Global 21cm Signal Estimation Using Neural Networks

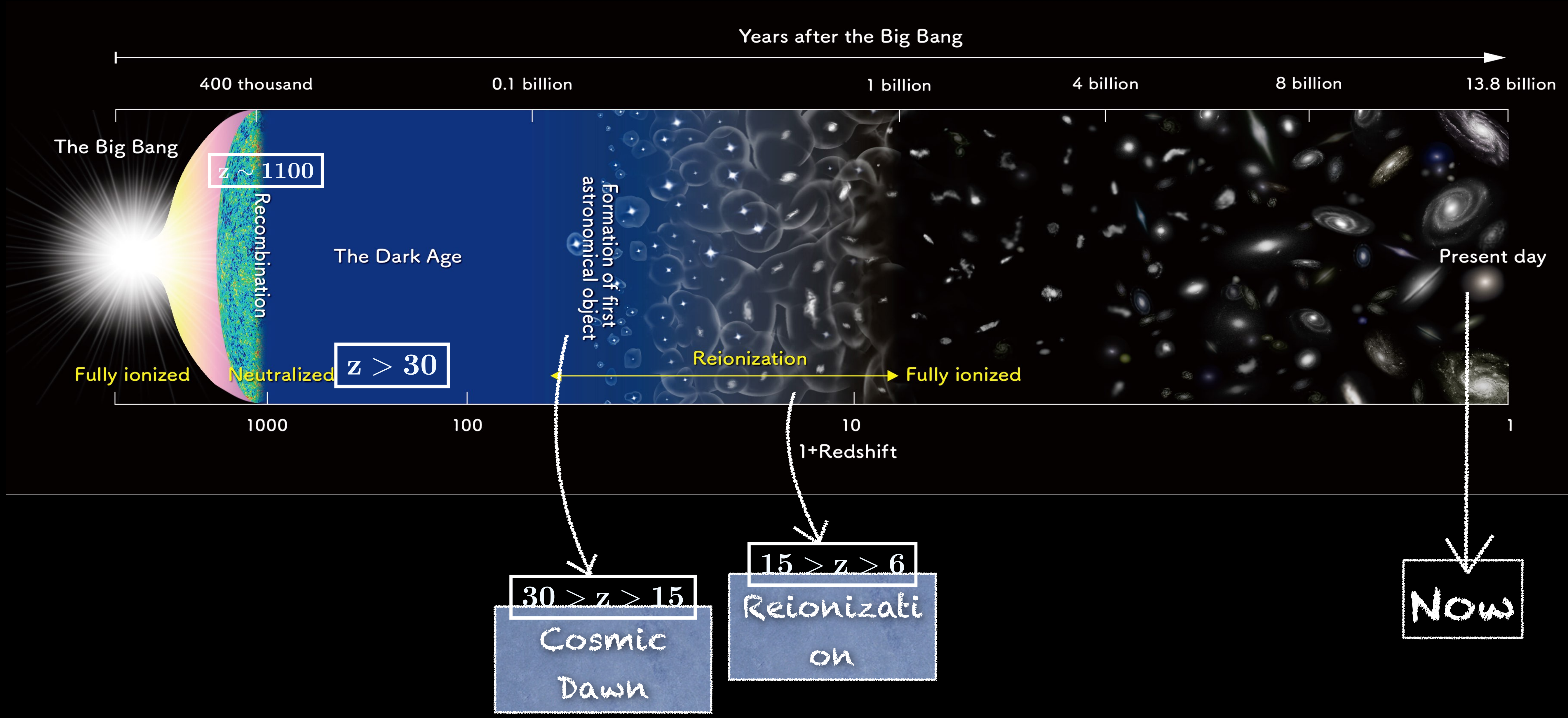
Abhirup Datta

Madhurima Choudhury, Arnab Chakraborty

Discipline of Astronomy, Astrophysics and Space Engineering
Indian Institute of Technology Indore

Choudhury, M. et al. (2018, 2019)
Chakraborty, A. et al. (2019a, 2019b)

The story of the universe until now...

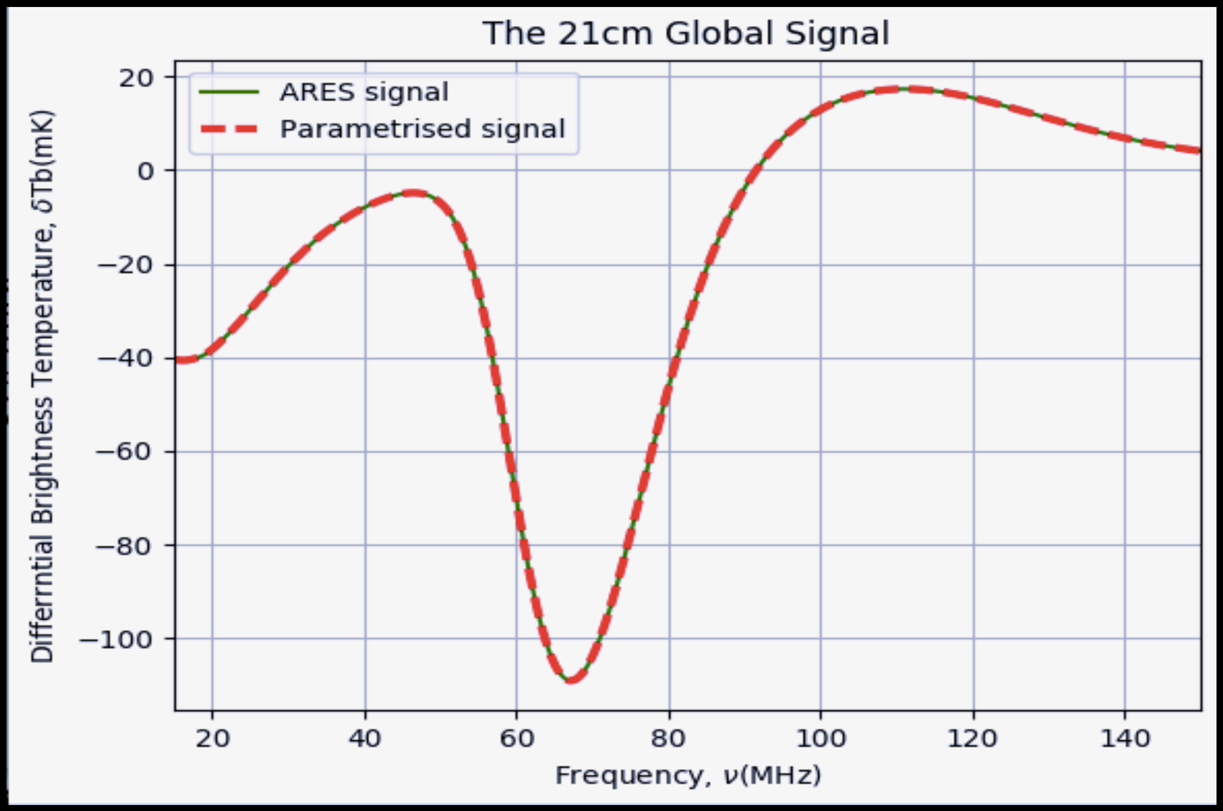


A MODEL 21CM GLOBAL SIGNAL

- We use the Accelerated Reionization Era Simulations (ares) code was designed to rapidly generate models for the global 21-cm signal (Mirocha et al, 2012, 2015).

- We have used the tanh model for parametrising the global signal, where the parameters , A(z) are the parameters for the global signal.

- $$A(z) = \frac{A_{ref}}{2}\{1 + \tanh[(z_0 - z)/\Delta z]\}$$



Parameters evolve according to a tanh model

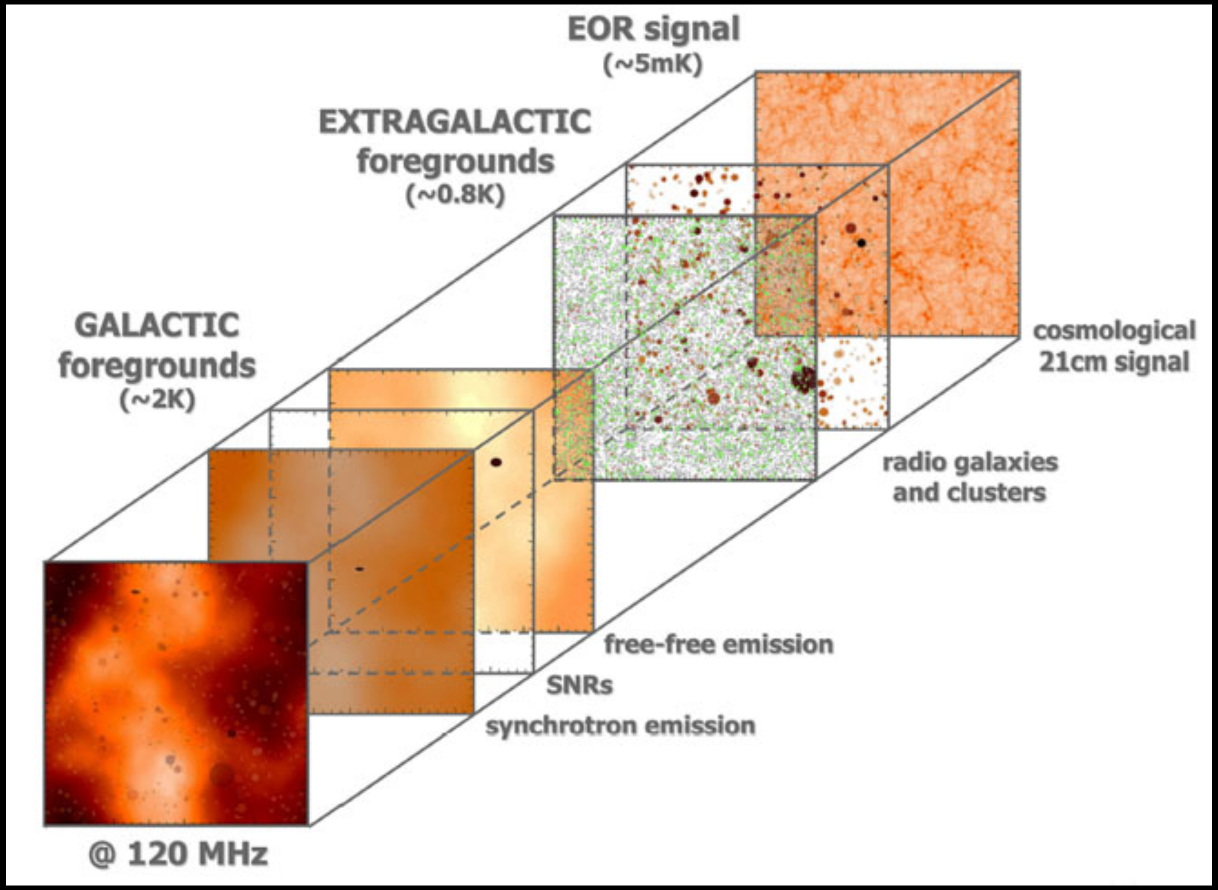
- J(z)—Lyman-alpha background (which determines the strength of W-F coupling)
- Xi(z)— Ionized fraction of hydrogen
- T(z)—temperature of the IGM

A MODEL FOREGROUND

- A typical foreground

$$\ln T_{FG} = \sum_{i}^{n} a_i \, [\ln(\nu/\nu_0)]^i$$

- Where, all temperatures are in K, and , is an arbitrary reference frequency, which is chosen to lie in the middle of our band.

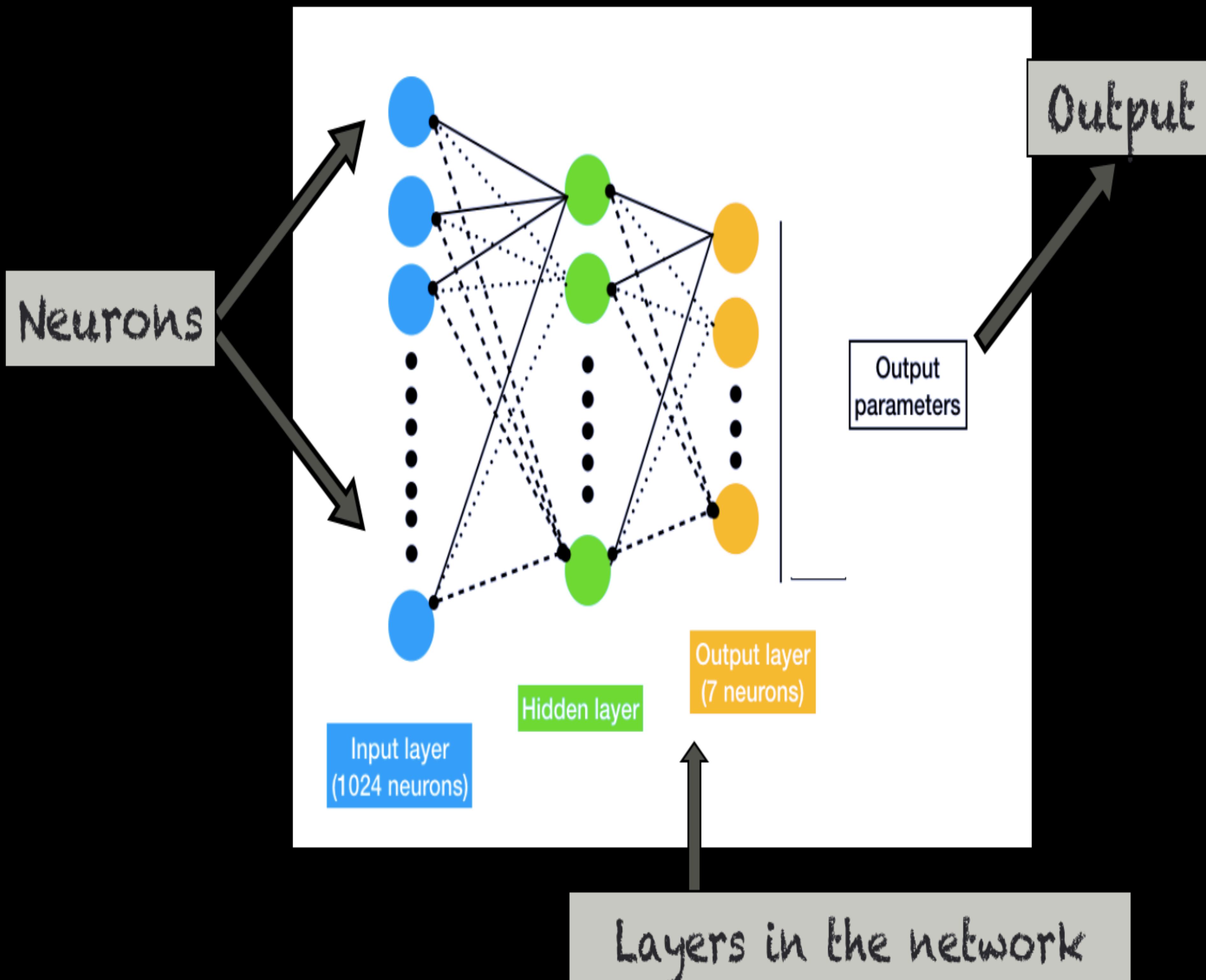


- The foreground parameters

$$\boxed{a_0, a_1, a_2, a_3}$$

Extracting the HI 21cm Global signal

Basic architecture of the network



- The ANN constructs functions, which associates the input with the output data.
- The basic neural network model is described by a series of functional transformations

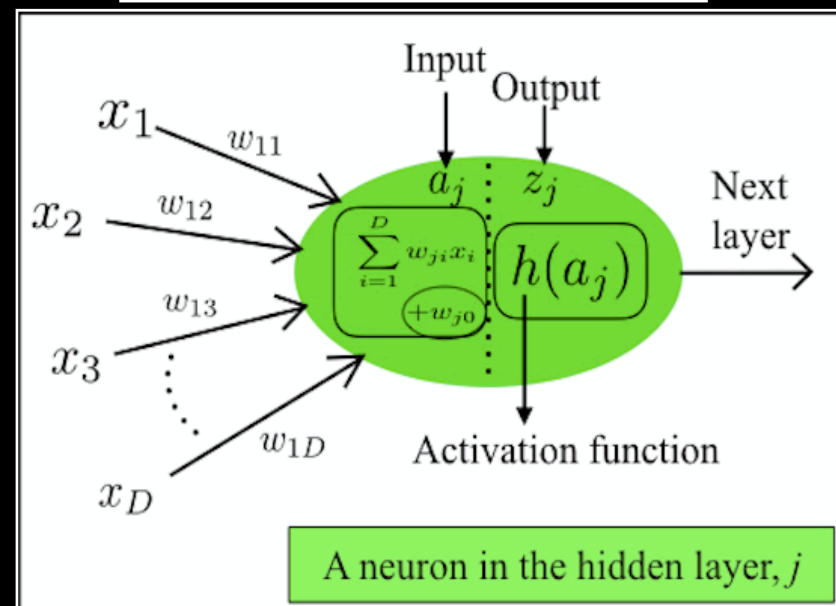
The training algorithm

- Each neuron in the input layer is connected to the next layer and a weight and a bias is associated with the connection.
- The input to each neuron in the hidden layer is a linear combination of all such possible connections.

$$a_j = \sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}$$

Weights

Biases



A neuron in the hidden layer, *j*

- This is then activated by a suitable activation function, h().

