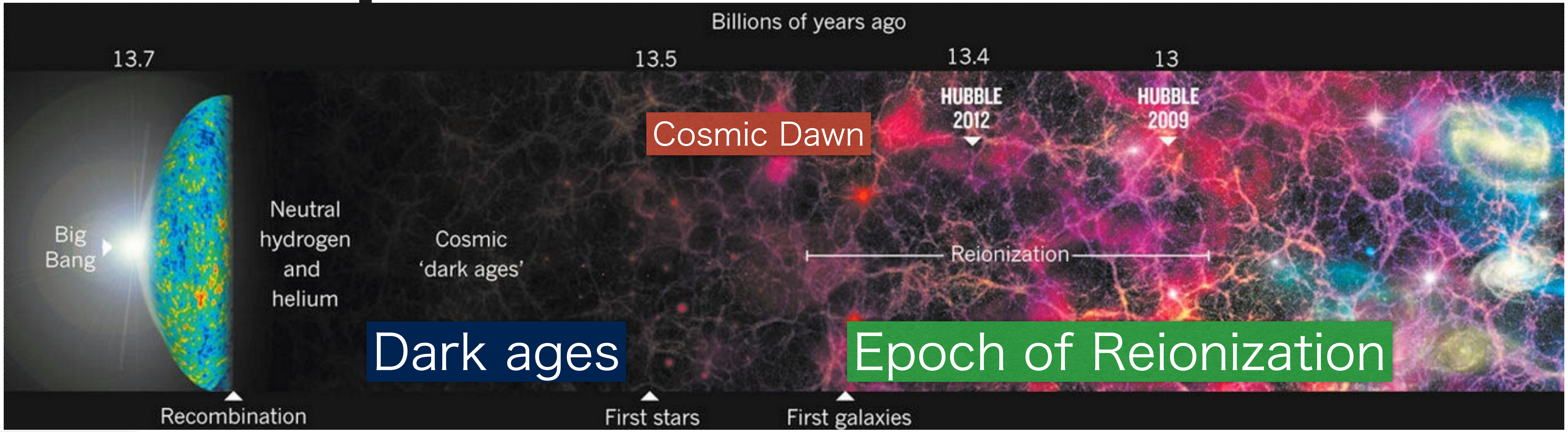
Foreground Removal with Gaussian Process Regression

MWA Project Meeting -2023

Takumi Ito , Keitaro Takahashi(Kumamoto University), Shintaro Yoshiura(NAOJ)

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.

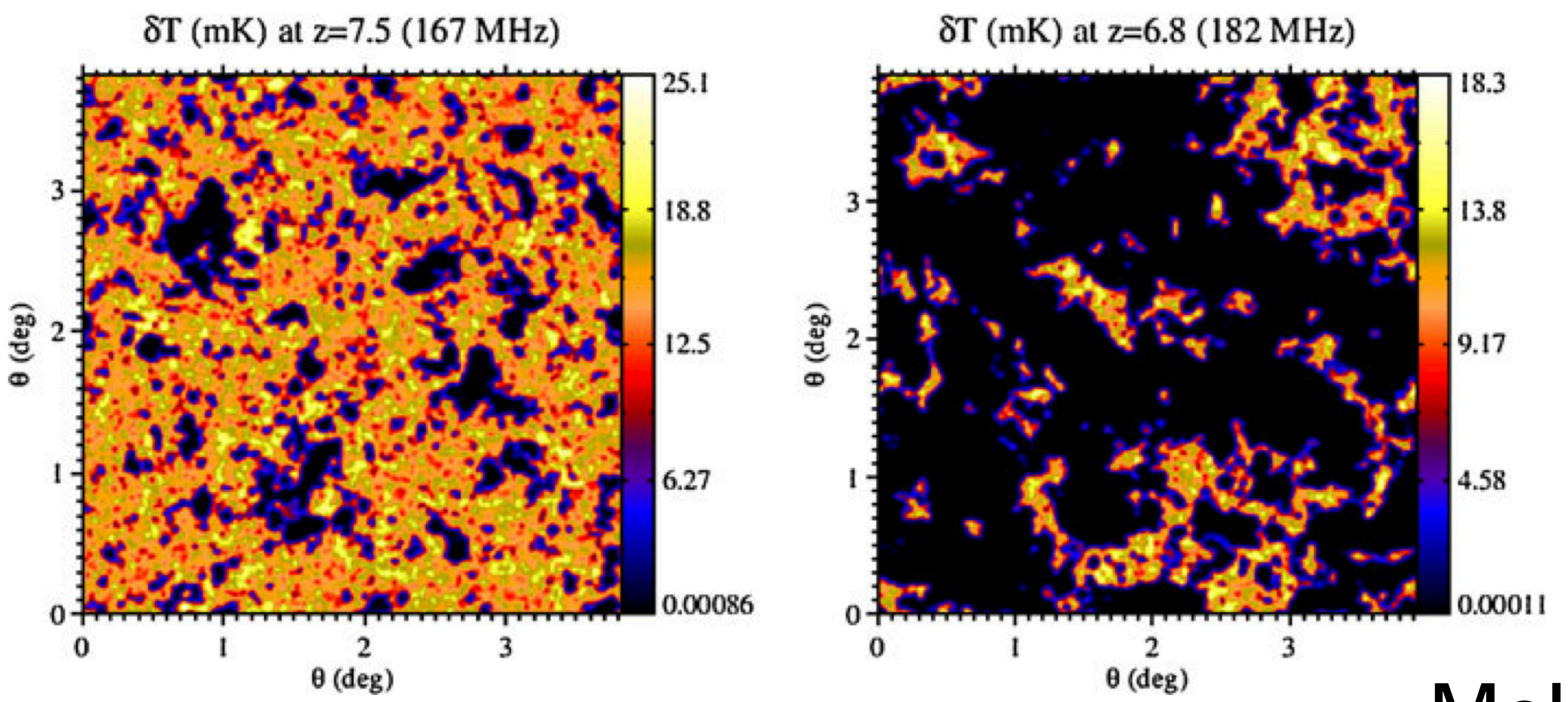
Epoch of Reionization

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

21 cm line

- The 21 cm line emission is due to the spin flip of HI
  →We can observe IGM at the EoR via 21cm line

- HI distribution at different redshifts can be observed by different frequencies
  →We can follow the evolution of IGM



