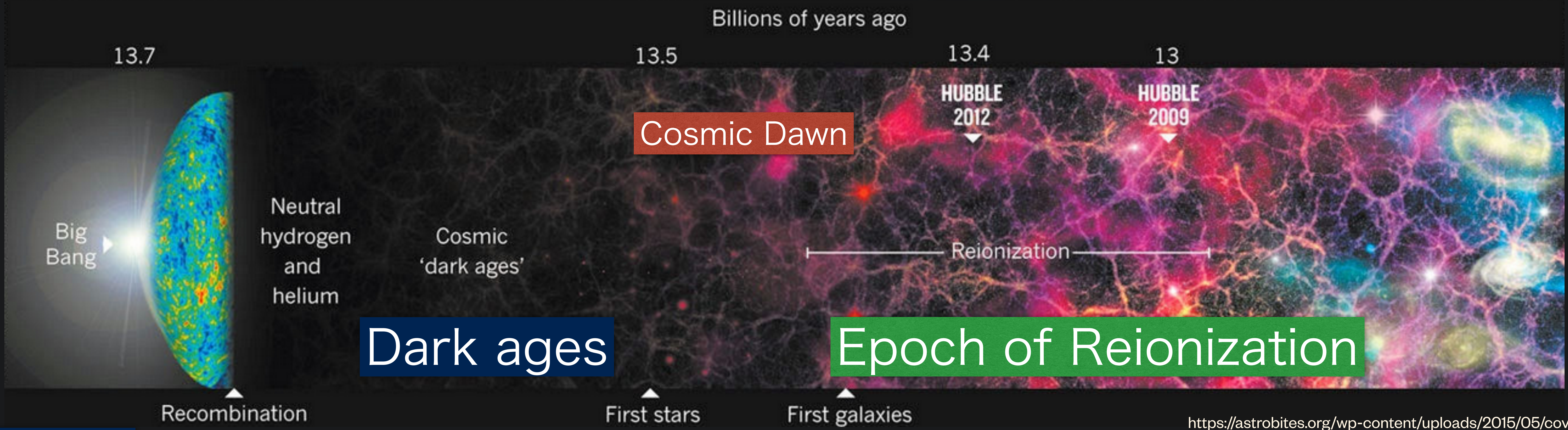
Foreground Removal with Gaussian Process Regression for observing the Epoch of Reionization

Takumi Ito (2nd year of Master course)
Kumamoto University

Shintaro Yoshiura(NAOJ), Keitaro Takahashi(Kumamoto univ)

What is the Epoch of Reionziation



https://astrobites.org/wp-content/uploads/2015/05/cover.png

Dark ages

→After the cosmic recombination, and there have been no luminous objects

Cosmic Dawn

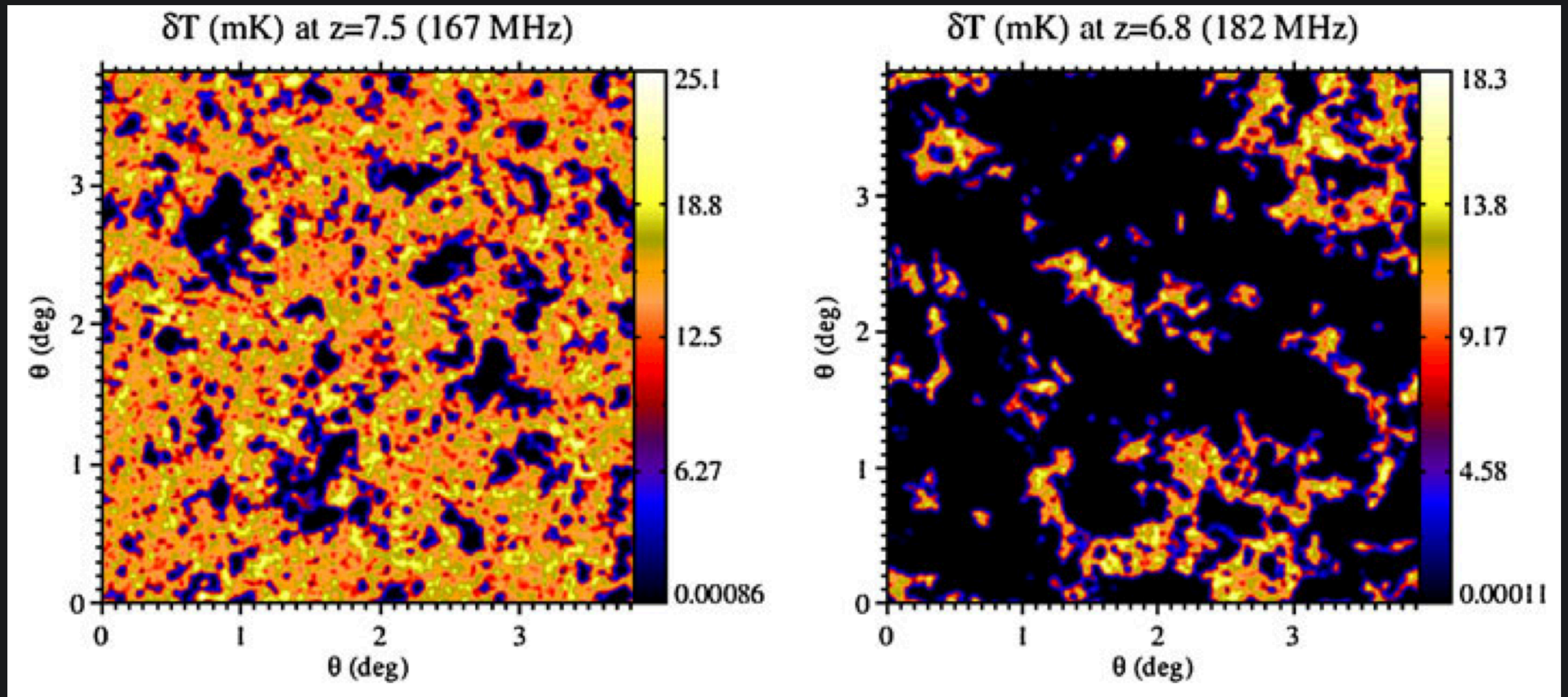
→First luminous objects are formed.

Epoch of Reionization

→ionizing photons from galaxies ionized the neutral hydrogen gas distributed in the Universe.

21 cm line

- We can observe IGM at the EoR via HI 21cm line.
- The 21 cm line emission is due to the spin flip of neutral hydrogen atoms
- HI distribution at different redshifts can be observed by different frequencies
  →We can follow the evolution of IGM

