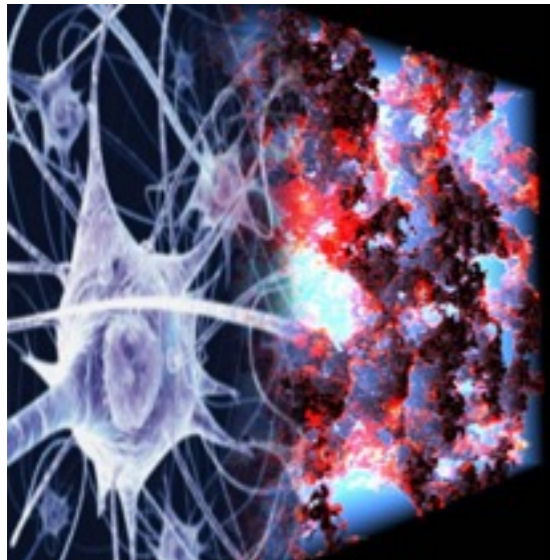


Peering towards cosmic dawn (2/10/2017-6/10/2017)

Analyzing 21cm signal at EoR with Artificial Neural Networks

Hayato Shimabukuro (Observatoire de Paris)

Based on Shimabukuro & Semelin
([astro-ph/1701.07026](https://arxiv.org/abs/astro-ph/1701.07026), Published in MNRAS)



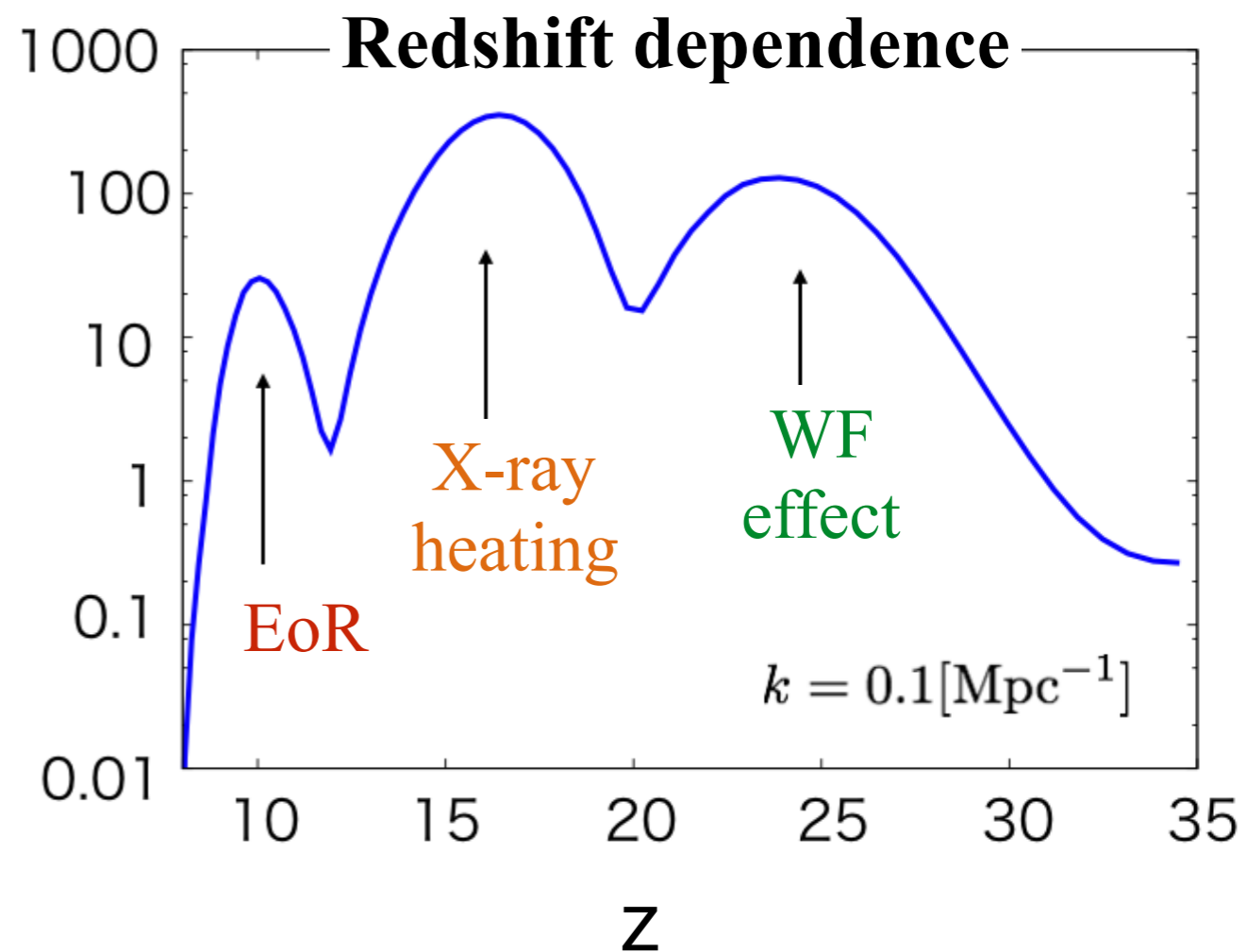
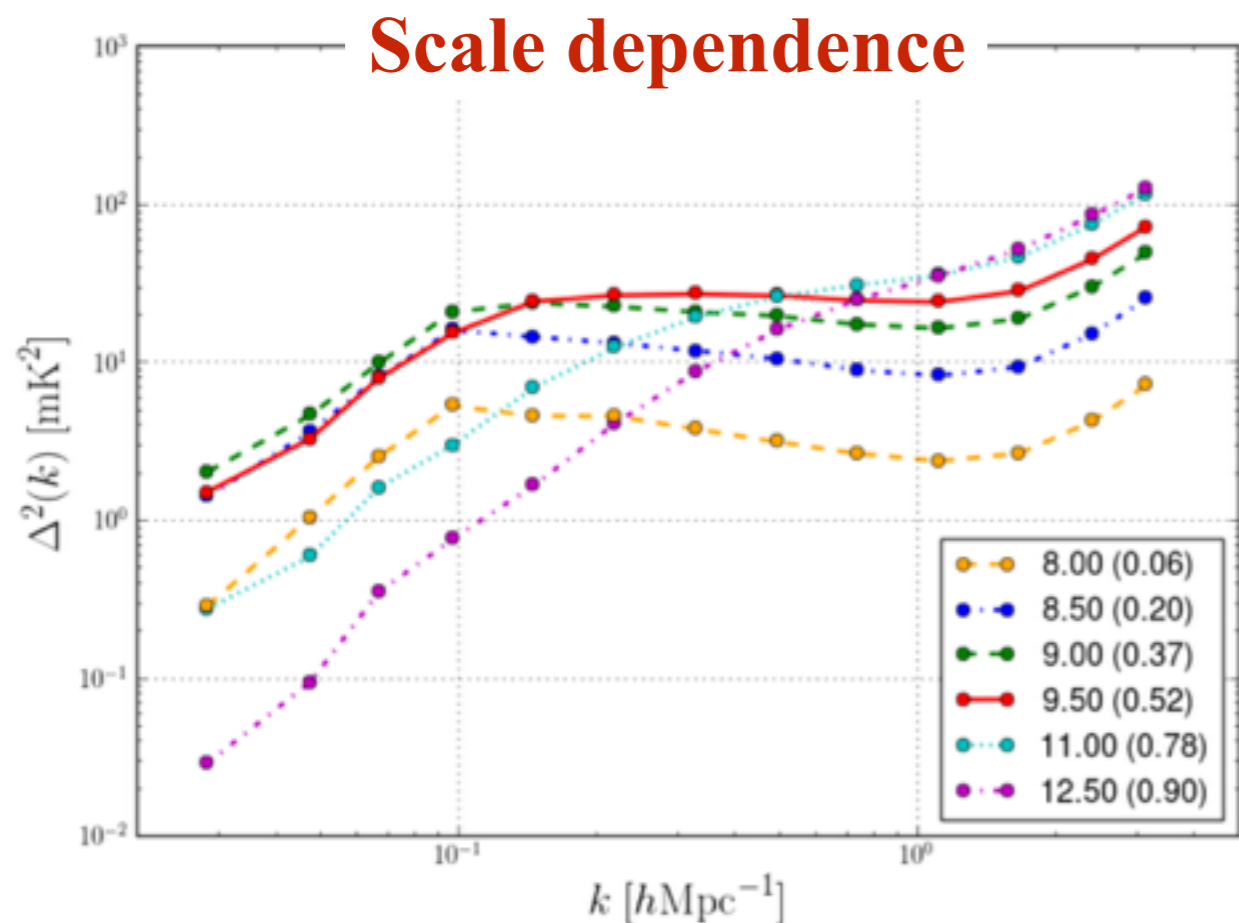
Cosmological 21 cm signal

21cm power spectrum

We often evaluate statistical property of the 21cm fluctuations by **power spectrum**.

21cm power spectrum (PS)

$$\langle \delta T_b(\mathbf{k}) \delta T_b(\mathbf{k}') \rangle = (2\pi)^3 \delta(\mathbf{k} + \mathbf{k}') P_{21}$$



Pober et al (2014)

Artificial neural networks

21cm signal with machine learning

“**Machine learning**” attracts attention as a method to tackle huge data.

Recently, some works apply machine learning techniques to analyse 21cm signal.

(e.g.)

- **Gaussian process**

(Kern et al. 2017)

- **Artificial Neural Network**

(Schmit et al. 2017)

Emulating 21cm power spectrum



Reduce calculation cost !!



easier calculation for MCMC

What is artificial neural network (ANN)?

ANN is one of the methods inspired by brain neural network which is used to establish approximate function between input and output data.

STEP

1. Prepare **known** data set (**training data**) \longrightarrow $(\vec{x}_{data}, \vec{y}_{data})$

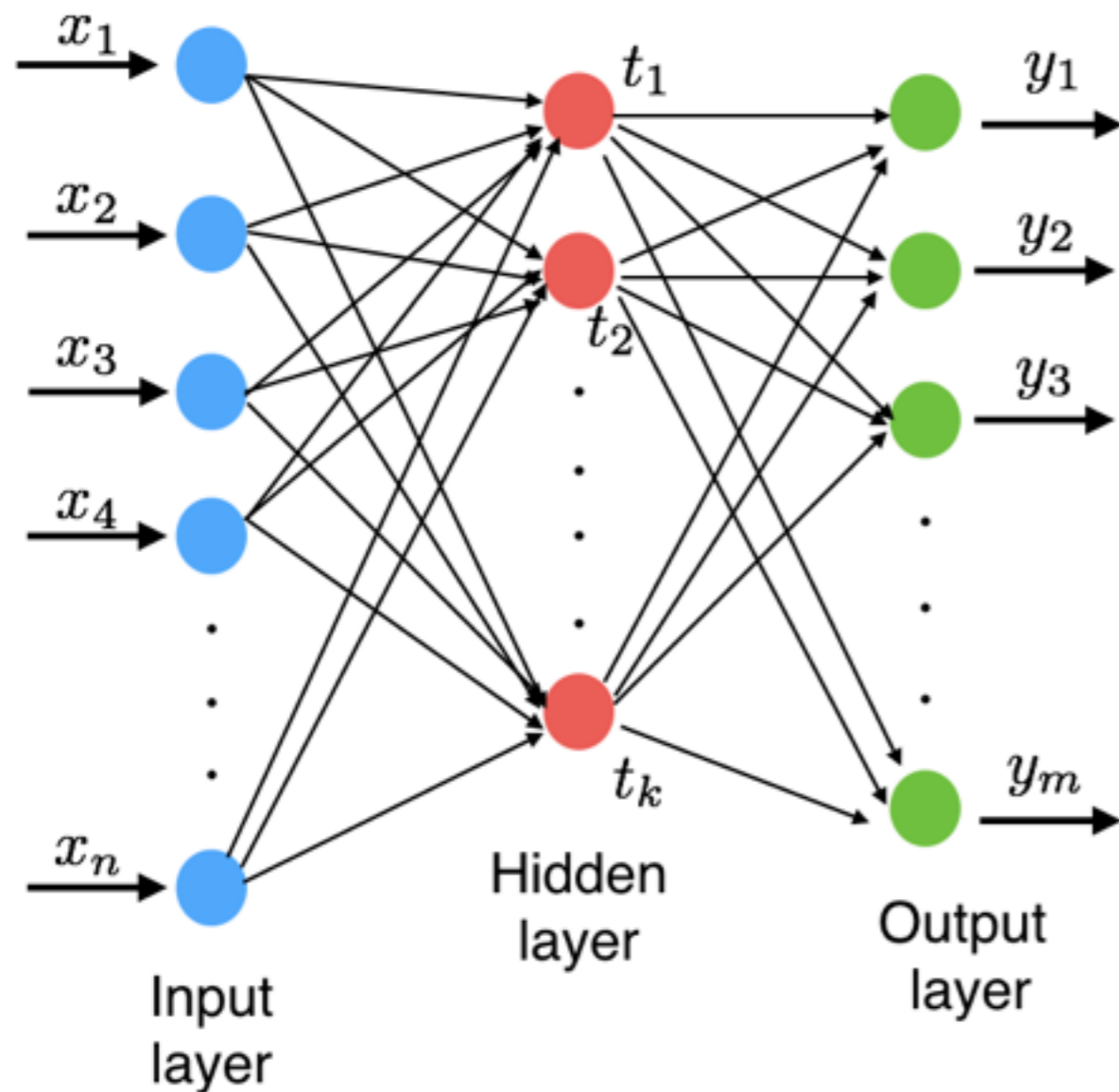
2. Train architecture of neural network by training data.

\longrightarrow $y = f(x)$

3. Apply trained network to unknown (**test data**) and can obtain expected output data.

\longrightarrow $y_{ANN} = f(x_{test})$

Artificial Neural Network (ANN)



- ANN consists of input layer, hidden layer and output layer. Each layer has neurons.