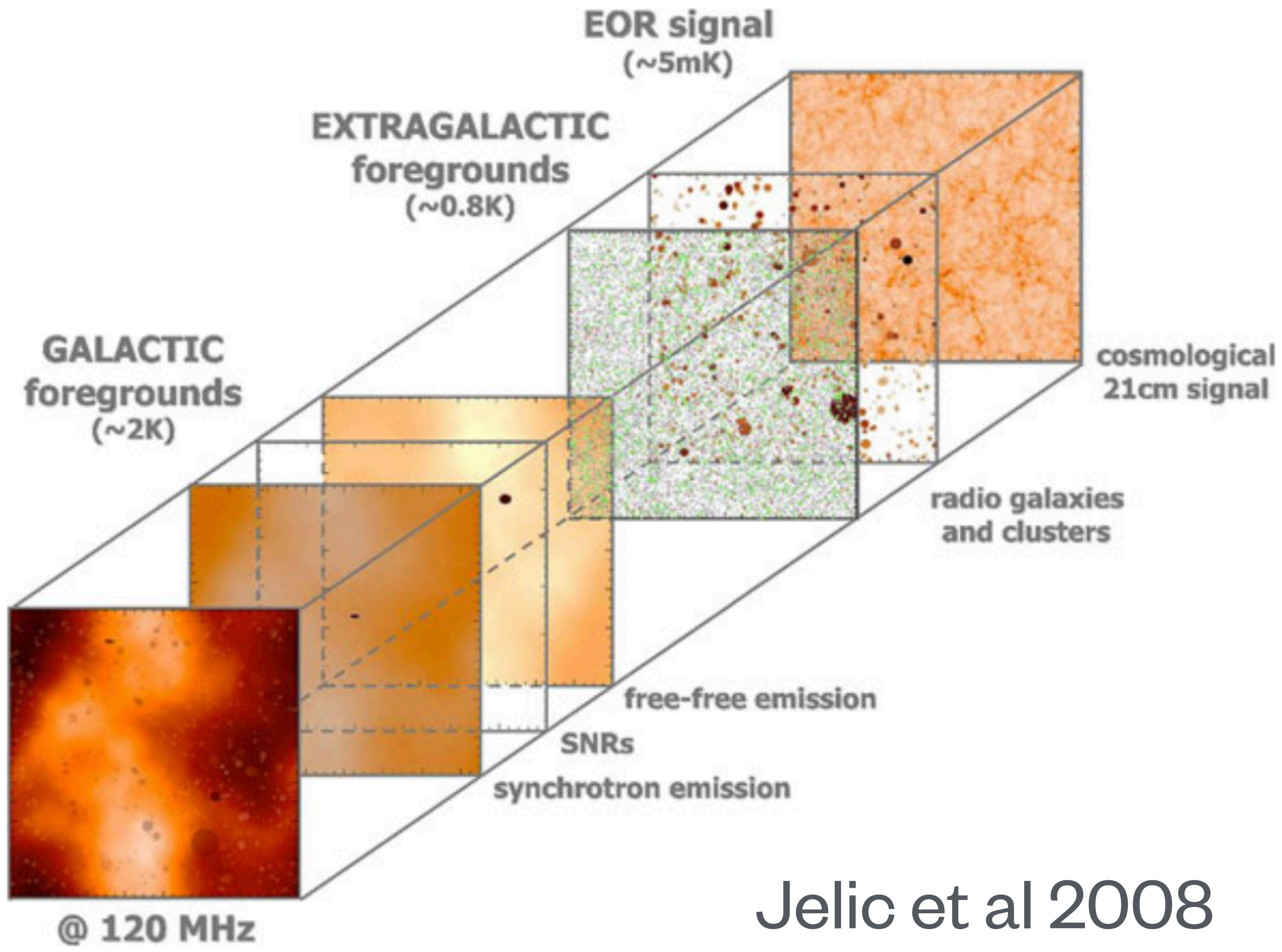
z=7.5 z=6.8

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

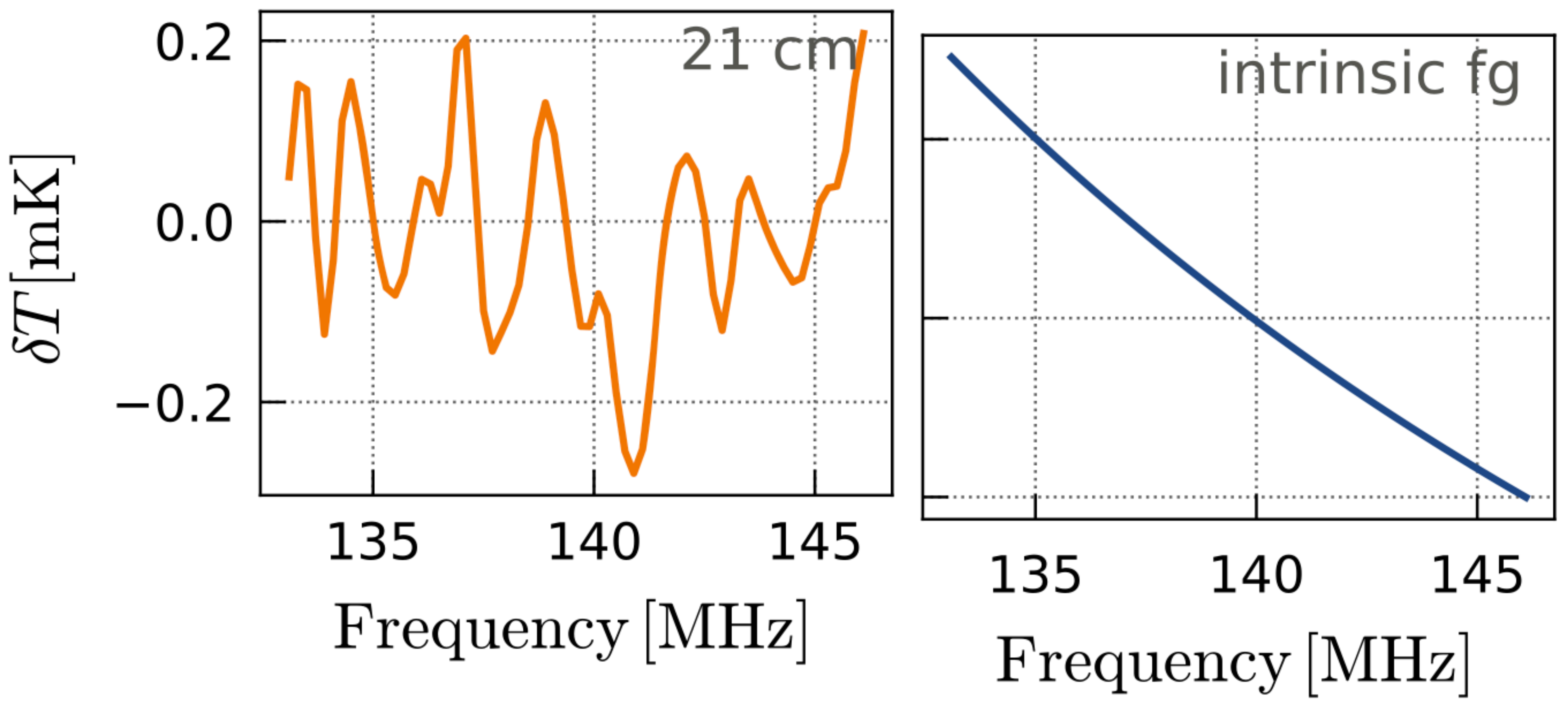
Mellema et al 2013

Foregrounds

- Observed signal
  -Foreground(FG) + 21cm line + Noise
- FG is brighter than EoR signal
  ($\sim 10^3$ in order)
- Avoidance or Removal of FG is important
  →How to remove foreground?
  →Use difference between FG and EoR signal
  -Emission strength(FG>>EoR signal)
  -Spectral behavior



Jelic et al 2008



Mertens et al 2018

Foreground Removal techniques

- There are various foreground removal techniques

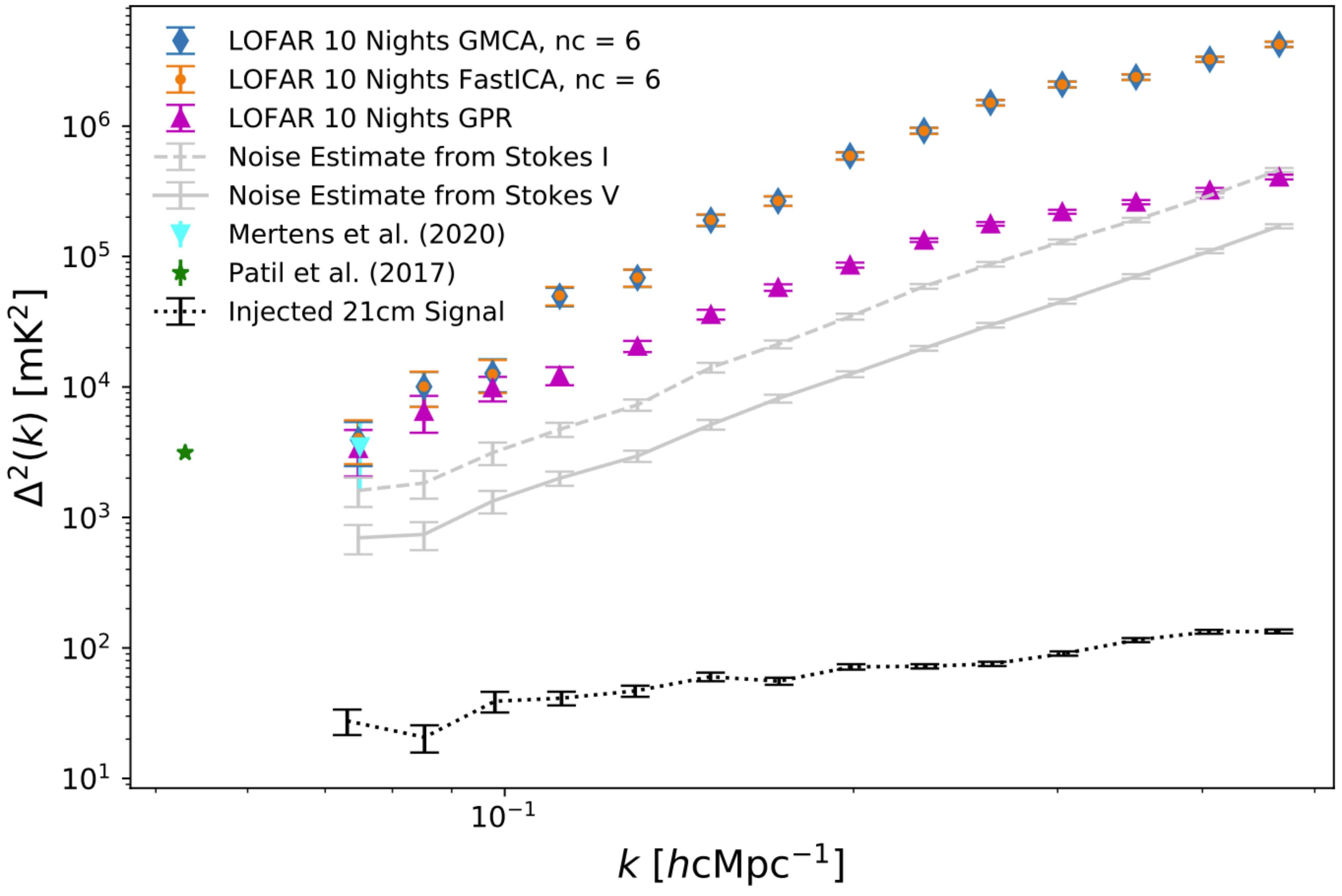
  -Generalized Morphological Component Analysis (GMCA)

  -FAST ICA

  -Principal Component Analysis (PCA)

  -Gaussian Process Regression (GPR)

- Hothi et al 2020 reports that GPR has better performance than FastICA,GMCA



My research

- GPR uses covariance called kernel to represent data

  →Best kernels set maybe different for each telescope

    →Compare some of the kernels combination by Bayesian evidence

- Imaging might be affect to FG removal

  →Visibility based FG removal(LOFAR analysis is image based)

I applied Gaussian Process Regression based FG removal to MWA observational data

Gaussian Process Regression(GPR)

- Our data consists of the foreground ($\mathbf{f}_{\mathrm{fg}}$),EoR signal ($\mathbf{f}_{21}$) and noise (n)

$$\mathbf{d} = \mathbf{f}_{\mathbf{fg}} + \mathbf{f}_{\mathbf{21}} + \mathbf{n}$$

- Assuming each component to be statistically uncorrelated,
  the covariance of the data $\mathbf{K}$ is given by

$$\mathbf{K} = \mathbf{K}_{\mathrm{fg}} + \mathbf{K}_{21} + \mathbf{K}_{\mathrm{noise}}$$

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)

Gaussian Process Regression(GPR)

- Assuming the data $\mathbf{d}$ is Gaussian distributed, we can model its probability distribution as

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

- We can write joint probability distribution of the GP at a series of other points in space ν' as

$$\begin{bmatrix} \mathrm{d} \\ \mathrm{d}' \end{bmatrix} \sim N\left(\begin{bmatrix} m(\nu) \\ m(\nu') \end{bmatrix}, \begin{bmatrix} \mathrm{K}(\nu, \nu) & \mathrm{K}(\nu, \nu') \\ \mathrm{K}(\nu', \nu) & \mathrm{K}(\nu', \nu') \end{bmatrix} \right)$$

Gaussian Process Regression(GPR)

- for the case of foreground removal, we want to estimate foreground model $\mathbf{f}_{\rm fg}$