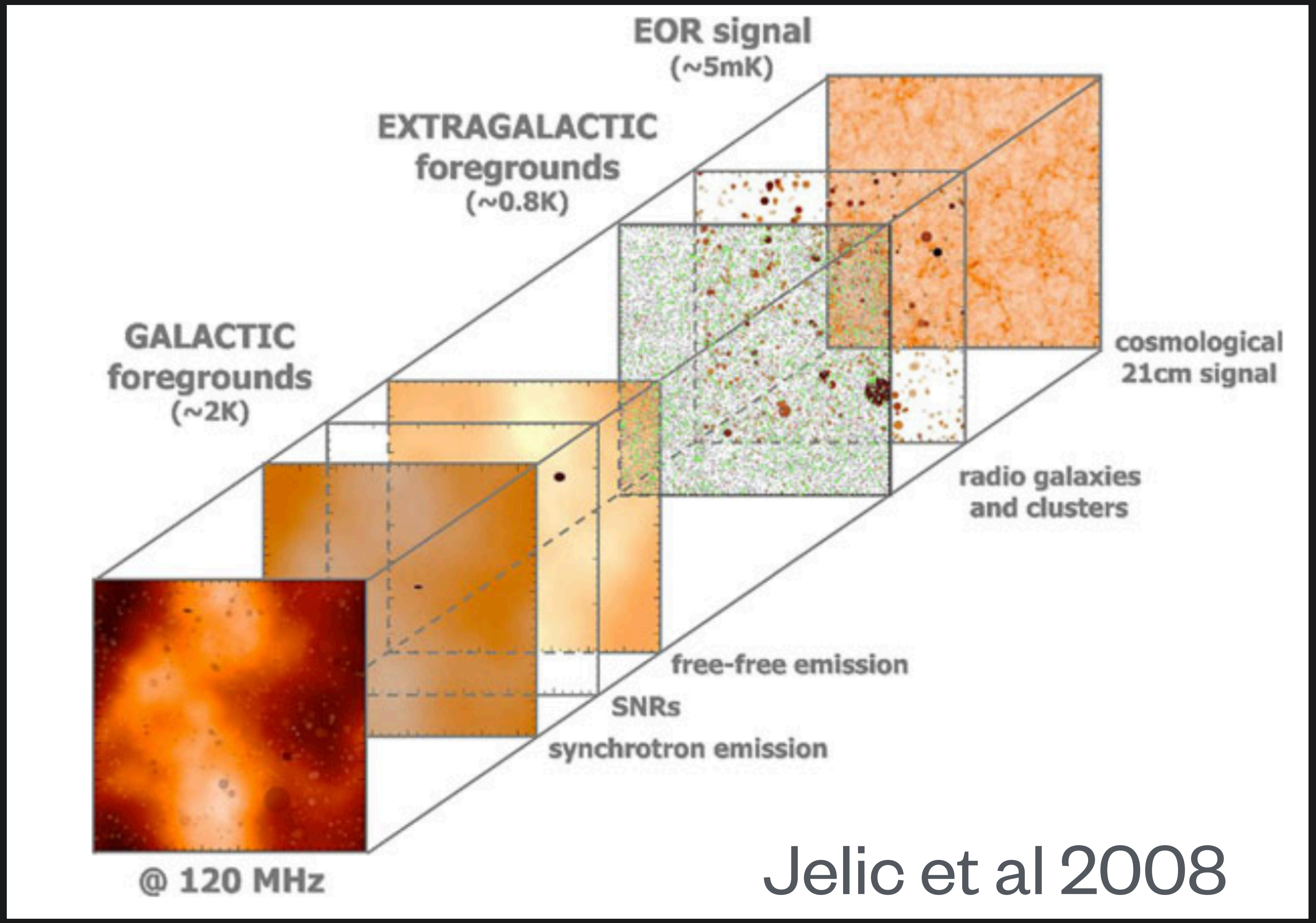


z=7.5 z=6.8

IGM simulation
orange: HI emission (not ionized)
black: no HI emission(ionized)

Mellena et al 2013

Foreground



- Signal = FG + EoR signal + noise
- FG is brighter than EoR signal ($\sim 10^3$ in order)
- Avoidance and Removal of FG are important

- How to remove FG?

  →use difference between FG and EoR signal

  - Emission strength (FG >> EoR signal)

  -Spectral behavior (FG:smooth)

- There are various foreground removal techniques
  - Generalized Morphological Component Analysis(GMCA)
  - FastICA
  - Principal Component Analysis(PCA)
  - Gaussian Process Regression(GPR)
- GPR has been applied to LOFAR foreground removal (Mertens et al 2020)

We apply GPR to the MWA data and try to remove FG

Murchison Widefield Array(MWA)

SKA low pathfinder

spec(phase1)

- Frequency 70~300MHz
- number of tiles 128
- number of baselines 8128
- field of view 610 sq deg @ 150MHz
- spectral resolution 20kHz

Code

- GPR4im is a package uses Gaussian Process Regression (GPR) as a foreground removal technique in the context single-dish 21cm intensity mapping. (Soares P. S., Watkinson C. A., Cunnington S., Pourtsidou A., 2021)
  → I apply GPR4im to MWA gritted visibility

Data

- MWA Simulation data(2h@EoR0)
- MWA Observational data (high band observation in 2014 (2h@EoR0))

Gaussian Process Regression(GPR)

- Gaussian Process(GP)→Multivariate Gaussian Distribution, N

- If we assume random value f follows GP, we write

$$f \sim N(m, K)$$

where m :mean, K: covariance(kernel)

Foreground Removal with GPR

- Observed data is described as a column vector, $\mathbf{d}$, containing the instrumental measurements at each frequency(column vector→visibility, pixel of intensity map…)

$$\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}; \nu = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_n \end{bmatrix}$$

- Our data vector $\mathbf{d}$ consists of the foreground ($\mathbf{f_{fg}}$), EoR signal ($\mathbf{f_{21}}$), noise(n)

$$\mathbf{d} = \mathbf{f_{fg}} + \mathbf{f_{21}} + \mathbf{n}$$

- Assuming each component to be statistically uncorrelated, the covariance of the data K is given by

$$K = K_{\mathrm{fg}} + K_{21} + K_{\mathrm{n}}$$

- Assuming the data vector is Gaussian distributed, we can model its probability distribution as

$$\mathbf{d} \sim N(m(\nu), K(\nu, \nu))$$

where m:mean function, K:covariance function

- We can write joint probability distribution of data and Foreground

$$\begin{bmatrix} \mathbf{d} \\ \mathbf{f}_{\mathrm{fg}} \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathrm{K}_{\mathrm{fg}} + \mathrm{K}_{21} + \mathrm{K}_{\mathrm{n}} & \mathrm{K}_{\mathrm{fg}} \\ \mathrm{K}_{\mathrm{fg}} & \mathrm{K}_{\mathrm{fg}} \end{bmatrix} \right)$$

- From above equation, the expectation value $\mathrm{E}[\mathrm{f}_{\mathrm{fg}}]$ and covariance $\mathrm{Cov}[\mathrm{f}_{\mathrm{fg}}]$ is driven as

$$\mathrm{E}[\mathbf{f}_{\mathrm{fg}}] = K_{fg}[K_{fg} + K_{21} + K_n]^{-1}\mathbf{d}$$

$$\mathrm{Cov}(\mathbf{f}_{\mathbf{fg}}) = K_{fg} - K_{fg}[K_{fg} + K_{21} + K_n]^{-1}K_{fg}$$

Flow outline

- Choose kernels of FG,21cm,noise (be described later)

$$K = K_{\mathrm{fg}} + K_{21} + K_{\mathrm{n}}$$

- Fit kernels to UV gritted visibilities and make FG model from the kernels

$$\mathrm{E}[\mathbf{f}_{\mathrm{fg}}] = K_{fg}[K_{fg} + K_{21} + K_{n}]^{-1}\mathbf{d}$$

- Subtract FG model from data

$$\mathbf{r} = \mathbf{d} - \mathrm{E}[\mathbf{f}_{\mathrm{fg}}]$$

Covariance (Kernel)

- Matern kernel is widely used kernels in GPR

$$K_{\mathrm{Matern}}(\nu,\nu') = \sigma^2 \frac{2^{1-\eta}}{\Gamma(\eta)} \left(\sqrt{2\eta} \frac{|\nu - \nu'|}{l} \right)^{\eta} K_{\eta} \left(\sqrt{2\eta} \frac{|\nu - \nu'|}{l} \right)$$

Γ:gamma function, K_η:modified Bessel function of the second kind,

σ^2:Variance(amplitude of the signal)