- By learning from training data, ANN can approximate any function which associates input and output values.

$$y = f(x)$$

- Applying trained network to unknown data in order to obtain expected value.

$$y_{\text{ANN}} = f(x_{\text{test}})$$

Train the network

We have to determine the weight to construct architecture of neural network.



How?

Answer

We adjust the weight to satisfy **minimizing the cost function E** which is difference between true value and output value. This procedure is called “**Training**”

cost function

$$E = \frac{1}{2} \sum_i^N (y_{\text{data},i} - y_{\text{ANN}})^2$$

N: the number of training datasets

We use “**Back propagation algorithm**” to determine weight. (Rumelhart et al 1986)

Result

Dataset

$$\vec{d} = [P(k), \vec{\theta}]$$



$$\theta_{\text{EoR}} = f(P_{21})$$

EoR parameter (**output**)

Inverse problem

21 cm power spectrum (**input**) ($0.04 \text{Mpc}^{-1} \leq k \leq 1.4 \text{Mpc}^{-1}$)

70 training datasets and 50 test datasets

(14 bins)

EoR Parameter

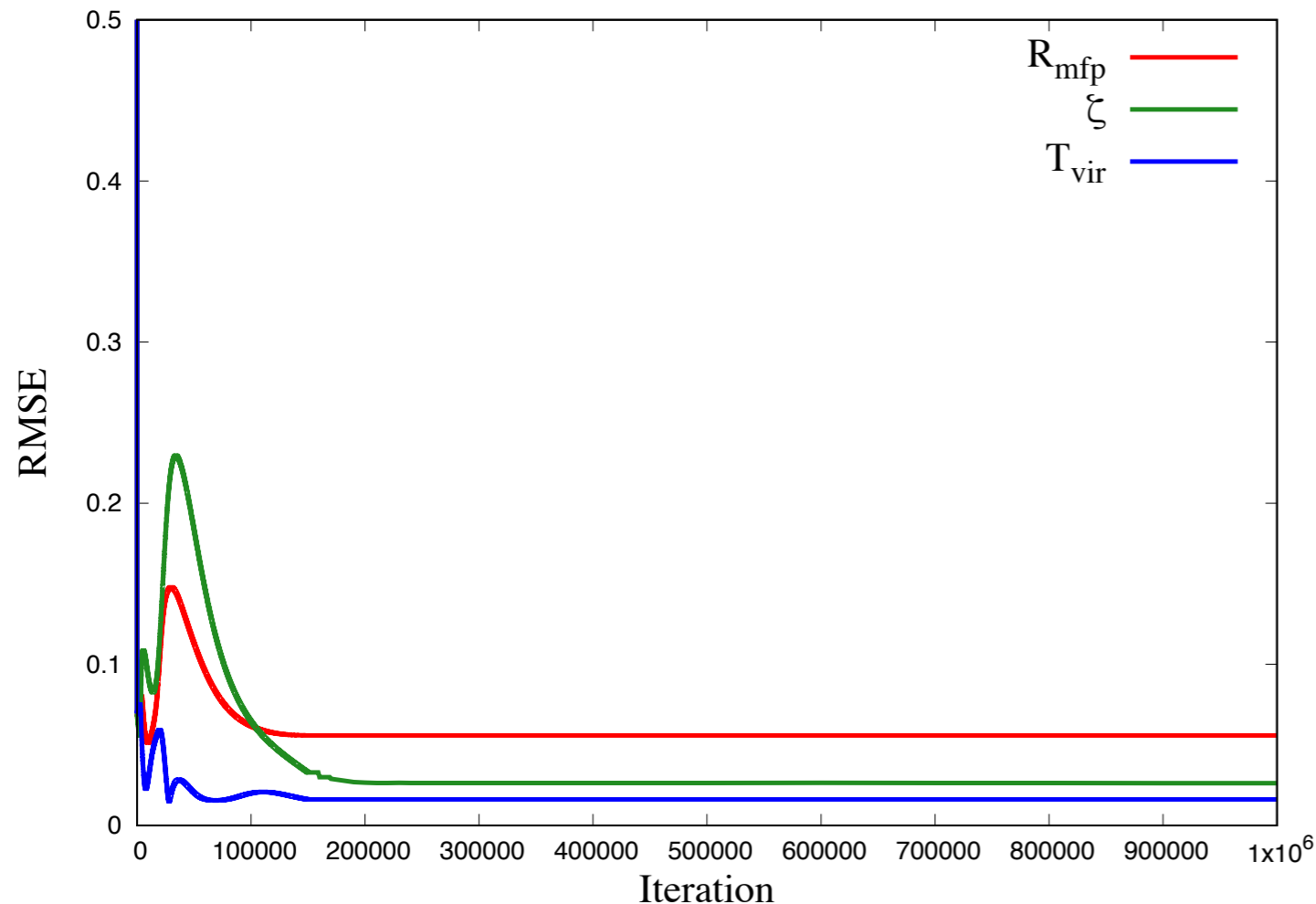
ζ : the ionizing efficiency.

T_{vir} : the minimum viral temperature of halos producing ionizing photons

R_{mfp} : the mean free path of ionizing photons through the IGM
(Maximum HII bubble size)

Convergence

We perform 10^6 iterations for back-propagation algorithm to see convergence.



$$\text{RMSE} = \sqrt{\frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} X^2}$$

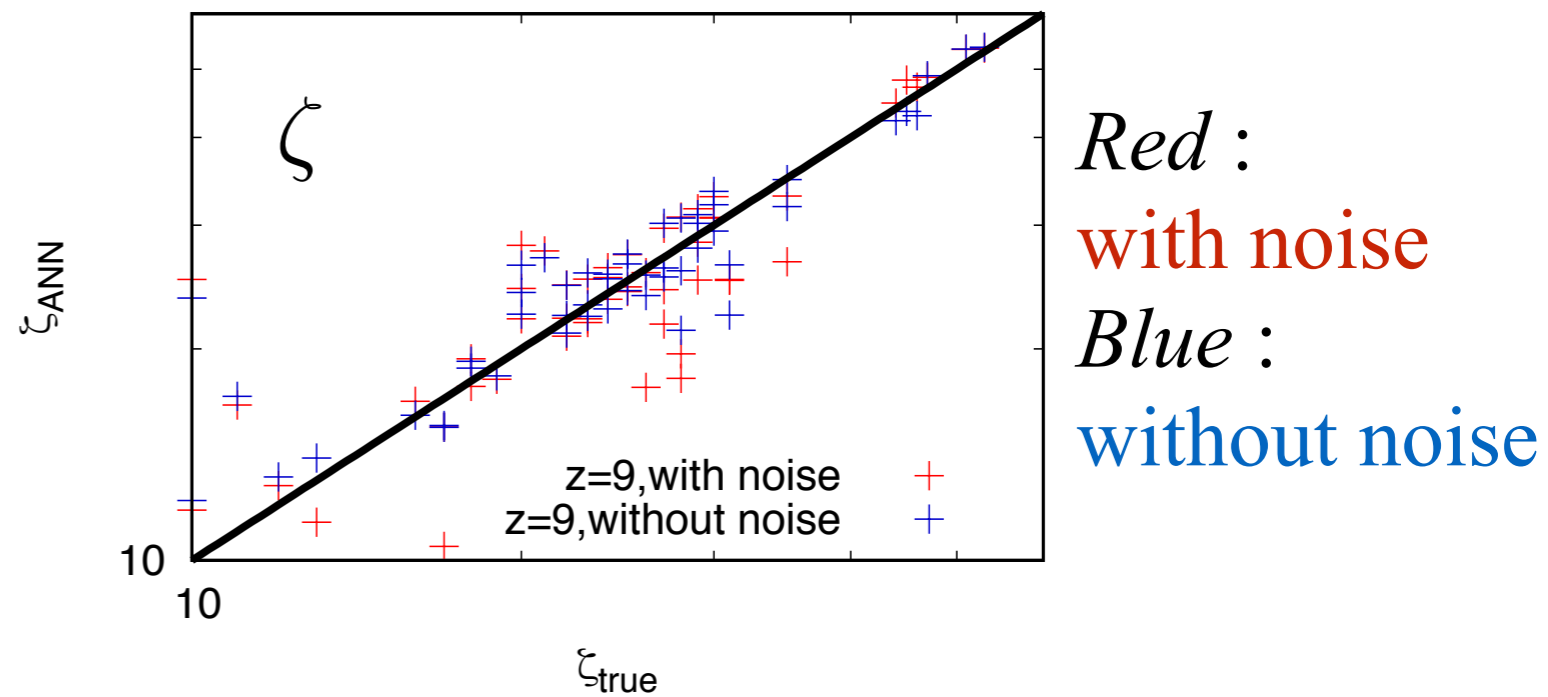
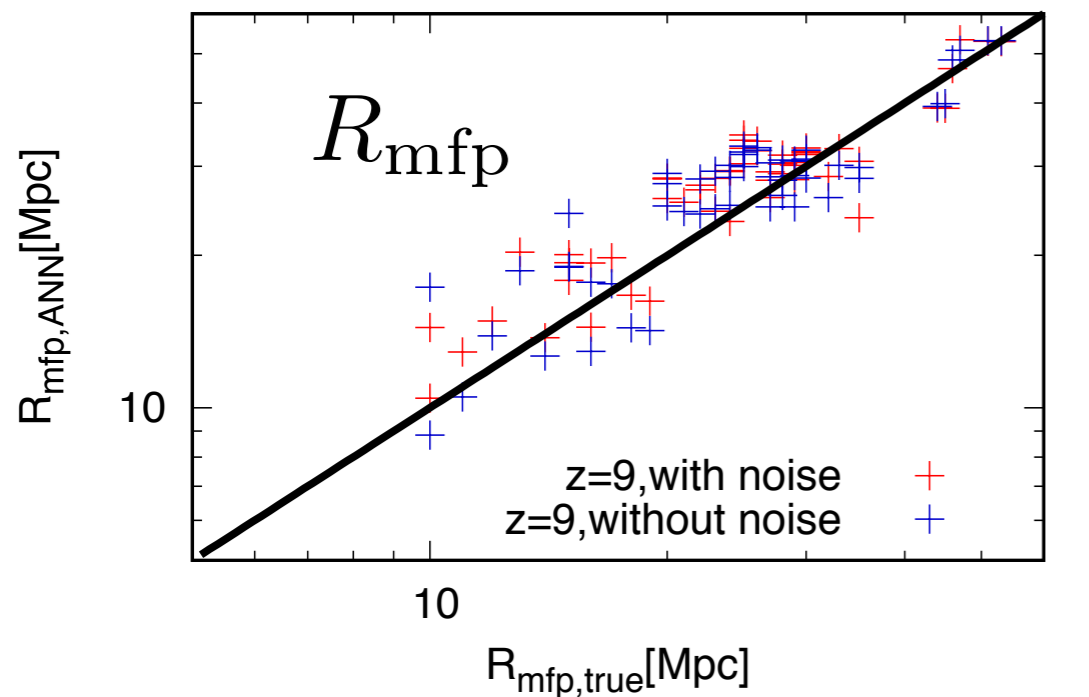
$$X = (\theta_{\text{ANN}} - \theta_{\text{data}}) / \theta_{\text{data}}$$

(θ is each EoR parameter.)

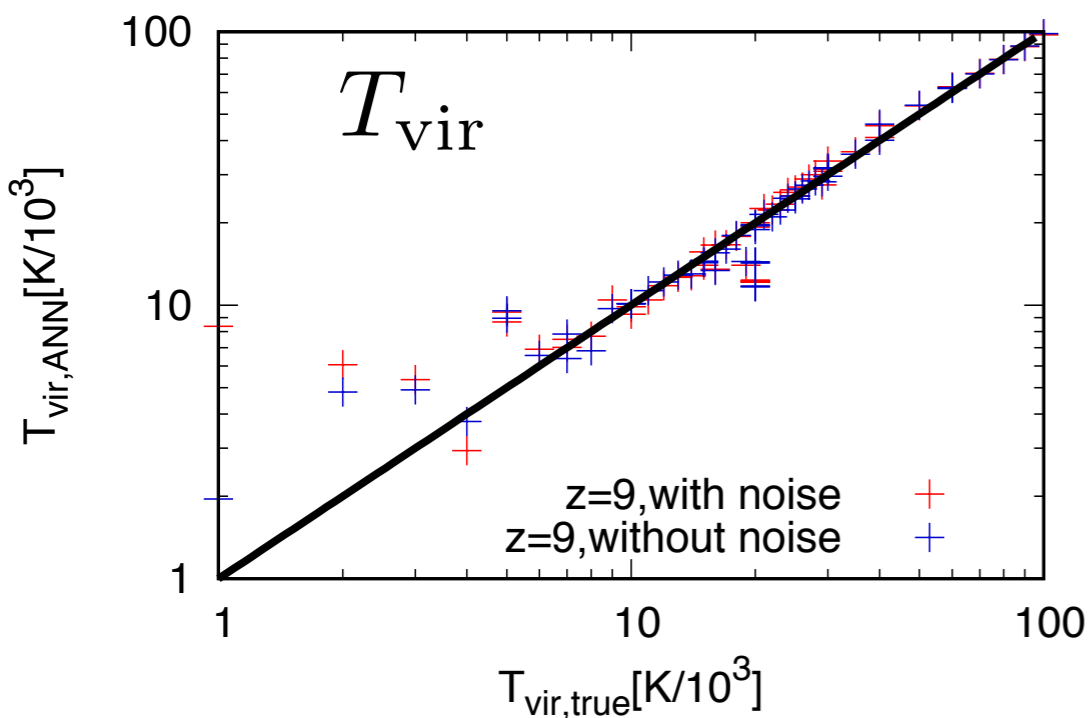
More than 200000 iterations, all values of RMSE converge.

