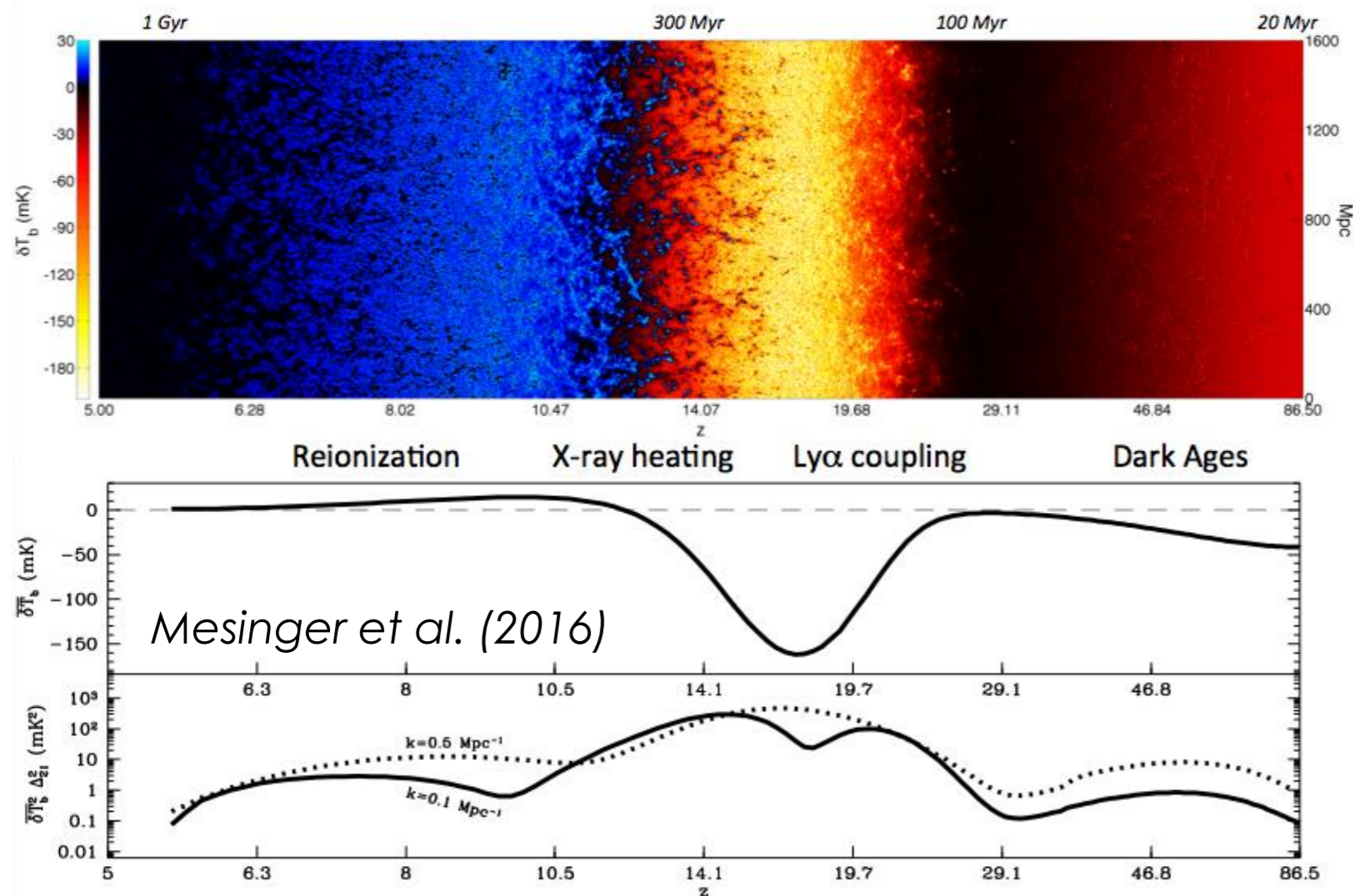
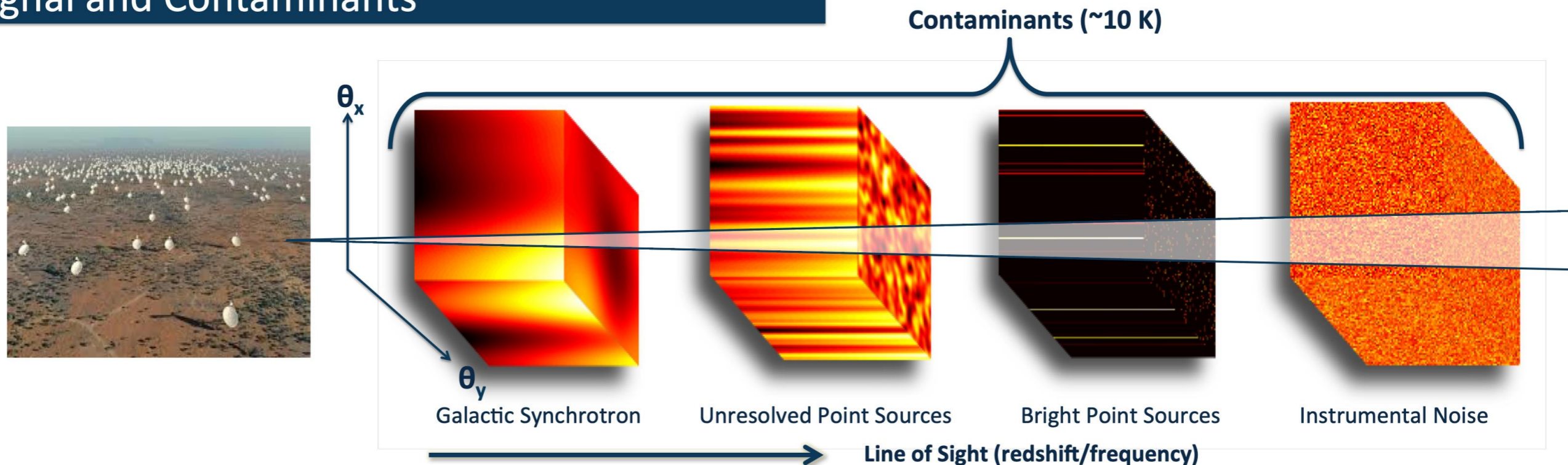


21 cm signal recovery in SKA Data Challenge 3

Le Zhang
Sun Yat-Sen University

21 cm Cosmology Workshop 2024 &
Tianlai Collaboration Meeting
@Hangzhou 2024.07.23

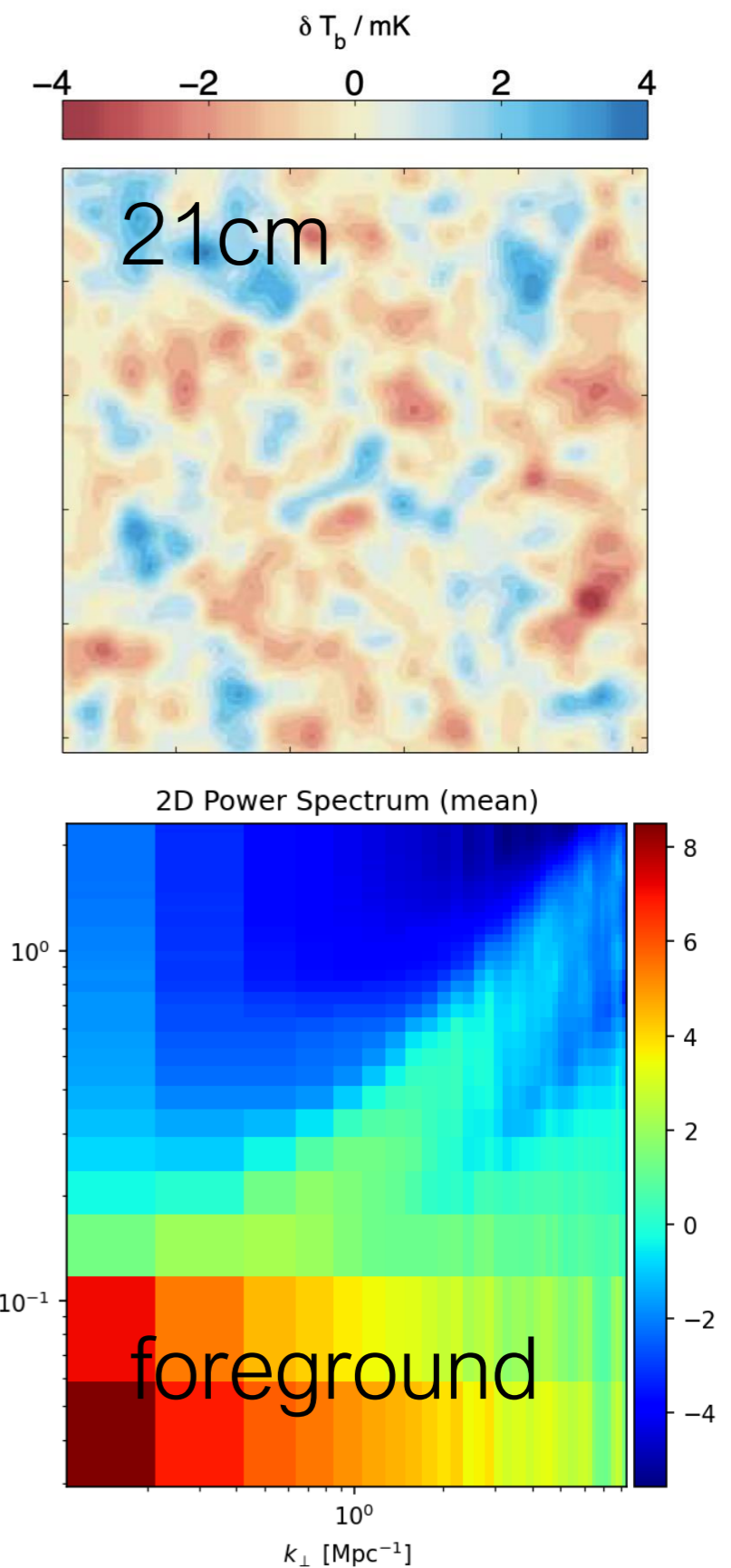
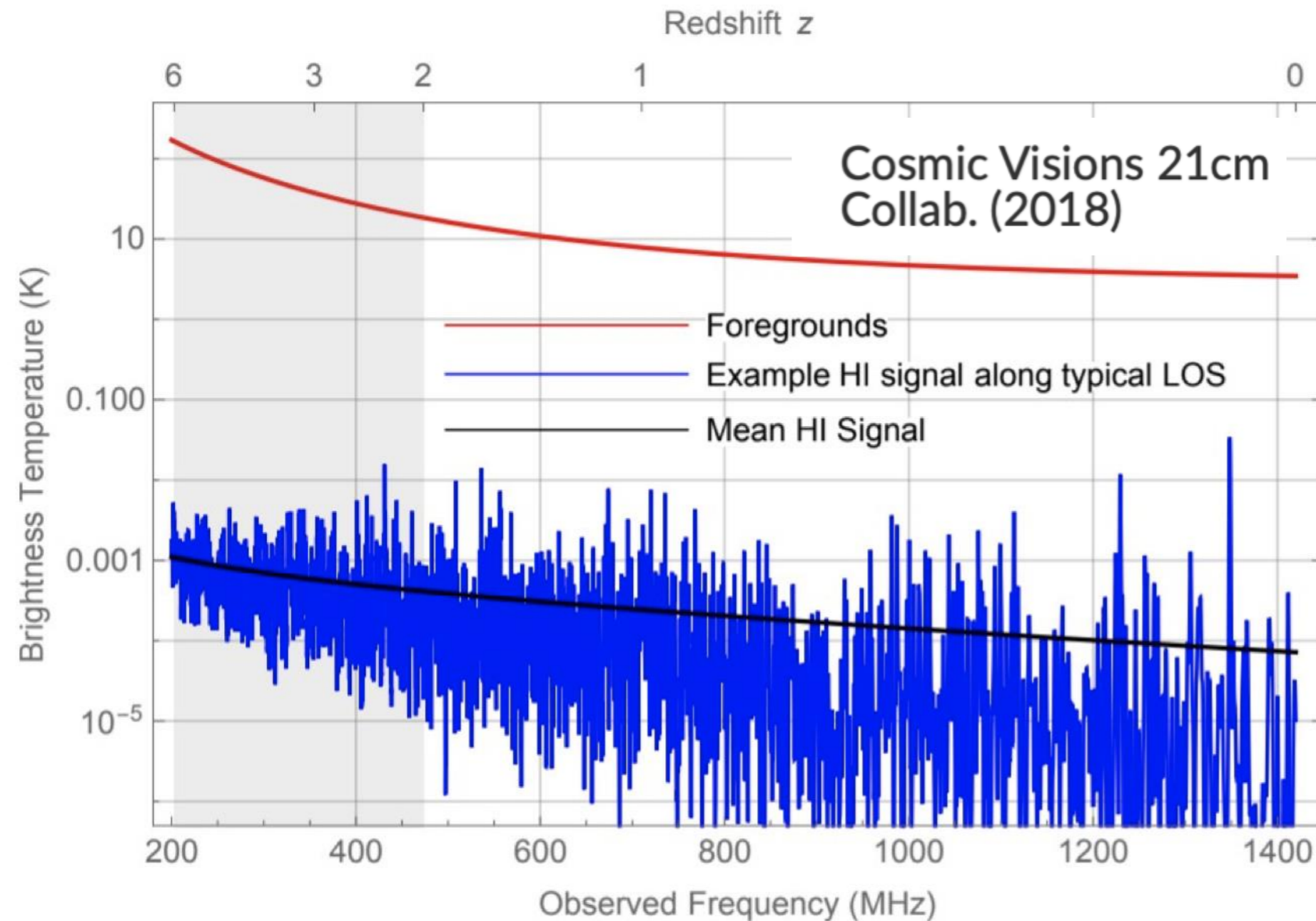
Signal and Contaminants



Strong foreground:
 $T_{\text{foreground}} \gg 10\text{K}$

Faint signal:
 $T_{21\text{cm}} \sim 10\text{mK}$

The key to separating out foregrounds: their spectral smoothness



For SKAO Science Director Dr Robert Braun, the teams' results provide a solid foundation for analysing real SKA observations.