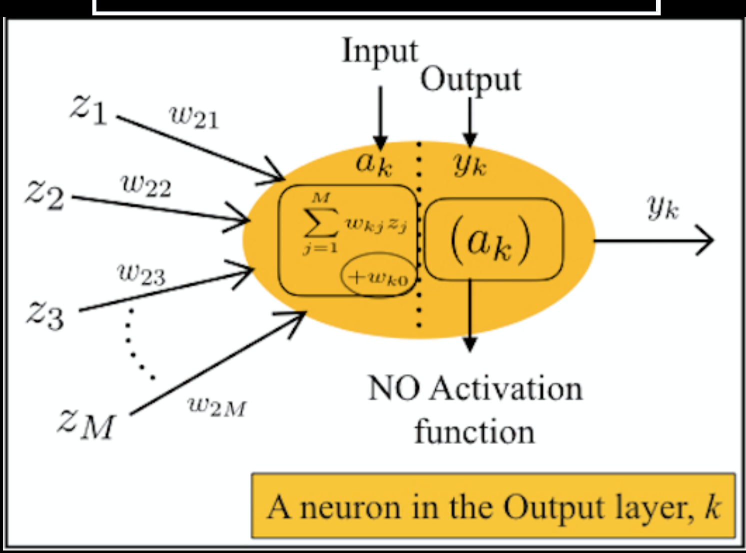
$$z_j = h(a_j)$$

Activation function: Logistic sigmoid

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

- These are the outputs of the neurons in the hidden layer.

$$a_k = \sum_{j=1}^{M} w_{kj}^{(2)} z_j + w_{k0}^{(2)}$$



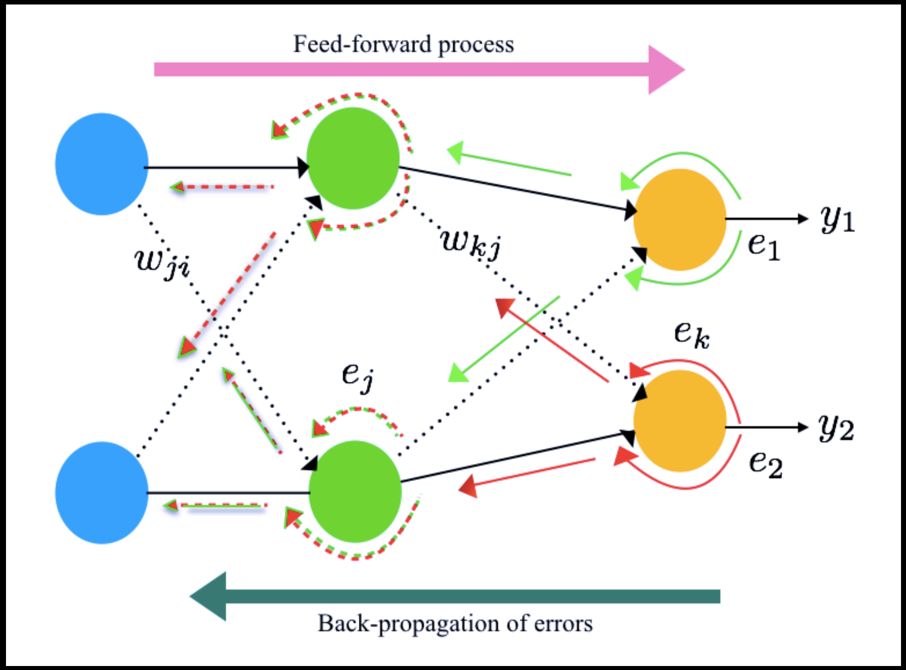
A neuron in the Output layer, *k*

- The input to the neurons in this layer is again a combination of all the z's from the previous step.
- Usually the output is not activated by any activation function, and we get the outputs from each output neuron.

The training algorithm continued

Two more steps are involved in this process.

- Optimization
- Back-propagation



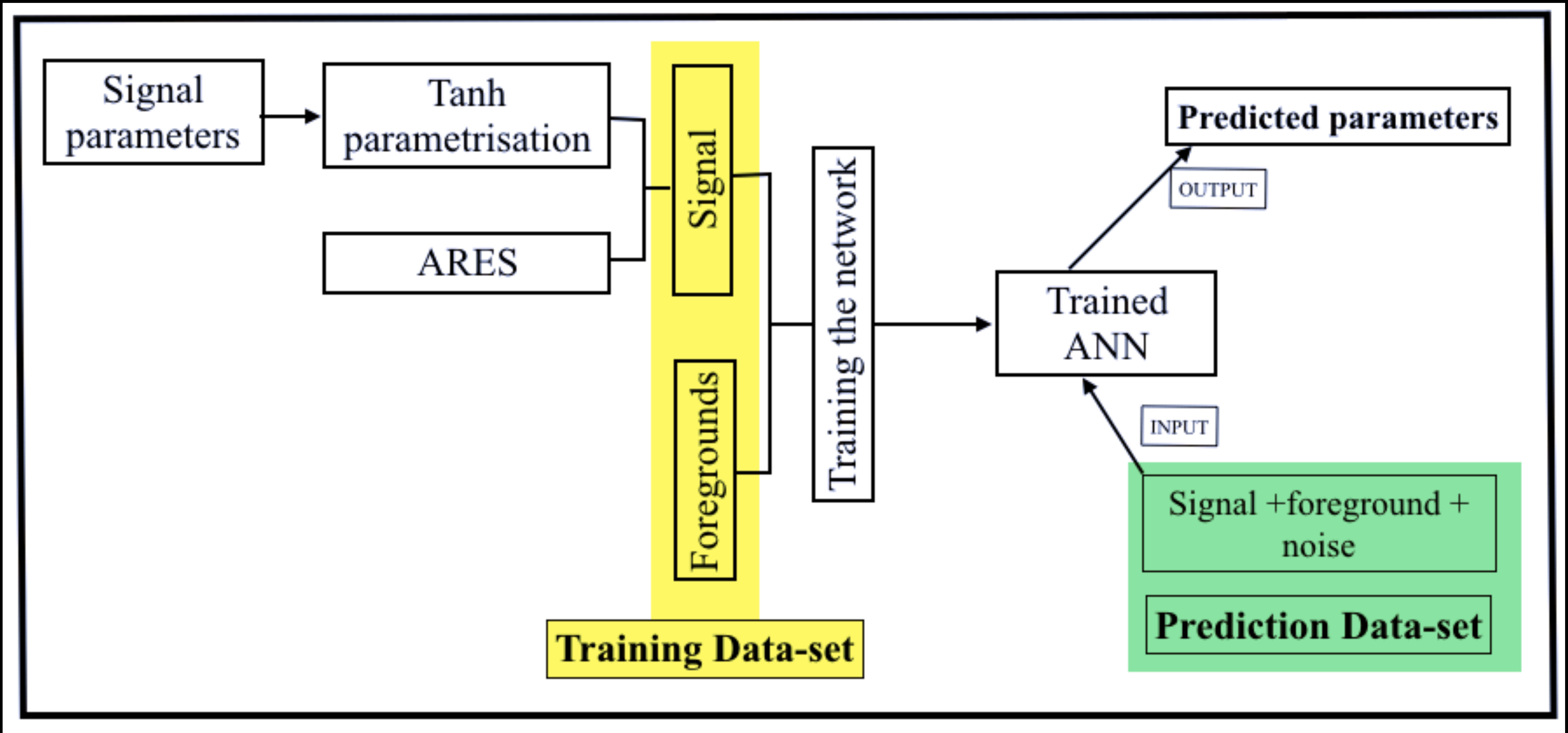
Choudhury et al (2018)

- An error/cost function is computed at the end of one feed-forward process.

$$\sqrt{\frac{1}{N_{train}}\sum^{N}\left(\frac{y_{pred}-y_{ori}}{y_{ori}}\right)^2}$$

- The idea is to minimise this error function by assigning suitable weights and biases at every step (Optimizing the weight and bias parameters).
- This process repeated till the error function is minimum. (This is called back-propagation)

Flowchart


Signal parameters
Tanh parametrisation
ARES
Signal
Foregrounds
Training the network
Trained ANN
Predicted parameters
OUTPUT
INPUT
Signal +foreground + noise
Prediction Data-set
Training Data-set

Building the training dataset

We need to simulate

- A model 21cm signal
- A model foreground
- Add effect of the instrument response.
- Add thermal noise

ARES

Log polynomial Model

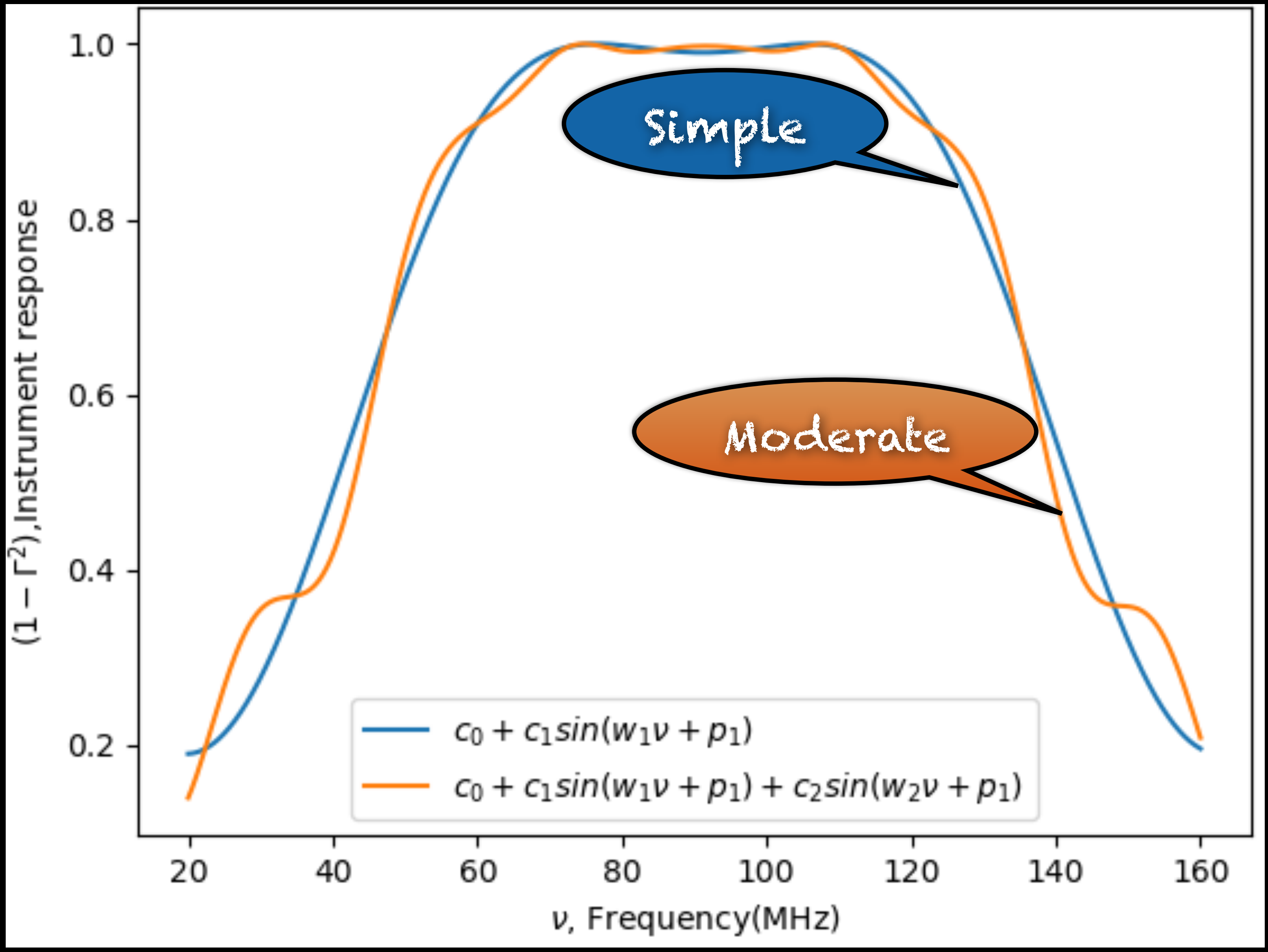
$$\textbf{Noise, } n = \frac{T_{FG}}{\sqrt{\Delta\nu * 10^6 * 3600 * N_t}}$$

