Analysing the 21cm signal with Artificial Neural Networks

Hayato Shimabukuro [島袋隼士] (Observatoire de Paris)

Based on Shimabukuro & Semelin
(astro-ph/1701.07026, Accepted in MNRAS)

l'Observatoire de Paris

Contents

- Introduction
- Cosmological 21cm signal
- Artificial neural networks (ANN)
- Result
- Summary

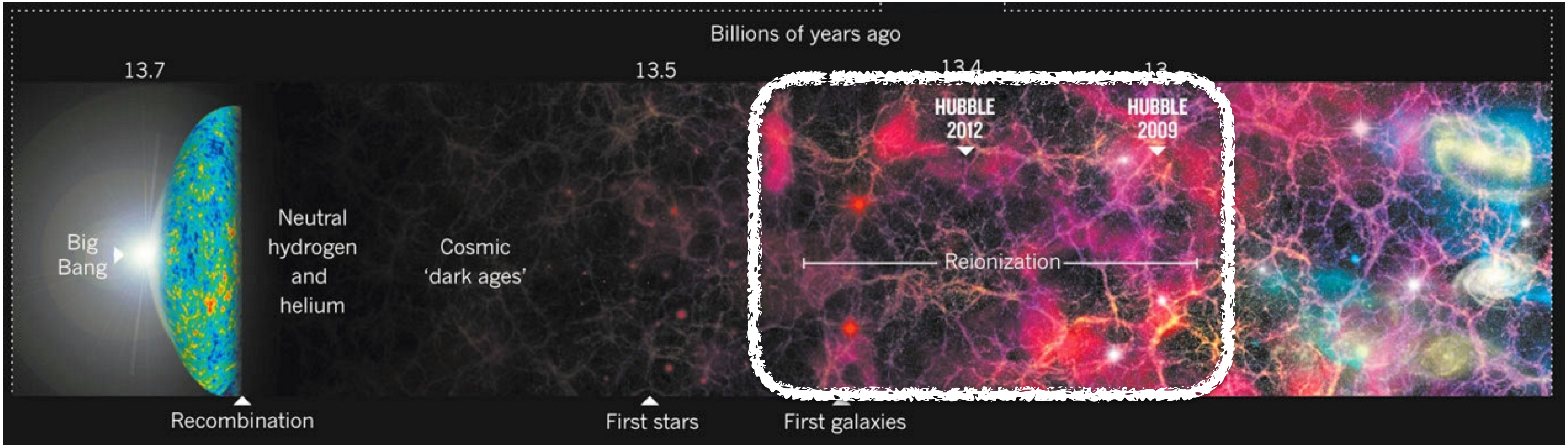
Introduction

Who am I?

- Okinawa-Sendai-Nagoya-Kumamoto-Paris
- Ph.D at Nagoya university (03/2016)
- Post-doc at Observatoire de Paris with Benoit Semelin (04/2016~)

Meridian
(子午線)

Introduction

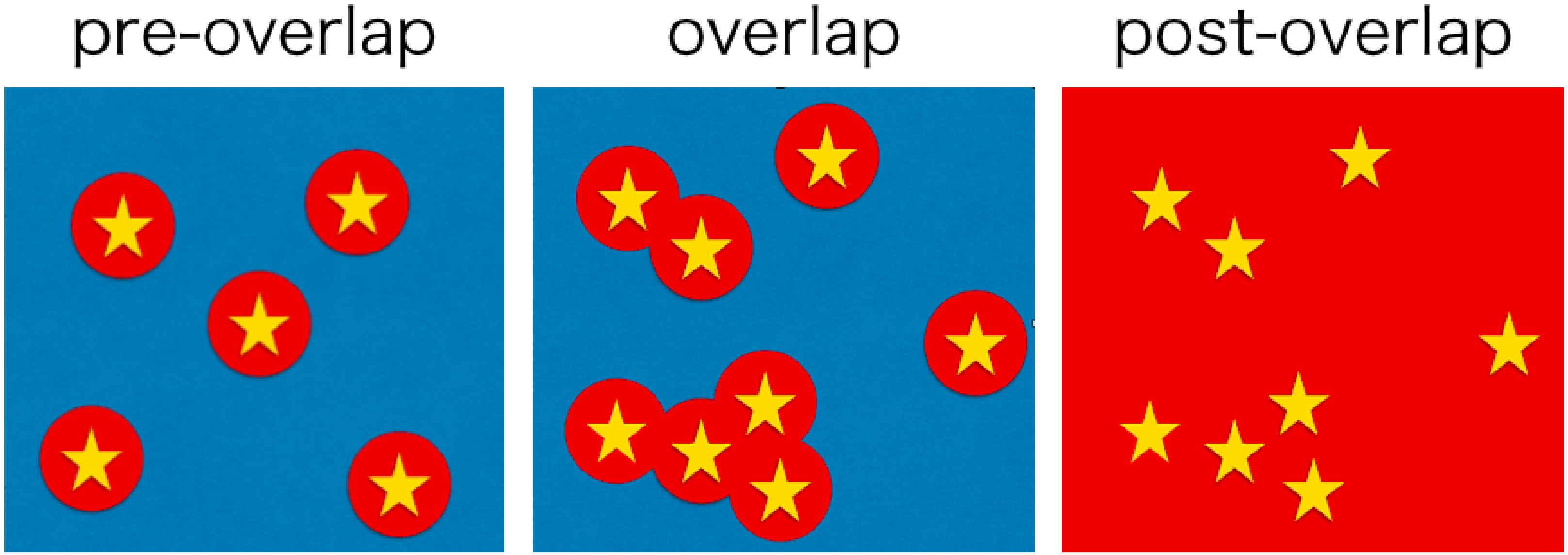


○z >30・・**Dark age**：Star formation has not started yet.

○15<z<30・・**Cosmic dawn**：After star formation, astrophysical effects become effective.

○7<z<15・・**Epoch of Reionization(EoR)**：Hydrogen is ionized by UV photons emitted from stars,galaxies.

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

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

(Fan et al 2006)

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

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

(Konno et al 2014)

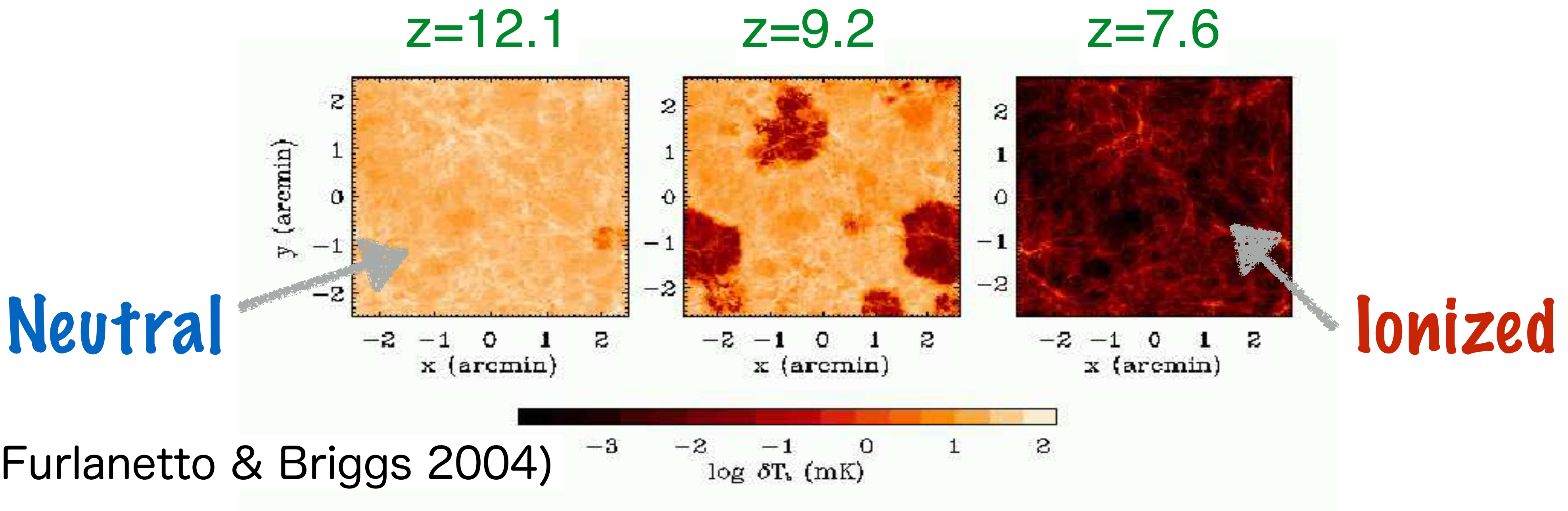
Beyond current constraints

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

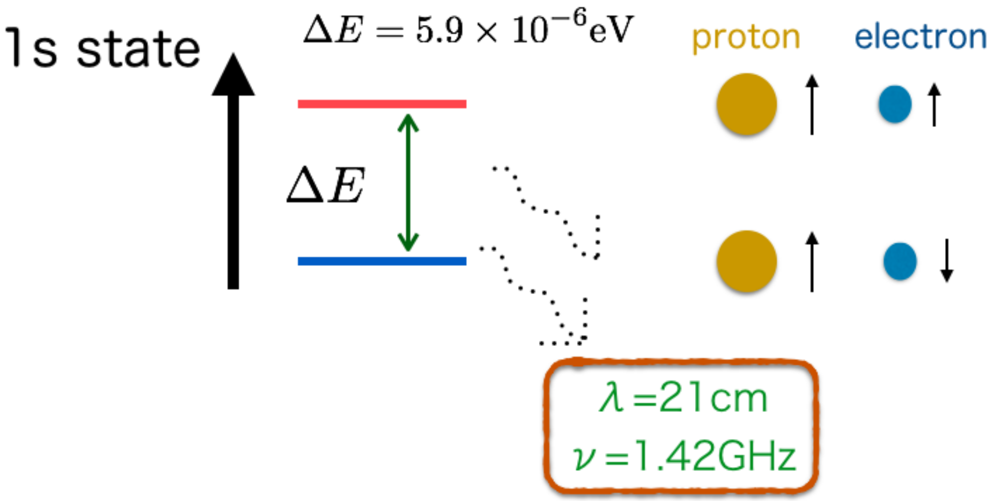


(Furlanetto & Briggs 2004)

Cosmological 21cm signal

21cm signal

○**21cm line radiation** : Neutral hydrogen emits the radiation due to the hyper fine structure.