Results

$z=9$, PS including **thermal noise**



Red :
with noise
Blue :
without noise

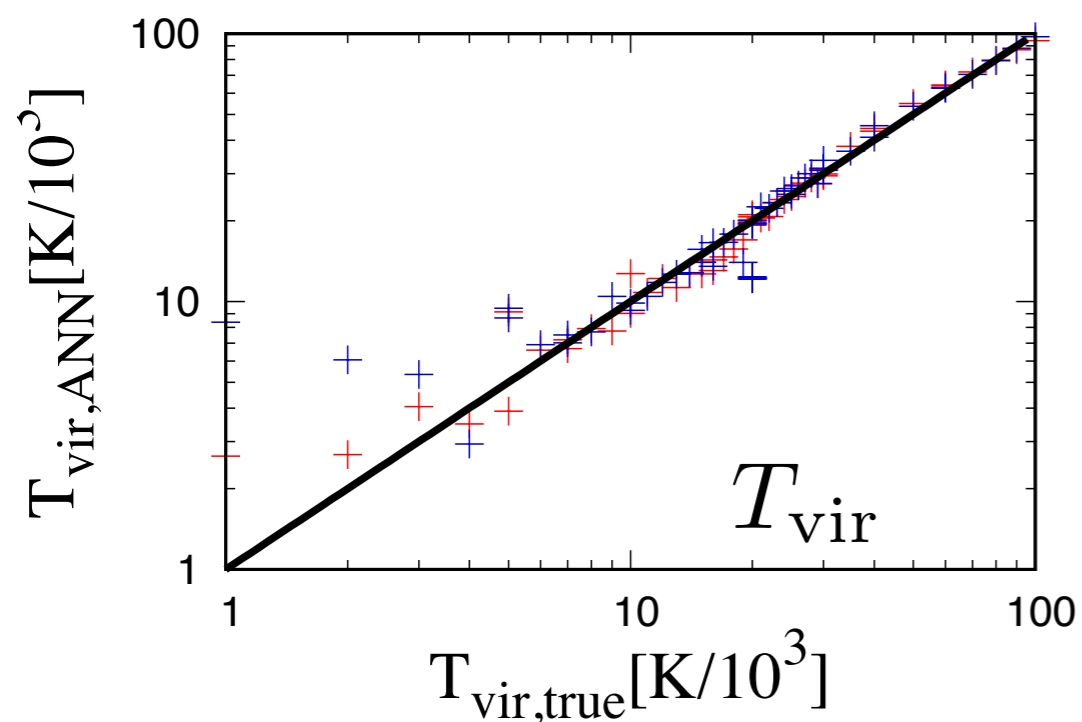
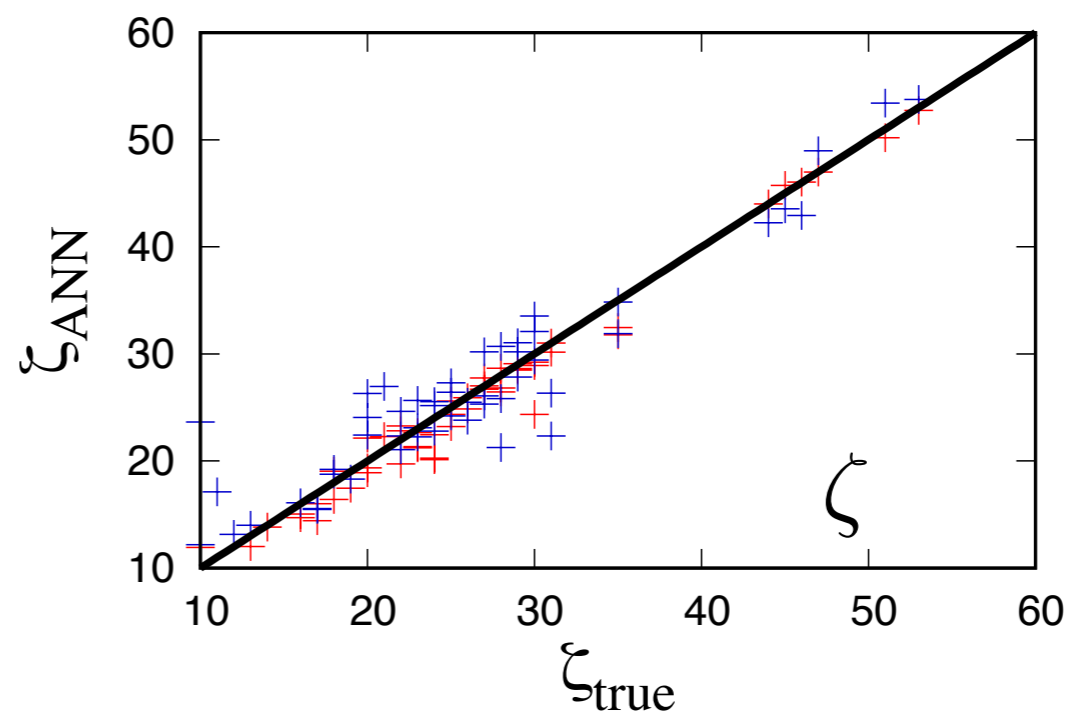
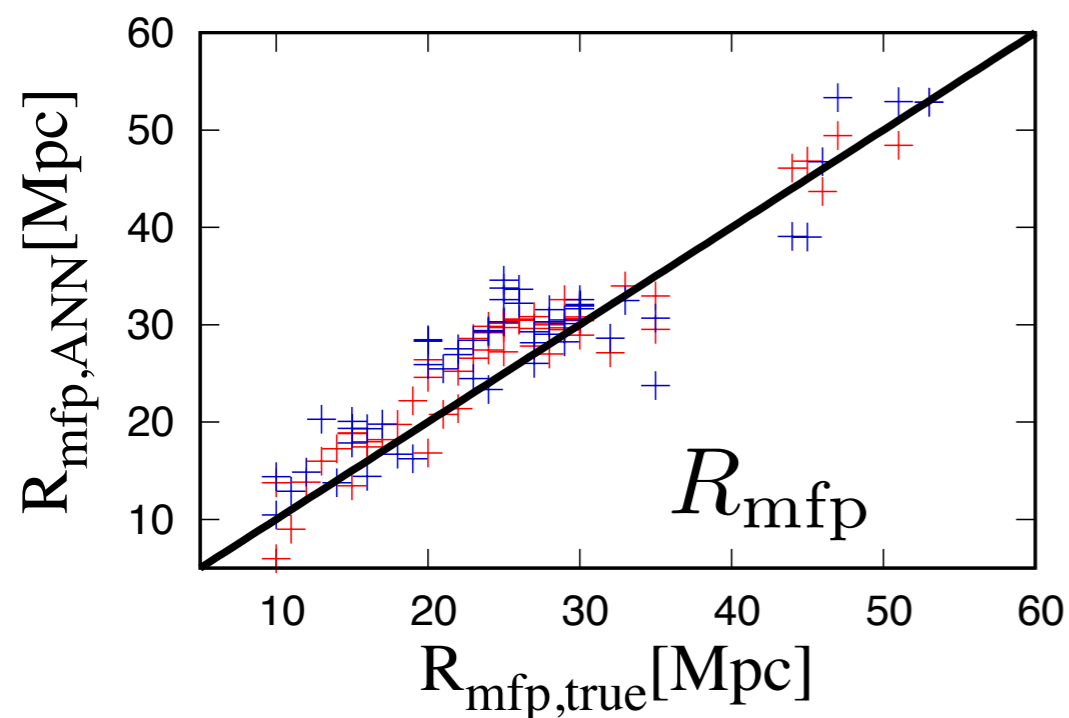


	RMSE _{wo/noise}	RMSE _{w/noise}
R_{mfp}	0.228	0.258
ζ	0.271	0.288
$\log(T_{\text{vir}})$	0.027	0.038

Accuracy is better when noises are not included.

Results

z=9, 10, 11. PS including **thermal noise**



42 neurons, 100000 iterations

Red : z=9,10,11

Blue : z=9

Visually, it is difficult to see whether the accuracy of parameter estimation is improved by taking redshift evolution into account.

Results

single z

multiple z

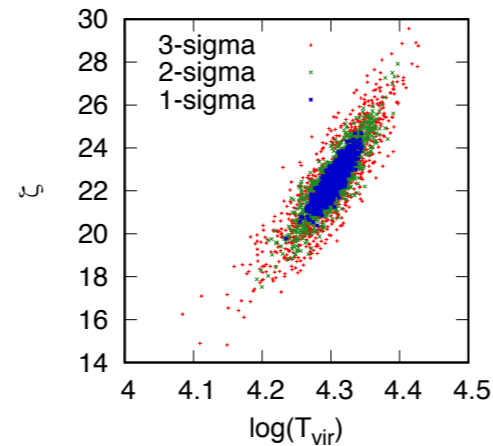
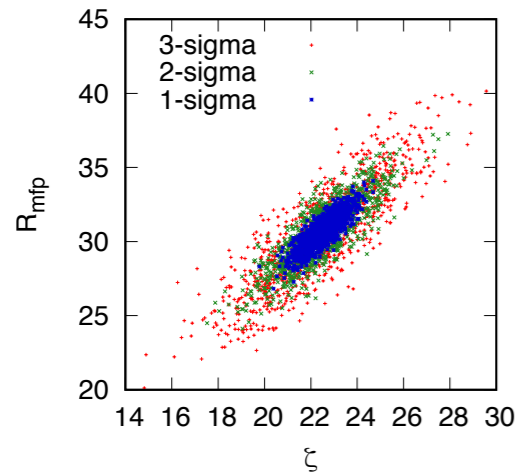
	RMSE _{wo/noise}	RMSE _{w/noise}	RMSE _{w/noise,zevolution}	RMSE _{w/noise,reduced}
R_{mfp}	0.228	0.258	0.172	0.262
ζ	0.271	0.288	0.168	0.290
$\log(T_{\text{vir}})$	0.027	0.038	0.019	0.029

Reduce the number of training datasets (N=20)

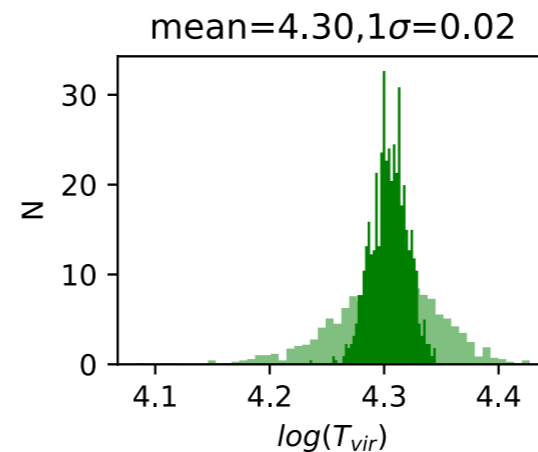
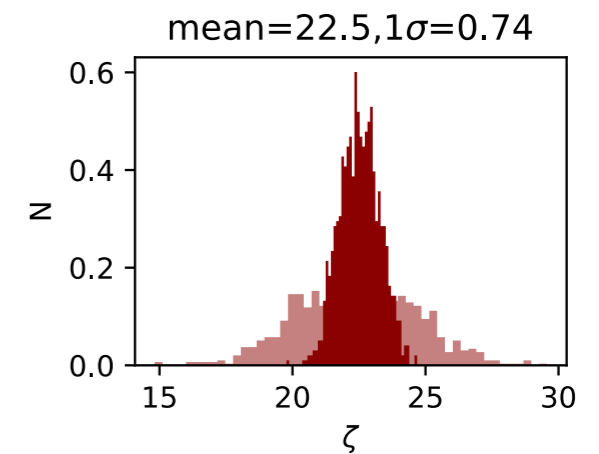
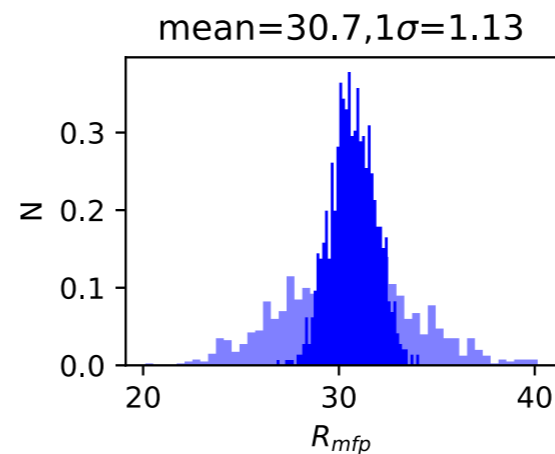
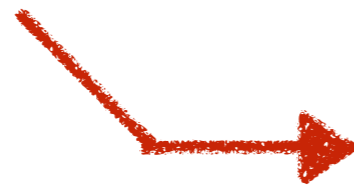
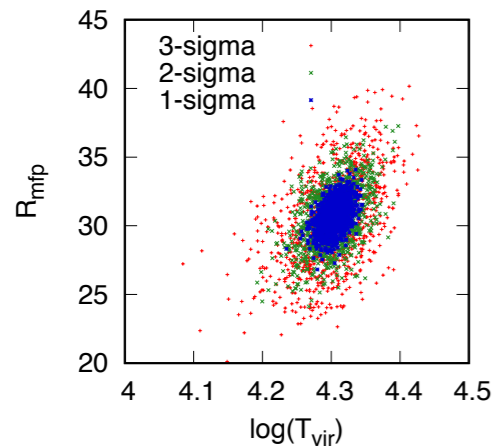
Estimate of the EoR parameter is improved by

- **Taking redshift evolution into account.**
- **Increasing the number of training datasets**

PDF for estimated parameters



1000 realizations for each parameter set with thermal noise drawn from gaussian distribution.



