Parameter estimation from Lyman-α forest in Fourier space using Information Maximising Neural Network

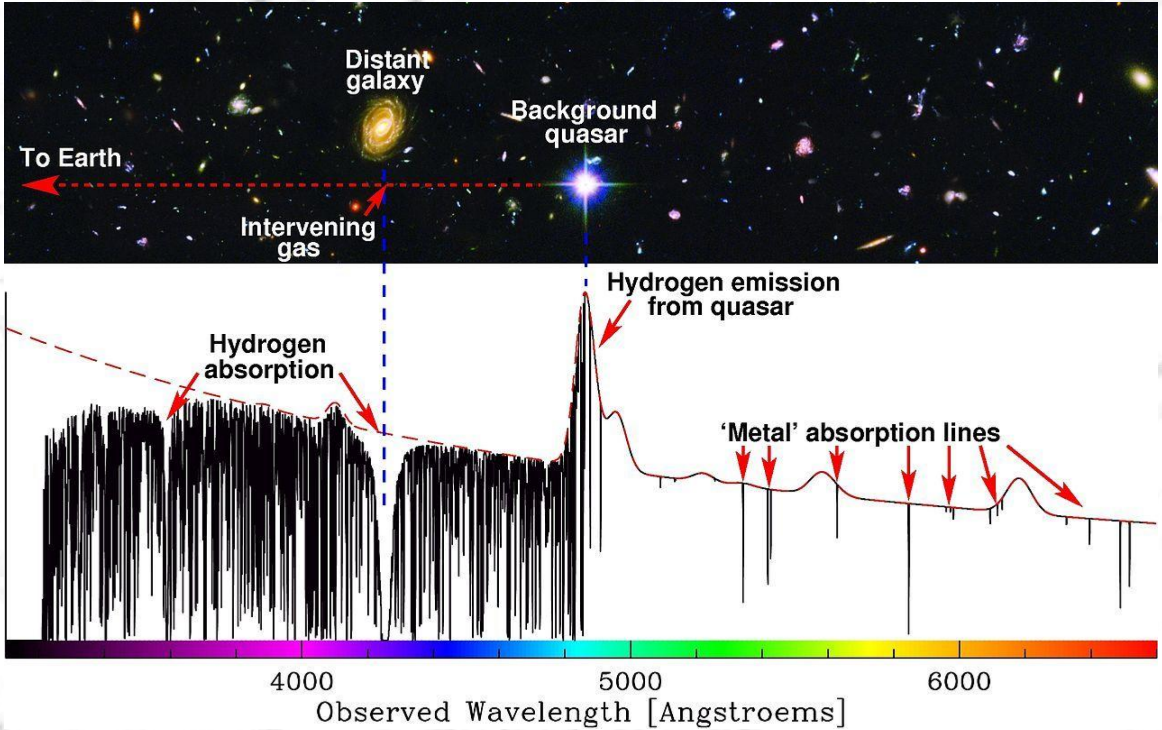
Soumak Maitra
Postdoctoral Researcher (TIFR-Mumbai)

6-10 January 2025

- Stefano Cristiani
- Matteo Viel (SISSA)
- Roberto Trotta (SISSA)
- Guido Cupani

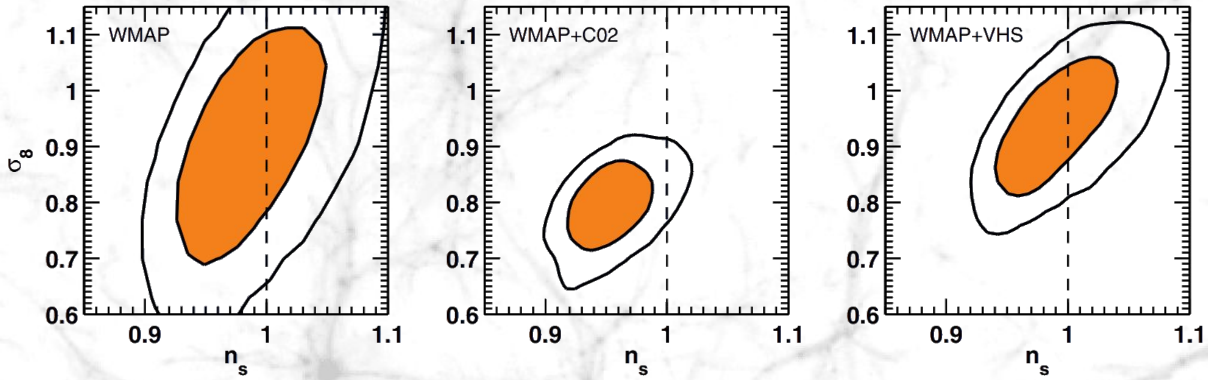
AI/ML Applications in Astronomy & Astrophysics

Lyman-α forest

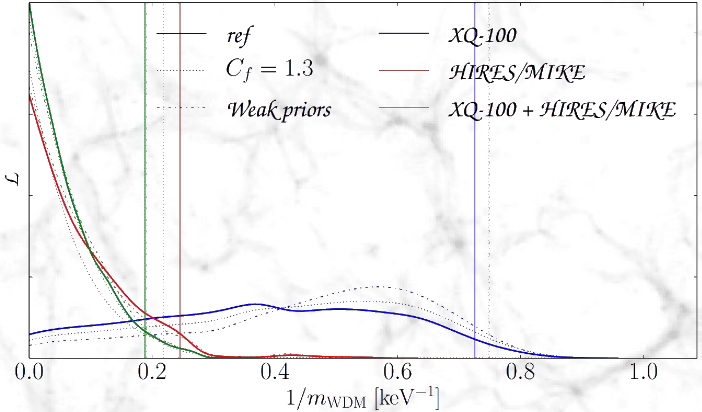


- 1s⟶2p HI Lyman-α absorption lines seen in the spectra of distant QSOs & galaxies.
- Provides a 1D map of the intervening gas and associated gas properties.

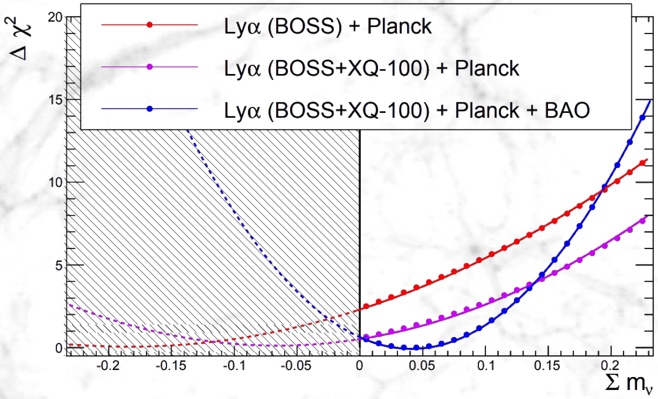
Lyman-α forest: Cosmological Utility



Primordial Power Spectrum (Viel+2004)

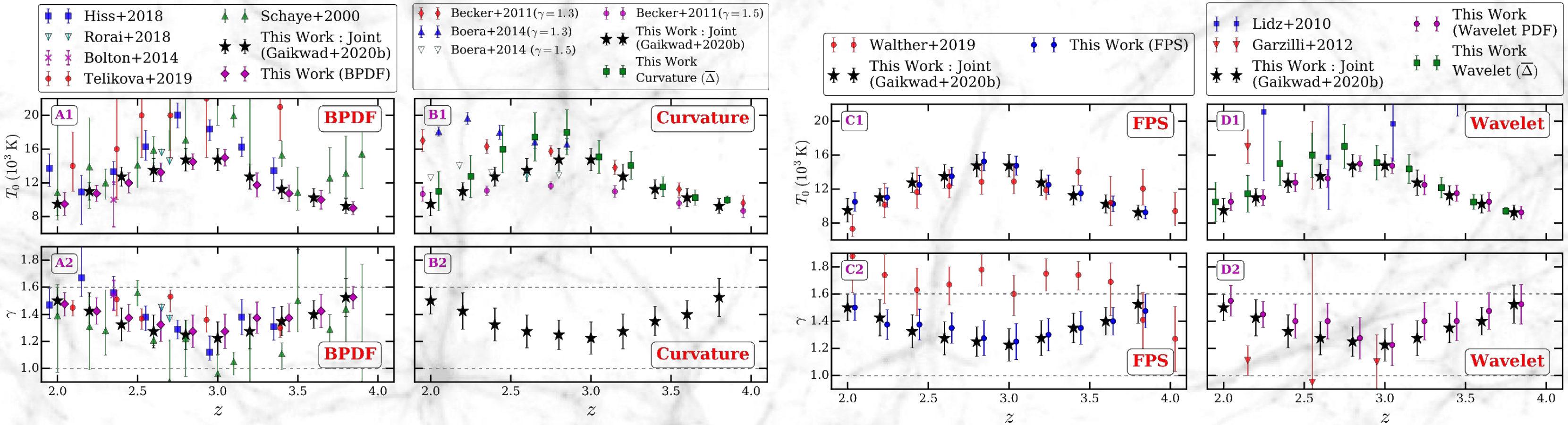


Warm Dark Matter model (Irsic+2017)