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

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

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

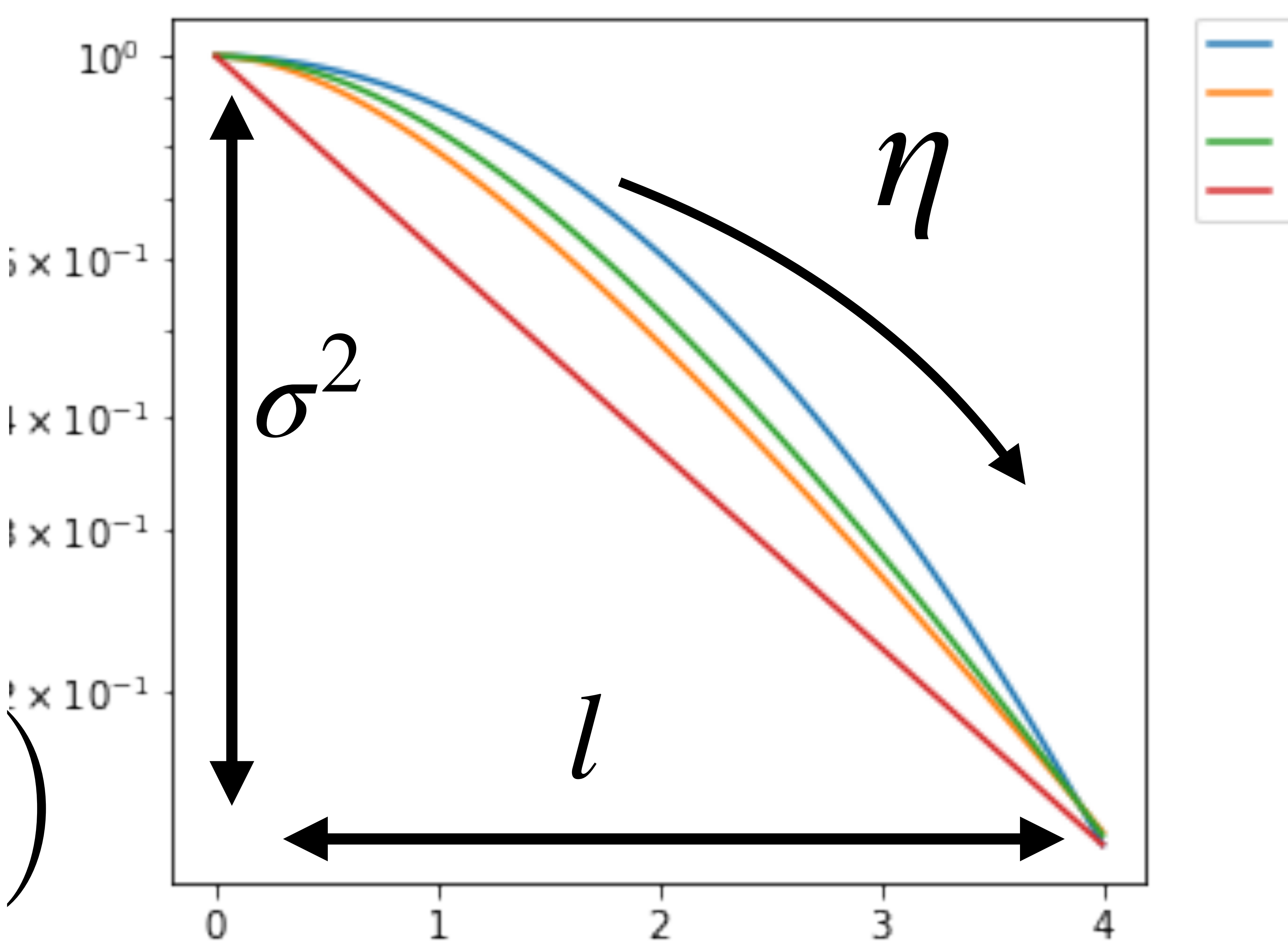
- To remove FG, subtract expected value of FG $\mathbf{E}[\mathbf{f}_{\rm fg}]$ from the data

$$\text{Residual} = \mathbf{d} - \mathrm{E}[\mathbf{f}_{\rm fg}]$$

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)

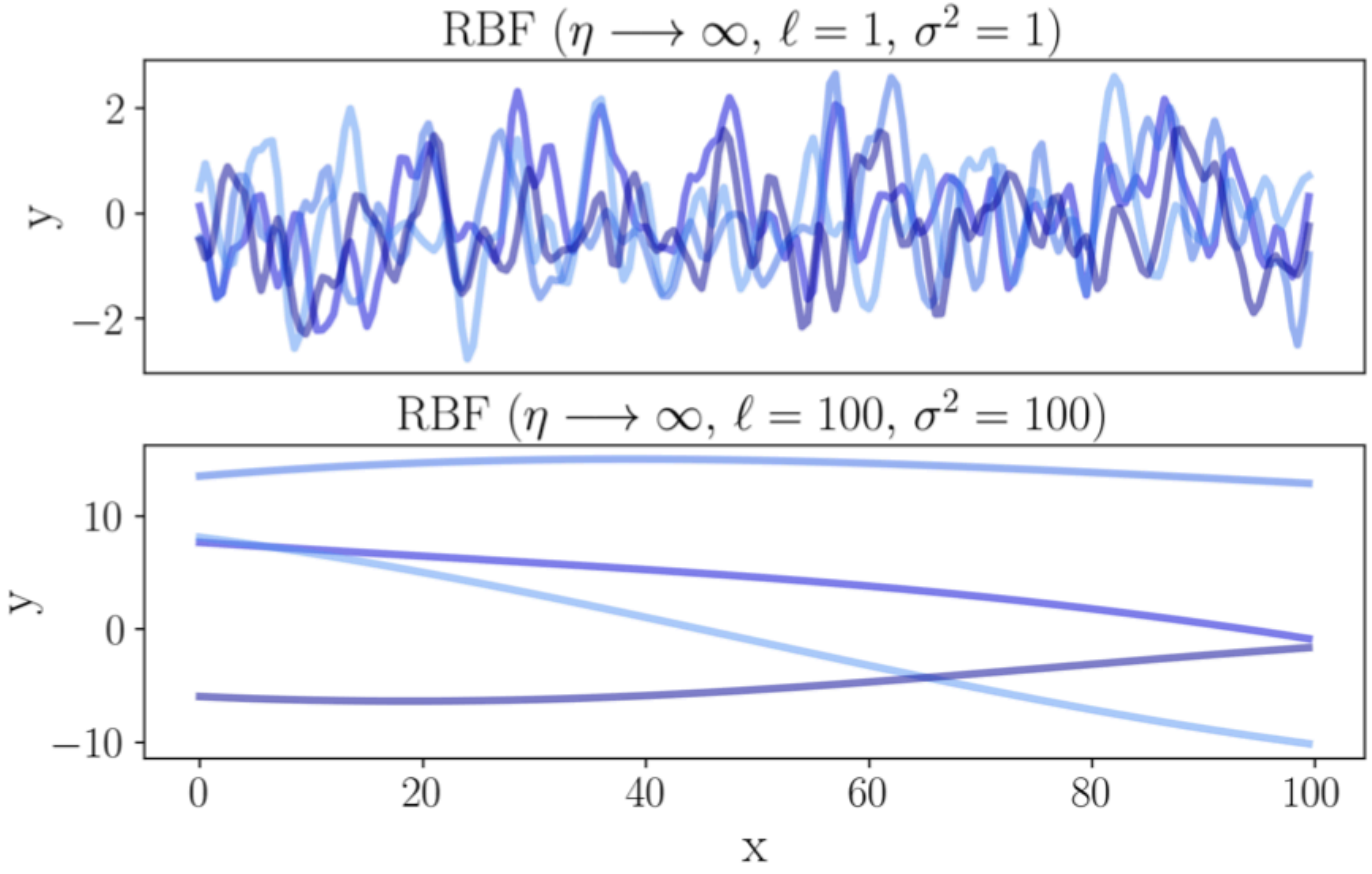
η:spectral parameter(It determines the overall "smoothness" of the data)

σ^2 and l of Matern kernel

- Larger σ^2 (Variance)
  → The signal is stronger.