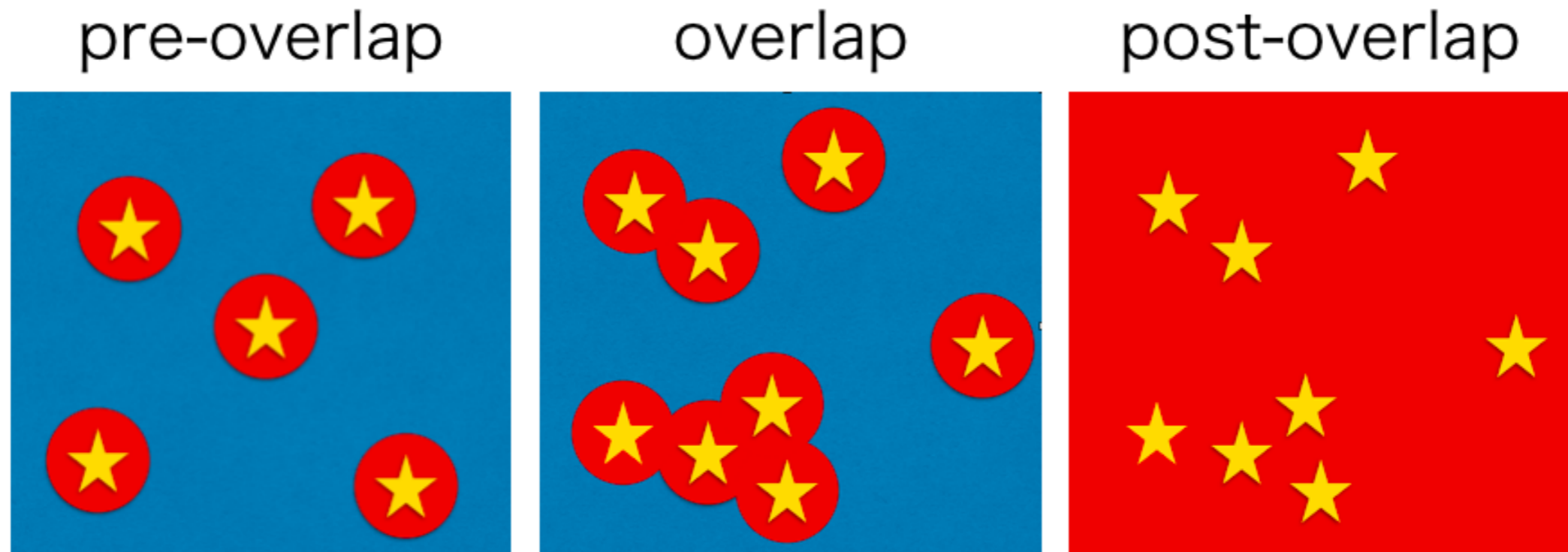
$$(R_{mfp}, \log(T_{vir}), \zeta) = (30, 4.3, 22)$$

Summary

- Artificial neural network (ANN) is one of the machine learning techniques based on brain architecture model.
- We applied the ANN to 21cm signal in order to extract EoR information.
- EoR parameters produced by ANN were good agreement with true values.
- Multiple redshift data improved accuracy.
- Now I apply the ANN to 21cm signal at Cosmic dawn (Cohen et al 2017)

Backup

Epoch of Reionization



Blue : Neutral IGM **Red** : HII region **Yellow** : Ionizing source (galaxy)

Pre-overlap : HII regions grow in relative isolation

Overlap : Once galaxies become common, HII regions intersect.

Post-overlap : Ionising of IGM advances sufficiently.

Observational constraints on EoR

† The polarisation of CMB photons by free electron.

→ $z_r = 9.9^{+1.8}_{-1.6}$ (Assuming instantaneous reionization model)

(Planck collaboration 2015)

† High-z QSO absorption spectra

→ $x_{\text{HI}} > \sim 10^{-4} (z > 6)$ (Put constraints on neutral fraction)

(Fan et al 2006)

† The luminosity function of high-z LAE (Lyman-alpha emitter galaxies)

→ $x_{\text{HI}} = 0.3 - 0.8 (z = 7.3)$ (Put constraints on neutral fraction)

(Konno et al 2014)

Beyond current constraints

Current observations can constrain state of the IGM at **late** stage of the EoR.

We would like to observe the higher redshift universe more directly !

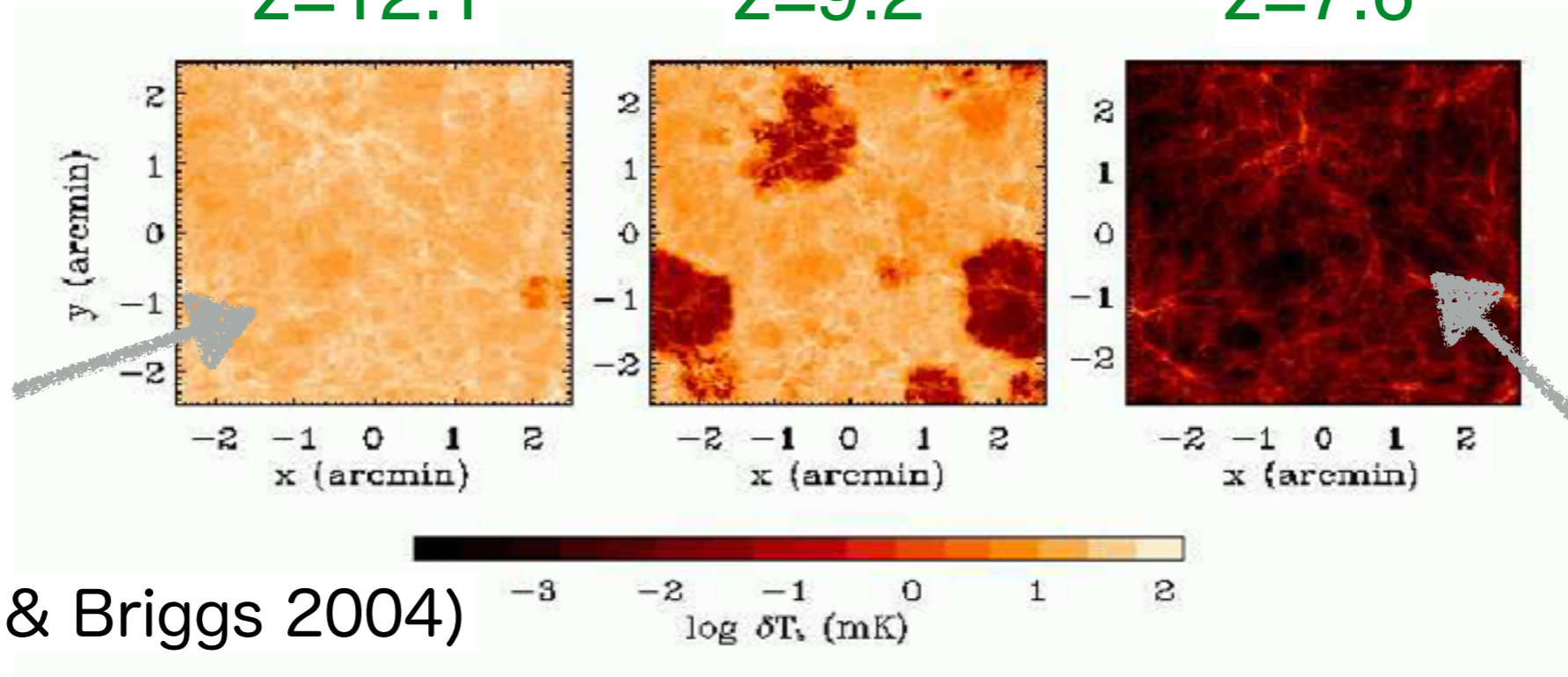
➔ **21cm line signal** from neutral hydrogen in the IGM

Simulated 21cm map

$z=12.1$

$z=9.2$

$z=7.6$



Neutral

Ionized

(Furlanetto & Briggs 2004)

Observations

Some on-going telescopes have started observation (MWA, LOFAR, PAPER).
Future observations are now planning (SKA, HERA) on 2020's.



MWA (Australia)



LOFAR (Netherlands)



PAPER (South Africa)



Commercial Message

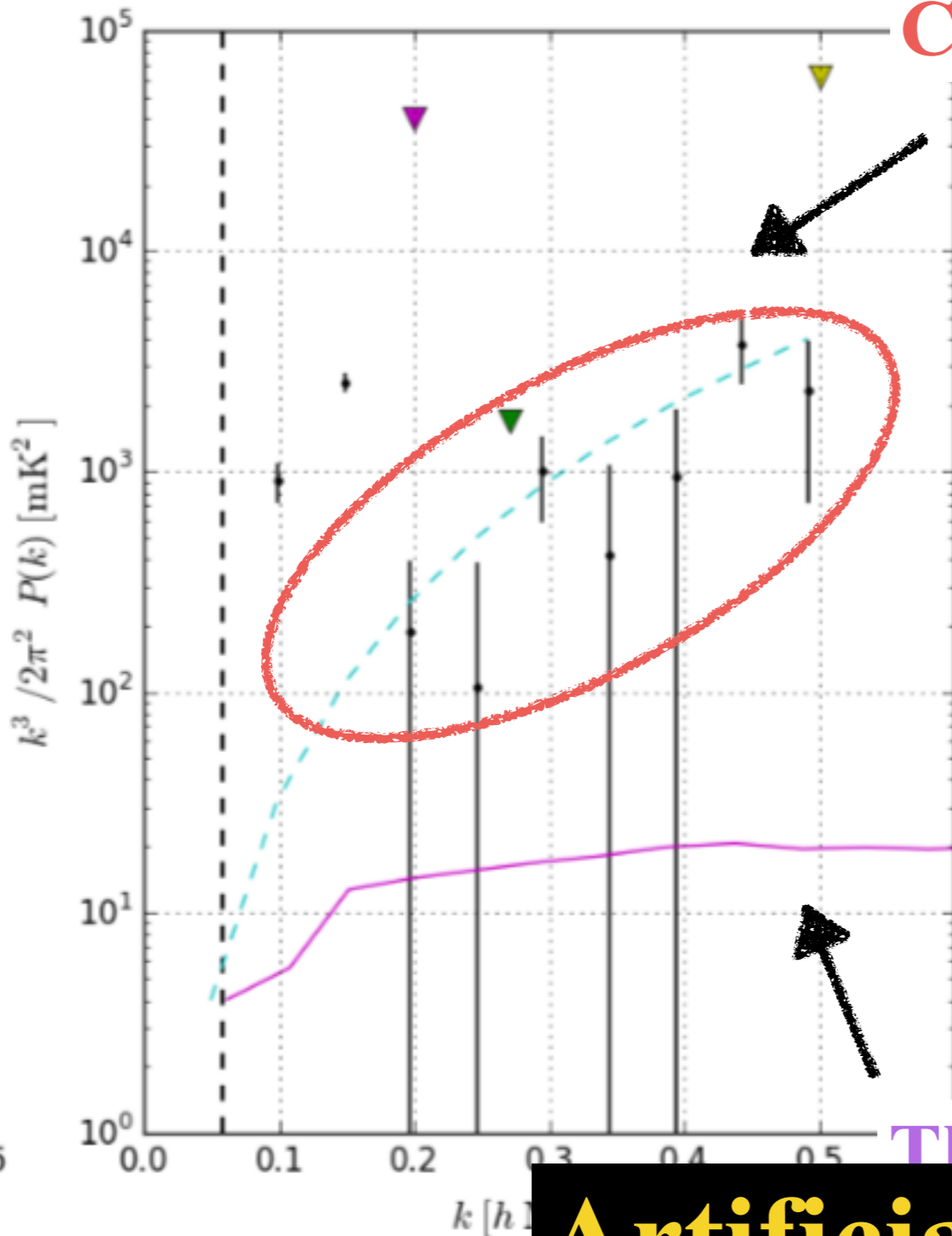
Collaboration with SKA-JAPAN has started (Kumamoto, Nagoya, ICRR).

➔ Cross-correlation between 21cm - LAE (Kubota et al, in preparation)

Upper limits on 21cm PS

$z=8.4$ (Ali et al 2015)

Current upper limit by PAPER



- 1-2 magnitude of order higher than theoretical prediction.

- But, we believe 21cm signal is detectable in near future.

After we actually detect 21cm signal,

How do we extract astrophysical information from 21cm PS ?

Theoretical prediction

Artificial neural network

Motivation

- We would like to extract the EoR information from 21cm PS.
- To determine EoR parameters helps this purpose.
- How precisely can we determine EoR model parameters?
- We train neural network architecture to learn association between 21cm PS and EoR parameters.
- Once we train ANN architecture, we can apply this to unknown data.