"I'm thrilled to see these results; I had some concerns when we issued this challenge that it might prove impossible to recover these faint simulated signals," Dr Braun said. "I am now very hopeful that the collaborative refinement of signal recovery techniques within our global community will allow the actual EoR signal to be detected with confidence."

Science Data Challenge 3a – Datasets

- General
 - Four-hour duration tracking observation
 - Thermal Noise Equivalent: 1000 hours
 - Field of View: One SKA1-Low pointing at RA, Dec = 0h, -30 deg
- Visibilities/[images](#)
 - Size: 7.5 TB (Visibilities)
 - Integration Time: 10 seconds
 - Channel Width: 100 kHz
 - Frequency Coverage: 106-196 MHz (6 cubes × 151 channels)
 - [Image Cube \(60 GB\): 2048×2048, 16 arcsec pixels, natural/uniform weighting](#)
- The Challenge
 - Determine EoR power spectrum as function of scale and frequency
 - Cleaning foreground (10^5 times brighter)
 - Presence of various residual calibration errors: DI, DD & bandpass

SKAO Science Data Challenge 3

WORLDWIDE PARTICIPATION



Participants



Computing facilities



THE CHALLENGE IN NUMBERS

Teams analysing

7.5 TB

of simulated telescope data and a corresponding

60 GB

of image cubes representing different radio frequencies

234

registered participants in

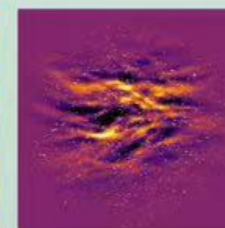
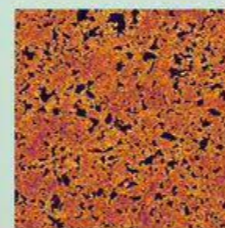
16

countries

12

supercomputing centres providing resources globally

33 teams



Teams are analysing data which simulates observations of the Epoch of Reionisation signal (left; bright areas are neutral hydrogen, and dark patches are ionised gas). It is obscured by foreground emission (right; orange dots are galaxies, and the ribbon-like shape is diffuse gas in our galaxy). While the features of each image appear equally bright here, in the data cube the background is millions of times fainter than the foreground.

SDC3a teams and FG-cleaning approaches

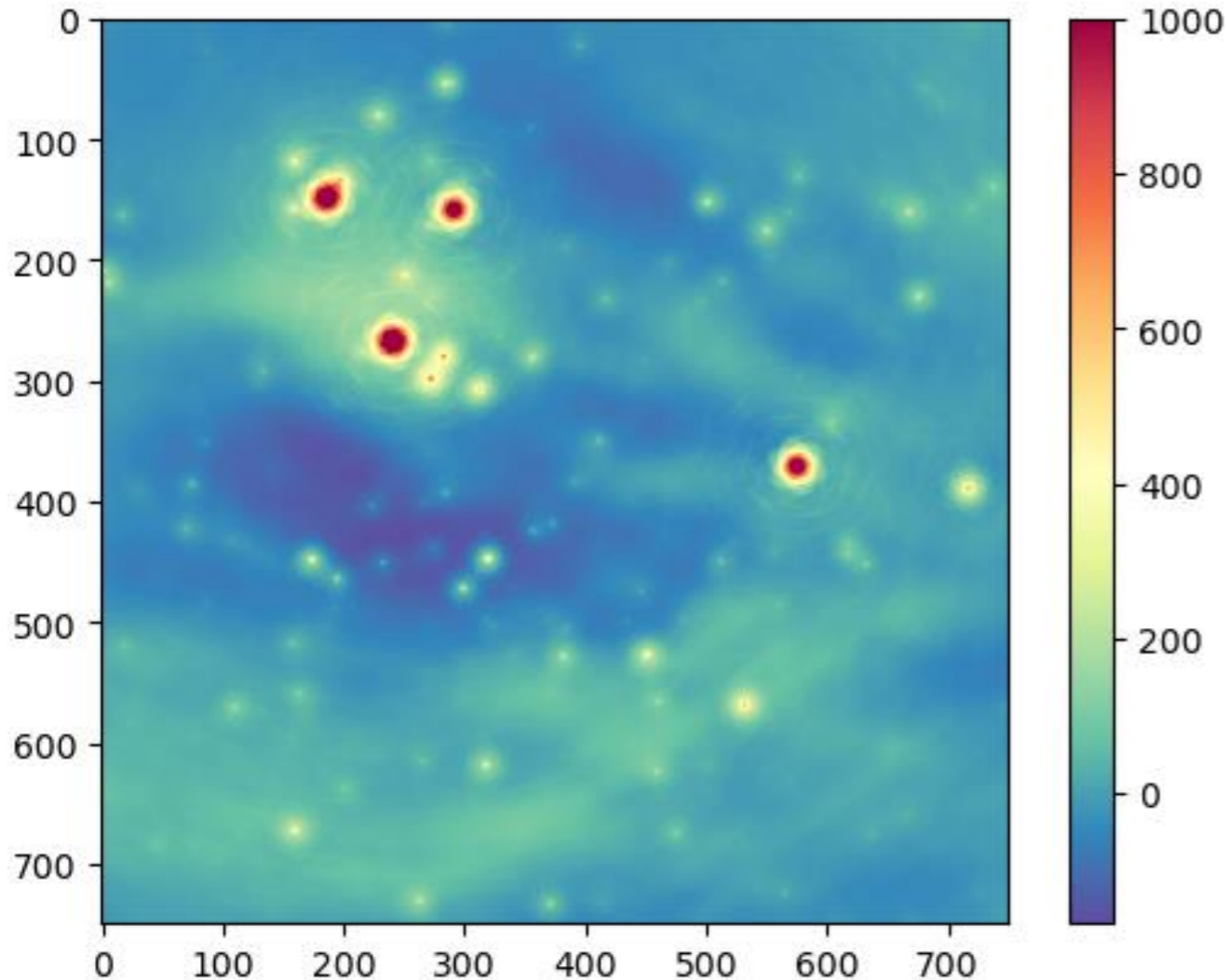
10 top teams:

- **HIMALAYA** (SYSU): **reconvolution + transfer function**
 - **DOTSS-21 (ML-GPR; Advanced_ML-GPR; Avoidance)**: Machine-learning+Gaussian Process Regression (GPR) to model the foregrounds to separate them from the 21-cm signal
 - **ERWA**: neural network applied in image; China SRC supported
 - **Shuimu-Tianlai** (Tsinghua-NAOC): oriented singular value decomposition (O-SVD)
 - **Wizards of Oz 3D**: improving the sky model to improve the quality of the sky-model-subtracted visibilities; 4th-order polynomial fitting; Pawsey Supercomputer supported
 - **Akashganga**: 2nd-order polynomial fitting+GPR
 - **REACTOR**: PI-AstroDeconv +PCA
 - **SKACH**: U-shaped 3D convolutional neural network to remove residual foreground; polynomial fitting in uv space
-
- **KUSANAGI**: PCA+neural network; **Cantabrigians**: Bayesian GPR model; **Hausos**: CNN-based approach; **Nottingham-Imperial**: point-source model+ FastICA; **Pisano Galaxy Moppers**: foreground avoidance; **Foregrounds-FRIENDS**: PCA+neural network; **HAMSTER**: delay spectrum approach; **KORSDC**: PSF deconvolution+ICA
 - **SROT**: GPR ...

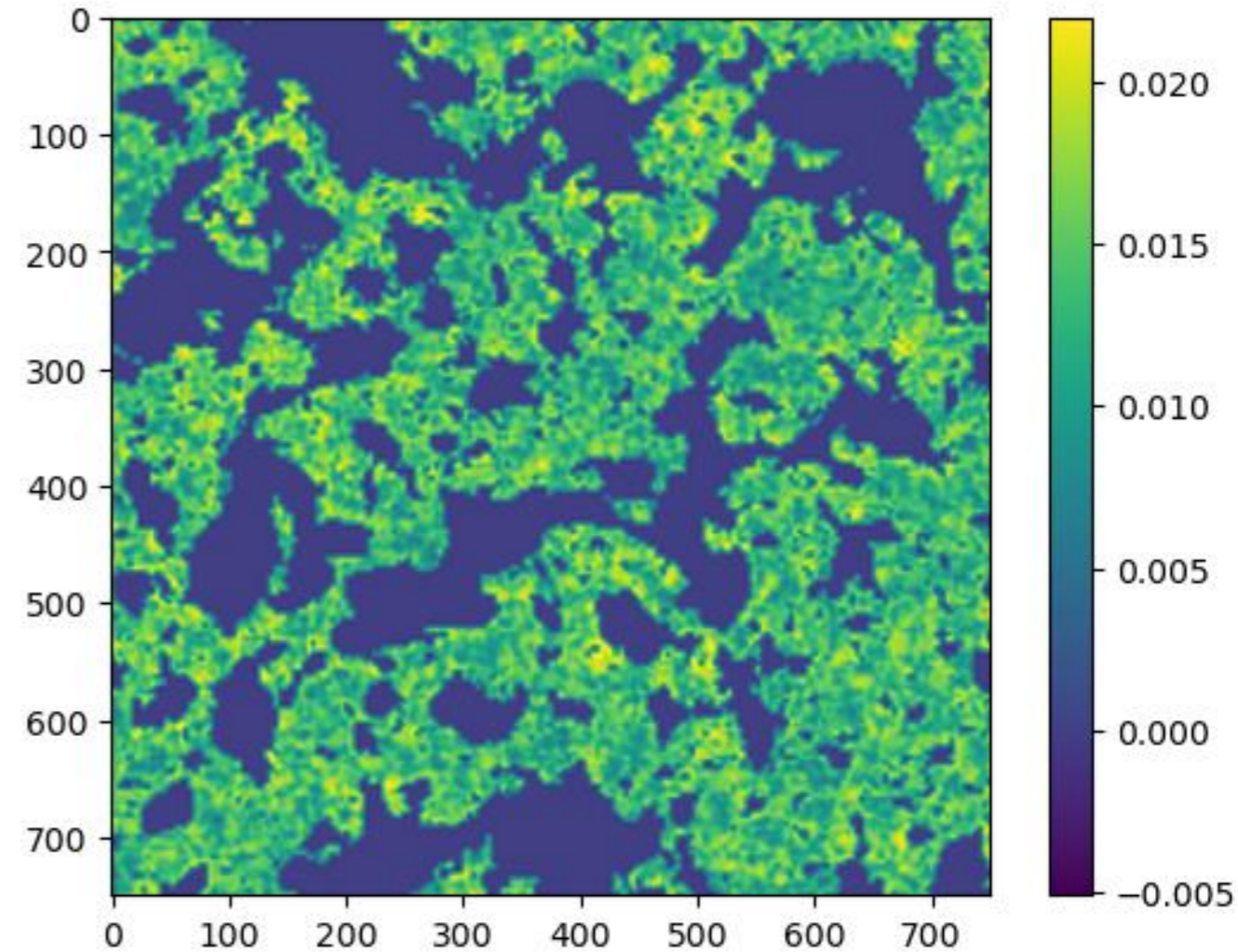
Challenge: FG contamination

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

Raw SDC3a image
(natural weighting)