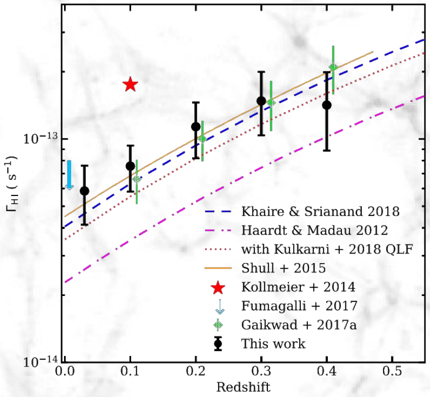


Neutrino Mass (Yeche+2017)

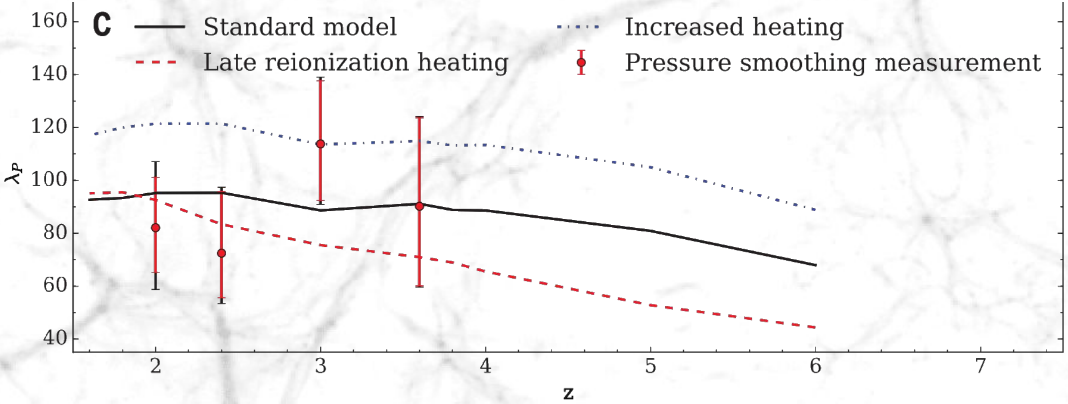
Lyman-α forest: Astrophysical Utility



Thermal Evolution of IGM (Gaikwad+2020)

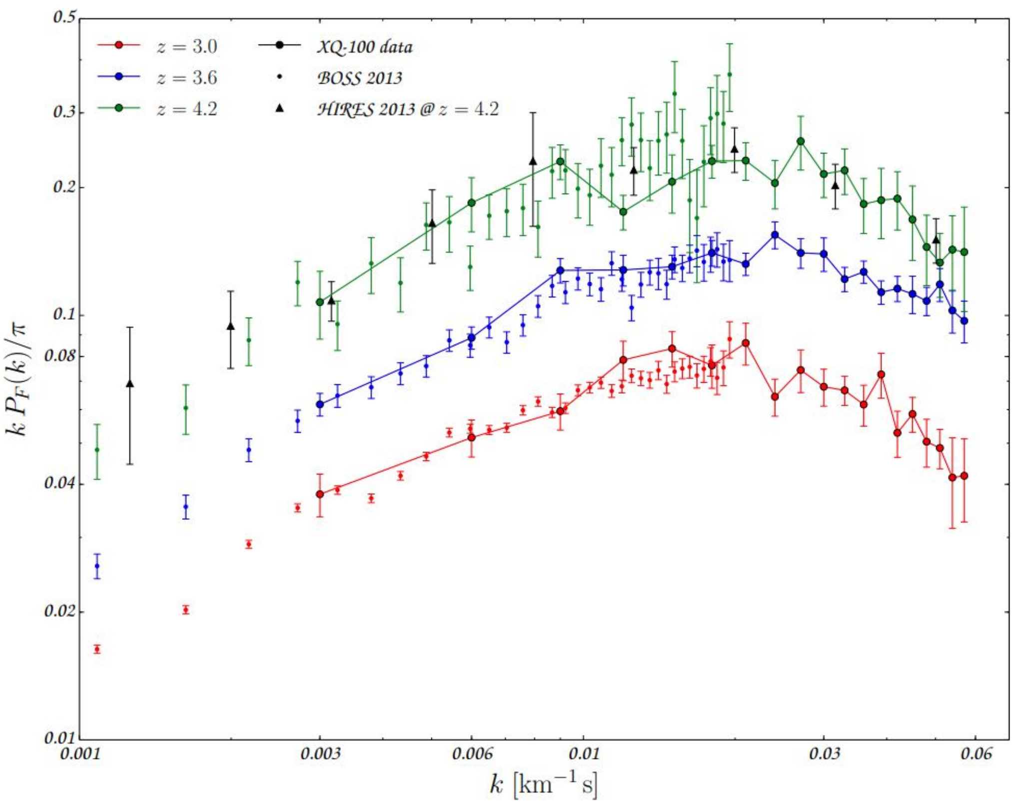


UV Background (Khaire+2019)



Pressure broadening scale (Rorai+2017)

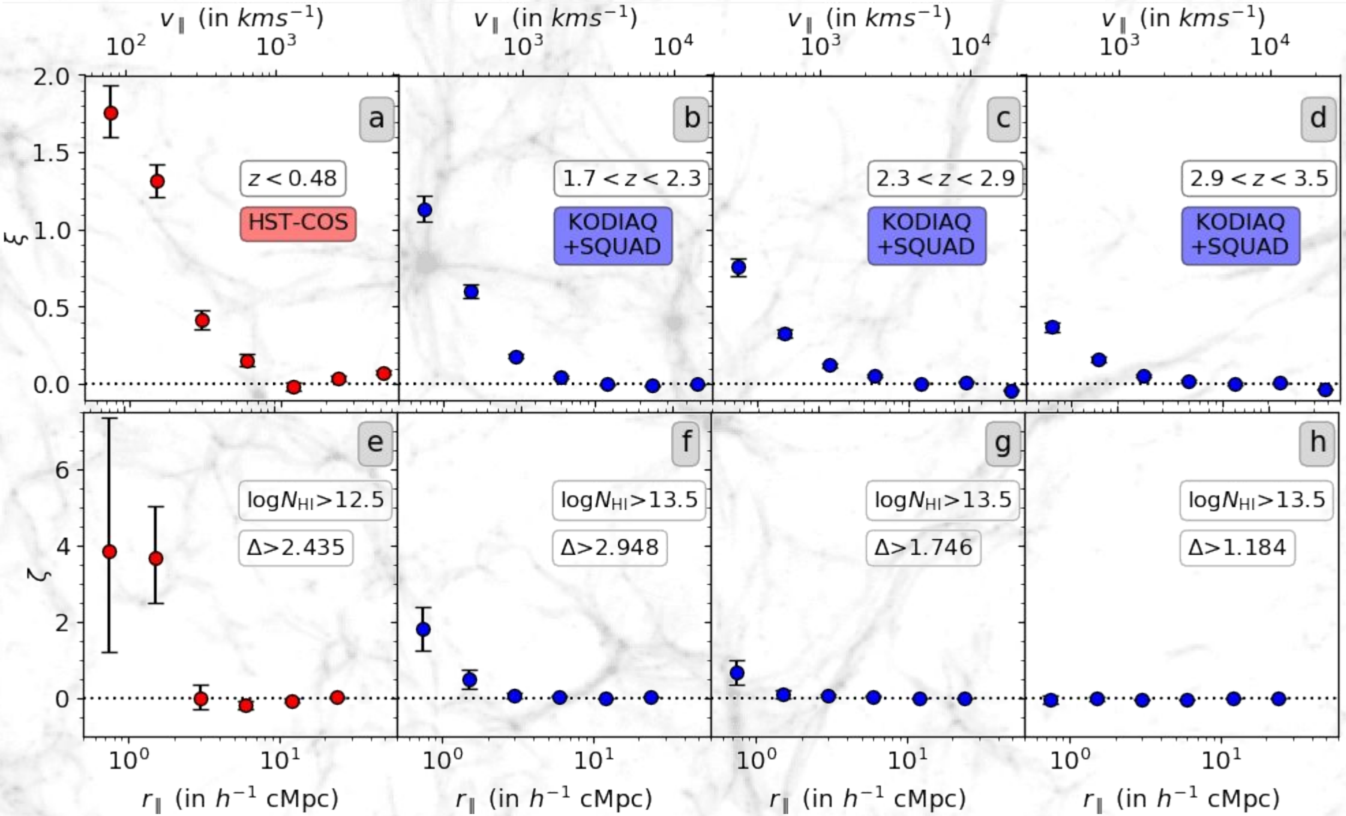
Lyman-α forest: 2-point correlation/power spectrum



Lyman-α forest power spectrum (Irsic+2017)

Lyman-α forest: 3-point correlation

- 82 quasar sightlines from HST-COS (z < 0.48) (Danforth+2016).
- 292 high resolution spectra from KODIAQ (HIRES) and SQUAD (UVES) surveys covering Lyman-α forest in 1.7 < z < 3.5



Maitra+2022a,b

But, are we getting the maximum information from clustering ?

