noise

Fast Independent Component Analysis (FastICA) helps by separating statistically independent components based on their non-Gaussianity:

- **Foregrounds: spectrally smooth, large-scale coherent patterns**
- **21cm signal: spectrally fluctuating, nearly Gaussian-like, spatially incoherent**

Core Principles of FastICA

- Central Limit Theorem:
  A sum of independent variables is more Gaussian than the original variables.

- To find sources, look for projections that are maximally non-Gaussian.

- Measure of Non-Gaussianity: Kurtosis, Neg-entropy, …

- FastICA (Hyvärinen, 1999) can find an unmixing matrix W such that: $Y=WX=WAS\approx S$

Advantages for 21cm Signal Recovery:

- Model-free: no need to assume specific foreground spectra

- Efficient: much faster than iterative optimization or MCMC methods

- project out foregrounds by removing strongest independent components, leaving residuals that contain EoR + noise

FastICA: Whitening → Optimization → Extraction

1. SVD whitening:

Transform the centered data to make components uncorrelated with unit variance

- $\mathbf{X}_c = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T, \quad \mathbf{Z} = \boldsymbol{\Sigma}^{-1}\mathbf{U}^T\mathbf{X}_c, \ \mathrm{Cov}(\mathbf{Z}) = \mathbf{I}$

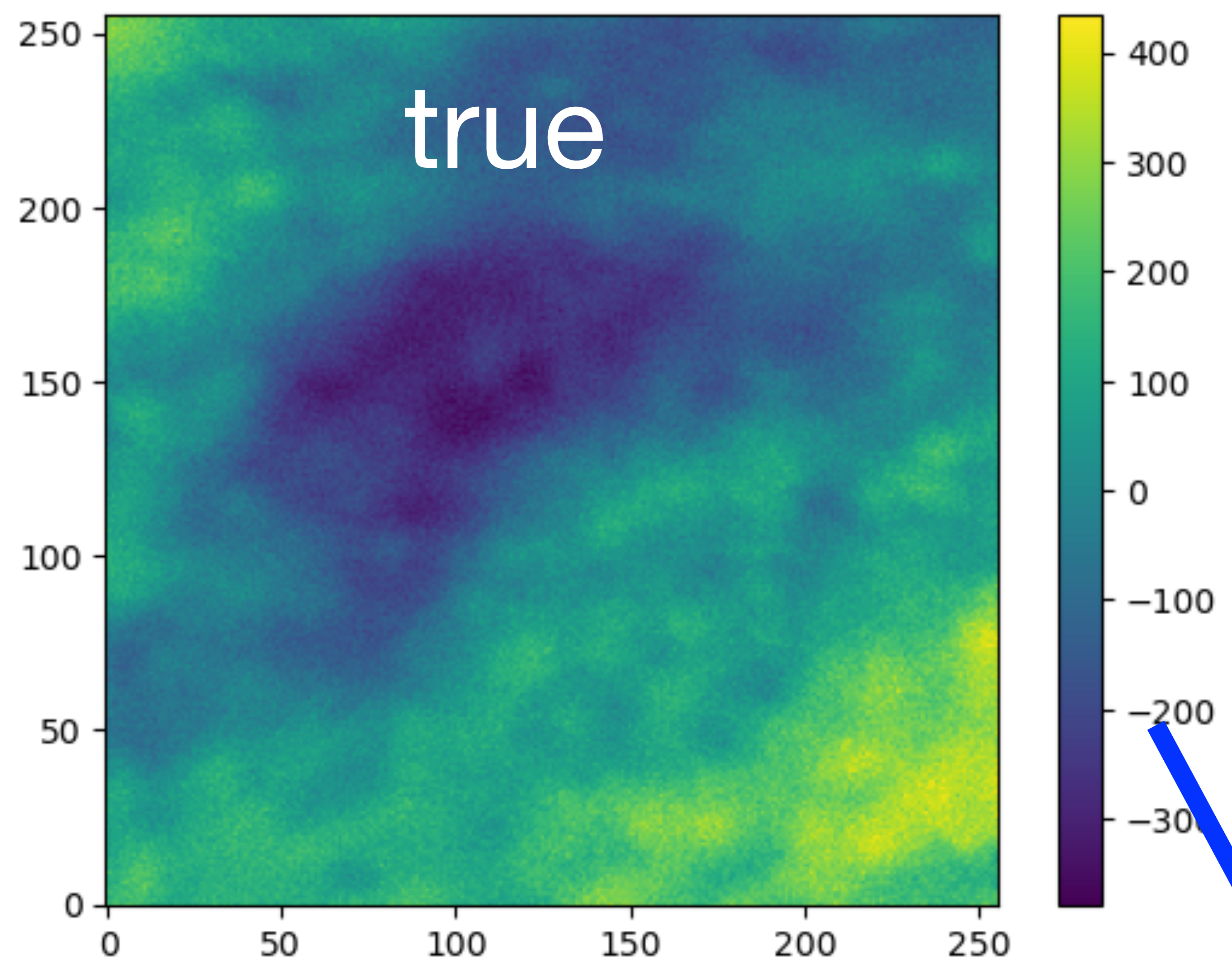
2. Iterative Optimization:

- **Find weight vector $\mathbf{w}$ that maximizes non-Gaussianity of projection $\mathbf{w}^T\mathbf{Z}$:**
- $\mathbf{w}_{\text{new}} \leftarrow \mathbb{E}\left[\mathbf{Z} \cdot g\left(\mathbf{w}^T\mathbf{Z}\right)\right] - \mathbb{E}\left[g'\left(\mathbf{w}^T\mathbf{Z}\right)\right] \cdot \mathbf{w}$
- $g(u)$ **: non-linear function (e.g., $\tanh(u), u^3$)**
- $\mathbb{E}[\ \cdot\ ]$ **: average over all data samples (pixels):** $\mathbb{E}[f(\mathbf{z})] = \dfrac{1}{N_p}\sum_{j=1}^{N_p} f\left(\mathbf{z}_j\right)$
- **Normalize $\mathbf{w}_{\text{new}}$ and repeat until convergence**

3. Extract Independent Components

- **S=WZ. Rows of S statistically independent components**

ICA on 3D Foreground Data Cube (n_comp=4)

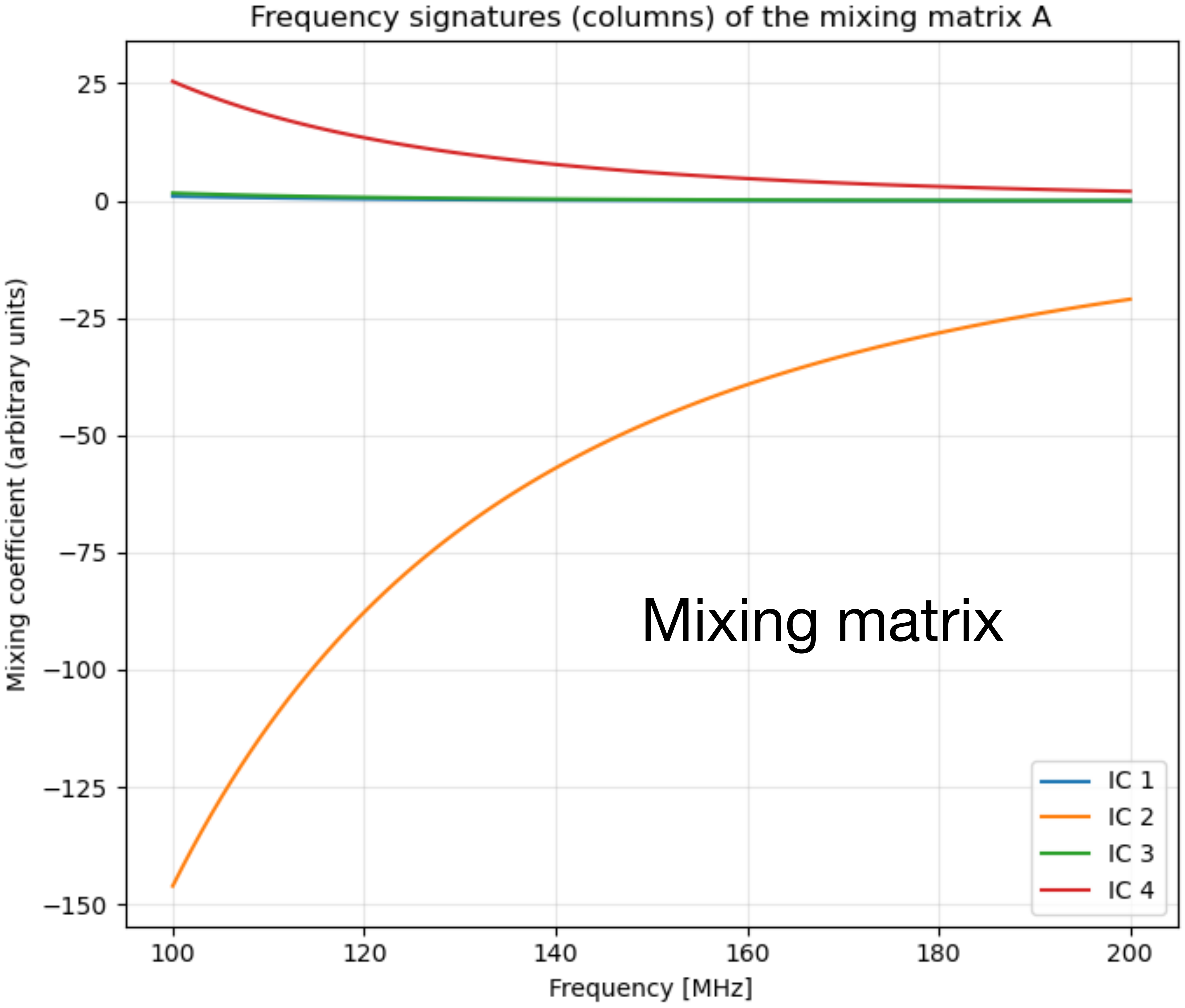


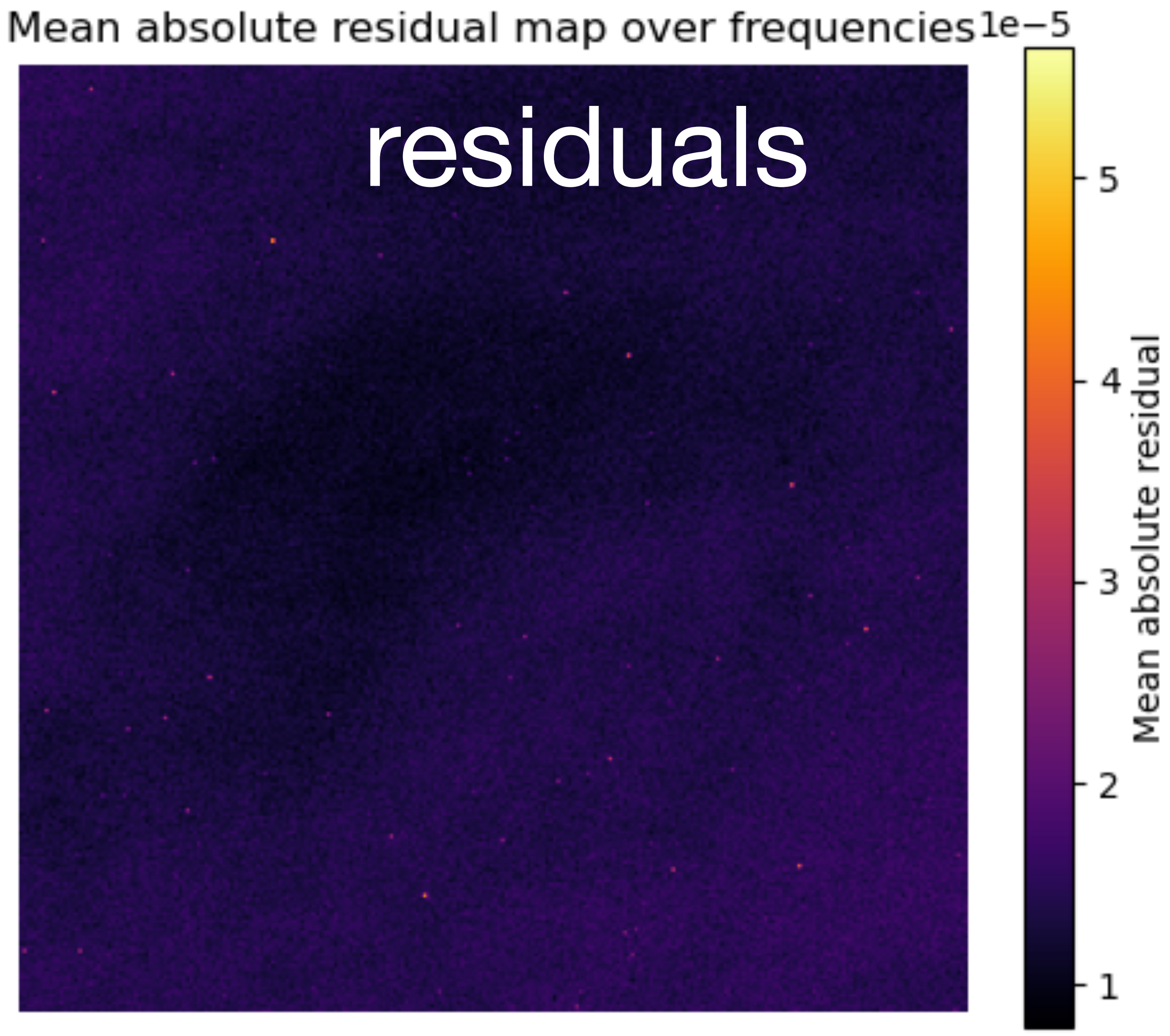
FastICA solves for statistically independent sources, but:

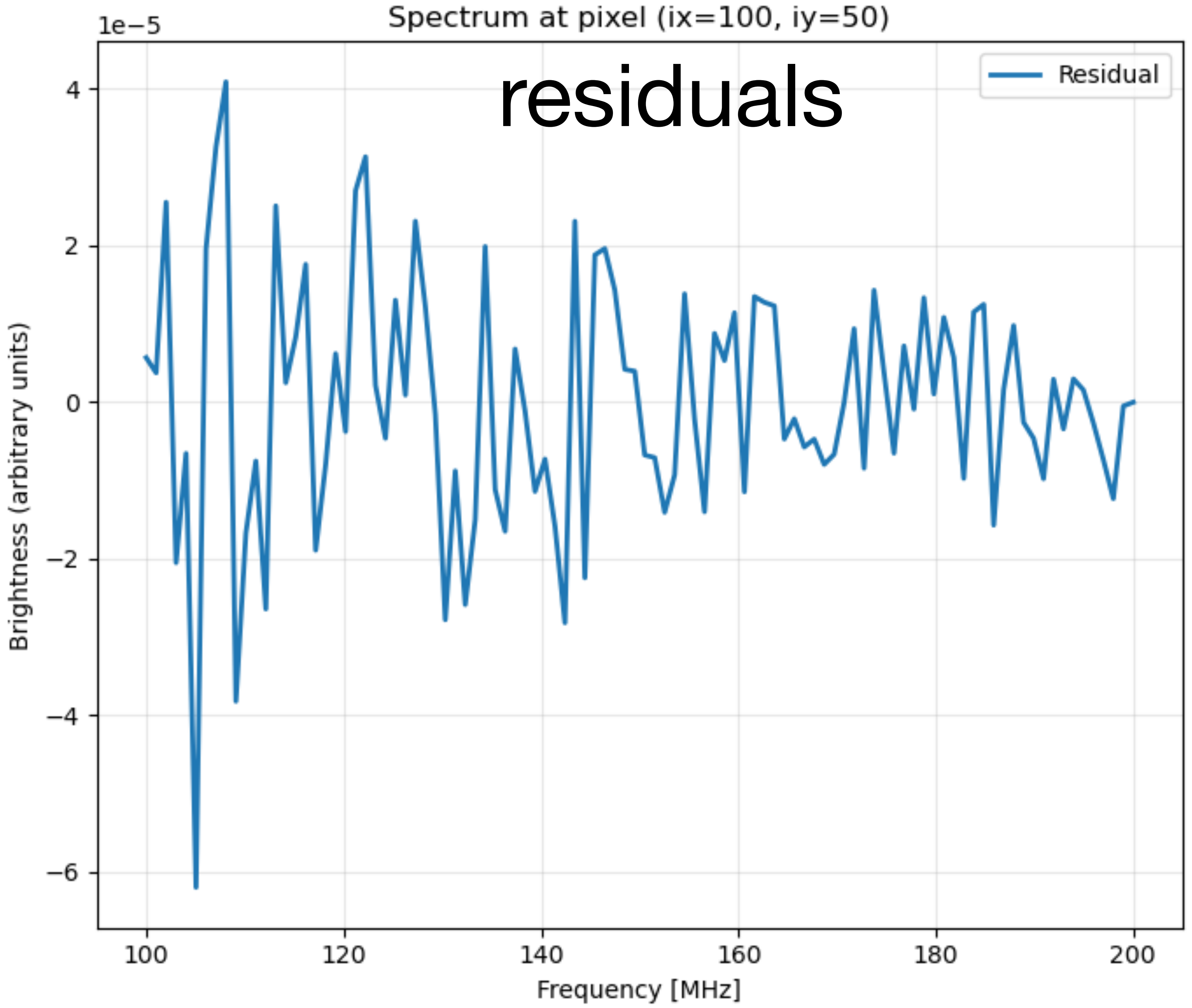
- **There's no natural order (like variance in PCA)**
- **ICA doesn't rank components**
- **The output depends on initial conditions**
- **Each component can flip sign**

