Estimation of Bias and Excess Variance of 21-cm Power Spectrum Due to Residual Gain Errors in Presence of Strong Foreground.

Saikat Gayen

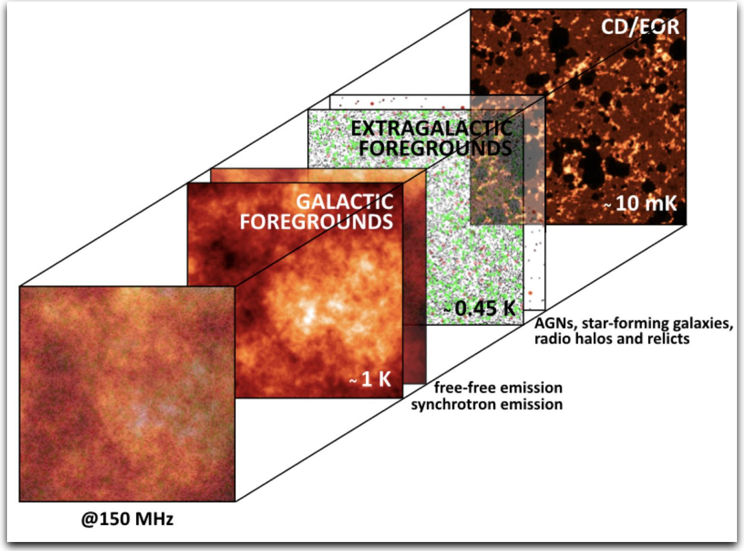
भारतीय प्रौद्योगिकी संस्थान काशी हिन्दू विश्वविद्यालय
INDIAN INSTITUTE OF TECHNOLOGY
BANARAS HINDU UNIVERSITY

Collaborators: Dr. Prasun Dutta, Dr. Abhirup Datta, Dr. Nirupam Roy, Dr. Samir Choudhuri, Dr. Jais Kumar, Mr. Sarvesh Mangla, Ms. Rashmi Sagar

Challenges of 21-cm Tomography

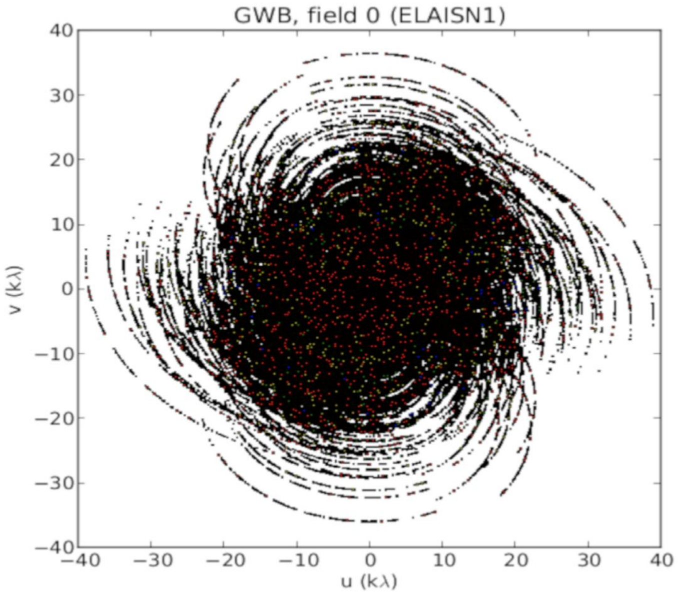
- In present resolution and sensitivity, 21-cm signal extraction using tomography is not possible for EoR and post EoR.
- Challenges: Strong foregrounds, Shortage of baseline coverage.
- Statistical analysis of fluctuation of redshifted 21-cm signal hold the key.
- **Study of 21-cm signal power spectrum.**

Illustration of different foreground components



Chapman et al. 2019

Baseline coverage of GMRT GWB data

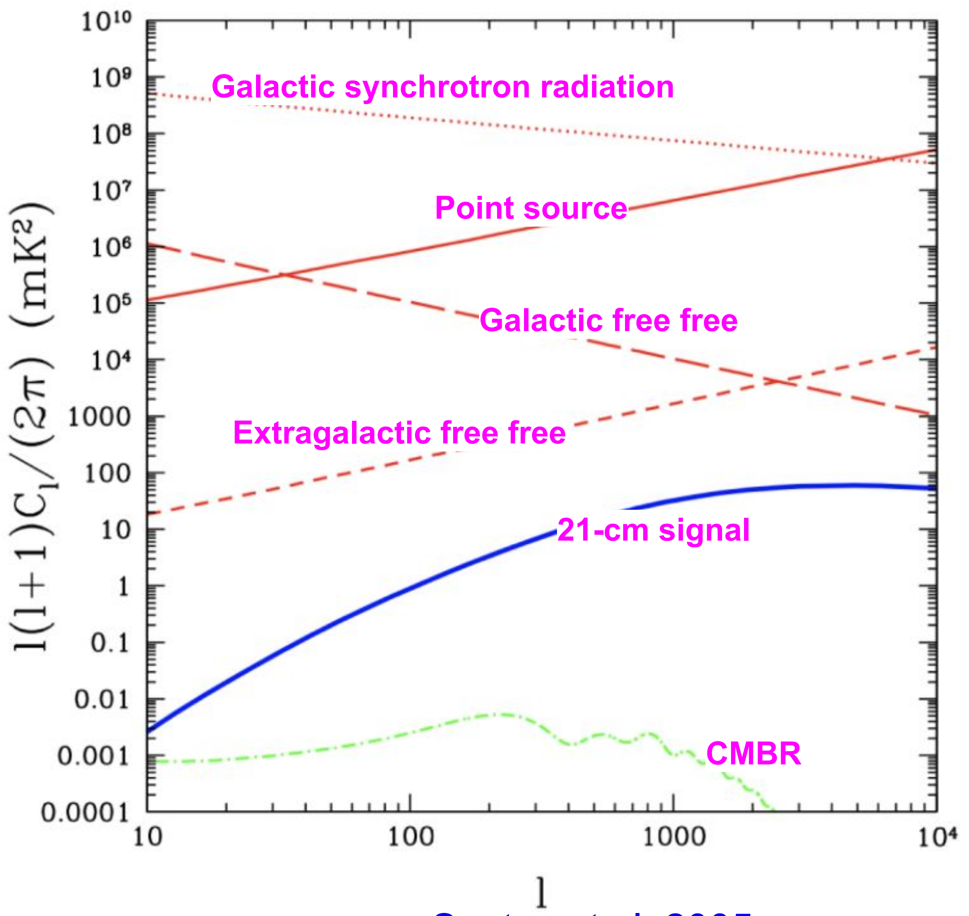


Chakraborty et al. 2019(b)

Foreground

- 21- cm power power spectrum is buried in foreground radiation from other astrophysical sources whose contribution is seven to eight orders of magnitude larger.
- Galactic synchrotron radiation is most dominant about 70%
- We should have good model of foreground.

Foreground comparison with HI power spectrum at 140 MHz



Santos et al. 2005

Foreground Mitigation

- To extract 21-cm signal power spectrum, somehow we have to remove foreground.