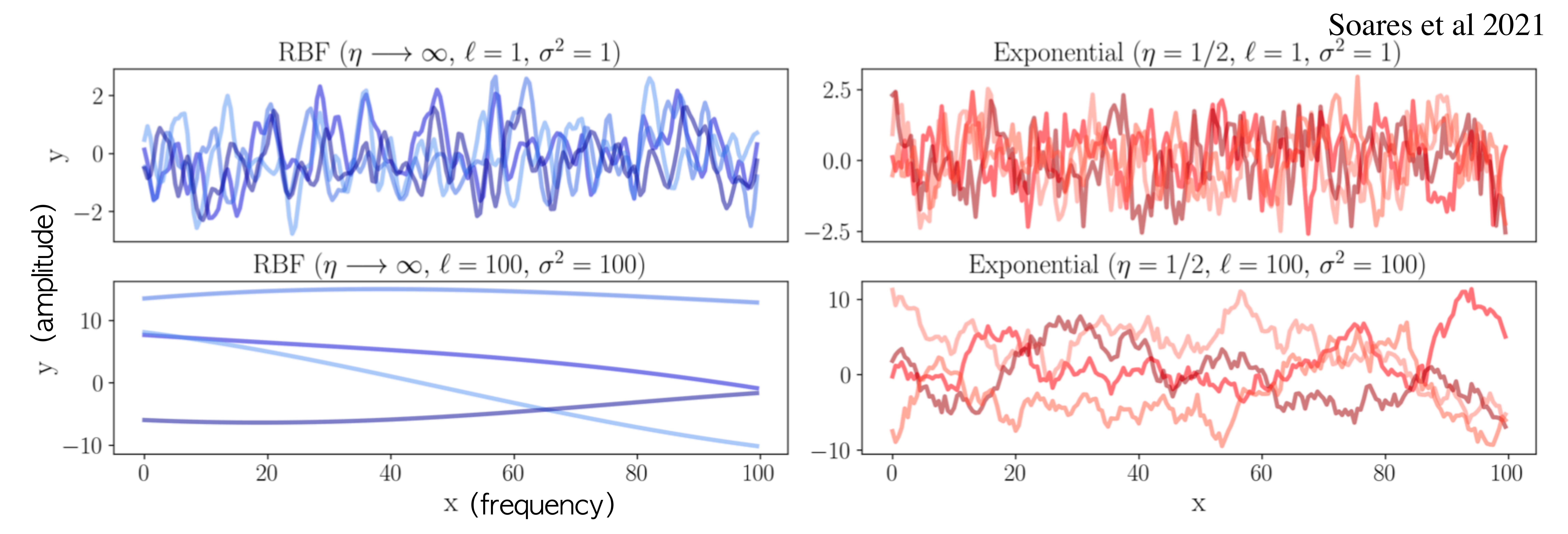
l:Length scale(topical scale of correlations in the data across frequency)

η:spectral parameter(It determines the overall “smoothness” of the data)

- Next page explains how these parameters work

Comparing parameters

randomly generated data plots with Metern kernels, with shown parameters



σ^2:Variance → A lager σ^2 means the signal is stronger.

l:Length scale → A lager l means the data is more correlated in frequency

η:spectral parameter → A lager η means the data is smoother

Choosing kernels(Mertens et al 2020)

- Foreground kernel

  -$\eta \to \infty$:Very smooth FG

  -$\eta = 3/2$:FG components that have medium frequency smoothness

$$K_{\mathrm{fg}} = K_{\infty} + K_{3/2}$$

- HI kernel

  -$\eta = 1/2$:spectrally varying signal such as HI signal

$$K_{21} = K_{1/2}$$

- Noise kernel

  - Assume that thermal noise is constant in frequency

$$K_{\mathrm{n}} = \sigma^2 \delta_{\nu,\nu'}$$

- Fit covariance $K = K_{\mathrm{fg}} + K_{21} + K_{\mathrm{n}}$ to data covariance(l, σ^2:free parameter)

  (I don't use K_{21} to simulation data since the simulated data consists of FG and noise)

Result(simulation)

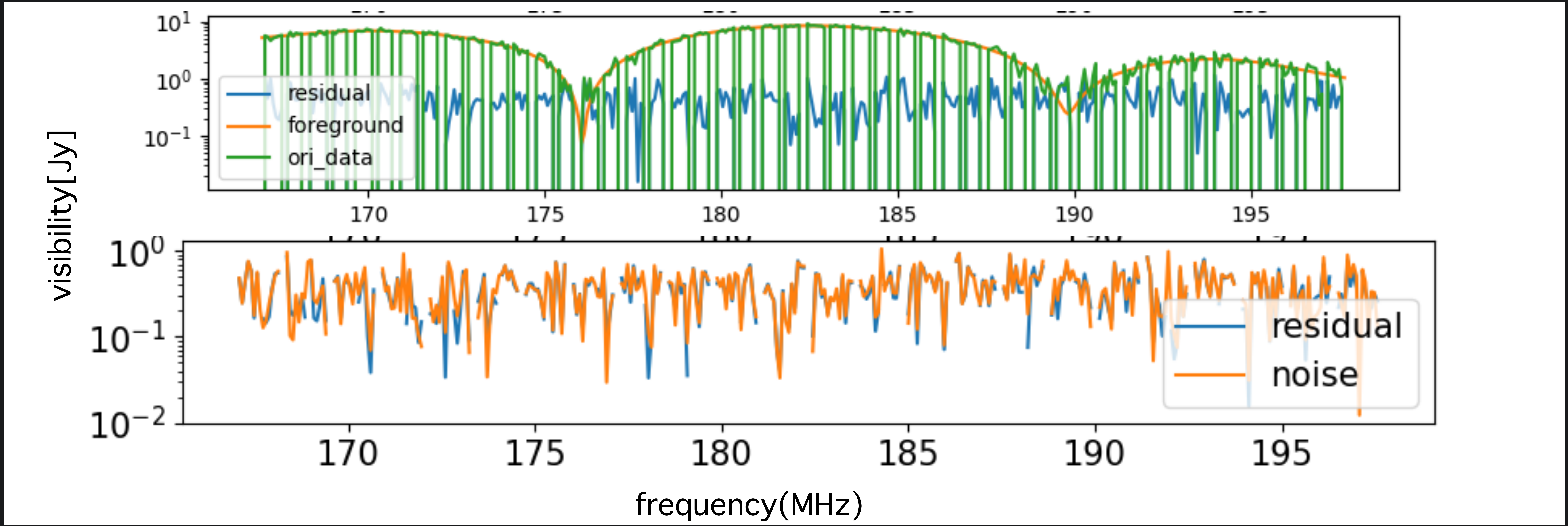
- MWA high band simulation using RTS (2h@EoR0)
- Foreground(2000 point sources) + Thermal noise

✓ I will show you …

- Visibility

-2D power spectrum

Results(simulation)