Information Maximizing Neural Network (Charnock+2018)

- Reduce data dimensionality. Entire dataset into a set of summaries equal to number of parameters to be estimated simultaneously. We estimate the parameters individually.

Drawing motivation from the MOPED algorithm, IMNN aims to find some transformation $f : \mathbf{d} \rightarrow \mathbf{x}$ which maps the data ($\mathbf{d}$) to the compressed summary (x_α, for the model parameter α). It transforms the original likelihood into the form

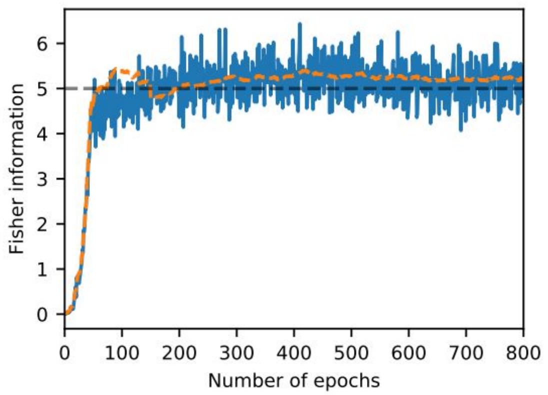
$$-2ln\mathscr{L}(\mathbf{x}|\vartheta) = (\mathbf{x} - \mu_f(\vartheta))^T \mathbf{C}_f^{-1} (\mathbf{x} - \mu_f(\vartheta)) \quad (1)$$

where

$$\mu_f(\vartheta) = \frac{1}{n_s} \sum_{i=1}^{n_s} \mathbf{x}_i^s \quad (2)$$

is the mean of n_s summaries. ϑ is the set of model parameters and C_f is the covariance matrix. The modified fisher information matrix can be expressed as

$$\mathbf{F}_{\alpha\beta} = Tr[\mu_{f,\alpha}^T \mathbf{C}_f^{-1} \mu_{f,\beta}] \quad (3)$$



$$\vartheta_\alpha = \vartheta_\alpha^{\text{fid}} + \mathbf{F}_{\alpha\beta}^{-1} \mu_{f,\beta}^T \mathbf{C}^{-1} (x - \mu_f).$$

Information Maximizing Neural Network (Charnock+2018)

- Reduce data dimensionality. Entire dataset into a set of summaries equal to number of parameters to be estimated simultaneously. We estimate the parameters individually.

Drawing motivation from the MOPED algorithm, IMNN aims to find some transformation $\boxed{f : \mathbf{d} \rightarrow \mathbf{x}}$ which maps the data ($\mathbf{d}$) to the compressed summary (x_α, for the model parameter α). It transforms the original likelihood into the form

$$-2 ln \mathscr{L}(\mathbf{x}|\vartheta) = (\mathbf{x} - \mu_f(\vartheta))^T \mathbf{C}_f^{-1} (\mathbf{x} - \mu_f(\vartheta)) \quad (1)$$

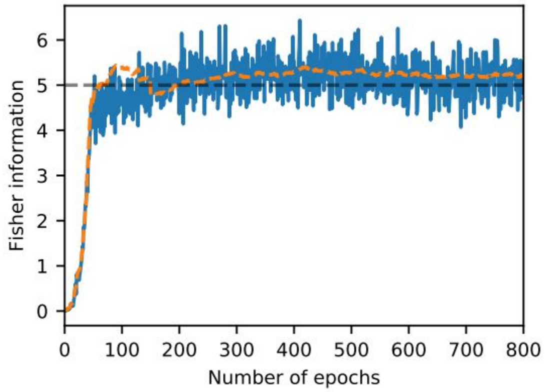
where

$$\mu_f(\vartheta) = \frac{1}{n_s} \sum_{i=1}^{n_s} \mathbf{x}_i^s \quad (2)$$

is the mean of n_s summaries. ϑ is the set of model parameters and C_f is the covariance matrix. The modified fisher information matrix can be expressed as

$$\boxed{\mathbf{F}_{\alpha\beta} = Tr[\mu_{f,\alpha}^T \mathbf{C}_f^{-1} \mu_{f,\beta}]} \quad (3)$$

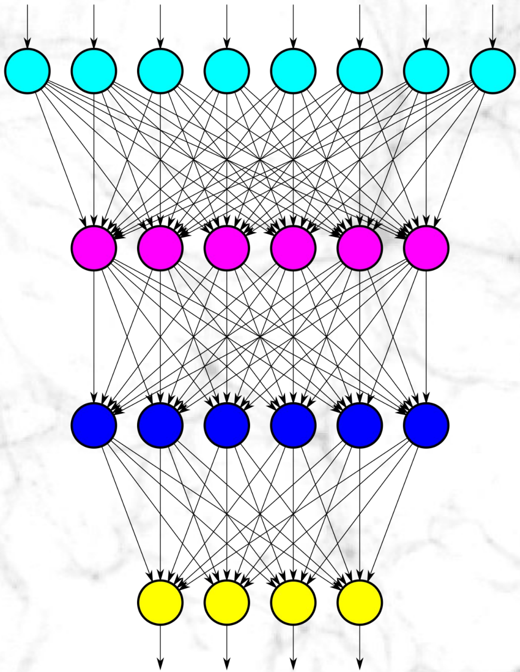
To be maximised by Neural network



$$\vartheta_\alpha = \vartheta_\alpha^{\text{fid}} + \mathbf{F}_{\alpha\beta}^{-1} \mu_{f,\beta}^T \mathbf{C}^{-1} (x - \mu_f).$$

SHERWOOD-Relics Simulations used for training

Input : Fourier transformed
Lyman-alpha Transmitted flux



Output: Summary

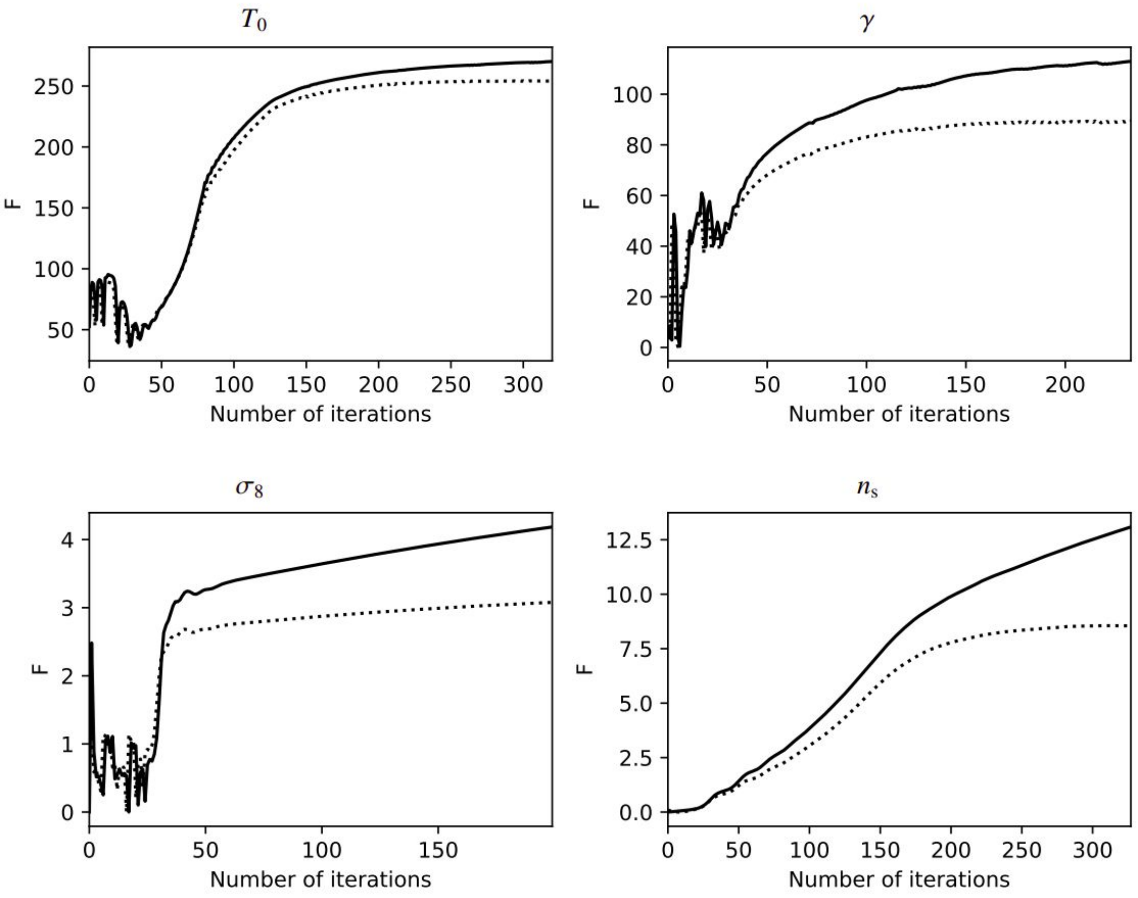
($40\ h^{-1}$cMpc)^3 with $2X(1024)^3$ particles (Puchwein+2023)