ICA on 3D Foreground Data Cube

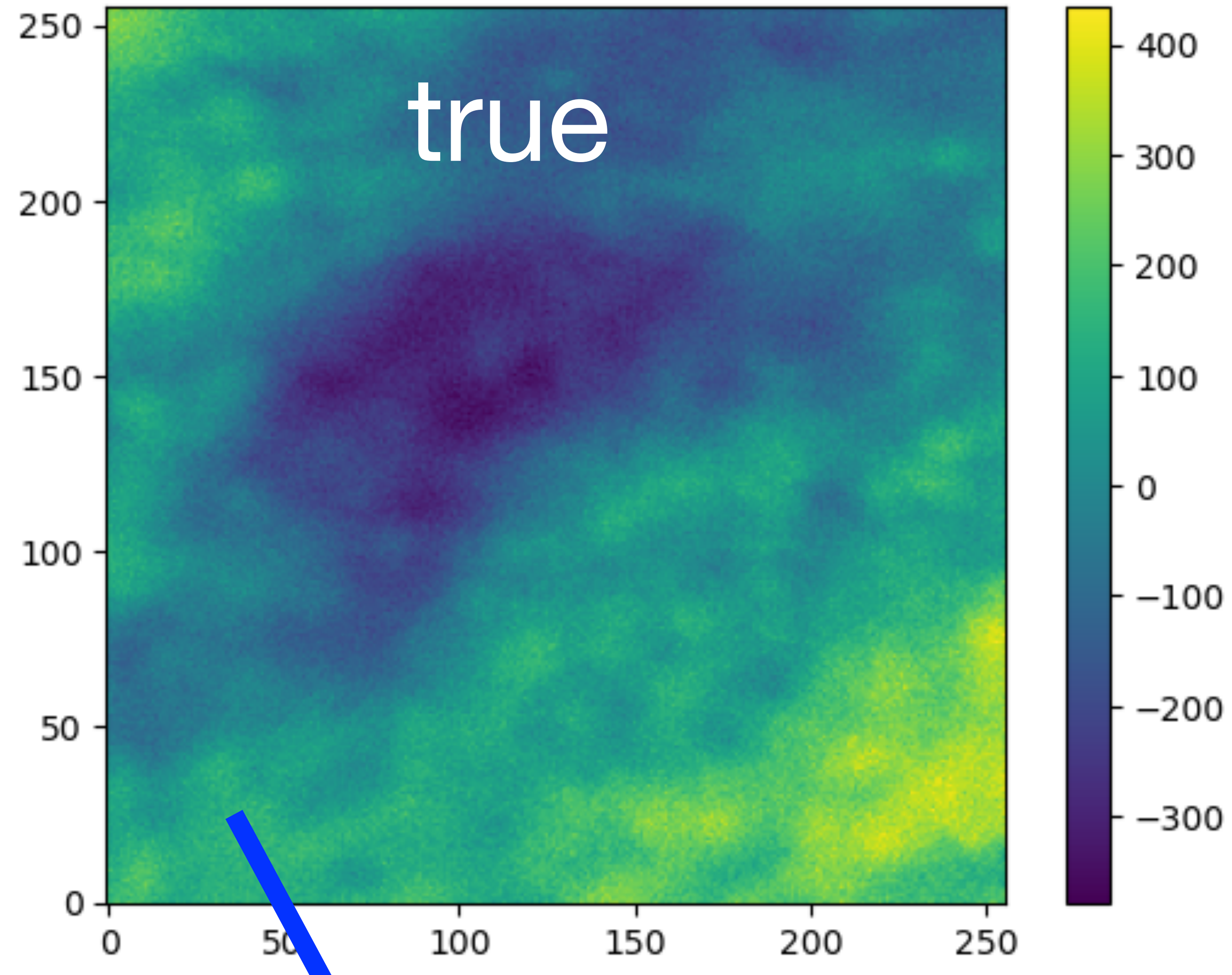
residuals ~ 10^{-5}K



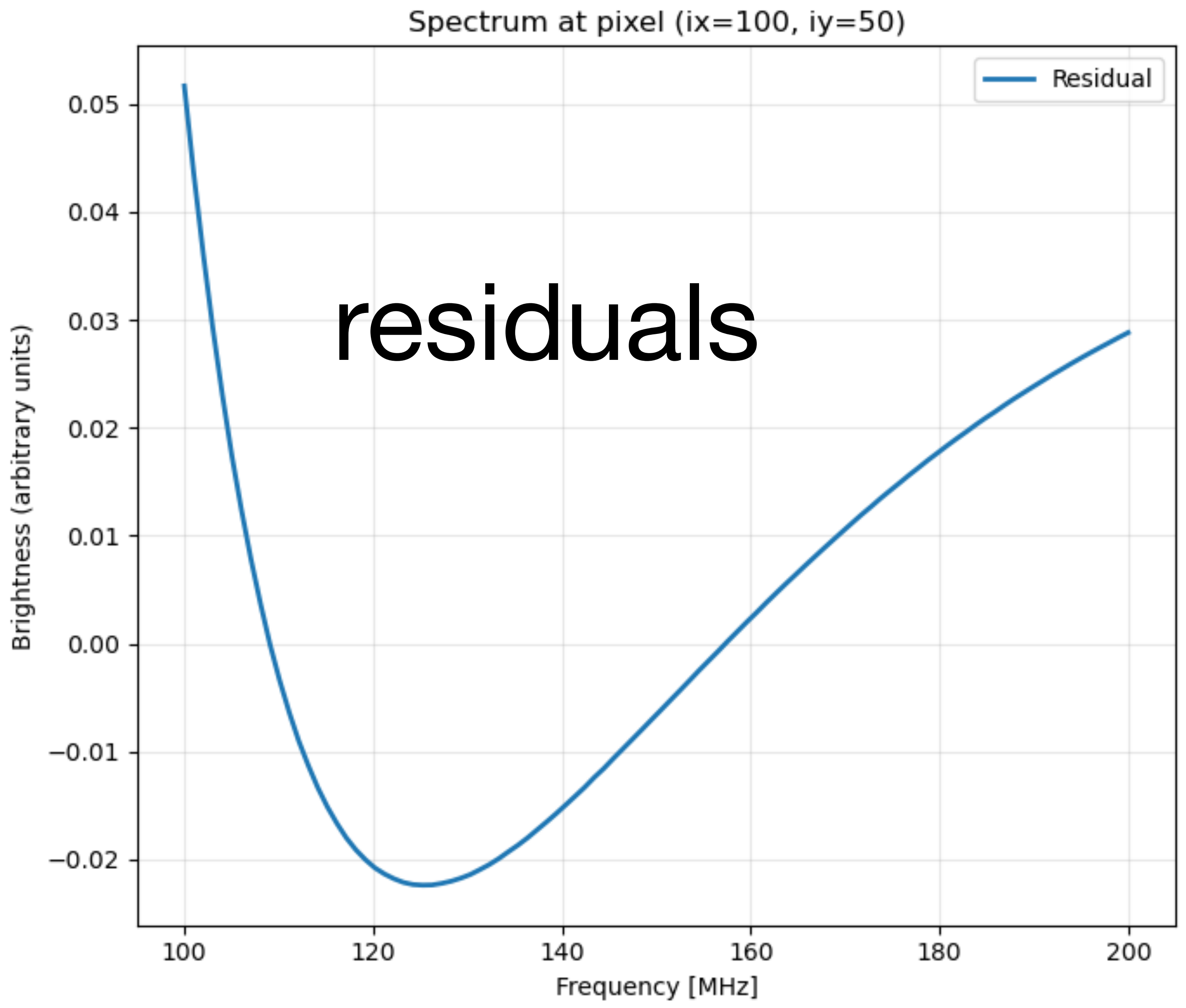


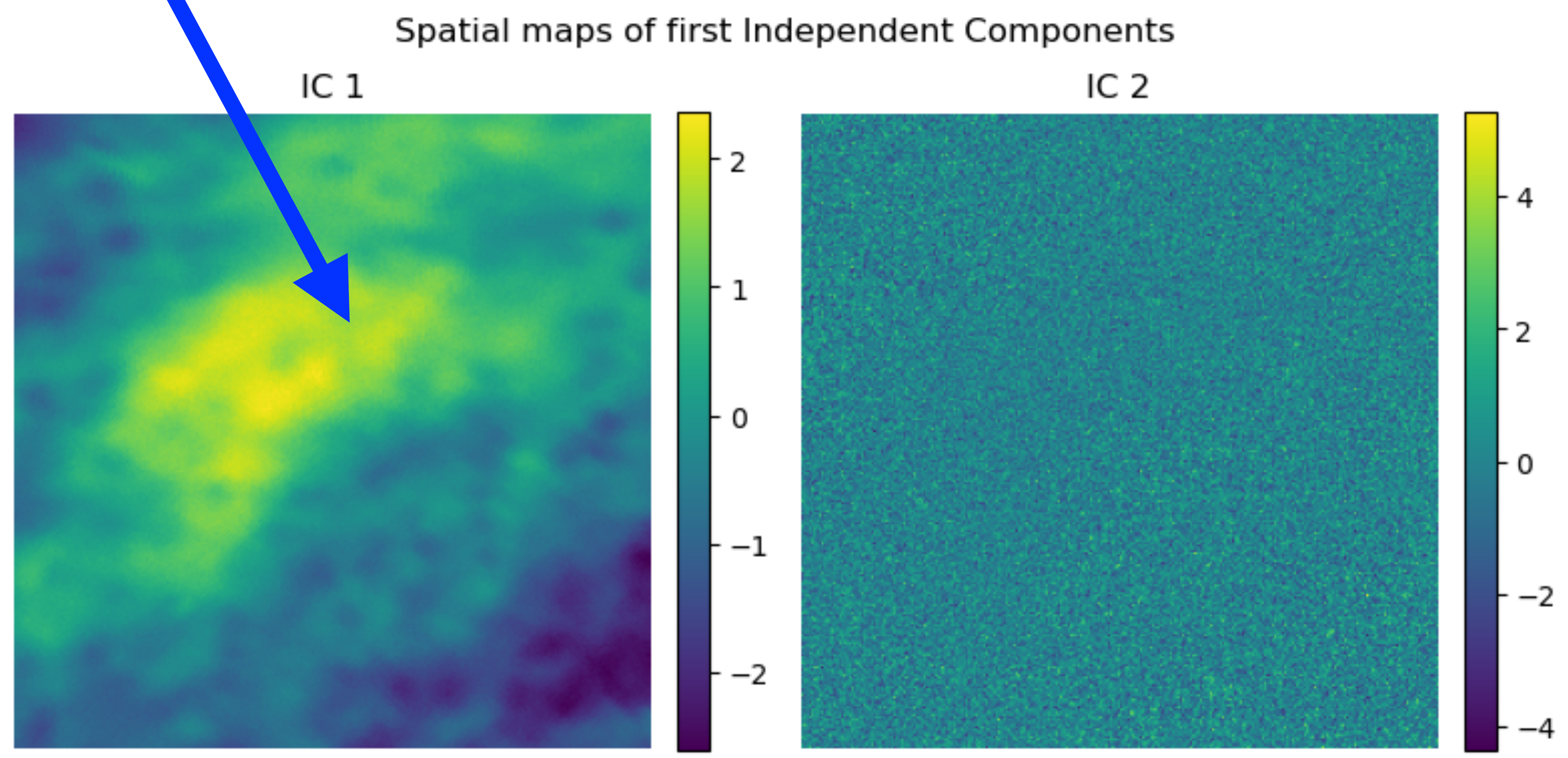


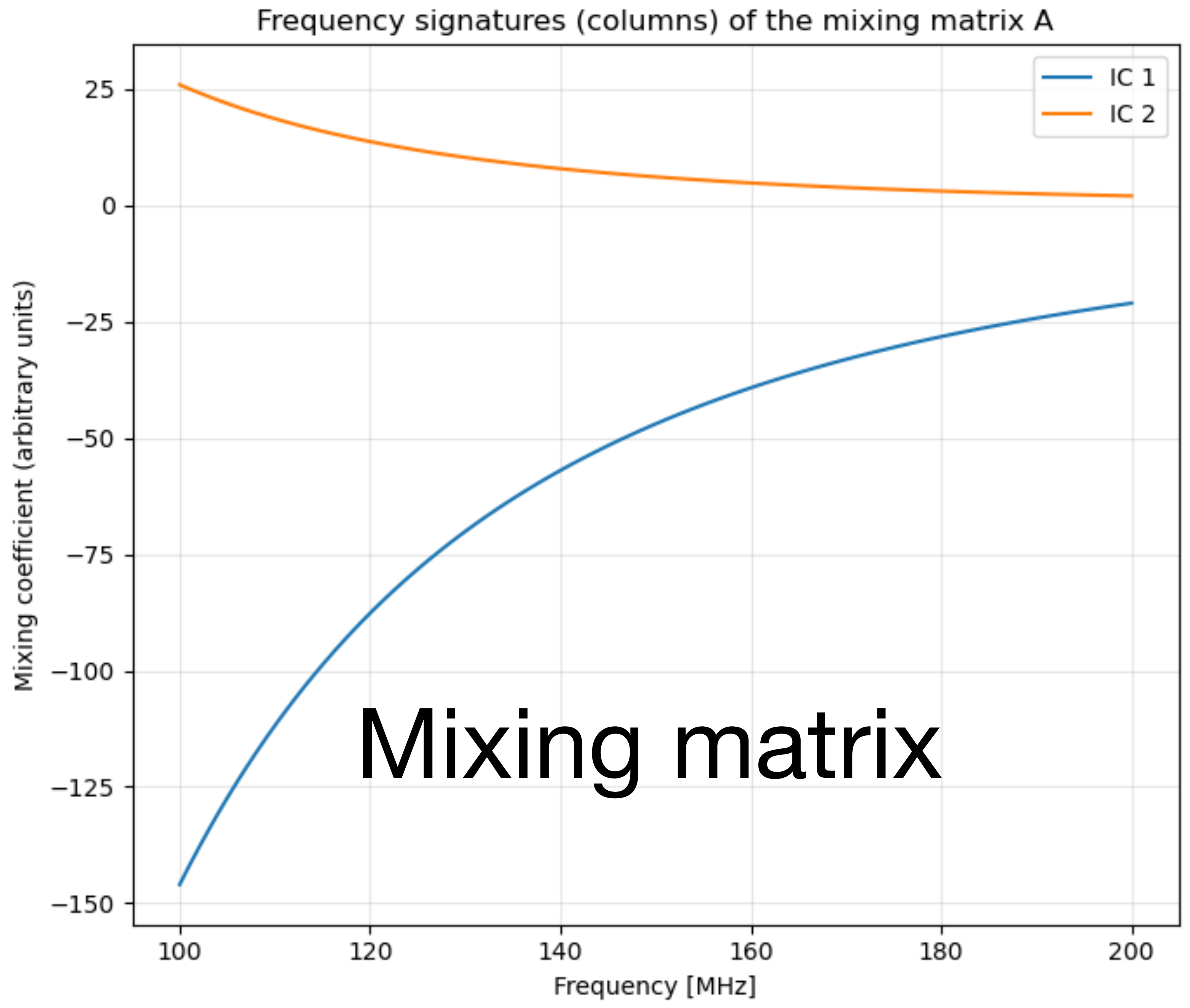
ICA on 3D Foreground Data Cube (n_comp=2)