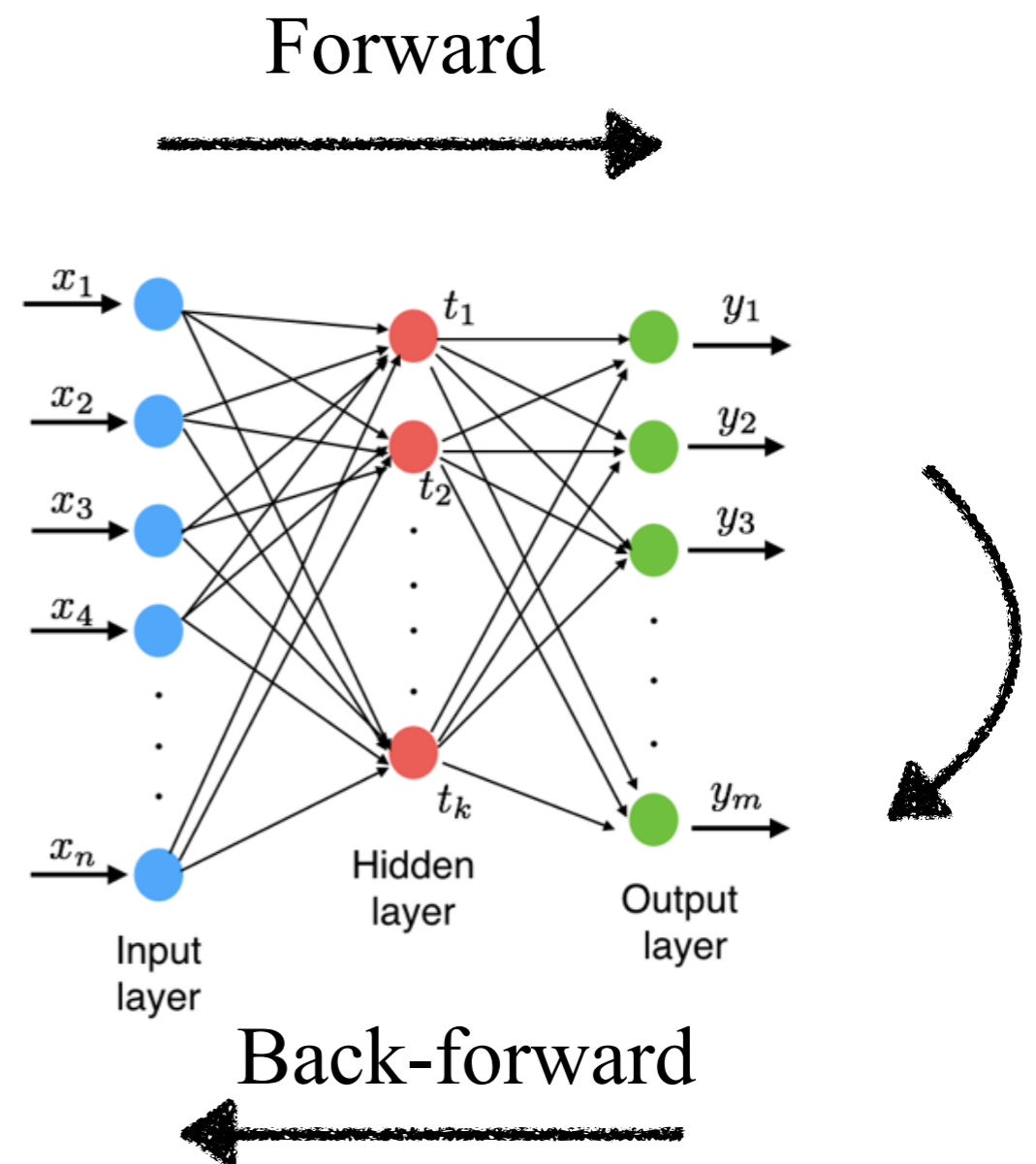
Back-propagation algorithm

We update the weights by gradient descent of cost function until they converge.

$$w(t + 1) = w(t) + \Delta w(t)$$

with

$$\Delta w_{ij}^{(l)} = -\eta \frac{\partial E}{\partial w_{ij}^{(l)}} = \eta \sum_{n=1}^{N_{\text{train}}} \frac{\partial E_n}{\partial w_{ij}^{(l)}}$$

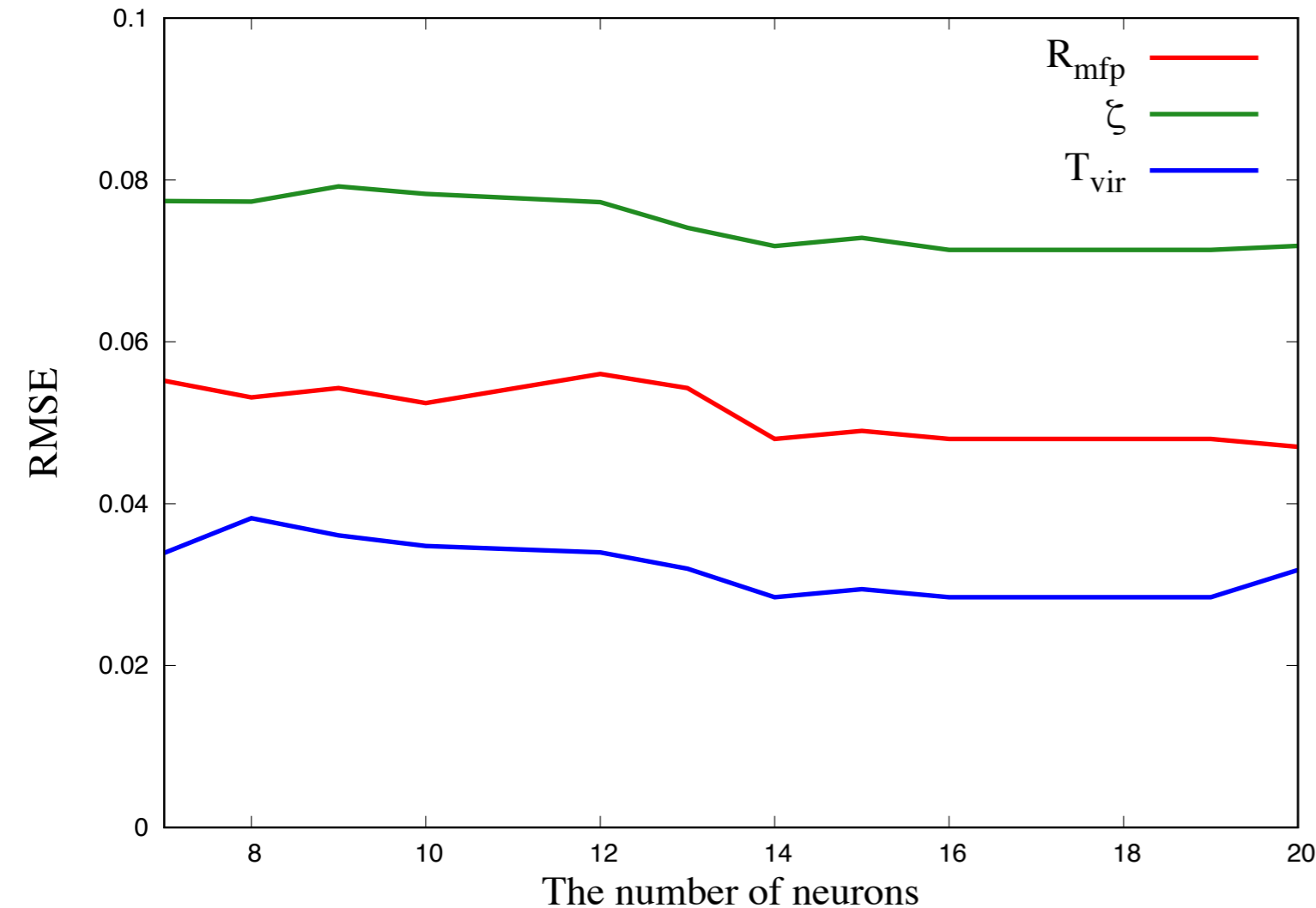


We calculate the derivative of cost function starting from output layer toward input layer.

Condition

- We used **70 training datasets** and **50 test datasets**.
- We calculate the 21cm PS by 21cmFAST
- We choose EoR parameters and 21cm PS datasets at $z=9,12$ for single z (and $z=9,10,11$ for multiple z).
- We train network with/without thermal noise and cosmic variance.
- $N_{\text{input}}=14, N_{\text{hidden}}=14, N_{\text{output}}=3$ for single z .

The number of neurons



- Fix 100000 iterations

- Change the number of neurons at hidden layer.

- Plot RMSE as function of the number of neurons.



(Note)

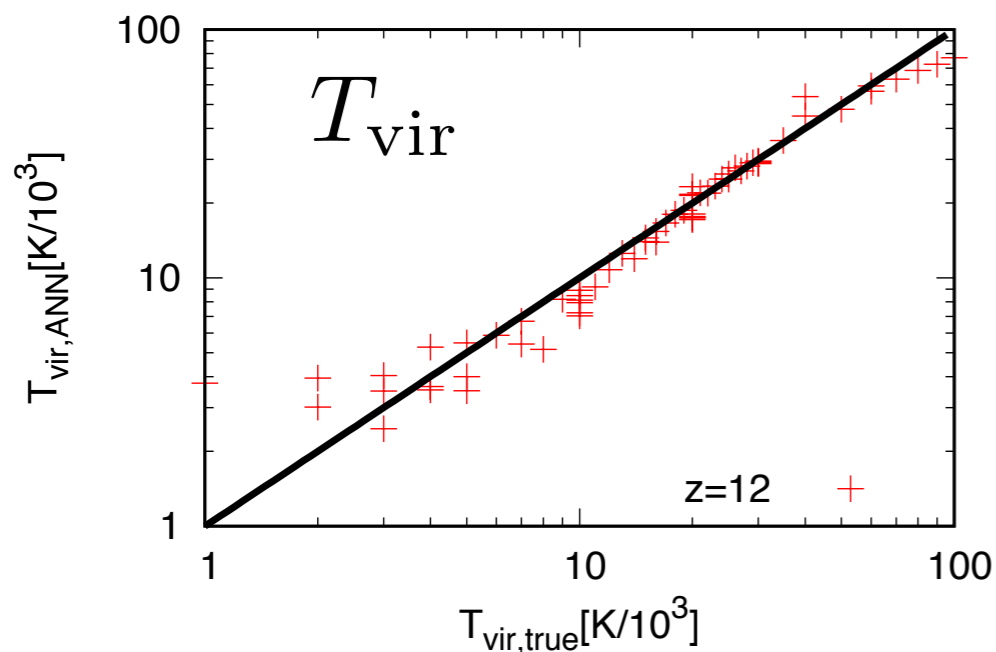
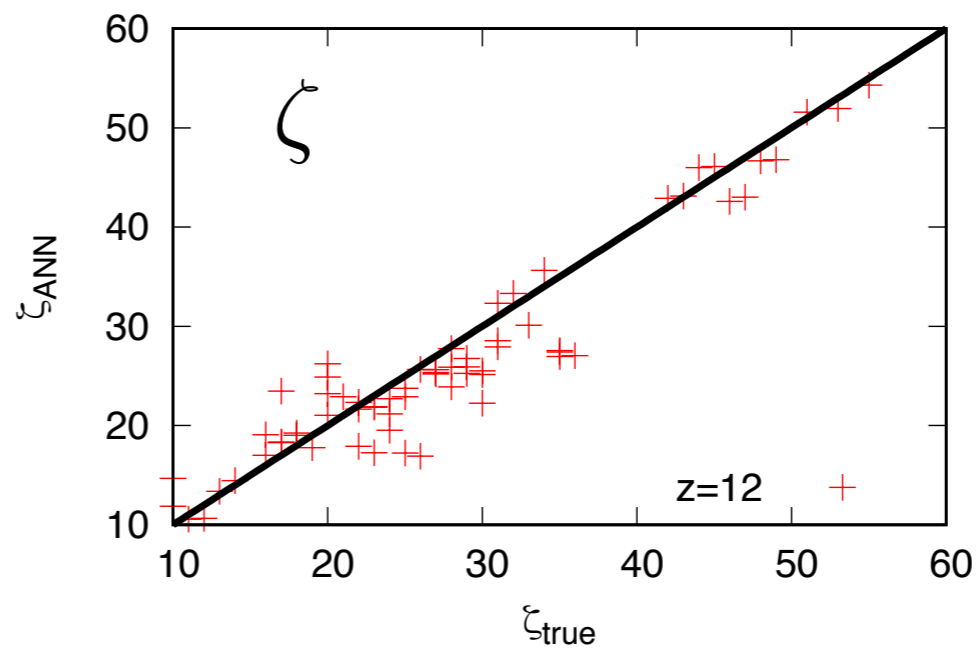
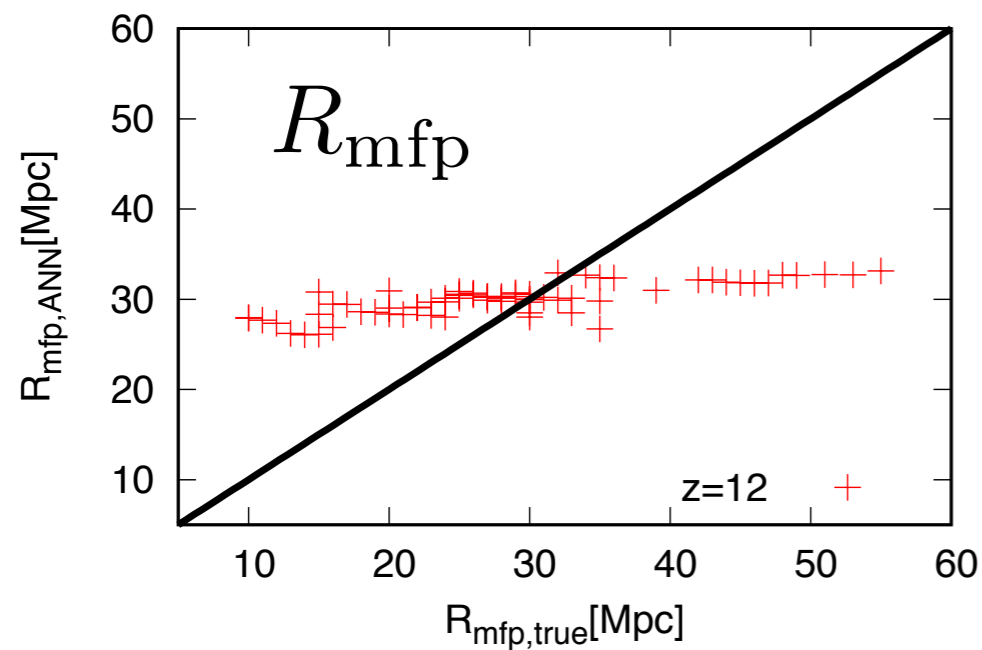
14 neurons at input layer.

Week dependence on the number of neurons.