Spin temperature

$$\frac{n_1}{n_0} = 3\exp\left(-\frac{h\nu_{21}}{kT_{\mathrm{S}}}\right)$$

$$T_{\mathrm{S}}^{-1} = \frac{T_{\mathrm{CMB}}^{-1} + x_\alpha T_\alpha^{-1} + x_{\mathrm{K}} T_{\mathrm{K}}^{-1}}{1 + x_\alpha + x_{\mathrm{K}}}$$

○Collision with **CMB photons**

○Collision with Ly-α photons

○Collision with atoms

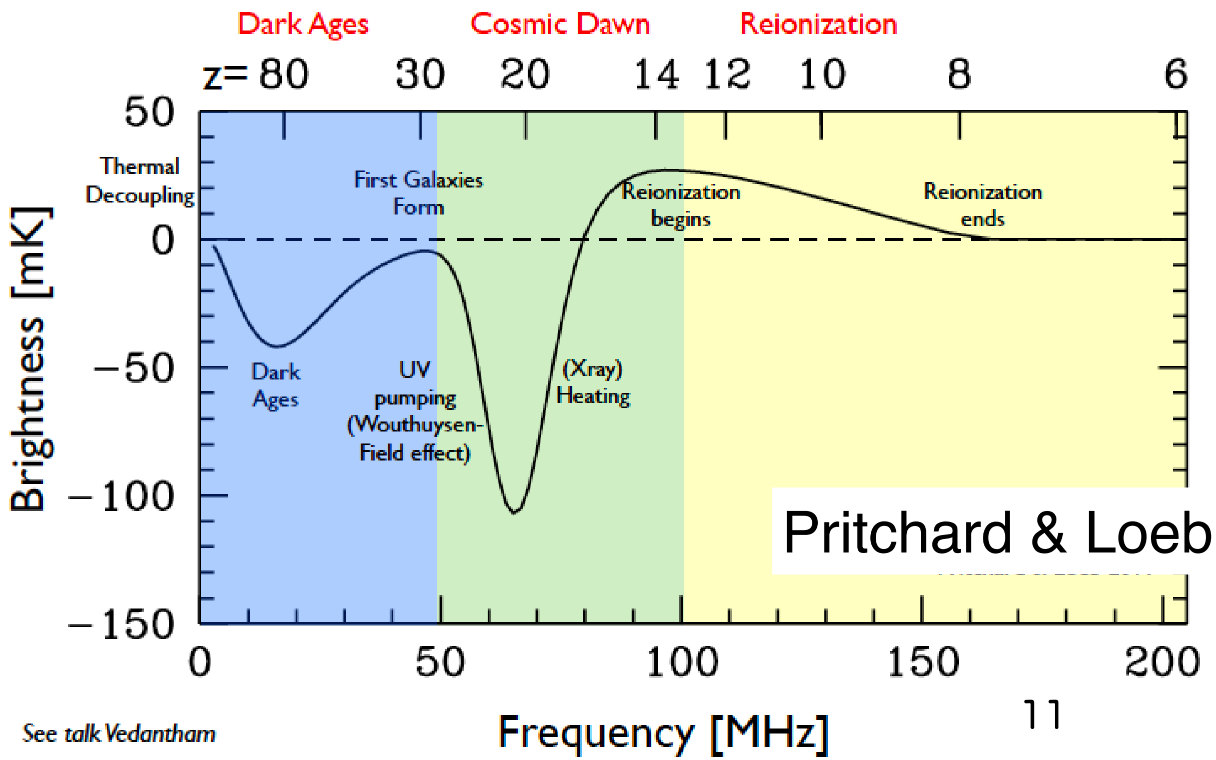
Brightness temperature

We actually observe brightness temperature instead of spin temperature

○Brightness temperature

$$\delta T_b(\nu) = \frac{T_S - T_\gamma}{1+z}(1 - e^{-\tau_{\nu_0}})$$

$$\sim 27 \underline{x_H} (1 + \underline{\delta_m}) \left(\frac{H}{\underline{dv_r/dr + H}} \right) \left(1 - \underline{\frac{T_\gamma}{T_S}} \right) \left(\frac{1+z}{10} \frac{0.15}{\Omega_m h^2} \right)^{1/2} \left(\frac{\Omega_b h^2}{0.023} \right) [\mathrm{mK}].$$



Pritchard & Loeb (2011)

Red : astrophysics.

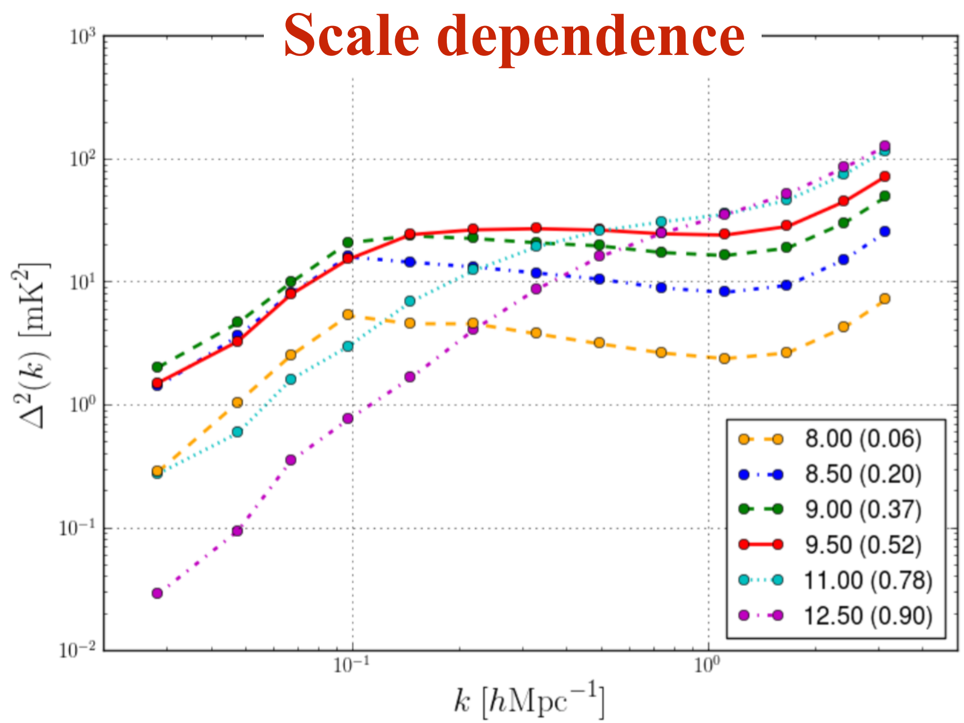
Blue : cosmology

21cm power spectrum

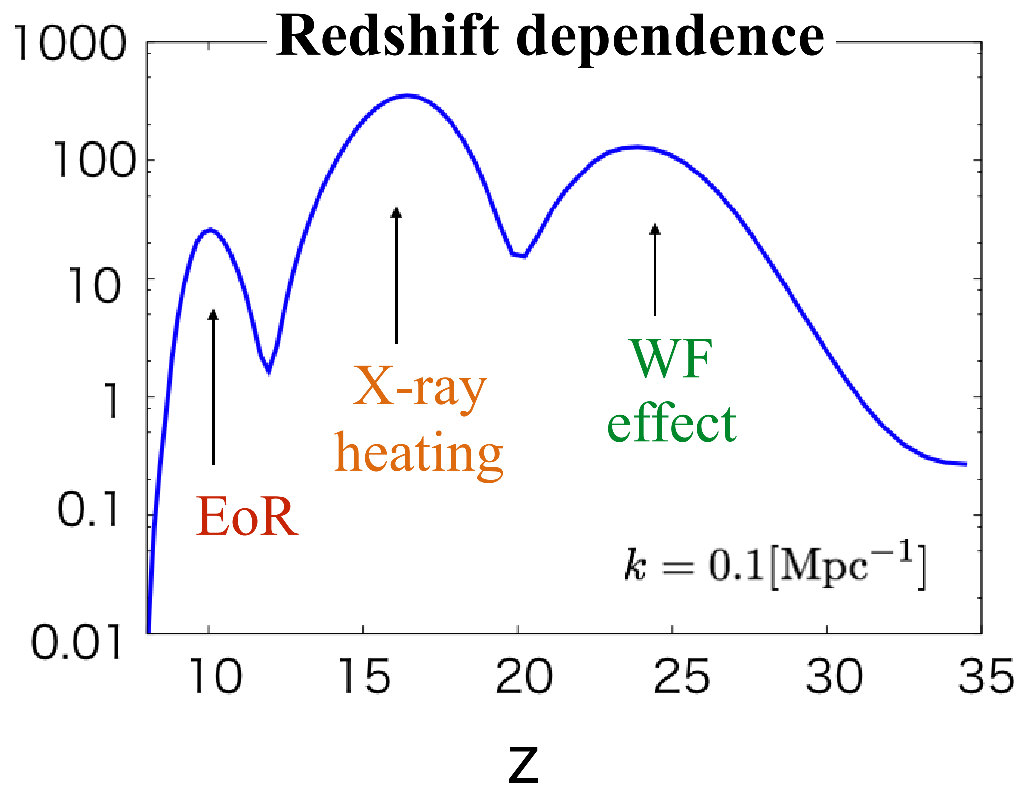
We often evaluate statistical property of the 21cm fluctuations by **power spectrum.**

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)

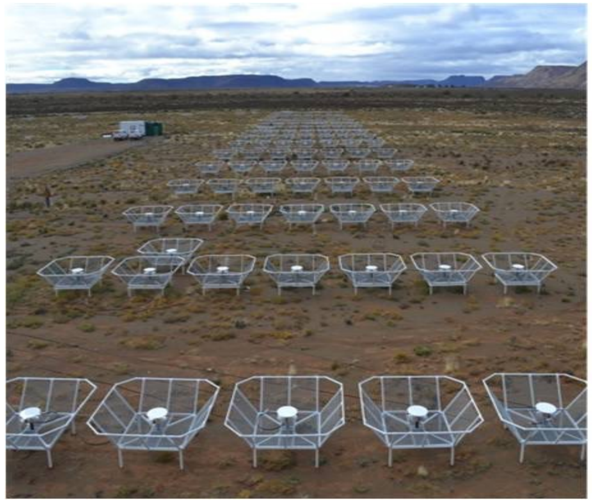


Observations

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

MWA (Australia)

LOFAR (Netherland)



PAPER(South Africa)

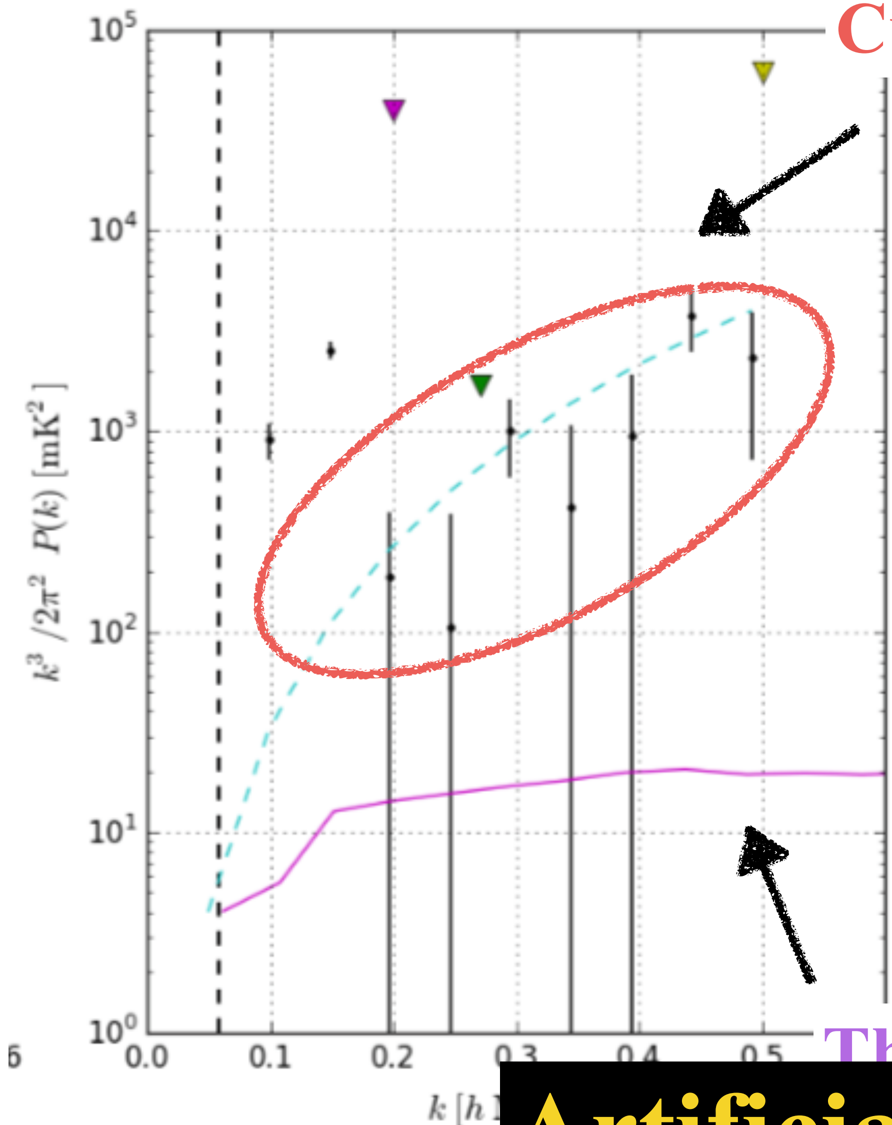
Commercial Message

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

Artificial neural networks

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? (ex.MCMC, Greig et al 2015).
- 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.