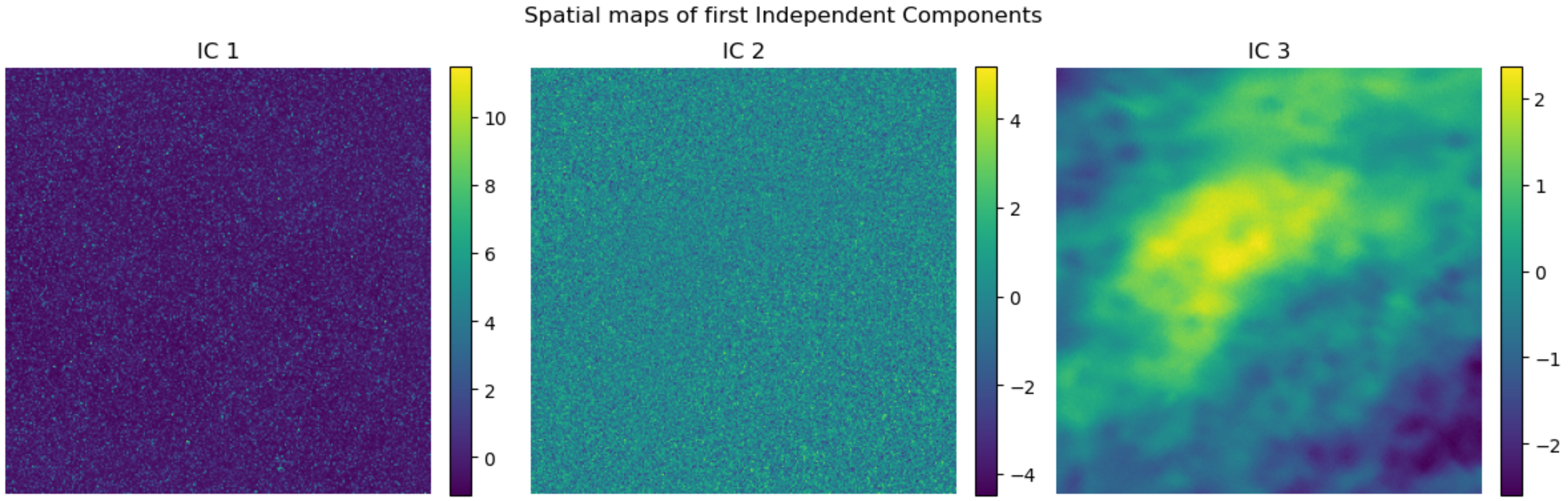
residuals ~ $10^{-2}\mathrm{K}$

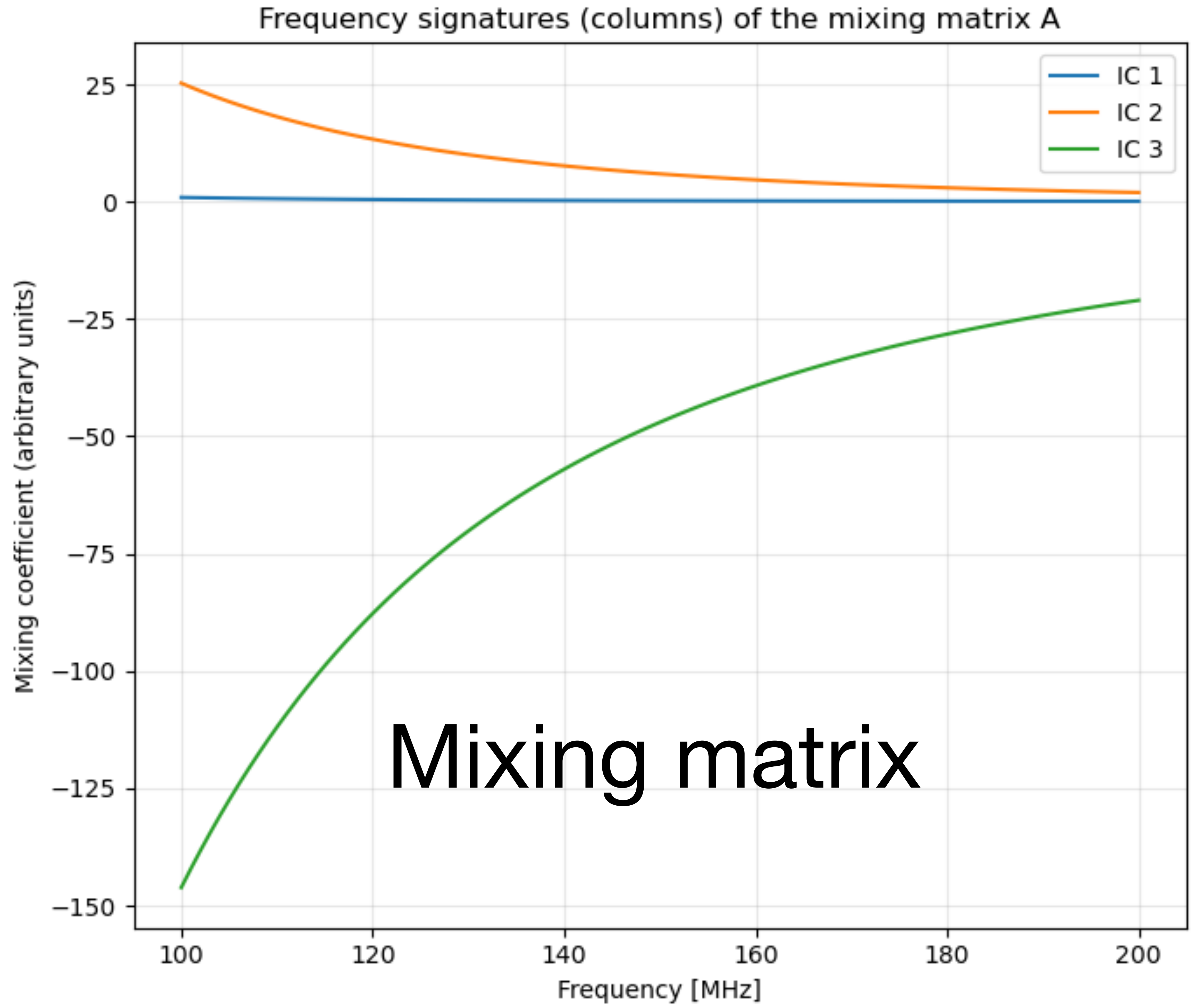


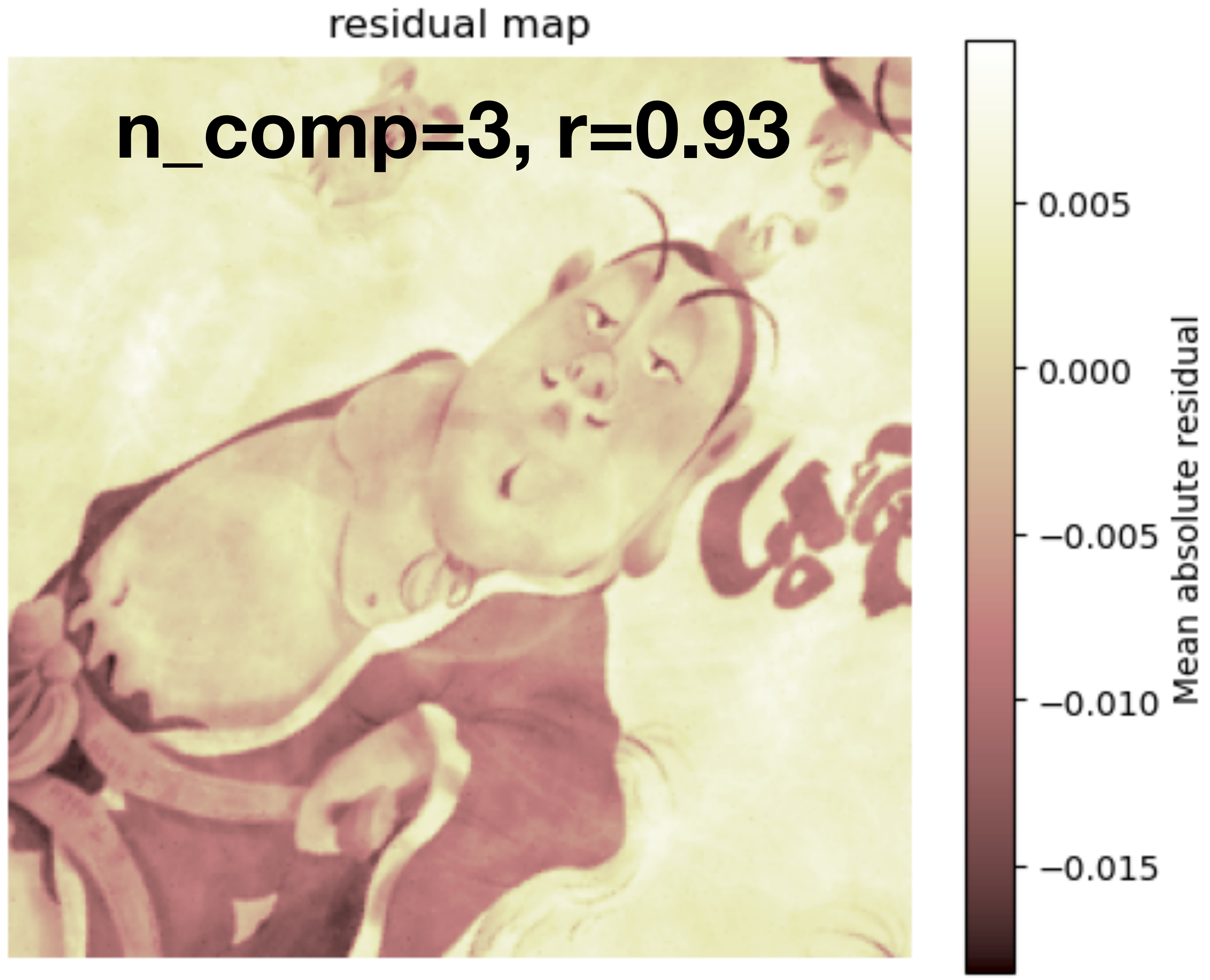




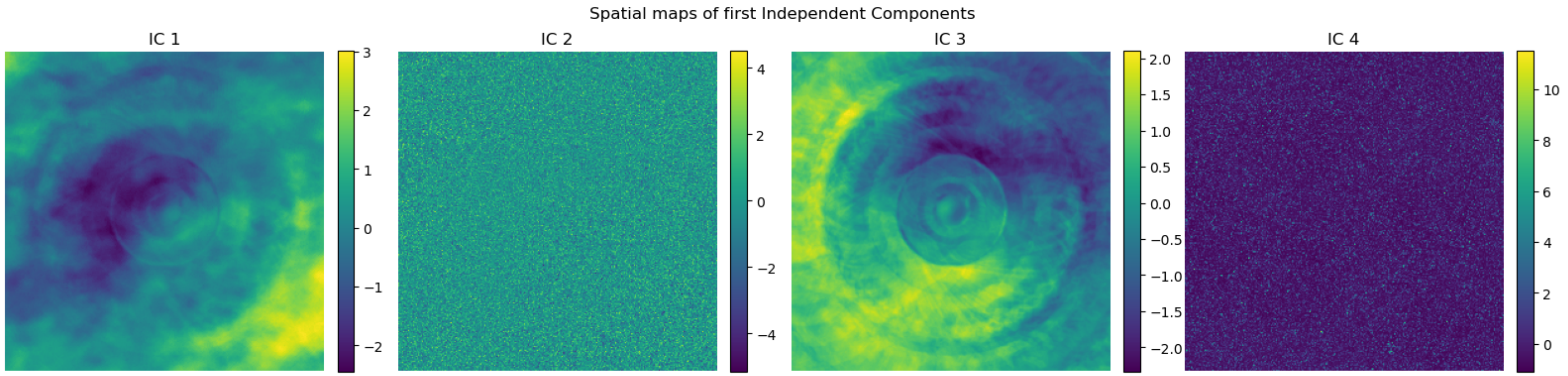
Adding EoR to data cube…

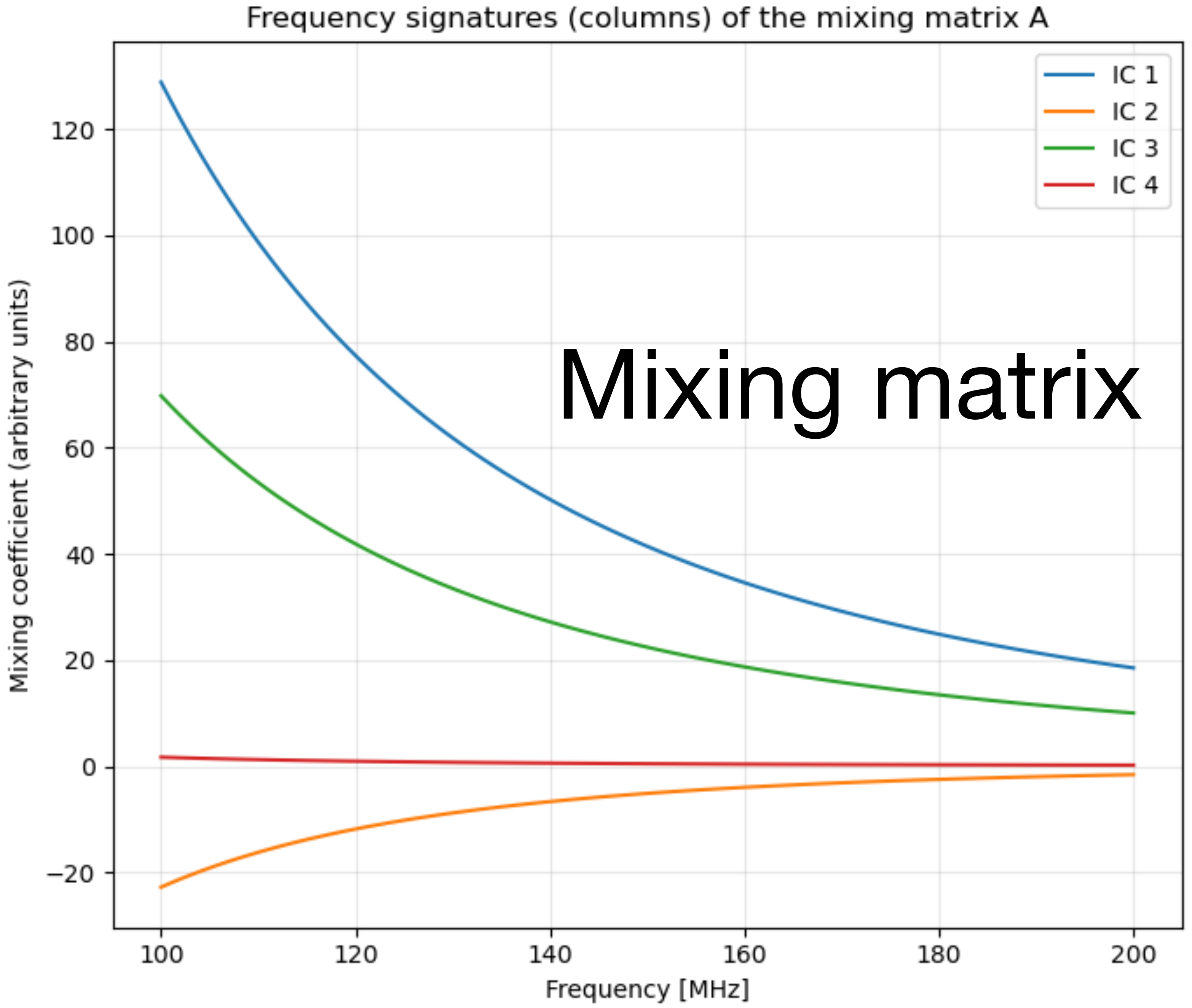


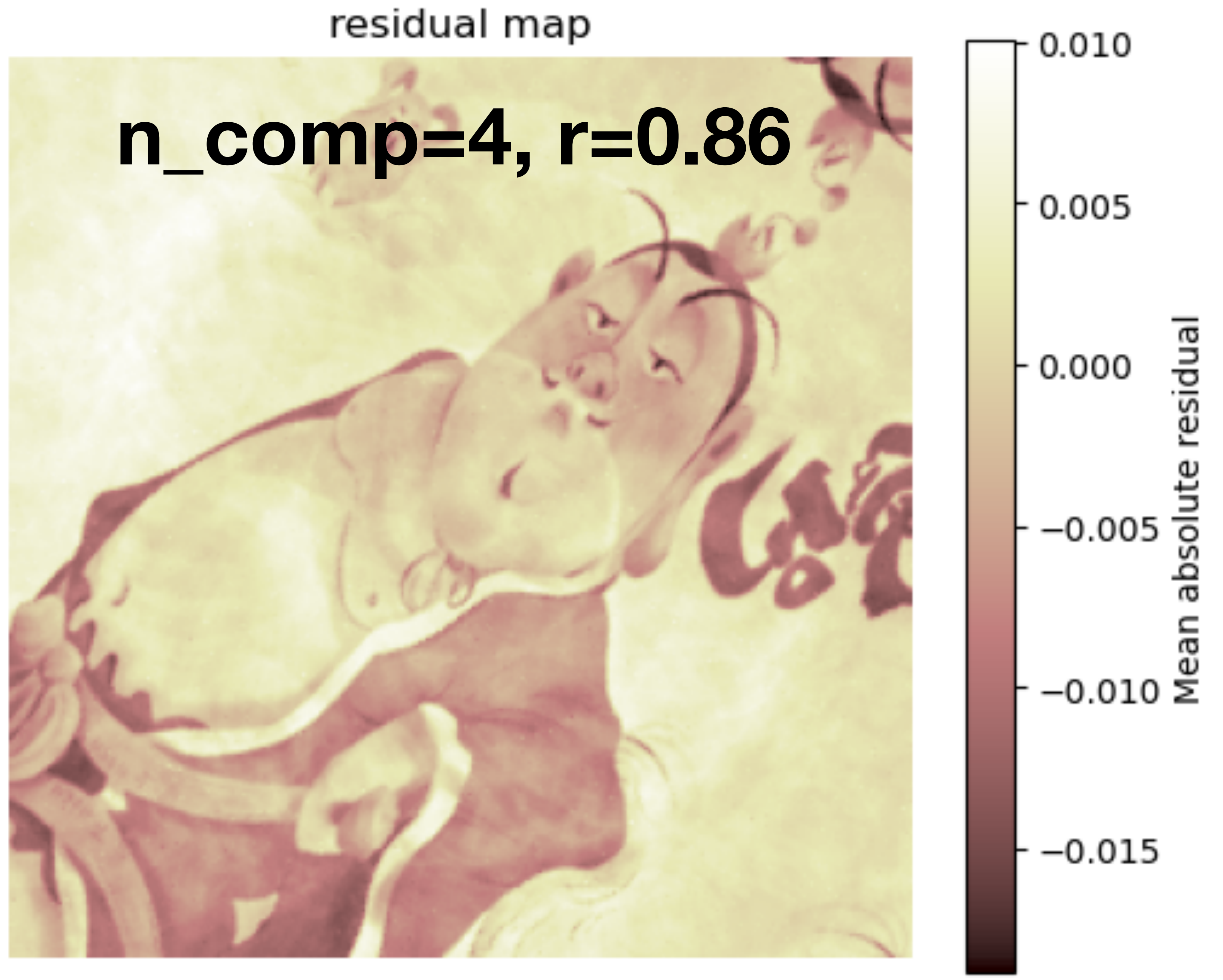




Adding EoR to data cube…



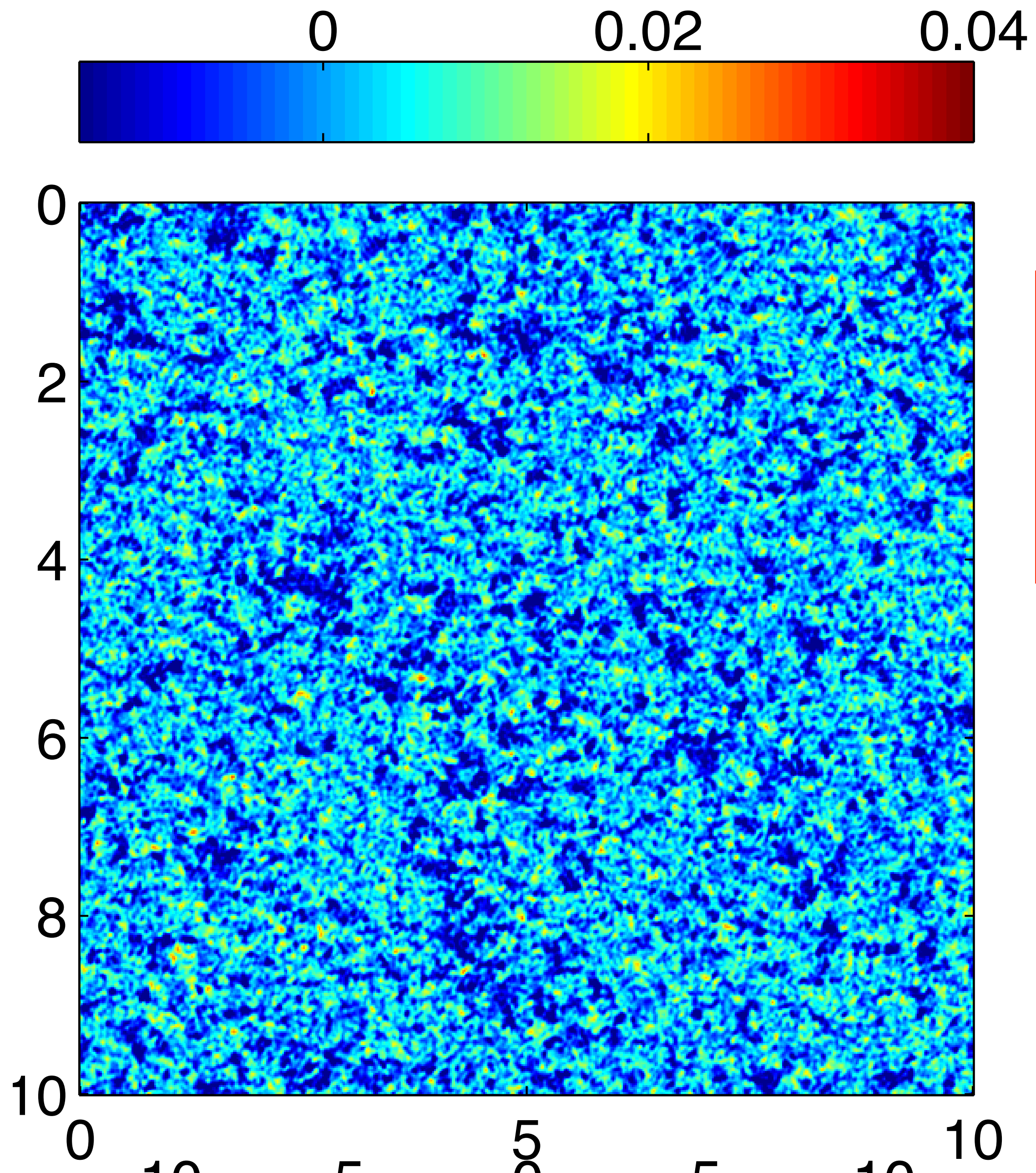




- Choosing the number of components in ICA is a critical and non-trivial decision
- Choose the smallest n_c after which residuals stop improving significantly

