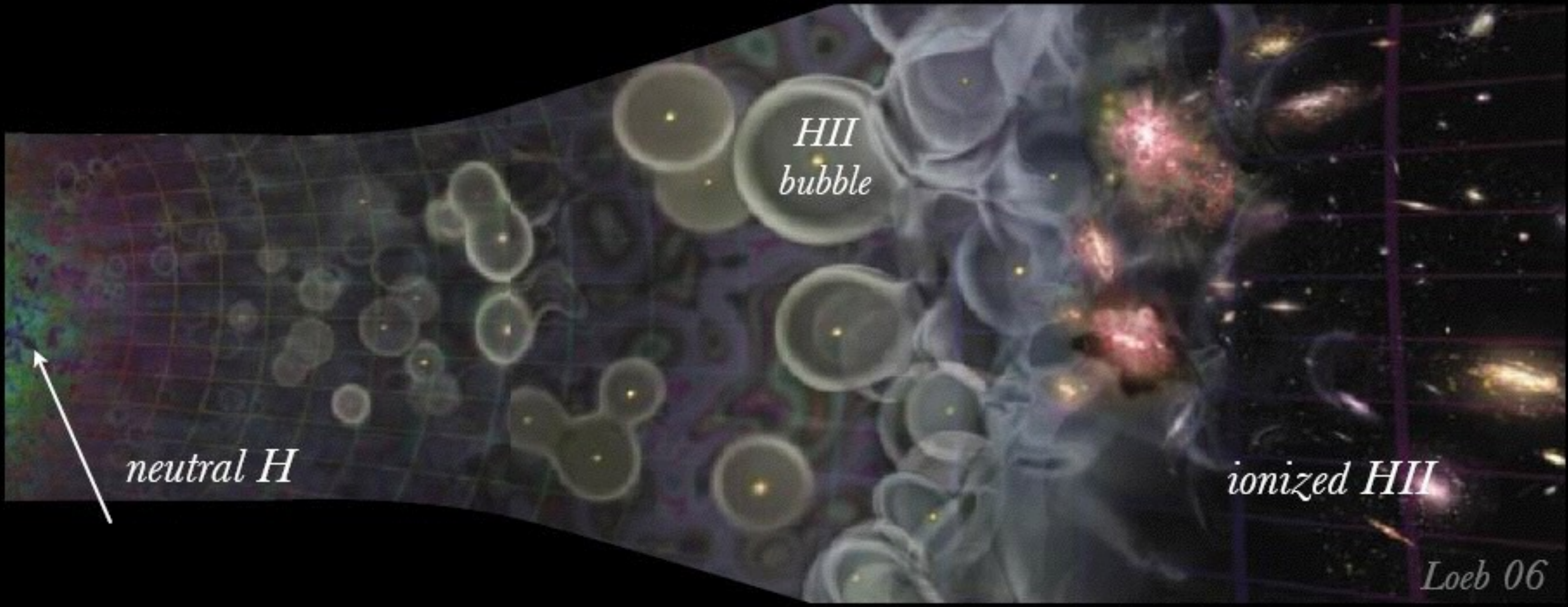


Recovering HII bubble size distribution with artificial neural network

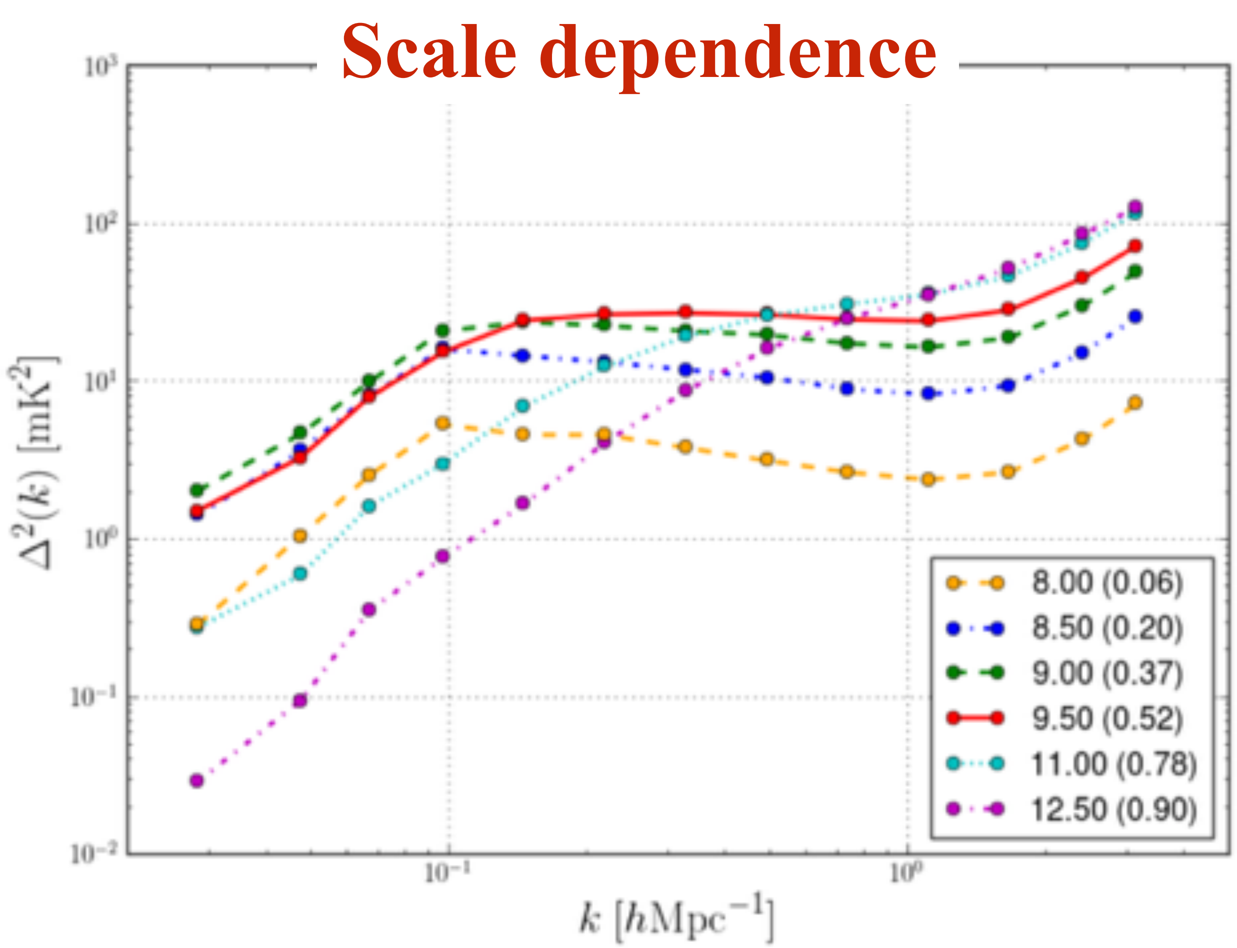
Hayato Shimabukuro (Tsinghua University)
with Yi Mao

21cm power spectrum

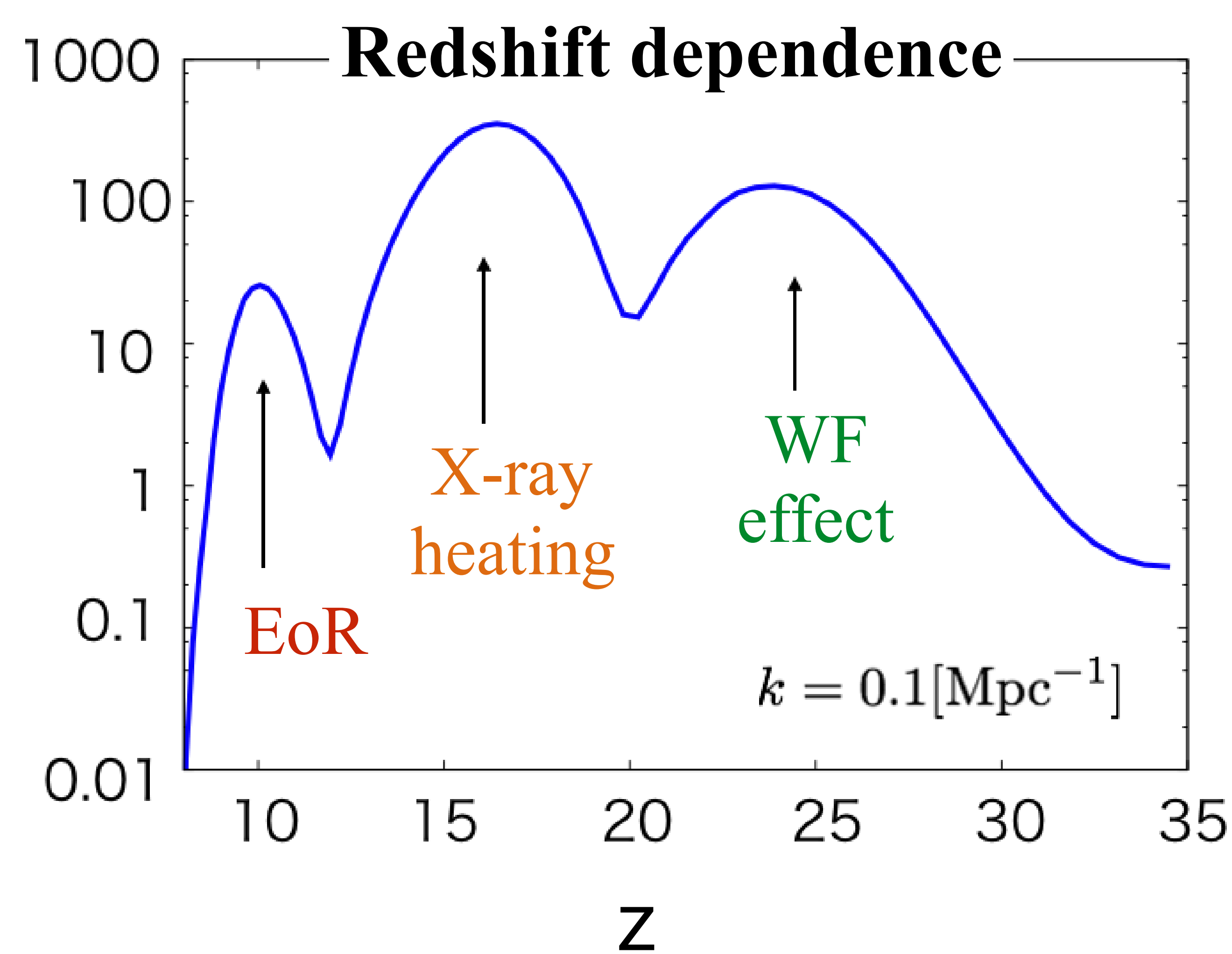
We first aim to detect 21cm signal **statistically**.