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) → $(\vec{x}_{data}, \vec{y}_{data})$

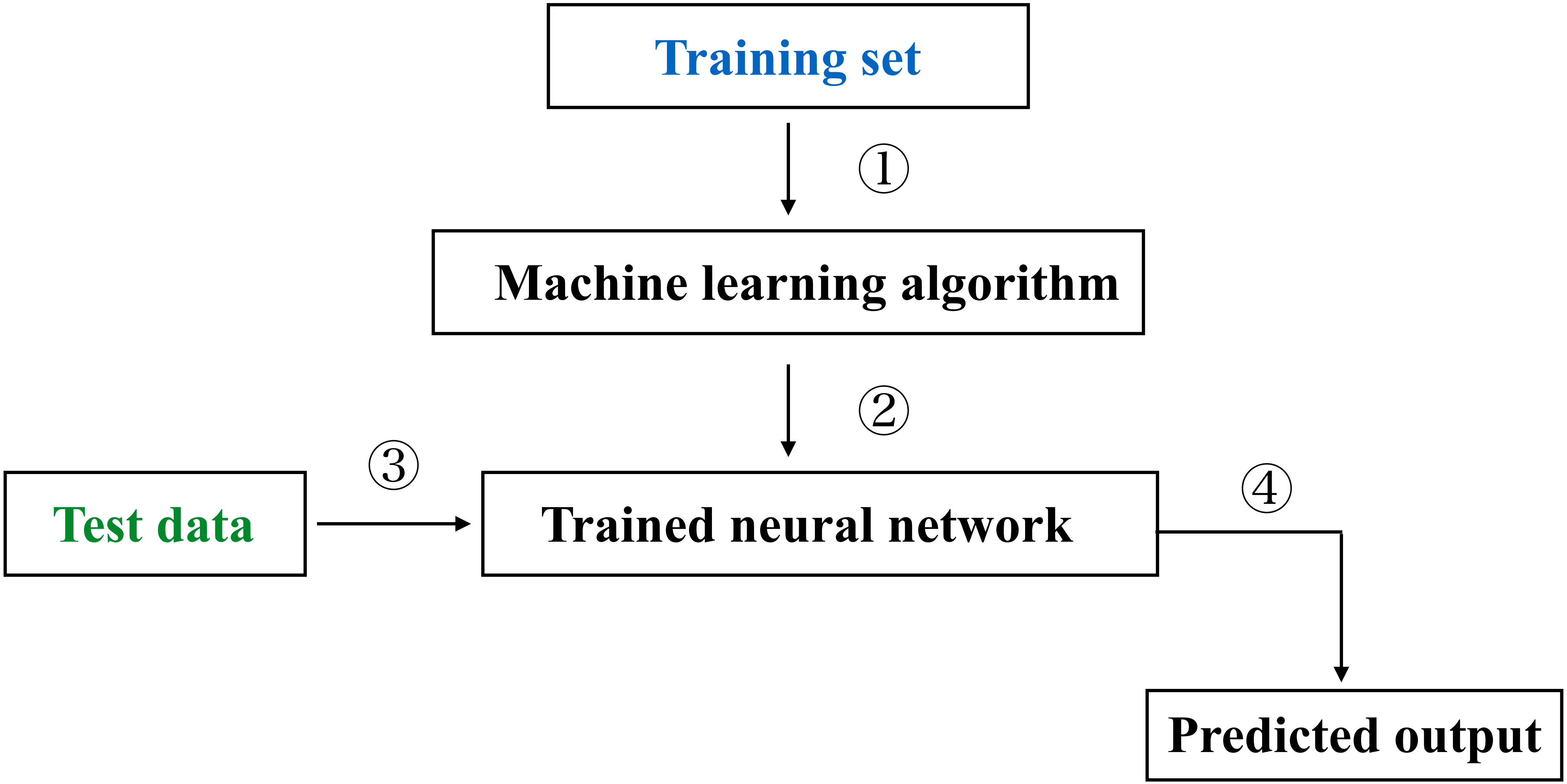
2. Train architecture of neural network by training data.

→ $y = f(x)$

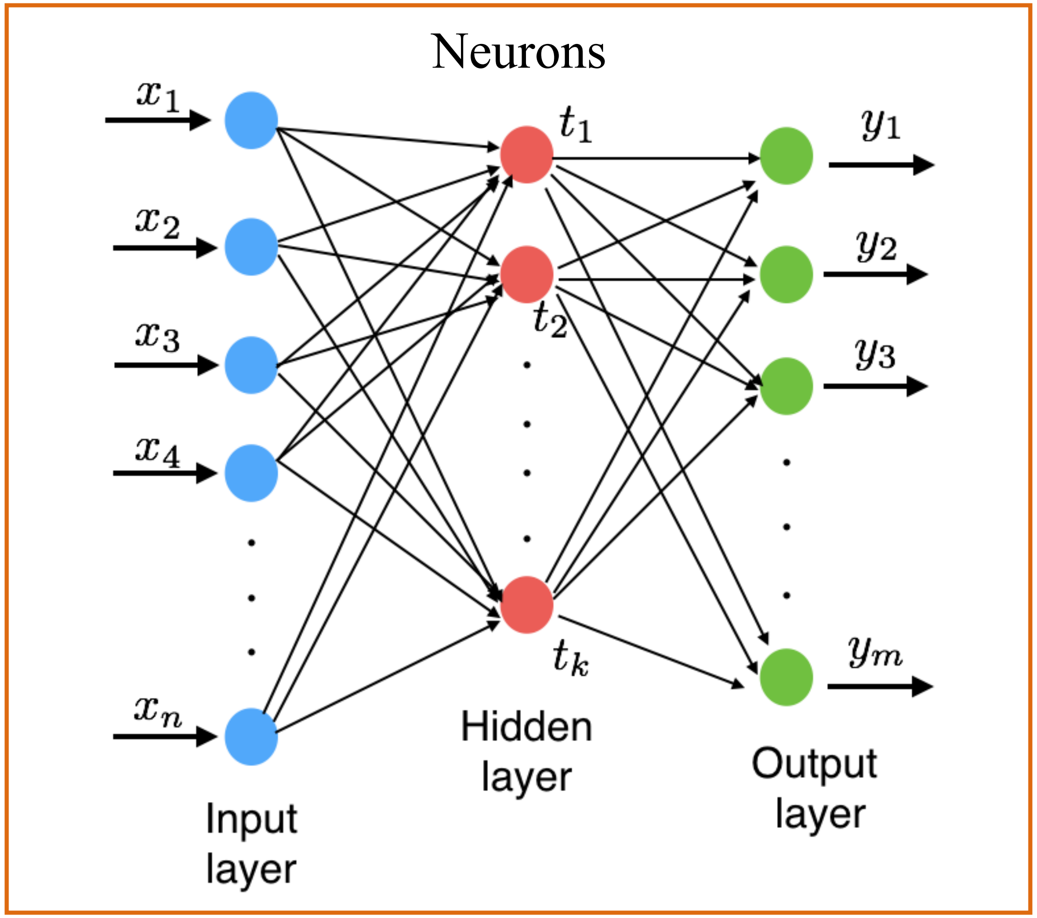
3. Apply trained network to unknown (test data) and can obtain expected output data.

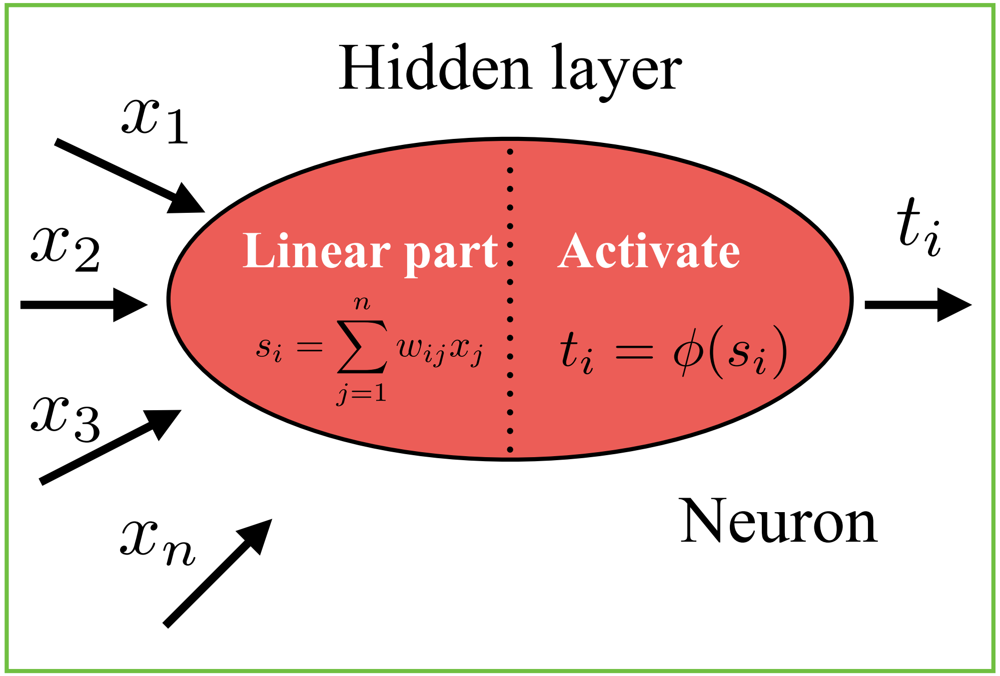
→ $y_{\mathrm{ANN}} = f(x_{\mathrm{test}})$

Flowchart


Training set
①
Machine learning algorithm
②
Test data
③
Trained neural network
④
Predicted output

What is ANN?