True EOR

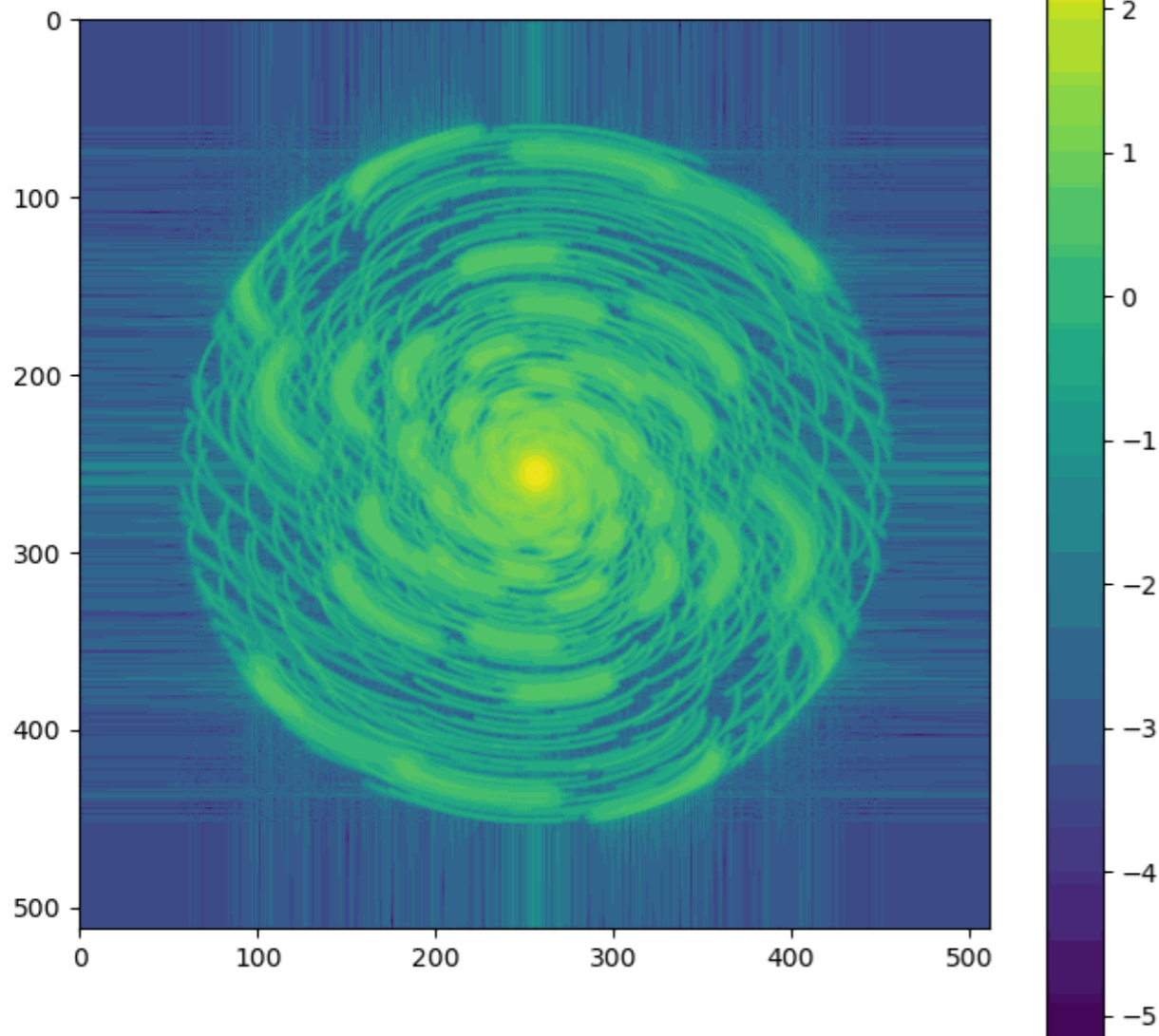


Key Challenge: frequency-dependent PSF

→ **uv coverage changed with frequency**

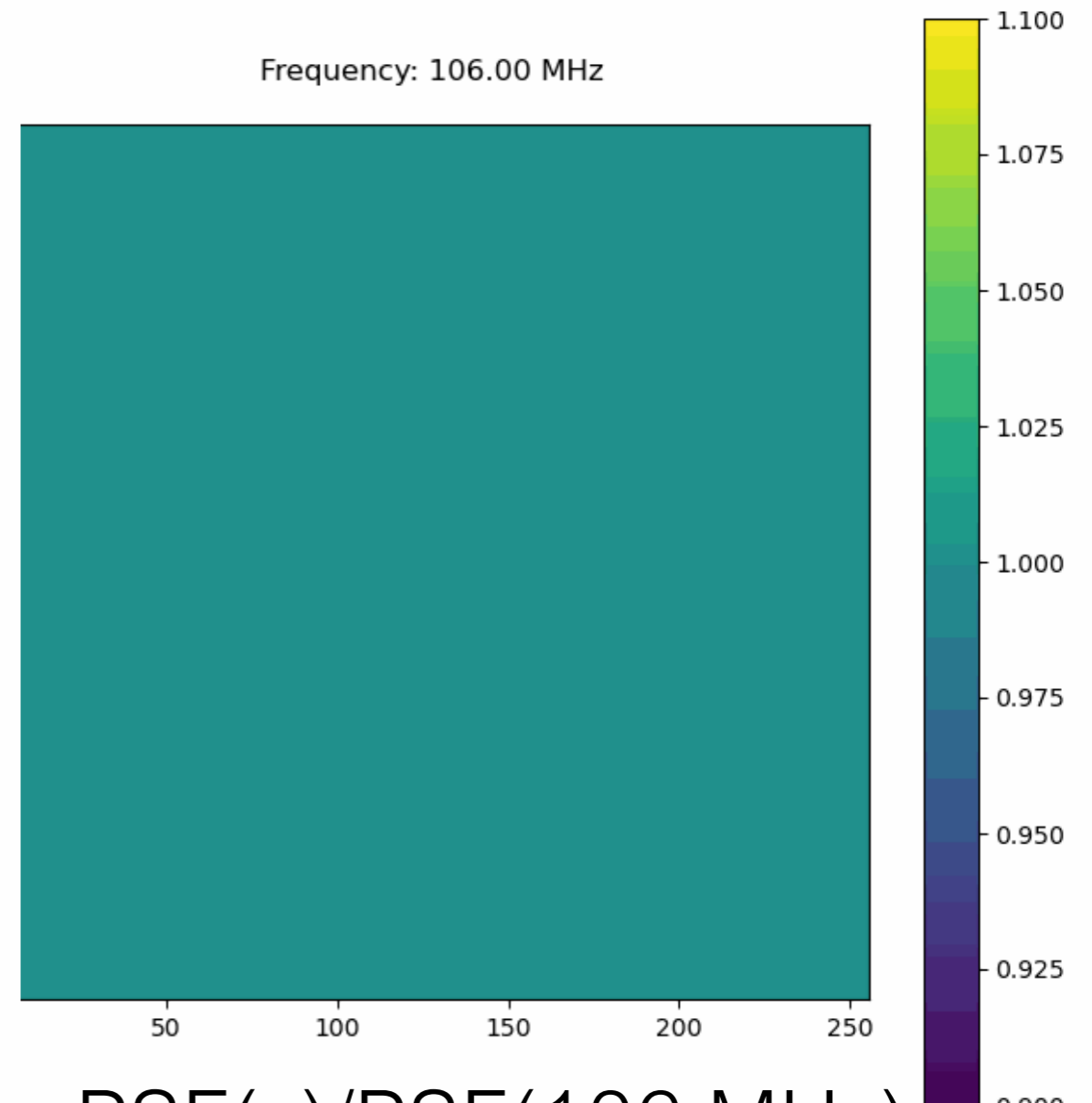
→ **notable change for PSF (esp. its sidelobe) with frequency**

Frequency: 106.00 MHz



uv coverage

Frequency: 106.00 MHz

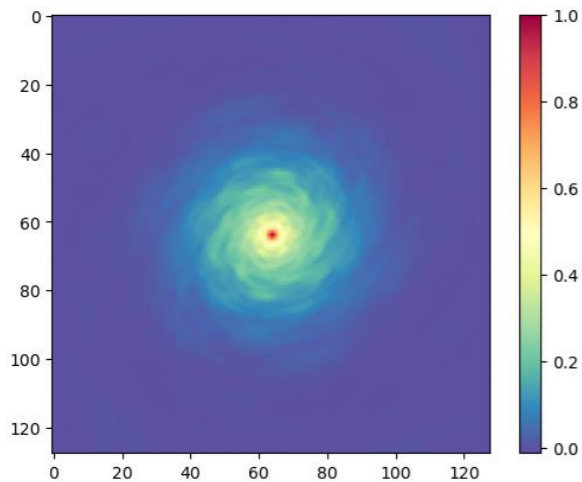


PSF(ν)/PSF(106 MHz)

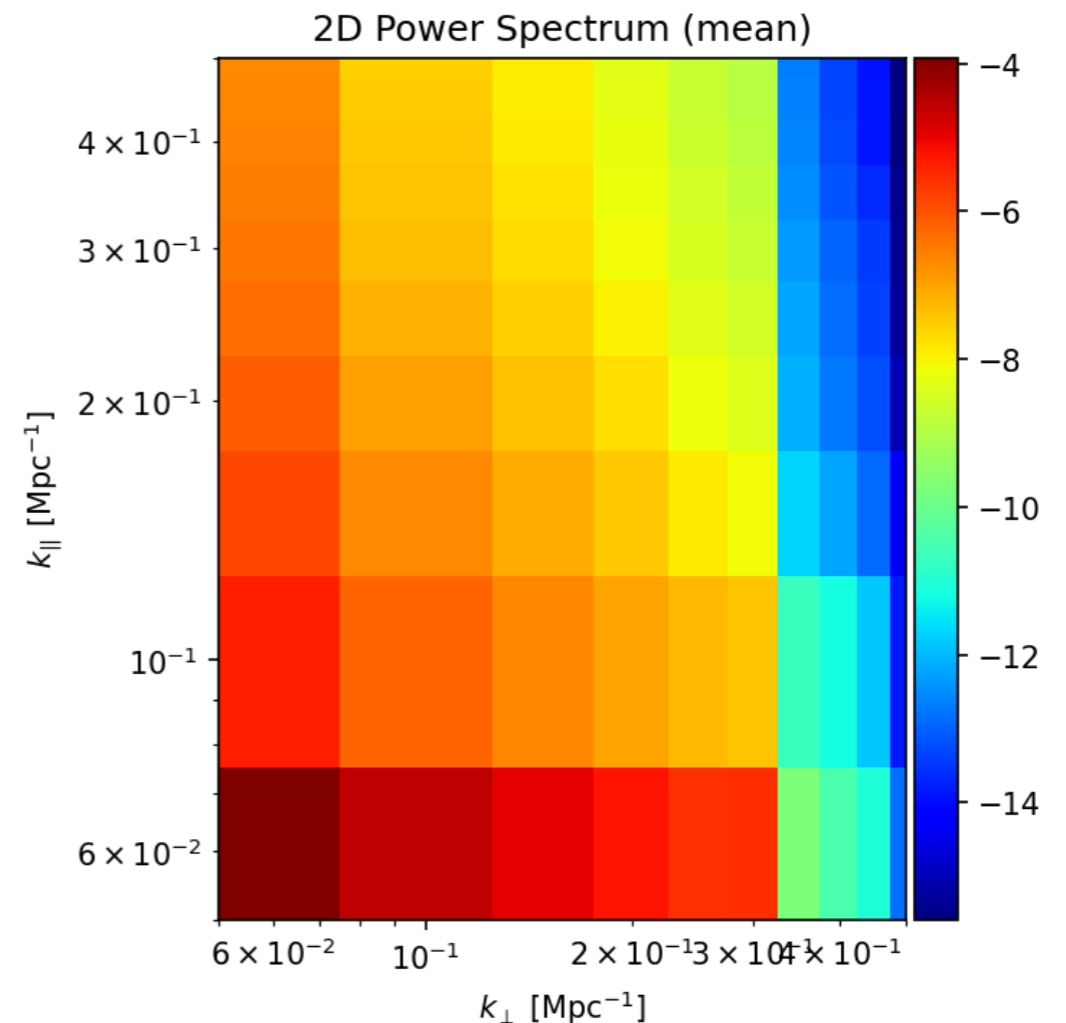
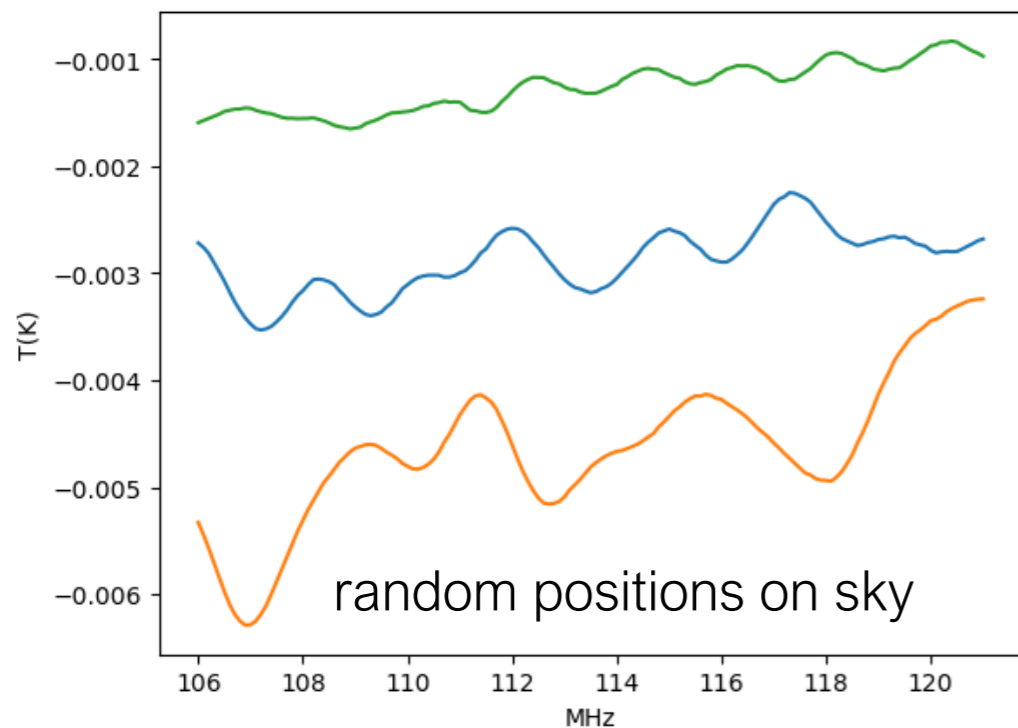
Key Challenge: frequency-dependent PSF