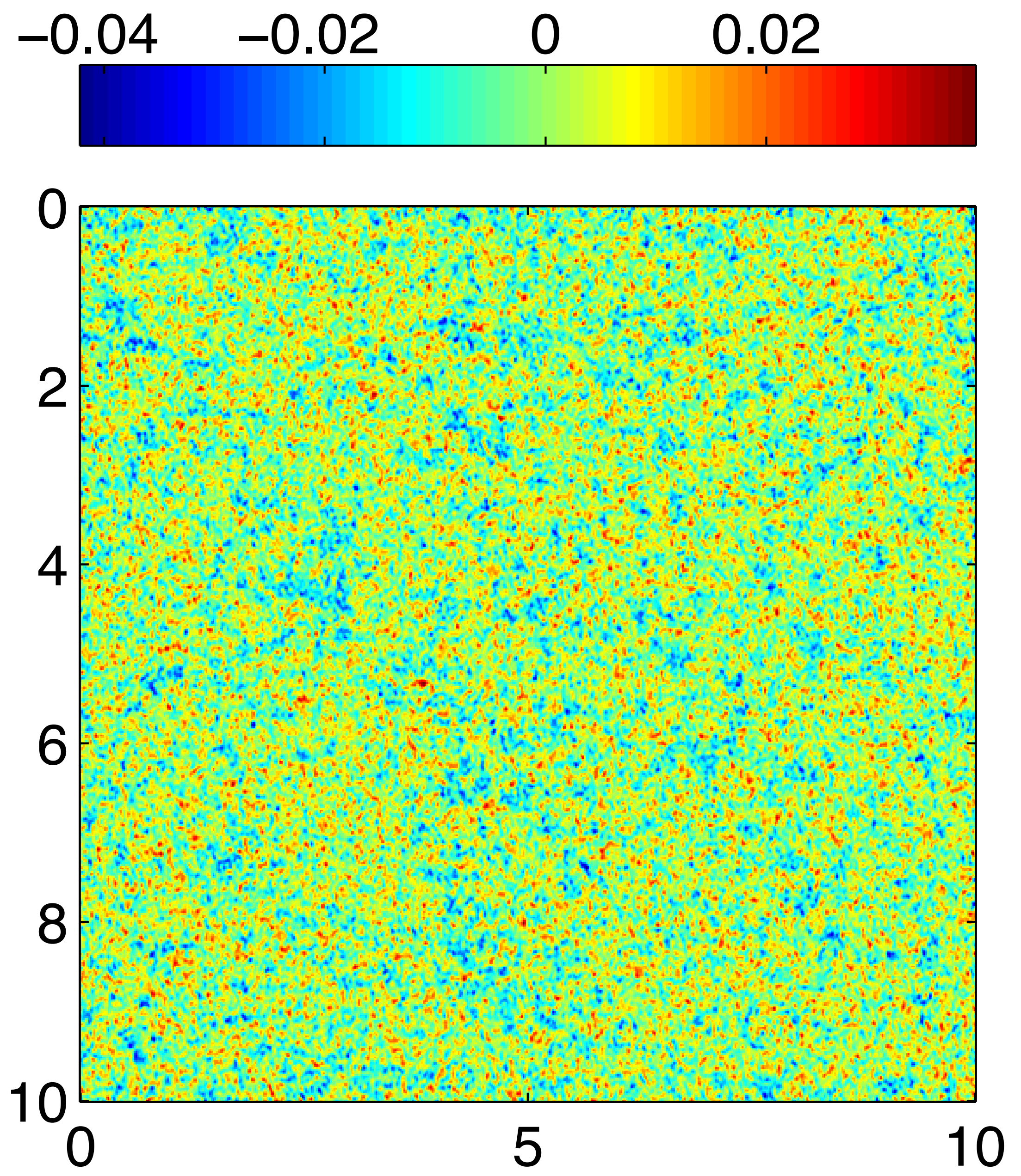
FASTICA

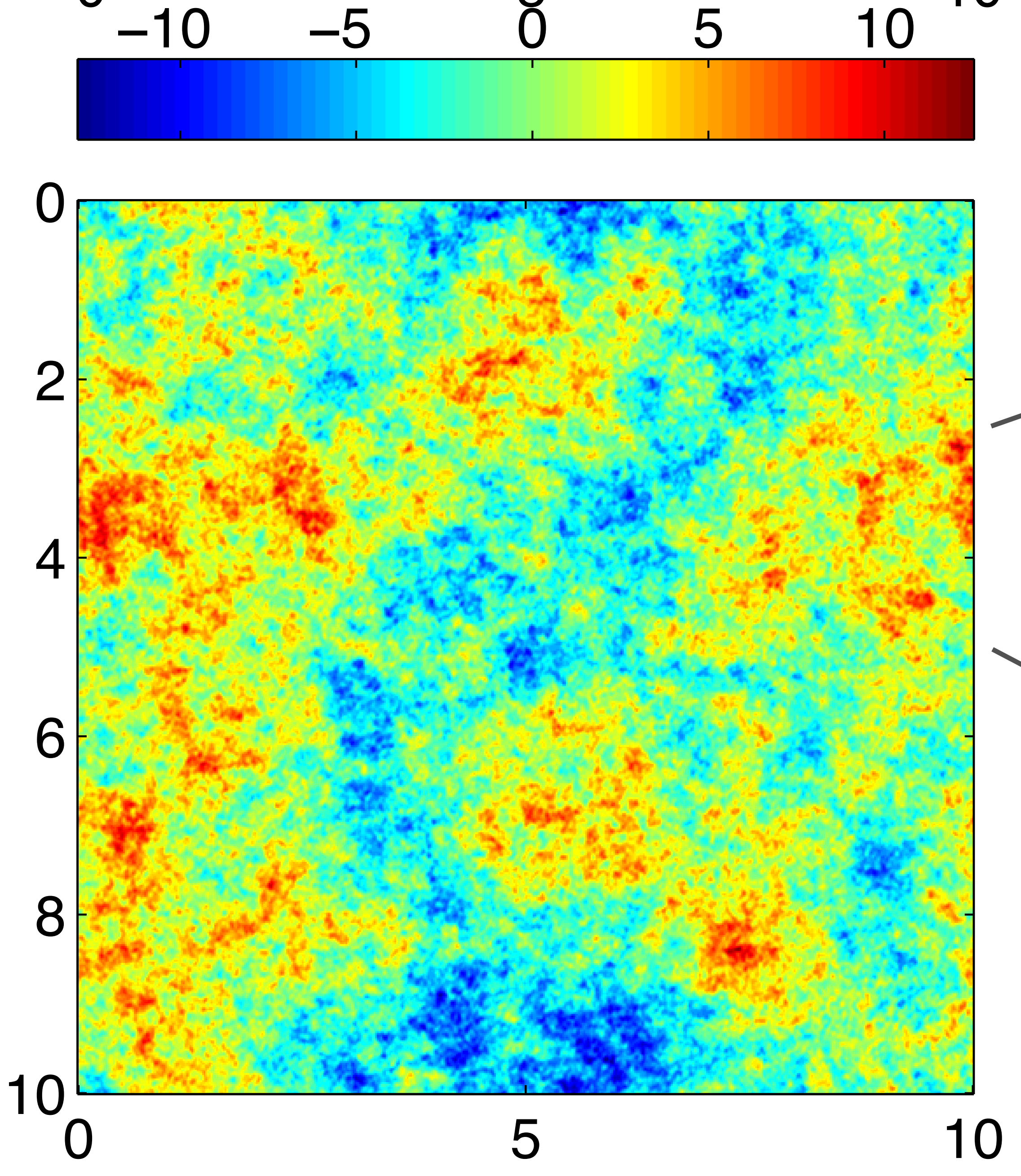
•Non-Gaussianity Maximization



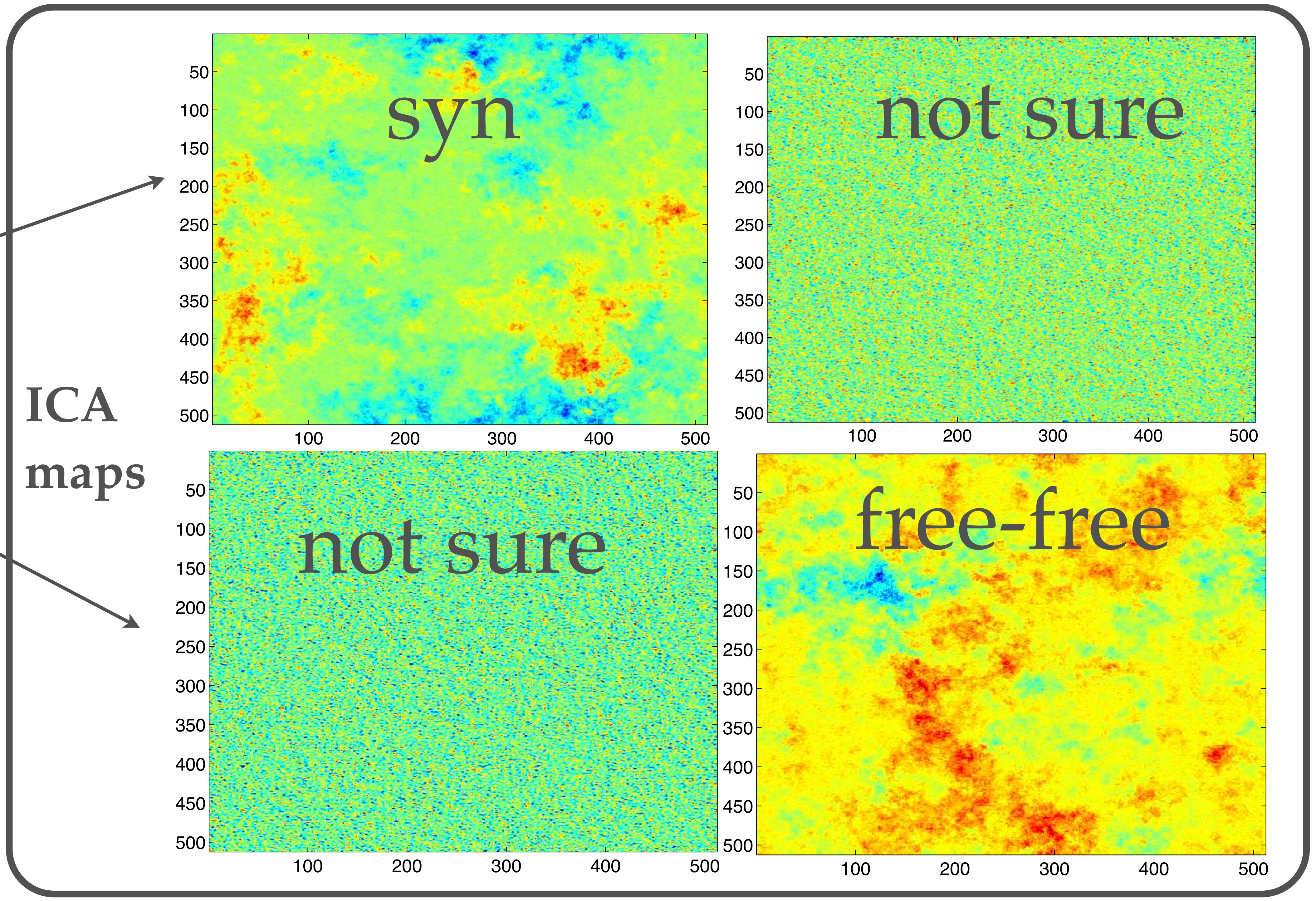
simulated 21cm

recovered 21cm





simulated foreground



Karhunen–Loève Transformation for foreground removal

- Generalization of PCA/SVD
- Extract 21cm signal from bright foregrounds rather than just removing high-variance foreground modes
- KL Basis: Eigenfunctions that jointly diagonalize signal and foreground covariance
- Lead to KL modes ranked by signal-to-foreground ratio
- large S/F modes preserve 21cm signal

KL transform is optimal if covariance S and F is accurately known!

KL Generalized Eigenvalue Problem in Frequency Space

Given:

- **Signal and foreground covariance:** $\mathbf{C}_s = \langle \mathbf{s}\mathbf{s}^T \rangle, \ \mathbf{C}_f = \langle \mathbf{f}\mathbf{f}^T \rangle$
- **Jointly diagonalize both matrices (eigenvalue problem) :** $\mathbf{C}_f \mathbf{e}_i = \lambda_i \mathbf{C}_s \mathbf{e}_i$

This gives:

- λ_i **: foreground-to-signal ratio for mode** i
- $\mathbf{e}_i$ **: KL eigen-mode**

Interpretation: $\mathrm{F/S} = \dfrac{\mathbf{e}_i^T \mathbf{C}_f \mathbf{e}_i}{\mathbf{e}_i^T \mathbf{C}_s \mathbf{e}_i} = \lambda_i$

- **Modes with small** λ_i **are signal-rich, foreground-poor**
- **Modes with large** λ_i **are foreground-dominated**

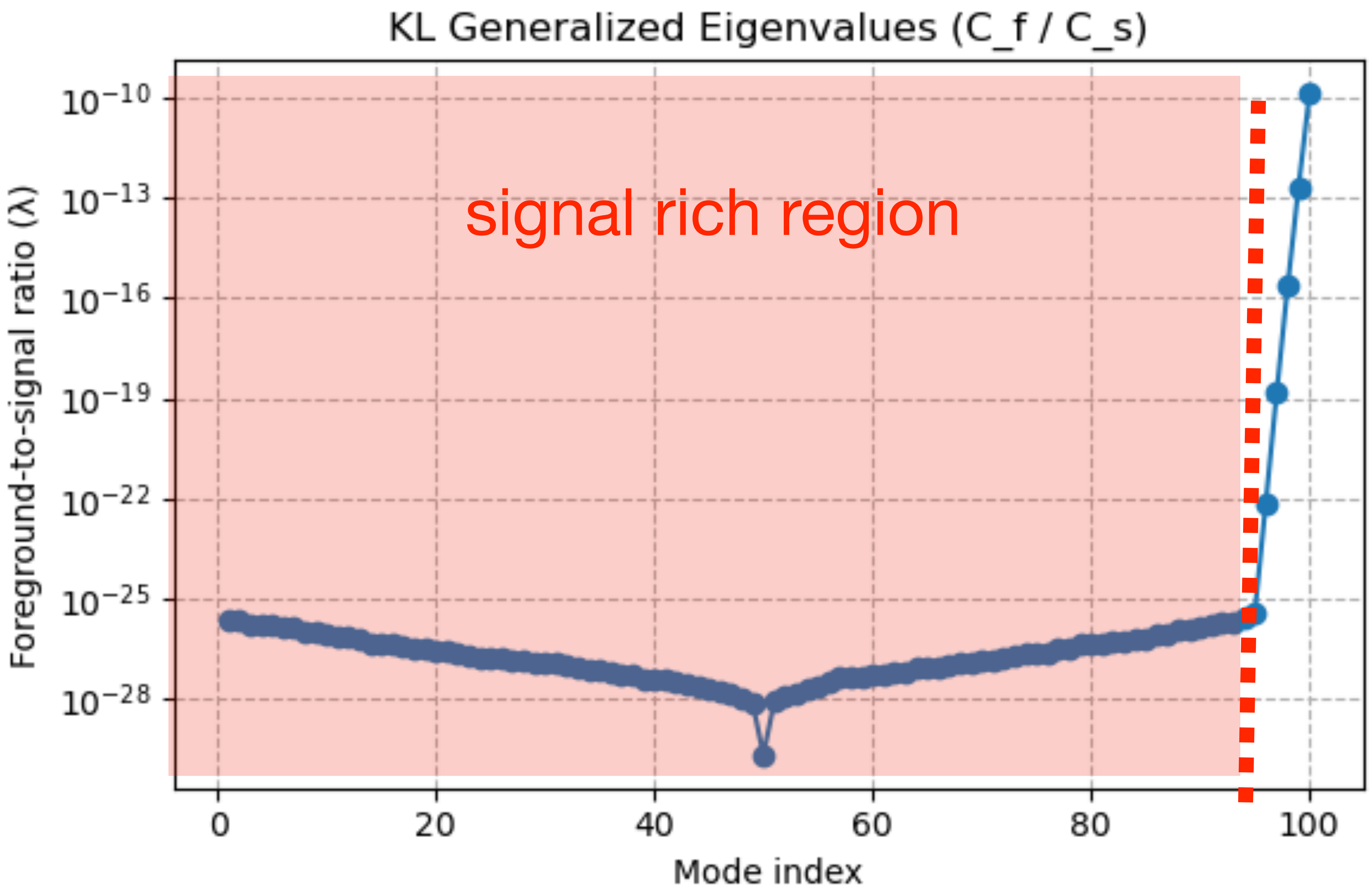
Implementation in 21cm Analysis

- C_s : modeled from simulations of EoR signal (Gaussian, isotropic)
- C_f : from astrophysical foreground models (e.g., correlated synchrotron model)
- Modeled analytically as $[C_f]_{\nu,\nu'} \propto \left(\frac{\nu\nu'}{\nu_0^2}\right)^{-\alpha} \exp\left(-\frac{\ln^2(\nu/\nu')}{2\xi^2}\right)$, $[C_s]_{\nu,\nu'} = A_s \exp\left(-\frac{(\nu-\nu')^2}{2\sigma_\nu^2}\right)$

Transforming the Data:

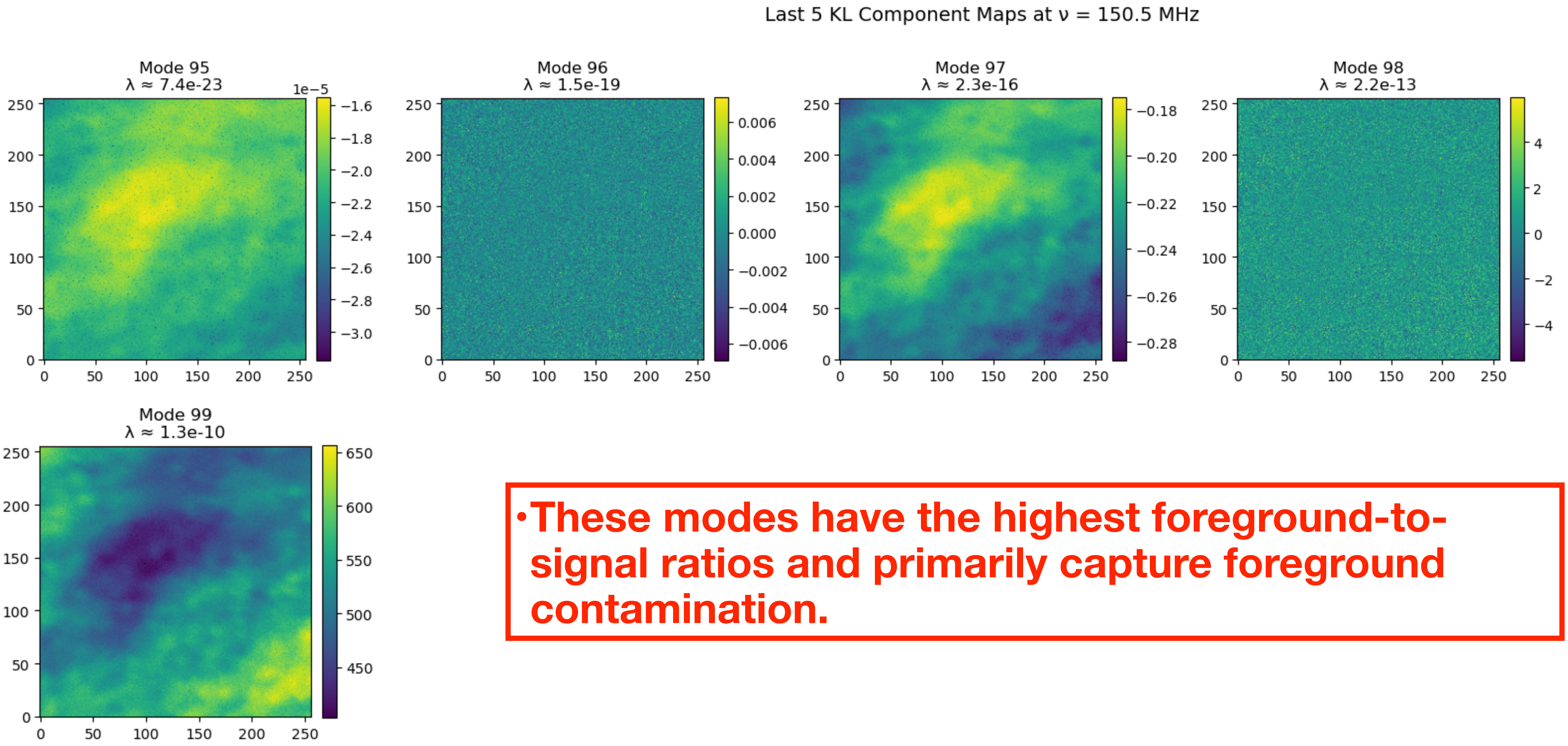
- Given observed spectrum $\mathbf{x} \in \mathbb{R}^{N_\nu}$, we project onto KL basis: $y_i = \mathbf{e}_i^T \mathbf{x}$ (KL coefficient for mode i)
- To reconstruct a cleaned signal: $\mathbf{x}_{\text{clean}} = \sum_{\lambda_i < \lambda_{\text{cut}}} y_i \mathbf{e}_i$