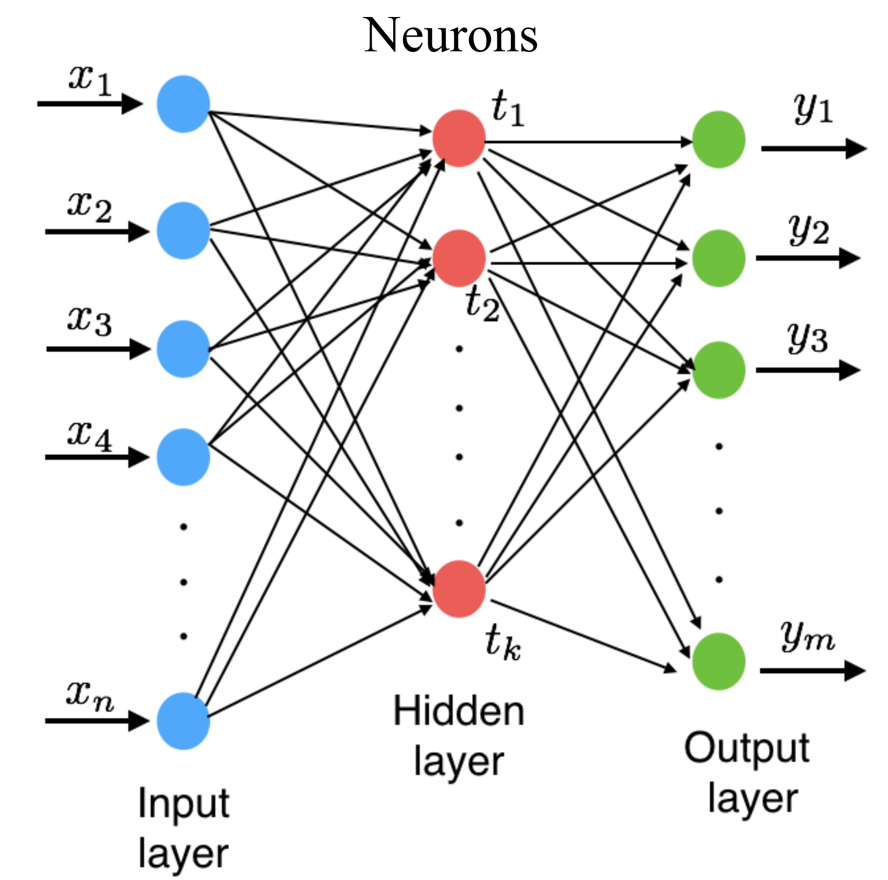
w_{ij} :Weight

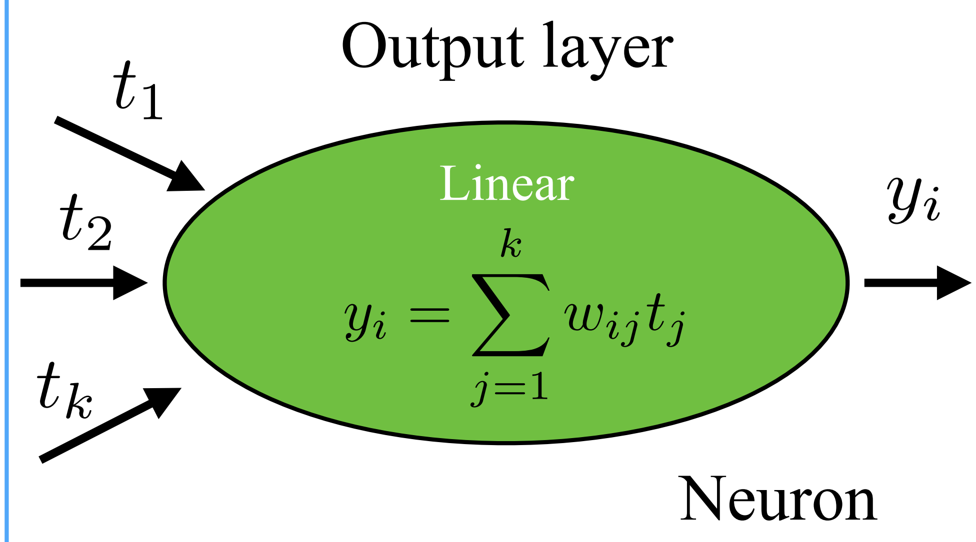
ϕ :Activation function

1.At input layer, we first calculate linear combination of input data with weight and transfer them to hidden layer.

2.Next, we activate that linear combination with activation function.

What is ANN?





w_{ij} :weight between hidden layer and output layer

At output layer, we consider linear combination of neurons at hidden layer. Unlike hidden layer, we do not need to activate linear combination at output layer.

What is ANN?

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

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

N: the number of training datasets

We use "Back propagation algorithm" to determine weight. (Rumelhart et al 1986)

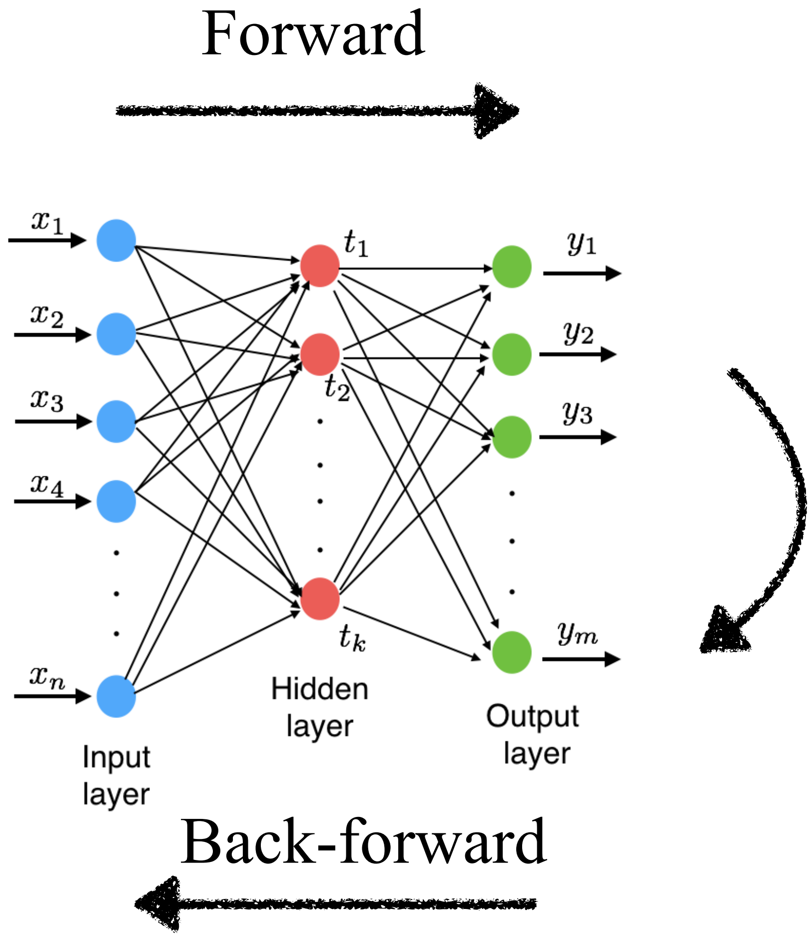
Back-propagation algorithm

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

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

with

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



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

Dataset

$$\vec{d} = [P(k), \vec{\theta}] \longrightarrow \boxed{\theta_{\mathrm{EoR}} = f(P_{21})}$$

EoR parameter (**output**)

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

(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)

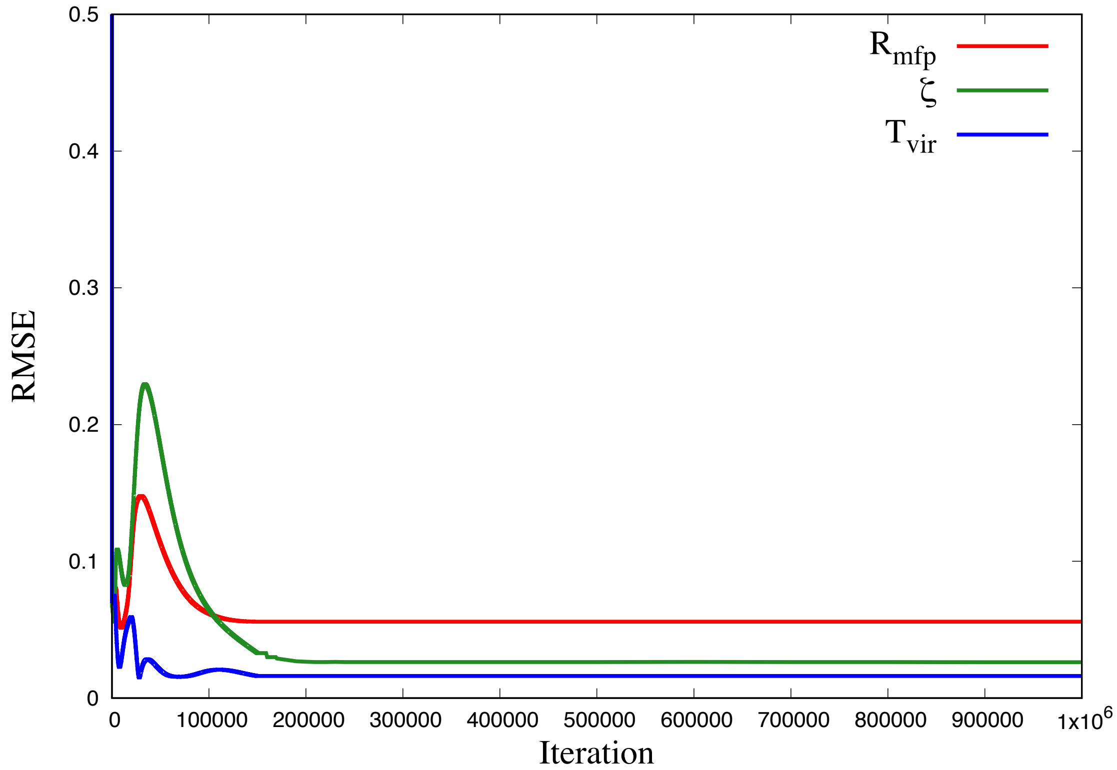
Condition

- 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_{input}=14,N_{hidden}=14,N_{output}=3 for single z.