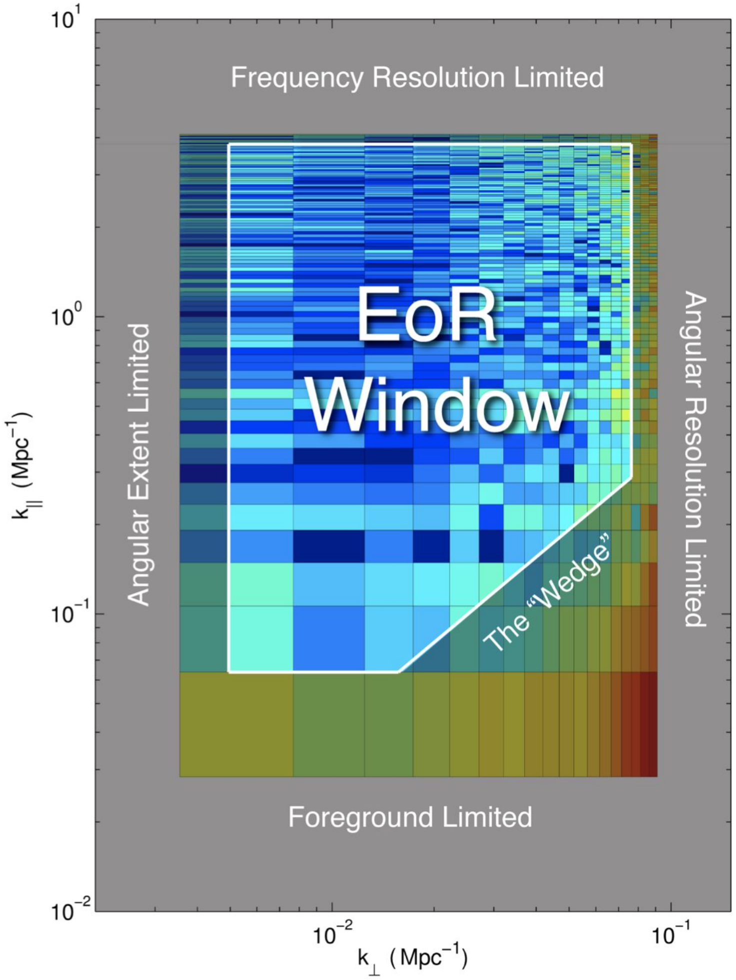
Foreground removal:

- Bright sources removal
- Foreground avoidance
- Foreground subtraction

EoR window:

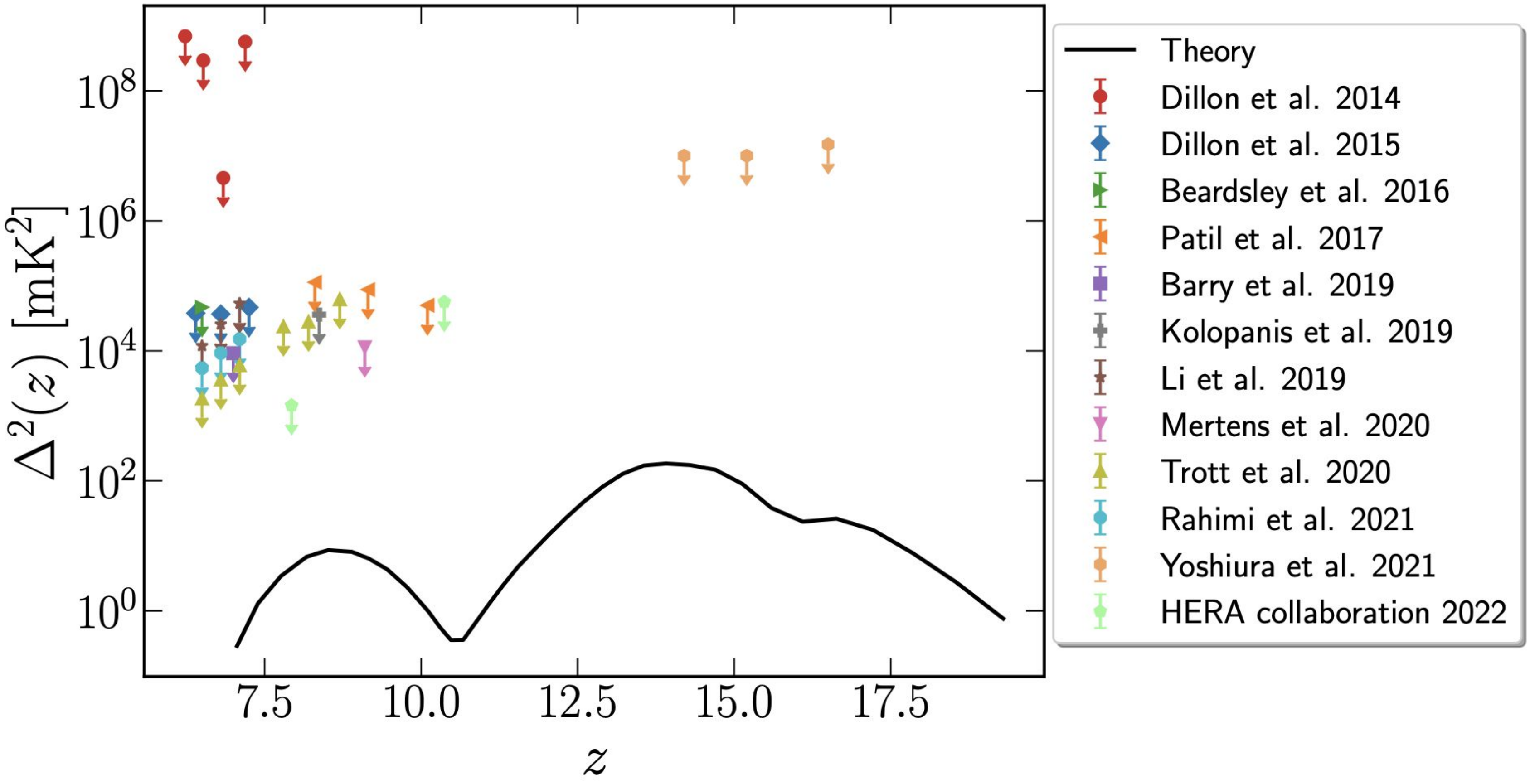
- In a perfect observation, with zero instrumental effects, the foregrounds would be entirely contained in the well defined horizontal band.
- In a realistic observation however, the chromaticity of the instrument results in a leakage of power up into the EoR window, into a region called the 'wedge'.

A schematic of EoR window



Dillon et al. 2014

Power Spectrum Upper Limit



Shaw et al. 2022

Visibility-Gain-Calibration

$$Visibility = Gain * \langle E_A^* E_B \rangle + Noise$$

$$\tilde{V}(\vec{U}_i) = \tilde{G}_i(t)\tilde{V}_i^S(\vec{U}_i) + \tilde{N}_i(\vec{U}_i)$$

Recorded visibility | Gain for i'th baseline | Sky visibility | Measurement noise

$$\tilde{G}_i = \langle \tilde{g}_A(t)\tilde{g}_B^*(t) \rangle = 1 + \tilde{G}_i^R$$