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

(We use 21cmFAST)



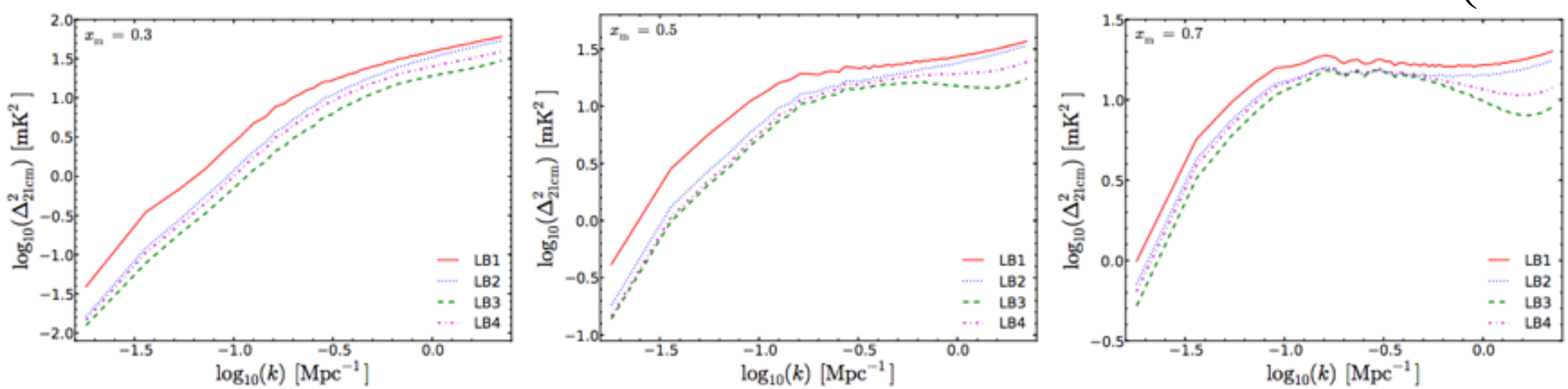
Pober et al (2014)



Distinguish EoR models

We often discuss distinguishing EoR models on the basis of 21cm power spectrum. (e.g. EoR parameters, feedback, ionising sources)

Dixon et al.(2016)



LB1: only high mass halo M>10^9 M_sun

LB2: including low mass (10^8-10^9 M_sun)

LB3: including low mass with radiative feedback

LB4: same with LB3 but mass dependent FB

Bubble size distribution (BSD)

HII bubble size distribution is another method to break degeneracy in EoR models.

However!!

Difficult to determine bubble size and measure BSD.

Some methods are suggested to measure BSD from **ionised map**

- ***Mean free path***

  Mesinger & Furlanetto (2007)

- ***Spherical average method***

  Zahn+ (2007)

- ***Friend-of-Friend (FoF)***

  Friedrich+ (2011)

- ***Watershed***

  Lin+ (2016)

BSD from 21cm tomography

We practically measure BSD from **21cm tomography data**.

Different from the case of measuring BSD from ionisation map.

•ionised map	**•21cm map**
binary map (ionized or neutral)	**Not** binary map

Therefore,

we need to transform 21cm map into binary map (ionized or neutral).

→ We need to fix threshold to distinguish fields.

It is difficult because 21cm map depends on both x, δ

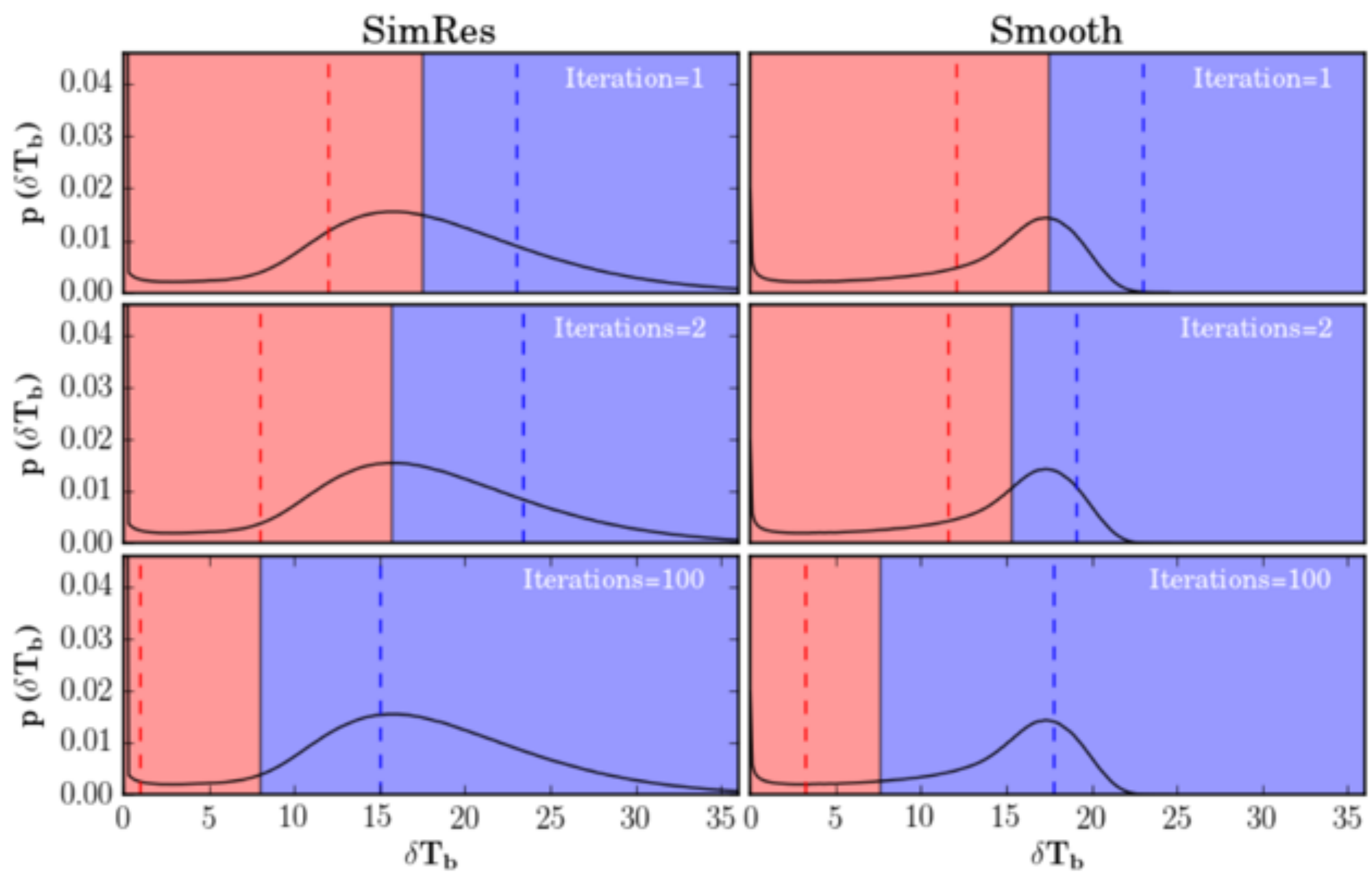
Making binary map

Giri+ 2017

Giri+ 2017 focuses on the fact that 21cm PDF is bimodal.

•K-means method

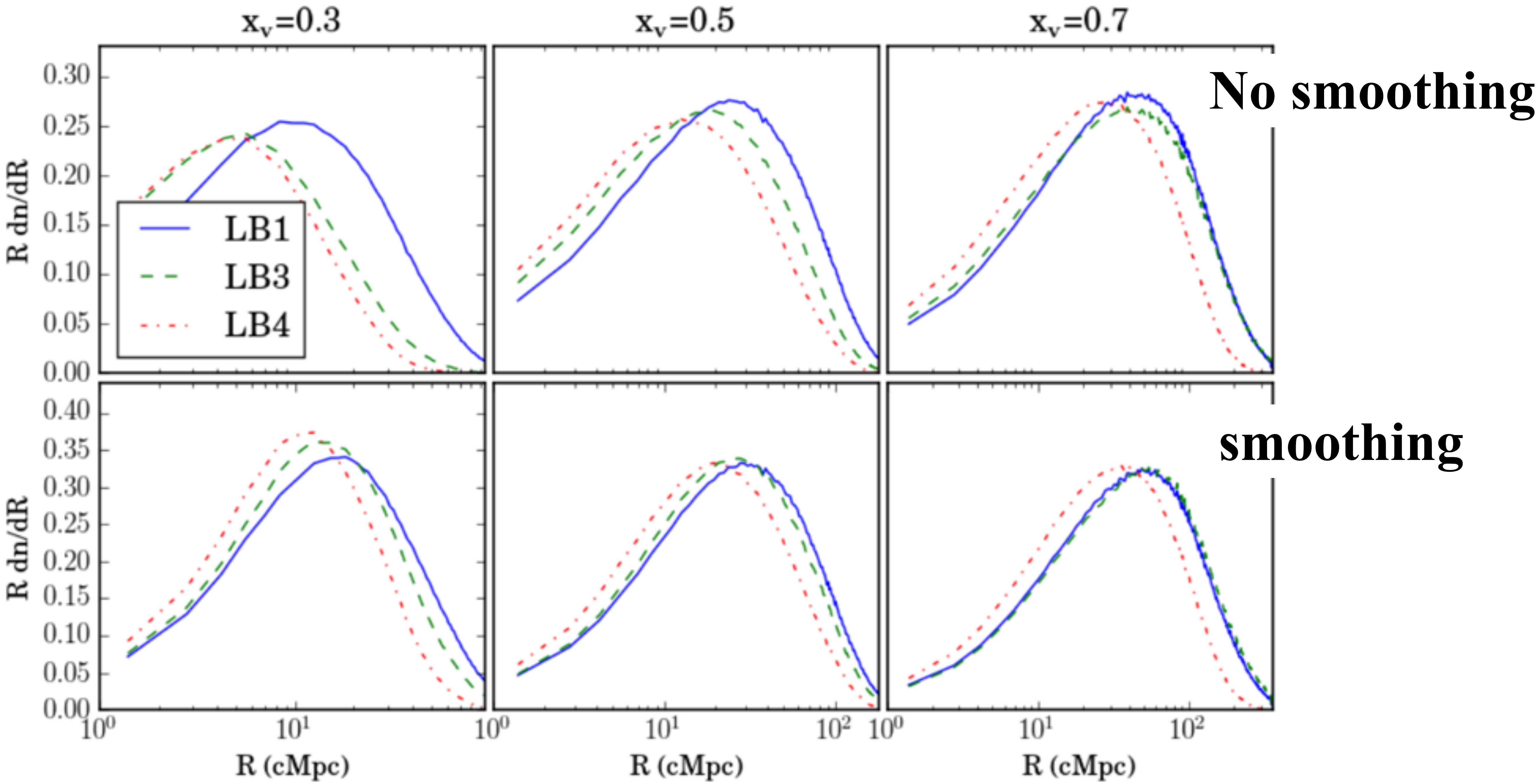
unsupervised learning algorithm for clustering problem