- Larger l (coherence scale)
  → The signal is more correlated in frequency

randomly generated data plots with Matern kernels, with shown parameters

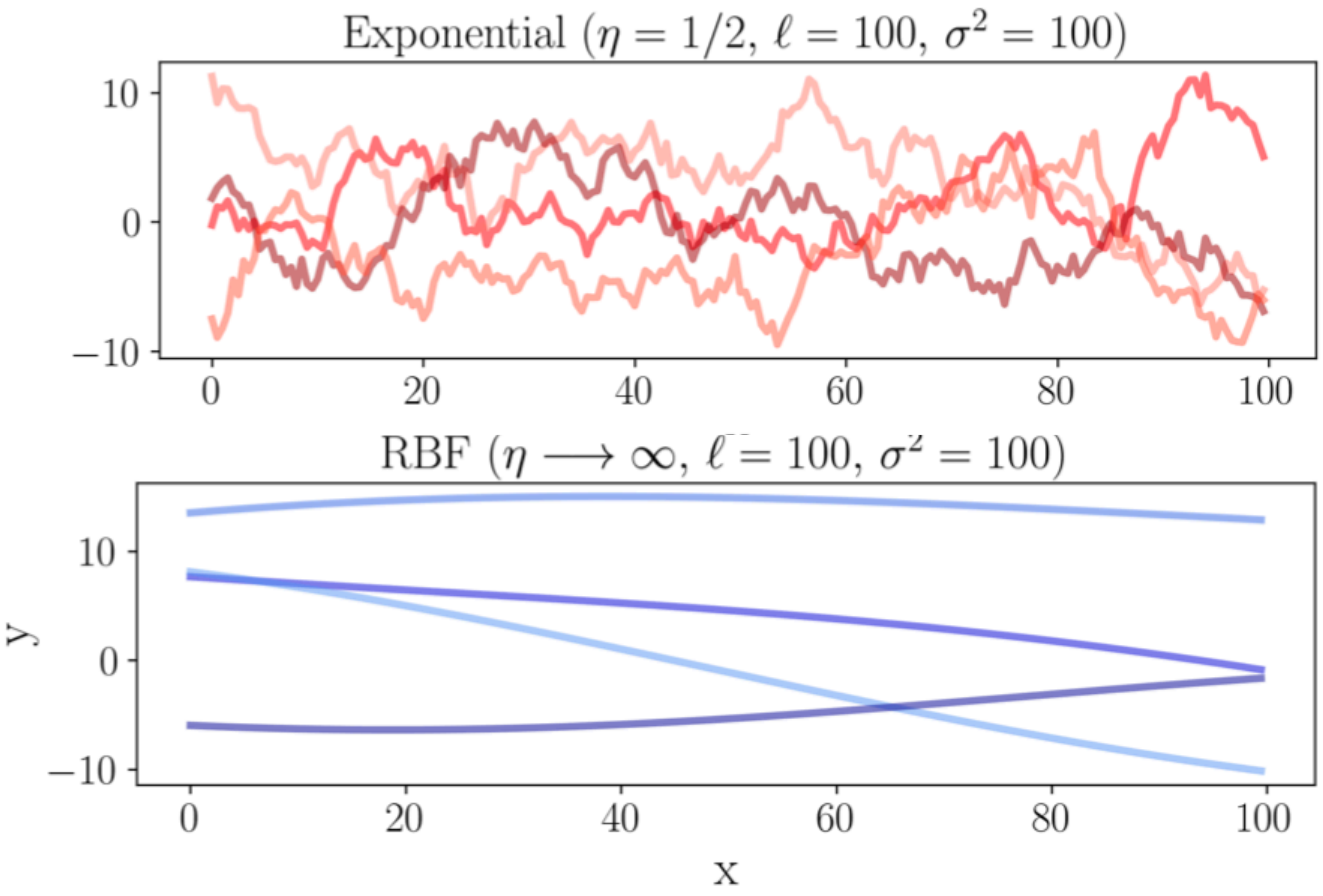


η of Matern kernel

- Larger η
  →data is spectrally smoother

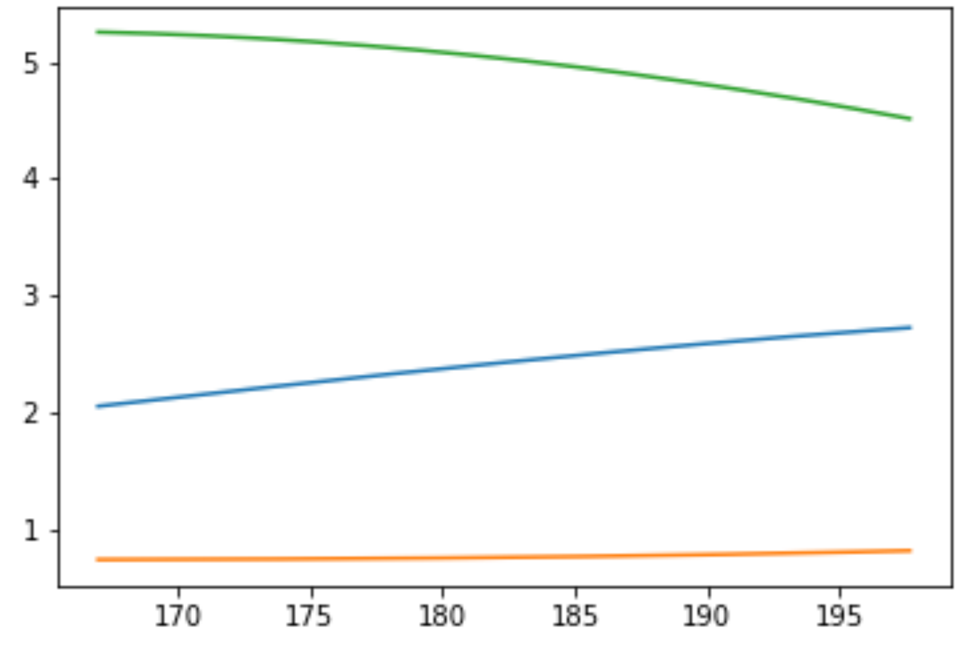
- FG is spectrally smoother!