Result

Convergence

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



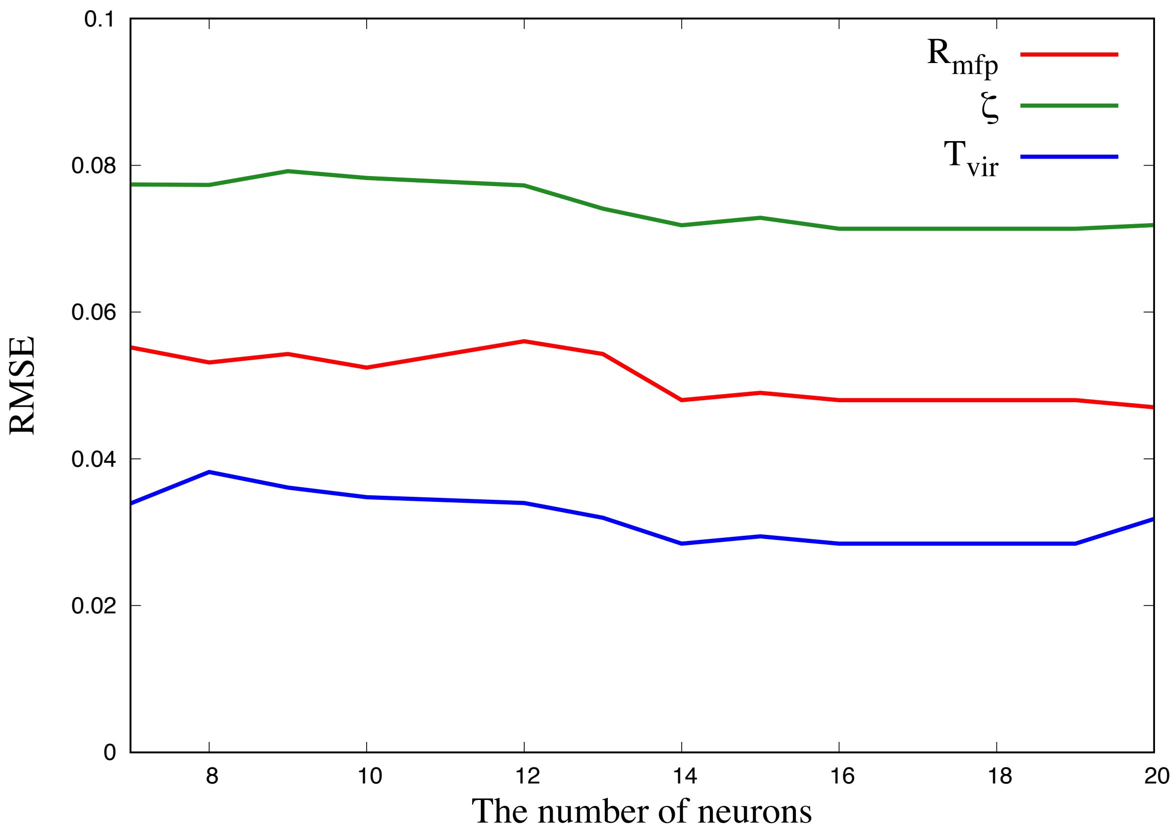
$$\mathrm{RMSE} = \sqrt{\frac{1}{N_{\mathrm{train}}} \sum_{i=1}^{N_{\mathrm{train}}} X^2}$$

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

(θ is each EoR parameter.)

More than 200000 iterations, all values of RMSE converge.

The number of neurons



- Fix 100000 iterations
- Change the number of neurons at hidden layer.
- Plot RMSE as function of the number of neurons.

Week dependence on the number of neurons.

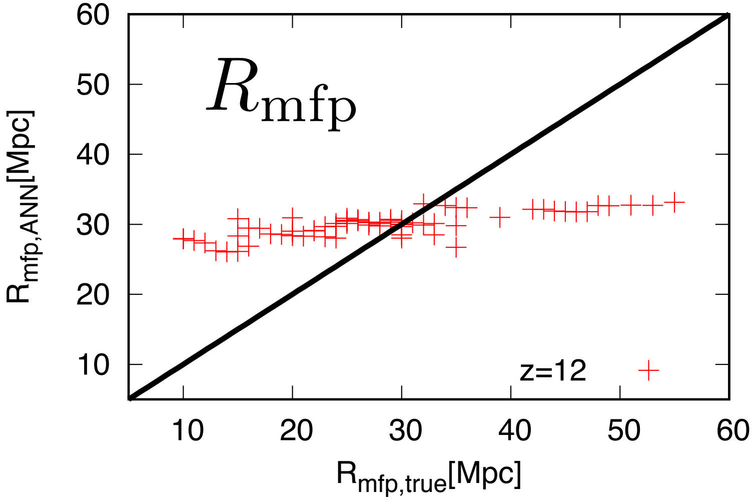
(Note)

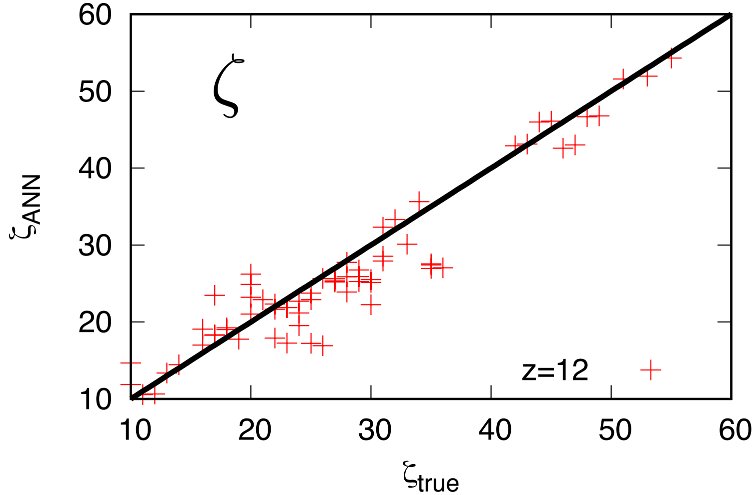
14 neurons at input layer.

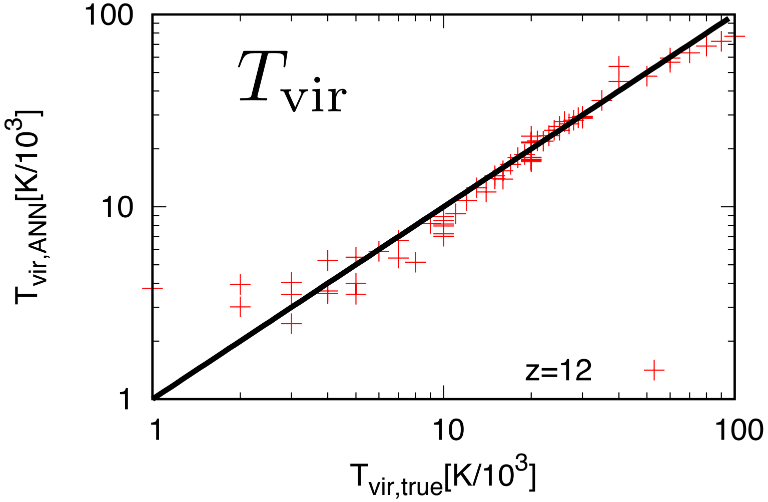
Results

z=12,PS without any noise

<u>14 neurons, 100000 iterations</u>





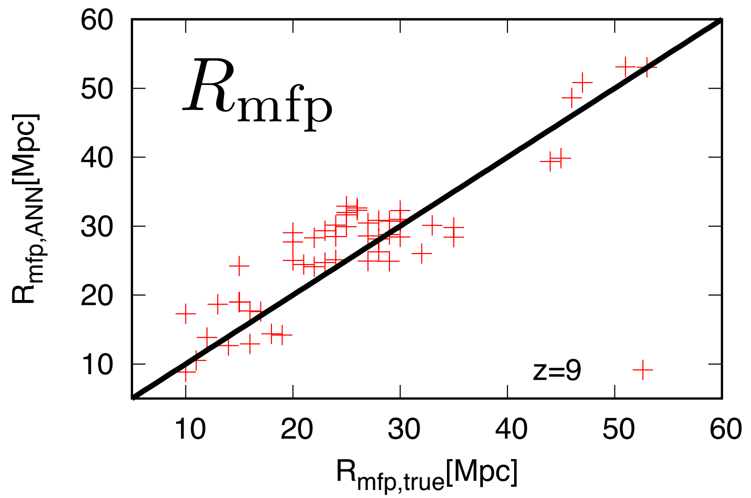


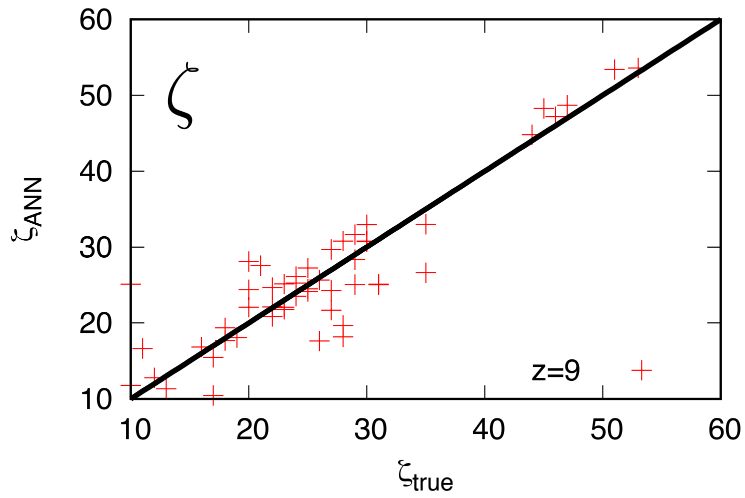
- True value .vs. Reconstructed value
- The scatter of R_mfp is large.
- Other reconstructed parameters match true one relatively well.

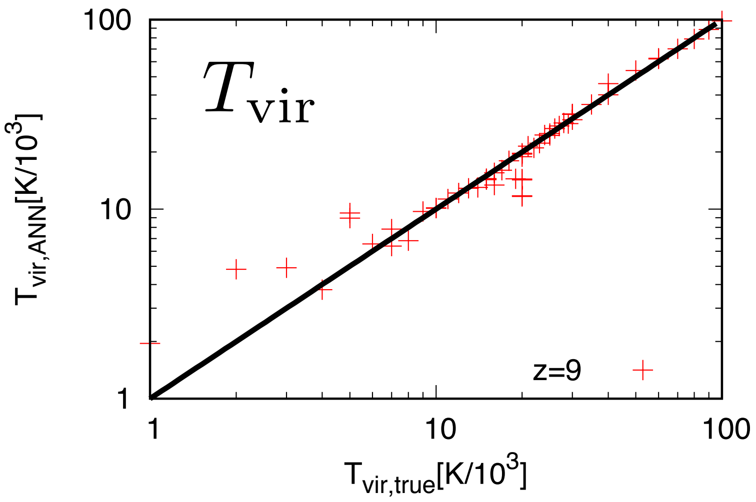
Results

z=9,PS without any noise

<u>14 neurons, 100000 iterations</u>







•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.

Noise & Cosmic variance

In order to evaluate the effect of noise, we include *both* **thermal noise** and **cosmic variance** for *test data*.

However,

We include *only* cosmic variance for *training data.*

How?

•We calculate comic variance from 10 realizations of initial condition.

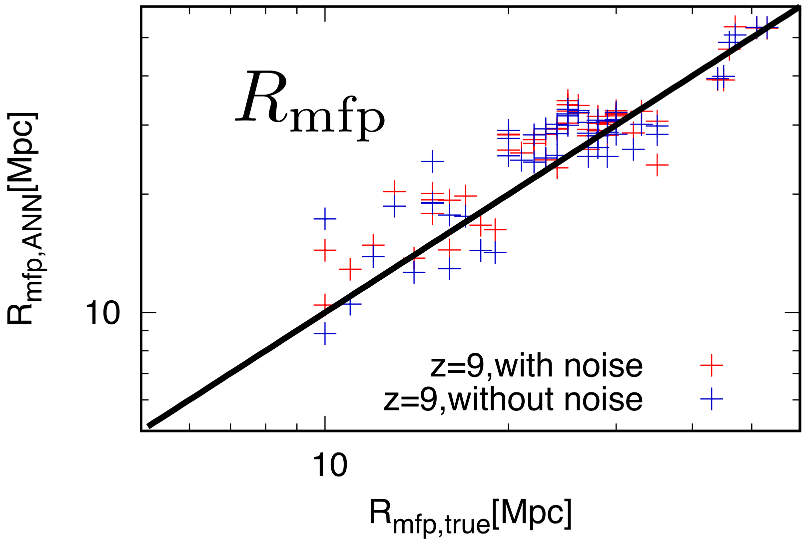
•Produce 50 noise realizations for each parameter set by adding random gaussian noise .