BSD for different sources

Giri et al 2017

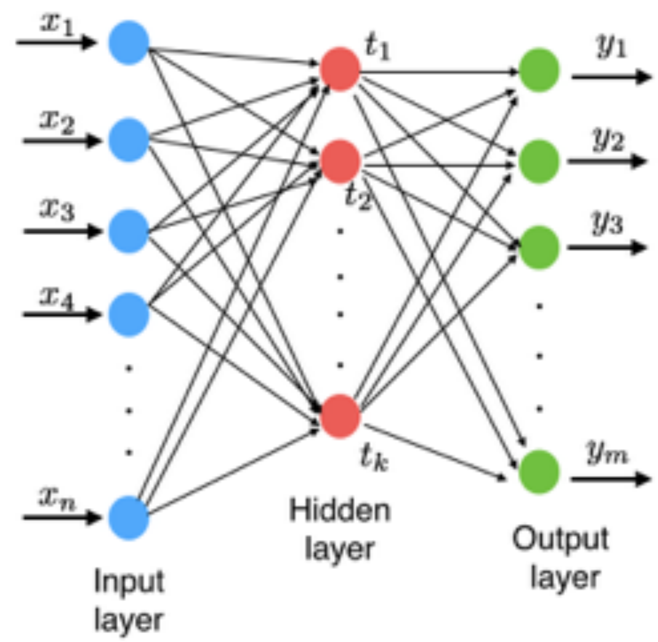
MFP-BSD



Our strategy

Input

21cm power spectrum

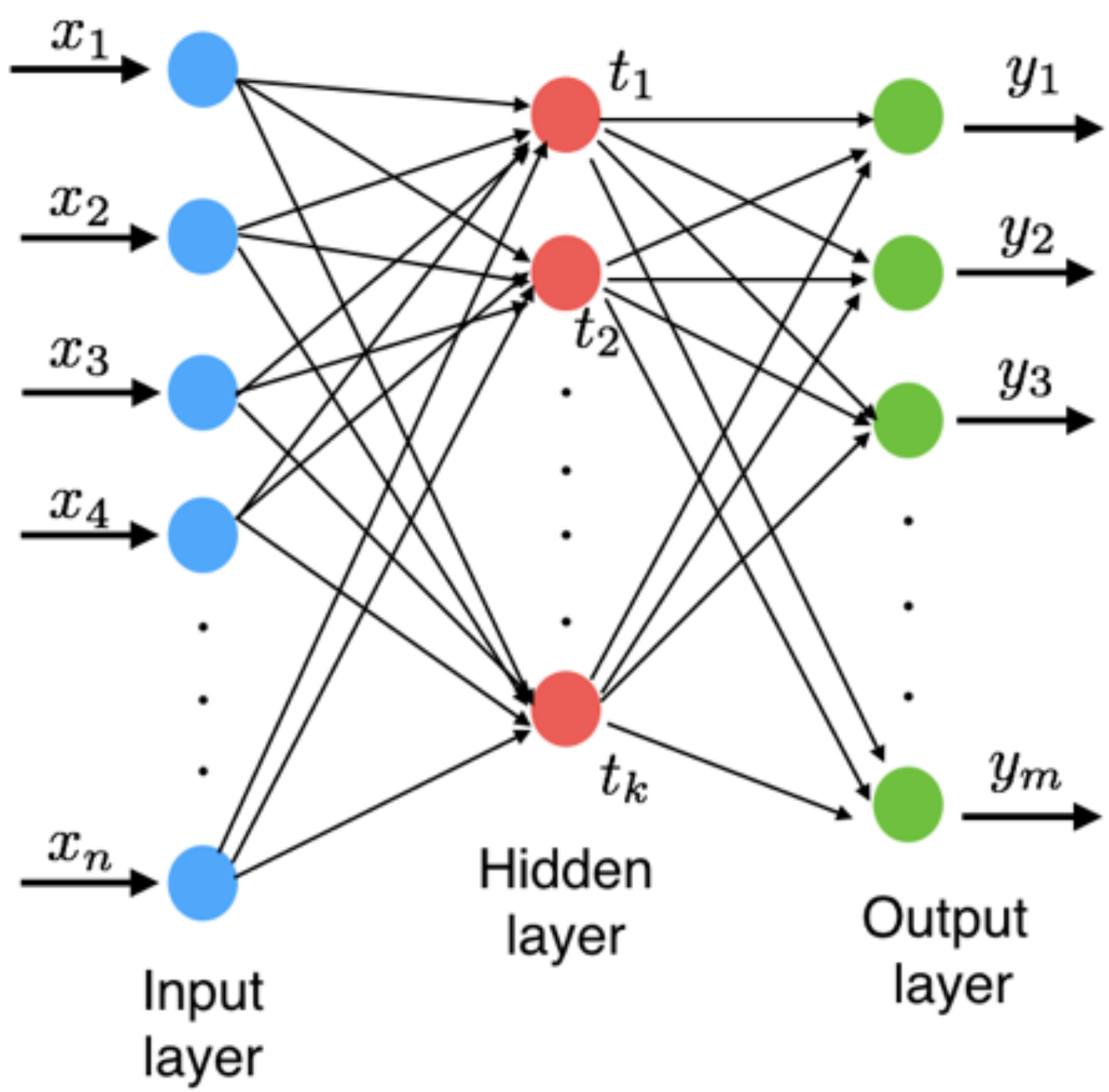


Output

HII size distribution

Our datasets consist of 21cm power spectrum as input data and bubble size distribution as output data.

Artificial Neural Network (ANN)



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

- Training network with training dataset, 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.

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

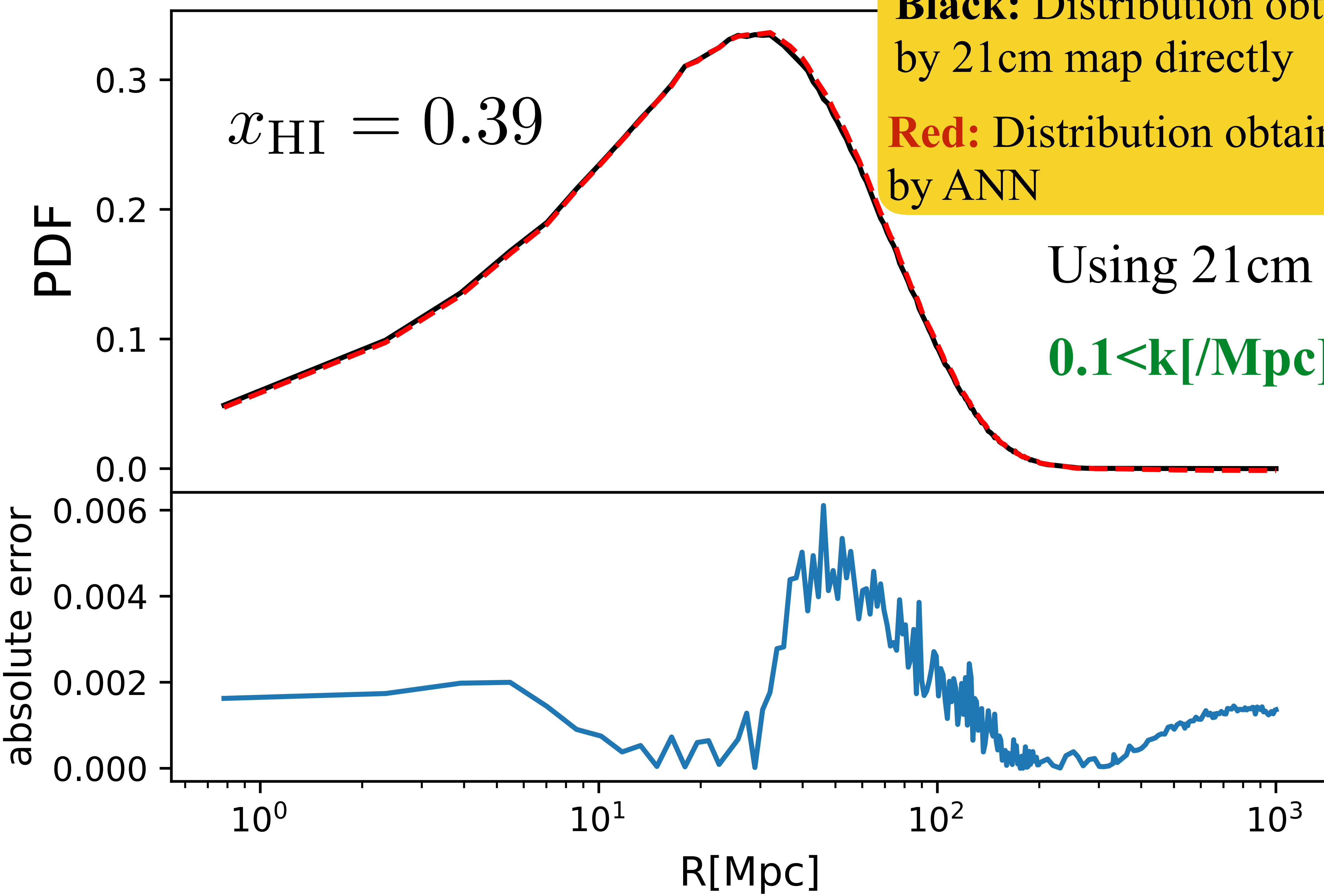
Motivation

- Some previous studies measure HII size distribution from 21cm 3D map directly.
- From observational aspect, we require good angular resolution to make 21cm map.
- Therefore, I attempt to recover HII size distribution from 21cm PS that does not require making 21cm map.