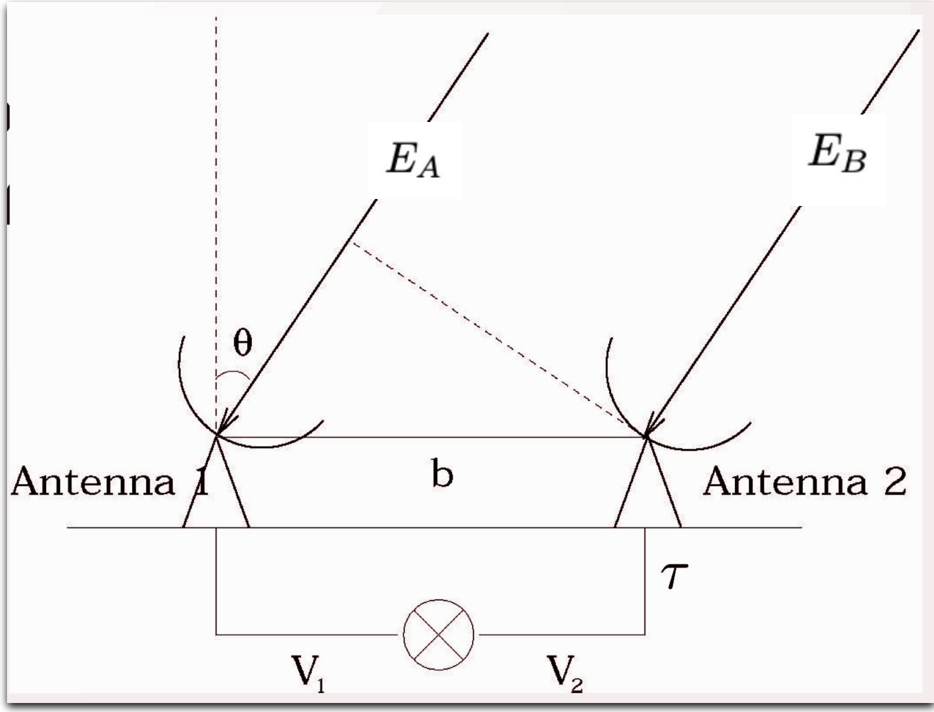
Gain for i'th baseline | Gain for antenna A | Gain for antenna B | **Residual gain**

- Residual gain for each antenna contribute to the gain term.

$$\tilde{g}_A(t) = [1 + \delta_{AR}(t) + i\delta_{AI}(t)]$$

Kumar et al. 2022

Visibility Measurement



Variance of power spectrum due to Noise vs due to residual gain errors

$$\sigma_P^2 = \frac{P_{HI}^2}{N_G} + \frac{2\sigma_N^2 P_{HI}}{N_B N_d} + \frac{2\sigma_N^4}{N_B N_d^2}$$

Due to Cosmic Variance

Due to thermal noise

Ali et al. 2008

- If there is no calibration error, power spectrum estimate has no bias and uncertainties can be written as above equation.
- First term is due to cosmic variance , third term is due to thermal noise and middle term is negligible as it is multiplied by 21-cm power spectrum.
- In presence of residual gain error, there will be bias in power spectrum and excess variance that will depends on baseline and there will be time and frequency correlation effect.

Gain Error Model

Residual Gain

$$\tilde{g}_A(t) = [1 + \delta_{AR}(t) + i\delta_{AI}(t)]$$

Standard Deviation

$$\sigma_{AC}^2 = \langle \delta_{AC}^2 \rangle$$

Correlation Function

$$\xi_{AC}(\tau) = \langle \delta_{AC}(t)\delta_{AC}(t+\tau) \rangle / \sigma_{AC}^2$$

Bias

$$\mathcal{B}_P = \left[(2n_1 + n_3) \frac{\chi_R\, \sigma_R^2 + \chi_I\, \sigma_I^2}{\sigma_R^2 + \sigma_I^2} + n_2 \right] \frac{\sigma_R^2 + \sigma_I^2}{N_d} C_\ell \,|_F$$

Variance

$$\sigma_P^2 = \left[\frac{C_\ell^2|_{HI}}{N_G} + \frac{2\sigma_N^2 C_\ell|_{HI}}{N_B N_d} + \frac{2\sigma_N^4}{N_B N_d^2} \right] + \left[\frac{4\sigma_N^2 (\sigma_R^2 + \sigma_I^2) C_\ell|_F}{N_B N_d^2} \right]$$
$$+ \left[\left[(4n_1^2 + n_3^2) \left(\frac{\chi_R \sigma_R^2 + \chi_I \sigma_I^2}{\sigma_R^2 + \sigma_I^2} \right)^2 + n_2^2 \right] \left[3\sigma_R^4 + 3\sigma_I^4 + 2\sigma_R^2 \sigma_I^2 \right] + 8(\sigma_R^4 + \sigma_I^4) \right] \frac{C_\ell|_F^2}{N_G N_d^2},$$

Kumar et al. 2022

Effect of Gain Error Model

Variance

Excess variance without residual gain errors

$$\sigma_P^2 = \left[\frac{C_\ell^2|_{HI}}{N_G} + \frac{2\sigma_N^2 C_\ell|_{HI}}{N_B N_d} + \frac{2\sigma_N^4}{N_B N_d^2}\right] + \left[\frac{4\sigma_N^2(\sigma_R^2+\sigma_I^2)C_\ell|_F}{N_B N_d^2}\right]$$

$$+ \Bigg[\left[(4n_1^2 + n_3^2)\left(\frac{\chi_R\sigma_R^2+\chi_I\sigma_I^2}{\sigma_R^2+\sigma_I^2}\right)^2 + n_2^2\right]\left[3\sigma_R^4 + 3\sigma_I^4 + 2\sigma_R^2\sigma_I^2\right]$$

$$+8(\sigma_R^4 + \sigma_I^4)\Bigg]\frac{C_\ell|_F^2}{N_G N_d^2},$$

Kumar et al. 2022

Excess variance due to residual gain errors.

- When there is no calibration errors , due to instruments noise we get only first term.
- In presence of residual calibration errors there will be additional two terms.

Gain Error Model

Residual Gain

$$\tilde{g}_A(t) = [1 + \delta_{AR}(t) + i\delta_{AI}(t)]$$

Standard Deviation

$$\sigma^2_{AC} = \langle \delta^2_{AC} \rangle$$

Correlation Function

$$\xi_{AC}(\tau) = \langle \delta_{AC}(t)\delta_{AC}(t+\tau) \rangle / \sigma^2_{AC}$$

Bias

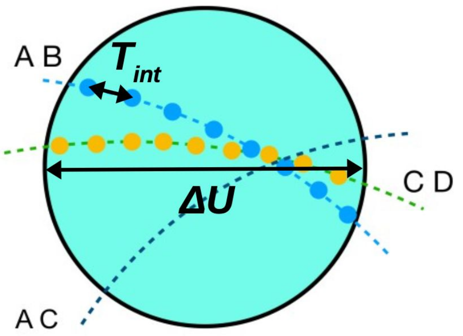
$$\mathcal{B}_P = \left[(2n_1 + n_3) \frac{\chi_R\, \sigma^2_R + \chi_I\, \sigma^2_I}{\sigma^2_R + \sigma^2_I} + n_2 \right] \frac{\sigma^2_R + \sigma^2_I}{N_d} C_\ell\, |_F$$