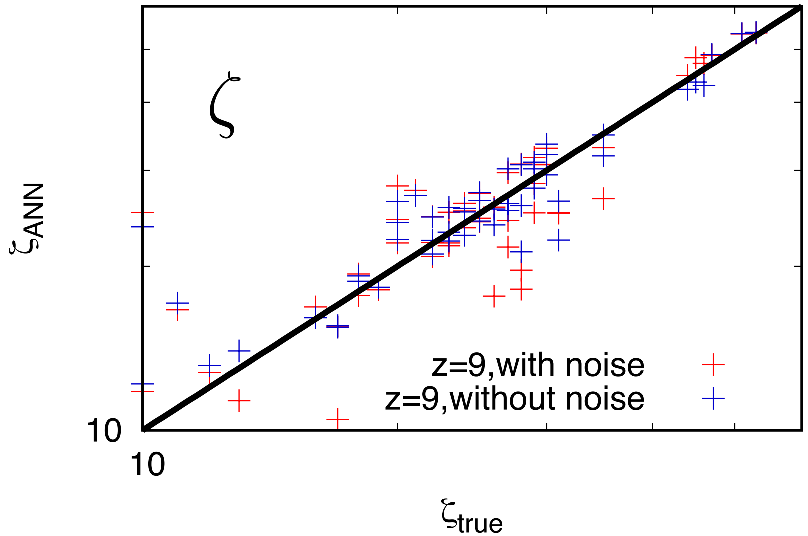
We calculate thermal noise assuming SKA observation based on (Morales et al 2005).

Results

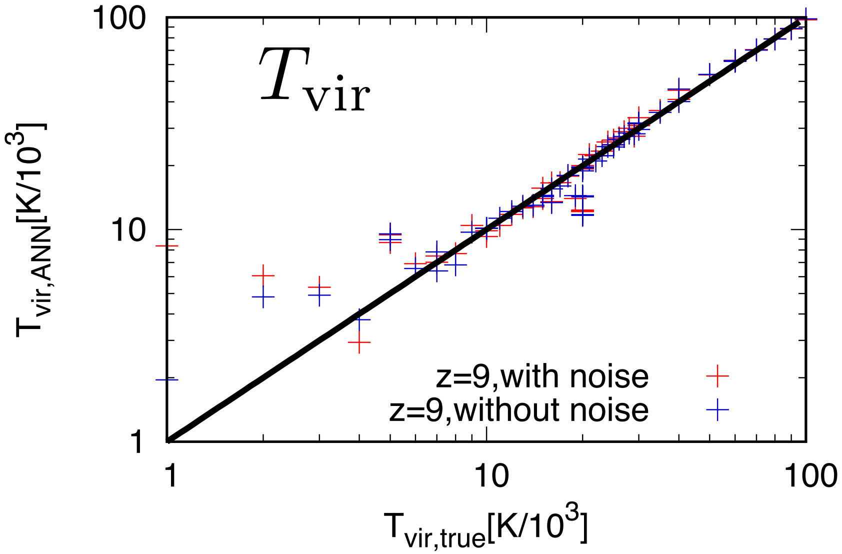
z=9, PS including thermal noise and cosmic variance

14 neurons, 100000 iterations





Red : with noise
Blue : without noise

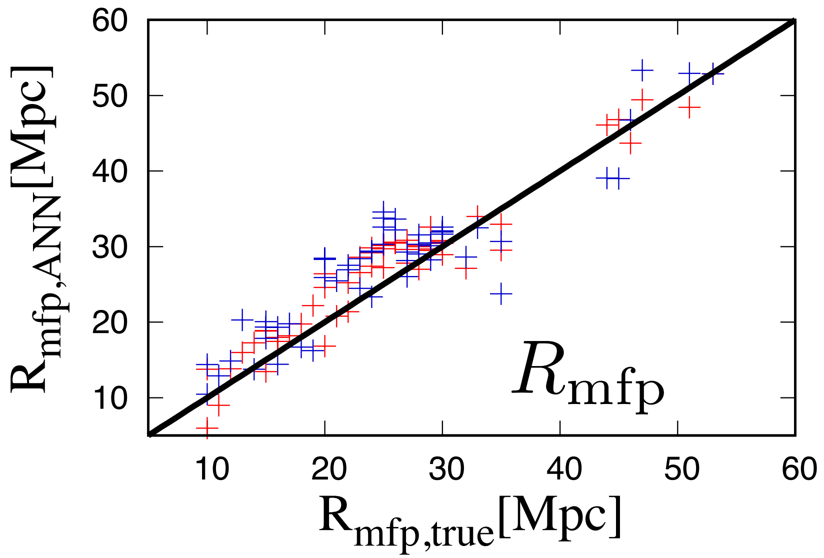


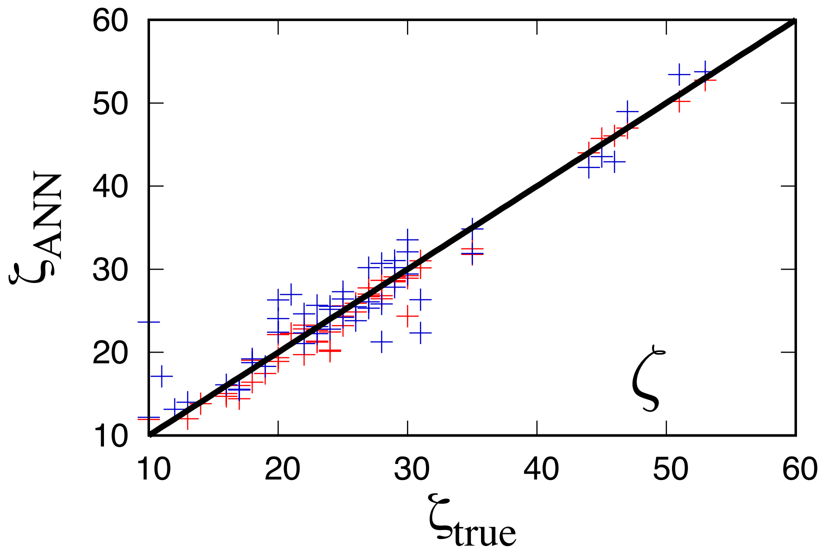
	$\mathrm{RMSE_{wo/noise}}$	$\mathrm{RMSE_{w/noise}}$
R_{mfp}	0.228	0.258
ζ	0.271	0.288
$\log(T_{\mathrm{vir}})$	0.027	0.038

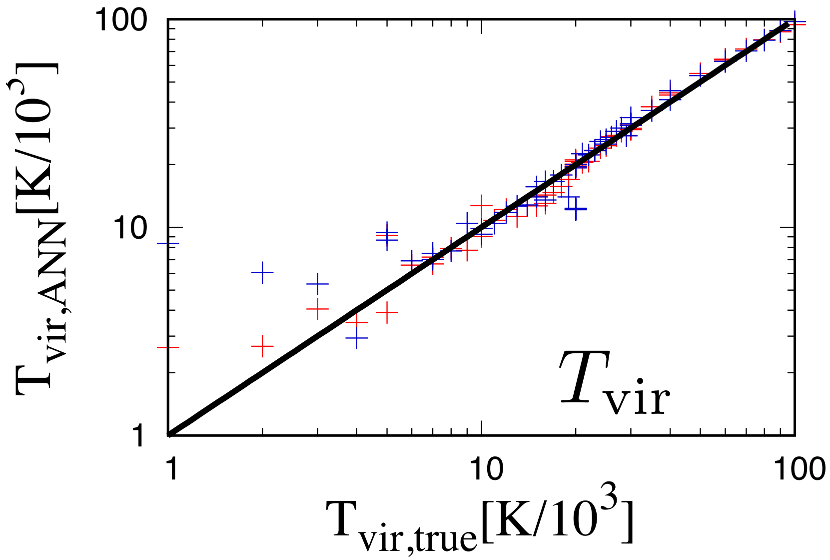
Accuracy is better when noises are not included.

Results

z=9, 10, 11. PS including thermal noise and cosmic variance







<u>42 neurons,100000 iterations</u>

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**

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

Compare with MCMC

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

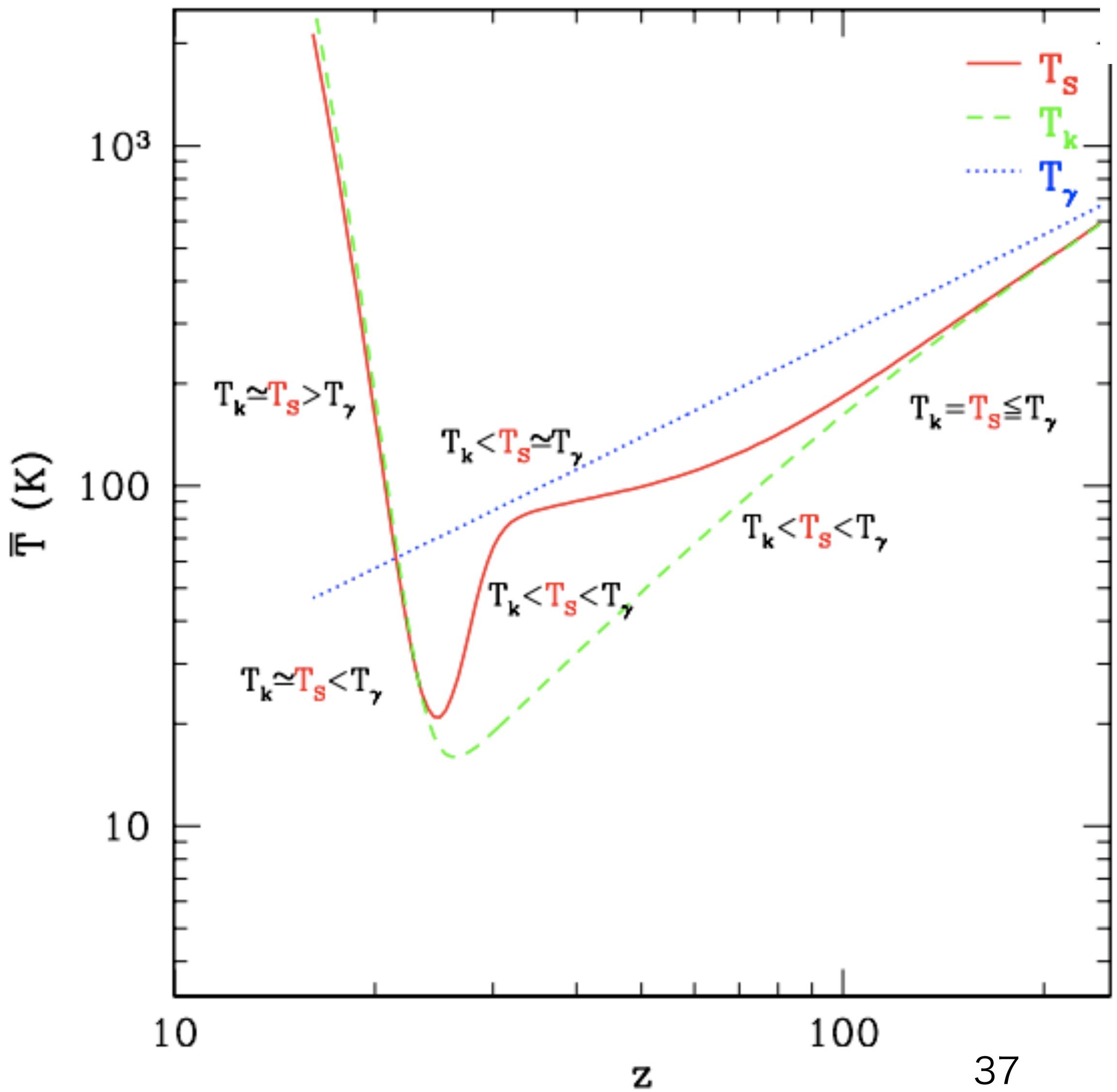
Summary

- 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.
- We’re now trying to analyse the epoch of cosmic dawn with more parameters (with A.Fialkov).