randomly generated data plots with Matern kernels, with shown parameters

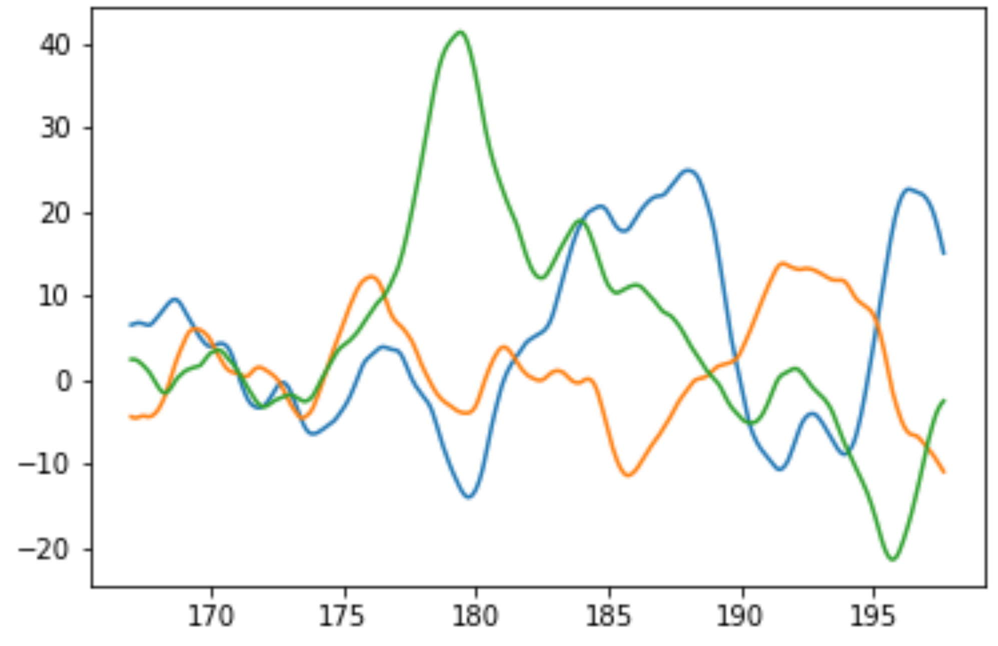


Components

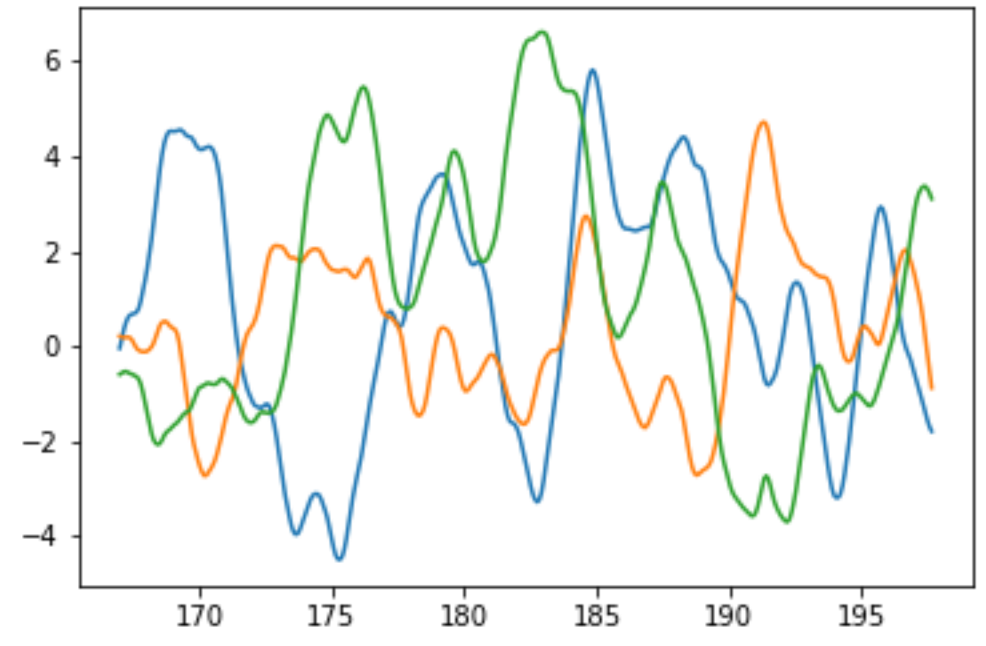
Smooth FG



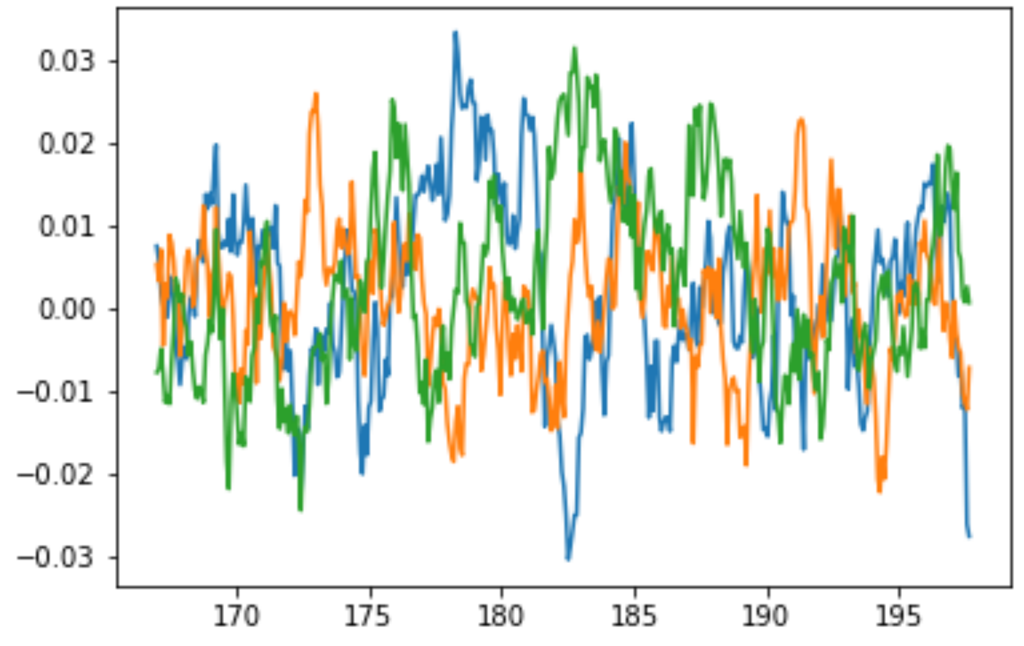
non-smooth FG



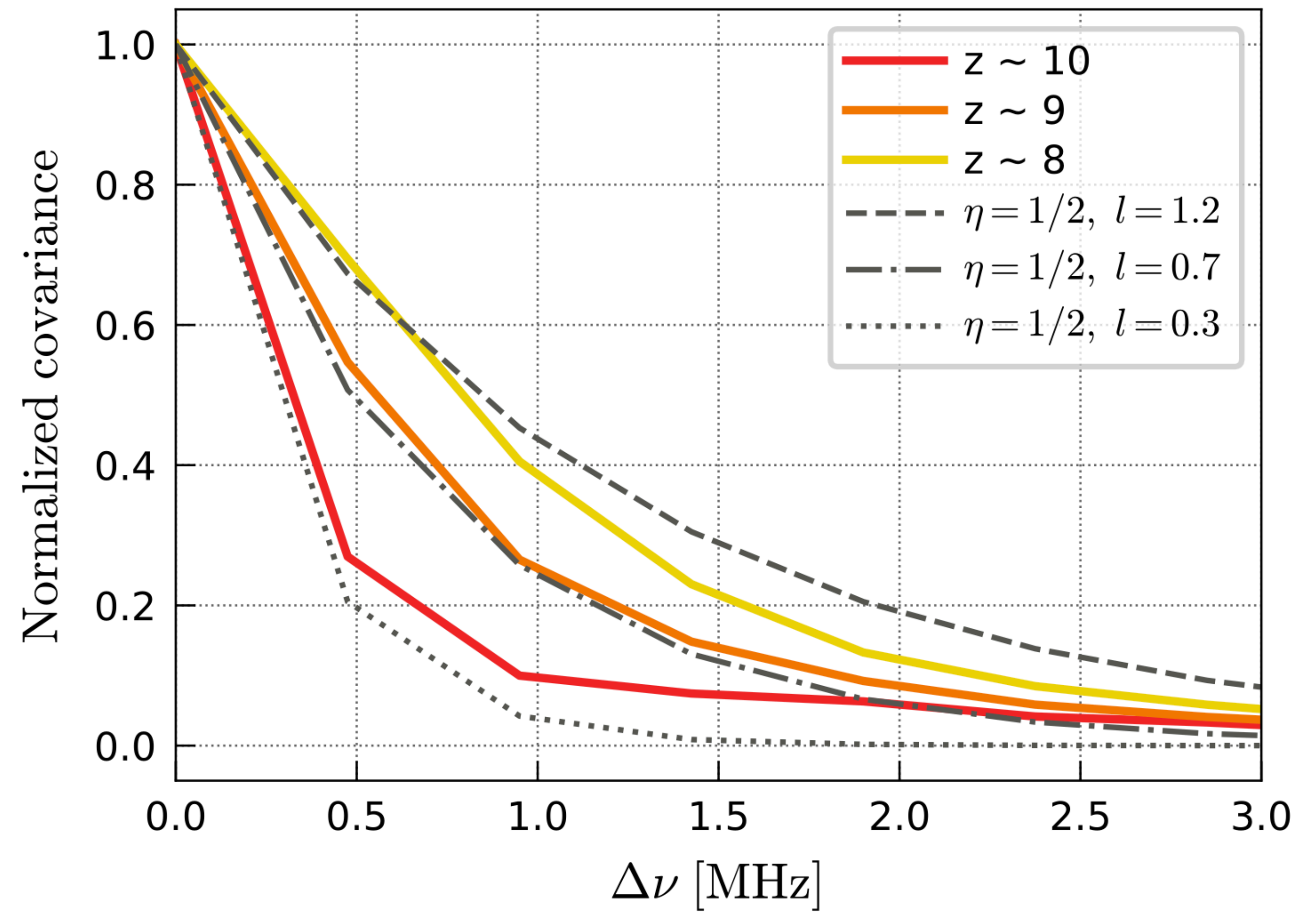
Additional component



HI



Mertens et al 2018



- Covariance of 21cm line

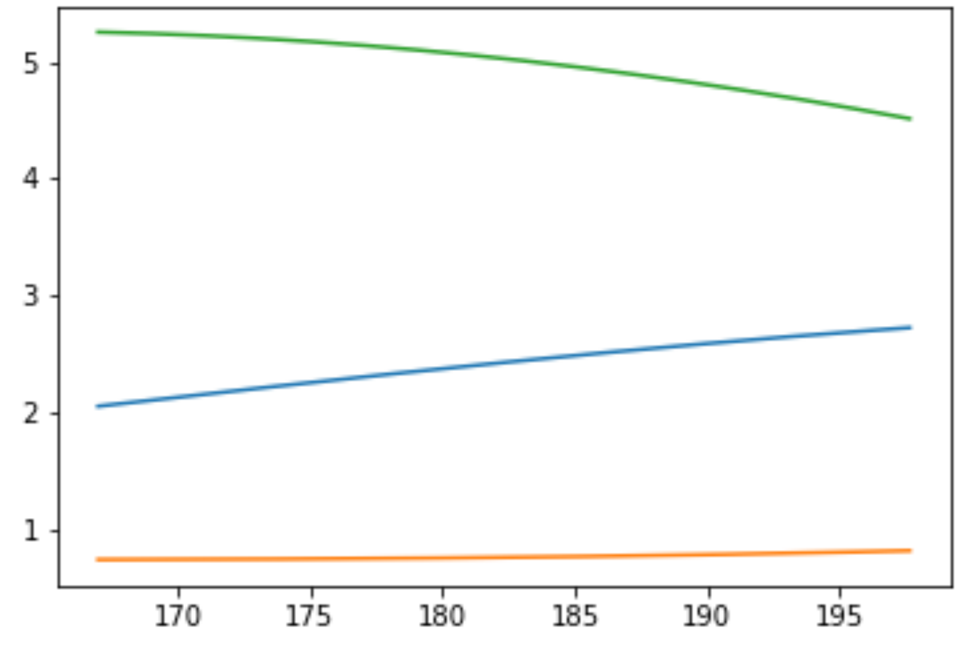
  →Well approximated by $\eta = 1/2$

- We don't know what are the best FG kernels for MWA!