An Instrument model

- We have considered two very simple models for the instrument.
- The instrument response is given by:

$$G(\nu) = |1 - \Gamma^2|$$

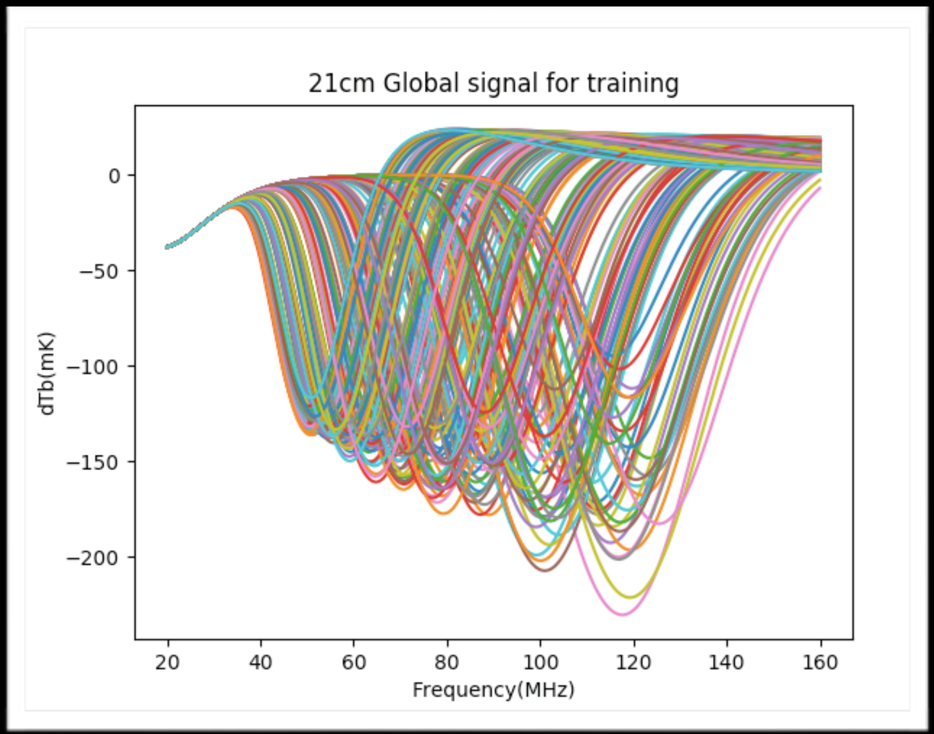
Antenna reflection coefficient

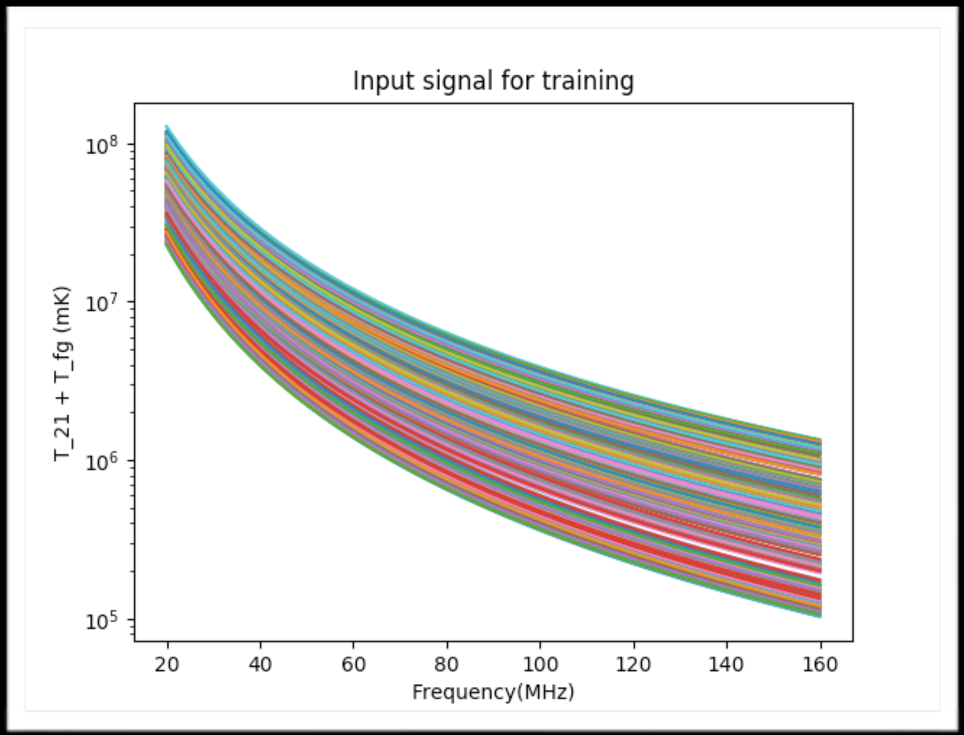


[DARE Radiometer calibration, R.Bradley, 2012]

Choudhury et al,2019 (in review)

The training dataset, for the perfect instrument





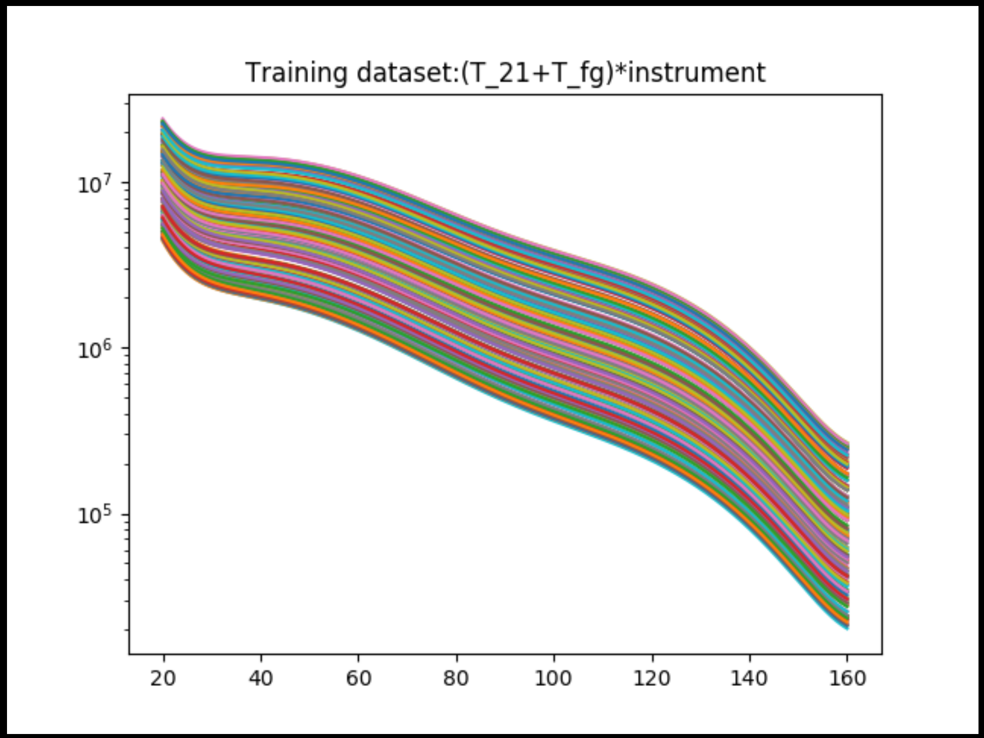
Total signal fed as input for training process.

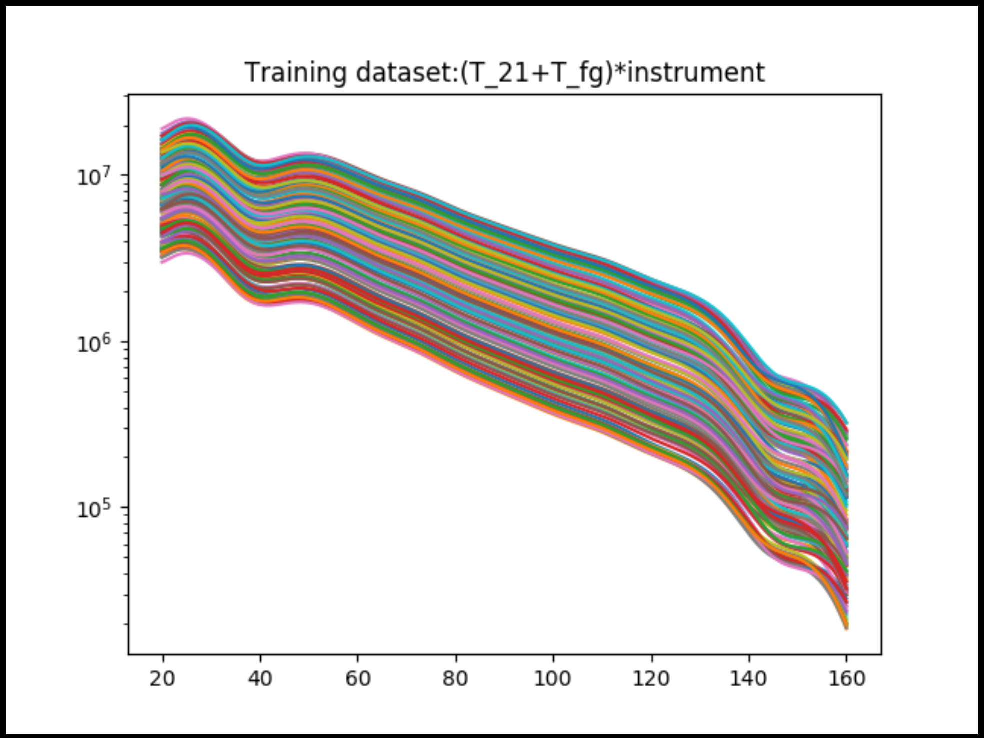
$$T_{tot} = T_{21} + T_{FG}$$

- The training dataset is constructed and the network is trained.
- We choose the optimum number of neutrons in the hidden layer and the number of iterations, till the error function is minimum.
- The network is tested and validated

Network is ready to be used!

The training dataset, when modified by a instrument model

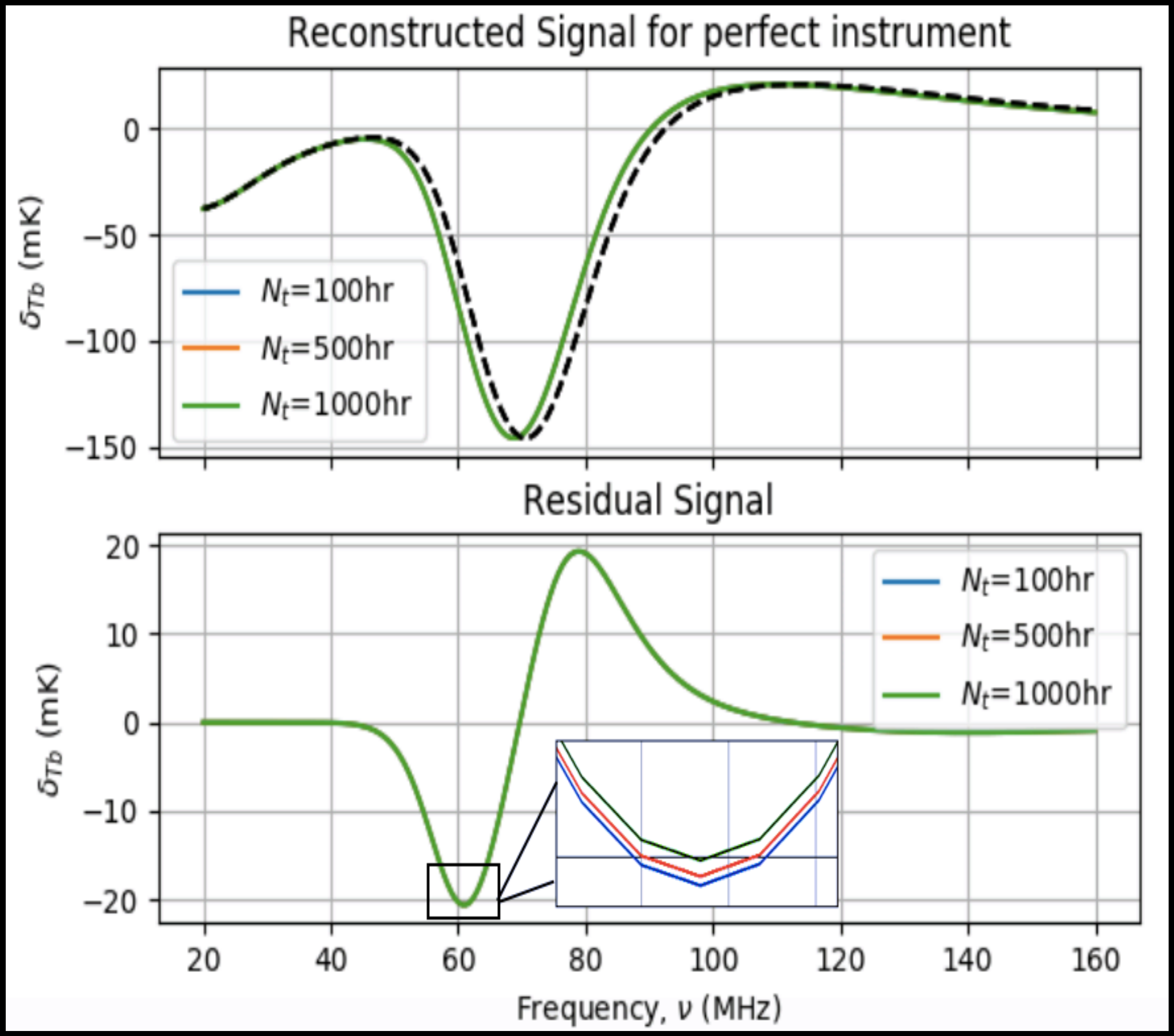




$$T_{tot} = (T_{21} + T_{FG}) * G(\nu)_{sim}$$

$$T_{tot} = (T_{21} + T_{FG}) * G(\nu)_{mod}$$

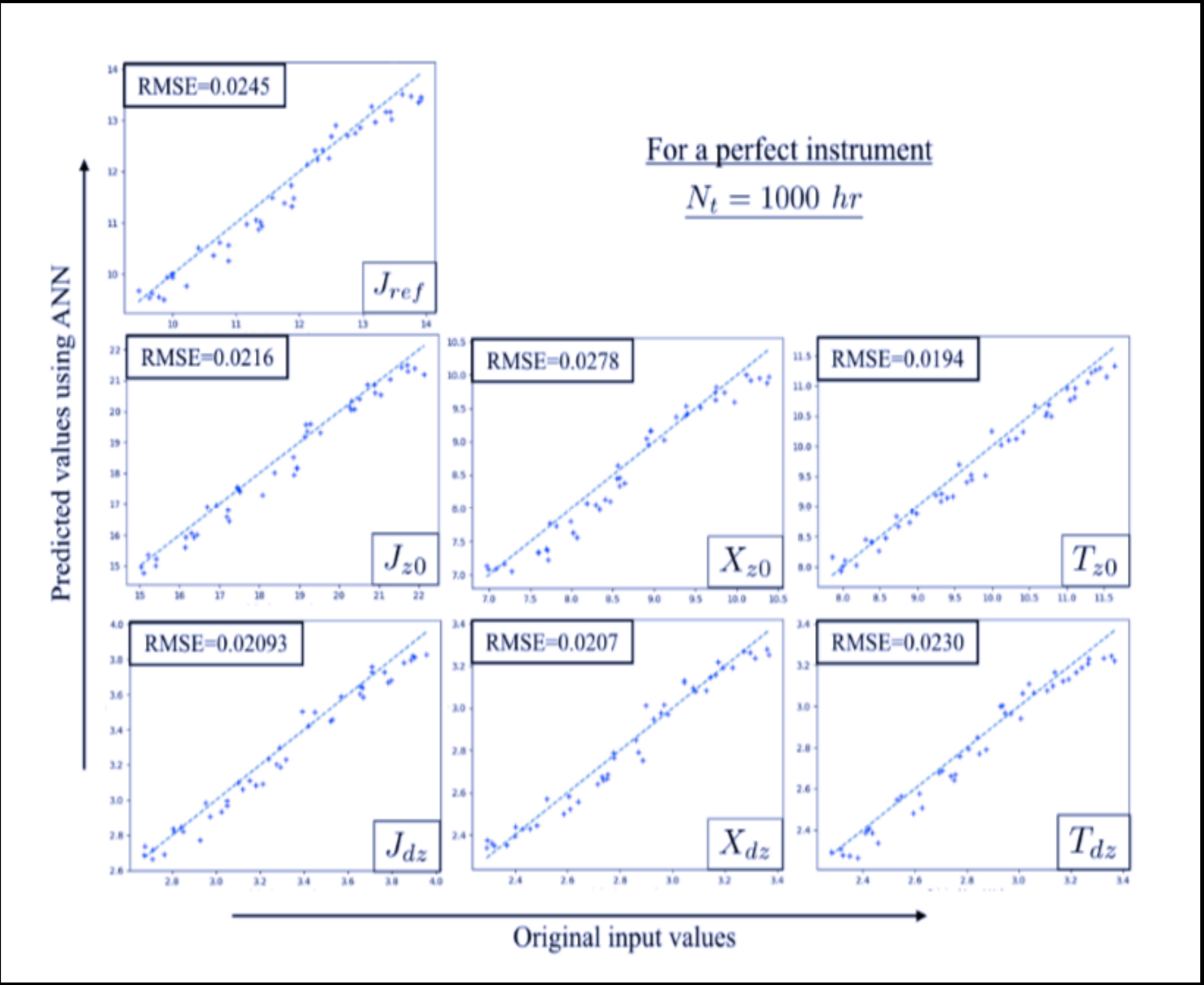
Case 1: Perfect Instrument



- A known input signal, along with foreground and noise is fed into the network, for different observation periods.
- The signal parameters are estimated and the signal is reconstructed
- Note that the reconstructed signal is very close to the original input signal
- The residual signal is the difference

$$res = (T_{21})_{org} - (T_{21})_{recon}$$

Training RMSEs



- A set of 90 test data is taken and is fed into the network
- The original vs the predicted values of the parameters are shown in this plot for the perfect instrument case.
- RMSE values are noted.