→ **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

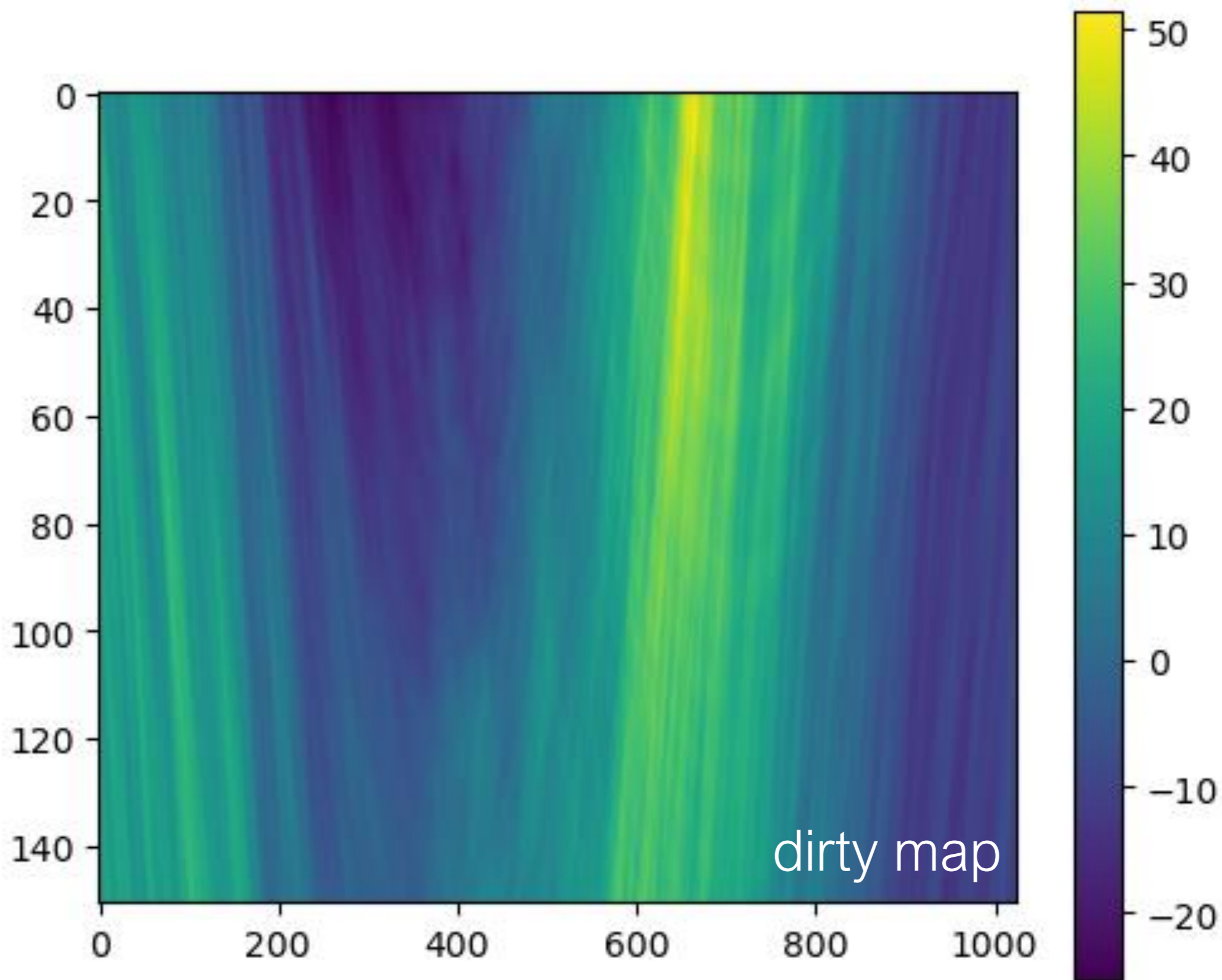


- mode mixing: spatial modes mixed up with frequency modes

Mode-mixing:

PSF varied with frequency, breaking down the smoothness

frequency
(LoS)



Challenges:

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

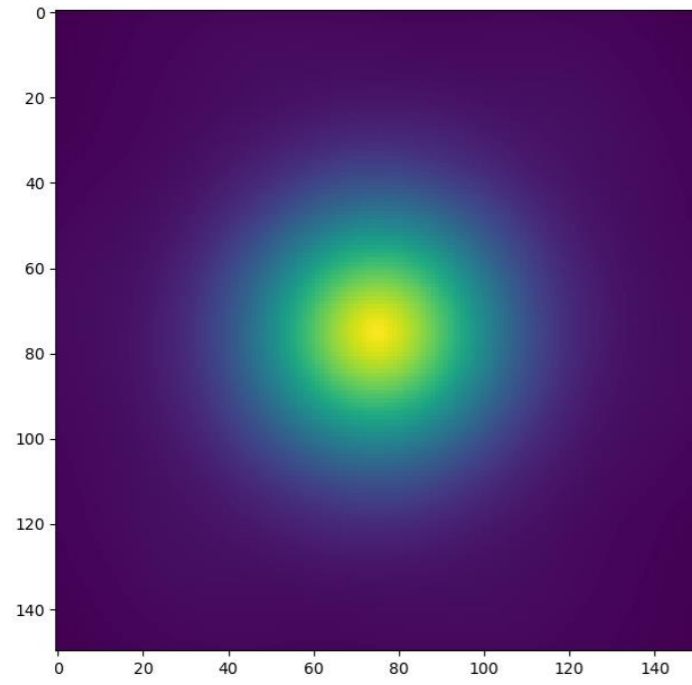
Solution: other way around?

Counterintuitive approach:

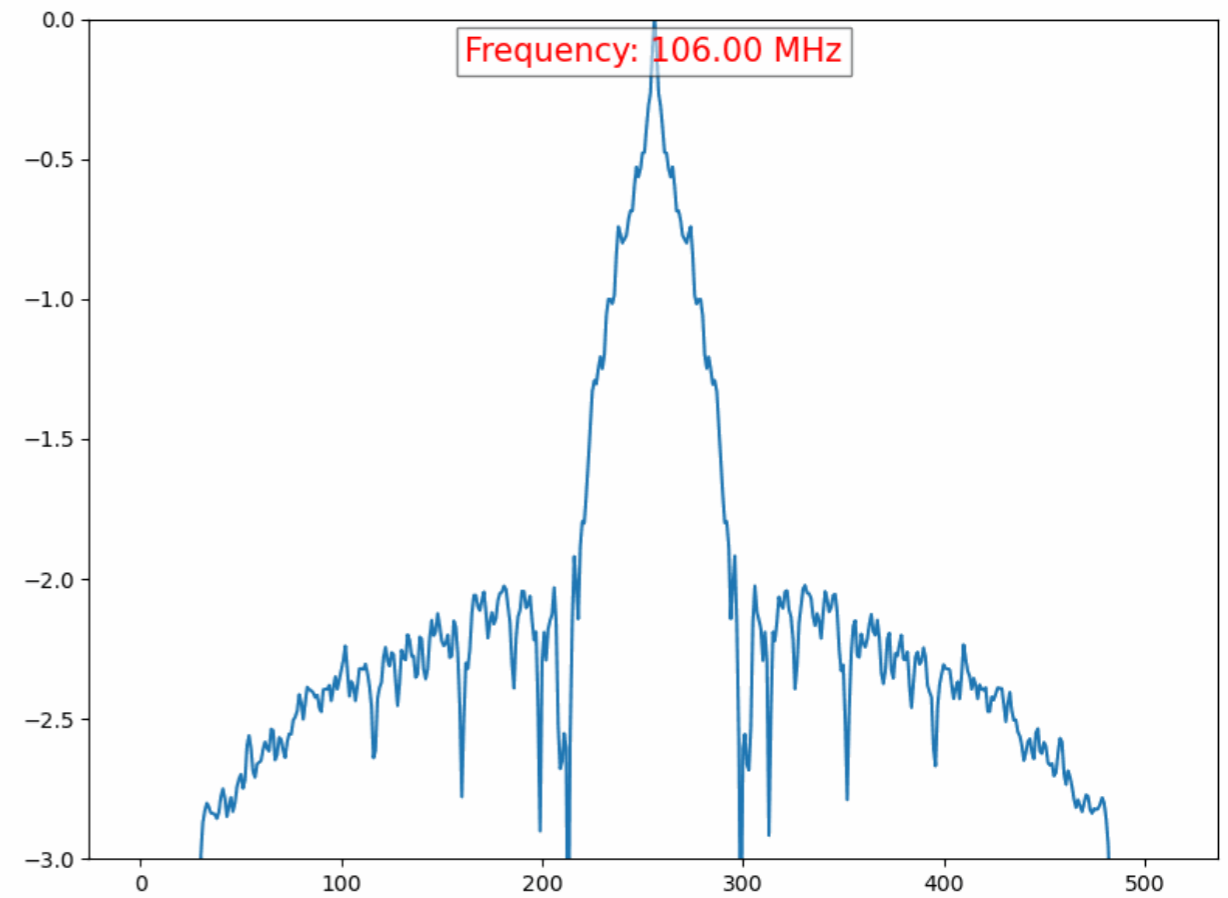
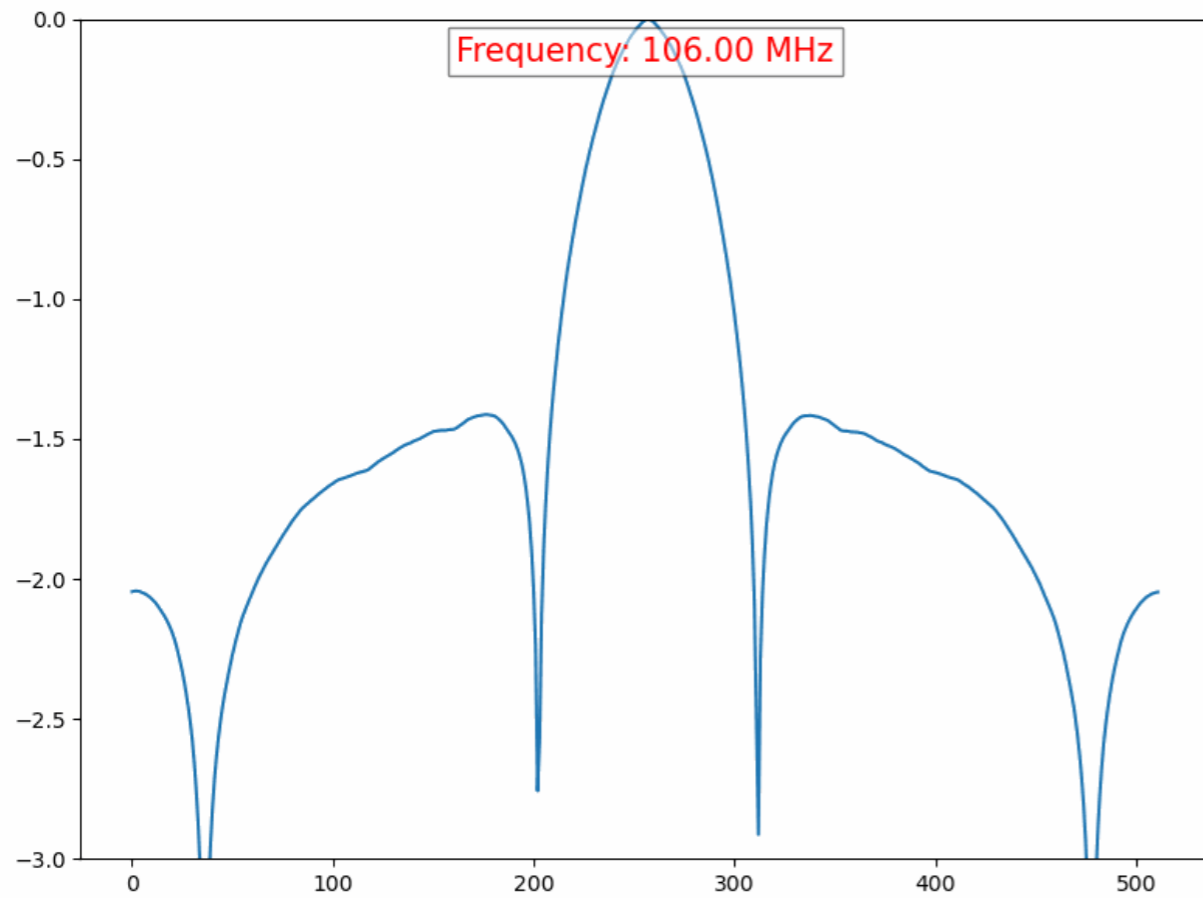
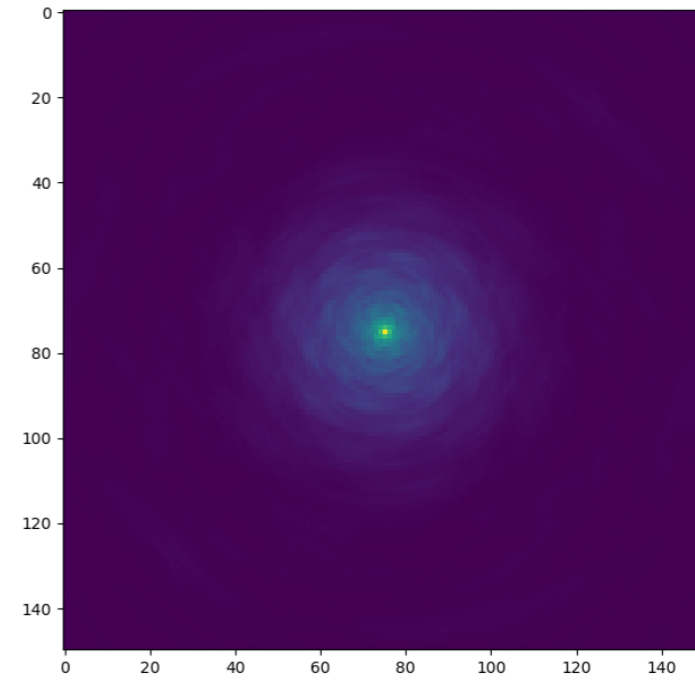
✓ reconvolution rather than deconvolution

Reduction of mode mixing: suppressing high- k_{\perp} modes that vary significantly with frequency, which dominate the effect of mode mixing.