  →Compare some kernels sets using Bayesian evidence

Applied kernel and its parameter priors

Smooth FG

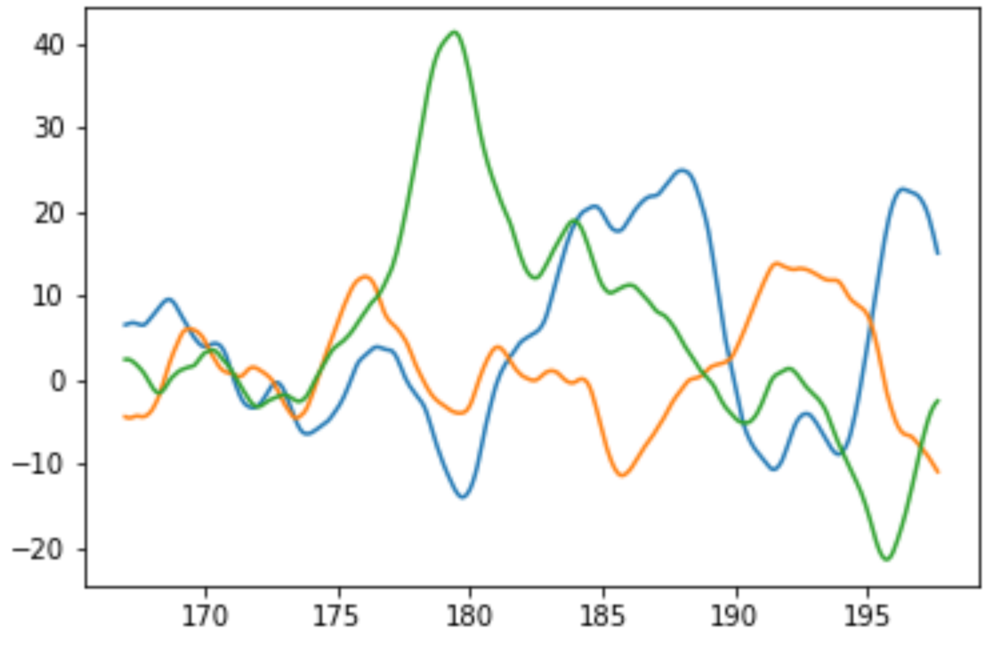


$\eta \to \infty$

$10 < l < 100[\mathrm{MHz}]$

—

non-smooth FG

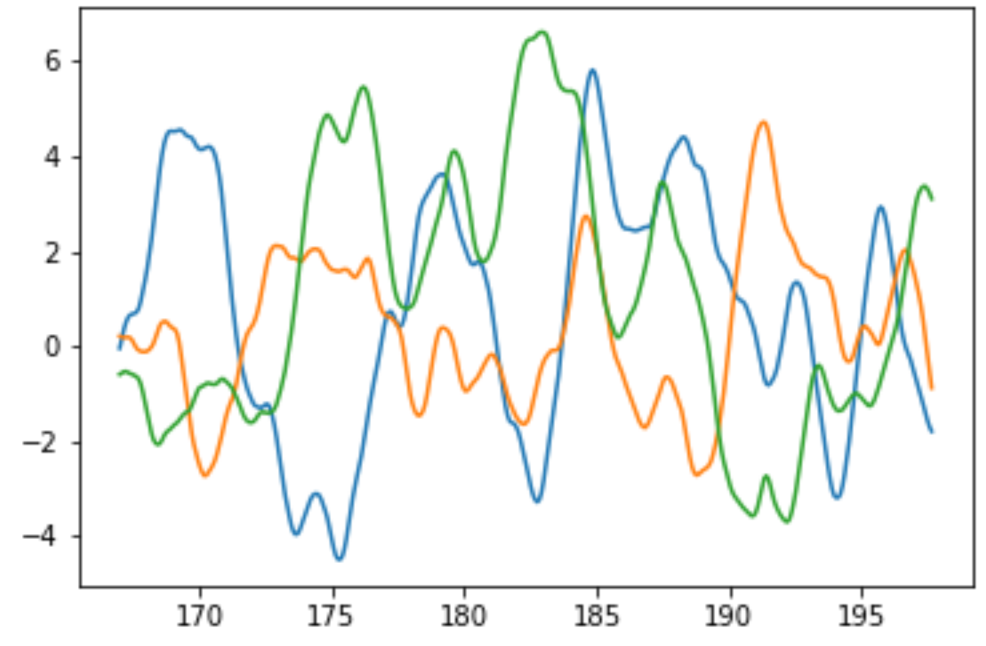


$\eta = 5/2$ or $3/2$

$1 < l < 10[\mathrm{MHz}]$

—

Additional component

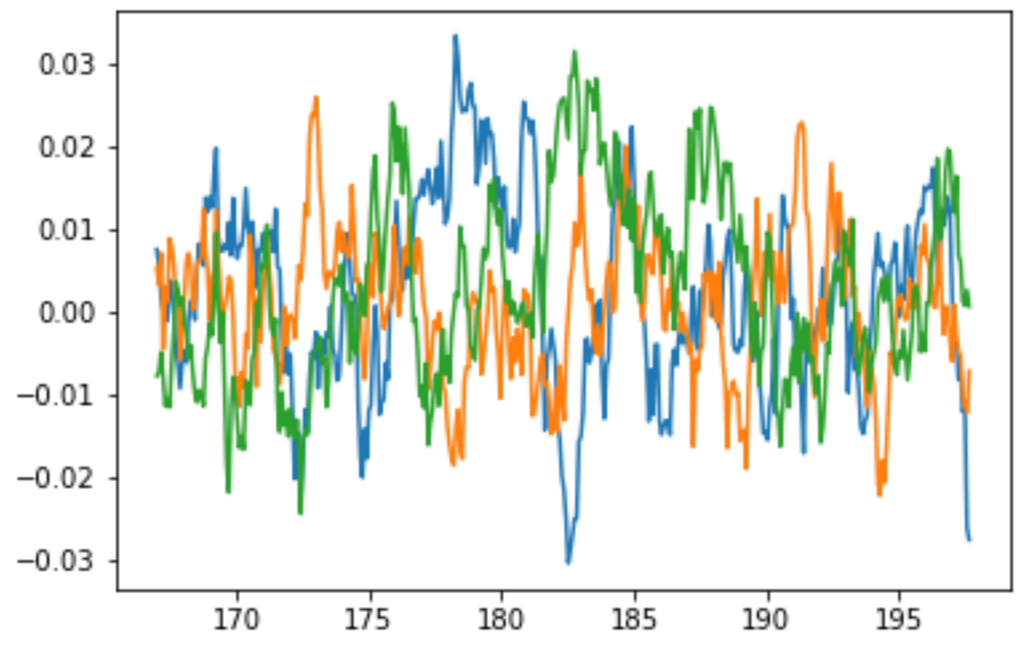


$\eta = 5/2$ or $3/2$

$0.64 < l < 1.92[\mathrm{MHz}]$

—

HI



$\eta = 1/2$

$0.1 < l < 1.2[\mathrm{MHz}]$

$10^{-6} < \sigma^2 < 0.5[\mathrm{Jy}^2]$

- Apply GPR to the data with each kernel set, and compare its evidence

  -MWA Observational data (high band observation in 2015 (2h@EoR0))

Result 1 ~Best kernel for MWA~

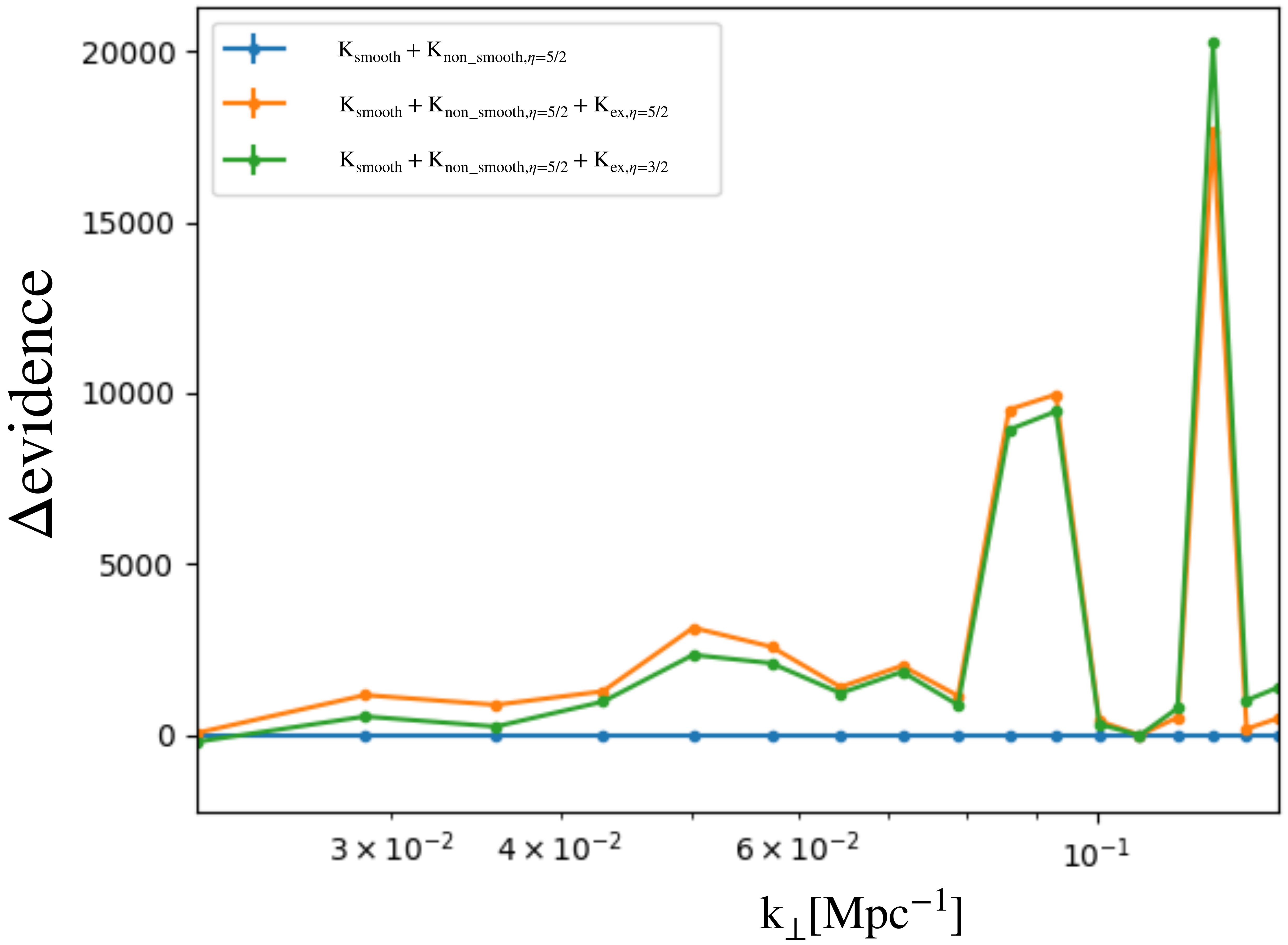
- Compare Bayesian Evidence of different kernel sets
- $\Delta\mathrm{evidence} = \log Z_{\mathrm{model}} - \log Z_{\mathrm{fiducial}}$

  - $|\ \Delta\mathrm{evidence}\ | > 10$:

    →Strong difference

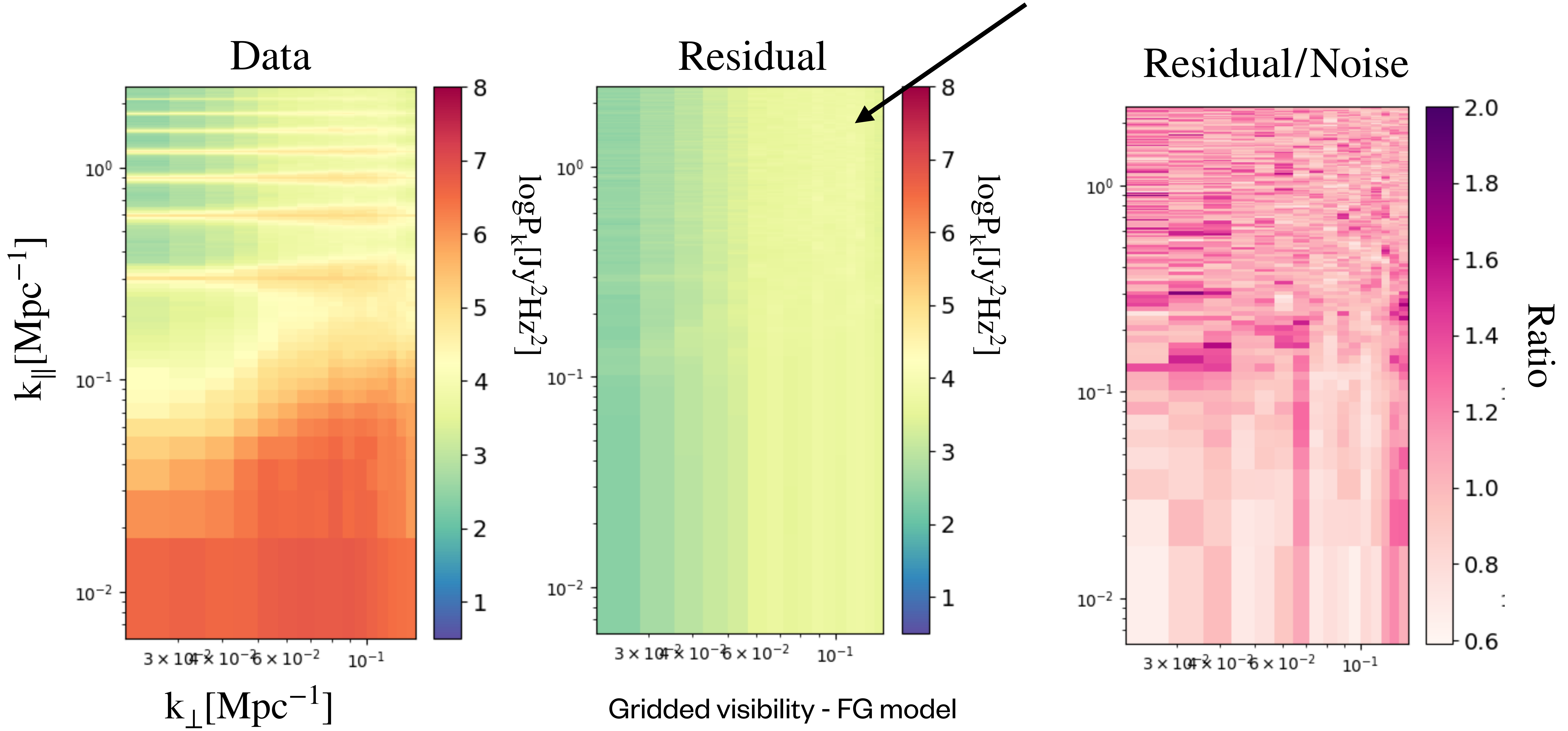
  -Higher Δ evidence

  → Better model
- Chosen kernels are depend on the value of $K_{\perp}$



Result 2~2D power spectrum~

Coarse band harmonics disappeared



Signal loss

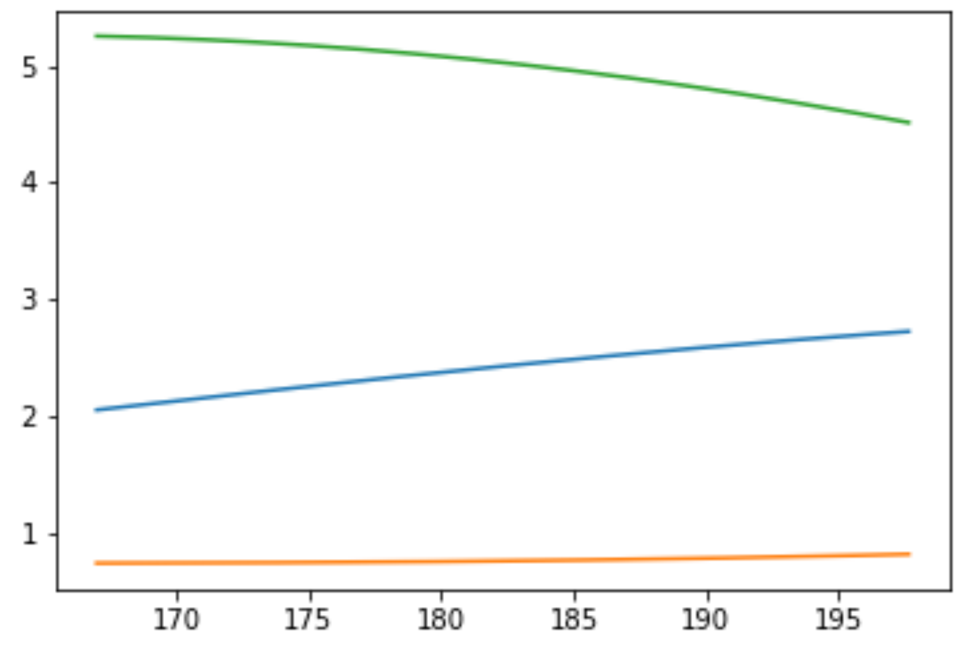
Additional component	HI
$0.64 < l < 1.92[\mathrm{MHz}]$	$0.1 < l < 1.2[\mathrm{MHz}]$

- Additional component has similar coherence scale to HI kernel
  →Subtracting additional component may cause signal loss

- Generate Multivariate Gaussian Distribution with K as signal component
  →Apply GPR based FG removal method to generated data and
  recover HI power spectrum