Table 1. List of Sherwood-Relics simulations used

Simulation Model	Purpose	Seed	Temperature	γ	σ_8	n_s	Γ_{HI}
40-1024 (Fiducial)	Training	181170	Fiducial	1.3	0.829	0.961	Puchwein et al. (2019)
40-1024-cold	Training	181170	Fiducial/1.5	1.3	0.829	0.961	Puchwein et al. (2019)
40-1024-hot	Training	181170	Fiducial*1.5	1.3	0.829	0.961	Puchwein et al. (2019)
40-1024-g10	Training	181170	Fiducial	1.0	0.829	0.961	Puchwein et al. (2019)
40-1024-g16	Training	181170	Fiducial	1.6	0.829	0.961	Puchwein et al. (2019)
40-1024-s754	Training	181170	Fiducial	1.3	0.754	0.961	Puchwein et al. (2019)
40-1024-s804	Training	181170	Fiducial	1.3	0.804	0.961	Puchwein et al. (2019)
40-1024-s854	Training	181170	Fiducial	1.3	0.854	0.961	Puchwein et al. (2019)
40-1024-s904	Training	181170	Fiducial	1.3	0.904	0.961	Puchwein et al. (2019)
40-1024-n921	Training	181170	Fiducial	1.3	0.829	0.921	Puchwein et al. (2019)
40-1024-n941	Training	181170	Fiducial	1.3	0.829	0.941	Puchwein et al. (2019)
40-1024-n981	Training	181170	Fiducial	1.3	0.829	0.981	Puchwein et al. (2019)
40-1024-n1001	Training	181170	Fiducial	1.3	0.829	1.001	Puchwein et al. (2019)
40-1024 (lowered Γ_{HI})	Training	181170	Fiducial	1.3	0.829	0.961	Puchwein et al. (2019)/1.25
40-1024 (elevated Γ_{HI})	Training	181170	Fiducial	1.3	0.829	0.961	Puchwein et al. (2019)*1.25
40-1024-seed001	Testing	965431	Fiducial	1.3	0.829	0.961	Puchwein et al. (2019)
40-1024-seed002	Testing	126642	Fiducial	1.3	0.829	0.961	Puchwein et al. (2019)
40-1024-seed003	Testing	140516	Fiducial	1.3	0.829	0.961	Puchwein et al. (2019)