Setup

- 1000 EoR models
- 48000 training datasets (20% of which is used for validation)
- 2000 test datasets
- 21cm PS is ranged from k=0.11/Mpc to 1.1/Mpc with 14 bins
- 5 hidden layers
- 212 neurons at each hidden layer

Recovered BSD



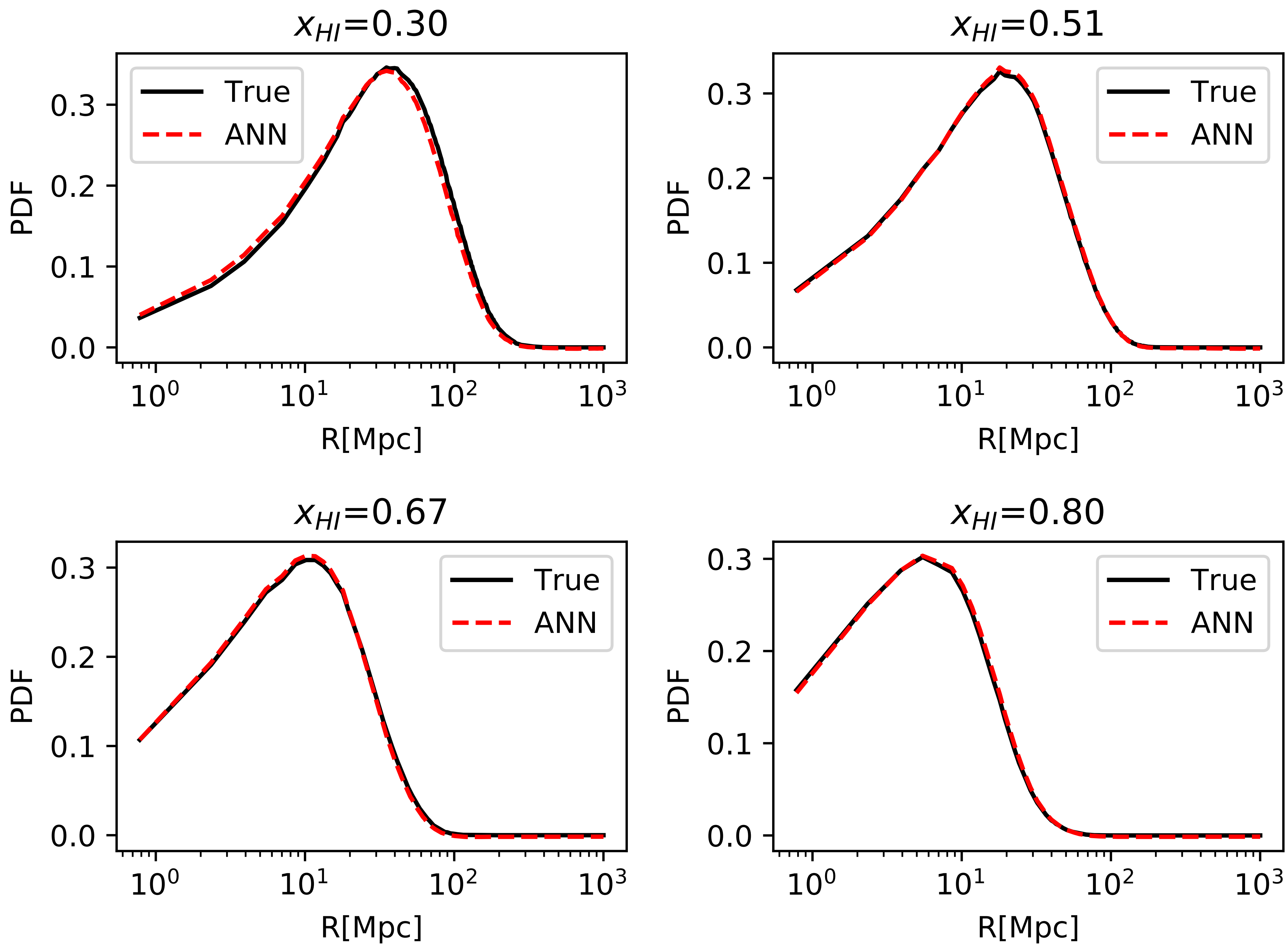
Black: Distribution obtained by 21cm map directly