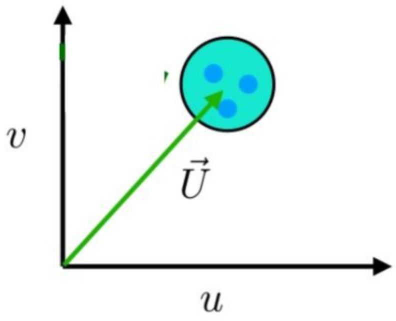
Variance

$$\sigma^2_P = \left[\frac{C^2_\ell|_{HI}}{N_G} + \frac{2\sigma^2_N C_\ell|_{HI}}{N_B N_d} + \frac{2\sigma^4_N}{N_B N^2_d} \right] + \left[\frac{4\sigma^2_N (\sigma^2_R + \sigma^2_I) C_\ell|_F}{N_B N^2_d} \right]$$
$$+ \Bigg[\left[(4n^2_1 + n^2_3) \left(\frac{\chi_R \sigma^2_R + \chi_I \sigma^2_I}{\sigma^2_R + \sigma^2_I} \right)^2 + n^2_2 \right] [3\sigma^4_R + 3\sigma^4_I + 2\sigma^2_R \sigma^2_I]$$
$$+ 8(\sigma^4_R + \sigma^4_I) \Bigg] \frac{C_\ell|^2_F}{N_G N^2_d},$$

Kumar et al. 2022

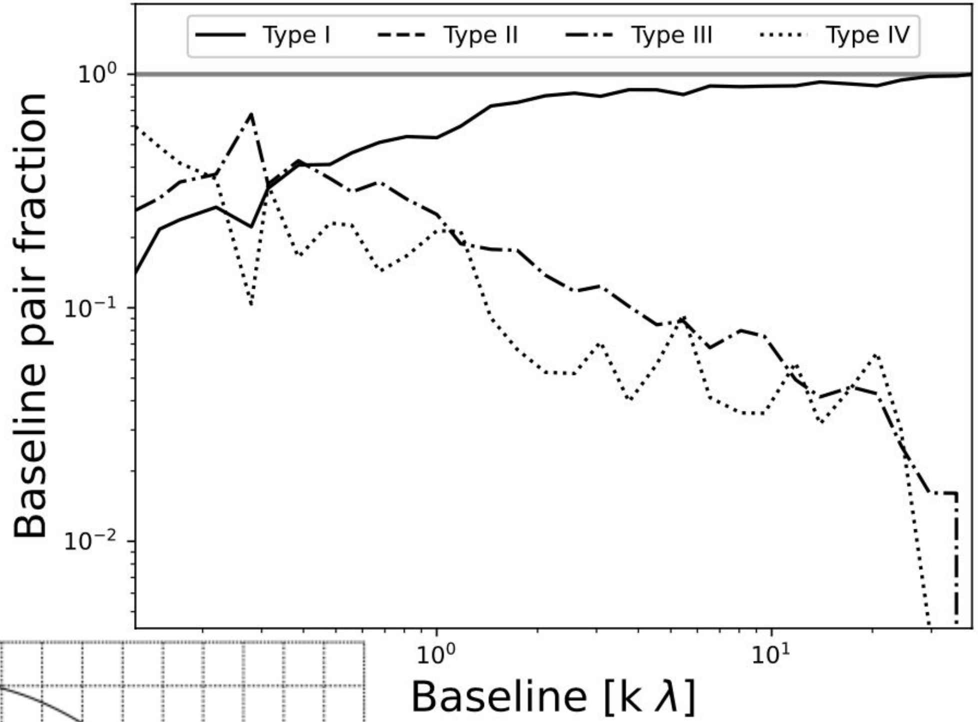
Baseline Pair Fractions

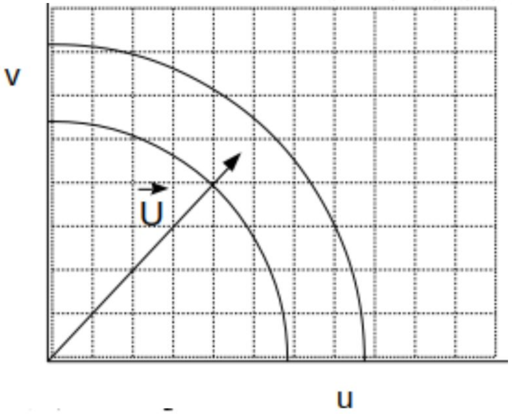




Type I	$n_I(\vec{U})$	$V^*_{AB}(\vec{U},t)\ V_{AB}(\vec{U}+\Delta\vec{U},t')$
Type II	$n_{II}(\vec{U})$	$V^*_{AB}(\vec{U},t)\ V_{AC}(\vec{U}+\Delta\vec{U},t)$
Type III	$n_{III}(\vec{U})$	$V^*_{AB}(\vec{U},t)\ V_{AC}(\vec{U}+\Delta\vec{U},t')$
Type IV	$n_{IV}(\vec{U})$	$V^*_{AB}(\vec{U},t)\ V_{CD}(\vec{U}+\Delta\vec{U},t')$

Kumar et al. 2022





$$\Delta U < \frac{1}{\pi\theta_0}$$

Gain Error Model

Residual Gain

$$\tilde{g}_A(t) = [1 + \delta_{AR}(t) + i\delta_{AI}(t)]$$

Standard Deviation

$$\sigma^2_{AC} = \langle \delta^2_{AC} \rangle$$

Correlation Function

$$\xi_{AC}(\tau) = \langle \delta_{AC}(t)\delta_{AC}(t+\tau)\rangle / \sigma^2_{AC}$$

Bias

$$\mathcal{B}_P = \left[(2n_1 + n_3) \frac{\chi_R \sigma_R^2 + \chi_I \sigma_I^2}{\sigma_R^2 + \sigma_I^2} + n_2 \right] \frac{\sigma_R^2 + \sigma_I^2}{N_d} C_\ell \,|_F$$

Variance

$$\begin{aligned} \sigma_P^2 \quad &= \left[\frac{C_\ell^2|_{HI}}{N_G} + \frac{2\sigma_N^2 C_\ell|_{HI}}{N_B N_d} + \frac{2\sigma_N^4}{N_B N_d^2} \right] + \left[\frac{4\sigma_N^2(\sigma_R^2+\sigma_I^2)C_\ell|_F}{N_B N_d^2} \right] \\ &+ \Bigg[\left[(4n_1^2 + n_3^2)\left(\frac{\chi_R\sigma_R^2 + \chi_I\sigma_I^2}{\sigma_R^2+\sigma_I^2}\right)^2 + n_2^2 \right] \left[3\sigma_R^4 + 3\sigma_I^4 + 2\sigma_R^2\sigma_I^2 \right] \\ &+8(\sigma_R^4 + \sigma_I^4) \Bigg] \frac{C_\ell|_F^2}{N_G N_d^2}, \end{aligned}$$