1D parameter estimation of T_0, γ, σ_8 and n_S

Training



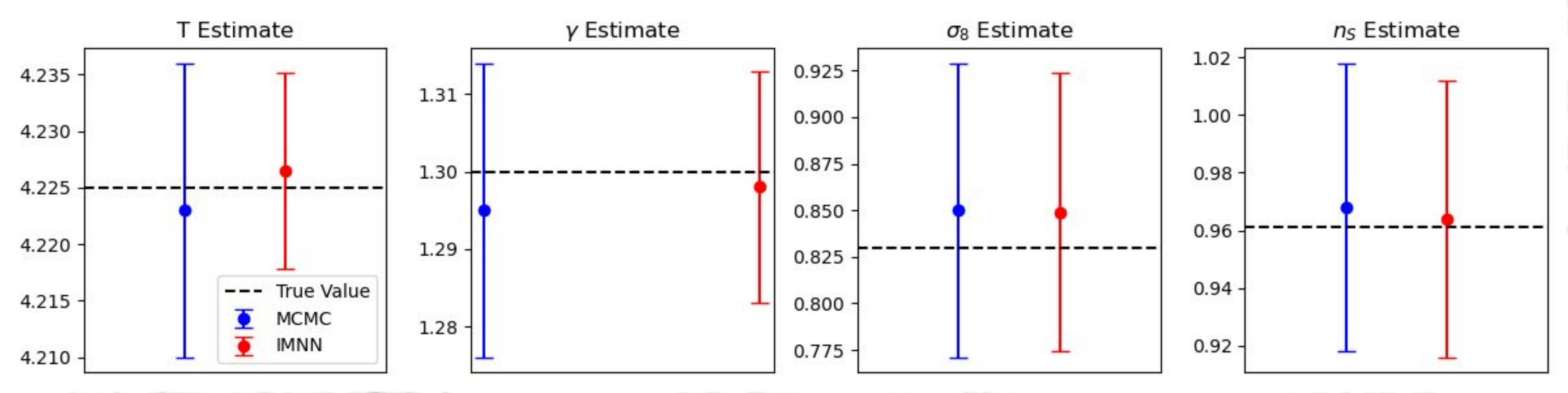
Solid lines: Training

Dashed lines: Validation/Testing

IMNN estimates vs MCMC estimates from Power Spectrum

IMNN estimates vs MCMC estimates from Power Spectrum

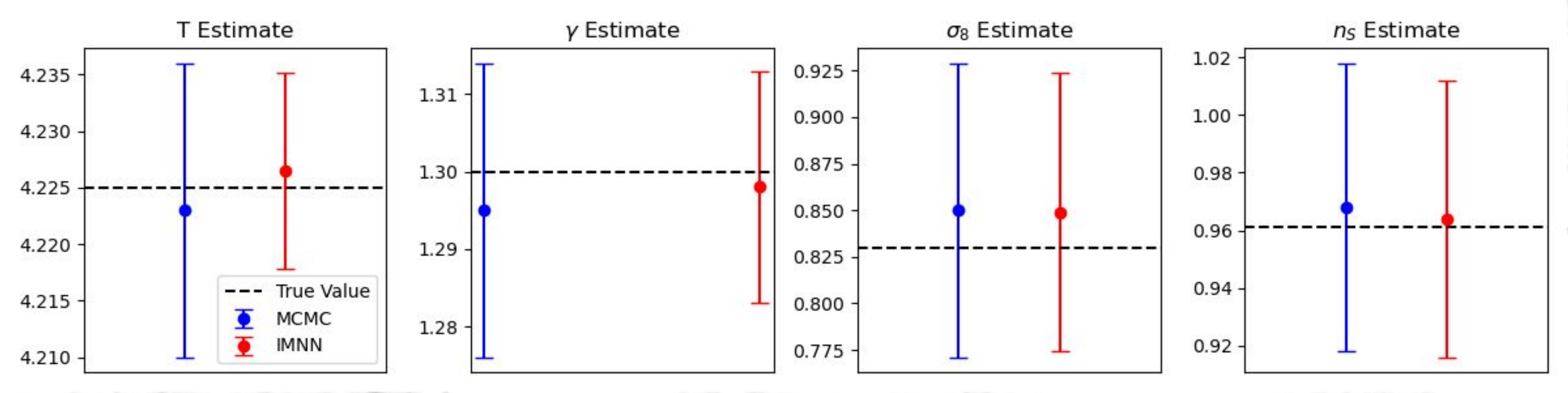
z=3



Maitra+2024

IMNN estimates vs MCMC estimates from Power Spectrum

z=3

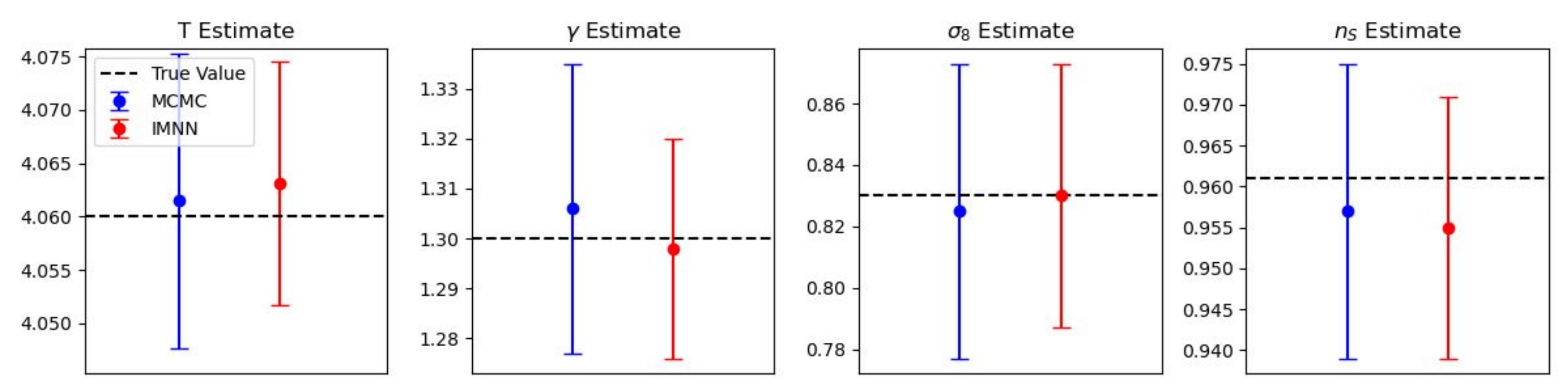


Estimated error ratios (MCMC/IMNN)	1.52	1.27	1.05	1.04

Maitra+2024

IMNN estimates vs MCMC estimates from Power Spectrum

z=4

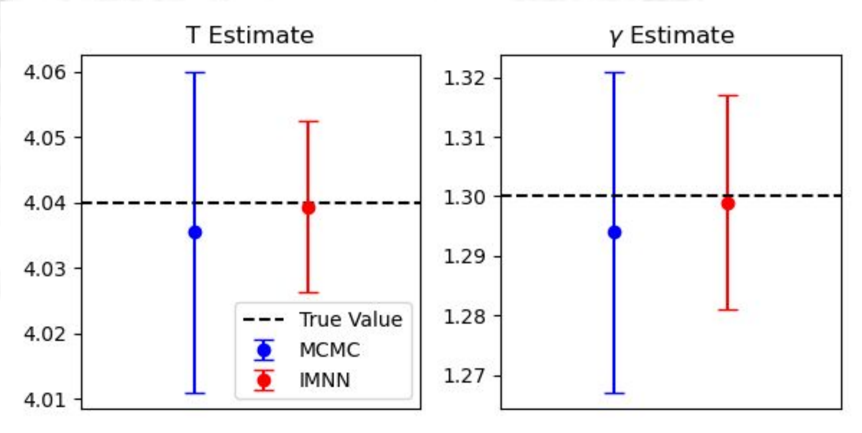


Estimated error ratios (MCMC/IMNN)	1.21	1.32	1.12	1.13

Maitra+2024

IMNN estimates vs MCMC estimates from Power Spectrum

z=2

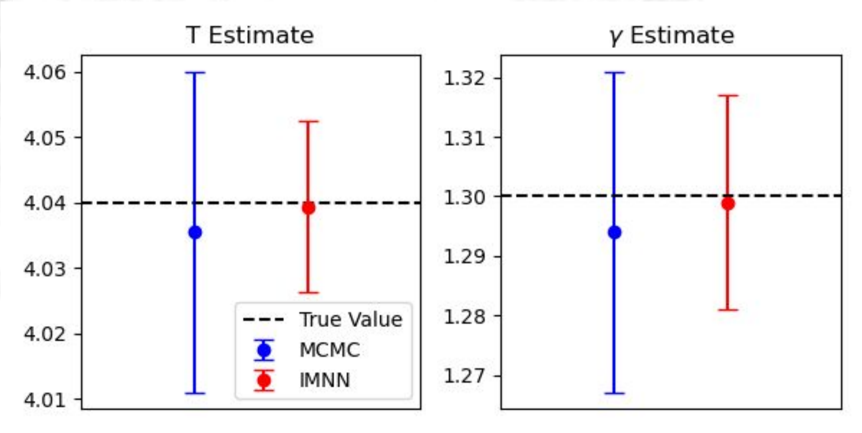


Estimated error ratios (MCMC/IMNN) 1.89 1.50

Maitra+2024

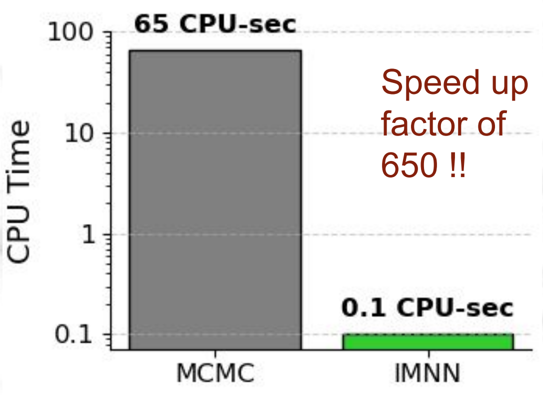
IMNN estimates vs MCMC estimates from Power Spectrum

z=2



Estimated error ratios (MCMC/IMNN) 1.89 1.50

Major Pro

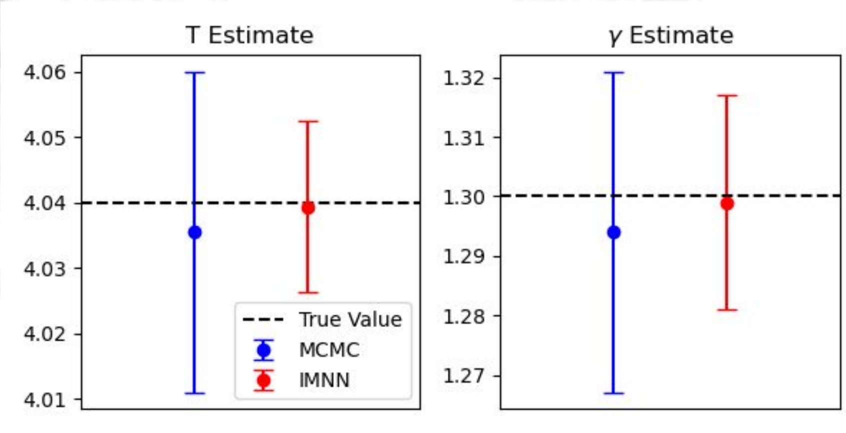


Speed up factor of 650 !!

Maitra+2024

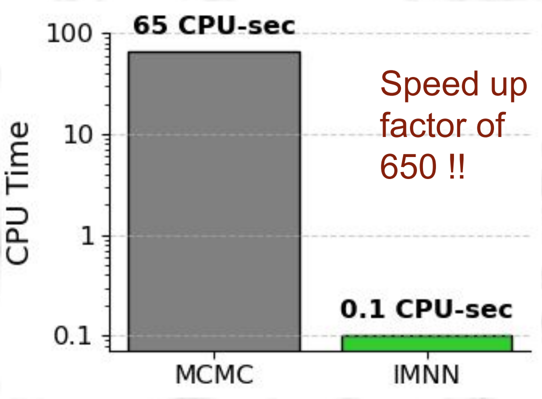
IMNN estimates vs MCMC estimates from Power Spectrum

z=2



Estimated error ratios (MCMC/IMNN) 1.89 1.50

Major Pro



Major Con

You might miss an episode of your favourite TV show….

Maitra+2024

Instrumental smoothing effects

Table 3. Estimation of parameters for different instrumental smoothing in both training and testing set (gaussian; z = 3.0).

Parameter	**Input Parameter**	**Output parameter (Error corresponding to 30 sightlines)**			
		No smoothing	FWHM=6 $\mathrm{km s^{-1}}$	FWHM=50 $\mathrm{km s^{-1}}$	FWHM=150 $\mathrm{km s^{-1}}$
$\log(T_0)$ (IMNN) (4.225 ± 0.176)	4.225	4.2265 ± 0.0087	4.2237 ± 0.0088	4.2230 ± 0.0360	4.3322 ± 0.4573
$\log(T_0)$ (MCMC) (4.225 ± 0.176)	4.225	4.2228 ± 0.0131	4.2226 ± 0.0135	4.2277 ± 0.0325	4.698 ± 0.2260
γ (IMNN) (1.3 ± 0.3)	1.3	1.298 ± 0.015	1.297 ± 0.015	1.297 ± 0.040	1.268 ± 0.168
γ (MCMC) (1.3 ± 0.3)	1.3	1.295 ± 0.019	1.297 ± 0.020	1.301 ± 0.041	1.279 ± 0.162

Instrumental smoothing effects

Table 3. Estimation of parameters for different instrumental smoothing in both training and testing set (gaussian; z = 3.0).

Parameter	**Input Parameter**	**Output parameter (Error corresponding to 30 sightlines)**			
		No smoothing	FWHM=6 $\mathrm{kms^{-1}}$	FWHM=50 $\mathrm{kms^{-1}}$	FWHM=150 $\mathrm{kms^{-1}}$
$\log(T_0)$ (IMNN) (4.225 ± 0.176)	4.225	4.2265 ± 0.0087	4.2237 ± 0.0088	4.2230 ± 0.0360	4.3322 ± 0.4573
$\log(T_0)$ (MCMC) (4.225 ± 0.176)	4.225	4.2228 ± 0.0131	4.2226 ± 0.0135	4.2277 ± 0.0325	4.698 ± 0.2260
γ (IMNN) (1.3 ± 0.3)	1.3	1.298 ± 0.015	1.297 ± 0.015	1.297 ± 0.040	1.268 ± 0.168
γ (MCMC) (1.3 ± 0.3)	1.3	1.295 ± 0.019	1.297 ± 0.020	1.301 ± 0.041	1.279 ± 0.162

Instrumental smoothing effects

Table 3. Estimation of parameters for different instrumental smoothing in both training and testing set (gaussian; $z = 3.0$).