Backup

Thermal history

Mesinger et al 2010



$$T_S^{-1} = \frac{T_{\rm CMB}^{-1} + x_\alpha T_\alpha^{-1} + x_{\rm K} T_{\rm K}^{-1}}{1 + x_\alpha + x_{\rm K}}$$

Blue：CMB temperature

$$\propto (1+z)$$

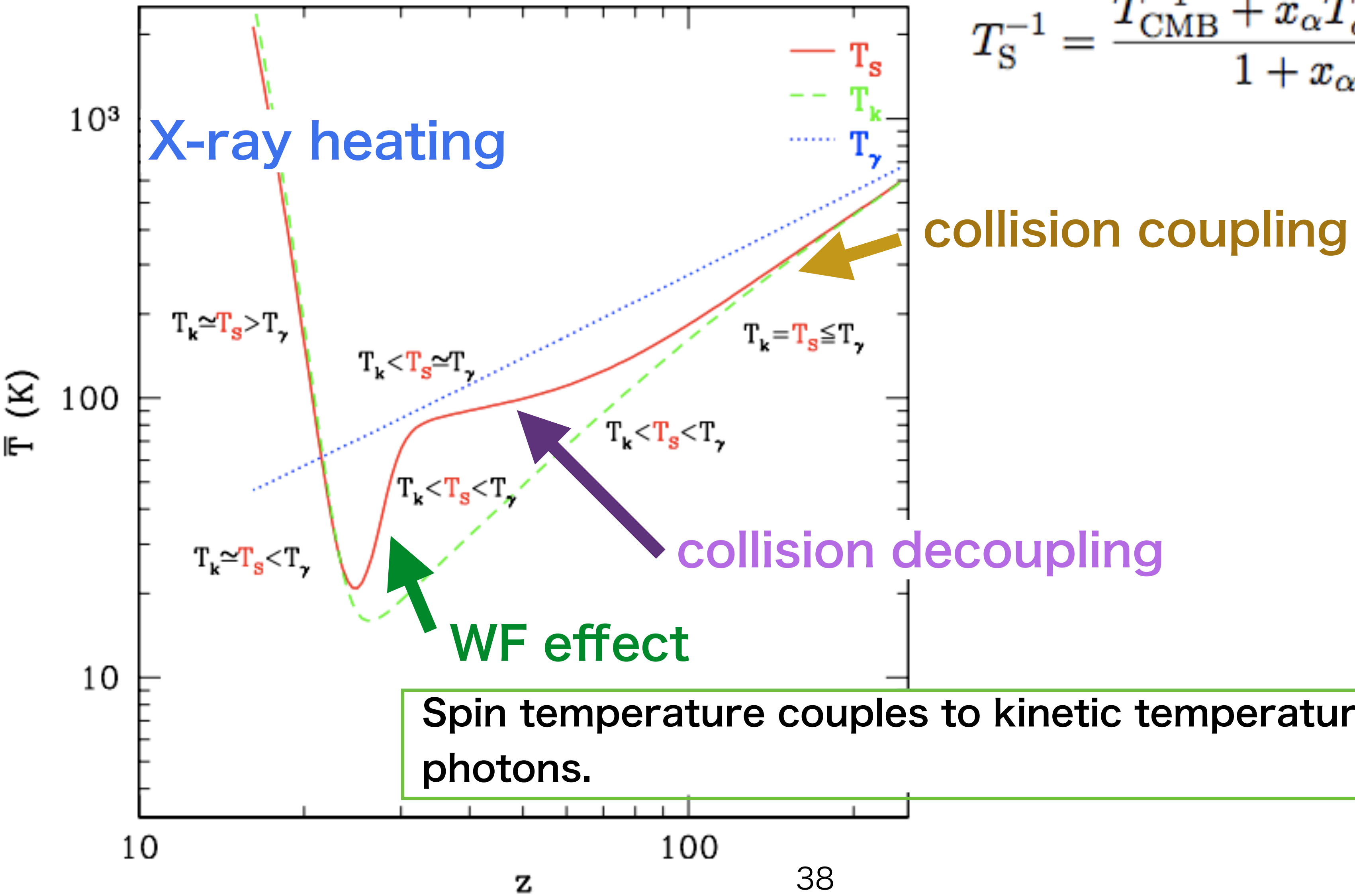
Green：kinetic temperature

$$\propto (1+z)^2$$

Red : spin temperature

Thermal history

Mesinger et al 2010



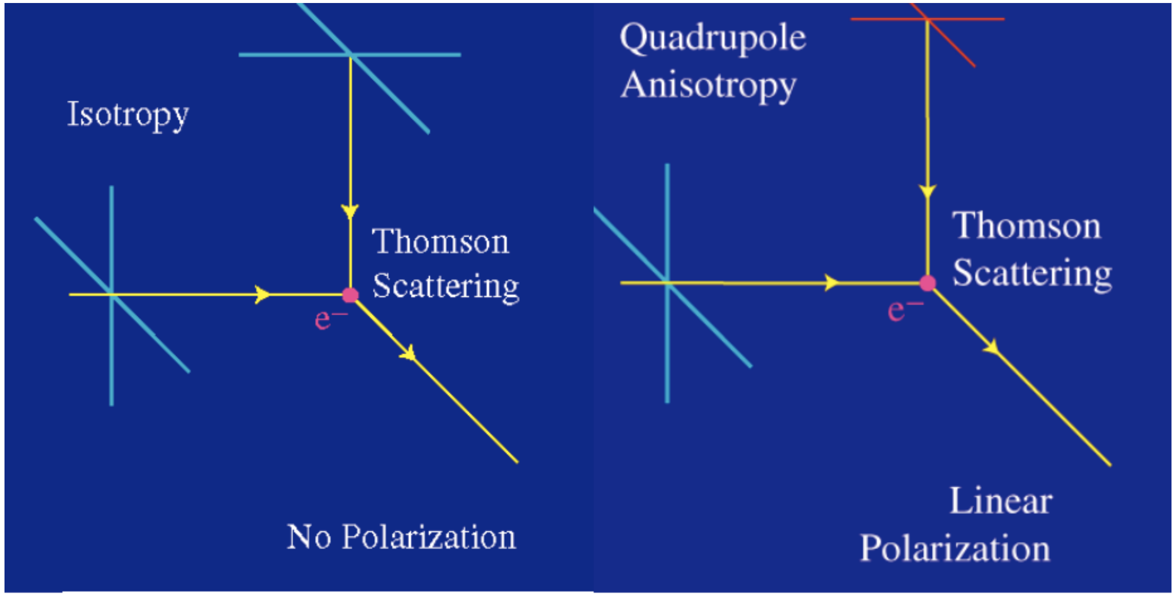
$$T_S^{-1} = \frac{T_{\rm CMB}^{-1} + x_\alpha T_\alpha^{-1} + x_{\rm K} T_{\rm K}^{-1}}{1 + x_\alpha + x_{\rm K}}$$

Spin temperature couples to kinetic temperature via Ly-α photons.

Constraints

The observation of the CMB polarization

→the optical depth of Thomson scattering



Optical depth

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

$$\tau_e = 0.078 \pm 0.0019$$

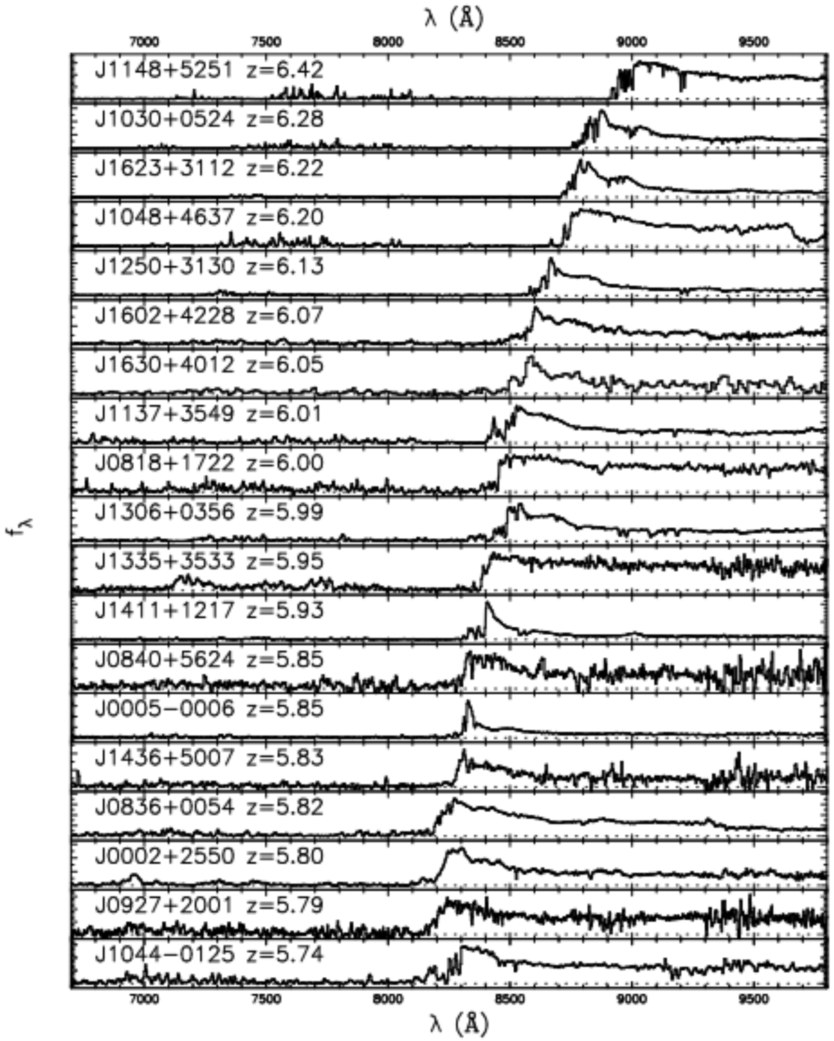
$$z_r = 9.9^{+1.8}_{-1.6}$$

(Planck collaboration 2015)

Constraints

high-z QSO absorption

→ constraint on the epoch where the EoR finishes.