KL on 3D Foreground Data Cube



- **If eigenvalues are well separated (large gaps), the eigenvectors and eigenvalues are stable under small perturbations**
- **The cleaned data cube is less sensitive to model uncertainties in $\mathbf{C}_s$ and $\mathbf{C}_f$**

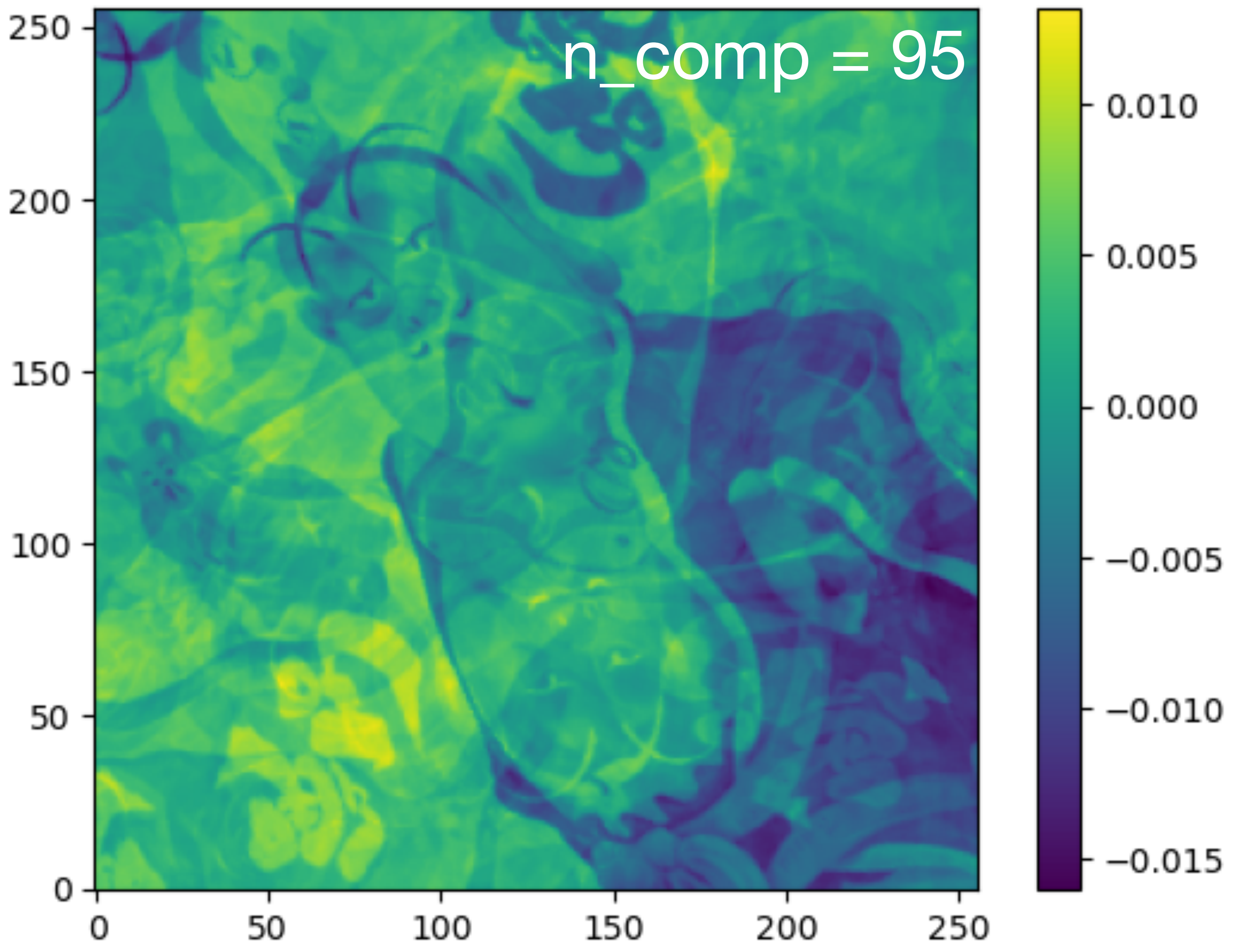
Last 5 KL Component Maps (Fg rich)



- These modes have the highest foreground-to-signal ratios and primarily capture foreground contamination.

Adding EoR to data cube…

Pearson correlation coefficient =0.95



- **Despite inaccuracies in the signal model, the KL method can still recover the signal effectively**

Adding noise to data cube…

• If noise covariance is known, KL can optimally separate signal, foreground, and noise—enhancing noise rejection and signal recovery.

- **Form the Total Noise + Foreground Covariance:**
- $\mathbf{C}_b = \mathbf{C}_f + \mathbf{C}_n$
- **Solve Generalized Eigenvalue Problem:**
- $\mathbf{C}_b\mathbf{e} = \lambda\mathbf{C}_s\mathbf{e}$
- **This finds eigenmodes that optimally separate signal from combined foreground + noise.**
- **Project data onto eigenmodes and filter modes with large λ (foreground + noise dominated) to clean the signal.**

Gaussian Process Regression (GPR) for 21cm Foreground Subtraction

Intuition:

- **Think of a Gaussian Process as an infinite collection of correlated Gaussian random variables, one for each value of ν.**

- **For any set of input points $\left\{\nu_1, \nu_2, \ldots, \nu_n\right\}$, the vector:**

  $\left[f\left(\nu_1\right), f\left(\nu_2\right), \ldots, f\left(\nu_n\right)\right]$ **is jointly Gaussian-distributed with:**

  Mean = $\mu(\nu)$**, Covariance** $= K\left(\nu_i, \nu_j\right)$

- **Foregrounds vary smoothly with frequency and the kernel K can capture this smoothness**

Gaussian Process Regression (GPR) for 21cm Foreground Subtraction

What is GPR?

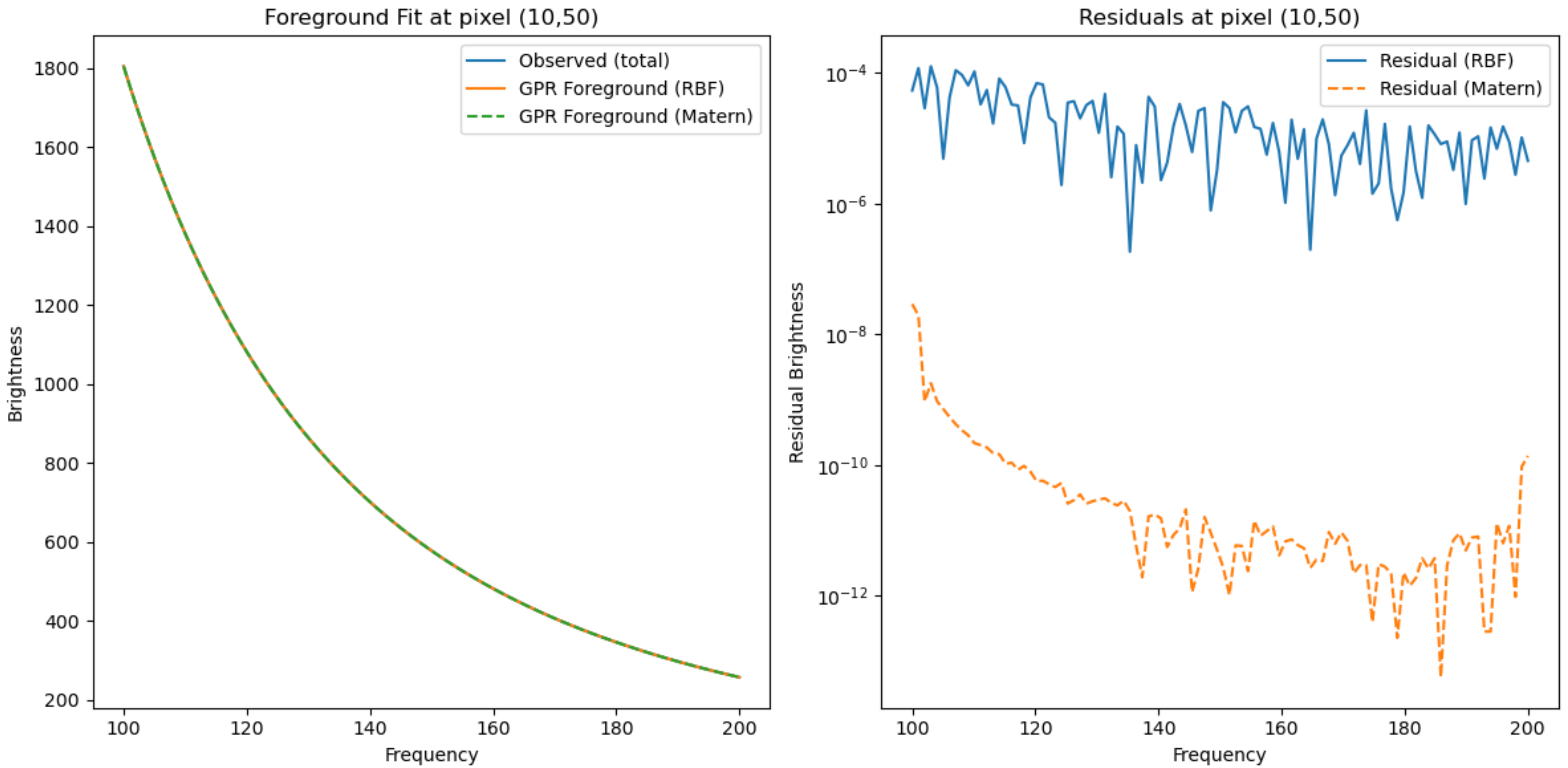
- **GPR is a Bayesian, non-parametric regression method for modeling smooth functions**
- **Models functions as distributions over smooth curves:** $f(\nu) \sim \mathbf{GP}\left(\mu, K\left(\nu, \nu'\right)\right)$
- **Uses a kernel function K to define correlations between data points (e.g., between frequencies)**
- **Covariance $K\left(\nu, \nu'\right)$: kernel encodes smoothness**
- **Outputs both a mean prediction (foreground model) and uncertainty**

GPR Kernel Choice – RBF vs. Matérn

- **RBF (Squared Exponential) Kernel：** $K_{\mathrm{RBF}}(\nu,\nu') = \sigma_f^2 \exp\left(-\frac{(\nu-\nu')^2}{2\ell^2}\right)$
  - **Infinitely differentiable (very smooth)**
  - **Can over-smooth and remove real 21 cm signal (especially low-$k_{\|}$)**
- **Matérn Kernel：** $K_{\mathrm{Matern}}(\nu,\nu') = \sigma_f^2 \cdot \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}\,|\nu-\nu'|}{\ell}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}\,|\nu-\nu'|}{\ell}\right)$
  - **v=1.5 :** $K(\nu,\nu') = \sigma_f^2\left(1+\frac{\sqrt{3}r}{\ell}\right)\exp\left(-\frac{\sqrt{3}r}{\ell}\right)$**, v=0.5:** $K(\nu,\nu') = \sigma_f^2 \exp\left(-\frac{r}{\ell}\right)$
- **For a dataset $\mathbf{y}$ observed at inputs $\mathbf{X}$, the $\log$ marginal likelihood is:**

$$\log p(\mathbf{y} \mid \mathbf{X}, \theta) = -\frac{1}{2}\mathbf{y}^\top \mathbf{K}^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K}| - \frac{n}{2}\log(2\pi)$$

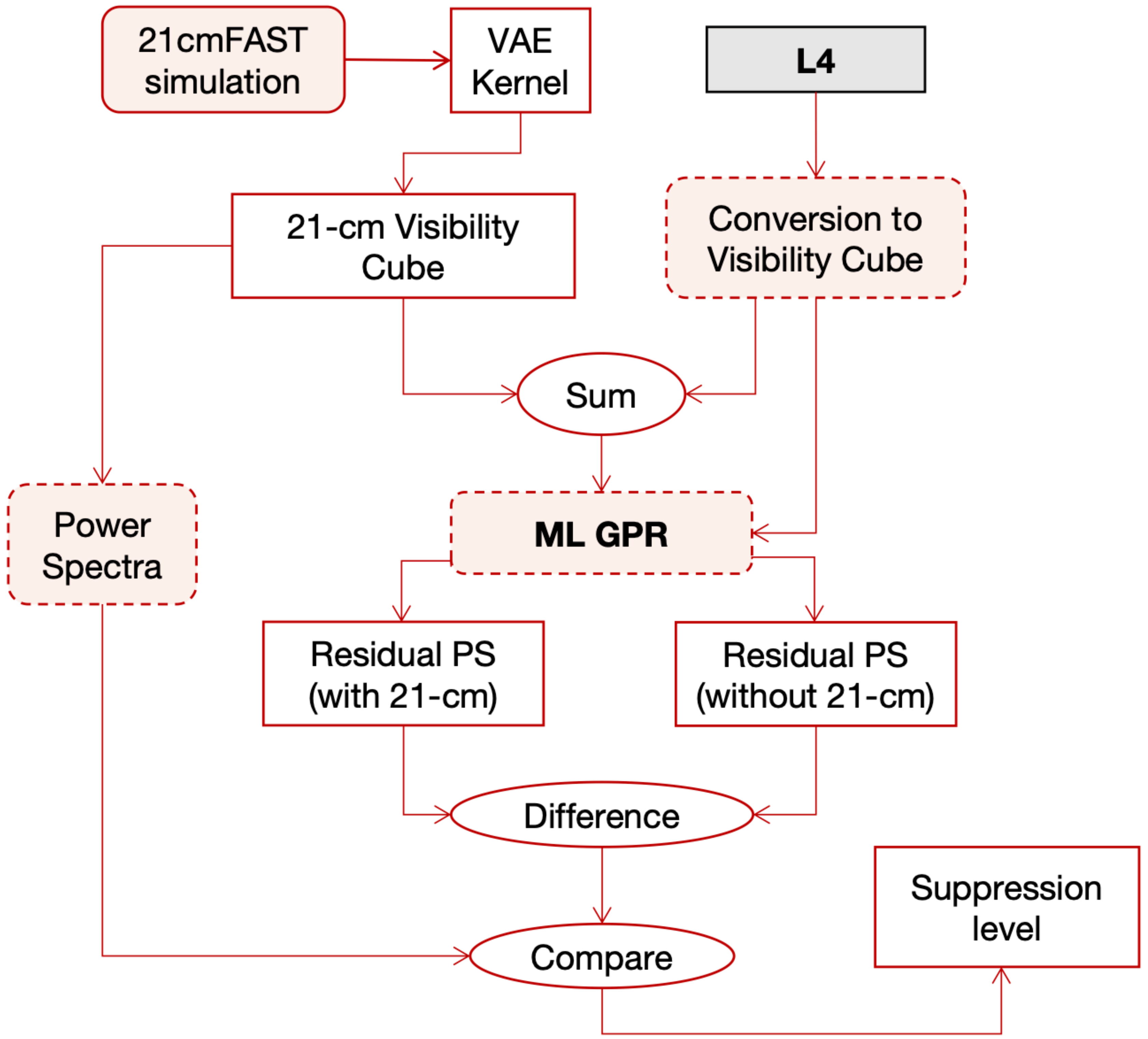
- **This acts as an objective function for hyperparameter optimization**

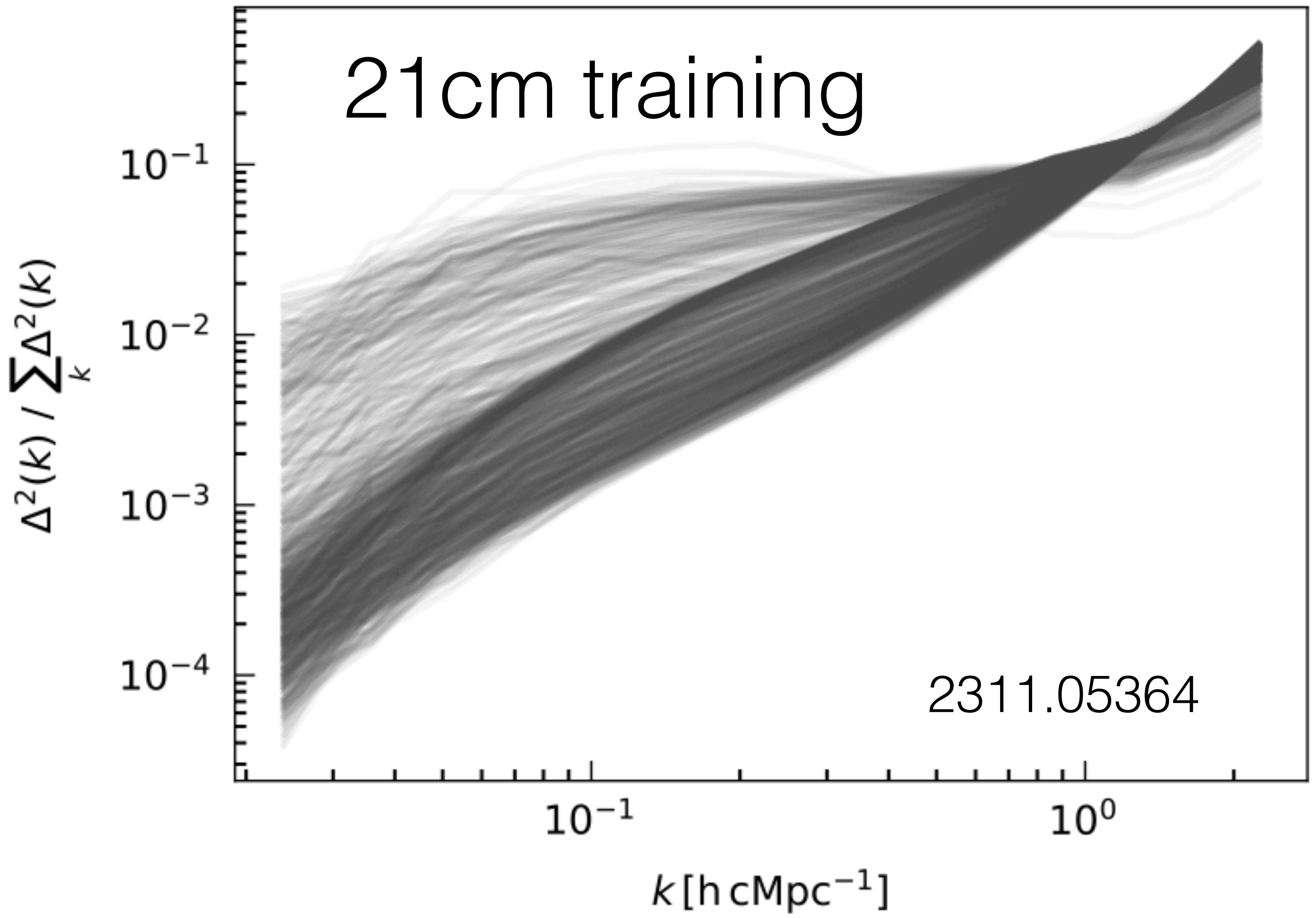


Foreground Fit at pixel (10,50)
Observed (total)
GPR Foreground (RBF)
GPR Foreground (Matern)
Brightness
Frequency
1800
1600
1400
1200
1000
800
600
400
200
100
120
140
160
180
200
Residuals at pixel (10,50)
Residual (RBF)
Residual (Matern)
Residual Brightness
Frequency
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}