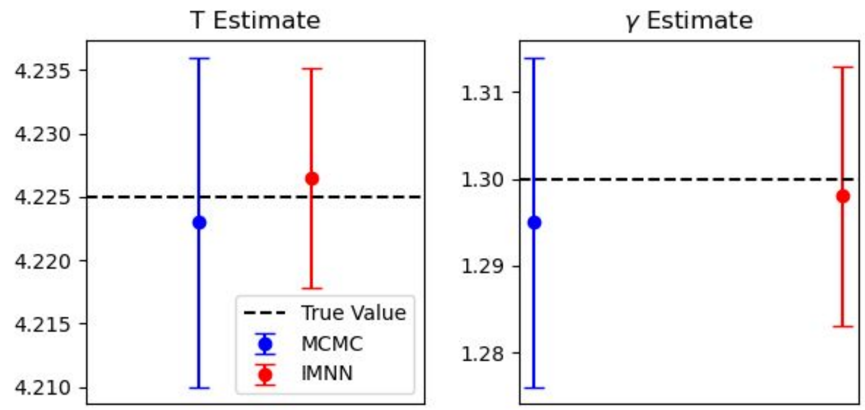
Parameter	Input Parameter	Output parameter (Error corresponding to 30 sightlines)			
		No smoothing	FWHM=6 kms^{-1}	FWHM=50 kms^{-1}	FWHM=150 kms^{-1}
$\log(T_0)$ (IMNN) (4.225 ± 0.176)	4.225	4.2265 ± 0.0087	4.2237 ± 0.0088	4.2230 ± 0.0360	4.3322 ± 0.4573
$\log(T_0)$ (MCMC) (4.225 ± 0.176)	4.225	4.2228 ± 0.0131	4.2226 ± 0.0135	4.2277 ± 0.0325	4.698 ± 0.2260
γ (IMNN) (1.3 ± 0.3)	1.3	1.298 ± 0.015	1.297 ± 0.015	1.297 ± 0.040	1.268 ± 0.168
γ (MCMC) (1.3 ± 0.3)	1.3	1.295 ± 0.019	1.297 ± 0.020	1.301 ± 0.041	1.279 ± 0.162

Improvements in IMNN from small-scale information!!

Continuum effect

z=3

Continuum ($\mu = 1.0, \sigma = 0$)



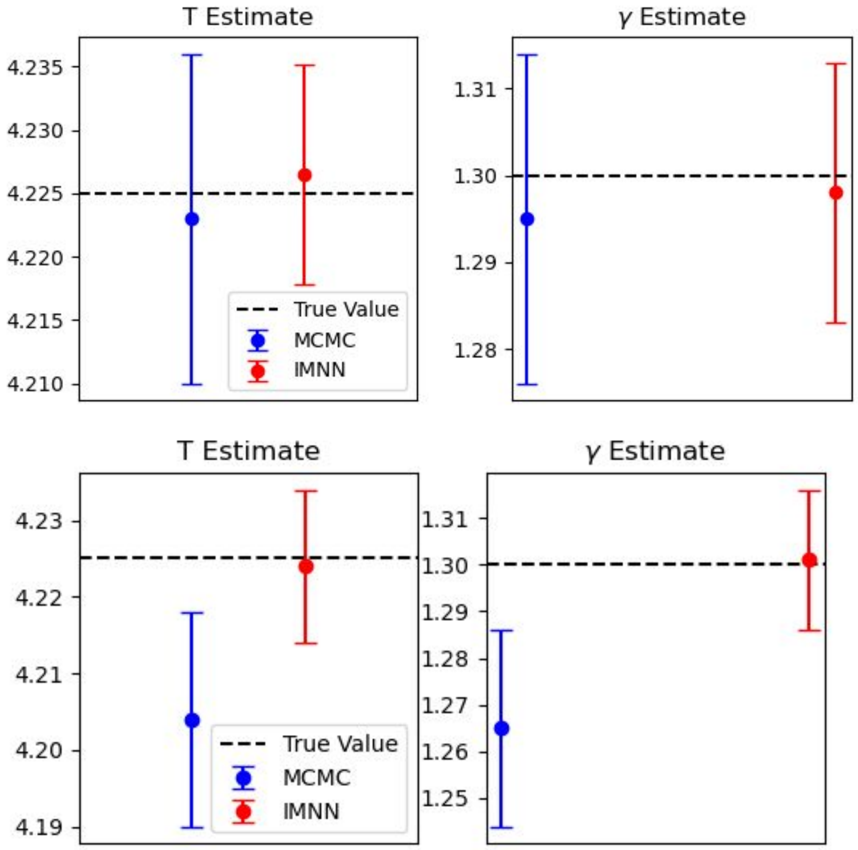
Maitra+2024

Continuum effect

z=3

Continuum ($\mu = 1.0, \sigma = 0$)

Continuum ($\mu = 0.9, \sigma = 0.1$)

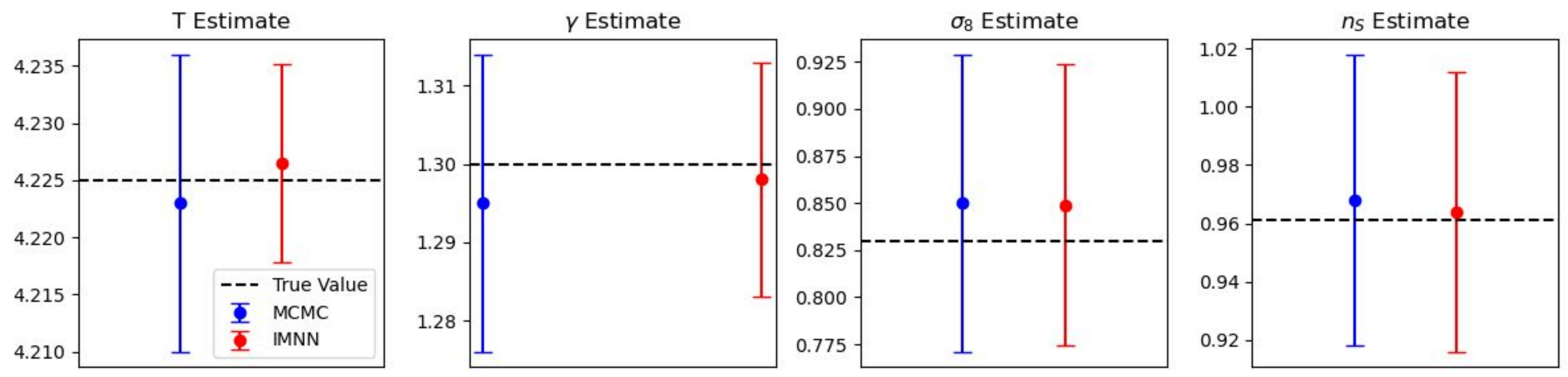


Maitra+2024

IMNN estimates vs MCMC estimates (SNR deviation by a factor 0.85)

z=3

Original SNR Distribution

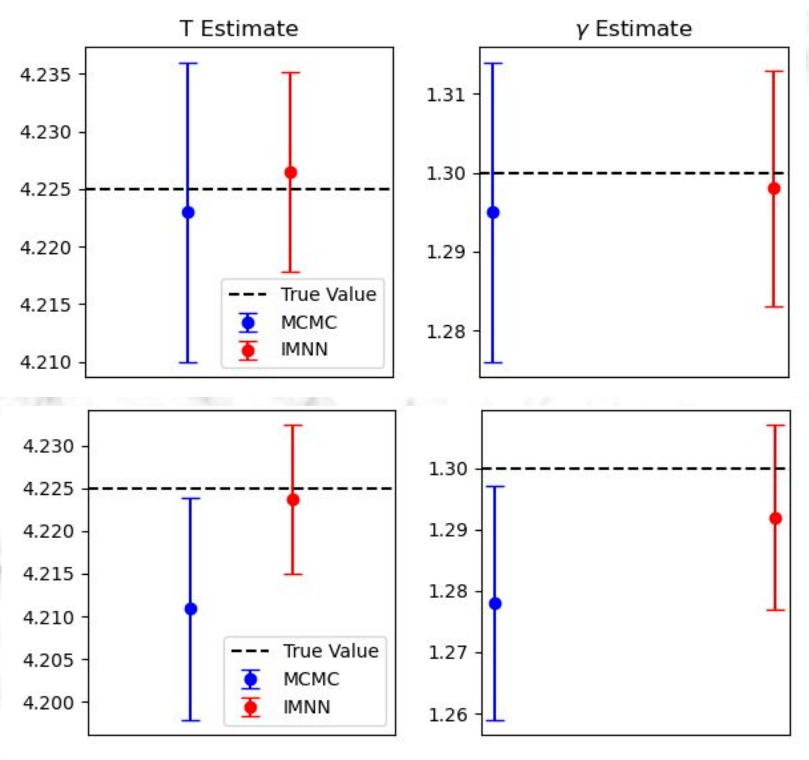


IMNN estimates vs MCMC estimates (SNR deviation by a factor 0.85)