Gun-Peterson test

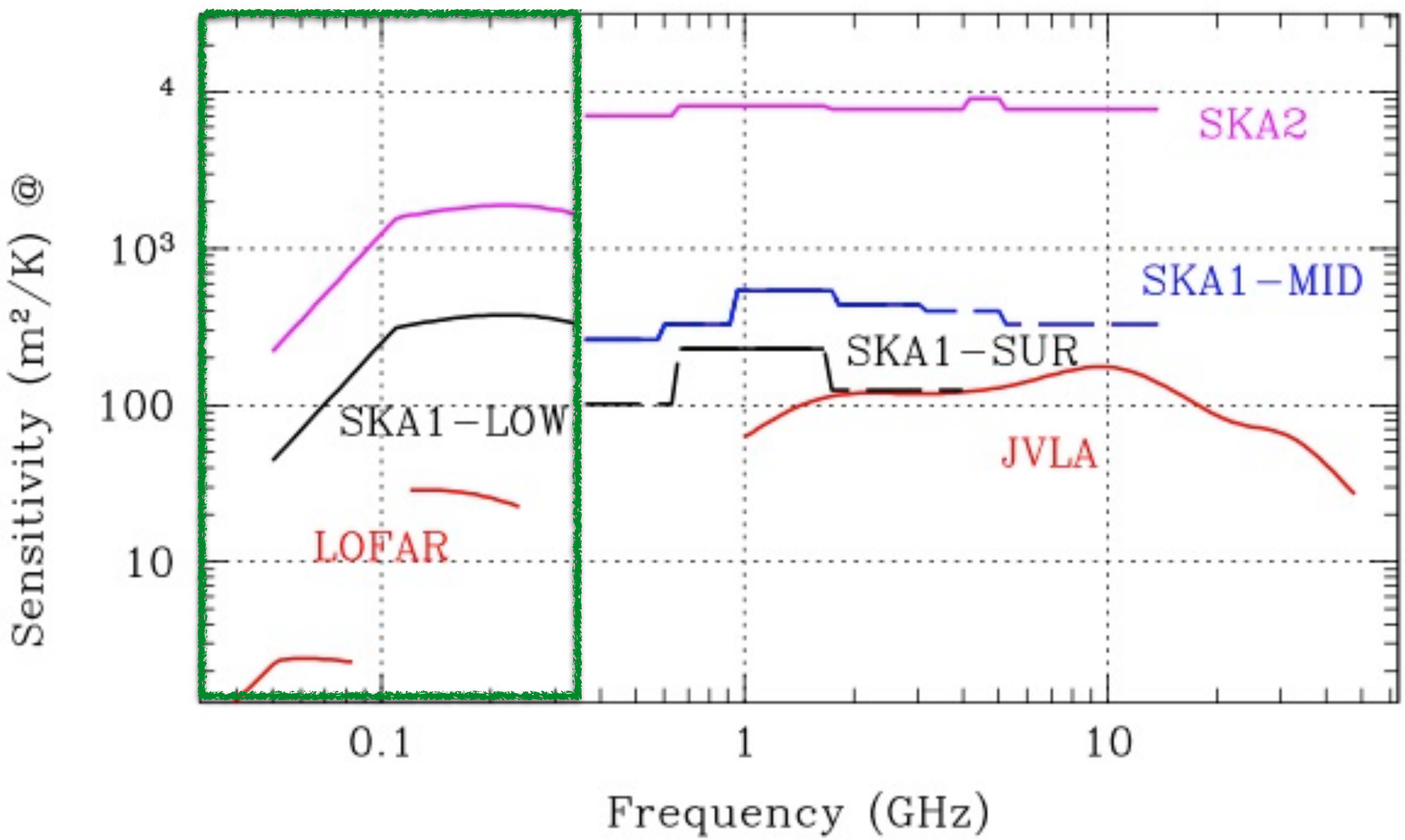
If the neutral hydrogen exists, it absorbs the Ly-alpha photons.

No emission line !

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

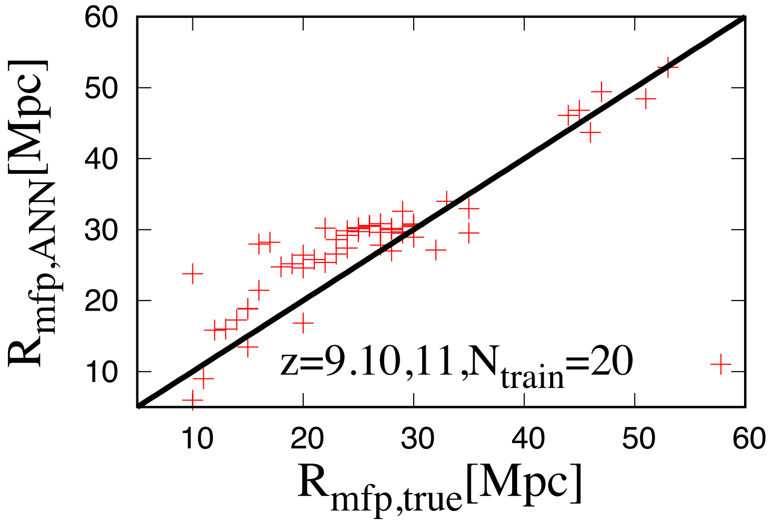
Sensitivity

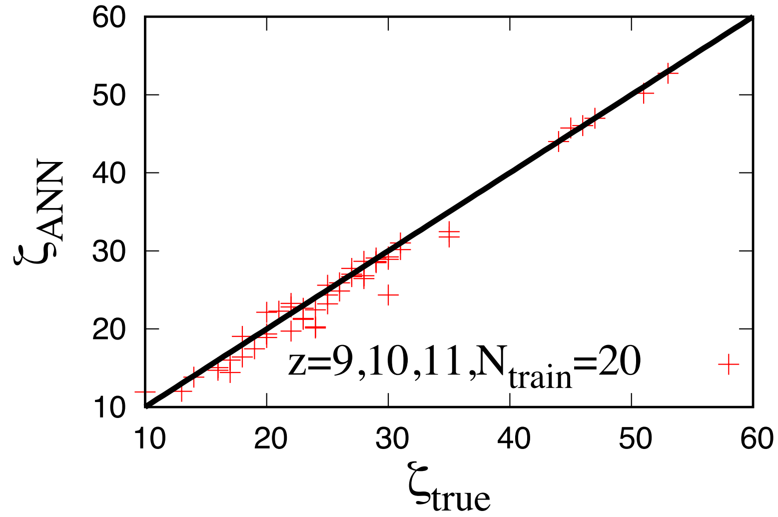
Sensitivity

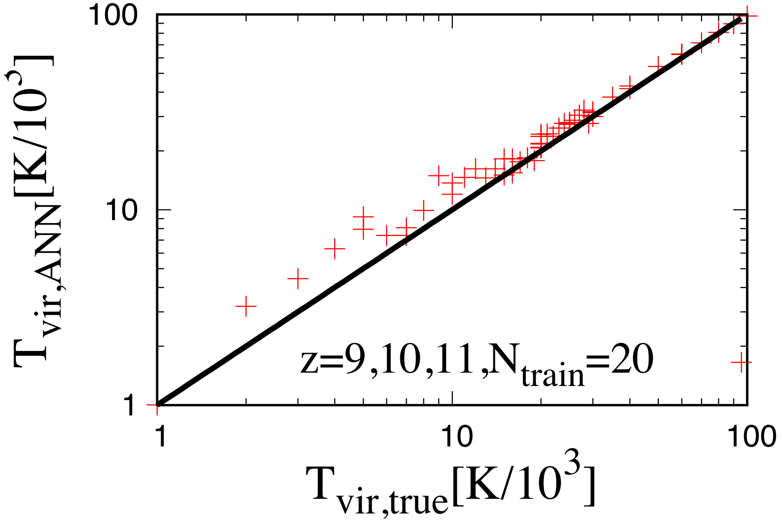


Results

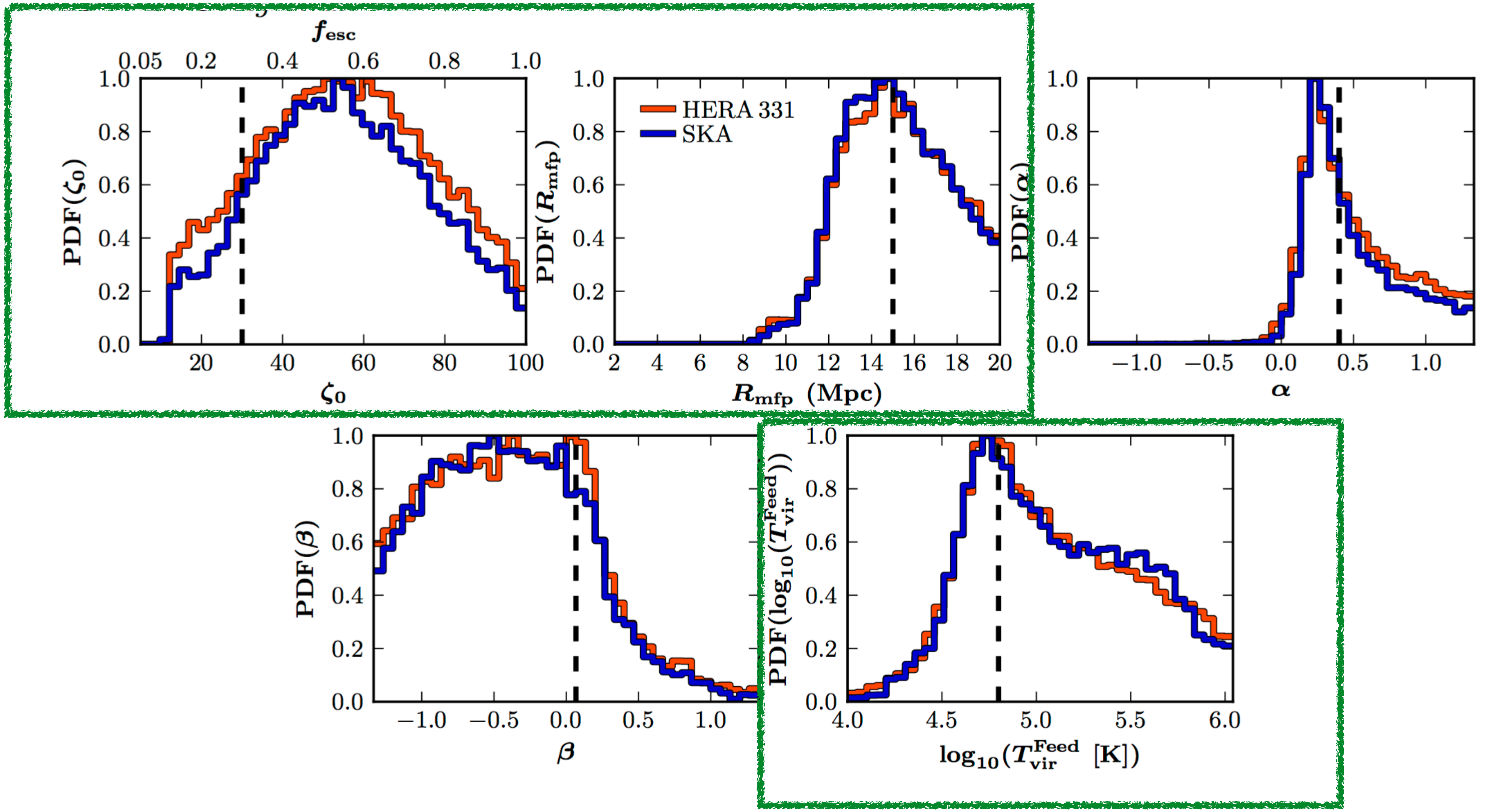
Reduce the number of training data. (N=20)







MCMC



The probability distribution function of each parameter.

Activation function