Results 3 ~signal loss~

K_{smooth}

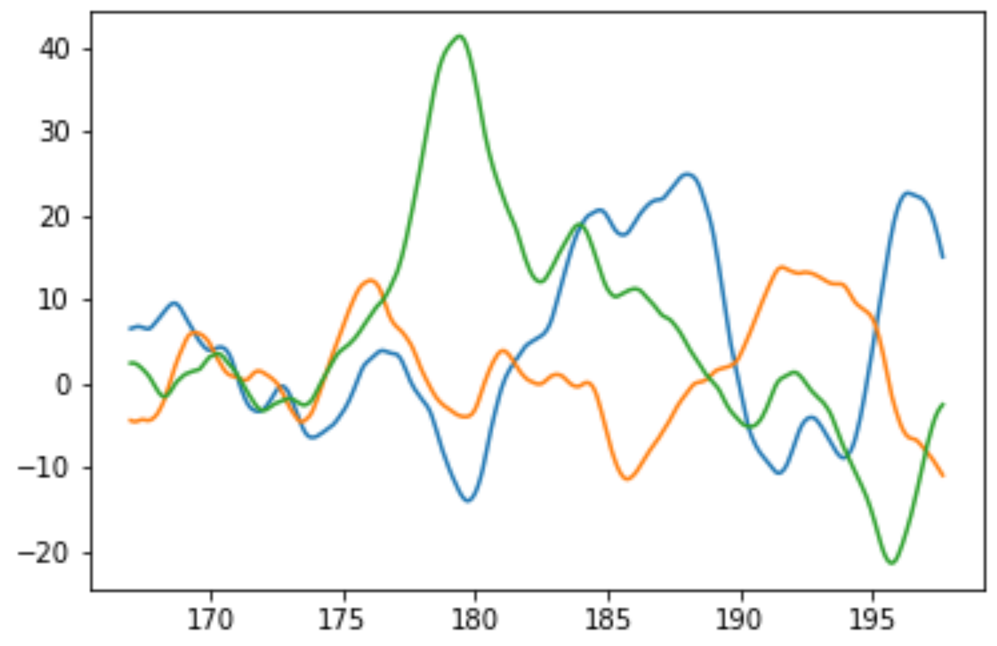


$\eta \to \infty$

$l = 100[\mathrm{MHz}]$

$\sigma^2 = 5[\mathrm{Jy}^2]$

$K_{\mathrm{non\_smooth},\eta=5/2}$

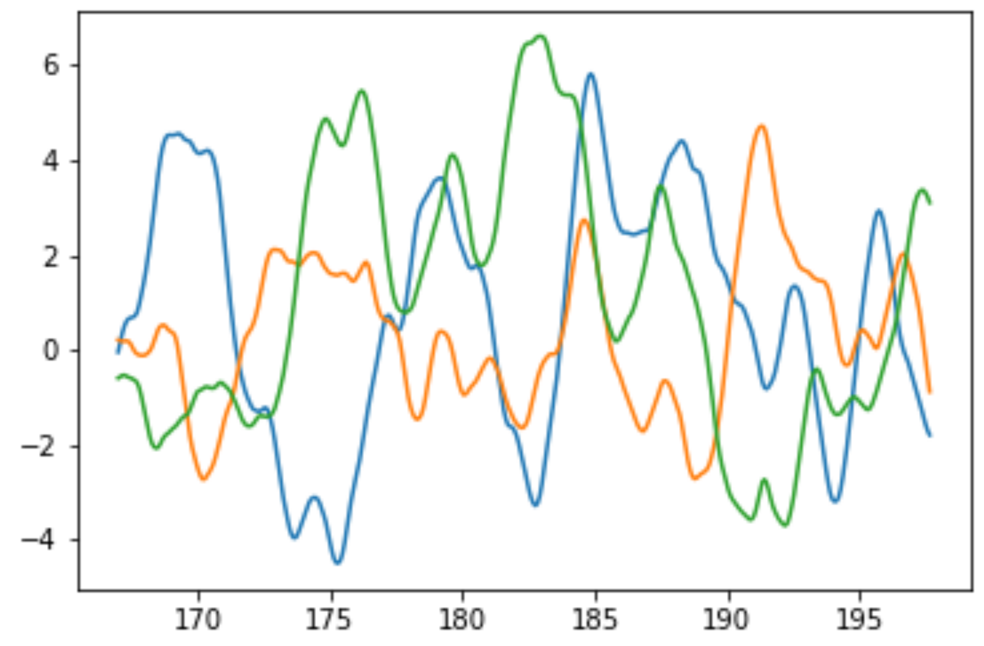


$\eta = 5/2$

$l = 5[\mathrm{MHz}]$

$\sigma^2 = 100[\mathrm{Jy}^2]$

$K_{\mathrm{add},\eta=5/2}$

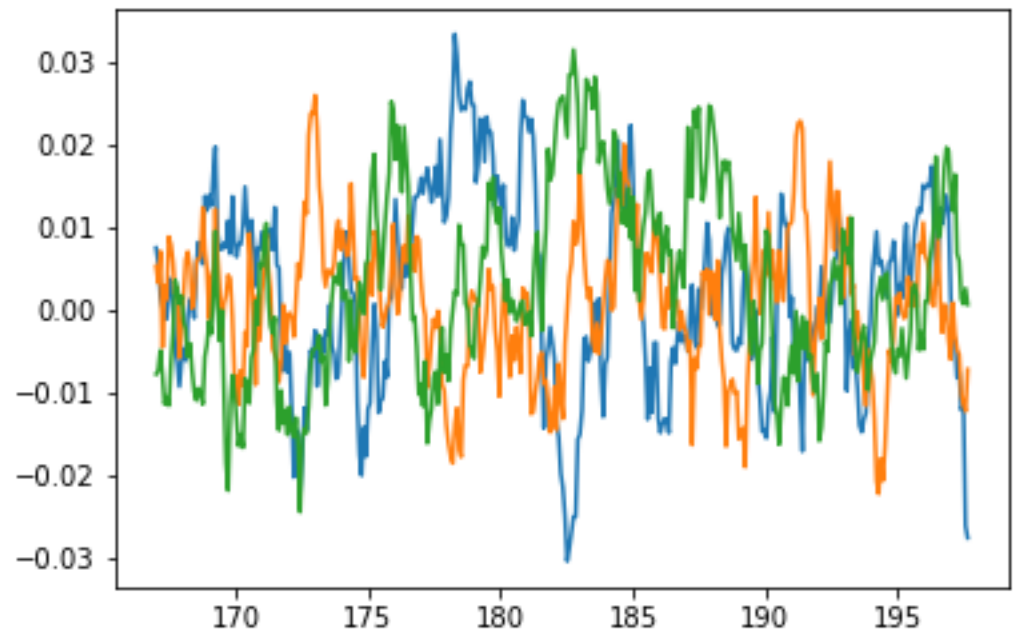


$\eta = 5/2$

$l = 1.28[\mathrm{MHz}]$

$\sigma^2 = 5[\mathrm{Jy}^2]$

K_{HI}



$\eta = 1/2$

$l = 0.75[\mathrm{MHz}]$

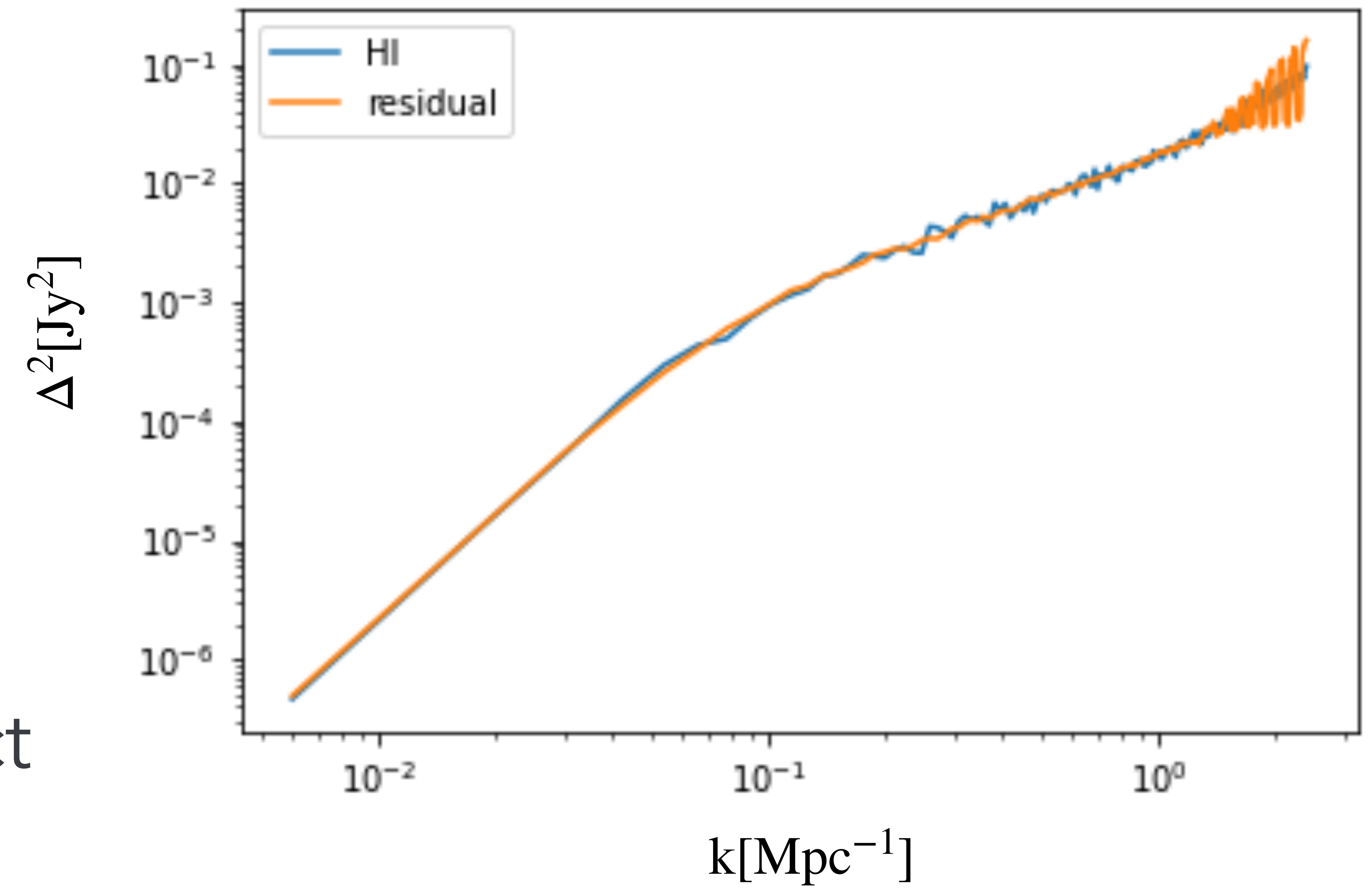
$\sigma^2 = 0.0001[\mathrm{Jy}^2]$

- Combine these signals together as simulated signal
  →GPR based FG removal

Results 3~Signal loss~

- HI power spectrum VS Residual

- If GPR work well
  -Ideal parameters are chosen
  -Signals are completely
  represented by kernel
  →No signal loss even if we subtract
  additional component



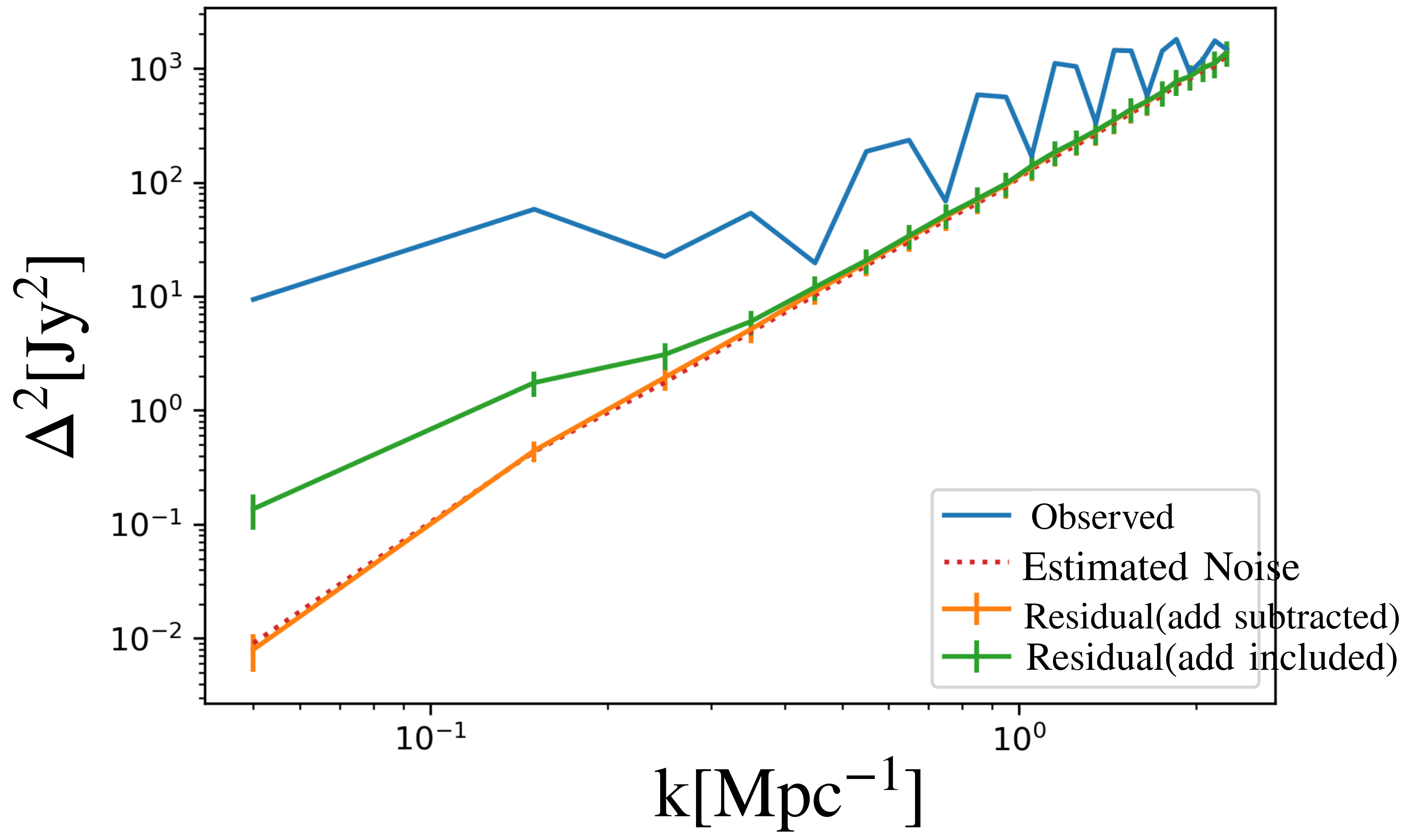
Conclusion

- I applied GPR based FG removal to MWA observational data(2h in 2015)

  →Kernel set with Additional power kernel has better Bayesian evidence

  →Residual is smaller than observed data

  - $\sim 10^2$ [Jy^2] (additional component included)
  - $\sim 10^3$ [Jy^2] (additional component subtracted)



- Maybe no signal loss even if we subtract additional component