z=3

Original SNR Distribution

0.85 X Original SNR Distribution

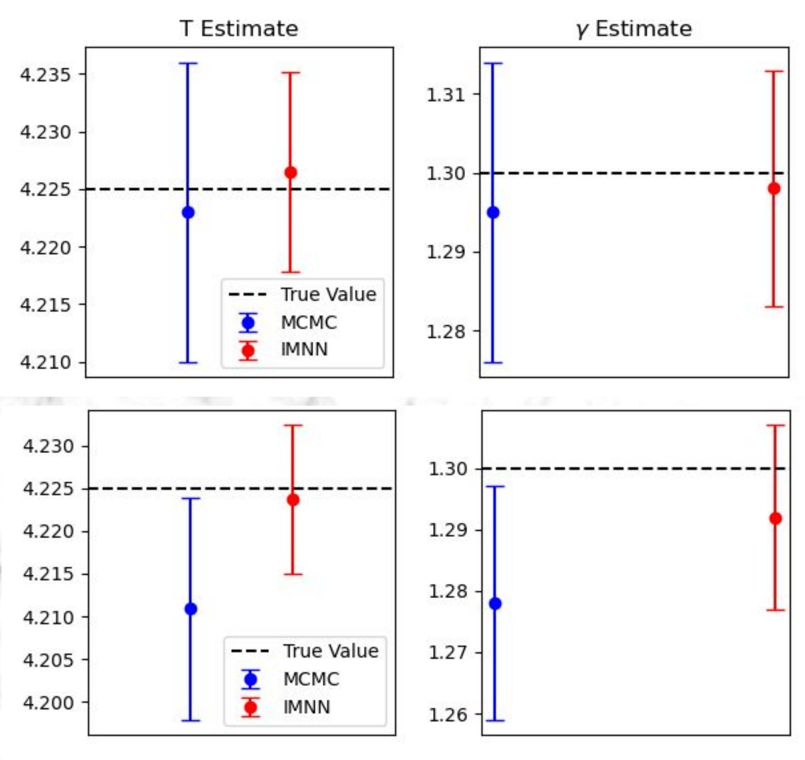


IMNN estimates vs MCMC estimates (SNR deviation by a factor 0.85)

z=3

Original SNR Distribution

0.85 X Original SNR Distribution



SNR deviation ratios (MCMC/IMNN)

4.17

2.86

Robustness check with CAMELS simulations (Villaescusa-Navarro+2021)

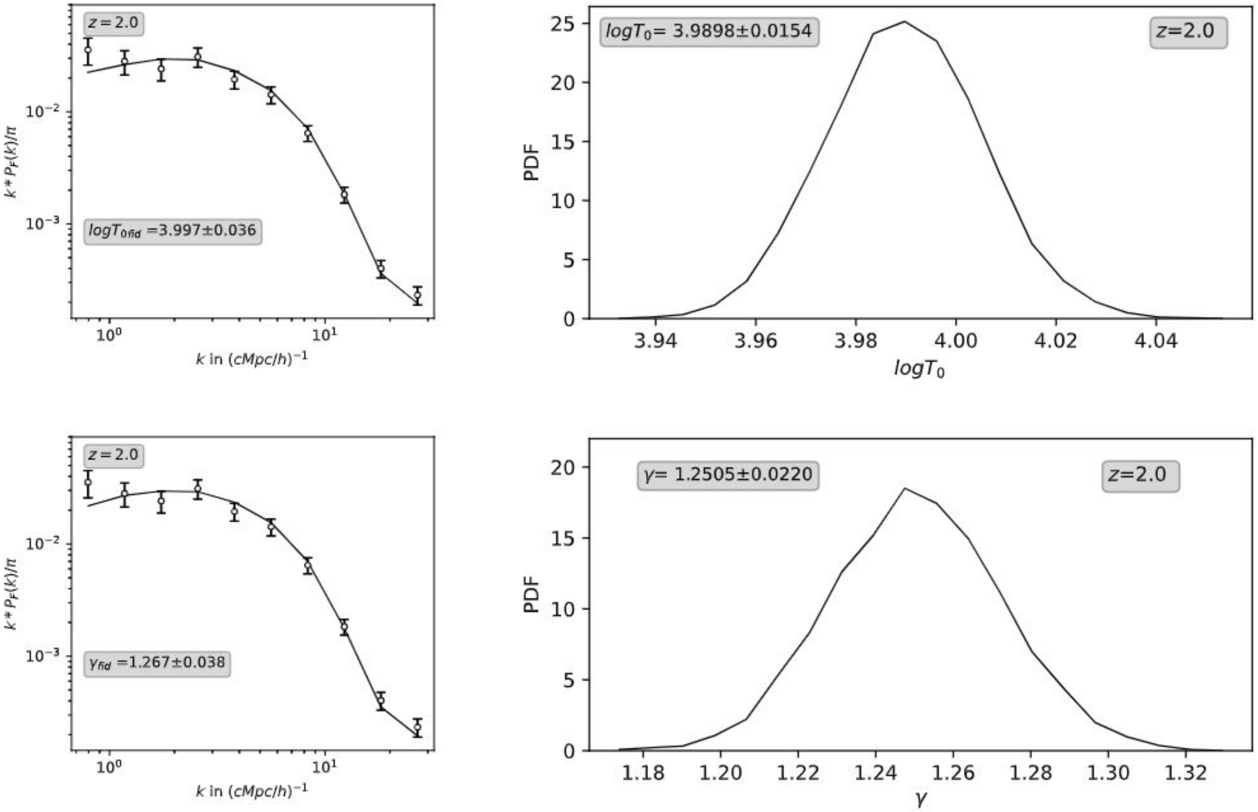
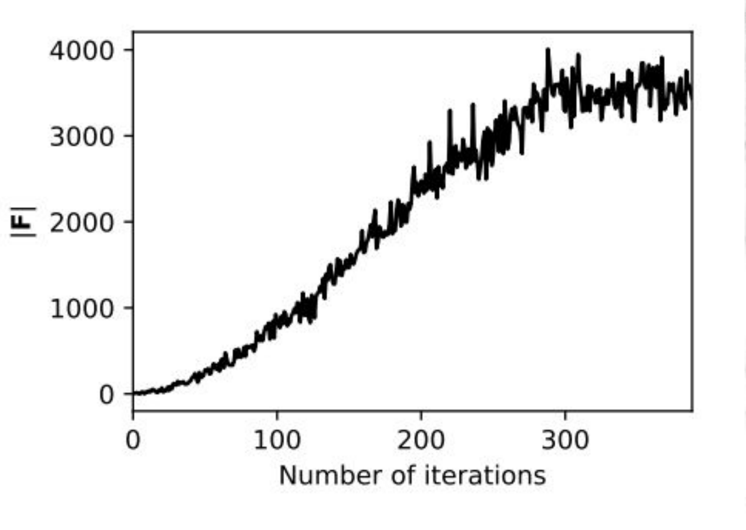
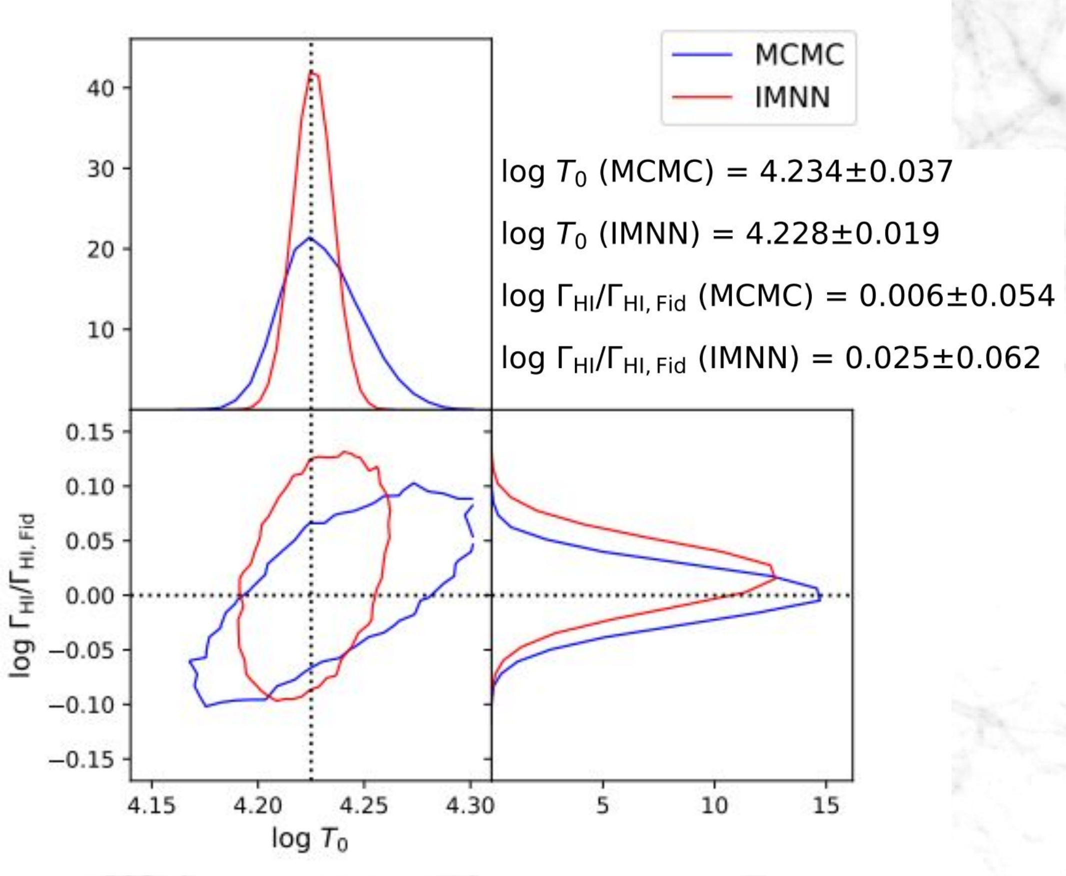


Figure 6. Estimation of T_0 and γ using MLE with power spectrum and IMNN on the Fourier transformed Lyα forest transmitted flux for CAMELS simulation at $z = 2$. The expected parameter values are $\log T_0 = 3.97$ and $\gamma = 1.26$. The power spectrum modeling for the MLE approach and the training for the IMNN approach were done with previous Sherwood-Relics simulations. We check the robustness of both the approaches by testing it on a different simulation.