Red: Distribution obtained by ANN

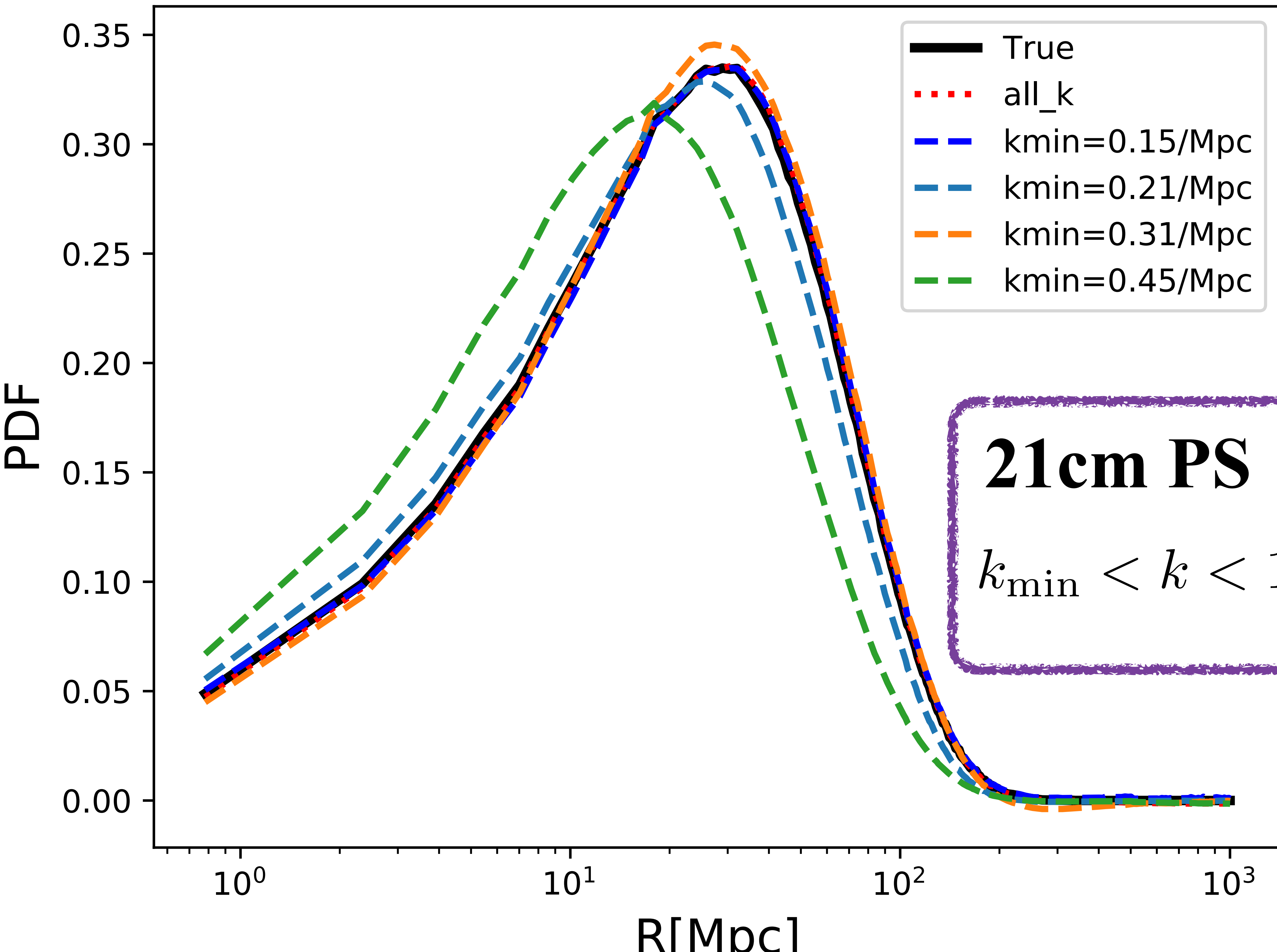
Using 21cm PS at

0.1<k[/Mpc]<1

Different epoch of reionization



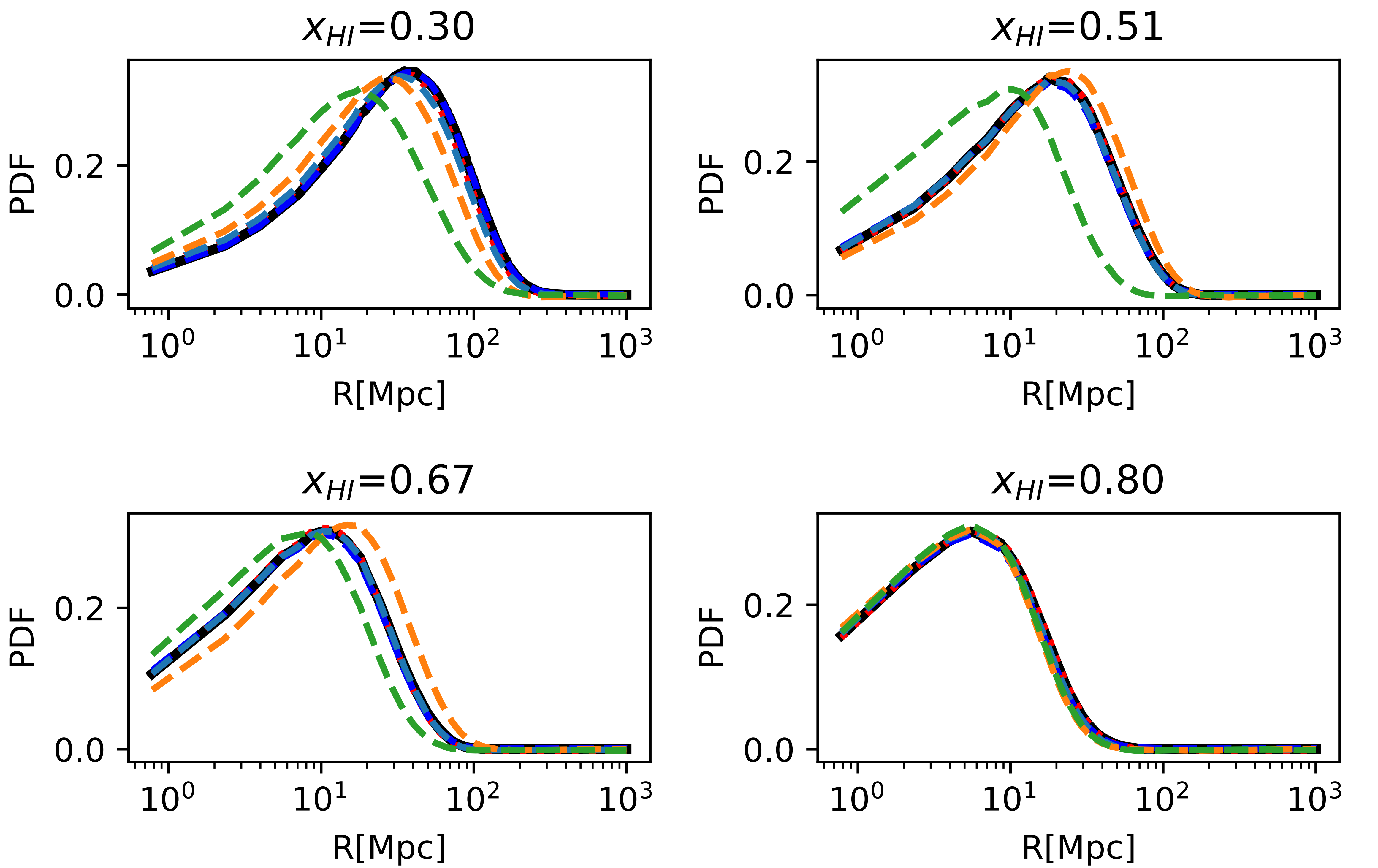
Scale dependence



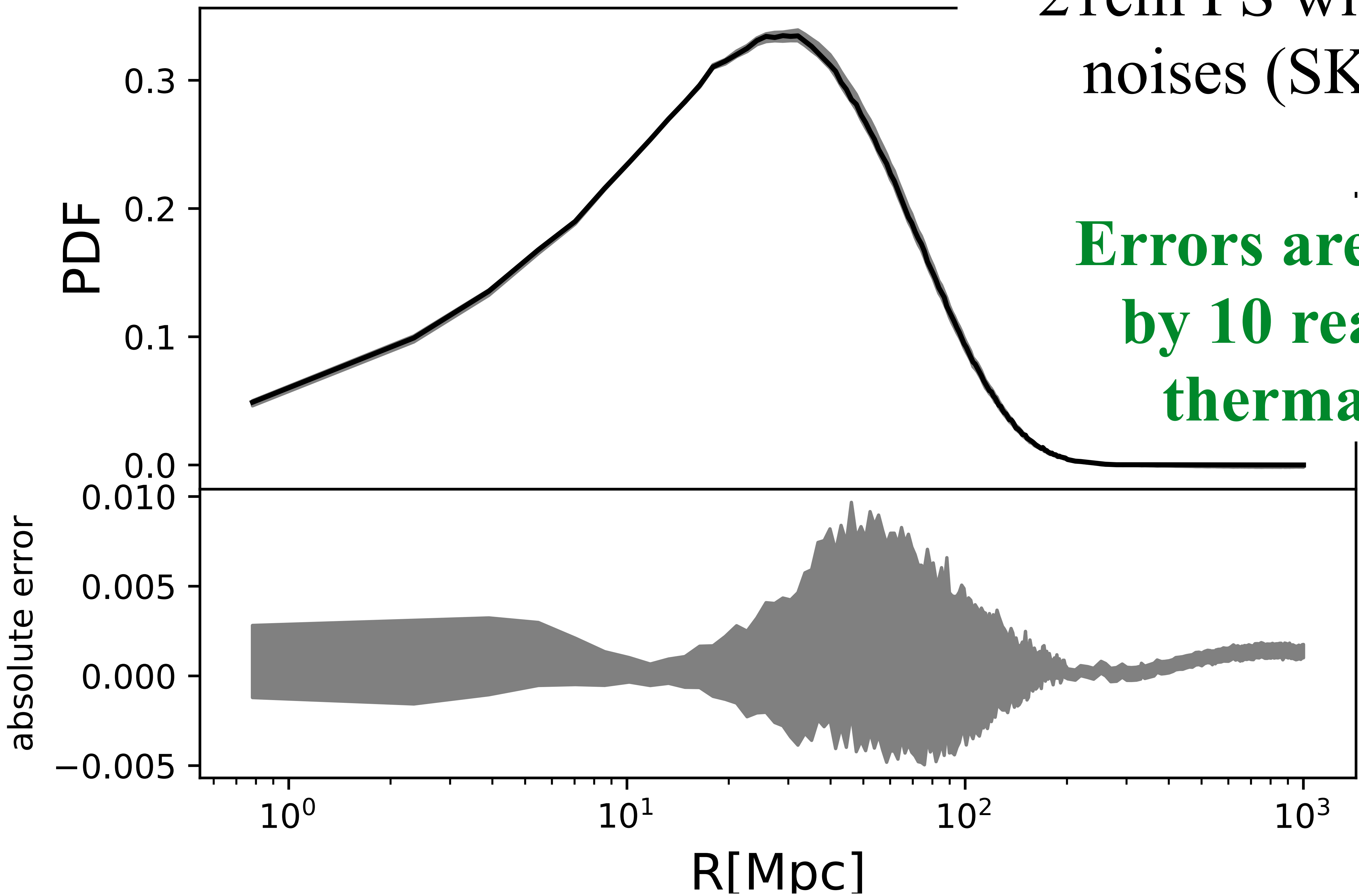
21cm PS

$k_{\min} < k < 1.1\mathrm{Mpc}^{-1}$

Scale dependence



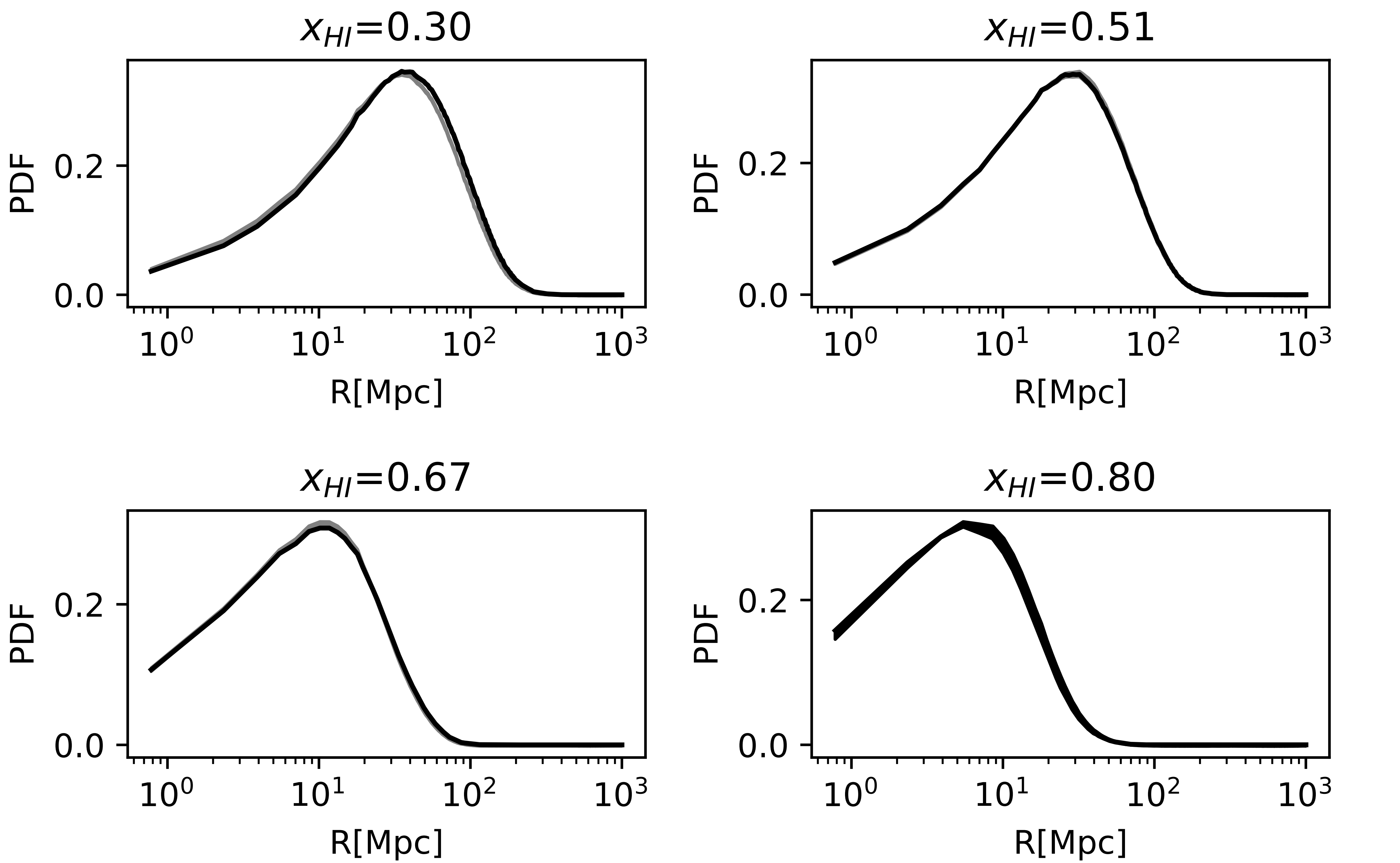
Thermal noise



21cm PS with thermal noises (SKA level)

Errors are estimated by 10 realizations thermal noises

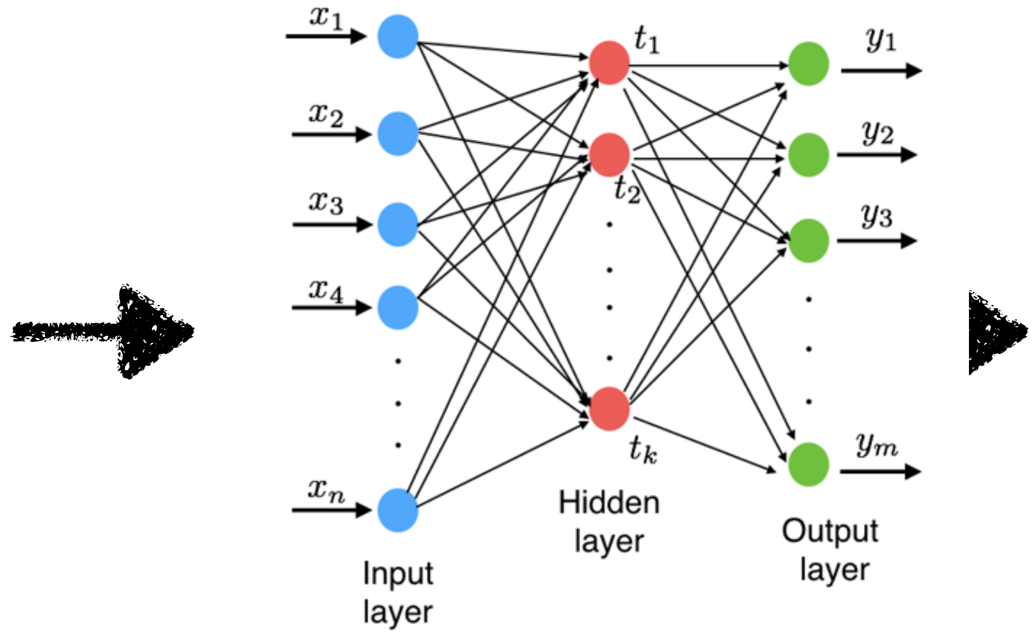
Thermal noise



Other ideas?