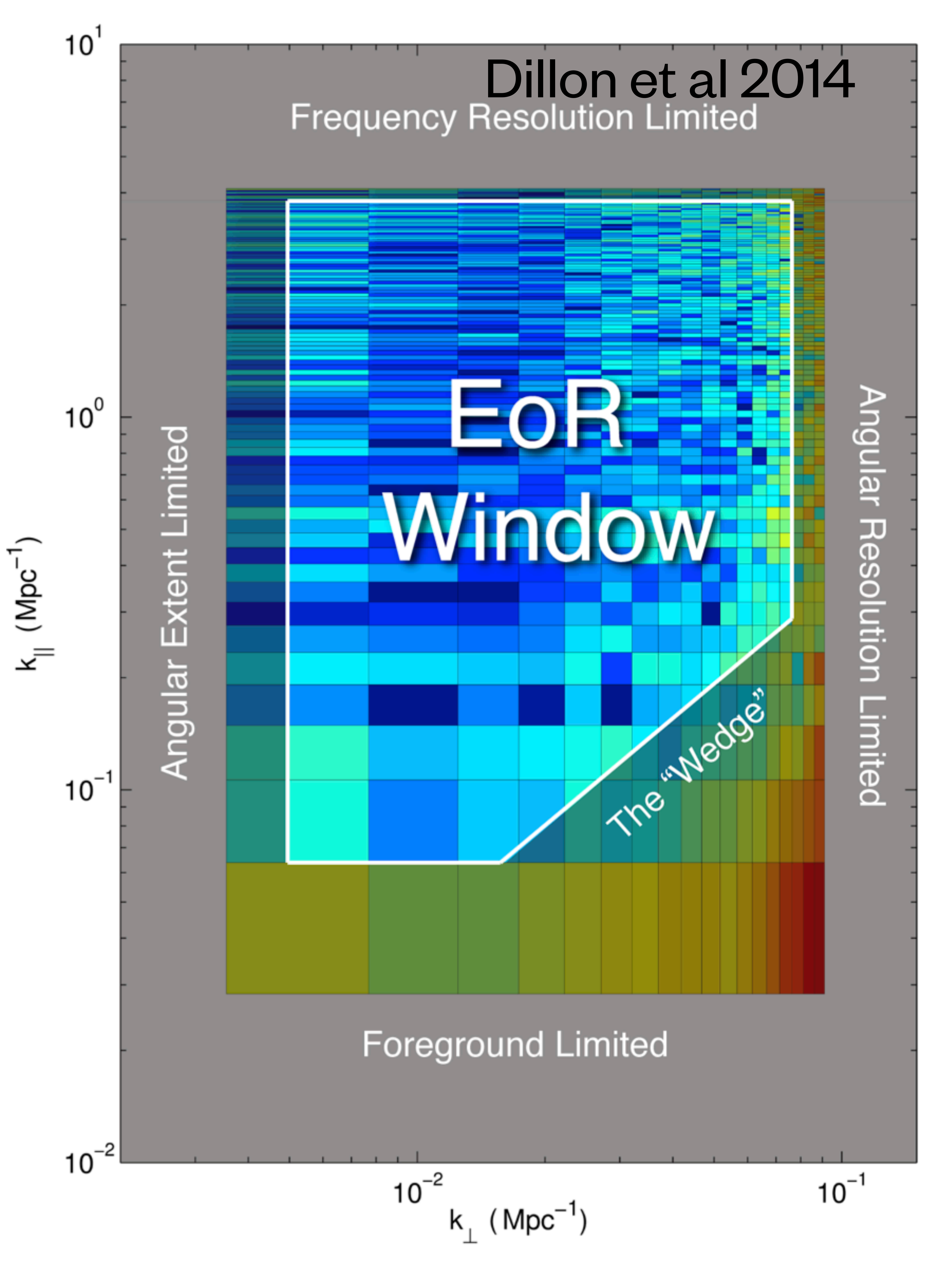


- Simulated visibilities before/after FG removal using GPR and its Residual

  -Signals getting weaker (order 10^1)

- Compare Residual and input noise

  -residual and the input noise are same order.



2D Power Spectrum

$k_{||}$: wavenumber parallel to line of sight

-spectrally smooth(FG)→ low $k_{||}$

-not smooth(HI) →high $k_{||}$

$k_{\perp}$: wavenumber perpendicular to line of sight

-diffuse emission(FG) → high $k_{\perp}$

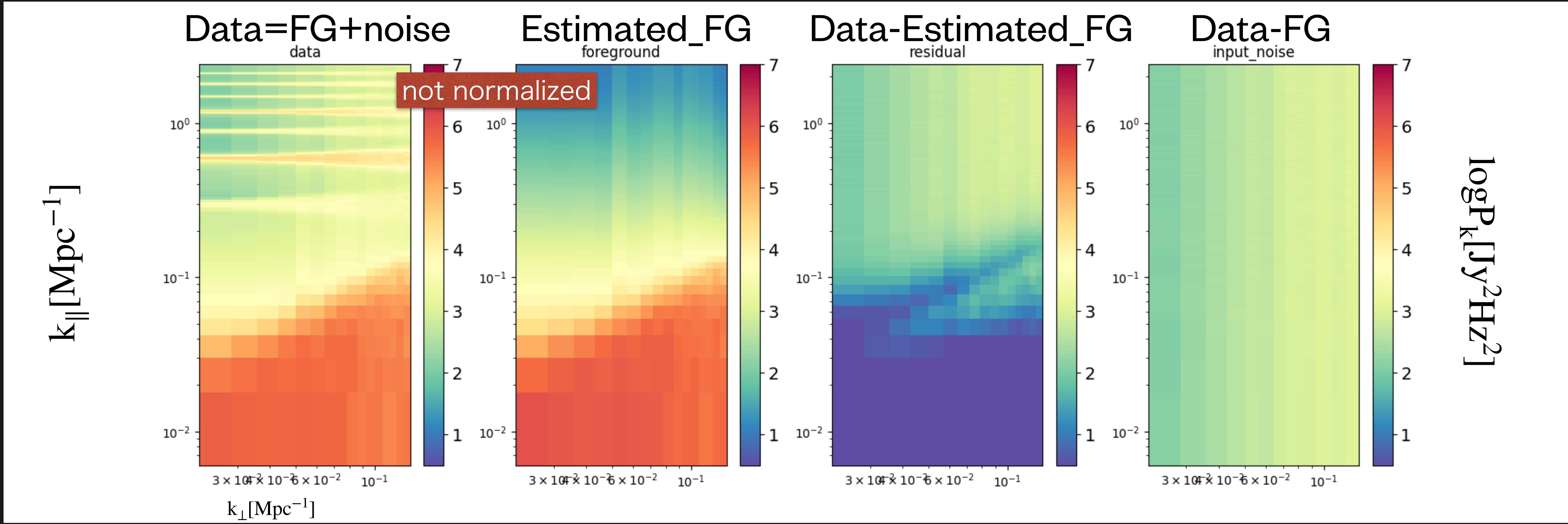
EoR window

- Lower foregrounds

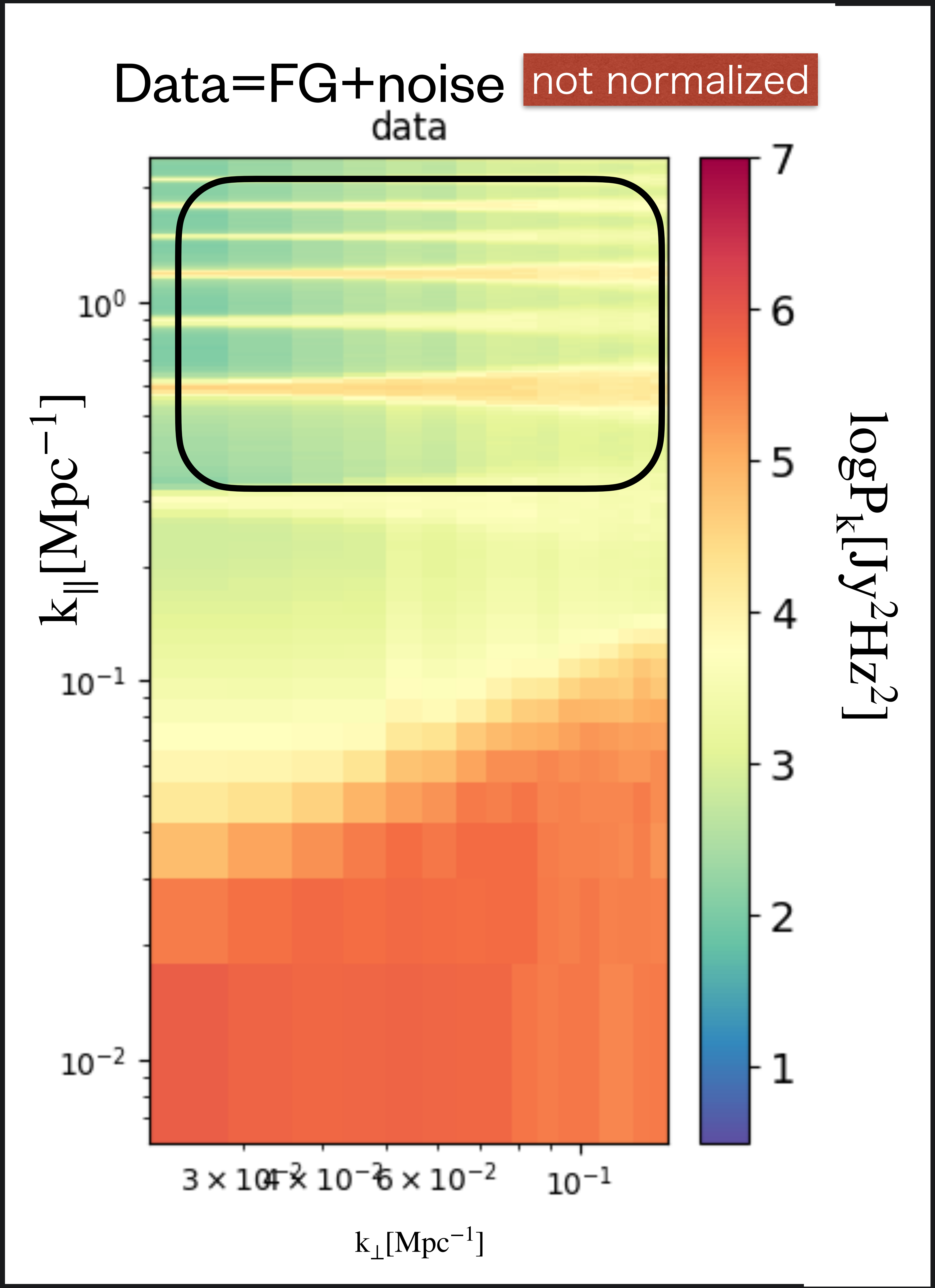
The "Wedge"