Choudhury, M et al. 2018

Tabulating the RMSE's

Parameters	Perfect Instrument	Fixed(1sine) Instrument	Varying(1sine) Instrument	Fixed(2sine) Instrument	Varying(2sine) Instrument
J_{ref}	0.0245	0.0705	0.0702	0.0616	0.0642
dz_J	0.0209	0.0575	0.0581	0.0502	0.0765
dz_T	0.0230	0.0668	0.0599	0.0531	0.0794
dz_X	0.0207	0.0709	0.0684	0.0614	0.0796
$z0_J$	0.0216	0.0550	0.0730	0.0607	0.0672
$z0_T$	0.0194	0.0650	0.0840	0.0535	0.0642
$z0_X$	0.0278	0.0739	0.0809	0.0556	0.0784

RMSE increases with more complexity of the dataset, but is still considerably small.
In other words, we get very good prediction of the parameters.

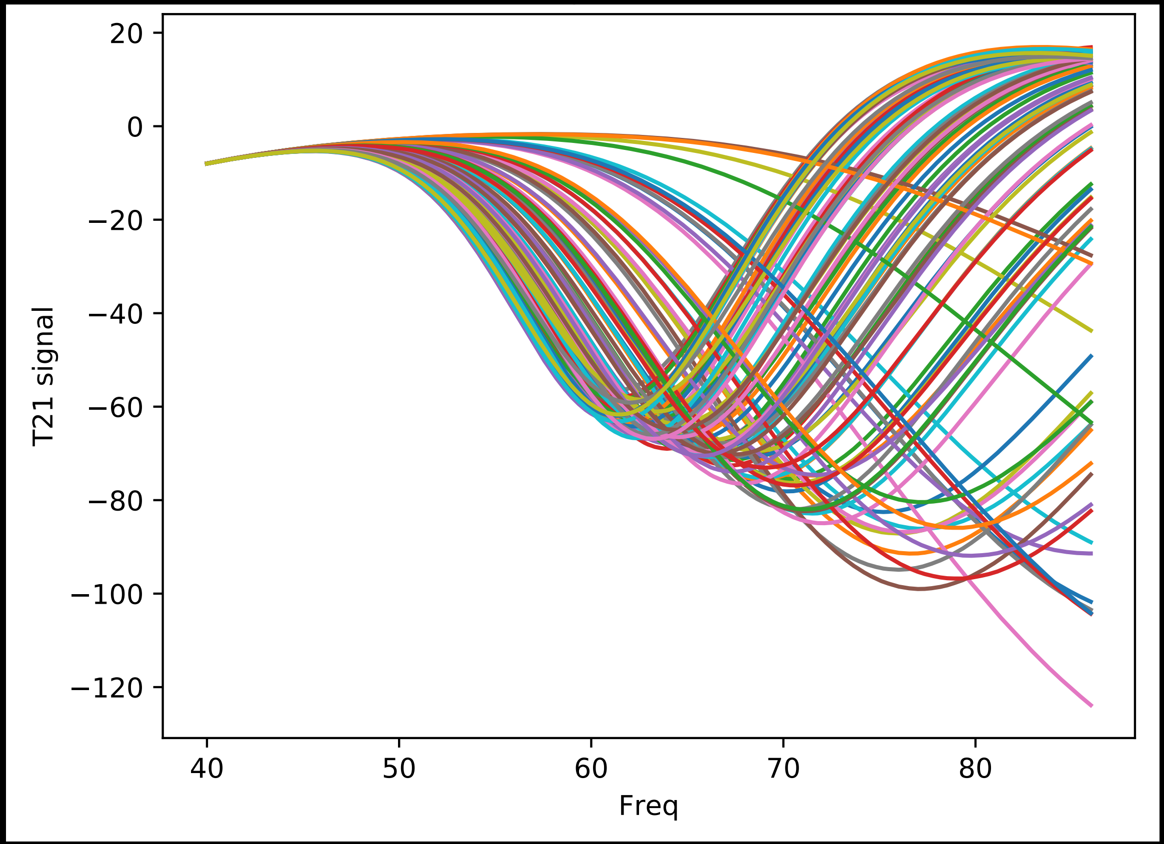
93-98%

Choudhury et al, 2019 (under review)

Un-parametrized models

fX
f_star
f_esc
N_ion
N_lw

The ARES code takes in values of the IGM parameters.



Choudhury, M. et al. (in prep)

Unparametrized signal from

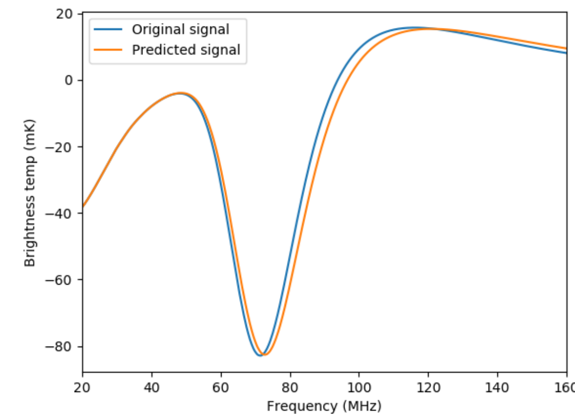
Signal + Foregrounds

Training the network

Model saved.
Network ready to use.

<u>Input</u>
Unknown Signal
+foreground + noise

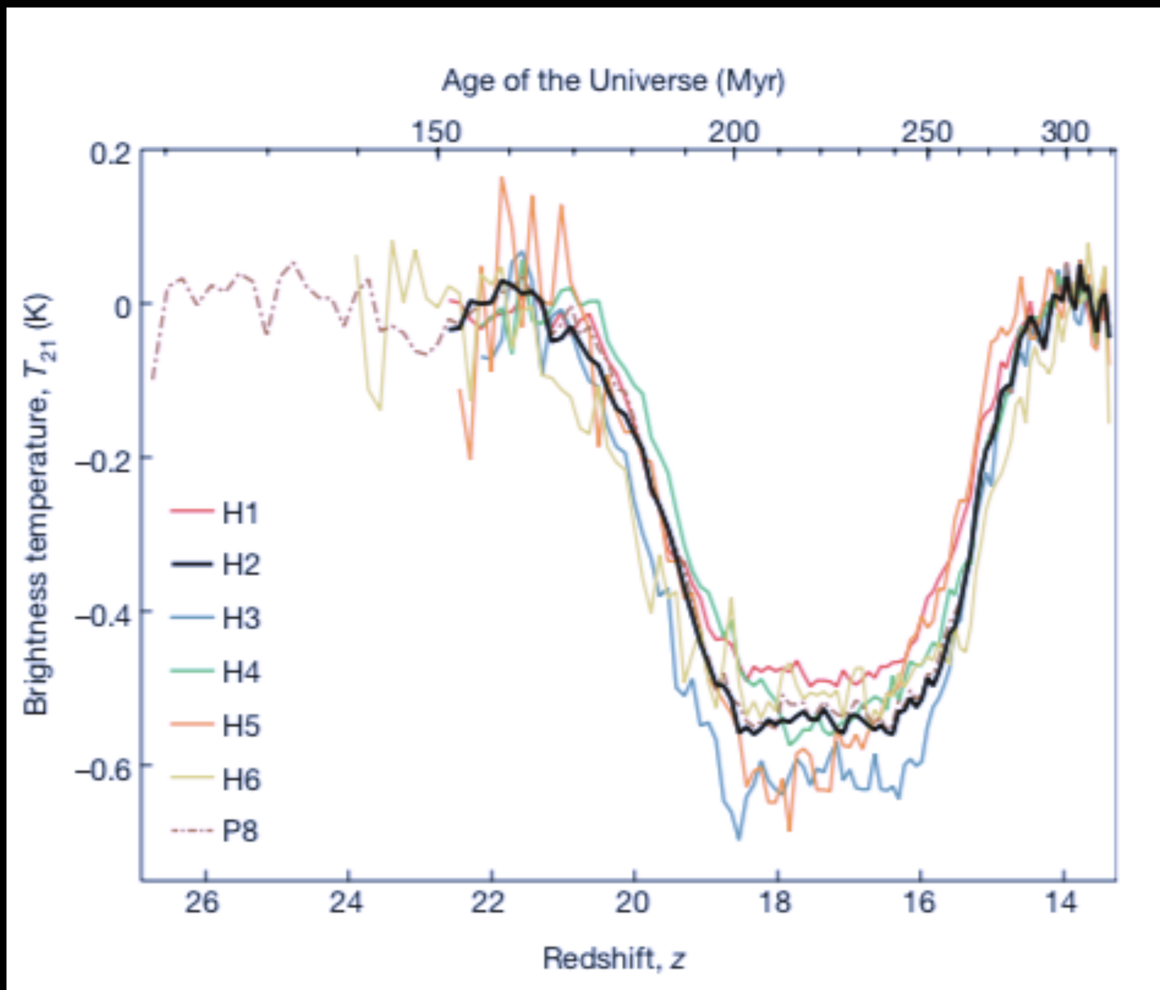
<u>Output</u>
Predicted signal &
foreground parameters



Parameter	Original	Predicted	Percentage Error
fX	3.81052237488	3.60149	5.48566763258
fstar	0.030428869624	0.0254143	16.4795310371
fesc	0.520263919378	0.523806	0.680761894243
nion	2666.66666667	2568.52	3.68036193848
nlw	7888.88888889	7771.27	1.49093103543
a0	3.13011351051	3.05784	2.30911535152
a1	-2.53626586951	-2.53478	0.058461717634
a2	-0.080900797364	-0.0810238	0.152001254055
a3	0.0288739941525	0.0288333	0.140866511004

84-99% accuracy!

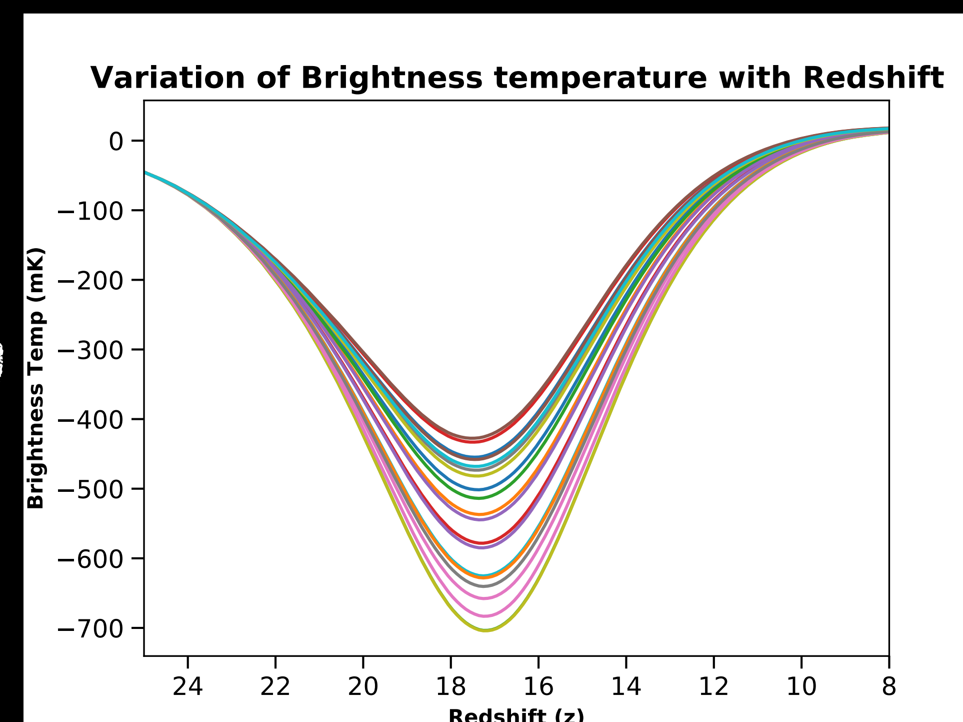
Recent detection from EDGES experiment



Models from :
Chatterjee, Dayal, Roy Choudhury and Hutter (2019)

We now want to include such models as well in our training sets

Unexpectedly large absorption dip



$$T_F(\upsilon) \approx a_0\ (\upsilon/\upsilon_c)^{-2.5} + a_1\ (\upsilon/\upsilon_c)^{-2.5}\log(\upsilon/\upsilon_c) + a_2\ (\upsilon/\upsilon_c)^{-2.5}[\log(\upsilon/\upsilon_c)]^2 + a_3\ (\upsilon/\upsilon_c)^{-4.5} + a_4\ (\upsilon/\upsilon_c)^{-2}\ . \quad (2)$$