Kumar et al. 2022

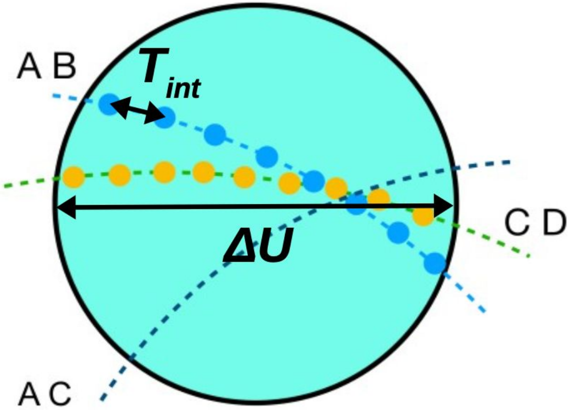
Effect of Time Correlation Function

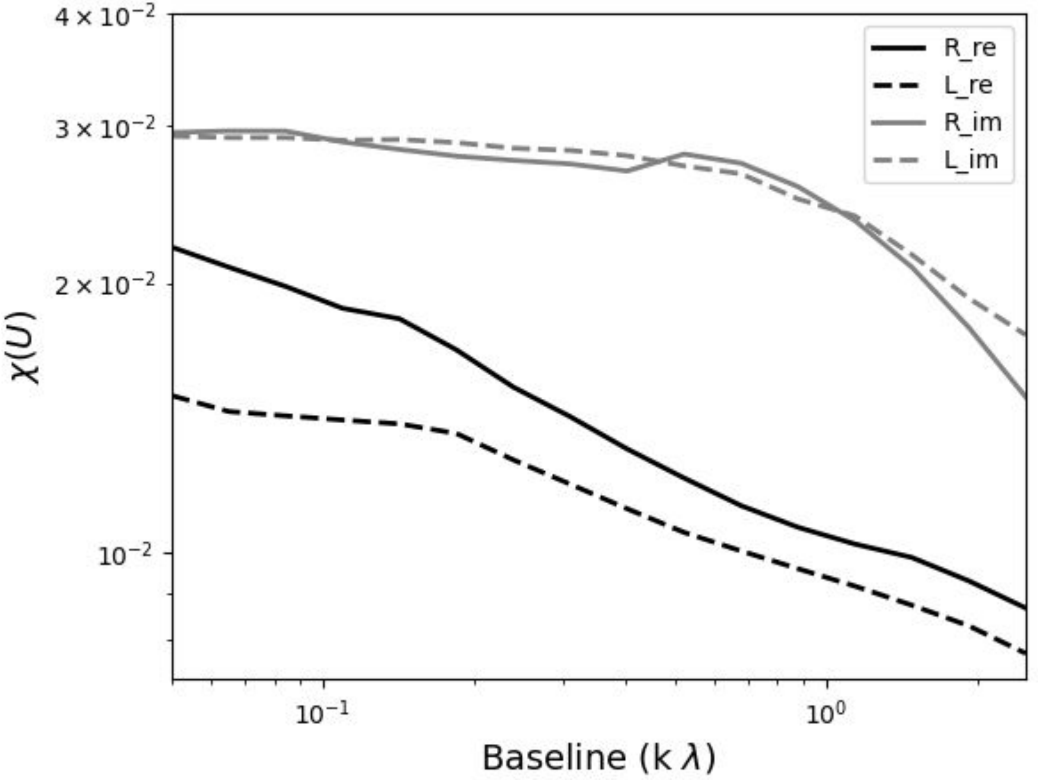
$$\xi_{AC}(\tau) = \langle \delta_{AC}(t)\delta_{AC}(t+\tau)\rangle / \sigma^2_{AC}$$

$$\chi(U) = \frac{1}{T_D^2}\int_{\Delta\tau}^{T_D} (T_D - \tau)\xi(\tau)d\tau$$

$$T_D = \frac{\Delta U \times T_{24}}{2\pi U}$$

Kumar et al. 2022





Assumptions and Simplifications

1. The gain errors are Gaussian random variables.
2. Antenna gains from different antennas are uncorrelated.
3. The real and imaginary parts of the residual gain errors are uncorrelated.
4. The electronic noise in different baselines are uncorrelated.

Also we consider following..

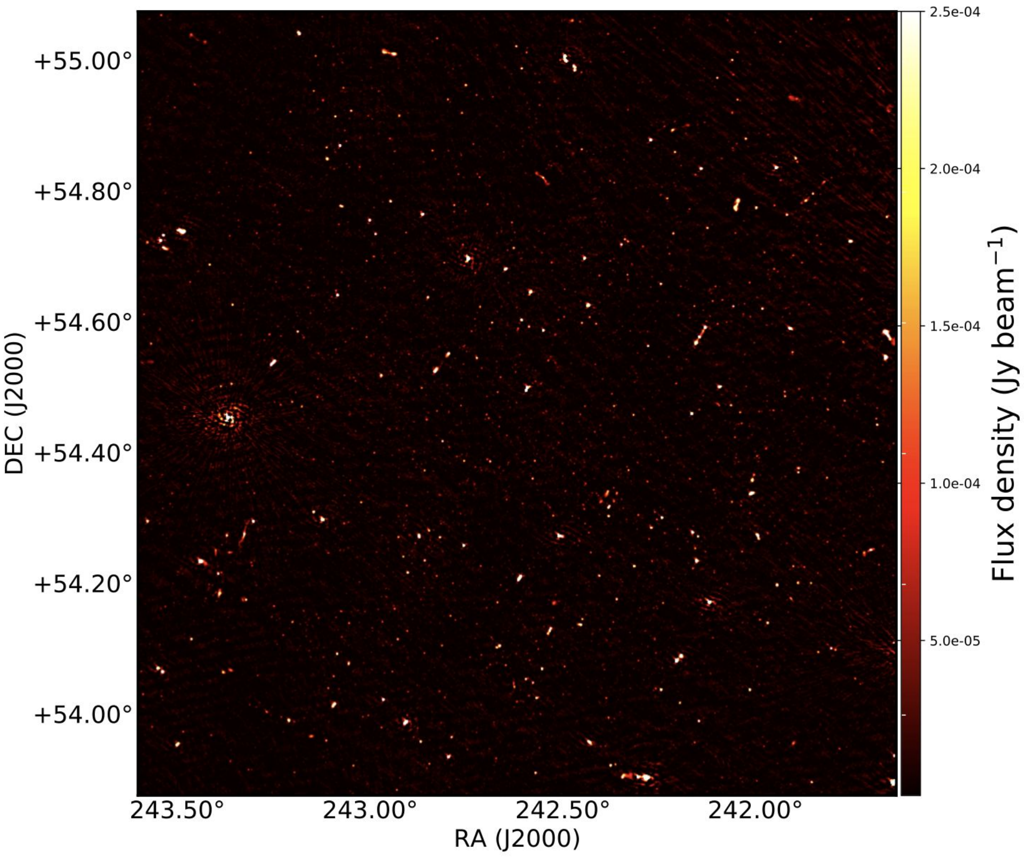
- Good estimate of the foreground
- The gain errors are independent of the sky signal
- The 21-cm signal and the foregrounds are independent to each other.

Kumar et al. 2022

Observation : DATA

Project code	32_120
Observation date	5,6,7 May 2017
	27 June 2017
Bandwidth	200 MHz
Frequency range	300-500 MHz
Channels	8192
Integration time	2s
Correlations	RR RL LR LL
Total on-source time	13 h (ELAIS N1)
Working antennas	28
Flux Calibrator	
Source	3C286
Flux Density	23 Jy
Source	3C48
Flux Density	42 Jy
Phase Calibrator	
Source	J1549+506
Flux Density	0.3 Jy
Target Field	
Source	ELAIS N1

uGMRT image is zoomed-in total intensity image of ELIAS N1 at 400MHz (bandwidth 200 MHz)