Results

z=12, PS without any noise

14 neurons, 100000 iterations

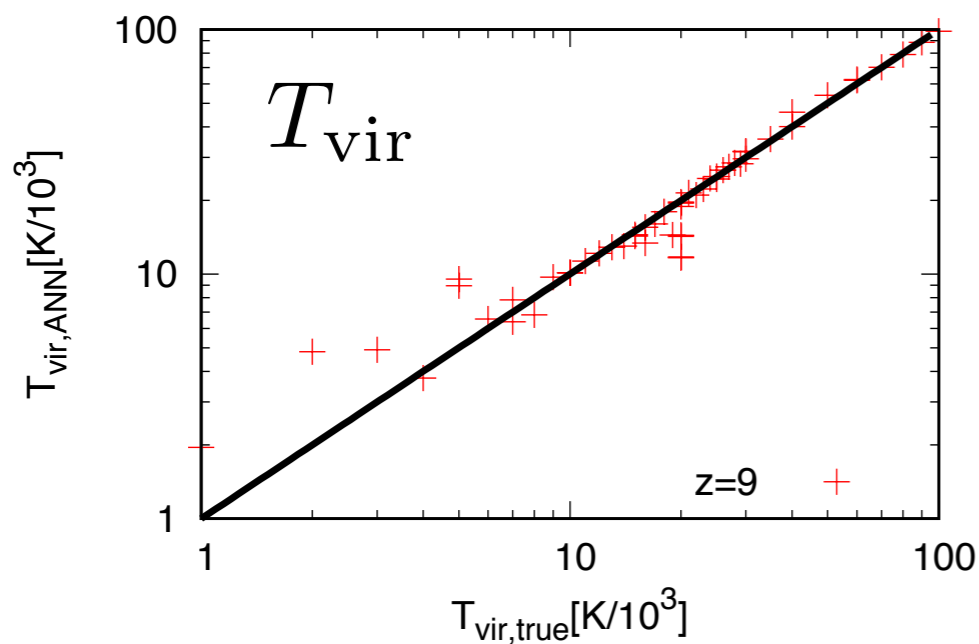
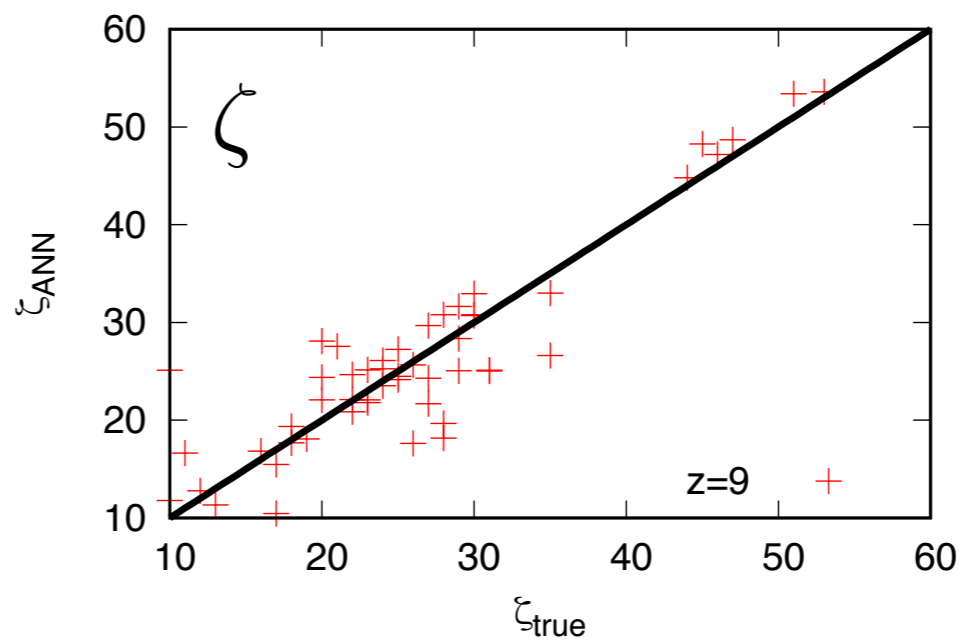
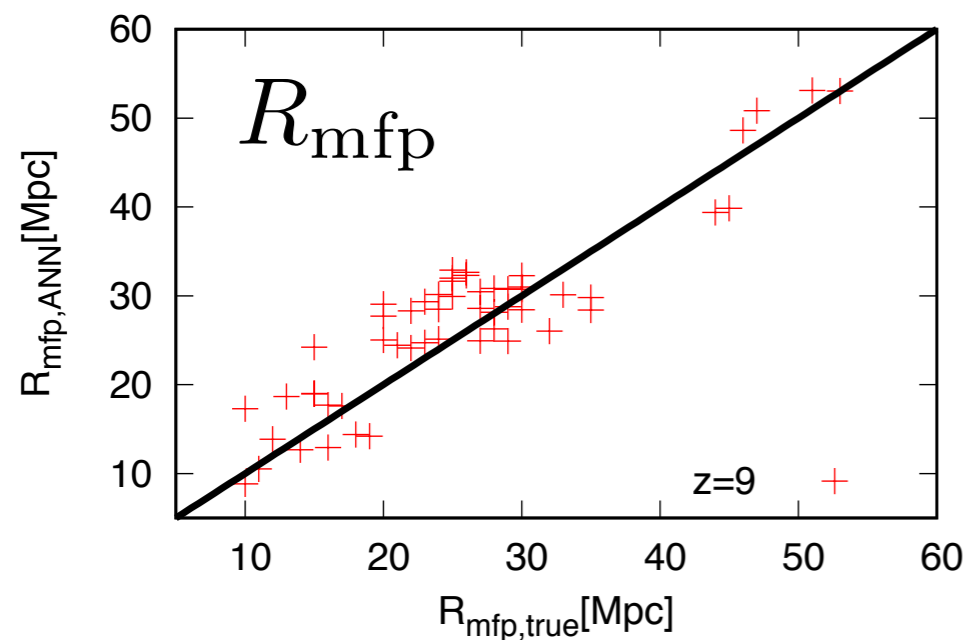


- True value .vs. Reconstructed value
- The scatter of R_{mfp} is large.
- Other reconstructed parameters match true one relatively well.

Results

z=9, PS without any noise

14 neurons, 100000 iterations



- Compared with $z=12$, reconstructed R_{mfp} match true value better.
- Because R_{mfp} expresses maximum size of HII bubble, it affects lower redshift when reionization advances.

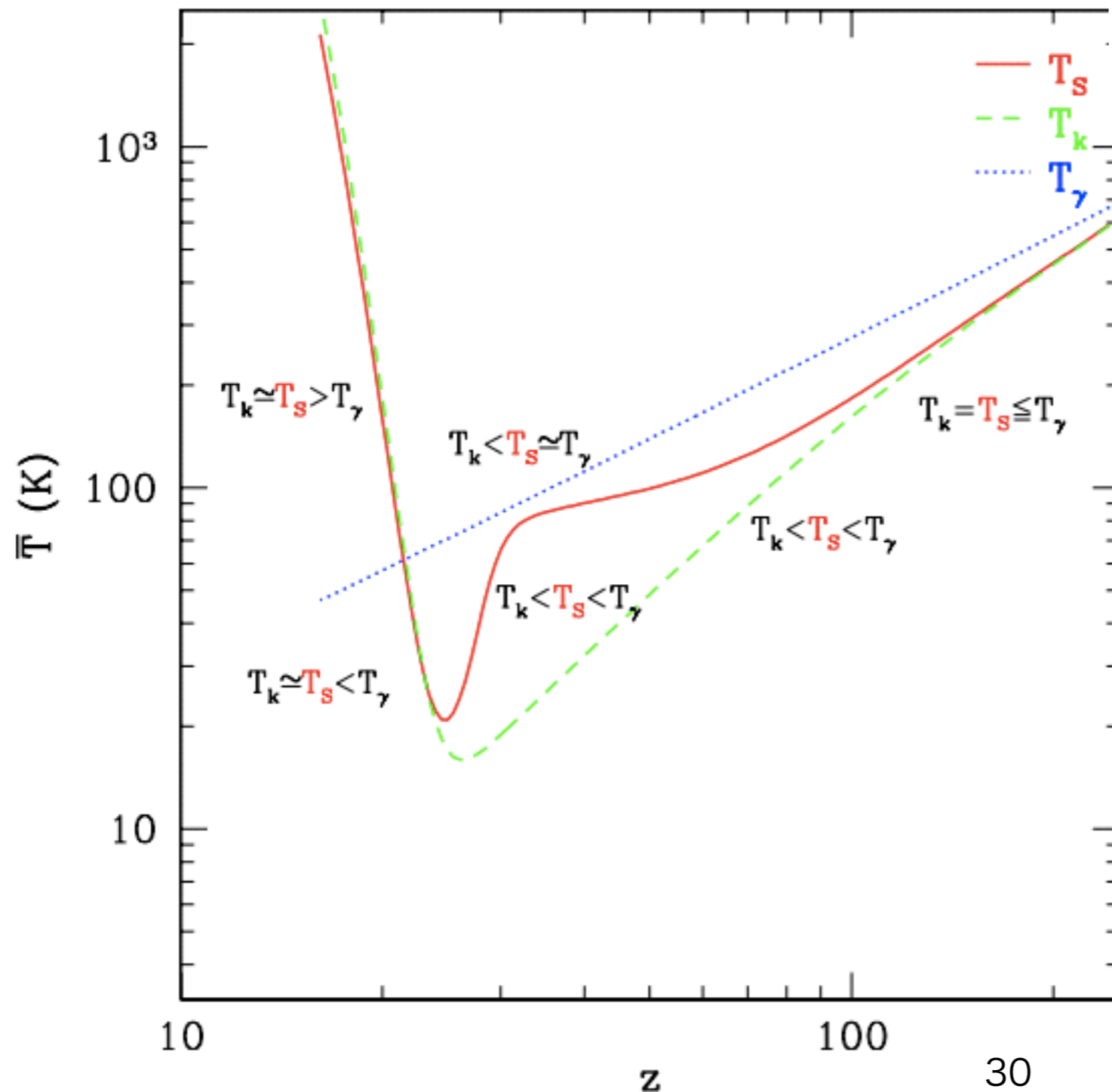
Compare with MCMC

	ANN		MCMC	
	RMSE _{SKA}	RMSE _{HERA}	$1\sigma_{\text{SKA}}$	$1\sigma_{\text{HERA}}$
R_{mfp}	0.258	0.278	0.178	0.184
ζ	0.288	0.354	0.167	0.220
$\log(T_{\text{vir}})$	0.038	0.040	0.024	0.033

- For comparison with MCMC, we also show 1sigma error obtained by MCMC.
- We compare the RMSE obtained by ANN with 1 sigma error obtained by MCMC in the both case of HERA and SKA observation.
- EoR parameters obtained by ANN have similar error level to those by MCMC.

Thermal history

Mesinger et al 2010



$$T_S^{-1} = \frac{T_{\text{CMB}}^{-1} + x_\alpha T_\alpha^{-1} + x_K T_K^{-1}}{1 + x_\alpha + x_K}$$

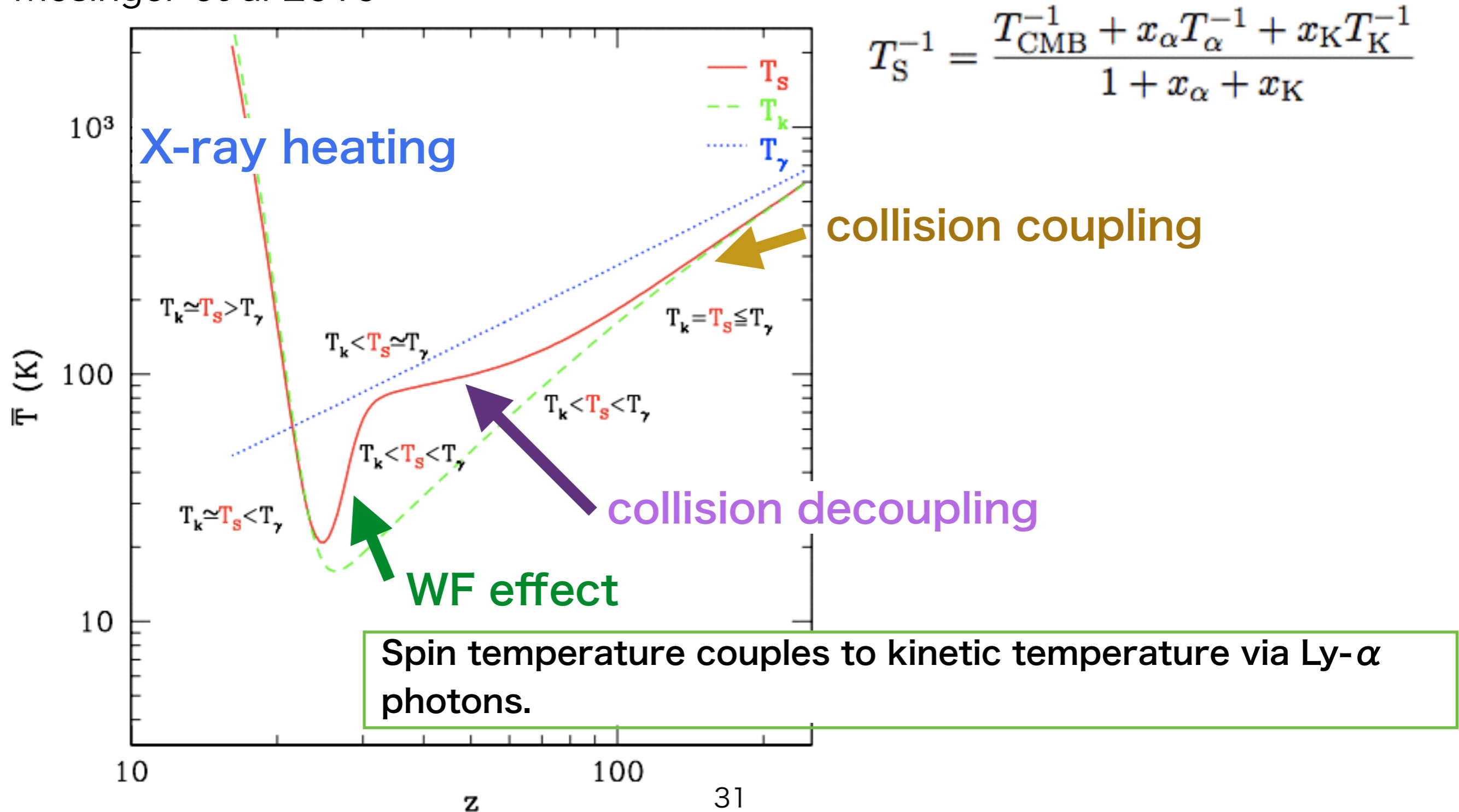
Blue : CMB temperature
 $\propto (1 + z)$