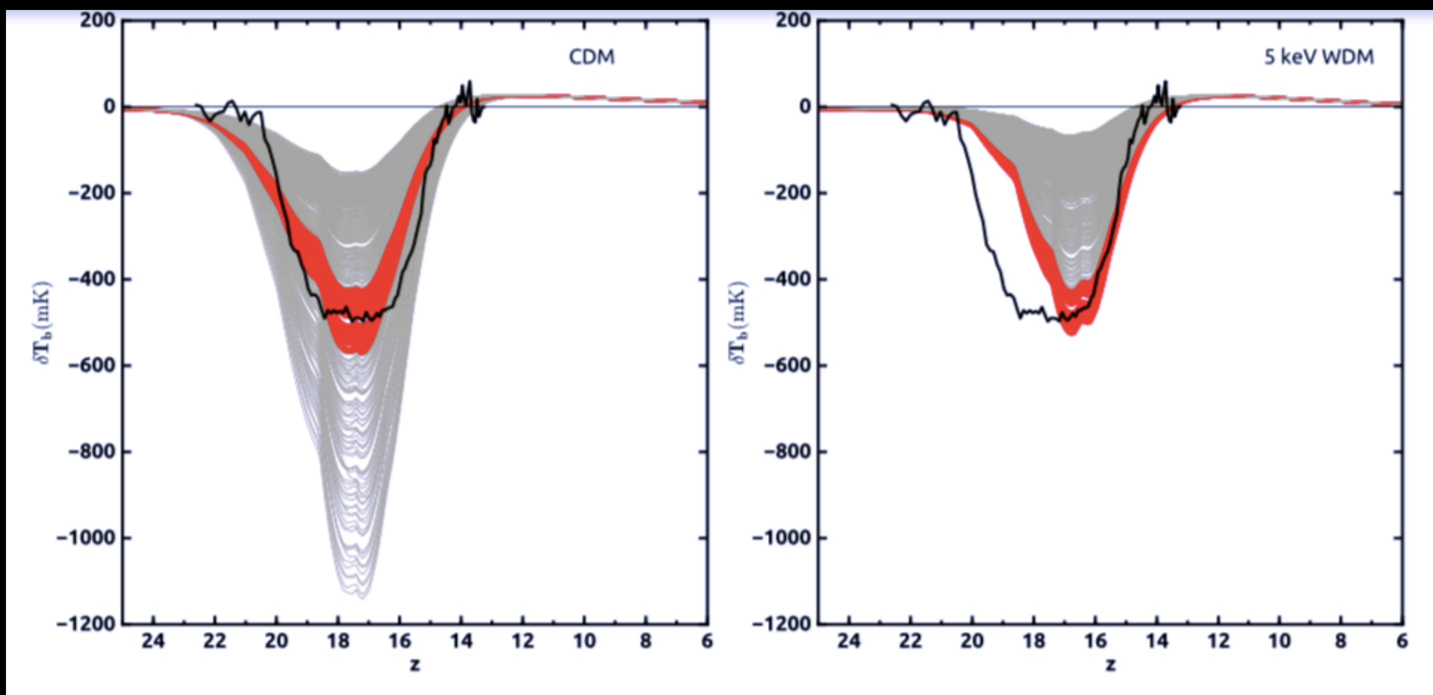
Foregrounds from EDGES results

Choudhury, M. et al. (in prep)

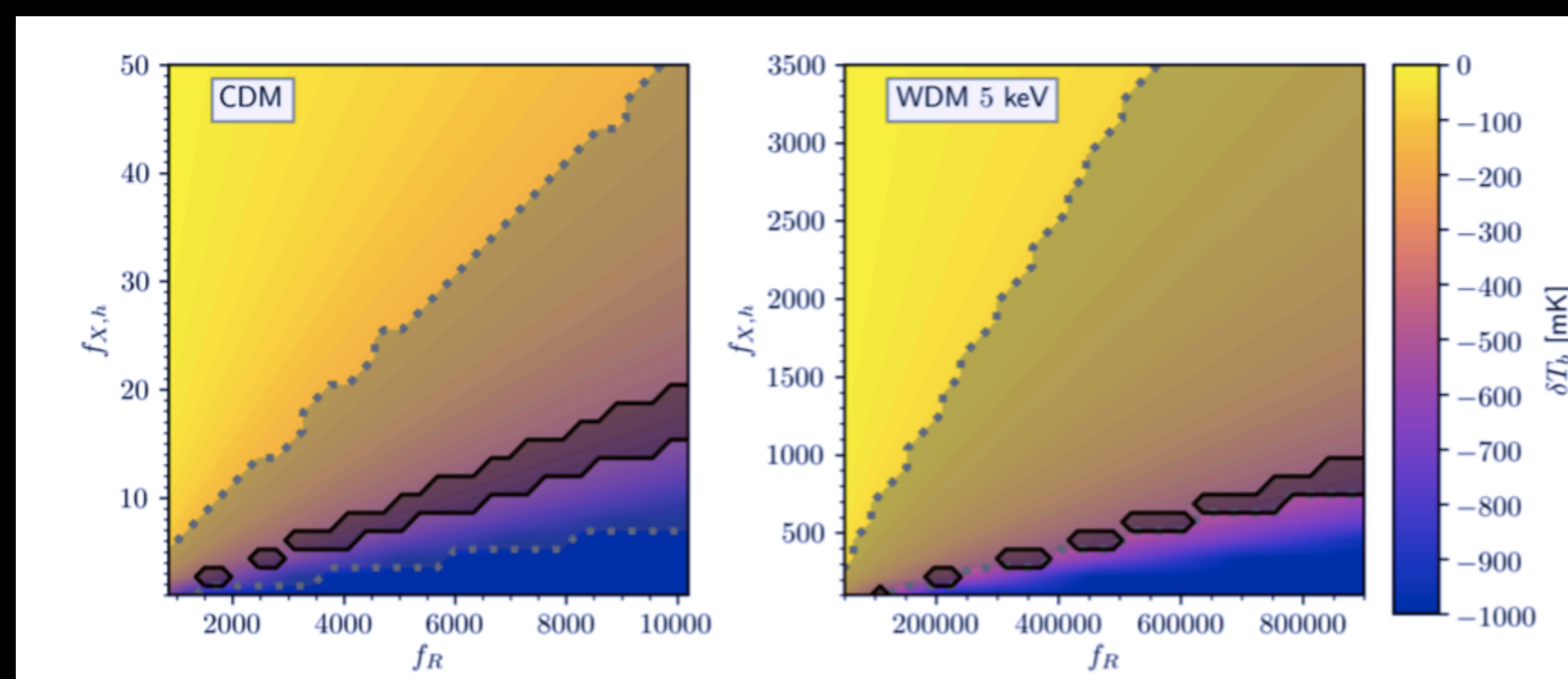
Models with excess radio backgrounds

The flux from high z sources is converted into a radio brightness temperature T_R, resulting in a total background temperature given by:

$$T_\gamma(z) = T_R(z) + T_{CMB}(z)$$



- **In each panel the black curve shows the EDGES result. The grey lines in each panel show models that satisfy the ARCADE-2 limits and where the signal is limited to z ~> 14.**
- **The red lines show models consistent with the EDGES result, both in terms of the redshift range of the signal as well as its amplitude (δTb = –500±75mK).**
- **As shown, the inclusion of an excess radio background results in free parameter combinations (fR and fX,h) yielding results in agreement with the EDGES data for the CDM and 5 keV WDM models.**

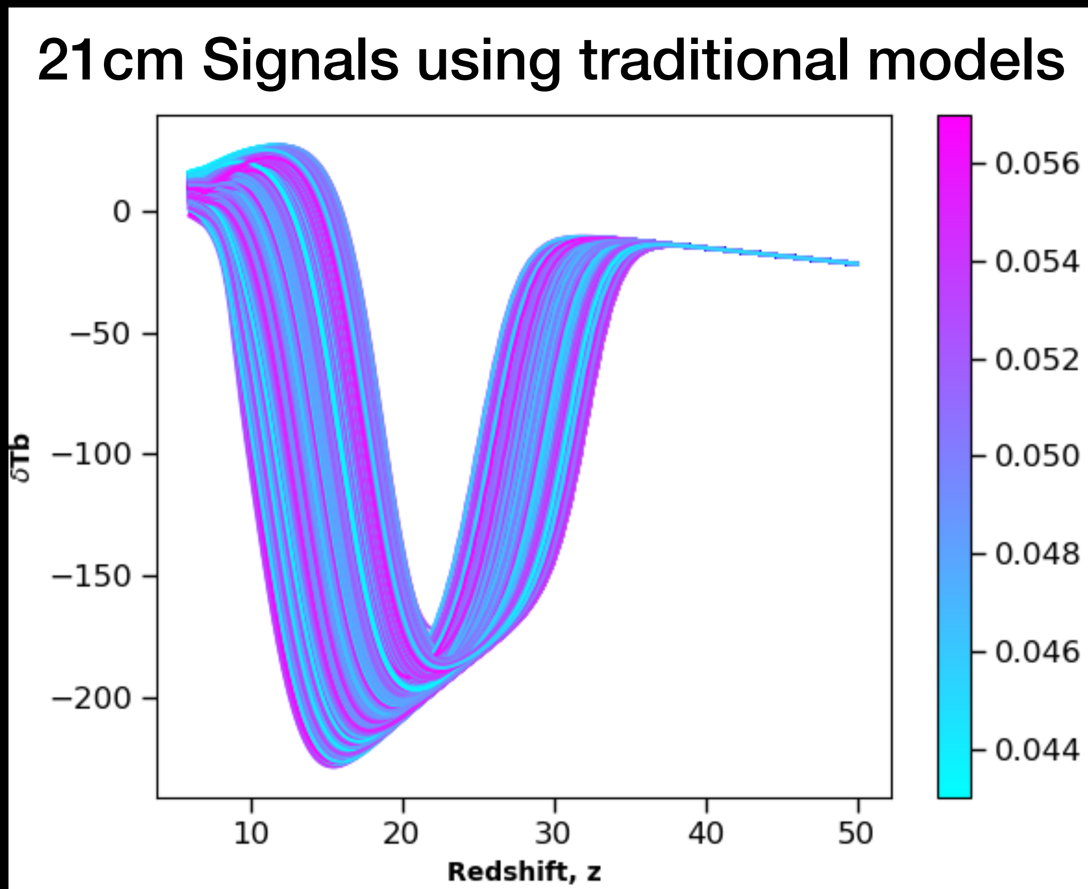


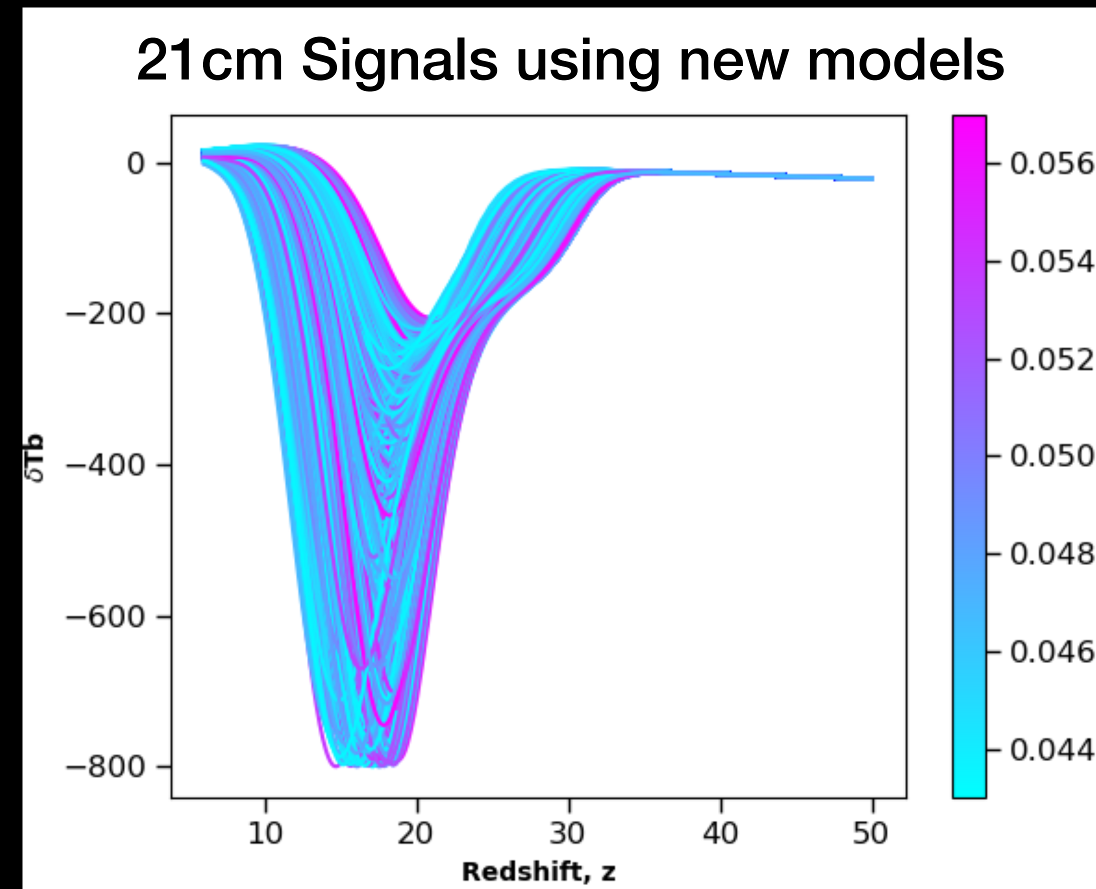
- **The dark-shaded areas show parameter combinations that, additionally, match the brightness temperature measured by EDGES (–500 ± 75mK).**