-EoR signal + Foreground contamination

Result 2DPS(simulation)



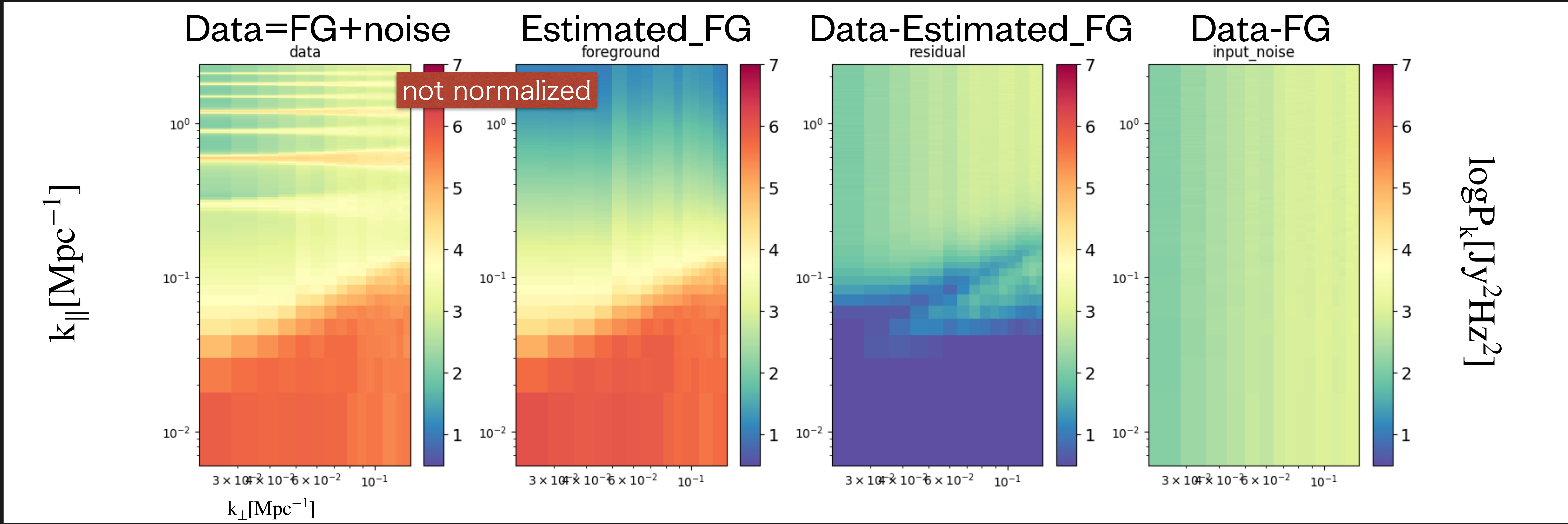
- Bright signals are subtracted and it makes coarse band harmonics weaker.
- Residual looks reproduce input noise on $k_{||}$>0.2
- $k_{||}$<0.2 are overfitted



MWA coarse band harmonics

These lines are systematics caused by flagging aliased channels from the polyphase filter banks

Result 2DPS(simulation)



- Bright signals are subtracted and it makes coarse band harmonics weaker.
- Residual looks reproduce input noise on $k_{\parallel}$>0.2
- $k_{\parallel}$<0.2 are overfitted

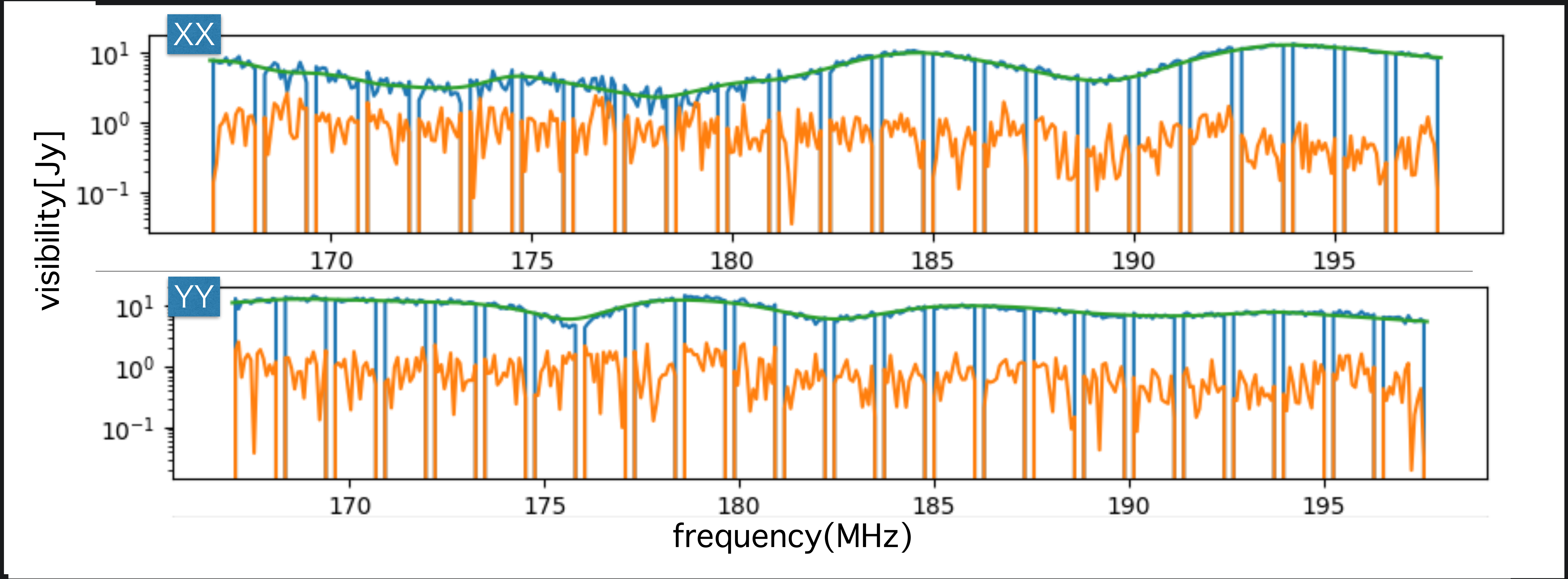
Result(Observation)

- MWA high band observation in 2014 (2h@EoR0)
  -Calibrated by RTS
  -Gridded by CHIPS

✓ I will show you …
  - Visibility
  -2D power spectrum

Result (observation)