Green : kinetic temperature
 $\propto (1 + z)^2$

Red : spin temperature

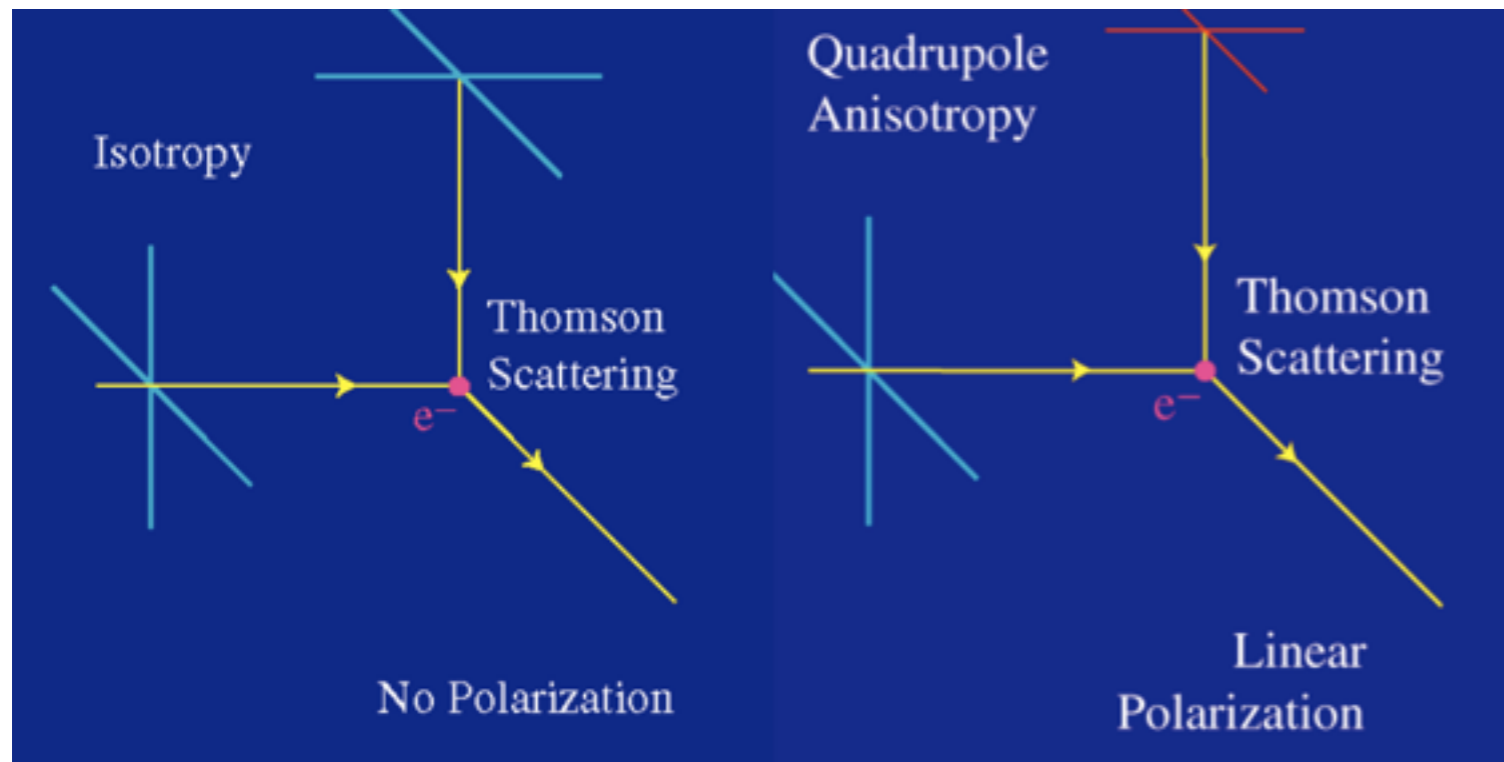
Thermal history

Mesinger et al 2010



Constraints

The observation of the CMB polarization
→ the optical depth of Thomson scattering



Optical depth

$$\tau_e \propto \int_{z_r}^0 n_e(z) \frac{dt}{dz} dz$$

$$\tau_e = 0.078 \pm 0.0019$$

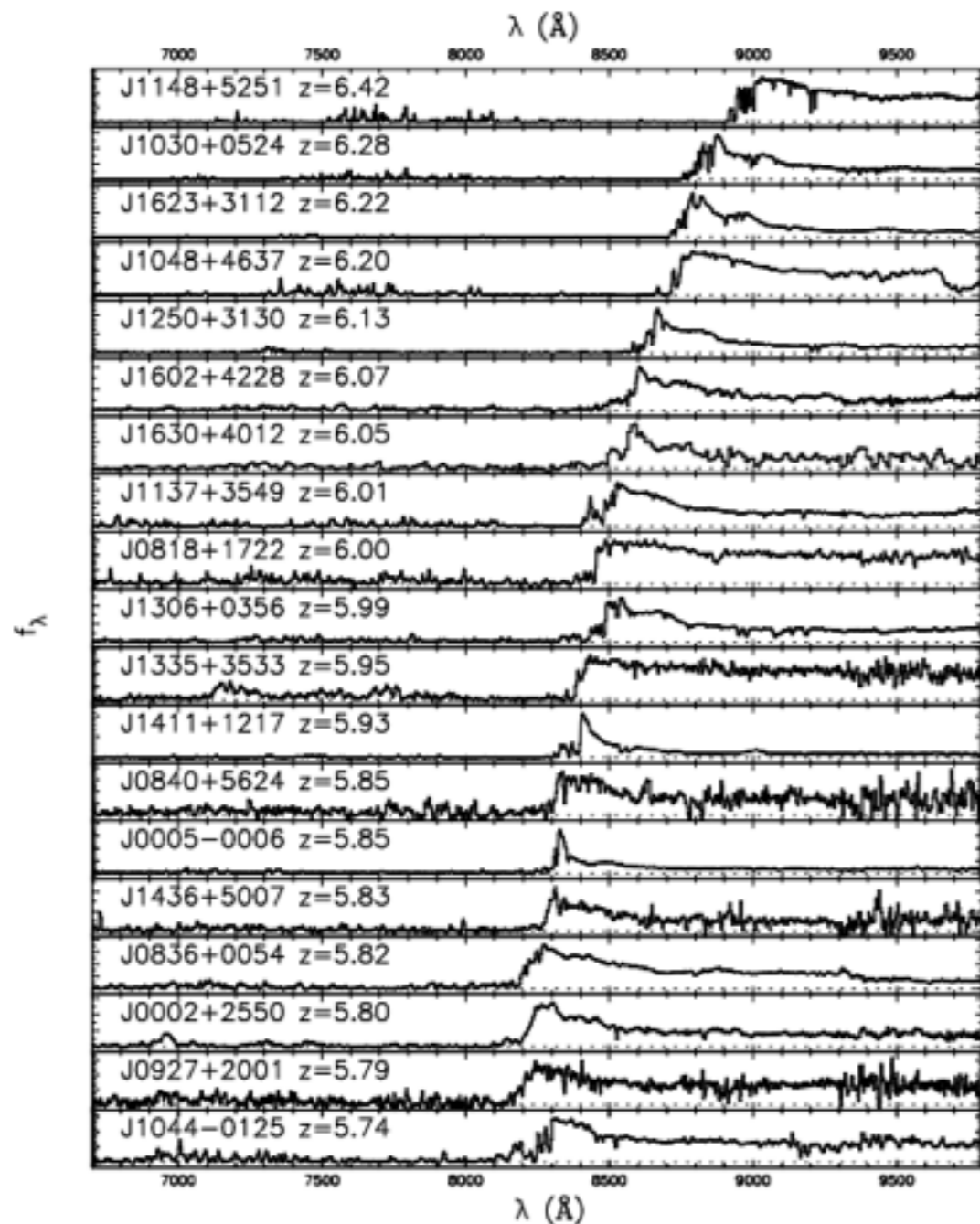
$$z_r = 9.9^{+1.8}_{-1.6}$$

(Planck collaboration 2015)

Constraints

high- z QSO absorption

→ constraint on the epoch where the EoR finishes.



Gun-Peterson test

If the neutral hydrogen exists, it absorbs the Ly-alpha photons.

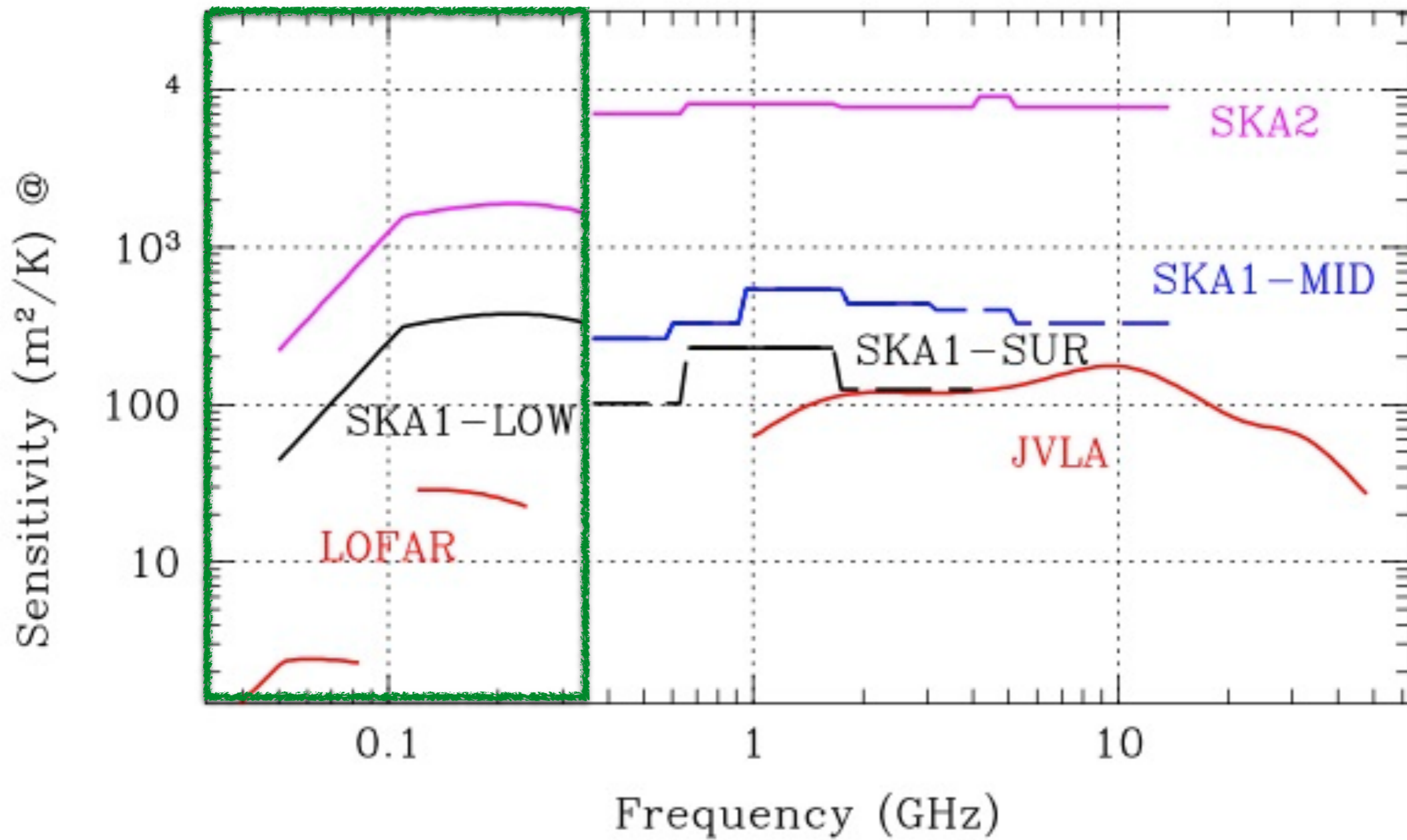
No emission line !



We can know the epoch where EoR finished via the QSO spectrum.