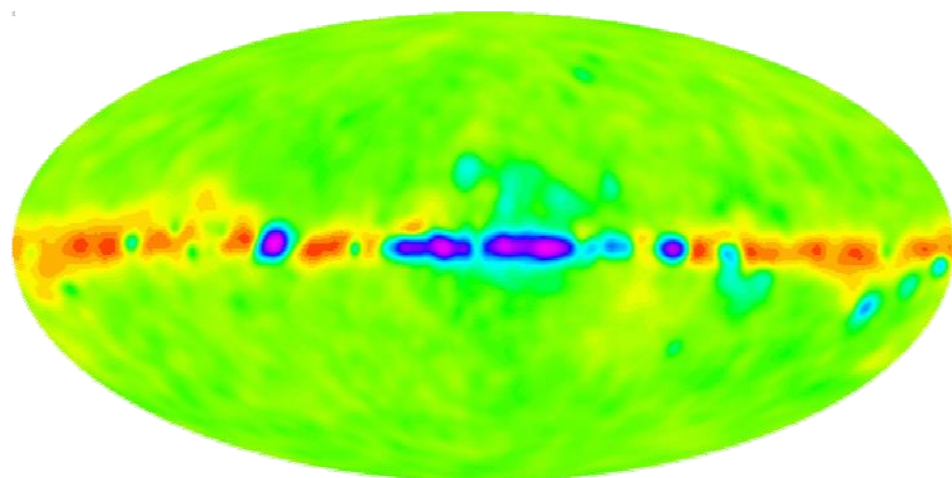
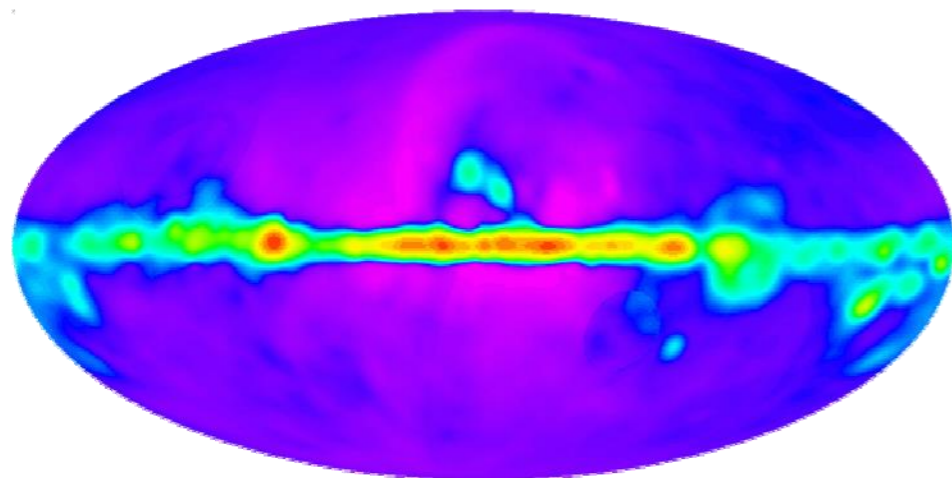
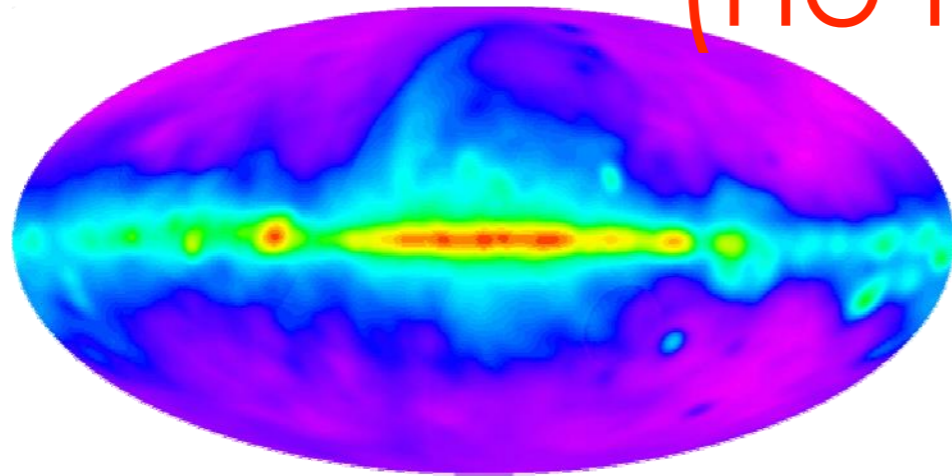
PSF ⊗ PSF



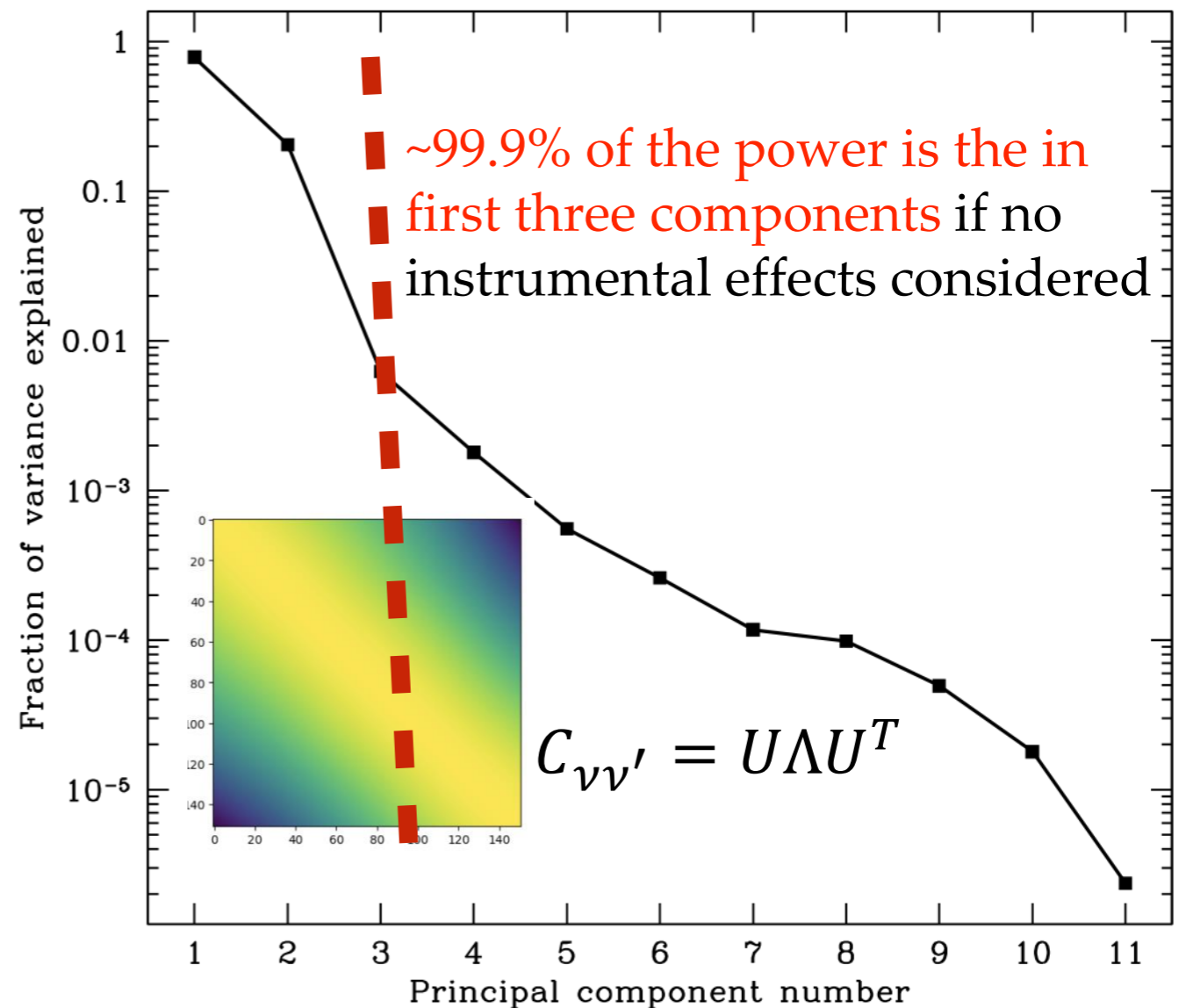
PSF



PCA/SVD method (no instrumental effects)

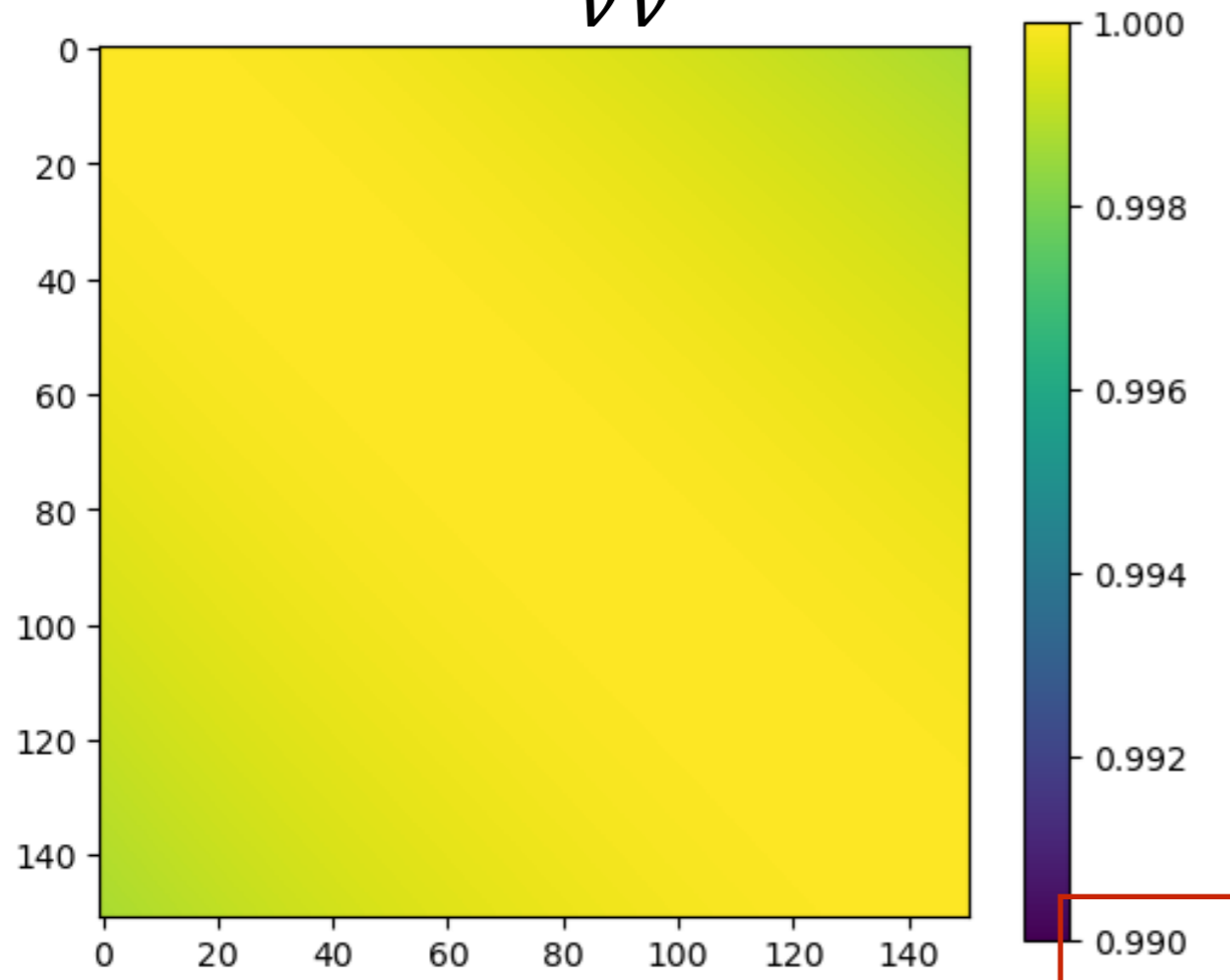


The first three principal components, which can be crudely interpreted as maps of total “stuff”, synchrotron fraction and thermal dust fraction.