Machine learning GPR

Component	Covariance	Parameter	Description	Prior Bounds	Estimated Value
Intrinsic foregrounds	Radial Basis Function	$\sigma^2_{\rm int}$ $l_{\rm int}$	Variance Lengthscale	[−1,1] [20,40]	−0.1 ± 0.02 32.9 ± 1.4
Mode-mixing foregrounds	Radial Basis Function	$\sigma^2_{\rm mix}$ $l_{\rm mix}$	Variance Lengthscale	[−2,0] [0.1,0.5]	−1.32 ± 0.01 0.275 ± 0.001
21-cm signal	Trained ML Kernel	x_1 x_2 σ^2_{21}	Latent space dimension Latent space dimension Variance	[−4,4] [−4,4] [−7,−1]	- - < -4.97
Excess power	Exponential Function	$\sigma^2_{\rm ex}$ $l_{\rm ex}$	Variance Lengthscale	[−3,−1] [0.2,2]	−1.88 ± 0.02 0.56 ± 0.03





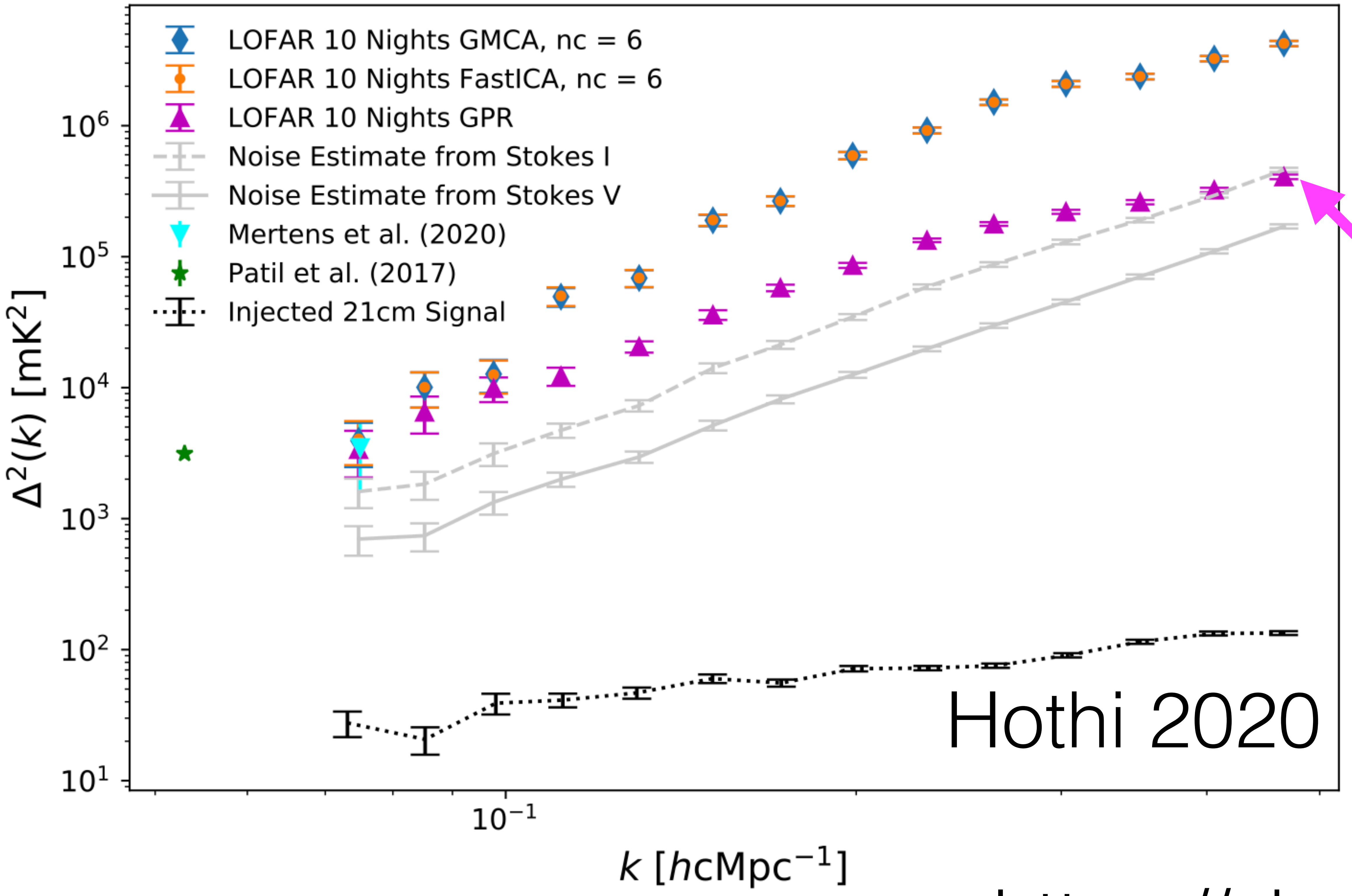
Gaussian process regression on LOFAR

$$\kappa_{Matern}(r) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu r}}{l}\right)^{\nu} K_{\nu}\left(\frac{\sqrt{2\nu r}}{l}\right)$$

$$E(\mathbf{f}_{\mathrm{fg}}) = \boldsymbol{K}_{\mathrm{fg}}\left[\boldsymbol{K}_{\mathrm{fg}} + \boldsymbol{K}_{21} + \boldsymbol{K}_{\mathrm{n}}\right]^{-1}\mathbf{d},$$

$$\begin{bmatrix}\mathbf{d}\\ \mathbf{f}_{\mathrm{fg}}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\boldsymbol{K}_{\mathrm{fg}} + \boldsymbol{K}_{21} + \sigma_n^2 I & \boldsymbol{K}_{\mathrm{fg}}\\ \boldsymbol{K}_{\mathrm{fg}} & \boldsymbol{K}_{\mathrm{fg}}\end{bmatrix}\right),$$

$$\mathrm{cov}(\mathbf{f}_{\mathrm{fg}}) = \boldsymbol{K}_{\mathrm{fg}} + \boldsymbol{K}_{\mathrm{fg}}\left[\boldsymbol{K}_{\mathrm{fg}} + \boldsymbol{K}_{21} + \boldsymbol{K}_{\mathrm{n}}\right]^{-1}\boldsymbol{K}_{\mathrm{fg}}.$$