Input

21cm power spectrum



Output

bubble size distribution

Escape fraction

global signal

Minkowski functionals

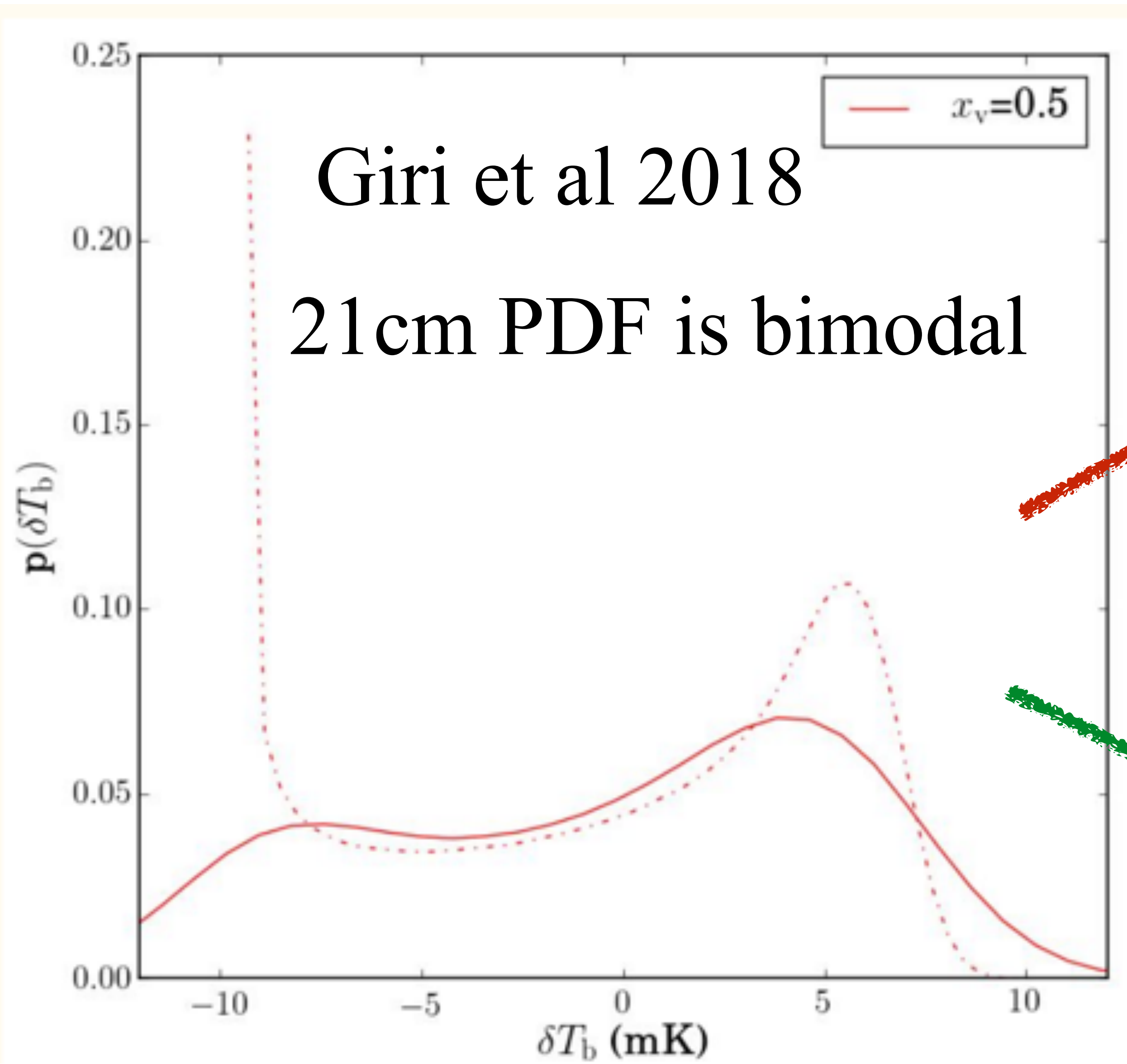
Summary

- We applied artificial neural networks (ANN) to analysis of 21cm signal.
- Reconstructed EoR parameters and HII size distribution are good agreement with true values.
- (Future work) Are there other 21cm observables which we can apply machine learning to ?

Non-Gaussianity of 21cm fluctuations

21cm PS analysis is useful, but it is imperfect to describe 21cm fluctuations.

21cm-PDF is highly non-Gaussian owing to astrophysical effects.



Higher order statistics !

Bispectrum

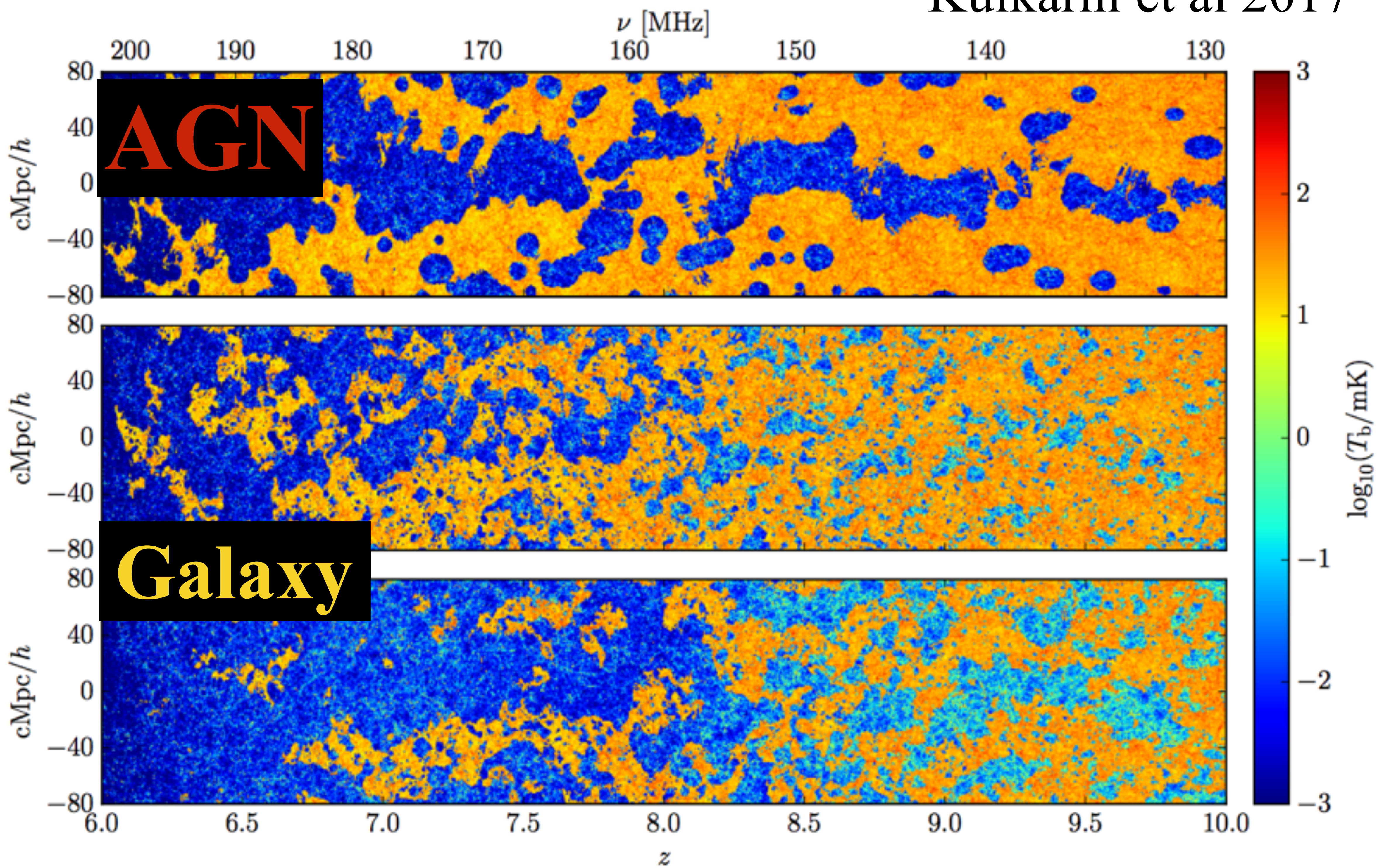
HS+ 2016, Watkinson+ 2017, Majumdar+ 2018