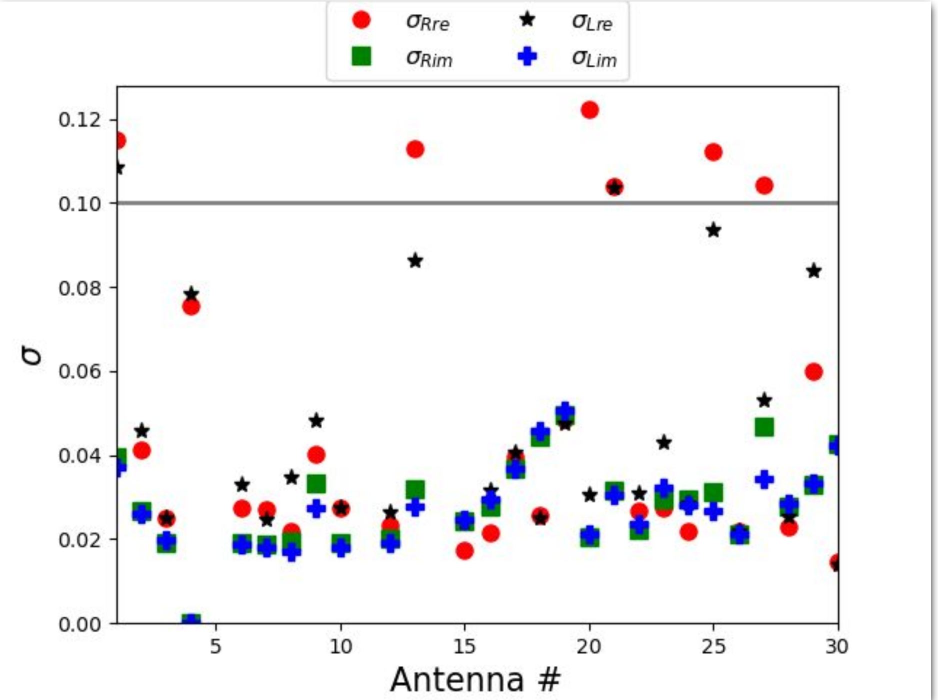


Chakraborty et al. 2019(b)

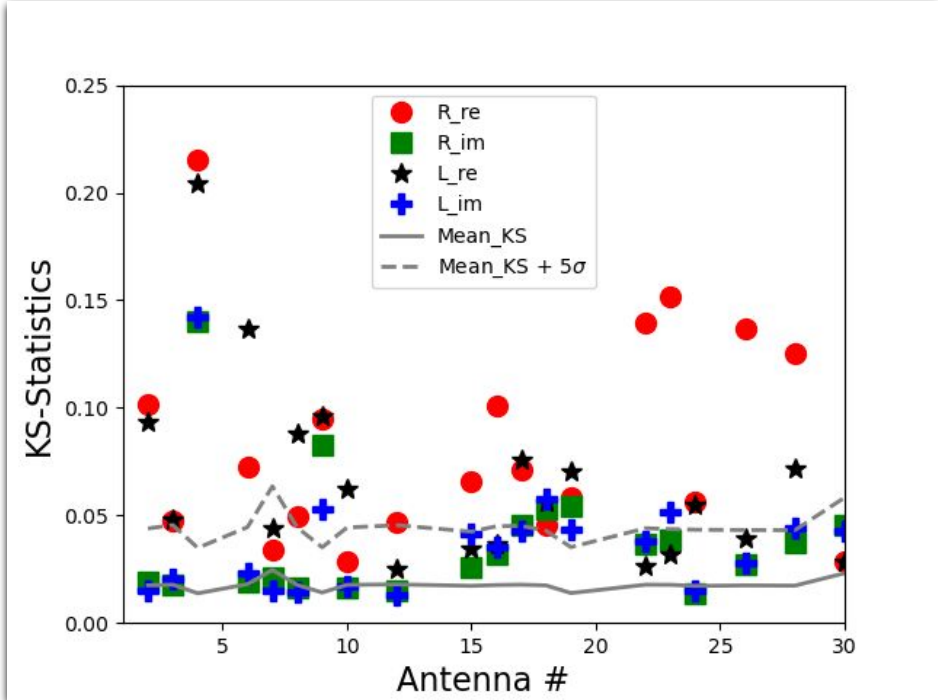
Gain Characteristics

- We set a threshold standard deviation and flagged bad antennae for further analysis.
- We perform 'kolmogorov-smirnov ' test to check gaussianity of residual gain errors of corresponding antennae.

Standard Deviation of residual gain of corresponding antennae



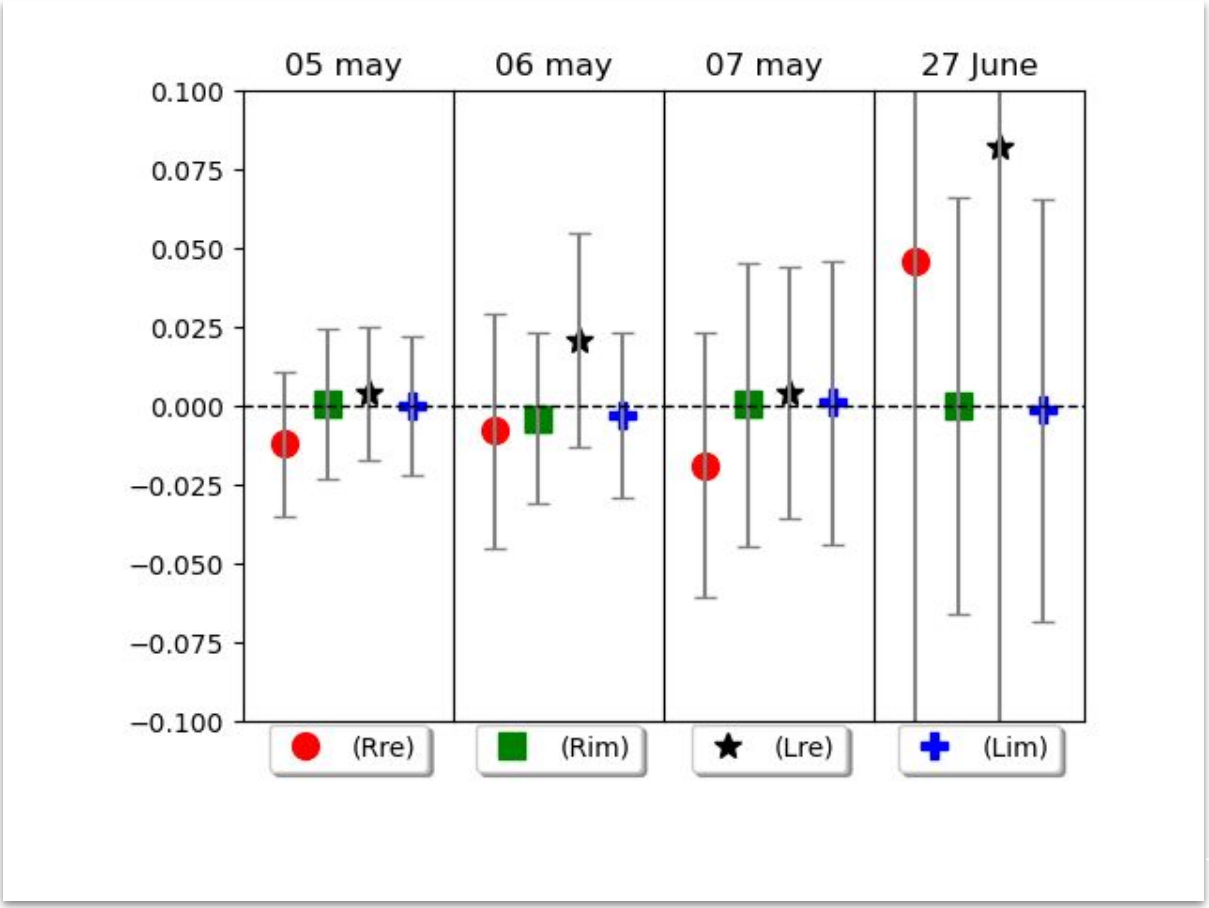
KS Statistics



Gain Characteristics

- Mean of residual gain errors for imaginary part are zero but for real part there are deviation.
- Error bar represents standard deviation and large error bar hints for bad data.