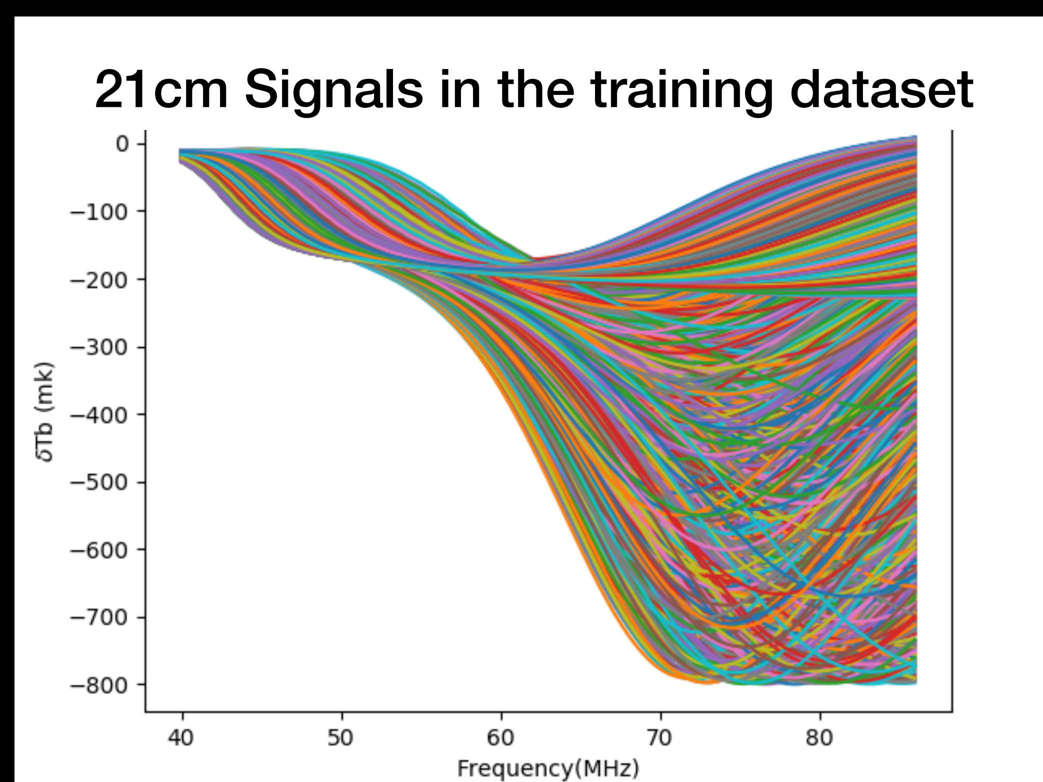
Ref: Chatterjee et al 2019

In this work, the authors have ruled out the 3keV WDM models using EDGES data

New training dataset

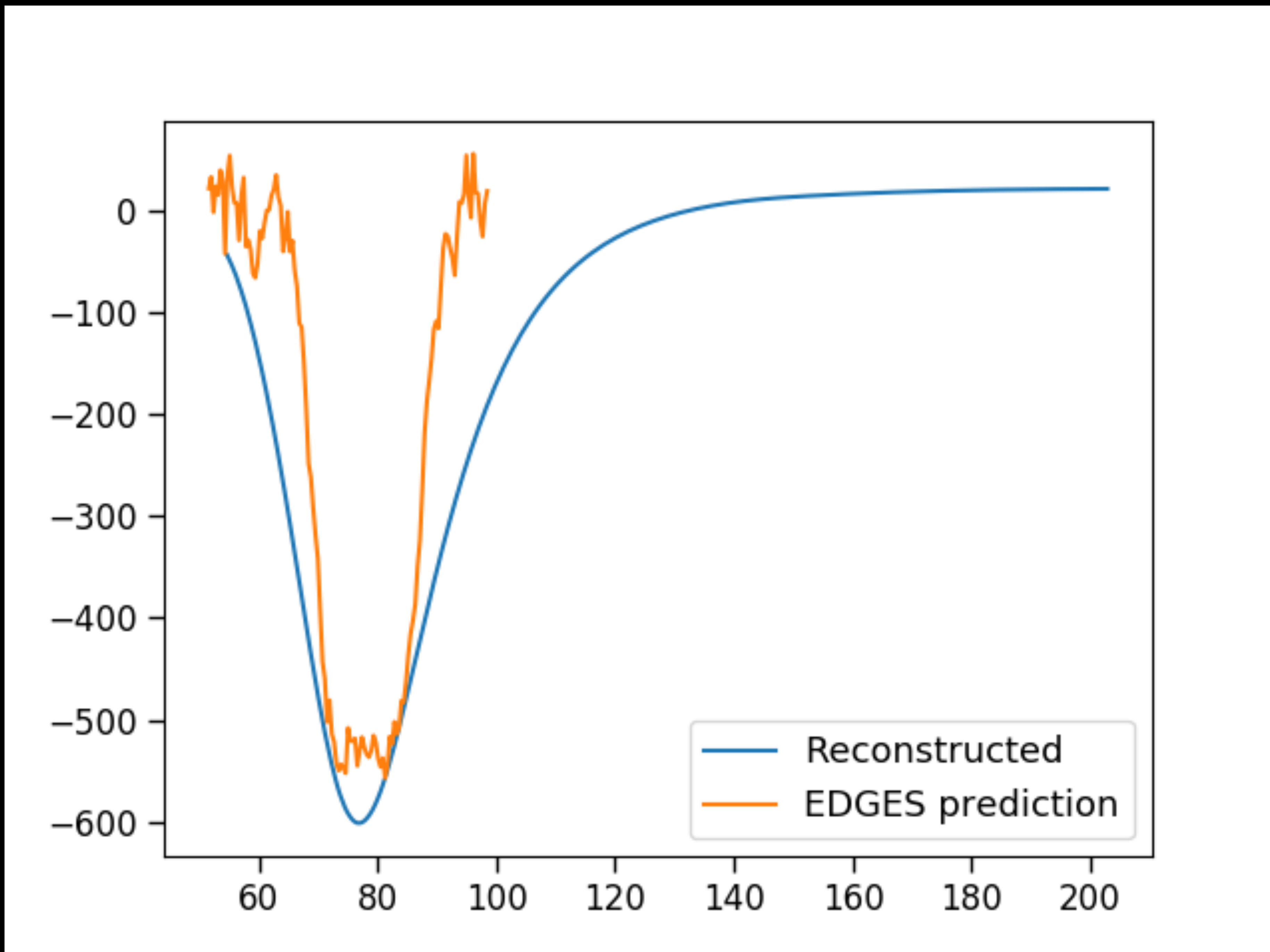






Choudhury, M. et al. (in prep)

Results - work in progress



Predicted parameters

$f_X = 15.29$

$f_* = 0.003$

$f_{esc} = 0.002$

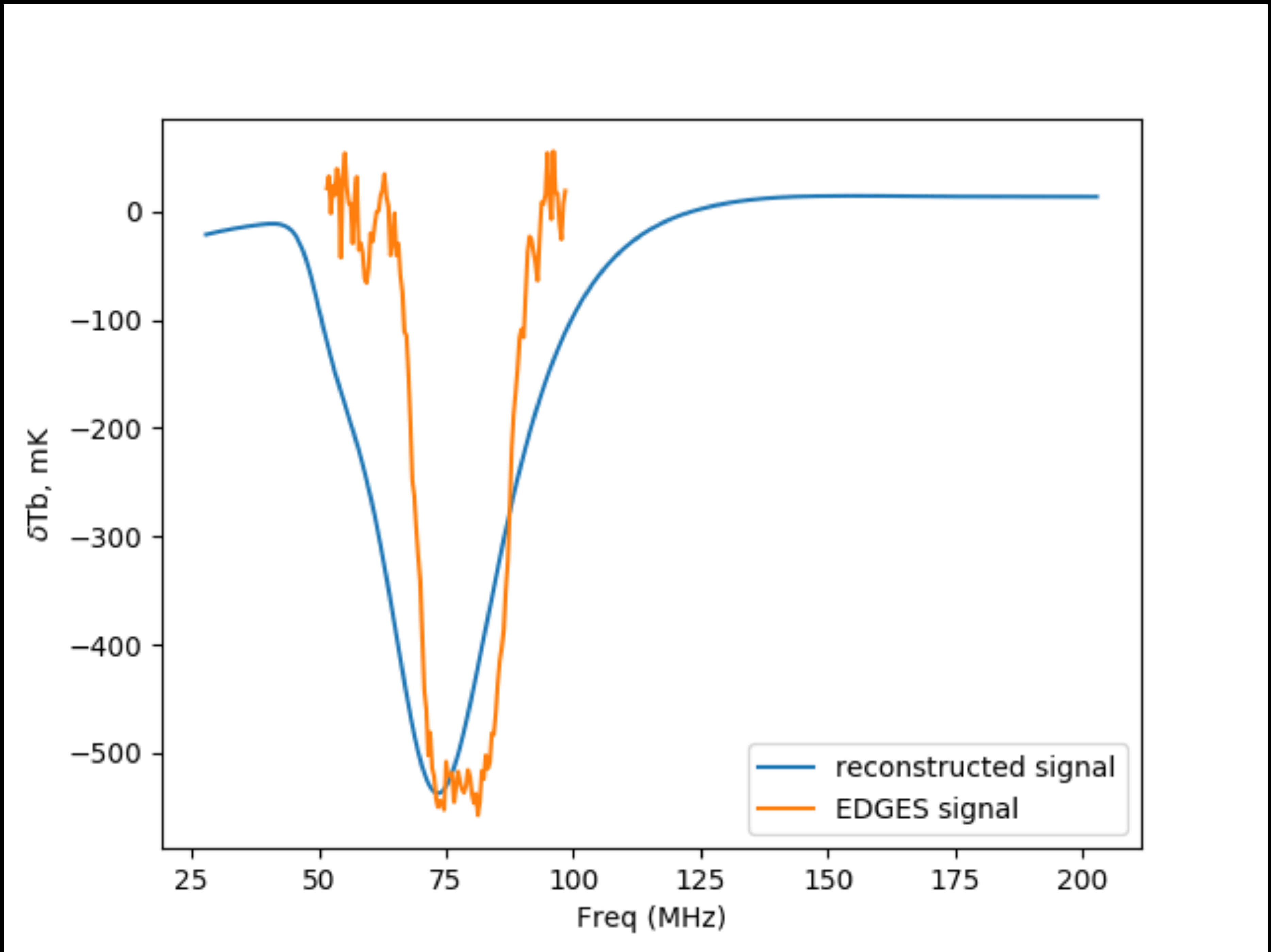
$fXh = 0.002$

$f_R = 0.35e4$

$N_\alpha = 9023$

EDGES DATA AS INPUT, WHEN TRAINED WITH ONLY EXOTIC MODELS + FOREGROUNDS (AS IN BOWMAN2018)

Results - work in progress



Predicted parameters

$f_X = 16.24$

$f_* = 0.006$

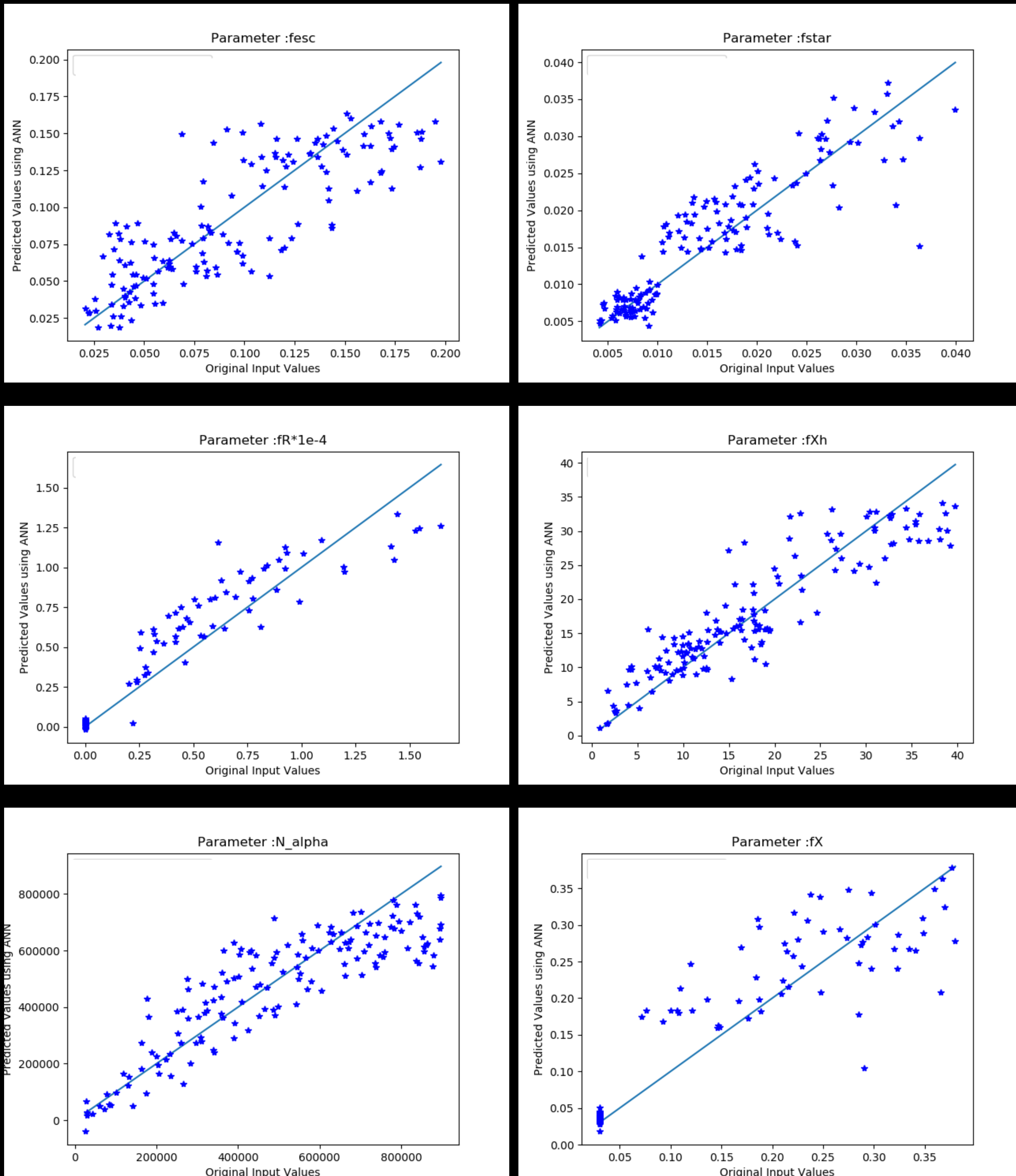
$f_{esc} = 0.139$

$fXh = 0.3$

$N_\alpha = 9728$

$f_R = 1e4$

EDGES DATA AS INPUT, WHEN TRAINED WITH ALL MODELS + FOREGROUNDS (AS IN BOWMAN2018)



PREDICTIONS WITH RMSES AS :