Skewness

HS+2015, Kubota + 2016

Morphology of ionised bubble

Kulkarni et al 2017



K-means method

Giri+ 20

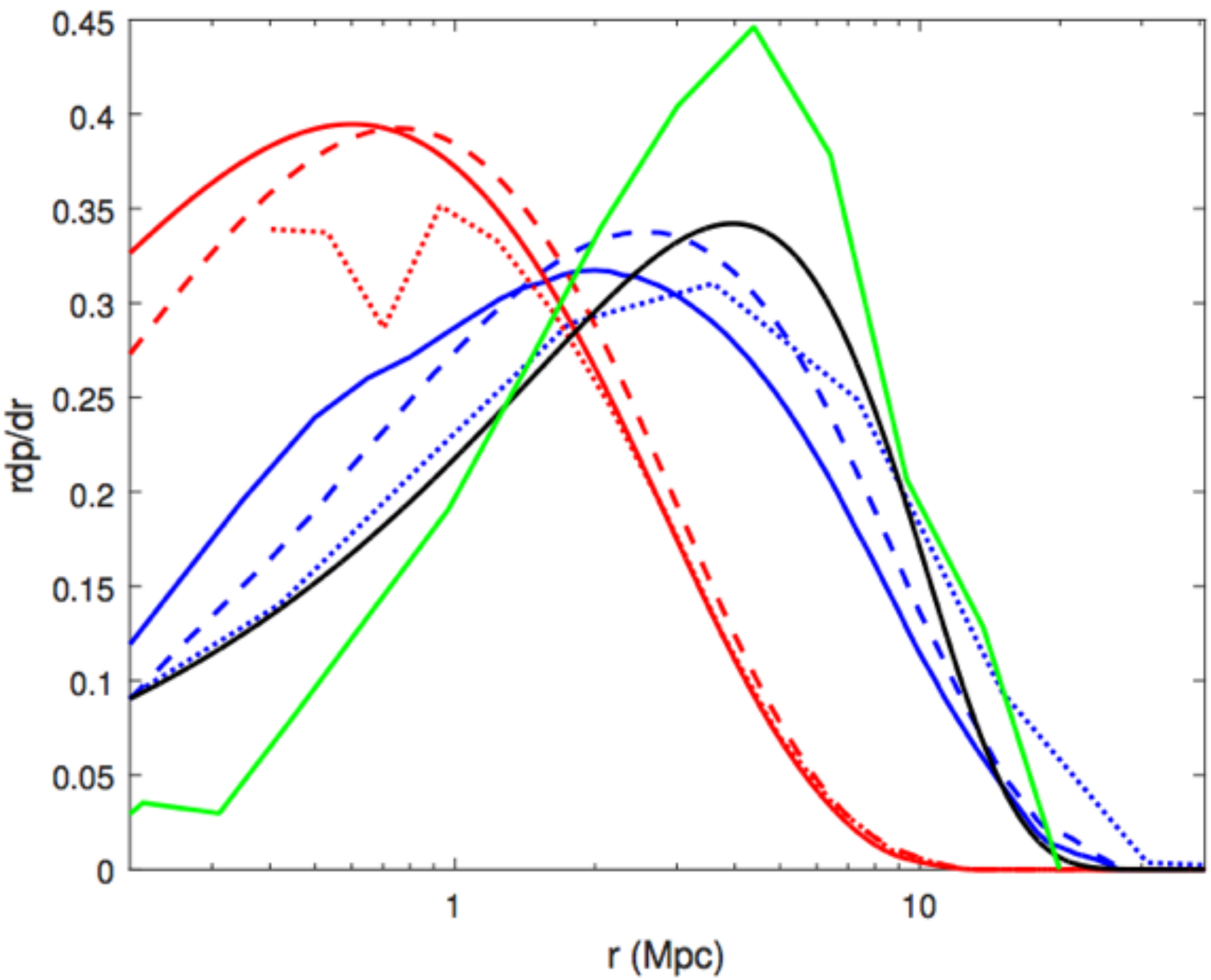
Giri+ 20

•K-mean

unsuperv

Comparison of methods

Yin+ (2016)



See solid line

Black : Reference

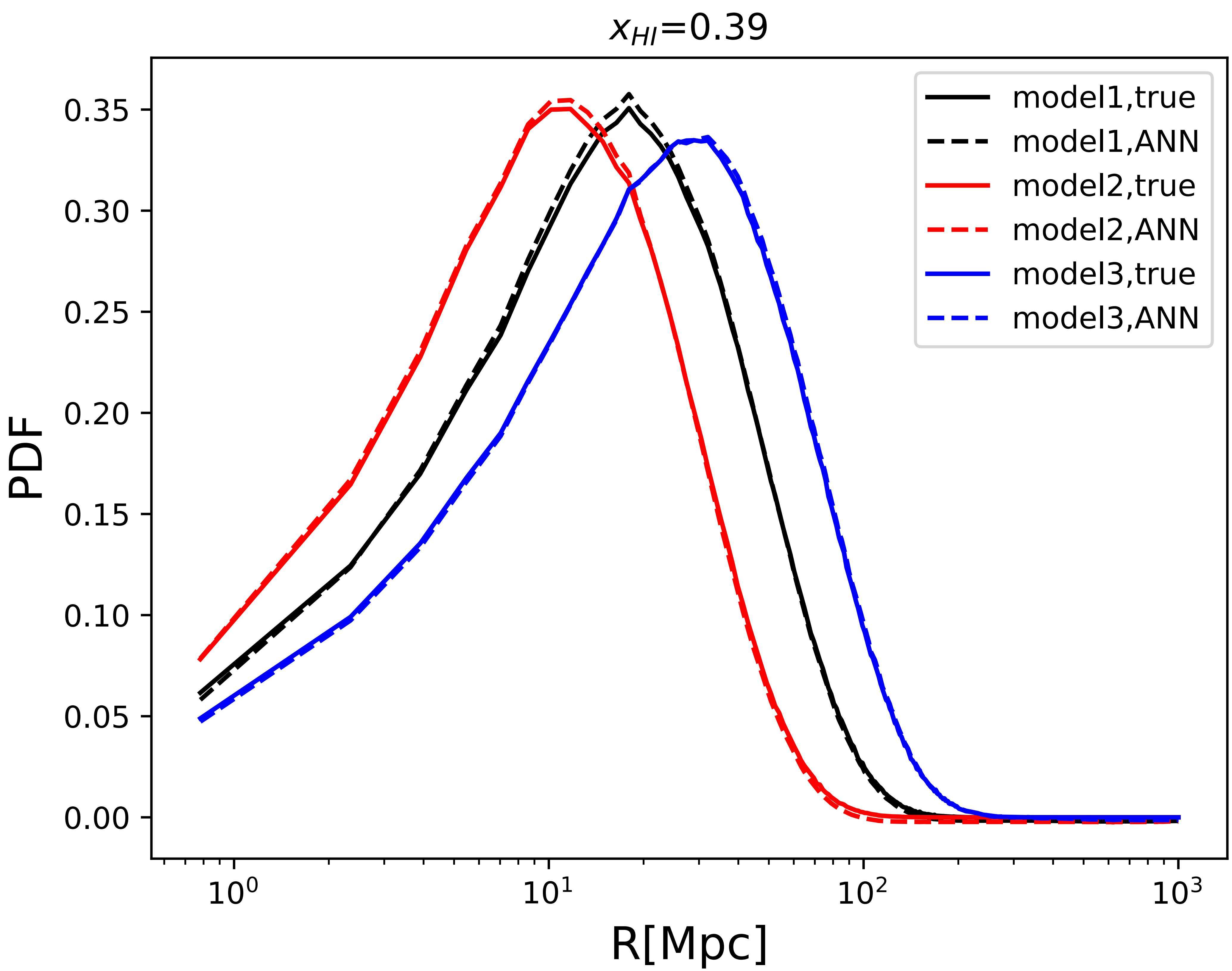
Green : Watershed (WS)

Blue : Mean free path (MFP)

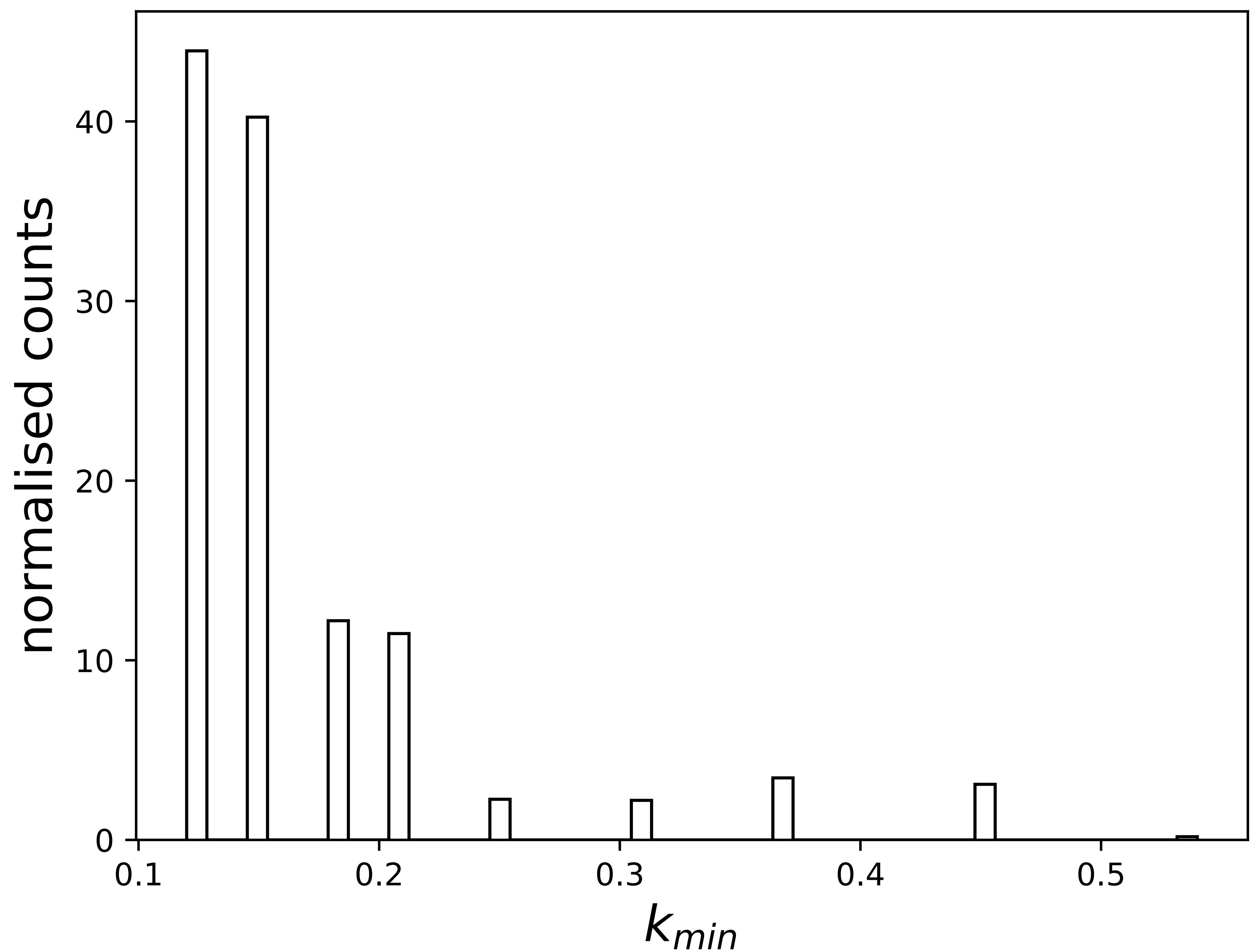
Red : Spherical average (SA)