Mean and Standard Deviation for four days together

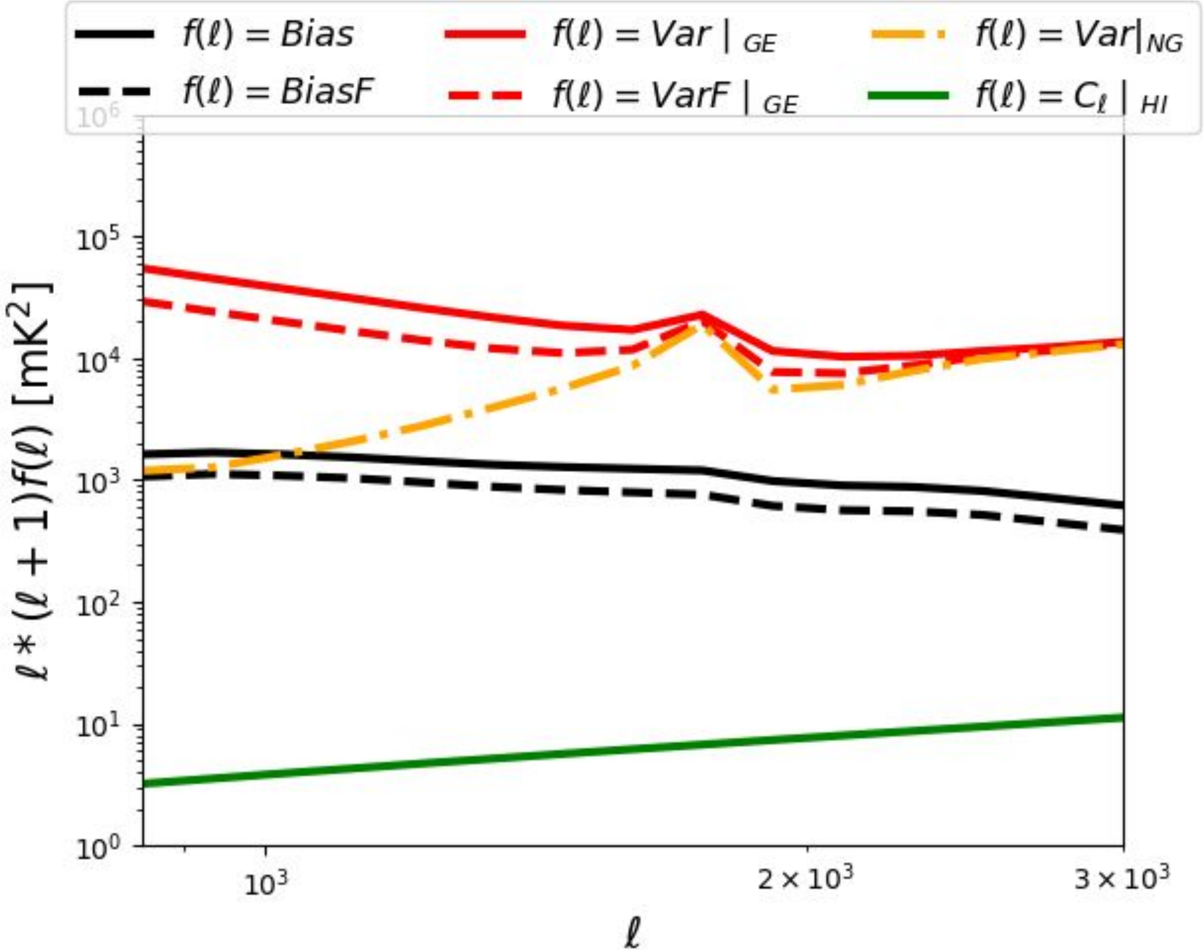


Results

Bias and variance of power spectrum due to residual gain errors for a single day

- In Spite of data points decrease due to flag, bias and variance decrease after flagging, as bad antenna are flagged.

- For shorter baseline residual gain error is crucial for excess variance of power spectrum.

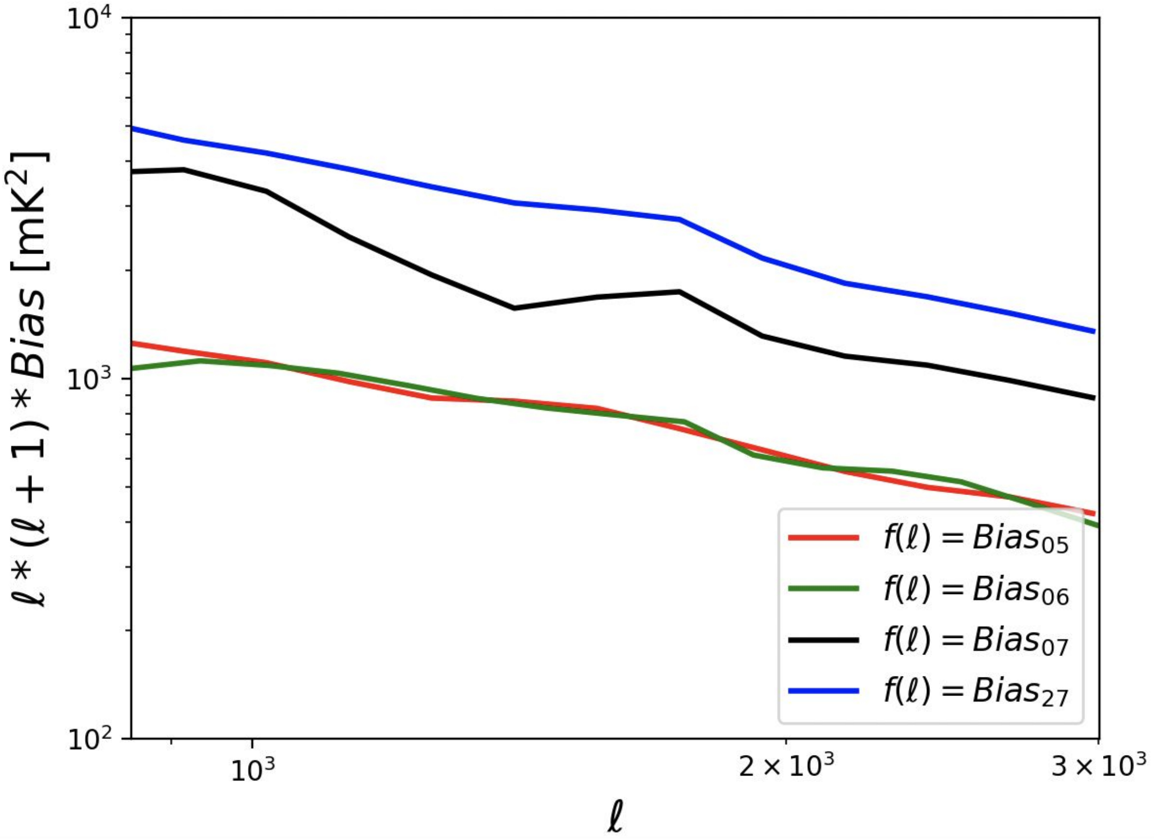


Results

- For first two day bias are low and varying almost in same manner with baseline.