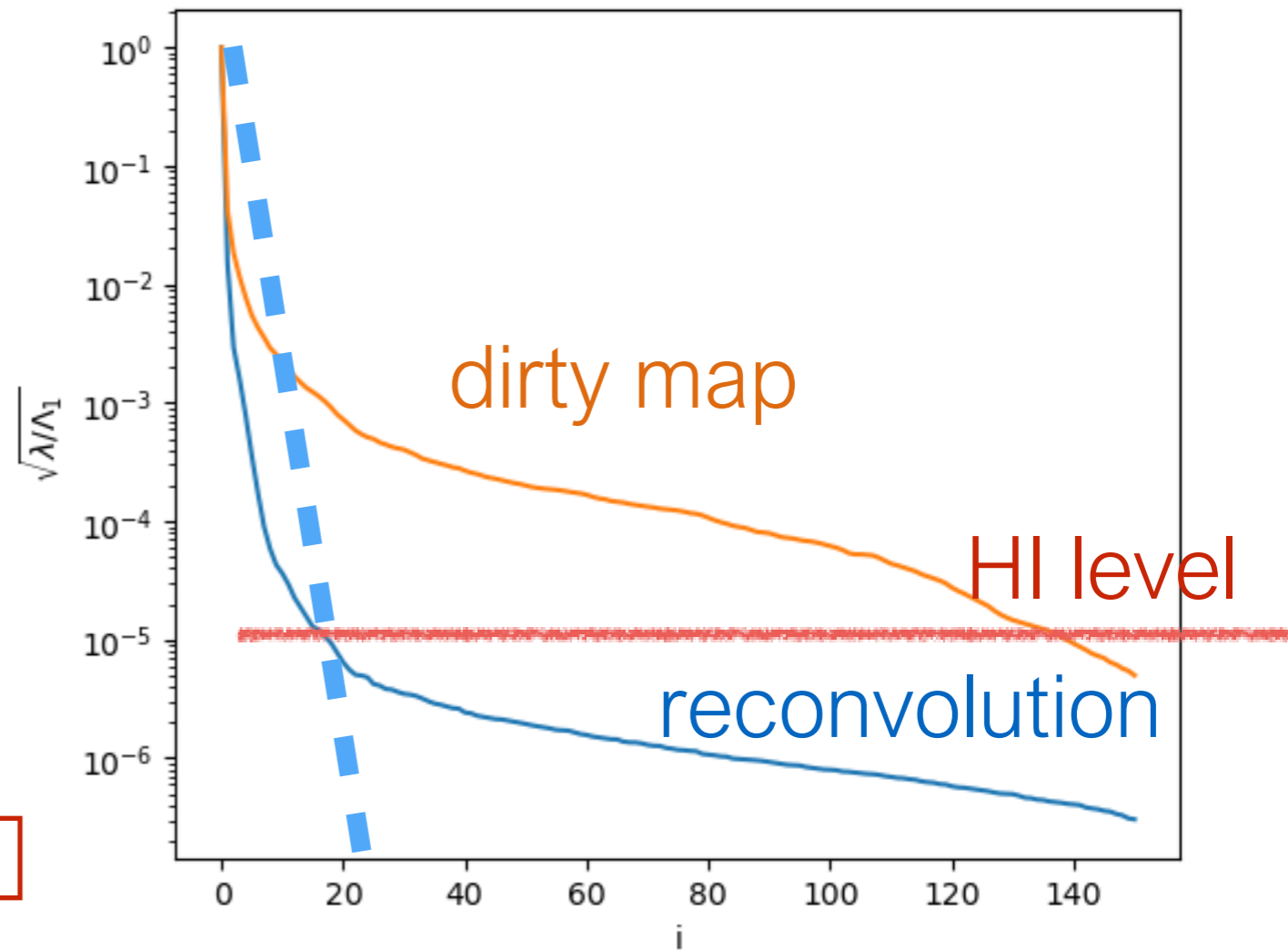


Smoothness recovered by “reconvolution”

$\text{COV}_{\nu\nu'}$

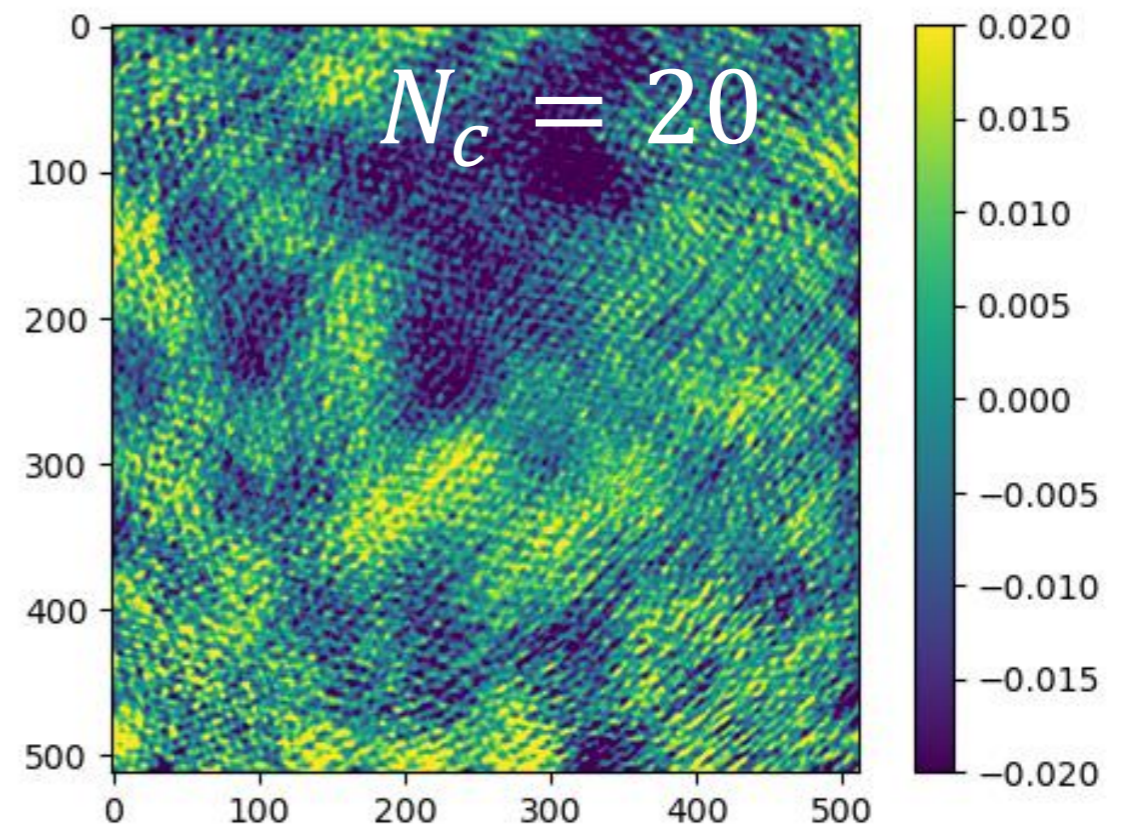
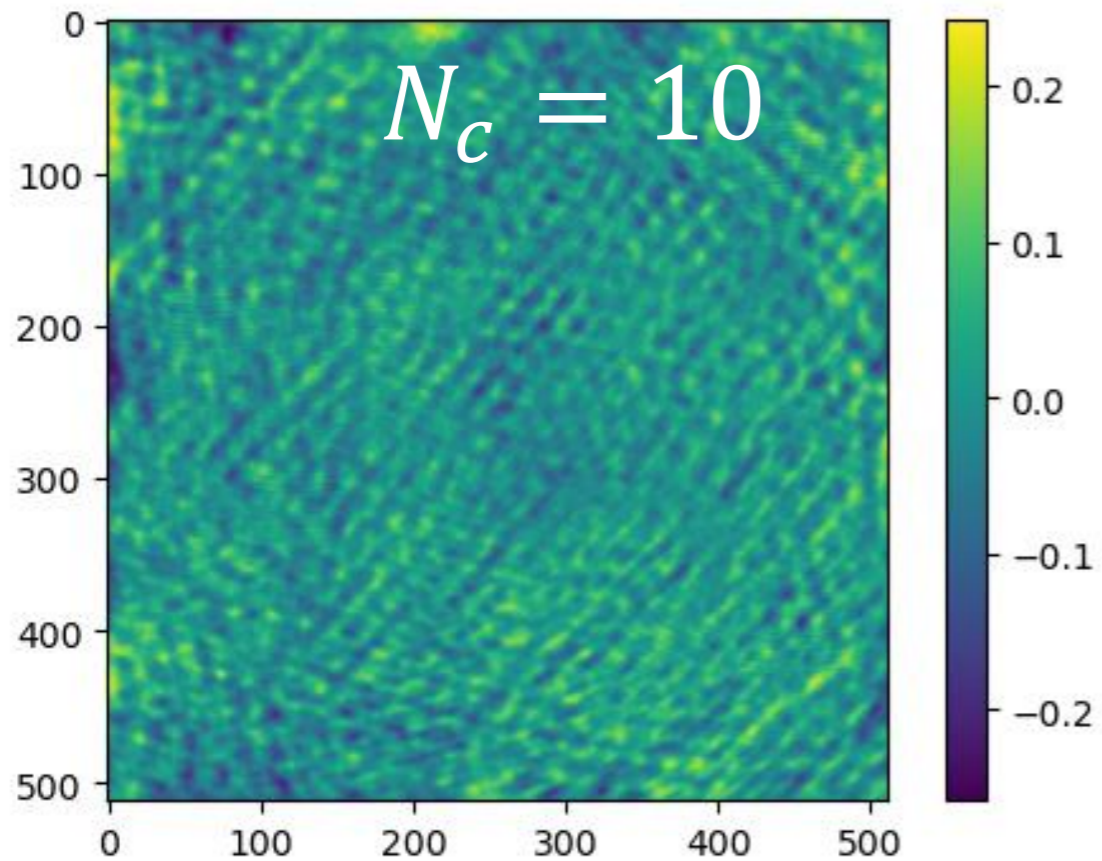
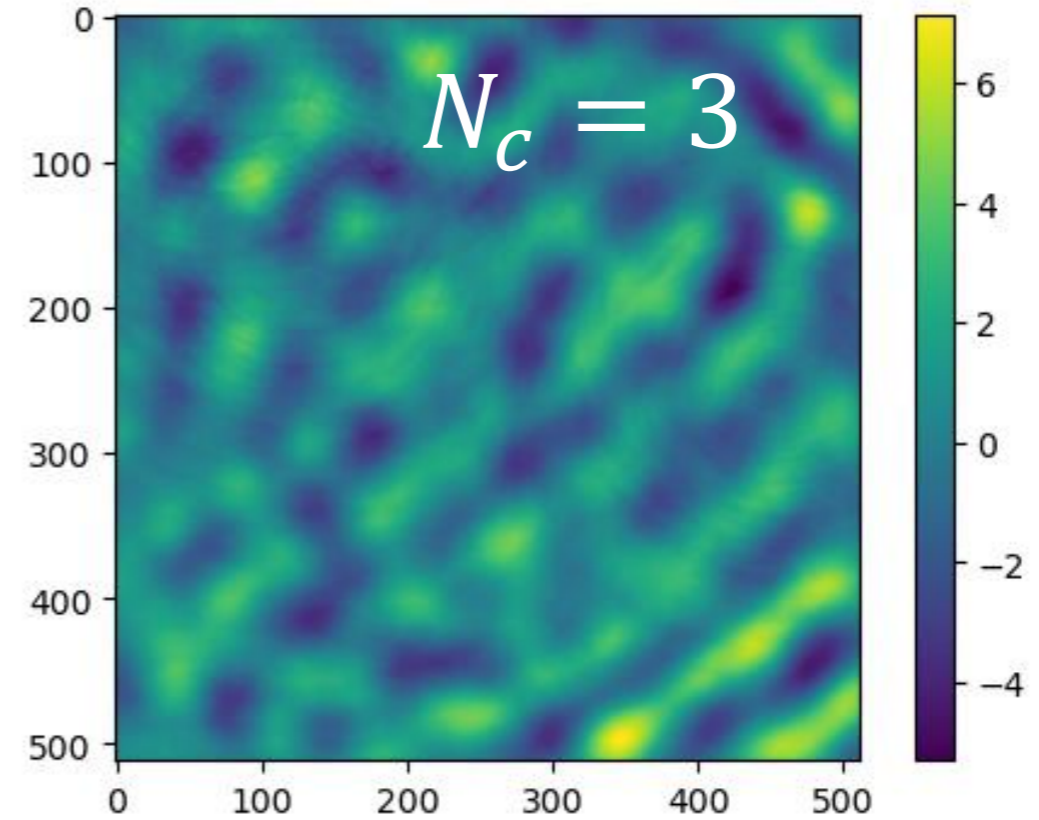
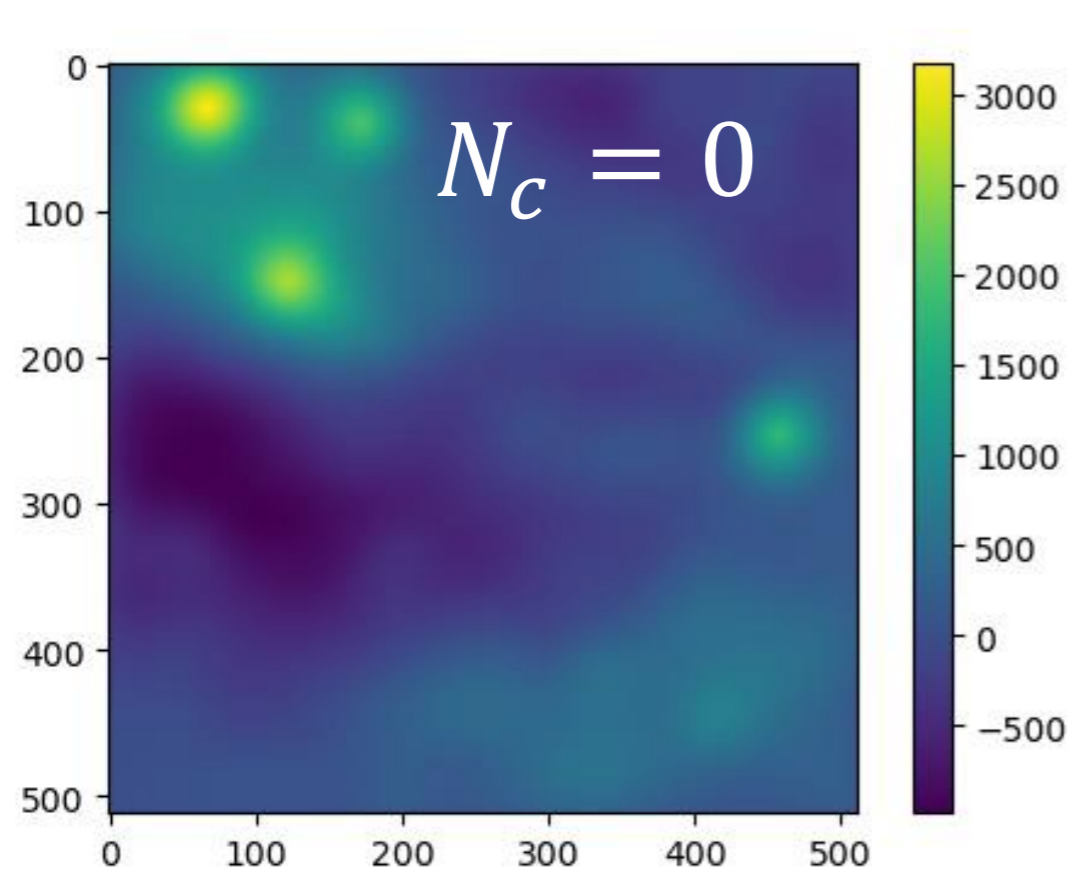