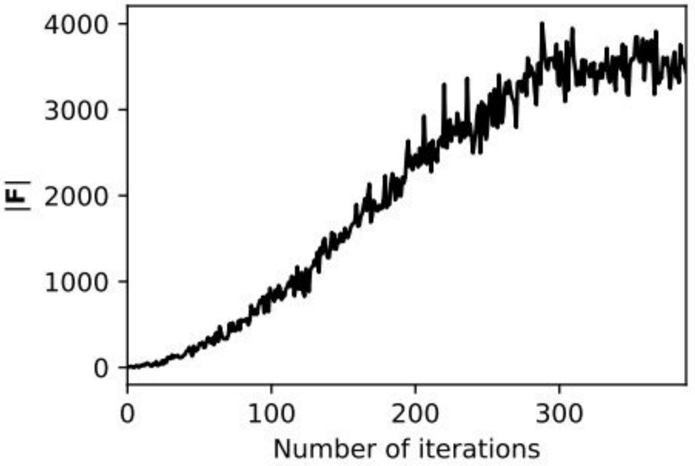
2D parameter estimation (with T_0 and Γ_{HI})

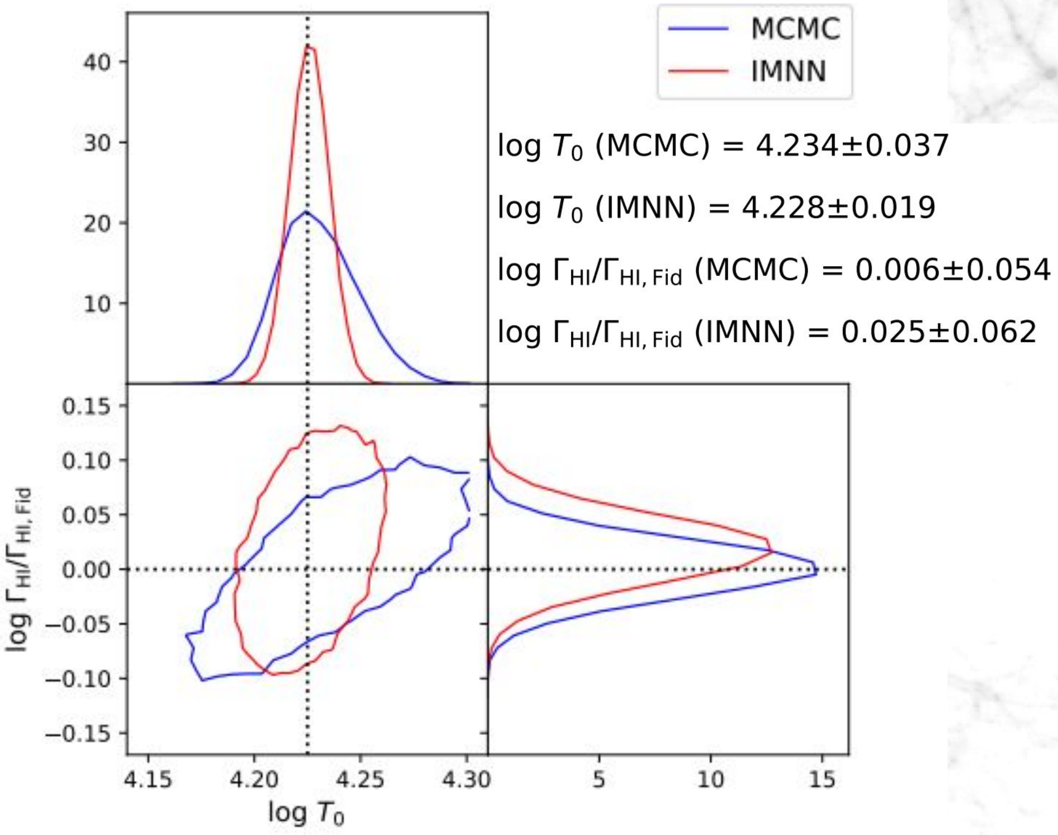
2D parameter estimation (with T_0 and Γ_{HI})





2D parameter estimation (with T_0 and Γ_{HI})





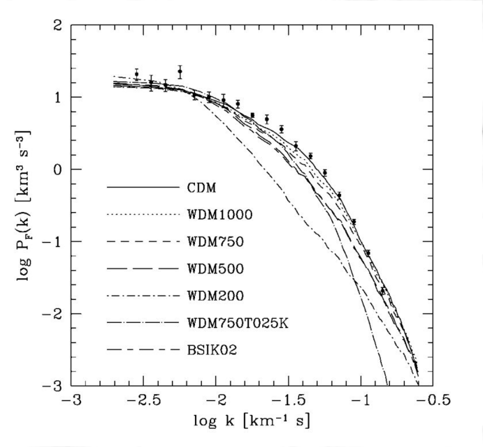
Improved estimates for T0 which has small scale effects.

Future perspective

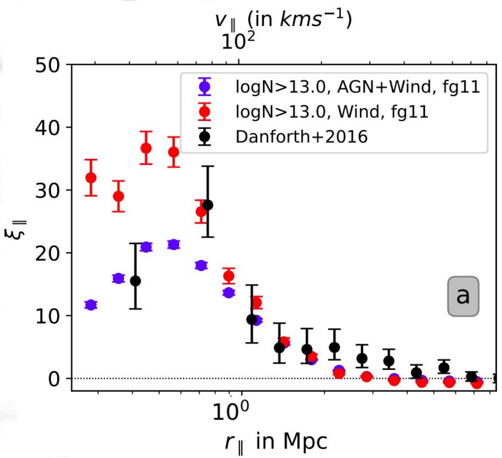
- Apply IMNN on observed high-resolution data to obtain thermal parameter estimates.

Future perspective

- Apply IMNN on observed high-resolution data to obtain thermal parameter estimates.
- Explore physics affecting small scale clustering.



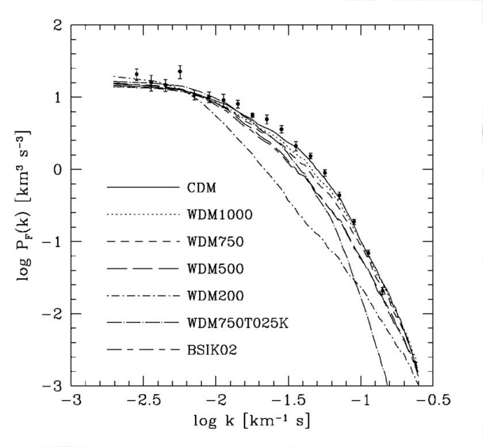
Warm dark matter models (Narayanan+2000)



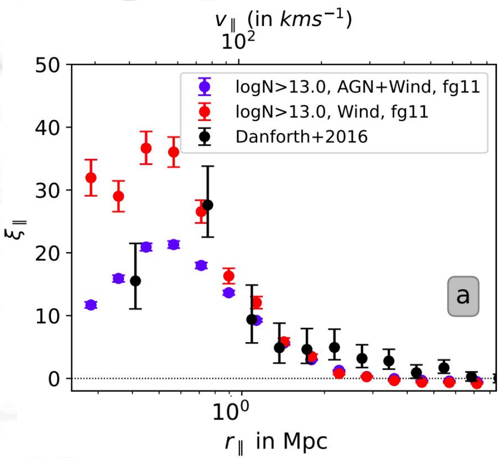
Feedback models (Maitra+2023)

Future perspective

- Apply IMNN on observed high-resolution data to obtain thermal parameter estimates.
- Explore physics affecting small scale clustering.



Warm dark matter models (Narayanan+2000)



Feedback models (Maitra+2023)

- Can be used with other statistics for parameter inference.