- As discussed last two days observation were not that good due to RFI.

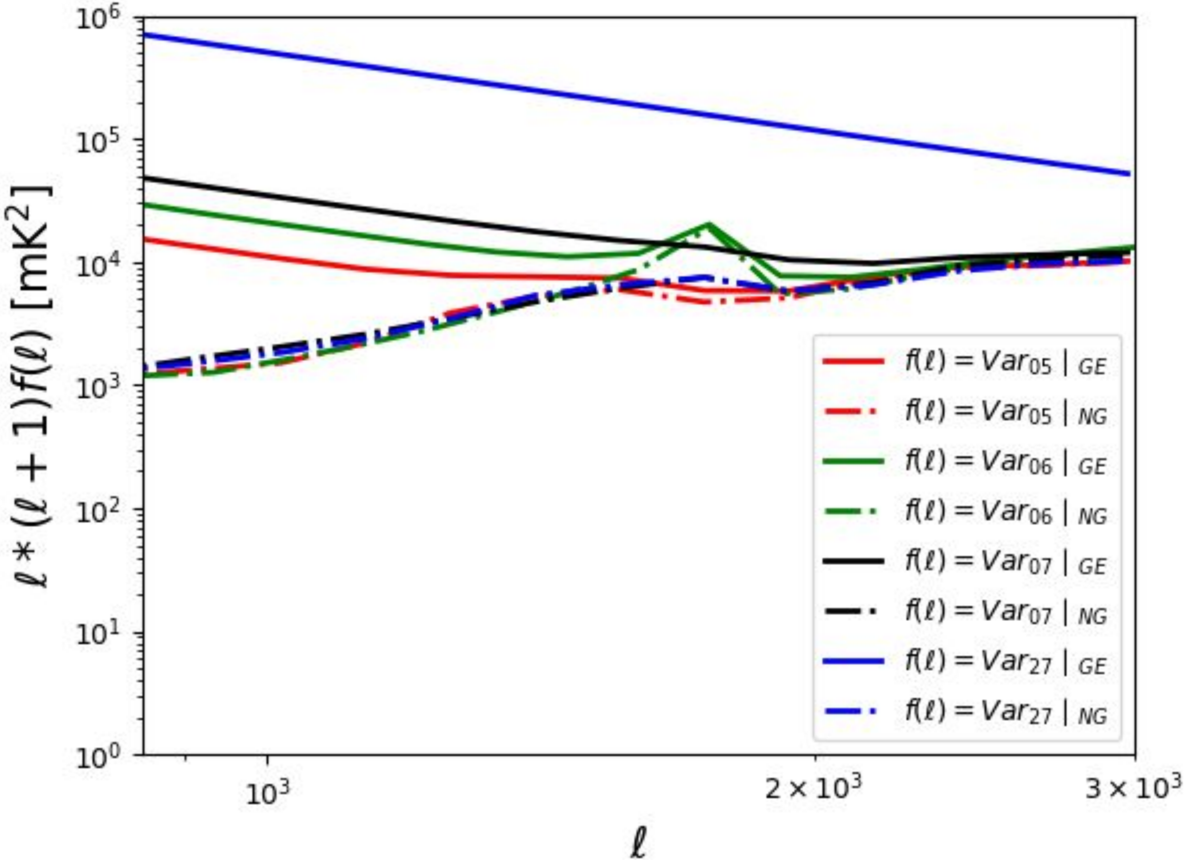
Bias of power spectrum due to residual gain errors for a single day for four different day together.



Results

Variance of power spectrum due to residual gain errors for a single day for four different day together.

- Variance without residual gain errors are same for all four day.
- Variance due to residual gain errors for first two days are comparatively low.

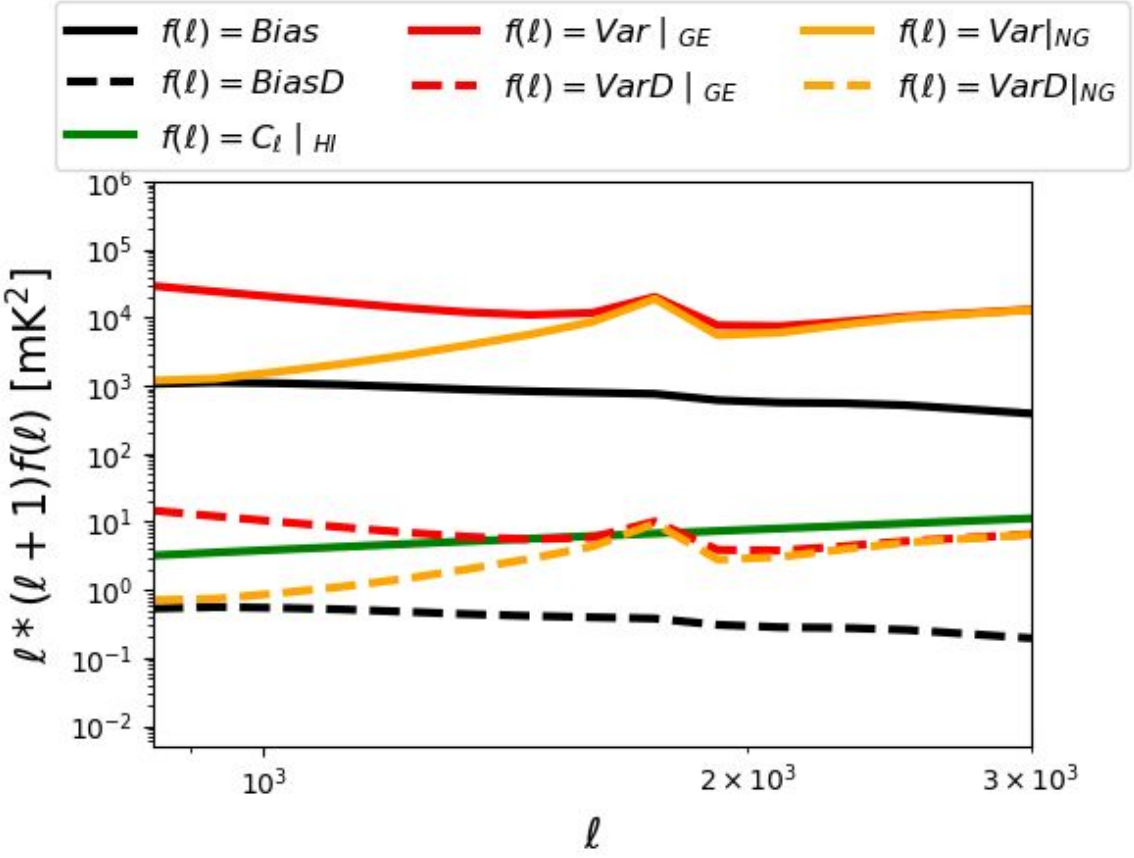


Results

Optimum number of days for **HI detection**

- Solid line is for , bias and variance of power spectrum for four days observation considering a good day.

- Dashed line represents optimum number of days to recover 21-cm power spectrum(green line) , here it is 2000 days of observation !



Conclusion

- Calculation of residual gain errors required for such high dynamic observation.
- Investigating gain characteristics we can detect bad observation and applying flagging we can decrease bias and uncertainty in power spectrum.
- We can analyse gain characteristics of observed data of different types of telescope and plan our observation accordingly in terms of antennae, baseline length etc.