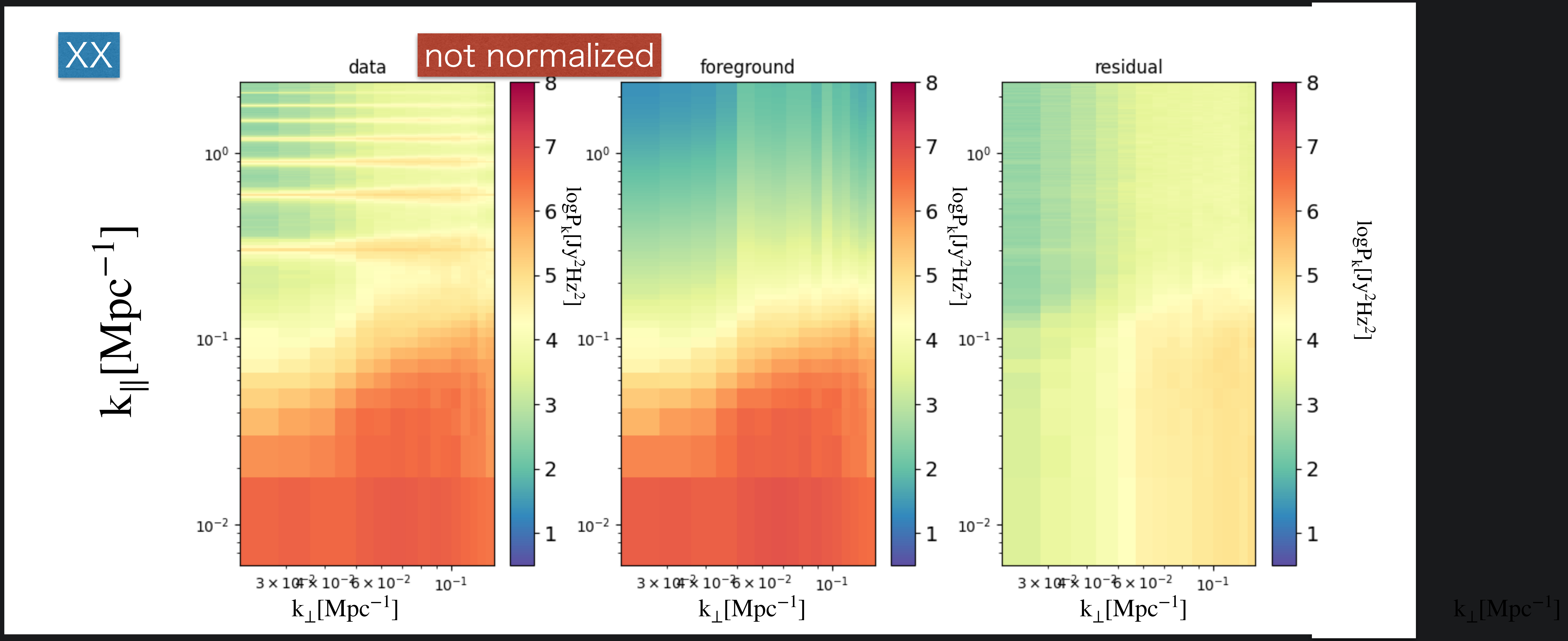
blue: Data
green: Estimated_FG
orange Residual



- one of the gridded visibilities before/after foreground removal and its Residual.
- Signal getting weaker(order 10^1)

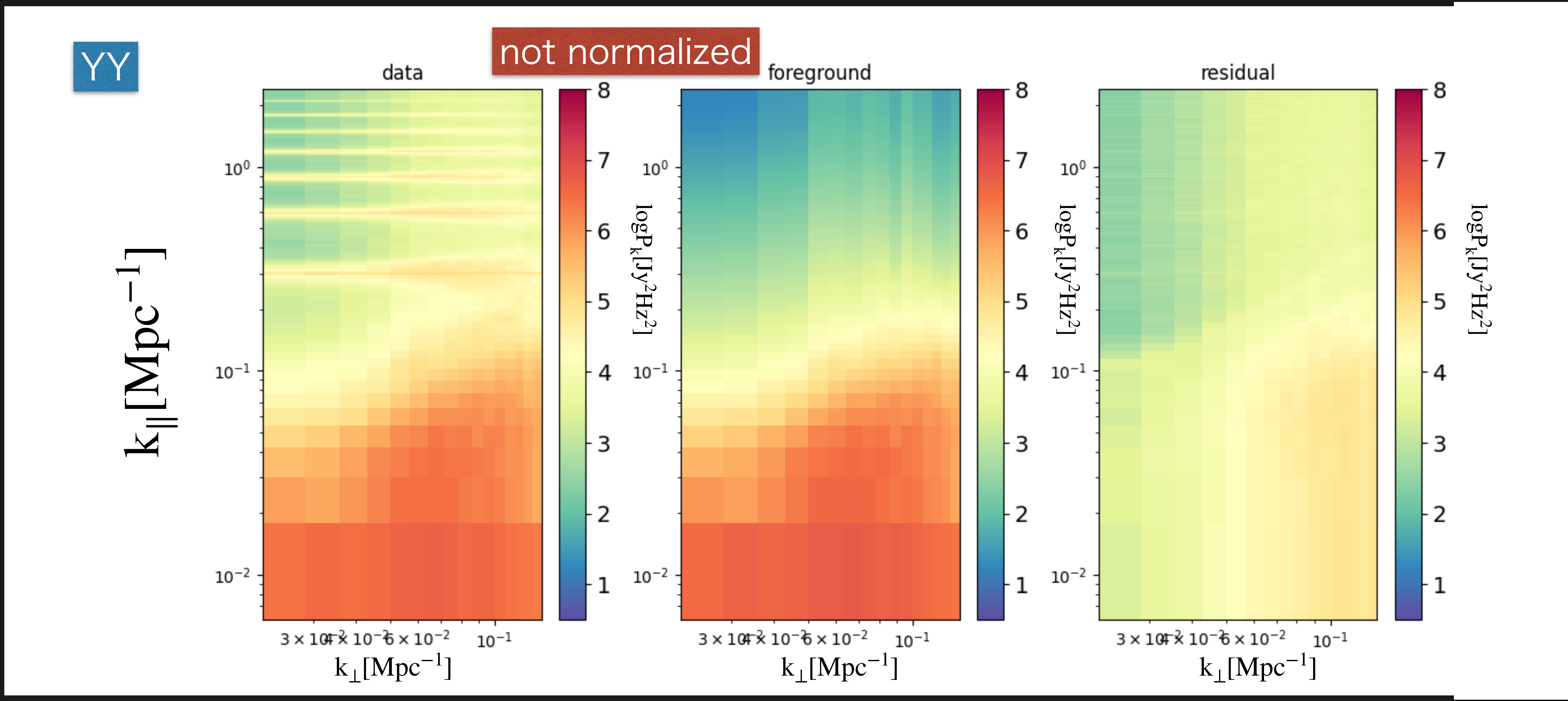
vis_20_1 (29)

Result 2DPS(observation)



- Bright signals are subtracted and it makes coarse band harmonics weaker.
- Signal getting weak on $k_{\parallel}$>0.1

Result 2DPS(observation)



$k_{\perp}[\mathrm{Mpc}^{-1}]$

- Bright signals are subtracted and it makes coarse band harmonics weaker.
- Signal getting weaker on $k_{\|}$>0.1