$f_X \rightarrow 0.35$

$f_* \rightarrow 0.26$

$f_{esc} \rightarrow 0.40$

$fXh \rightarrow 0.42$

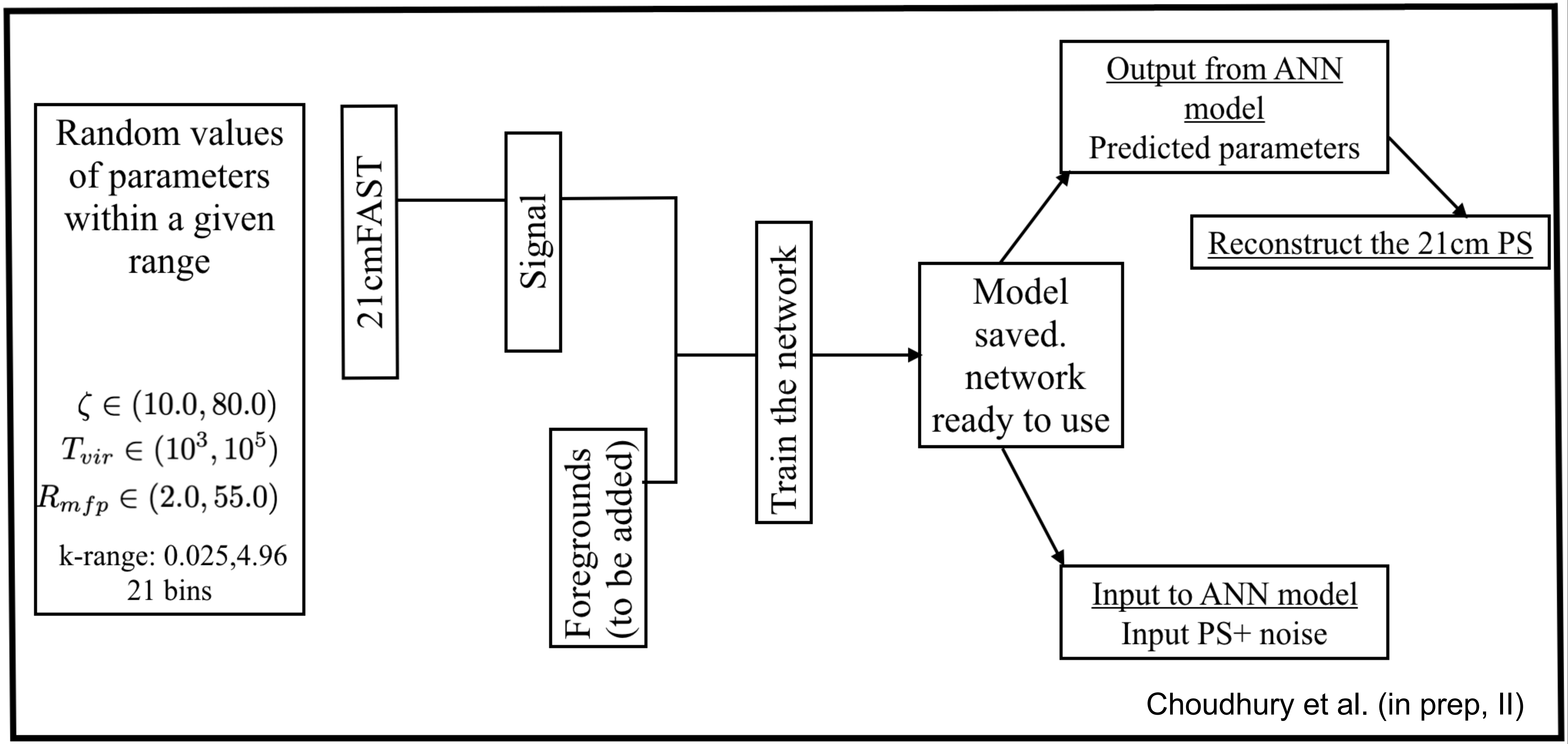
$N_\alpha \rightarrow 0.38$

$f_R \rightarrow 0.36$

Extracting the HI 21cm Power Spectra

Flowchart

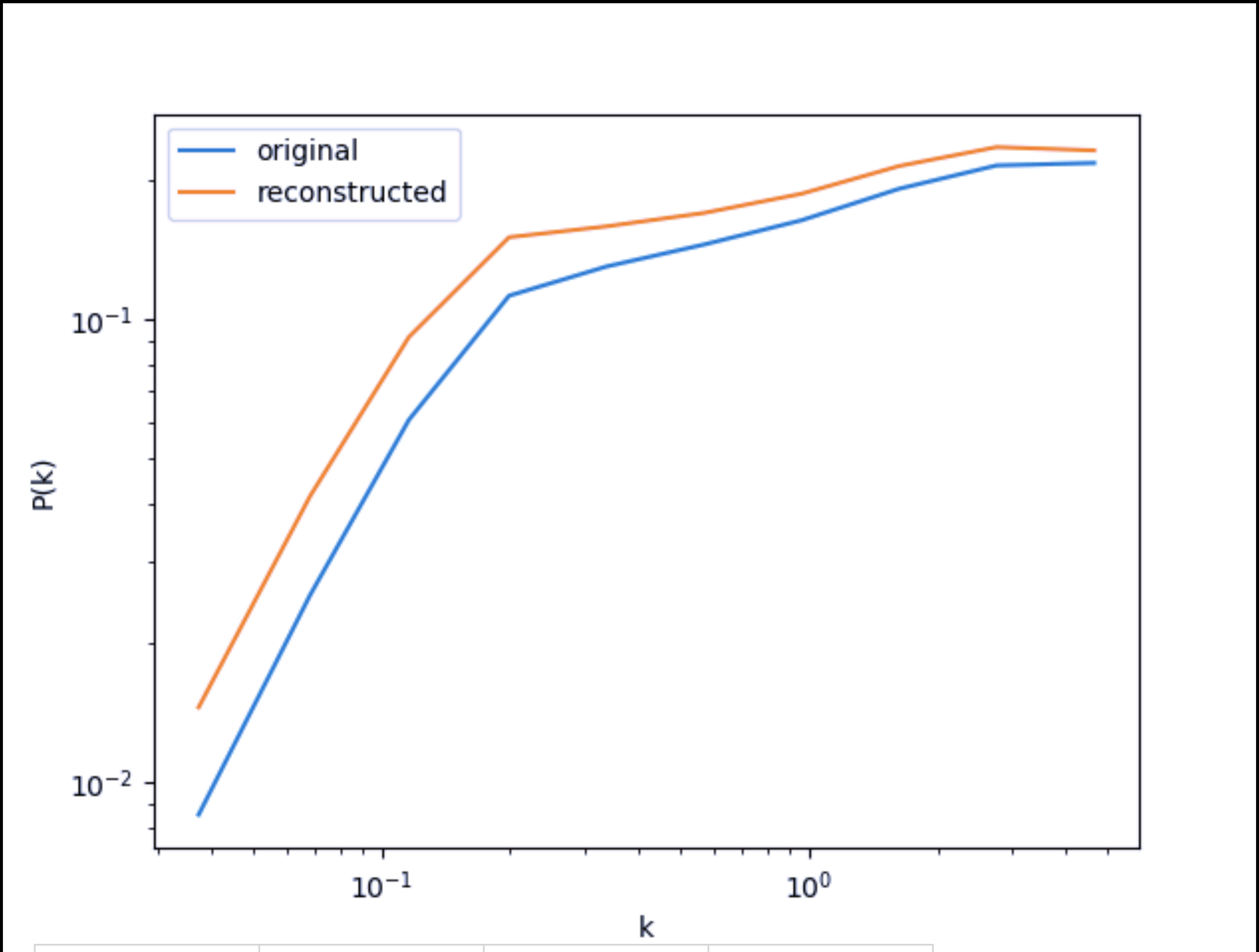
Power Spectrum Detection – Using ANN



Choudhury et al. (in prep, II)

Reconstruction of 21cm Power spectra

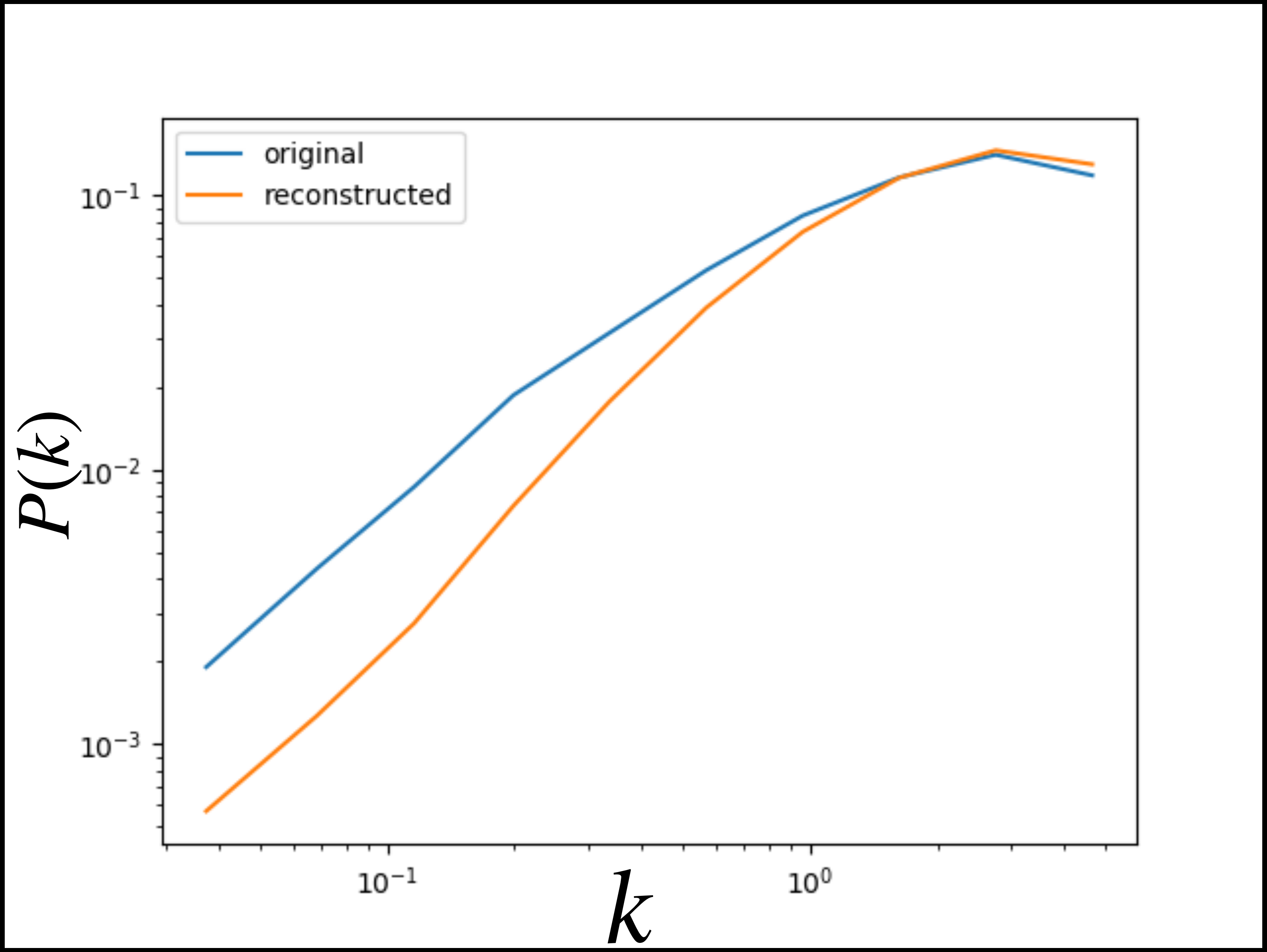
without any foregrounds



	Nion	Rmfp	M_min
Original	50	25	9.7*e8
Predicted	57.0	31.65	11.1*e8

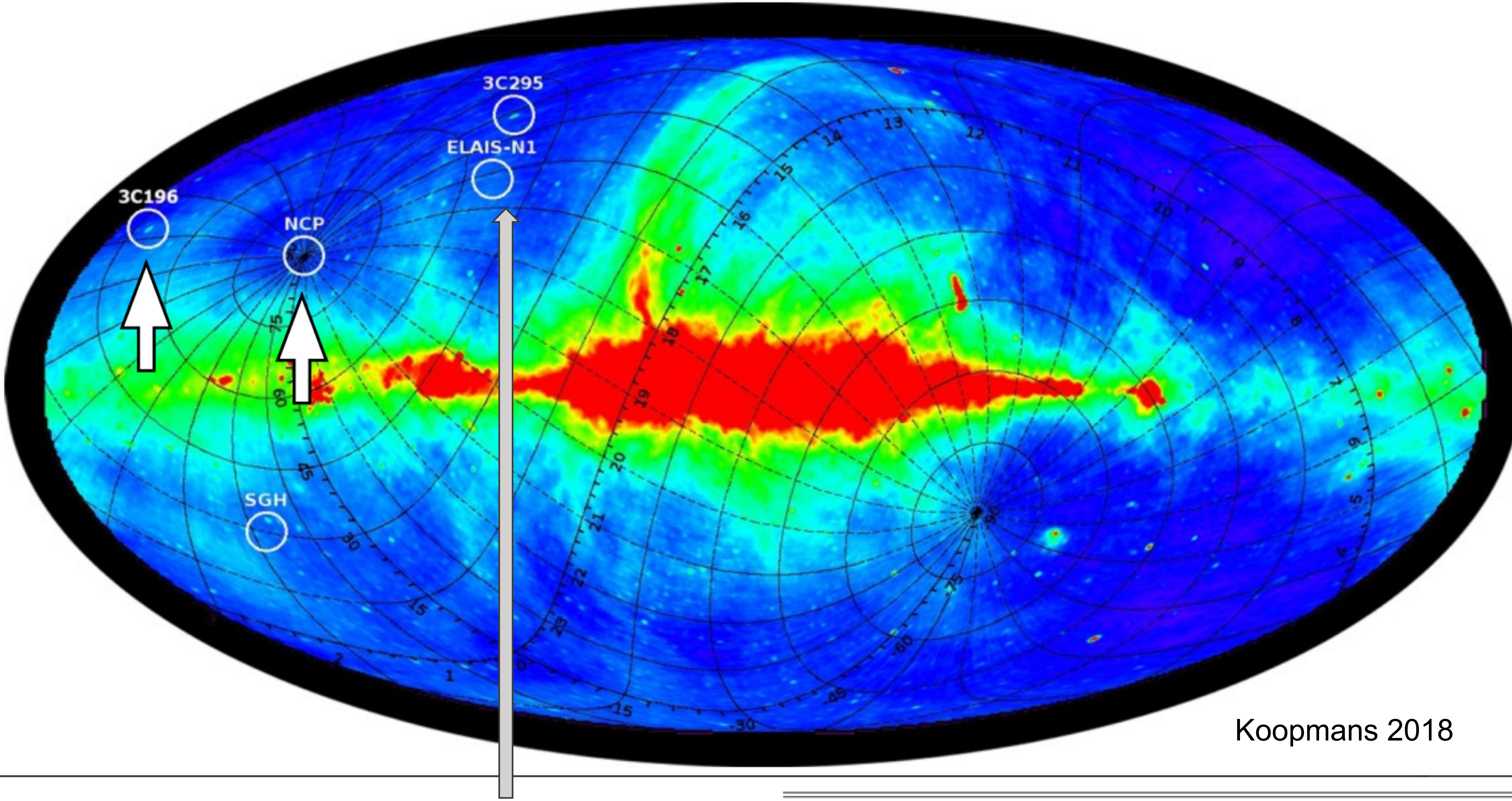
error:~30-40% for the signal parameters

Reconstruction of 21cm Power spectra
in presence of foregrounds



	Nion	Rmfp	M_min
Original	30	45	5.435 e 9
Predicted	53.2	52.6	8.6 e 9

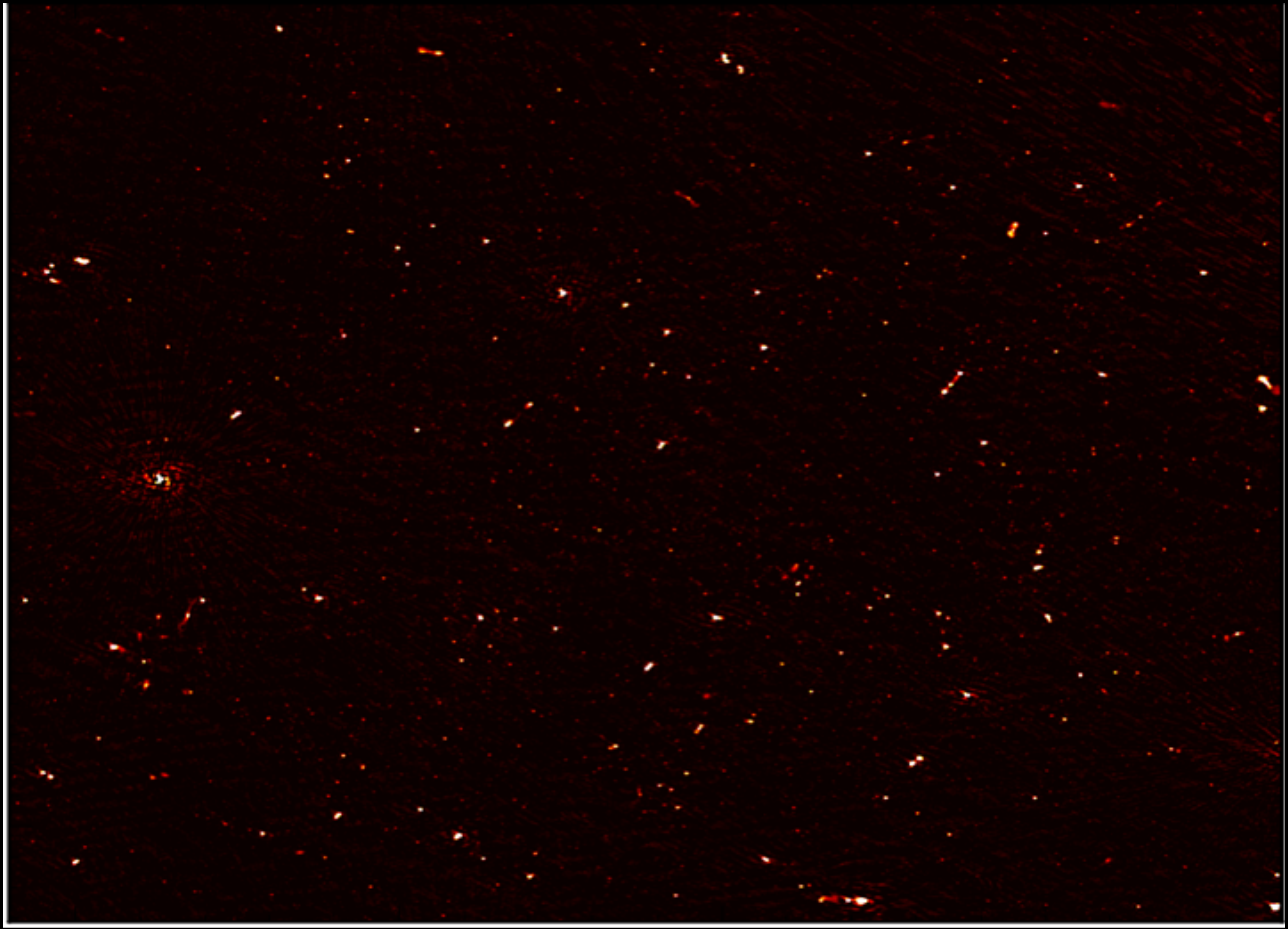
Work in progress - further optimization of the network is underway