WS and MFP seem better

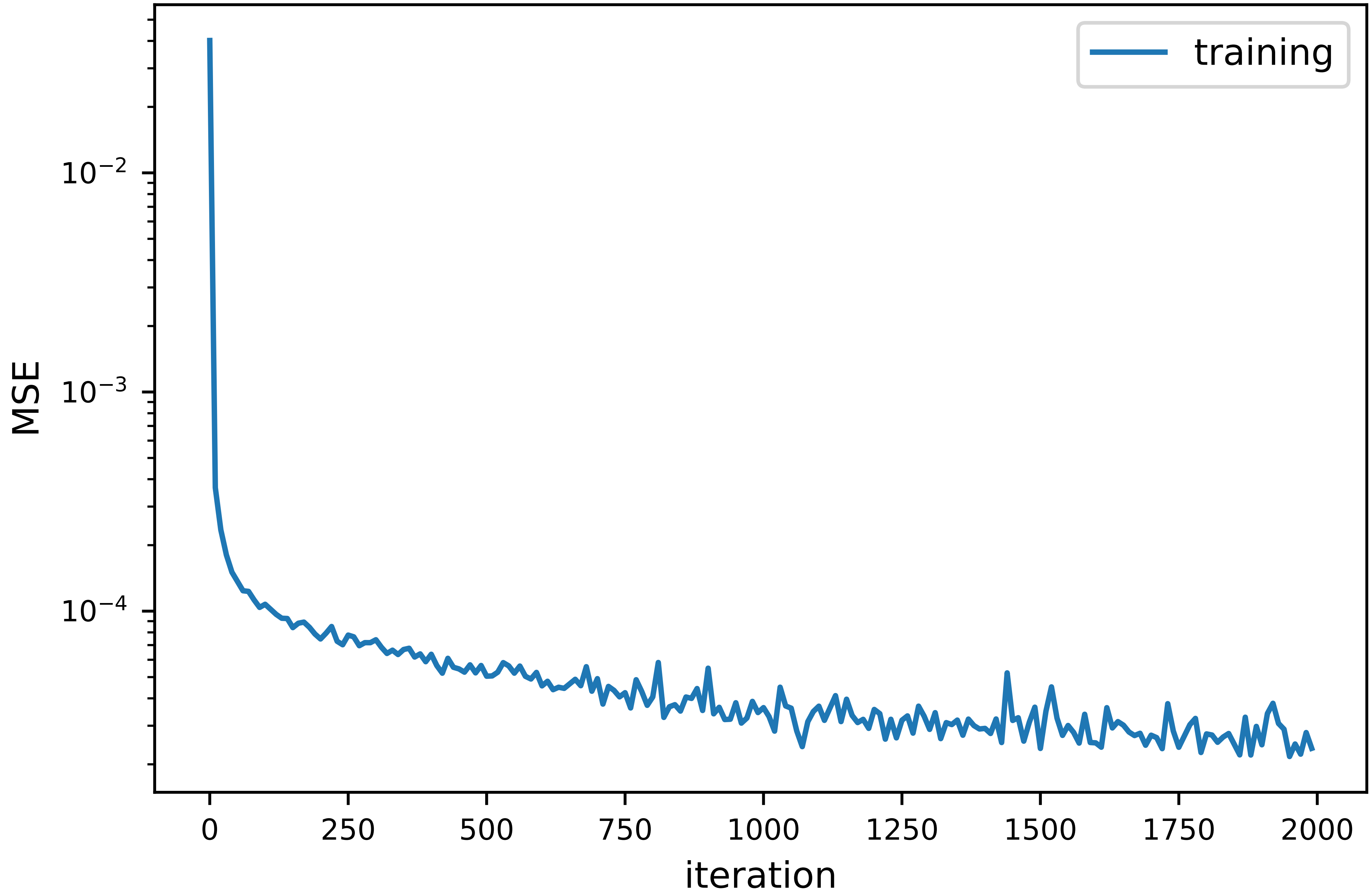
Different models



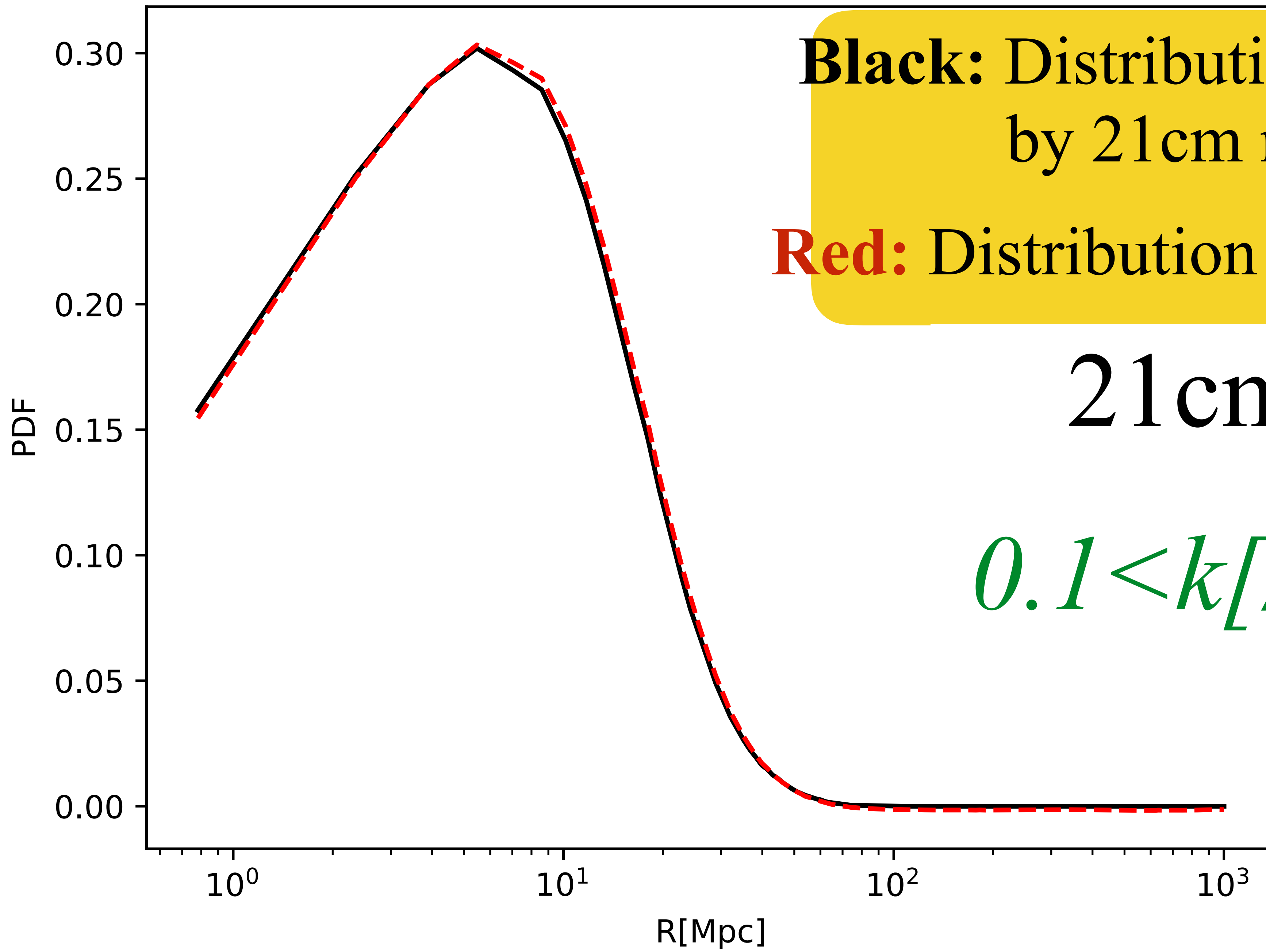
What does k_min give best accuracy?



Training accuracy



HII bubble size distribution



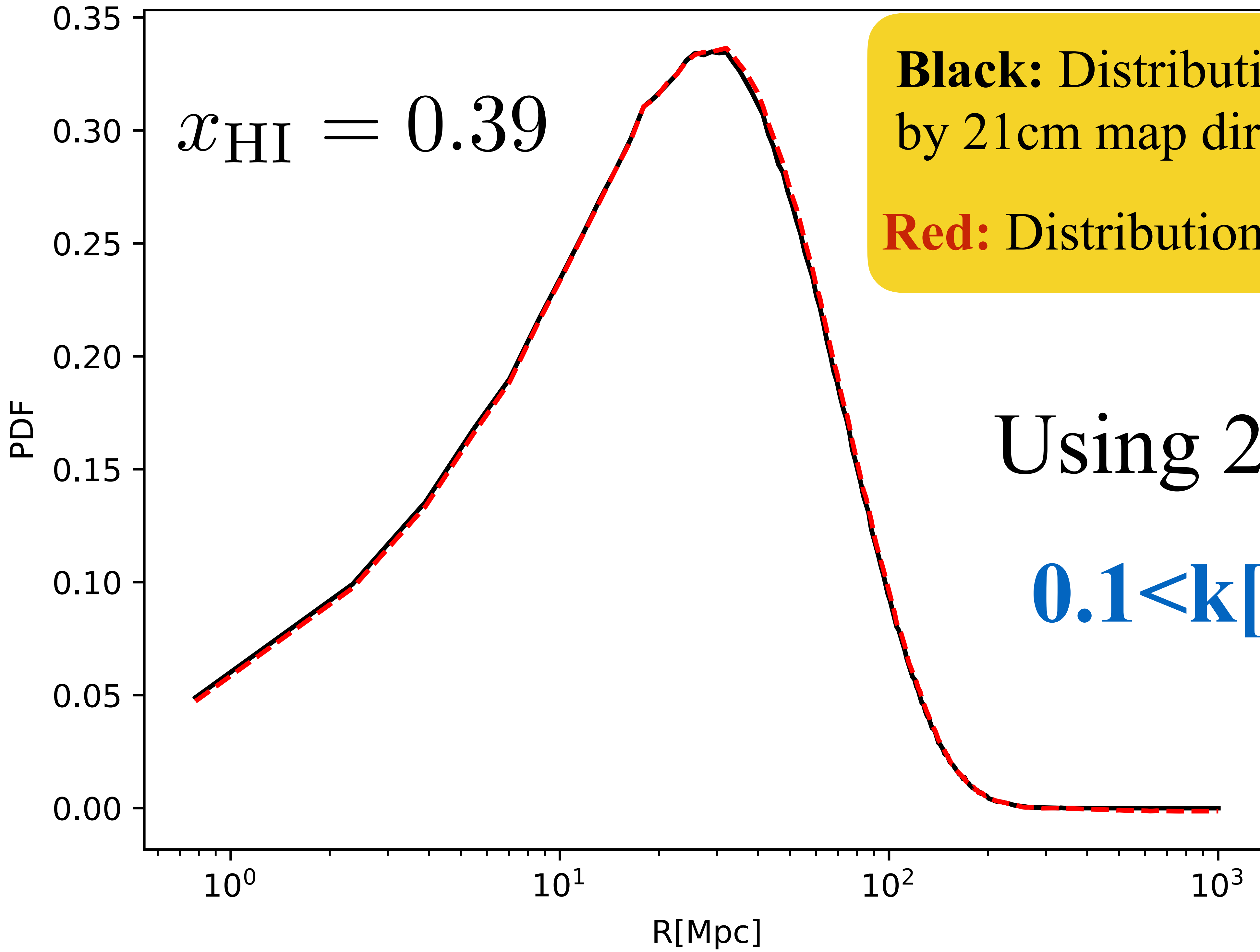
Black: Distribution obtained by 21cm map

Red: Distribution obtained by ANN

21cm PS at

0.1<k[/Mpc]<1

Recovered BSD



Black: Distribution obtained by 21cm map directly

Red: Distribution obtained by ANN

Using 21cm PS at

0.1<k[/Mpc]<1