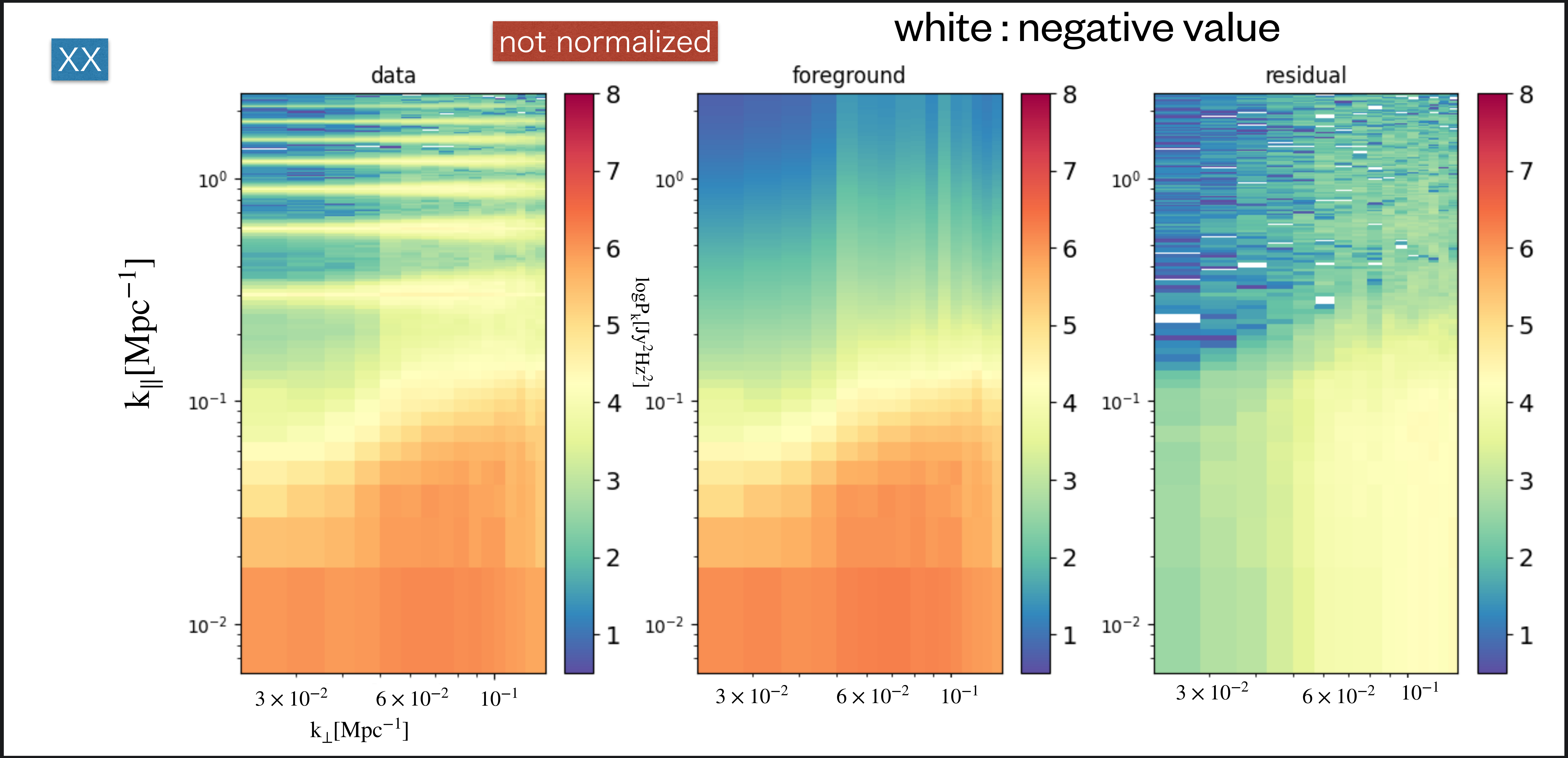
Even/odd cross power spectrum

- Even/odd cross power spectrum
  -lower noise power

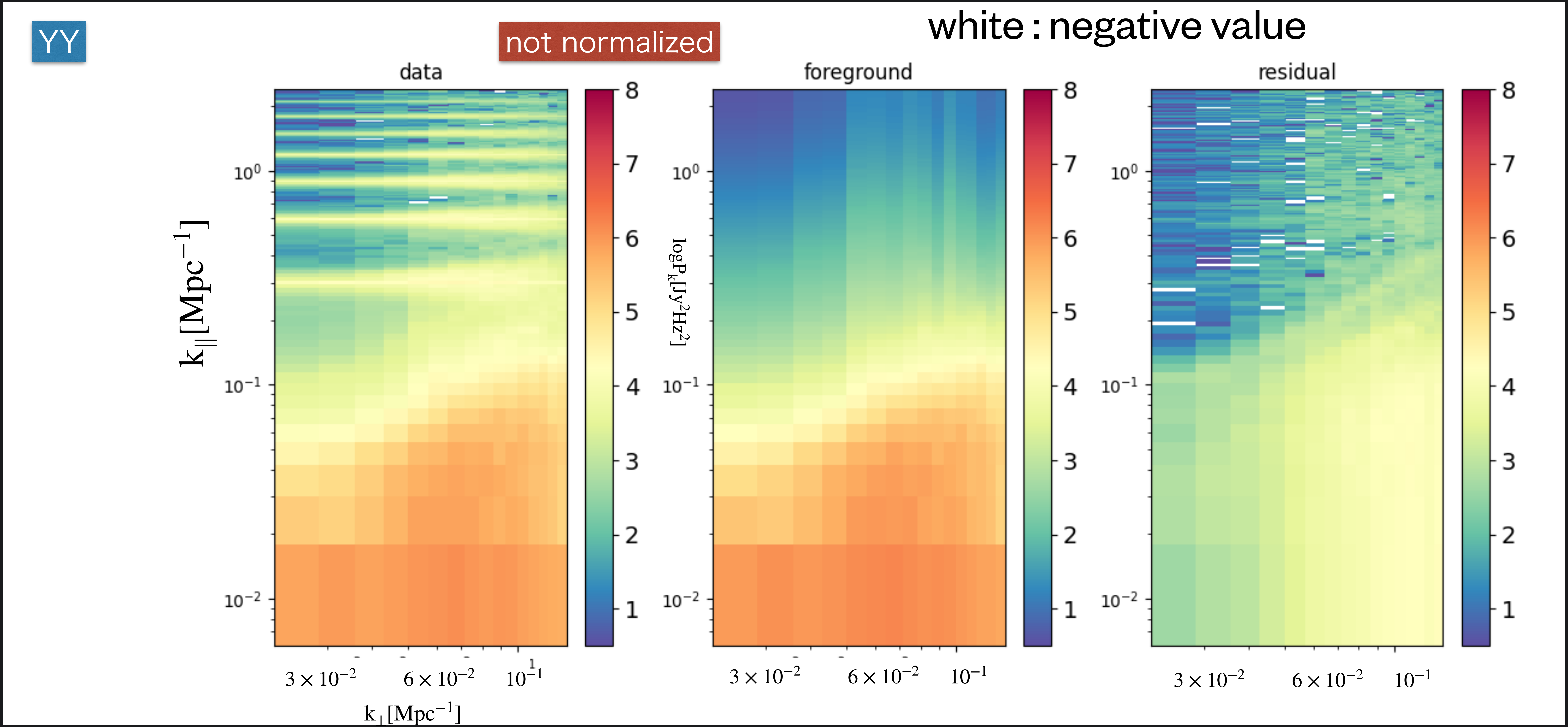
✳ even data cube+odd data cube contains sky signal and noise

✳ even data cube - odd data cube contains only noise

$$((E+O)(E+O)^\dagger - (E-O)(E-O)^\dagger)\frac{1}{4}$$
$$= (EE^\dagger + OO^\dagger + EO^\dagger + OE^\dagger - (EE^\dagger + OO^\dagger - EO^\dagger - OE^\dagger))\frac{1}{4}$$
$$= (2EO^\dagger + 2OE\dagger)\frac{1}{4}$$
$$= R(EO^\dagger)$$