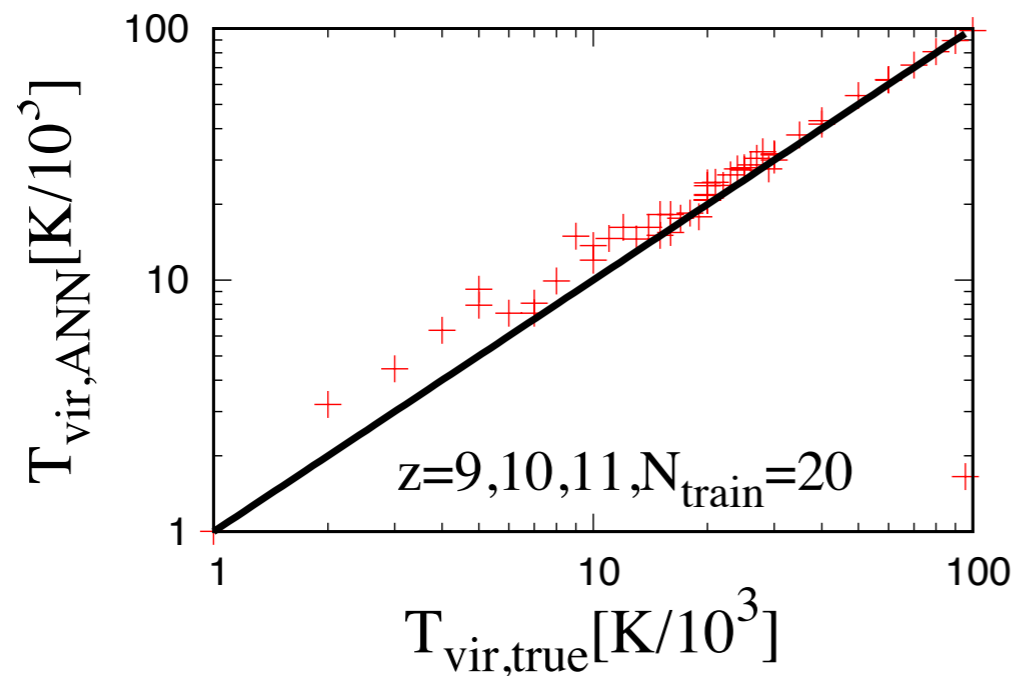
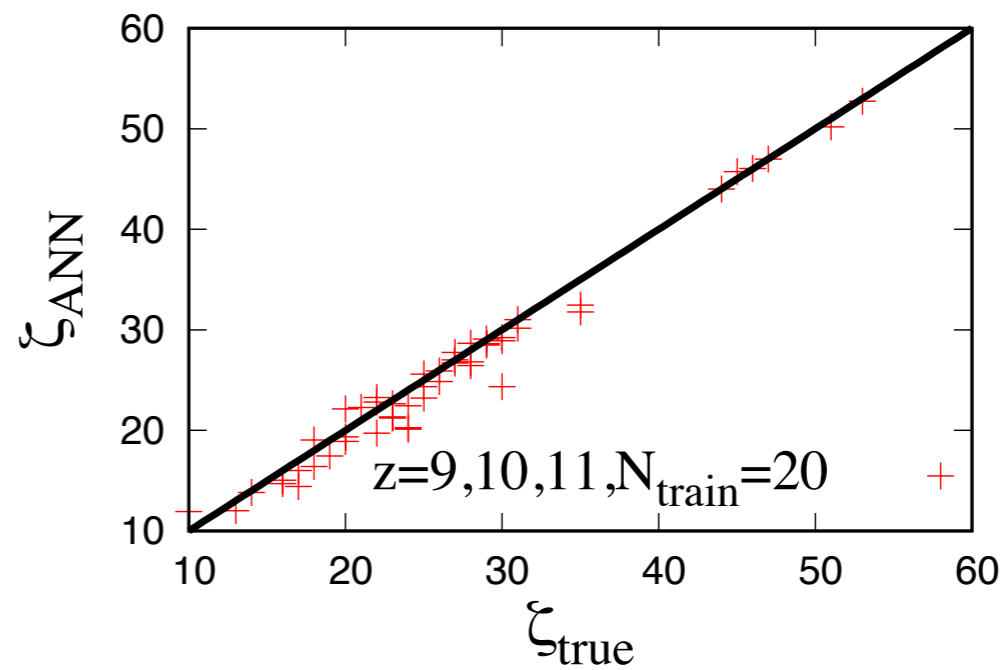
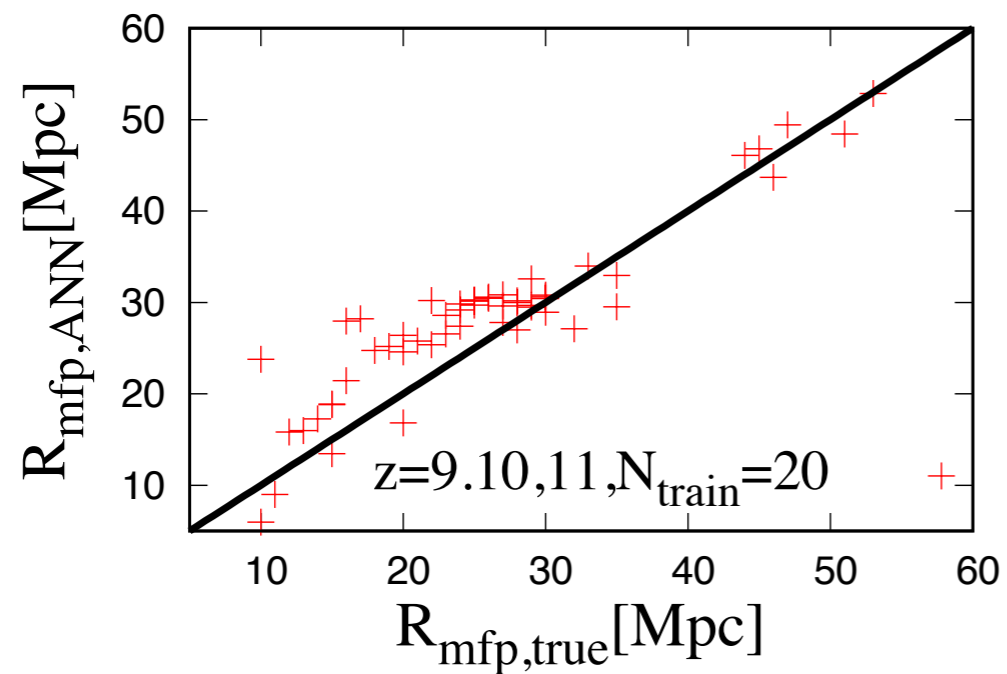
Sensitivity

Sensitivity

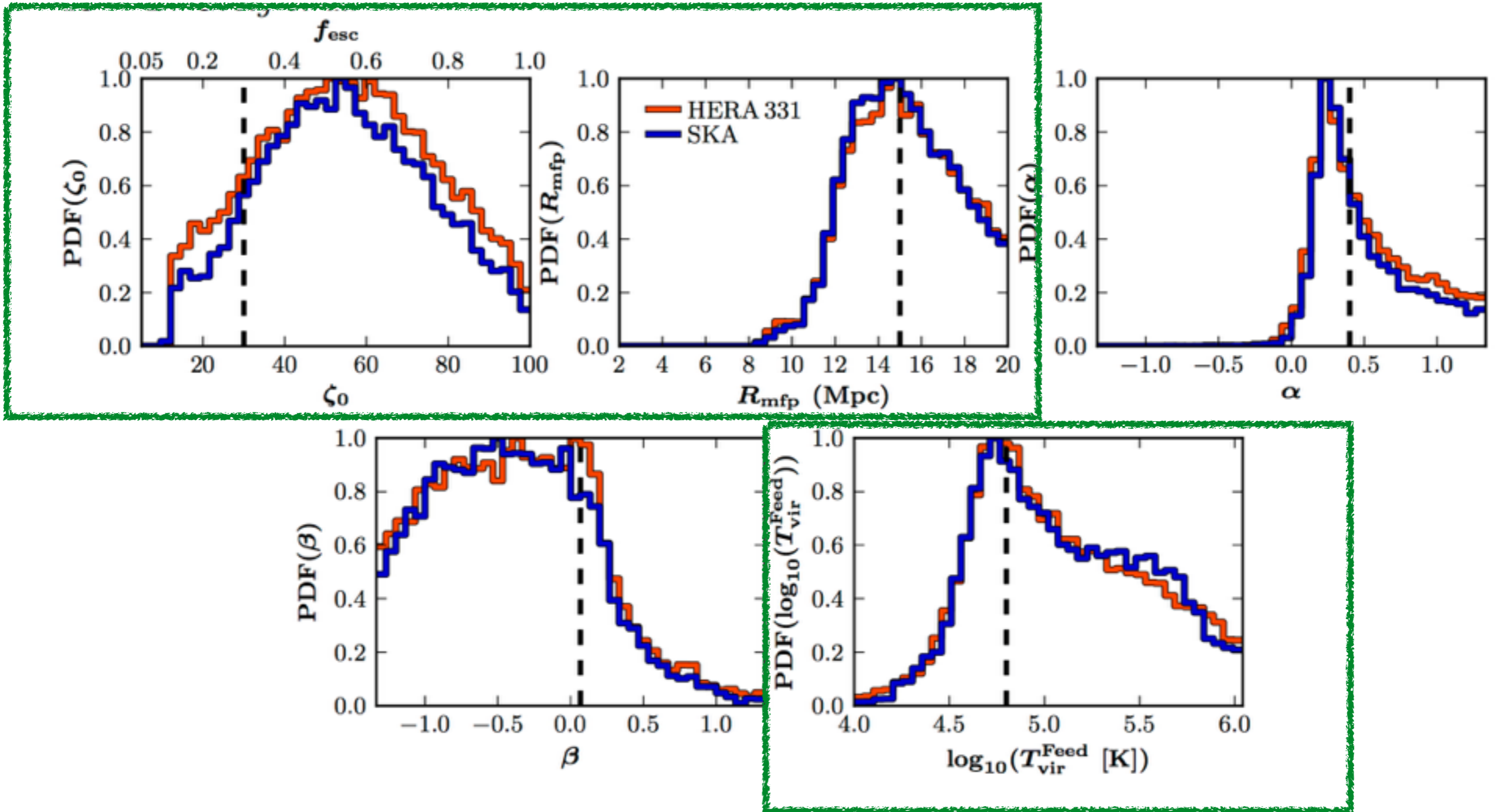


Results

Reduce the number of training data. ($N=20$)



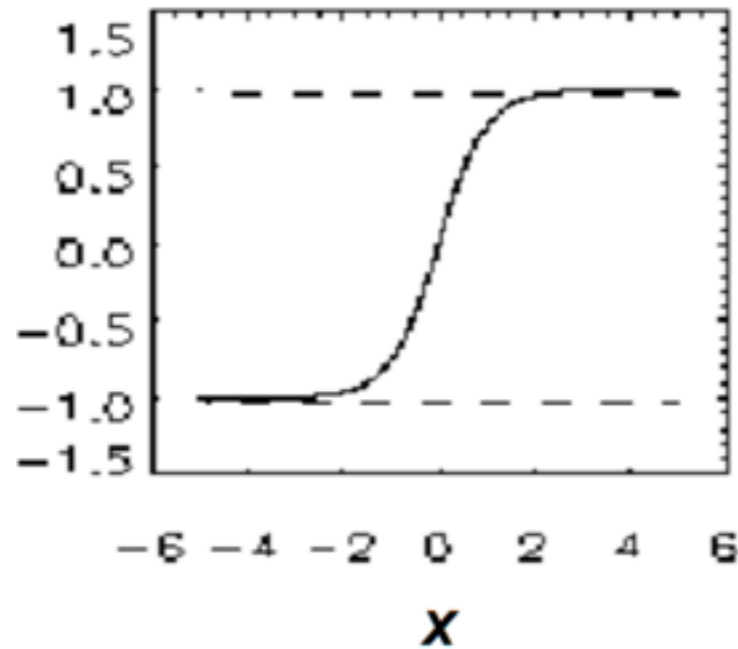
MCMC



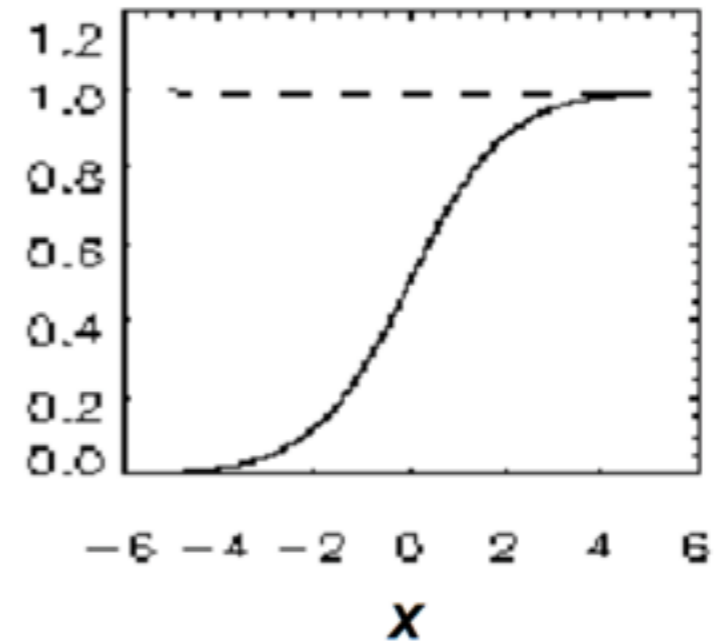
The probability distribution function of each parameter.

Activation function

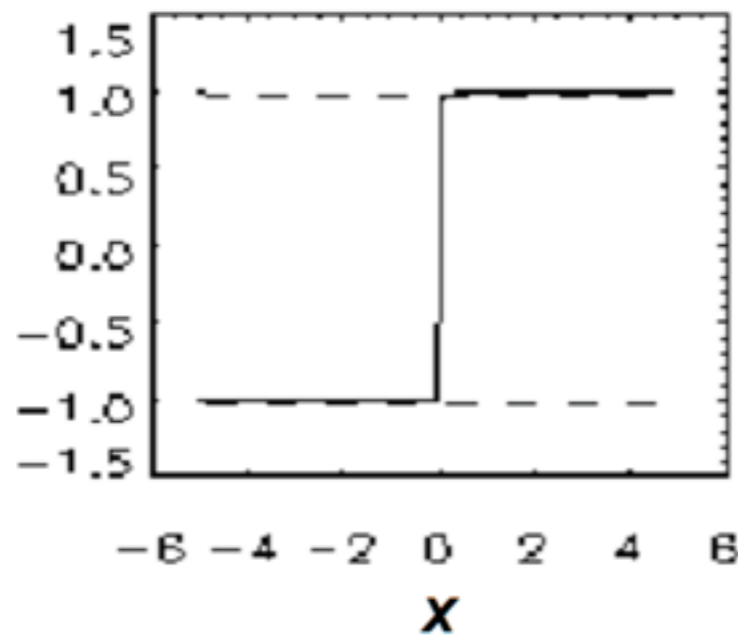
tanh(x)



Sigmoid, $(1 + \exp(-x))^{-1}$



Hard Limiter



Ramp Function