Koopmans 2018

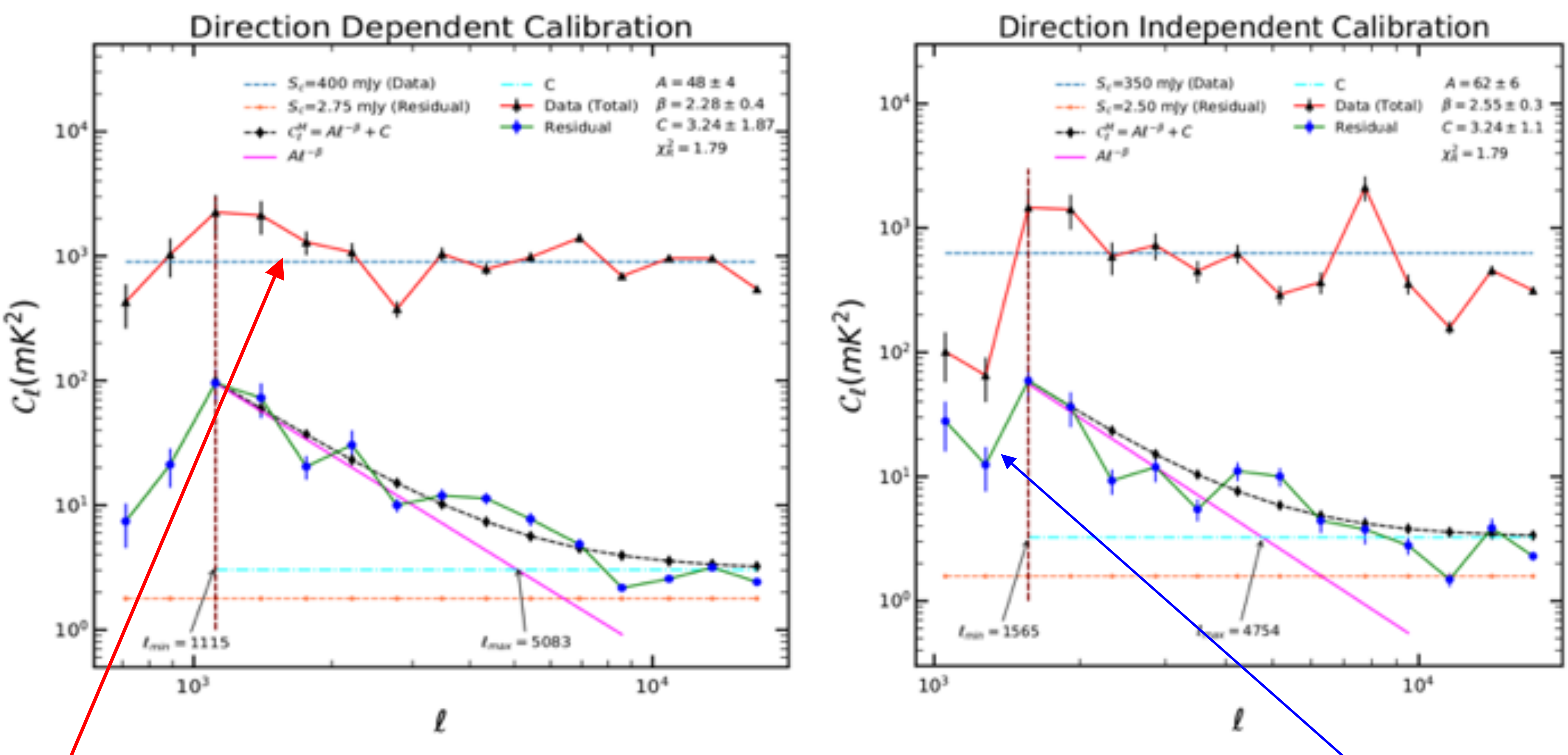
	GWB	GSB
Working antennas	28	28
Central Frequency	400 MHz	325MHz
Bandwidth	200 MHz	32MHz
Visibility integration time	2 sec	8 sec
Number of Channels	8192	512
Total Observation time	25 hours	25 hours

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	26
Pointing centres	$13^h31^m08^s$ $+30^d30^m32^s$ (3C286) $15^h49^m17^s$ $+50^d38^m05^s$ (J1549+506) $16^h10^m01^s$ $+54^d30^m36^s$ (ELAIS N1) $01^h37^m41^s$ $+33^d09^m35^s$ (3C48)



$RMS \sim 15\mu Jy/beam$ $\quad$ $Dynamic\ range \sim 18,000$

Foreground Power Spectra- Taper Grided Estimator



f = 1

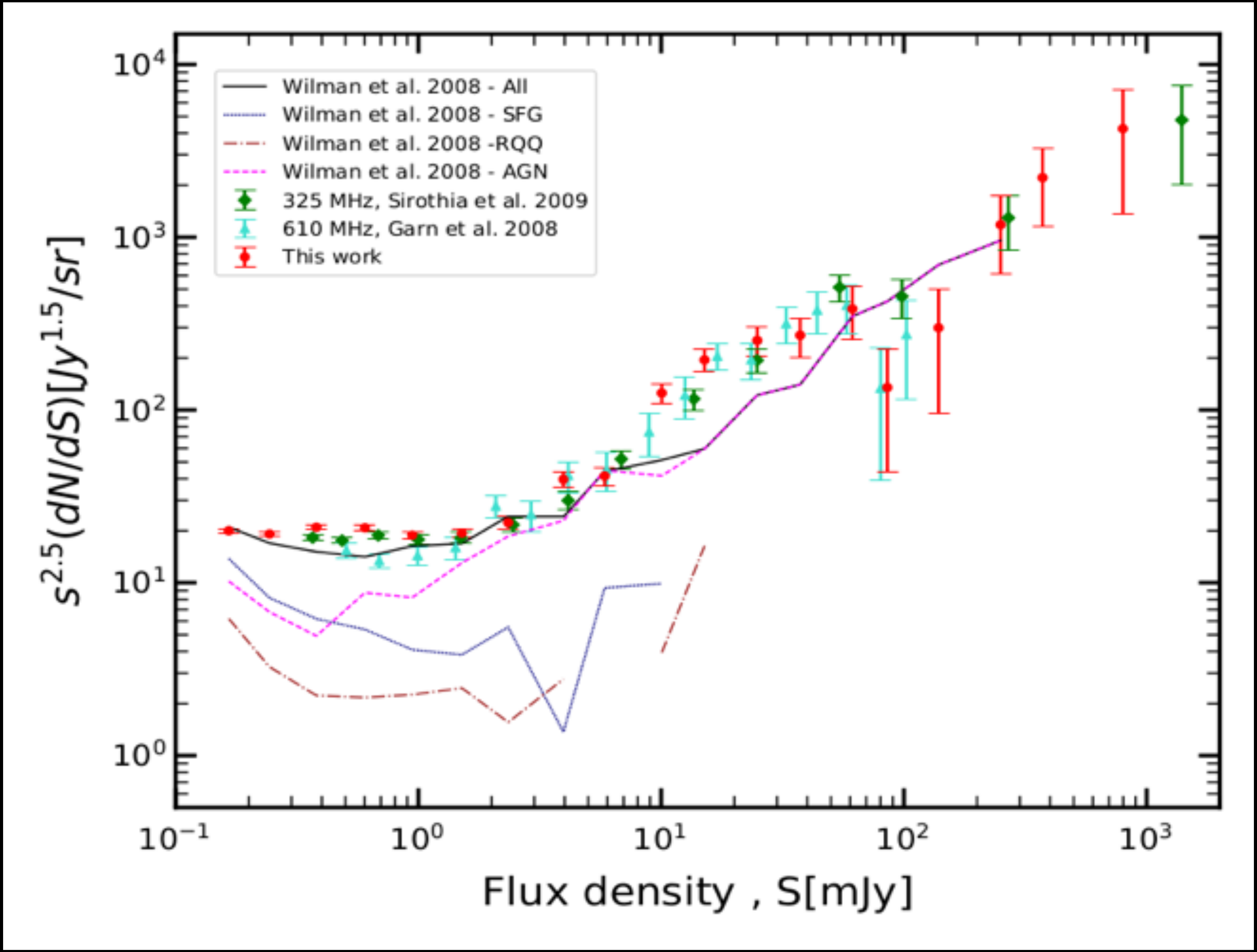
f = 0.5

Data = point source + DGSE + artefacts

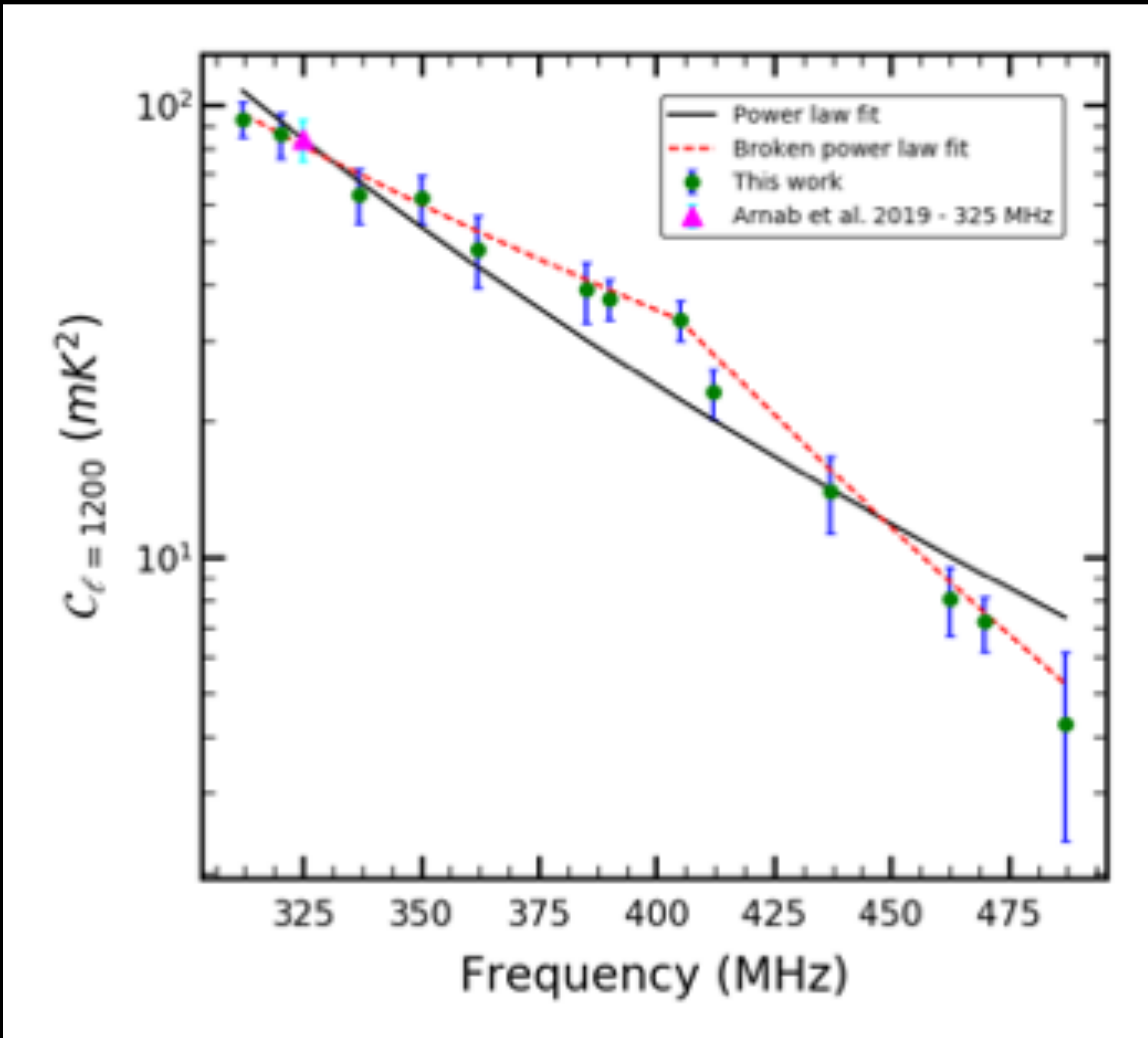
After subtraction of point source model :
Data = DGSE + residual point sources + artefacts

Chakraborty, A. et al. (2019a)

Euclidian-normalized differential source counts



Spectral Behaviour of Foreground



Chakraborty, A. et al. (2019b)

$$\mathcal{C}_{\ell=1200}(\nu) = \begin{cases} A\left(\frac{\nu}{\nu_{break}}\right)^{-2\alpha_1}, \text{for} \nu < \nu_{break} \\ A\left(\frac{\nu}{\nu_{break}}\right)^{-2\alpha_2}, \text{for} \nu > \nu_{break} \end{cases}$$

$$\alpha_1 = 2.1 \pm 0.2$$

$$\alpha_2 = 4.8 \pm 0.4$$

$$\chi^2_{red} = 0.3$$

Conclusions

- ANN as an alternate to the Bayesian Framework for 21cm signal extraction.
- Preliminary Results shows agreement with EDGES results
- Foregrounds can be interesting as well !