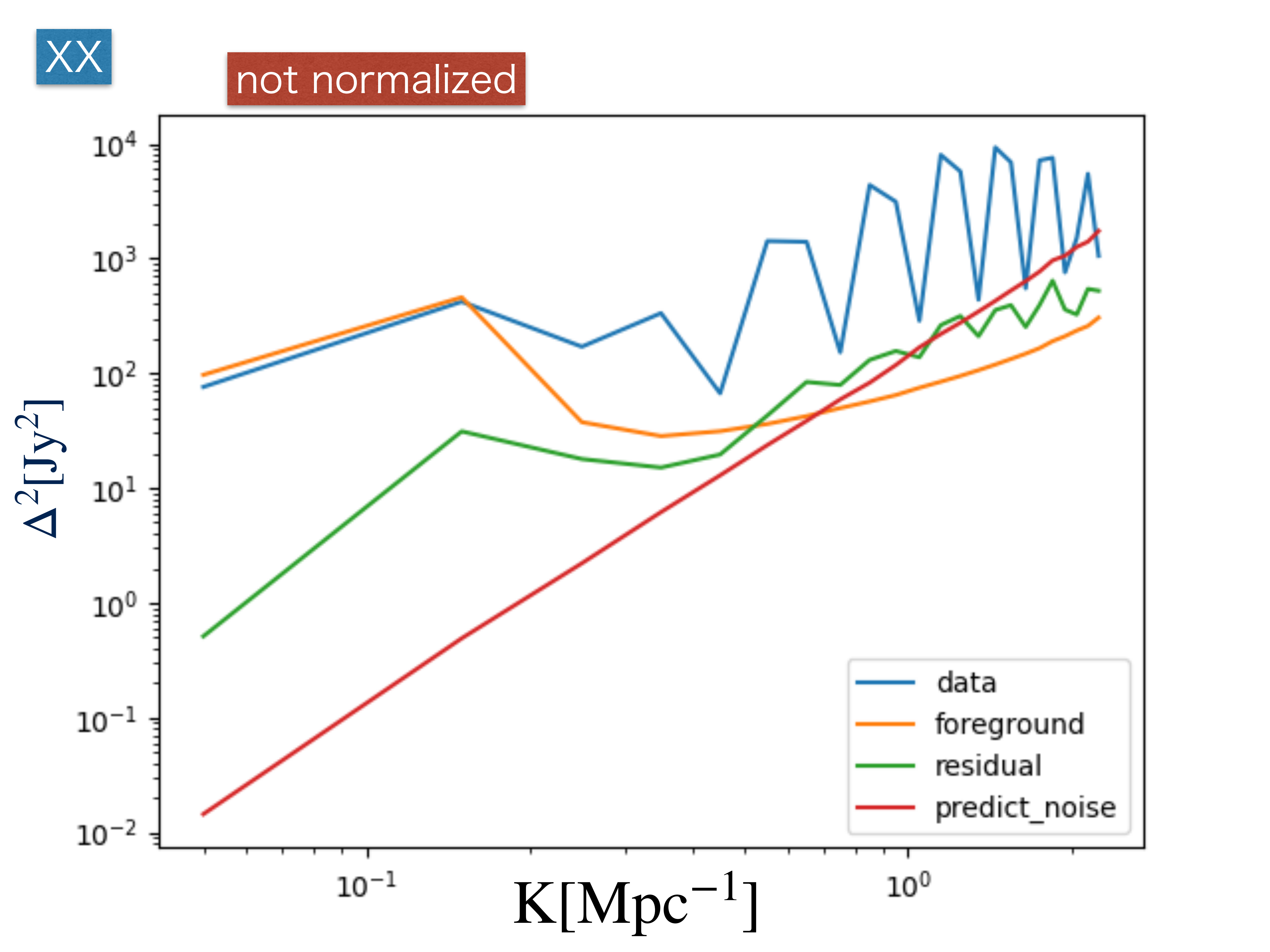


- GPR remove coarse band harmonics ,
  and Residual is consistent with noise in EoR window



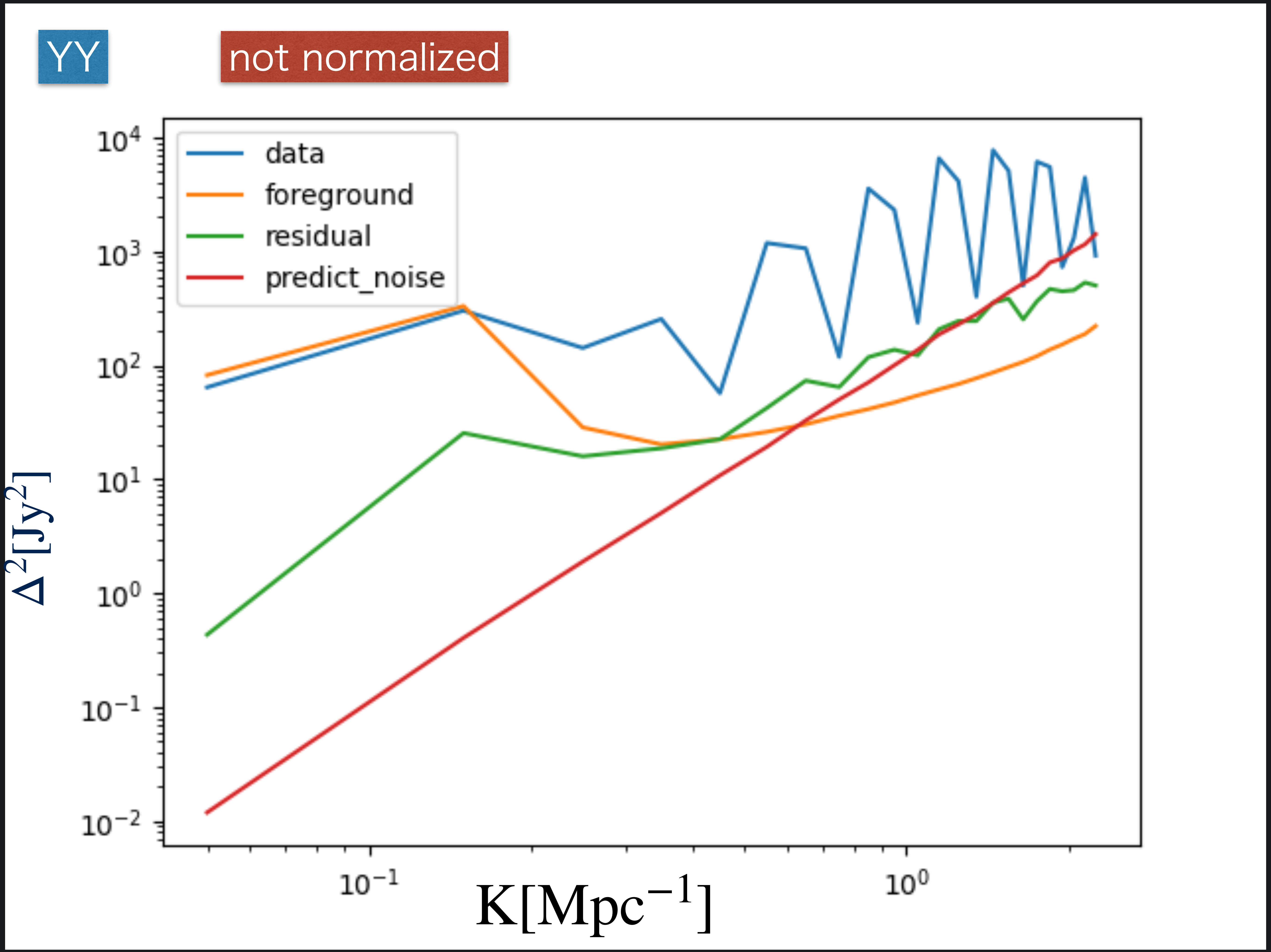
- GPR remove coarse band harmonics ,
  and Residual is consistent with noise in EoR window

Diagnostic power spectra



$$\mathrm{K} = \sqrt{\mathrm{K}_{\mathrm{para}}^2 + \mathrm{K}_{\mathrm{pepr}}^2}$$

- even data cube - odd data cube contains only noise
  →Assuming it's power shows the power of thermal noise.
- Diagnostic PS before/after FG removal using GPR and its Residual
  -Signals getting weaker (order $10^1 \sim 10^2$)
  -Power of data < Power of FG on $\mathrm{K} < 0.1$
  →fitting might be wrong?



$$\mathrm{K} = \sqrt{\mathrm{K}_{\mathrm{para}}^2 + \mathrm{K}_{\mathrm{pepr}}^2}$$

Diagnostic power spectra

- even data cube - odd data cube contains only noise
  →Assuming it's power shows the power of thermal noise.
- Diagnostic PS before/after FG removal using GPR and its Residual
  -Signals getting weaker (order $10^1 \sim 10^2$)
  -Power of data < Power of FG on $\mathrm{K} < 0.1$
  →fitting might be wrong?

Results and FutureWorks

- We tested GPR foreground removal to MWA observational data
  →Signals are getting lower (~10^1 Jy in order at each wavenumber)

- Foregrounds are stronger than data in $K < 10^{-1}$
  →fitting might be wrong
  -Test other kernels
  -Calculate uncertainties