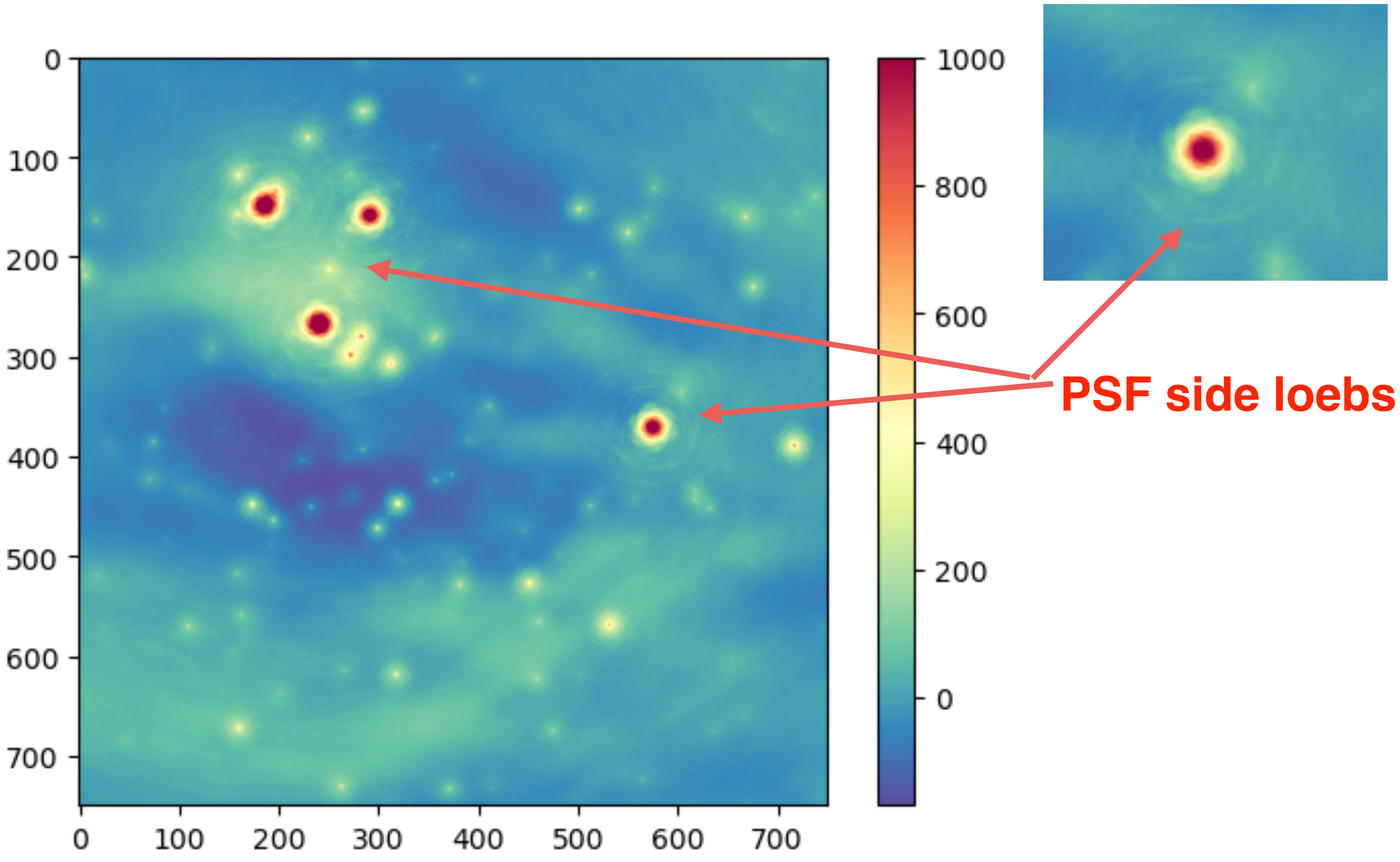


https://sheffieldml.github.io/GPy/

In a realistic situtation…

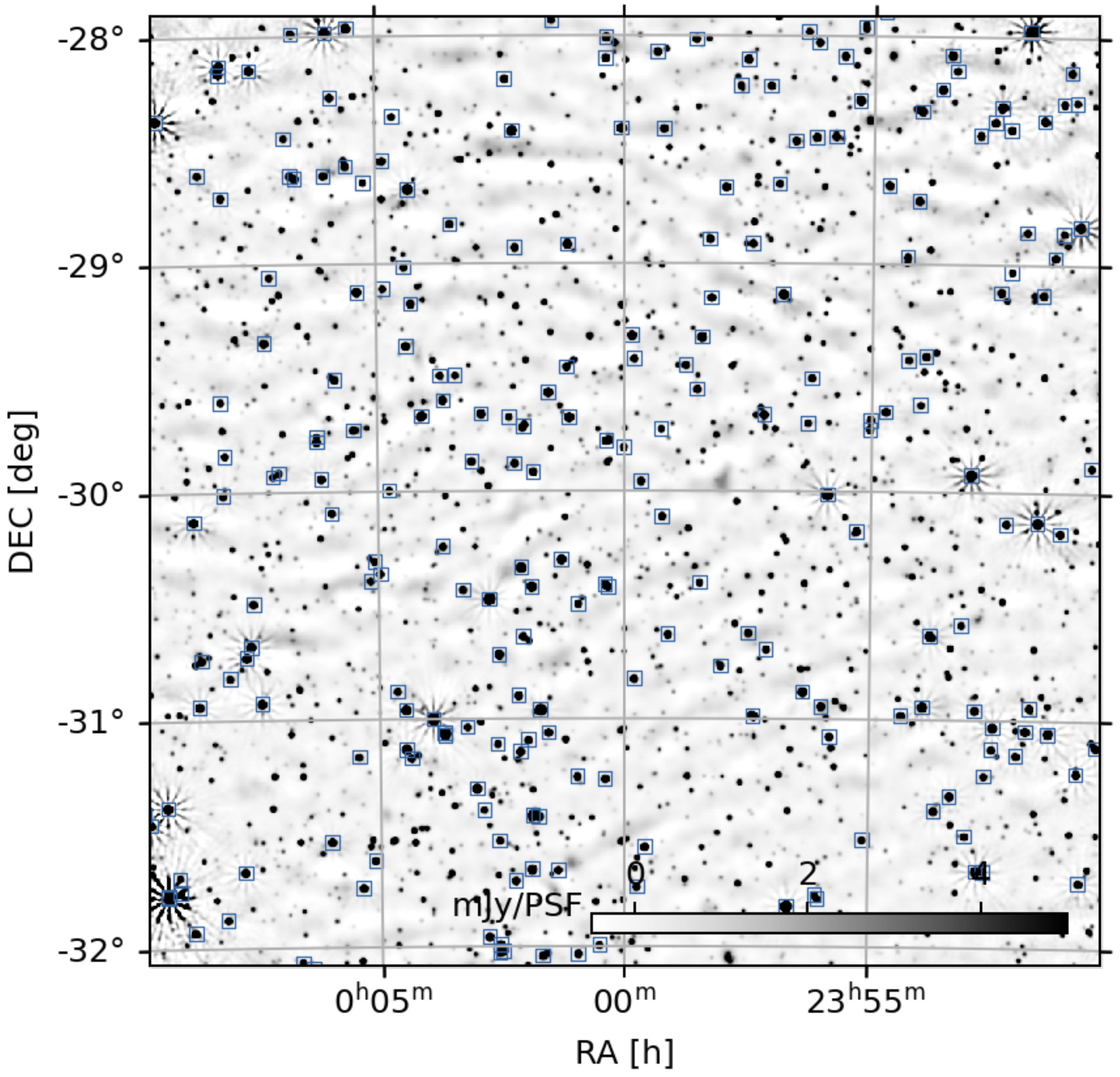
Challenge: frequency-dependent PSF

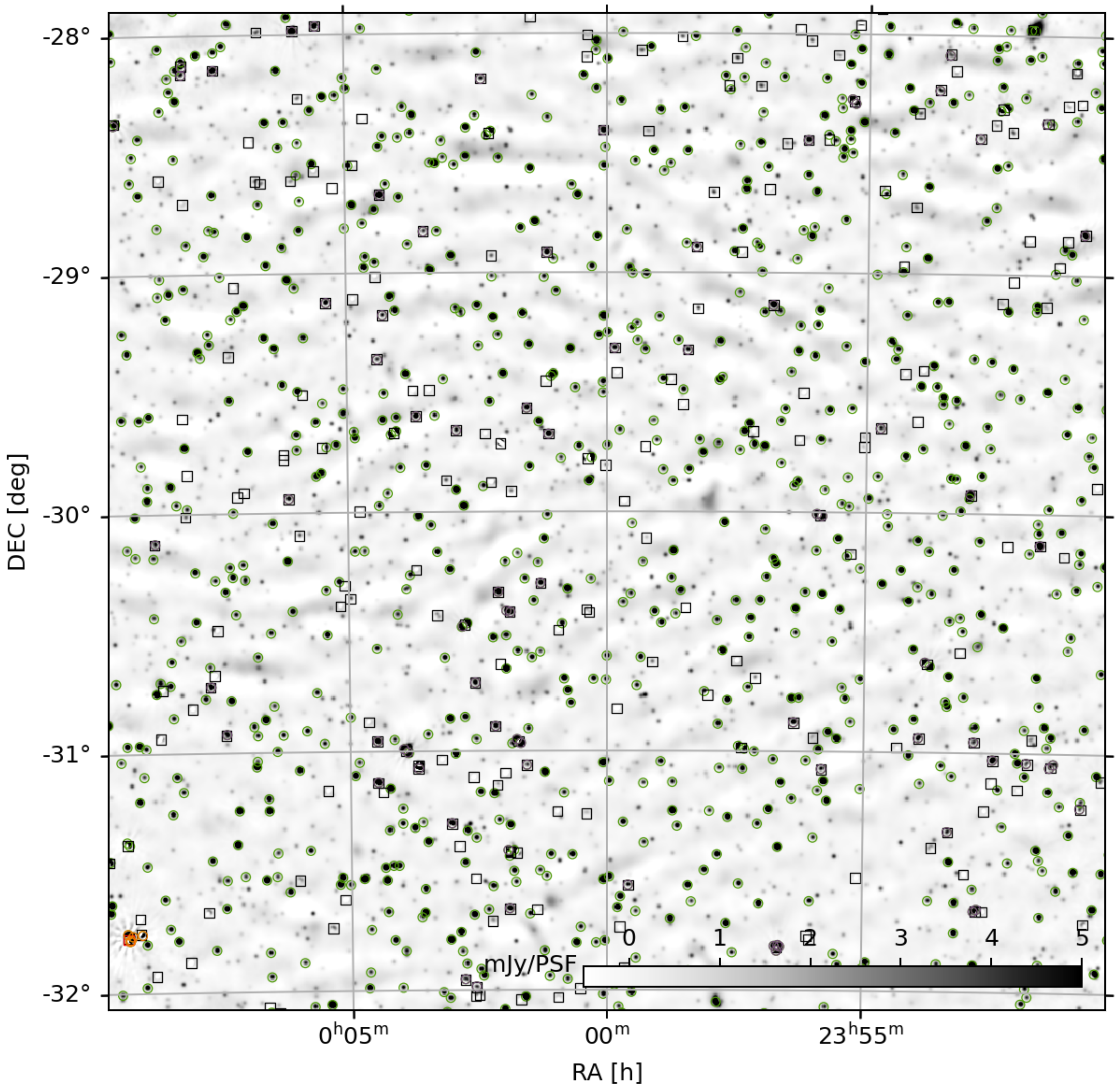
→ mode-mixing → non-smooth FG spectrum



Compact sources detected for subtraction —**DOTSS-21 team**

Iteratively subtract point sources and refine the FG model





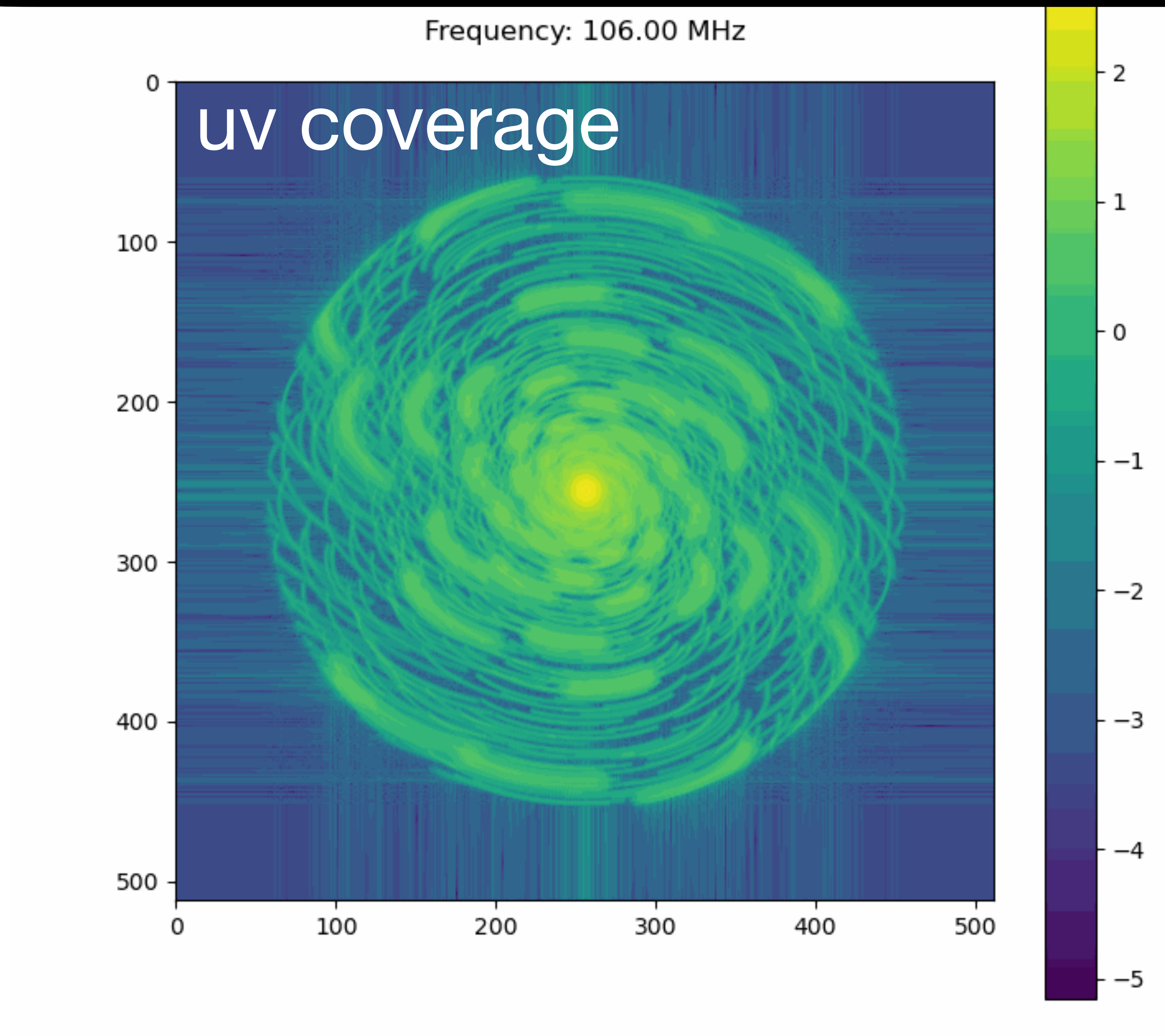
very massive!

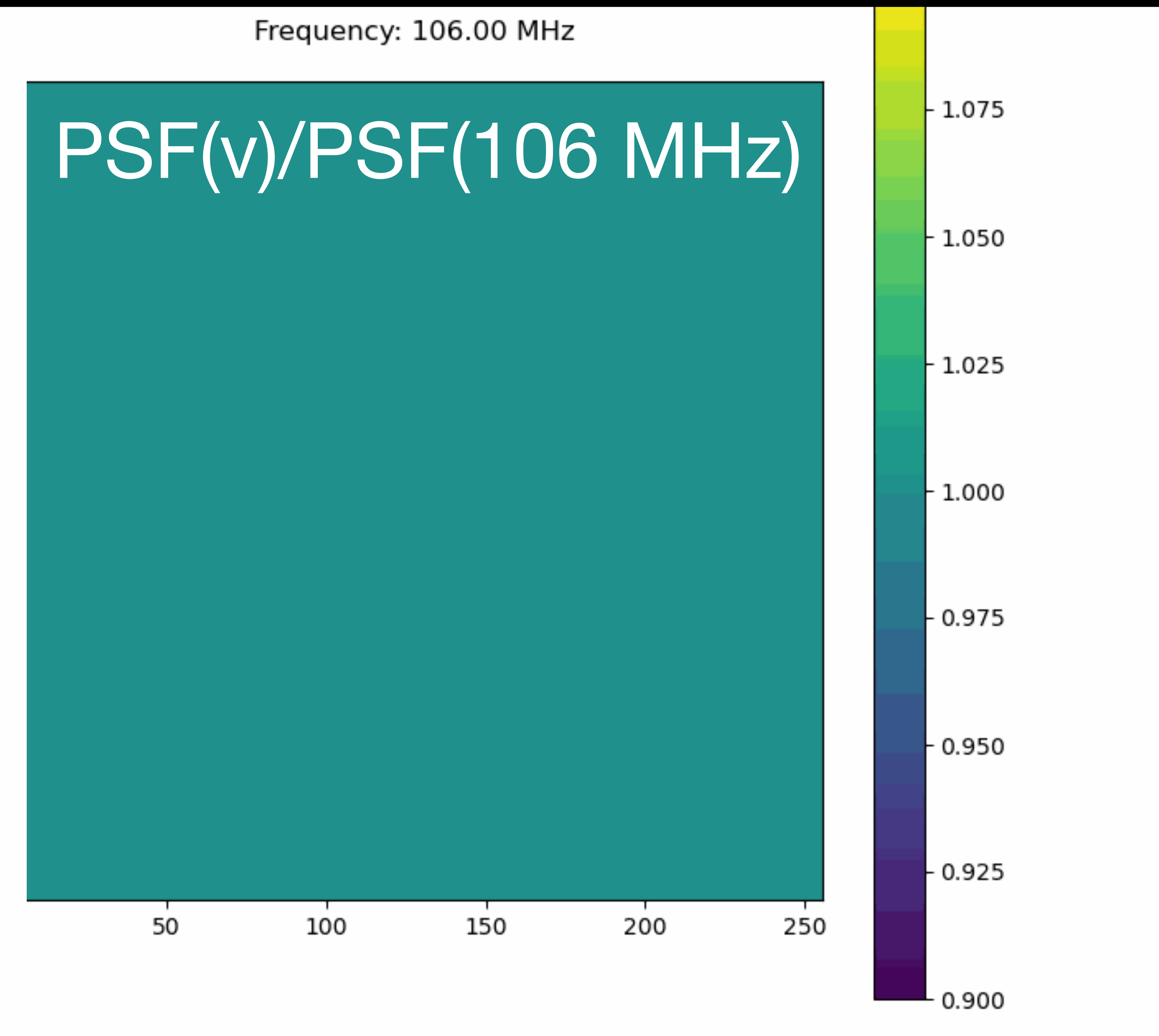
All FG cleaning algorithms are valid if no instrumental effects!

However…

Real-world Challenges:

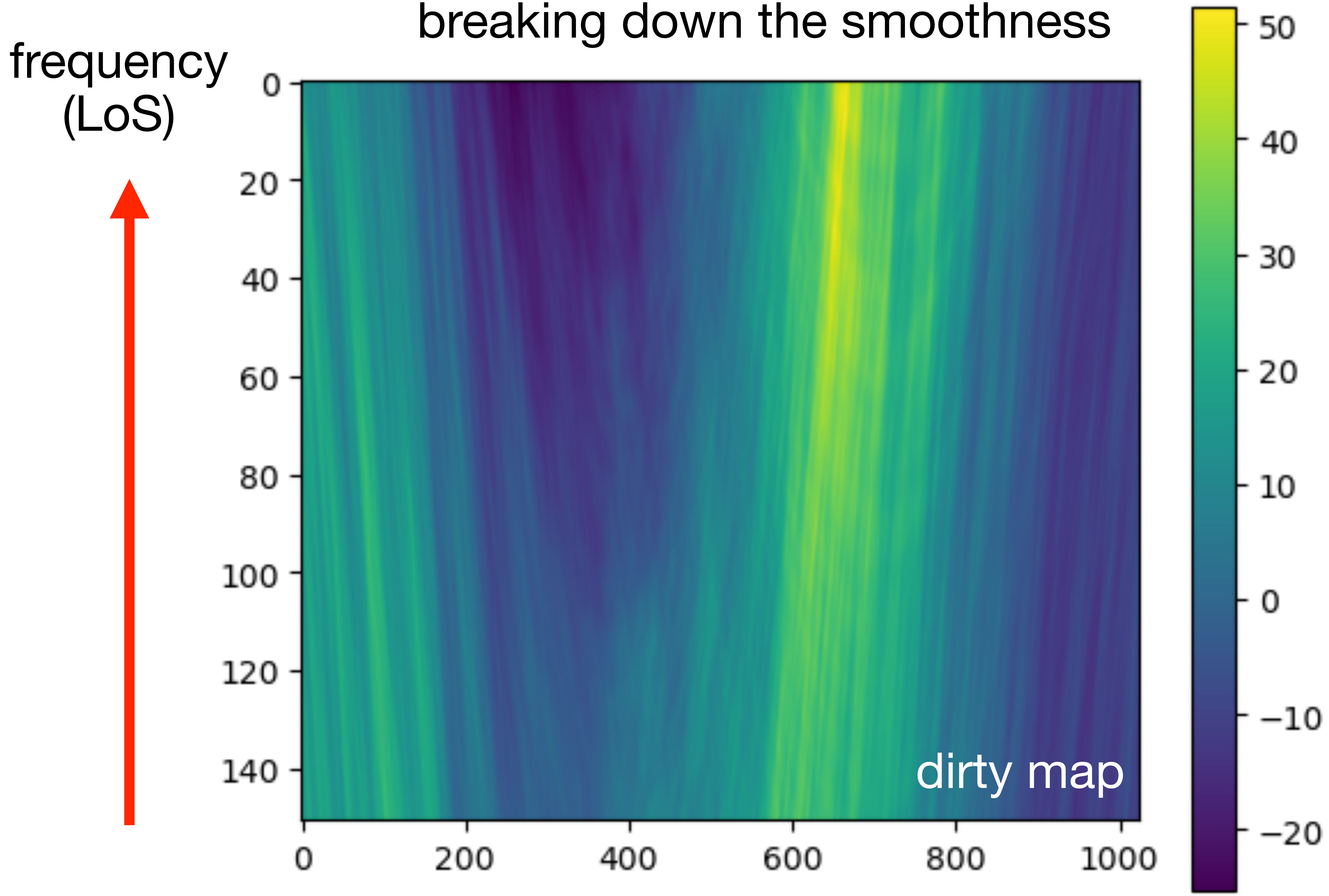
- “Mode-mixing” breaks the smoothing and prevents foreground removal
- “Mode-mixing” producing non-smooth FG spectrum
- PSF deconvolution — an ill-posed inverse problem; achieving the desired precision of 1 in 10,000 is not feasible





Real-world Challenge: Mode-mixing

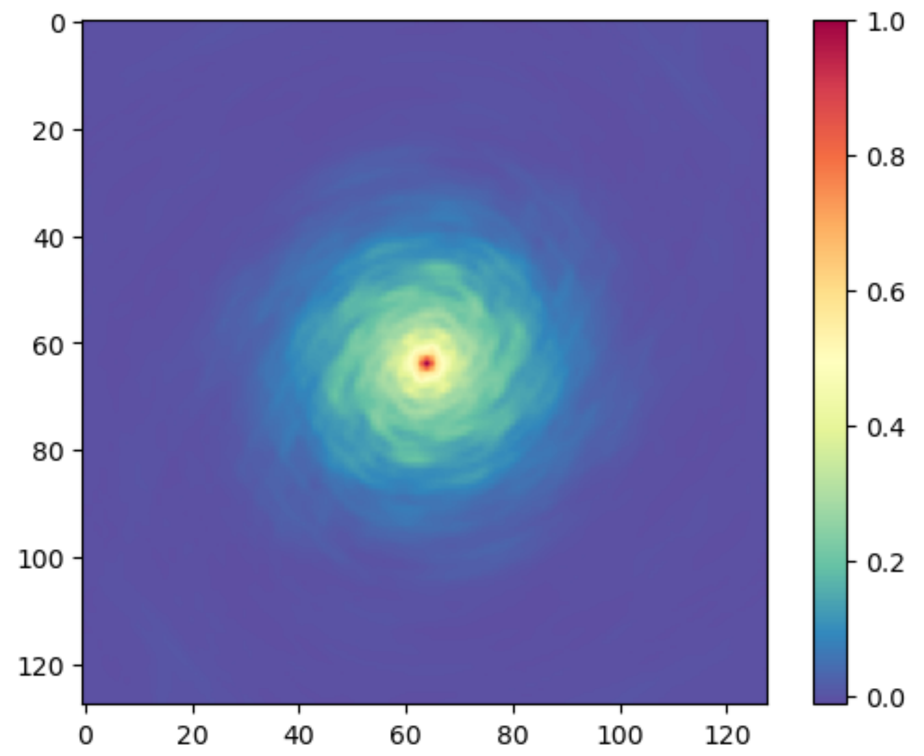
PSF varied with frequency,
breaking down the smoothness



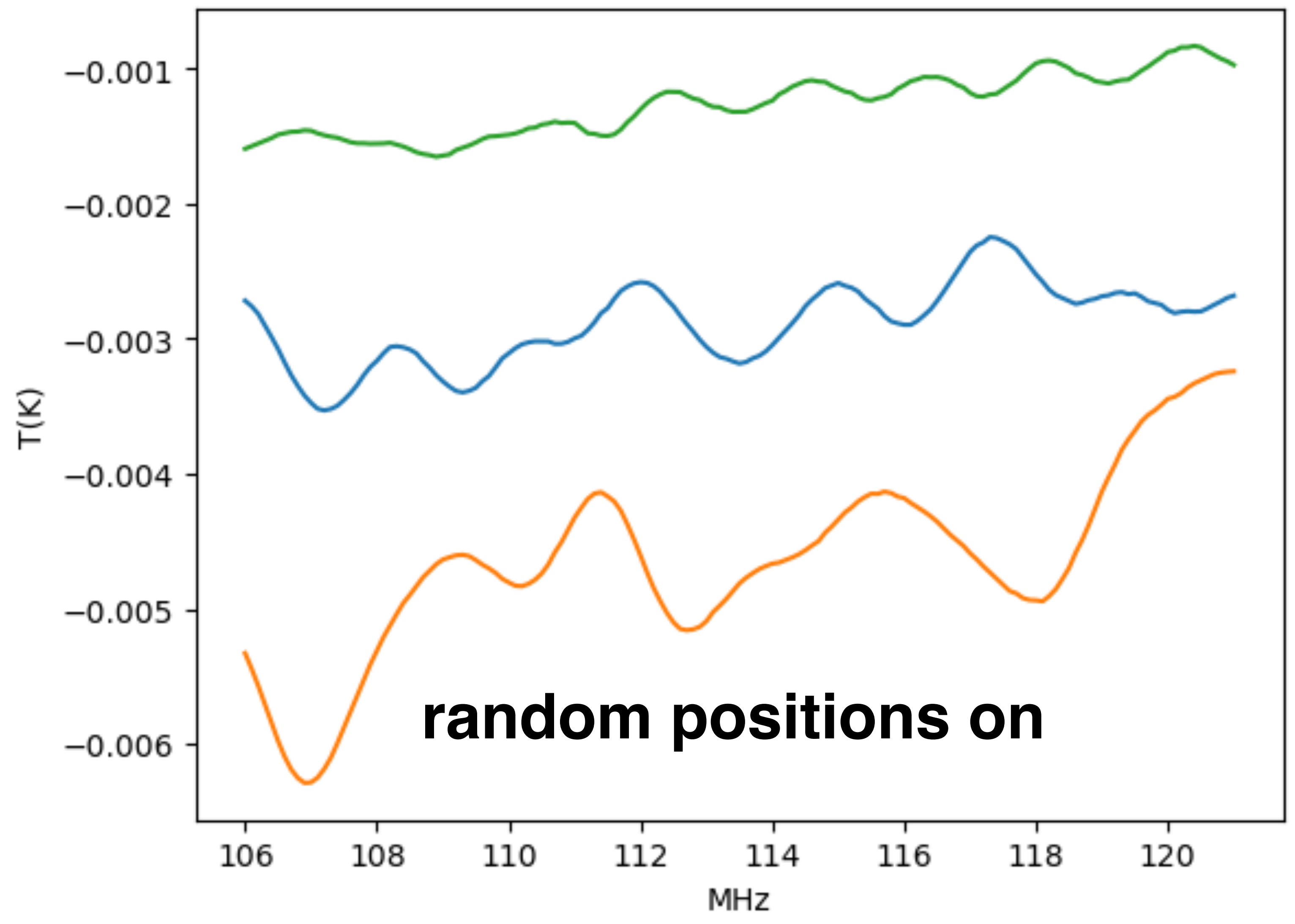
Key Challenge: frequency-dependent PSF

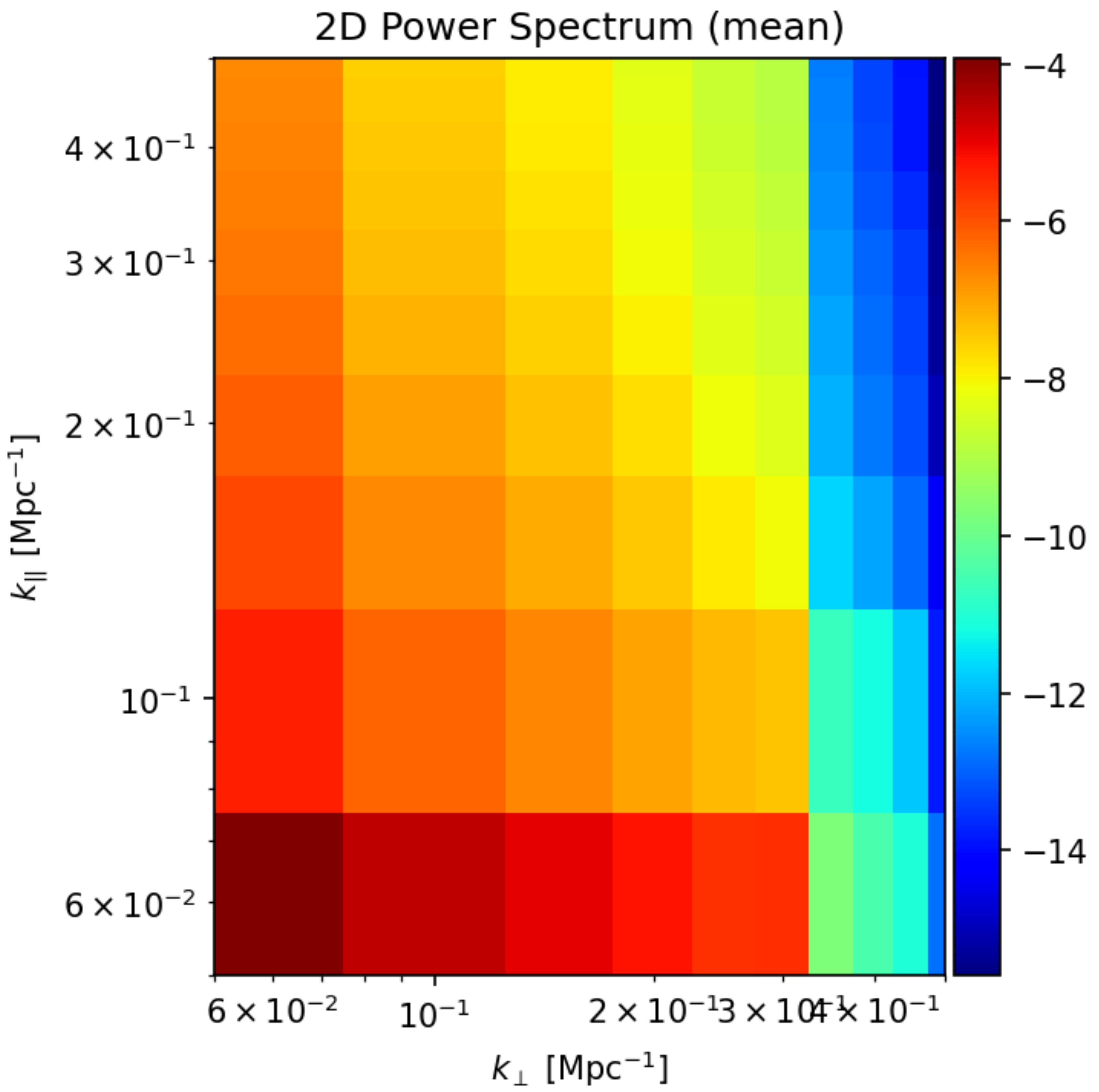
→ mode-mixing → producing non-smooth FG spectrum