eigenvalues



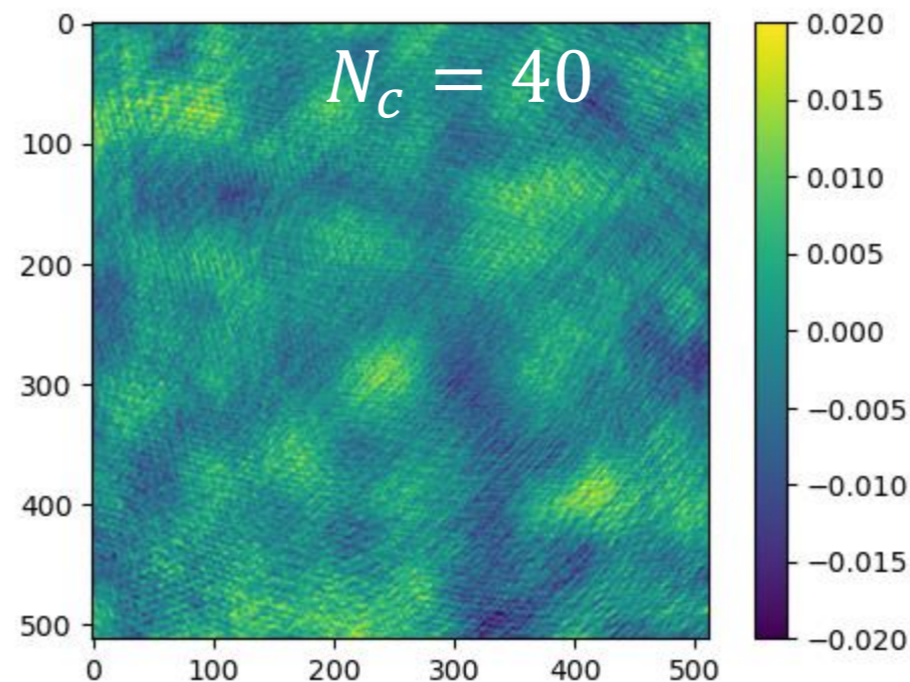
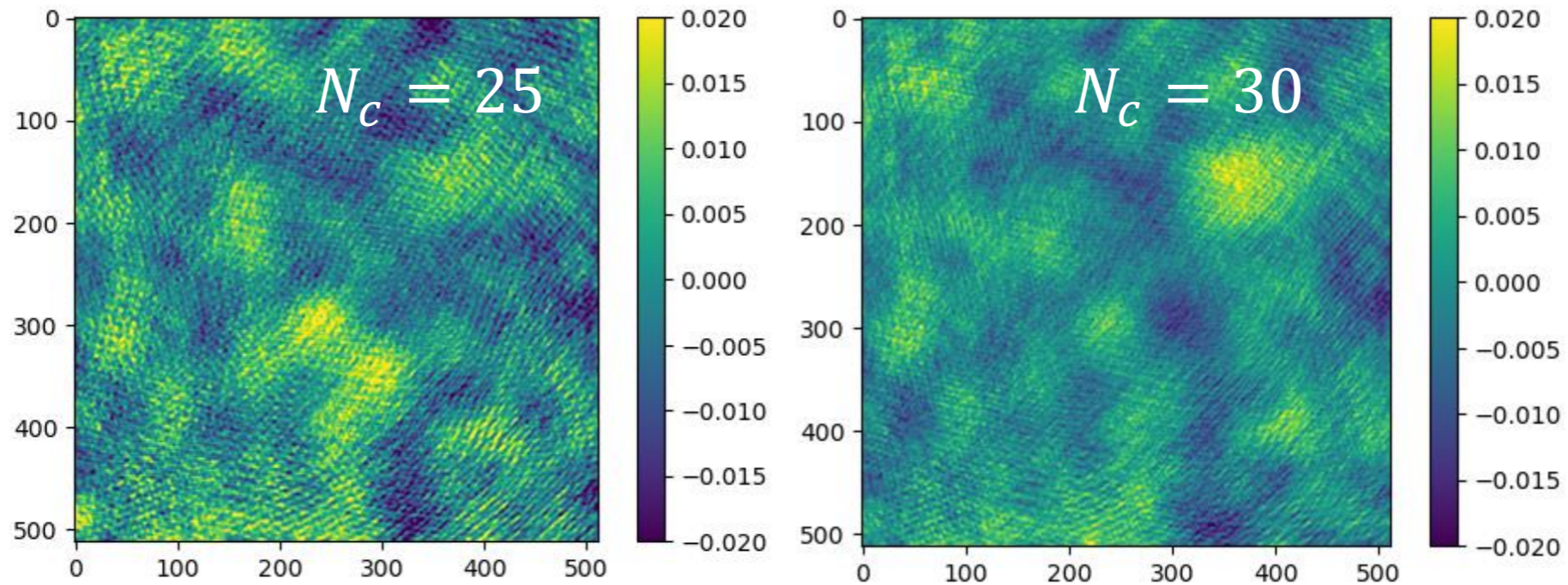
- PCA project out the first ~ 20 modes - remove foreground effectively!

Residuals after PCA projection



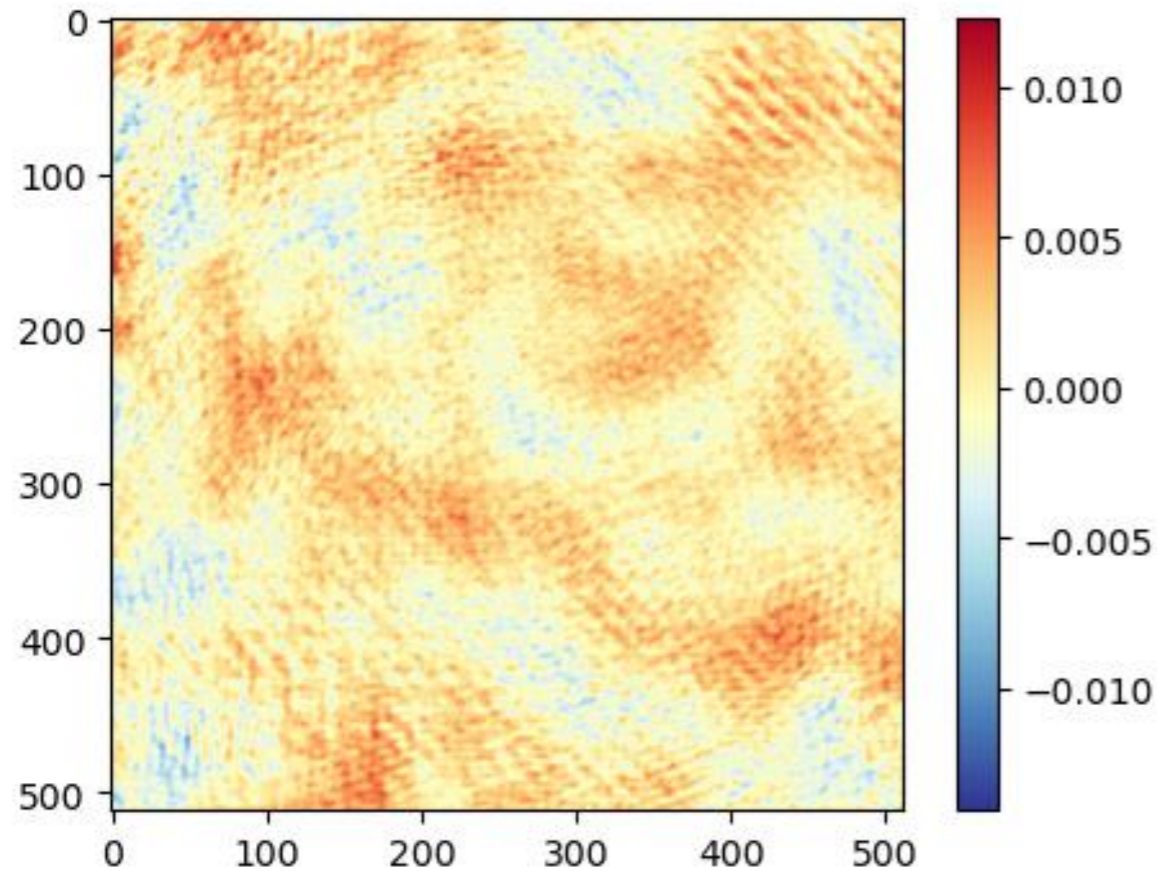
Convergence test:

- further modes projected out
- yet the residual patterns remained largely unaltered