- **set up a point source on the sky without frequency dependence**



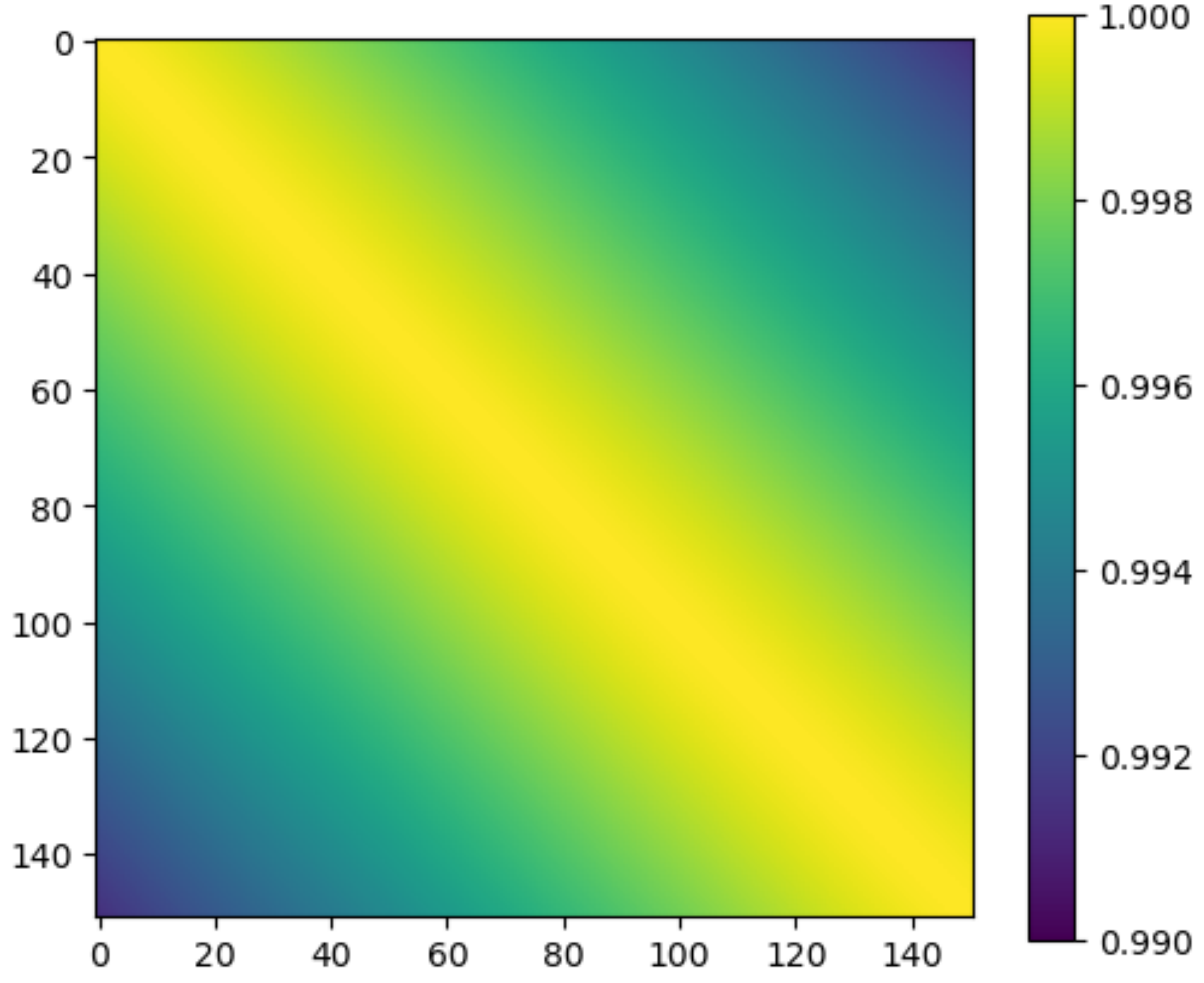
- **following interferometric observation, the spectrum undergoes an oscillation**



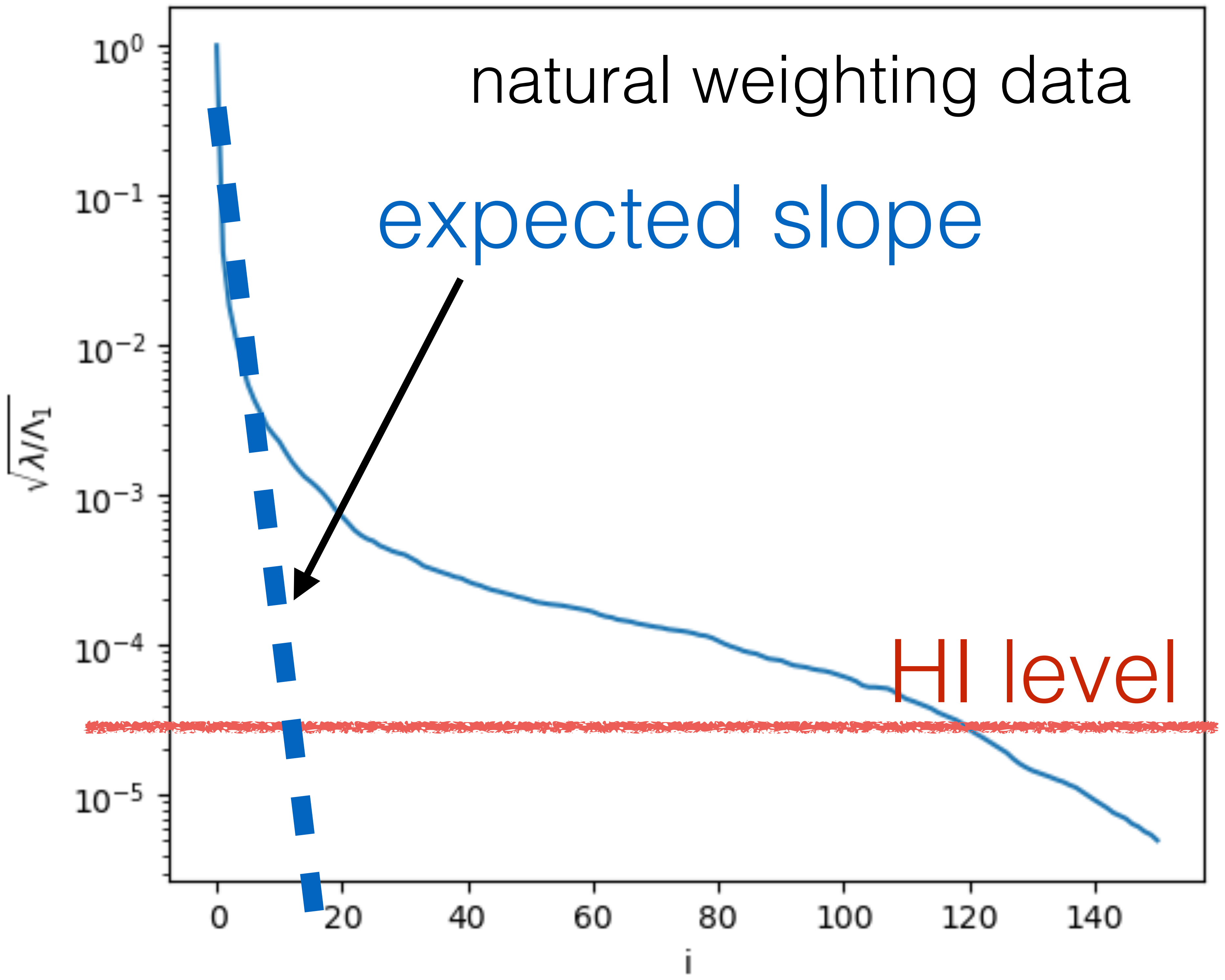


- **mode mixing: spatial modes mixed up with frequency modes**

Instrumental response (PSF) breaking down the smoothness:



$M_{\nu\nu'}$ correlation coefficients (150x150)



SVD — eigenvalues

Deduction of ~100 modes: impractical and physically meaningless

Mode mixing:

leakage of foreground power into cosmological modes due to instrument effects

→ contaminates high line-of-sight ($k_\parallel$) modes where the 21 cm signal lives

Why It Happens:

- **Interferometer chromaticity**

  Baseline length varies with frequency → beam changes with frequency

  Smooth-spectrum sources appear non-smooth after imaging

- **Calibration & imaging imperfections**

  Bandpass errors, gridding, pixelization, wide-field distortions

Mode Mixing $k_\perp \to k_\parallel$ from Frequency-Dependent PSF

1. Point Source Sky Model

We assume a **flat-spectrum point source** at angular position $I(\nu, \boldsymbol{\theta}) = \delta\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)$

2. Visibility from a Single Baseline b:

$$V(\nu) = \int d^2\boldsymbol{\theta} I(\nu, \boldsymbol{\theta}) e^{-2\pi i \mathbf{u}_\nu \cdot \boldsymbol{\theta}} = e^{-2\pi i \mathbf{u}_\nu \cdot \boldsymbol{\theta}_0}, \quad \textbf{with:} \quad \mathbf{u}_\nu = \frac{\nu \mathbf{b}}{c}$$

3. Phase Evolution with Frequency

- ν_0 **central frequency**
- $\nu = \nu_0 + \Delta\nu$

$$V(\nu) = e^{-2\pi i \frac{(\nu_0 + \Delta\nu)}{c} \mathbf{b} \cdot \boldsymbol{\theta}_0} = e^{-2\pi i \left(\frac{\nu_0}{c} \mathbf{b} \cdot \boldsymbol{\theta}_0\right)} \cdot e^{-2\pi i \left(\frac{\Delta\nu}{c} \mathbf{b} \cdot \boldsymbol{\theta}_0\right)}$$

•The 1st term is a constant; the 2nd term is a frequency-dependent phase.

4. Fourier Transform Over Frequency

Take 1D Fourier transform over frequency, over a bandwidth B

$$\tilde{V}(k_{\|}) = \int_{\nu_0-B/2}^{\nu_0+B/2} d\nu e^{-2\pi i\frac{(\nu-\nu_0)}{c}\mathbf{b}\cdot\boldsymbol{\theta}_0} e^{-2\pi i\nu k\|}$$

$$= e^{-2\pi i\nu_0 k\|}\int_{-B/2}^{+B/2} d\Delta\nu e^{-2\pi i\left(\frac{\Delta\nu}{c}\mathbf{b}\cdot\boldsymbol{\theta}_0 + \Delta\nu k\|\right)}$$

$$= e^{-2\pi i\nu_0 k\|}\cdot\int_{-B/2}^{+B/2} d\Delta\nu e^{-2\pi i\Delta\nu\left(\frac{\mathbf{b}\cdot\boldsymbol{\theta}_0}{c} + k\|\right)}$$

$$\Rightarrow \tilde{V}\left(k_{\|}\right) \propto \mathrm{sinc}\left[\pi B\left(\frac{\mathbf{b}\cdot\boldsymbol{\theta}_0}{c} + k_{\|}\right)\right]$$

The sinc is peaked at: $k_{\|} = -\dfrac{\mathbf{b}\cdot\boldsymbol{\theta}_0}{c}$

- **A clear leakage of angular position into LOS Fourier modes**
- **Even with a spectrally flat source, PSF variation introduces frequency structure**

Mode-mixing equation in real space:

Assume I(x) has no dependence on ν

$$I_d(\nu, \boldsymbol{x}) = \int d^2\boldsymbol{x}' \, \mathrm{PSF}\,(\nu, \boldsymbol{x} - \boldsymbol{x}')\, I\,(\boldsymbol{x}')$$

perform a Fourier transform along ν

$$\tilde{I}_d(x, \tau) = \int d\nu I_d(x, \nu) e^{-2\pi i \nu \tau} = \int d\nu \left[\int dx' \, \mathrm{PSF}\,(\nu, x - x') \cdot I\,(x') \right] e^{-2\pi i \nu \tau}$$

Swap the order of integration:

$$\tilde{I}_d(x, \tau) = \int dx' I\,(x') \cdot \left[\int d\nu \, \mathrm{PSF}\,(\nu, x - x') \cdot e^{-2\pi i \nu \tau} \right]$$

Mode-mixing equation in real space:

Define the delay-transformed PSF:

$$\widetilde{\mathrm{PSF}}(\tau, x) = \int d\nu \, \mathrm{PSF}(\nu, x) \cdot e^{-2\pi i \nu \tau}$$

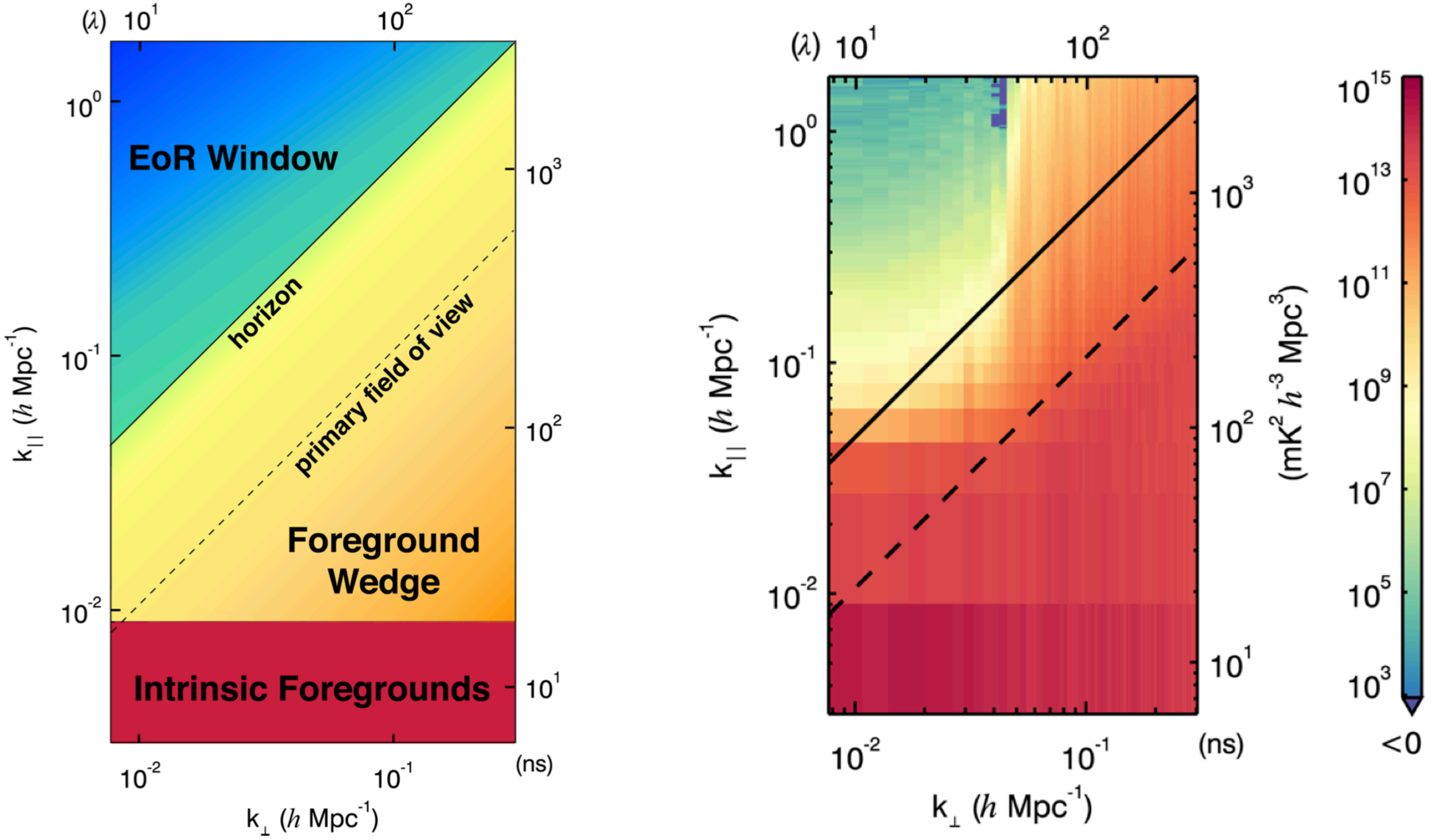
Then $\tilde{I}_d(x, \tau) = \int dx' \widetilde{\mathrm{PSF}}\,(\tau, x - x') \cdot I\,(x')$

- **the delay-transformed dirty image is a convolution of the sky with a delay-transformed PSF.**

EOR window

Use the mapping from observing parameters to comoving k:

$$k_{\|} = \frac{2\pi}{r_\nu}\eta \; ; \; k_{\perp} = \frac{2\pi}{D(z)}|\mathbf{u}|$$



MNRAS **461,** 3135–3144 (2016)

Summary

21 cm Cosmology

- Map the Cosmic Dawn & Reionization
- Trace neutral hydrogen over time
- Constrain dark matter, inflation, and structure formation

Foreground Challenges

- Foregrounds $\gg$ Signal:
- Spectrally smooth, but mode mixing leaks into signal modes
- Creates the EoR Wedge/Window

Thanks!

Foreground removal

- Avoidance, subtraction, and precise calibration

Questions:

1. How would adding components like free-free emission or spectral curvature challenge the assumptions made by PCA, ICA, or polynomial fitting?

2. If you simulate a data cube with point sources, synchrotron, and free-free emission, how would you evaluate which foreground removal method (PCA, ICA, GPR, KL) performs best?

3. How can you design a test to quantify signal loss or foreground leakage when applying these methods to complex foreground models?