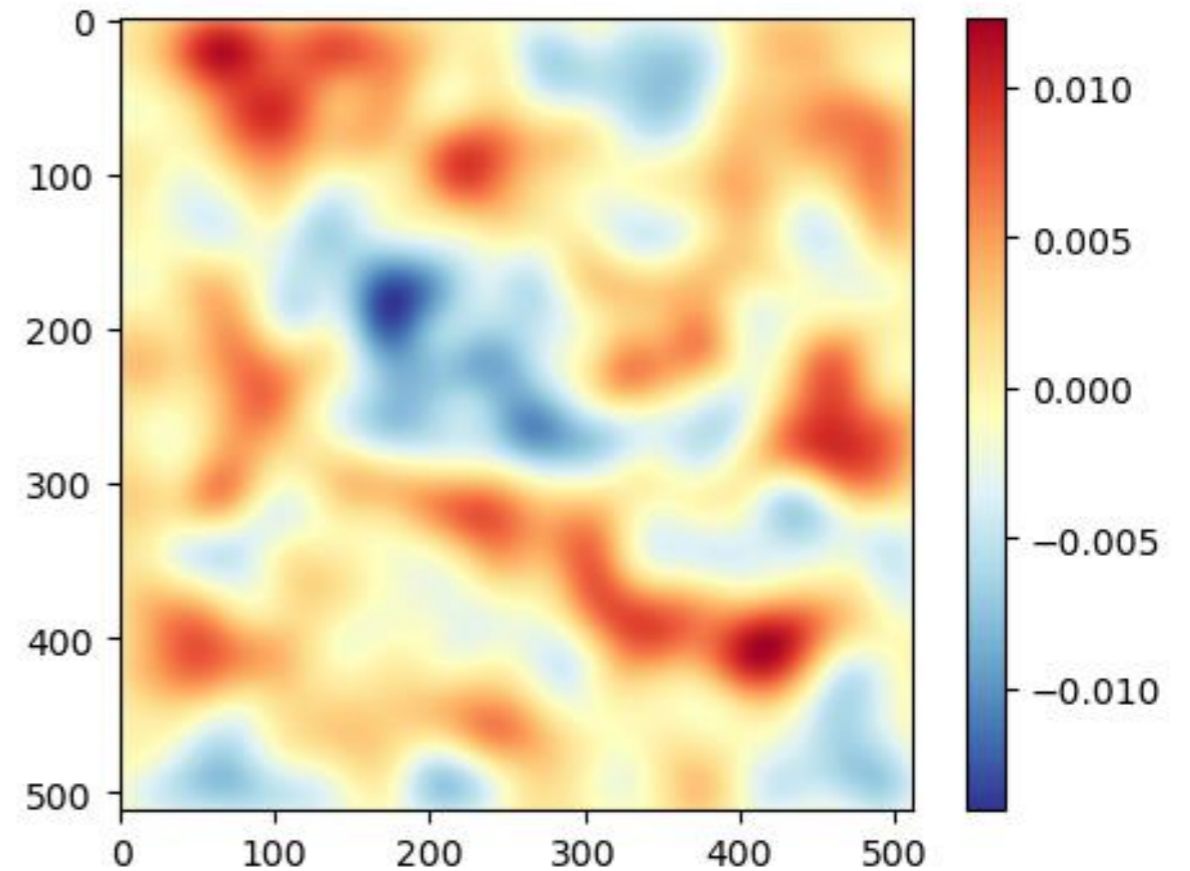


Compared with 21cm image @ 121 MHz

Recon



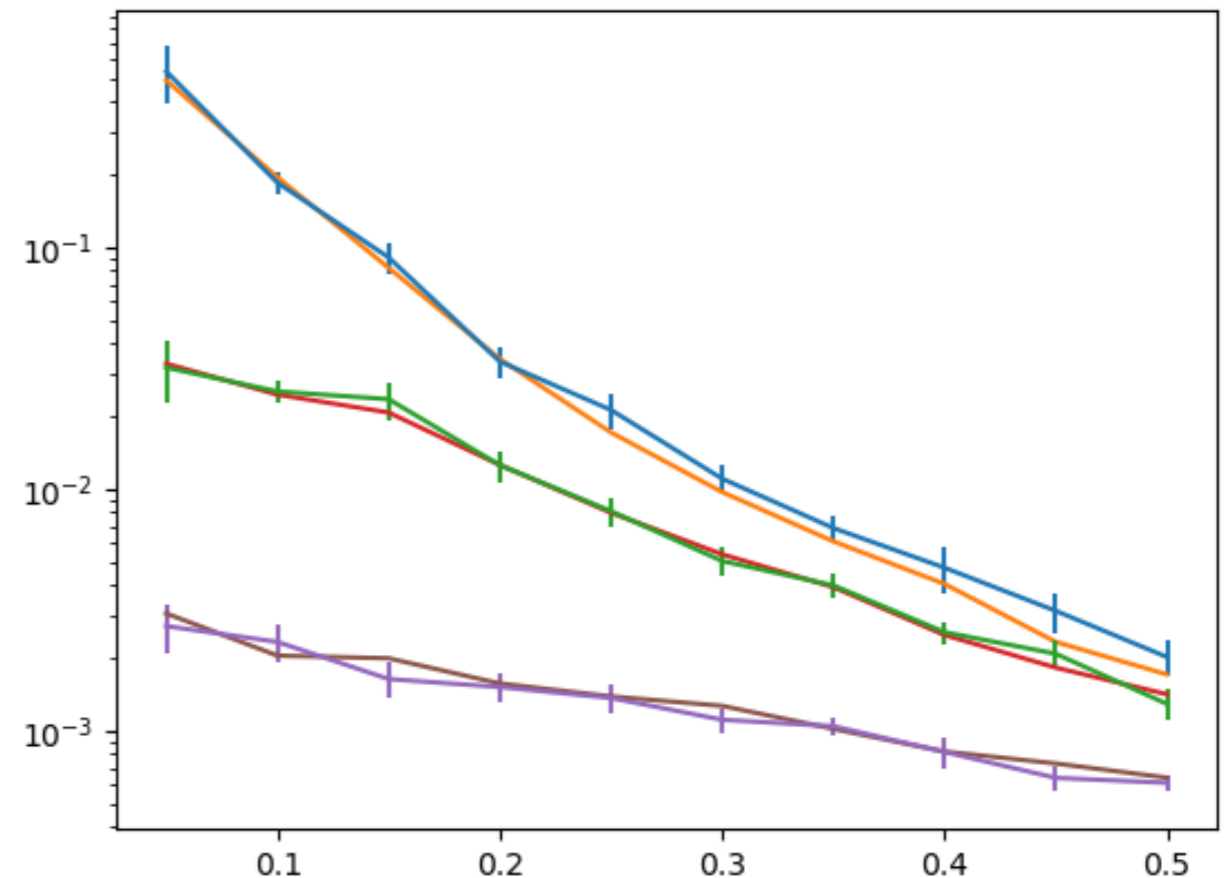
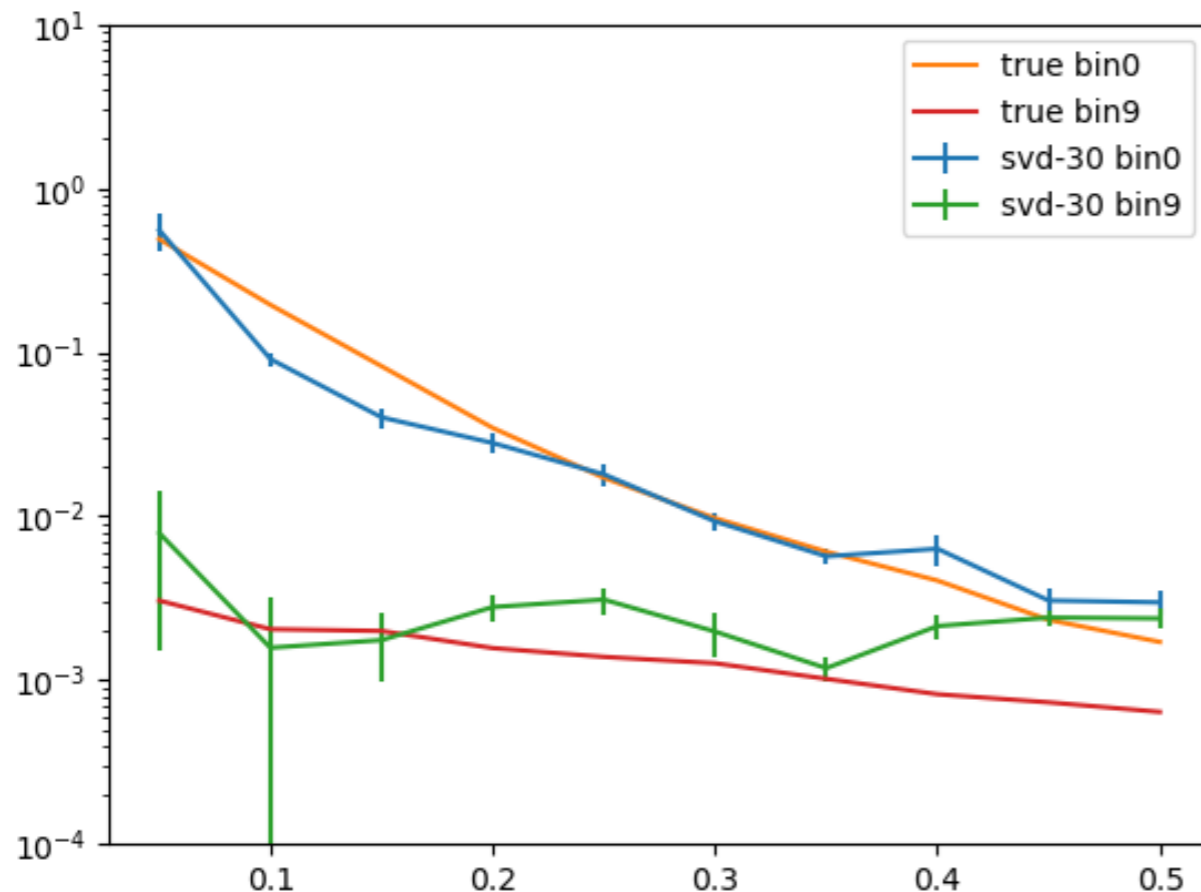
True



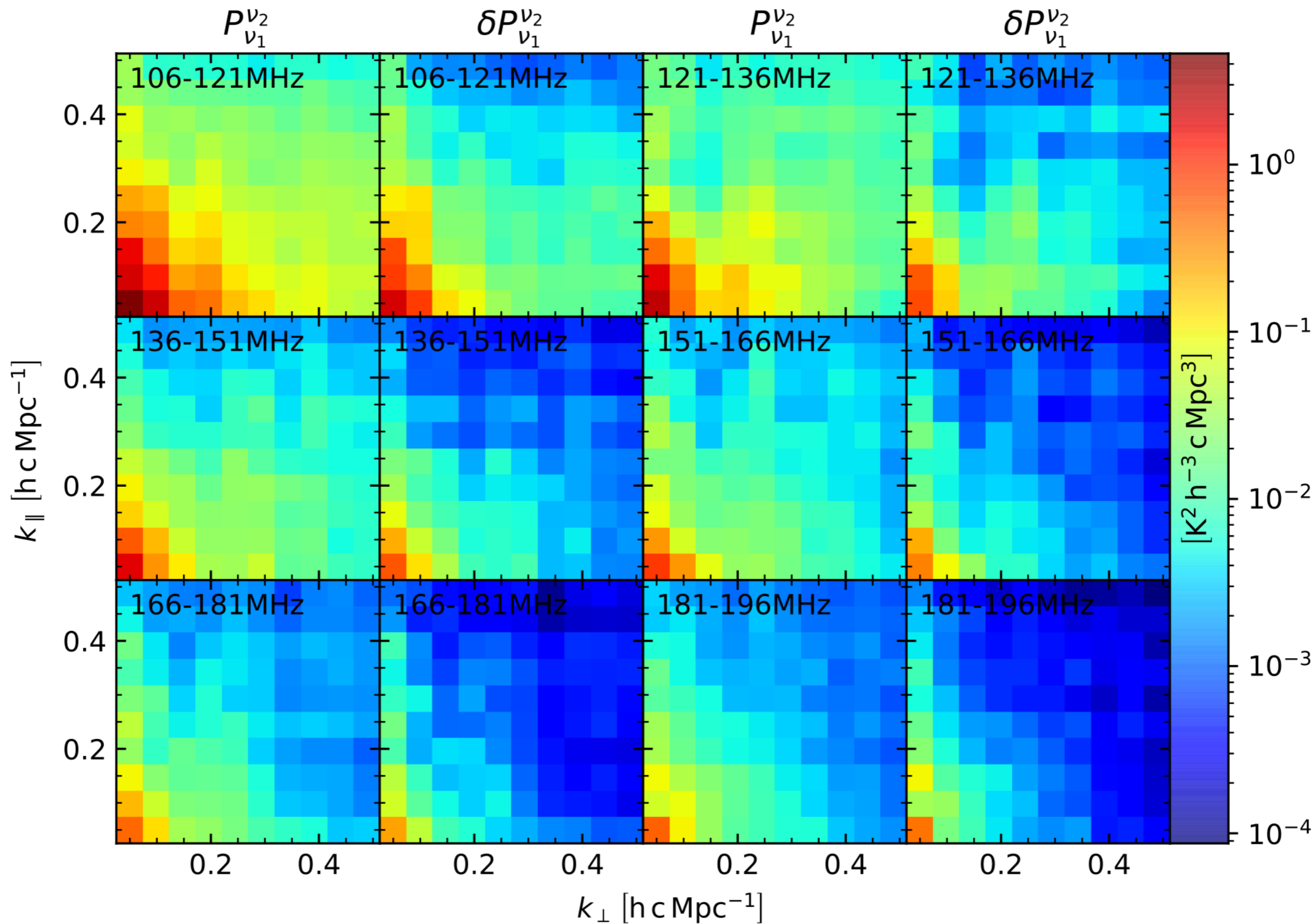
- Correlation coefficient of image $\sim 0.5-0.6$
- SKA Direct imaging of EoR at large scales is promising!

Transfer function: corrected for amplitude

- transfer function — $T(k_{\parallel}, k_{\perp}) \equiv P_{21}^{\text{true}}(k_{\parallel}, k_{\perp}) / P_{21}^{\text{rec}}(k_{\parallel}, k_{\perp})$
- correction for the amplitude of missing modes generated by PCA and normalizations of various instrumental effects.
- use mock realizations to estimate its mean and std
- EoR reconstruction: $P^{\text{reonc}} = T(k_{\parallel}, k_{\perp}) \times P^{\text{svd}}$



Submitted results—HIMALAYA



Collaborative paper in preparation ...

Square Kilometre Array Science Data Challenge 3a: foreground removal for an EoR experiment

TBC

8 April 2024

ABSTRACT

Key words:

1 INTRODUCTION

2 DATA SIMULATION

The SDC3a simulation was undertaken with the OSKAR ([Dulwich et al. 2009](https://ska-telescope.gitlab.io/sim/oskar/), <https://ska-telescope.gitlab.io/sim/oskar/>) package. This makes use of a telescope model and a sky model to generate visibilities with a specified sampling in time and frequency. There is provision to include a variety of instrumental errors into the simulation, which we refer to as the error model.

2.1 Telescope Model

The basic telescope model makes use of the SKA-Low configuration of 512 stations (SKA-TEL-SKO-0000422_04_SKA1Low_Configuration_Coordinates). The station layout is the so-called *Vogel* layout, a one arm spiral configuration with a uniform areal density of antennas and a maximally diverse azimuthal sampling ([Vogel 1979](#)). Only a single station layout has been adopted to reduce computational expense, rather than a full set of 512 diverse station layouts. The implication is that the effective station beam for all visibilities is simply the auto-correlation beam of this one station layout. The most important consequence of this simplification is an elevated far side-lobe response. An attempt to compensate for this effect is made within the sky model definition.

2.2 Sky Model

within the far side-lobe regime of a tracking observation, they would normally be modelled and removed with a so-called “de-mixing” process within a calibration and imaging pipeline. We have consequently attenuated the amplitude of all such sources to simulate both the additional signal attenuation that a cross-correlation station beam would provide (relative to the simplified auto-correlation beam that has been used) as well as a partially successful de-mixing process. The assumed net attenuation due to both of these effects was taken to be a factor of 10^{-3} . This is in excess of the attenuation that is subsequently provided by OSKAR via the auto-correlation station beam defined in the Telescope Model. Each source in the model is represented by an elliptical Gaussian approximation spatially and with a frequency-dependent flux density determined from an amplitude and spectral index defined at some reference frequency.

2.2.2 Inner sky model

The inner sky model, defined within the first null of the station beam pattern at the lowest observing frequency, has been constructed from a number of components.

The first of these is the composite GLEAM and LoBES catalogue mentioned previously. All sources with a 150 MHz flux density greater than 100 mJy (the nominal completeness limit of that survey and some 1900 in number) were included with a Gaussian representation as noted previously. The extragalactic source population at flux densities less than 100 mJy and down to $1 \mu\text{Jy}$ (at 150 MHz) over a spatial extent of 8×8 degrees was modelled with the T-RECS

Conclusions

- SDC3 brings together all interested scientists to join the EoR/21-cm community
- Inventing and sharing EoR reconstruction methods is the most interesting part in SDC3
- Methods can be applied to data that have been observed so far, and when tested, it is expected to show power in SKA data

Thanks !