Constraining the X-ray heating and reionization using 21-cm power spectra with Marginal Neural Ratio Estimation

Saxena, A.; Cole, A.; Gazagnes, S.; Meerburg, P.D.; Weniger, Christoph; Witte, S.J.

DOI
10.1093/mnras/stad2659

Publication date
2023

Document Version
Final published version

Published in
Monthly Notices of the Royal Astronomical Society

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **525,** 6097–6111 (2023) https://doi.org/10.1093/mnras/stad2659
Advance Access publication 2023 September 6

Constraining the X-ray heating and reionization using 21-cm power spectra with Marginal Neural Ratio Estimation

Anchal Saxena [1★] Alex Cole,[2] Simon Gazagnes [3] P. Daniel Meerburg,[1★] Christoph Weniger[2] and Samuel J. Witte[2,4]

[1]*Van Swinderen Institute, University of Groningen, Nijenborgh 4, NL-9747 AG Groningen, the Netherlands*
[2]*Gravitation Astroparticle Physics Amsterdam (GRAPPA), Institute for Theoretical Physics Amsterdam and Delta Institute for Theoretical Physics, University of Amsterdam, Science Park 904, NL-1098 XH Amsterdam, the Netherlands*
[3]*Department of Astronomy, The University of Texas at Austin, 2515 Speedway, Stop C1400, Austin, TX 78712-1205, USA*
[4]*Departament de Física Quàntica i Astrofísica and Institut de Ciencies del Cosmos Universitat de Barcelona, Diagonal 647, E-08028 Barcelona, Spain*

Accepted 2023 August 31. Received 2023 August 4; in original form 2023 March 30

ABSTRACT

Cosmic Dawn (CD) and Epoch of Reionization (EoR) are epochs of the Universe which host invaluable information about the cosmology and astrophysics of X-ray heating and hydrogen reionization. Radio interferometric observations of the 21-cm line at high redshifts have the potential to revolutionize our understanding of the Universe during this time. However, modelling the evolution of these epochs is particularly challenging due to the complex interplay of many physical processes. This makes it difficult to perform the conventional statistical analysis using the likelihood-based Markov-Chain Monte Carlo (MCMC) methods, which scales poorly with the dimensionality of the parameter space. In this paper, we show how the Simulation-Based Inference through Marginal Neural Ratio Estimation (MNRE) provides a step towards evading these issues. We use `21cmFAST` to model the 21-cm power spectrum during CD–EoR with a six-dimensional parameter space. With the expected thermal noise from the Square Kilometre Array, we are able to accurately recover the posterior distribution for the parameters of our model at a significantly lower computational cost than the conventional likelihood-based methods. We further show how the same training data set can be utilized to investigate the sensitivity of the model parameters over different redshifts. Our results support that such efficient and scalable inference techniques enable us to significantly extend the modelling complexity beyond what is currently achievable with conventional MCMC methods.

Key words: methods: data analysis – methods: statistical – dark ages, reionization, first stars.

1 INTRODUCTION

The Cosmic Dawn (CD) marks the formation of the first sources of light, which produced high-energy X-ray and ultraviolet (UV) radiation. The radiation from these sources heated up the intergalactic medium (IGM) and initiated the Epoch of Reionization (EoR), during which the IGM transitioned from a neutral to ionized state (Barkana & Loeb 2001; Furlanetto, Oh & Briggs 2006a; Pritchard & Loeb 2012). The astrophysics driving the heating and reionization process is still poorly understood, with large uncertainties on the properties of the sources which dominantly contributed to these epochs (e.g. their star formation efficiency, ionizing efficiency, and their X-ray luminosity). Observations of the high-redshift quasar spectra (Becker et al. 2001; Fan et al. 2003; Boera et al. 2019), electron scattering optical depth from the Cosmic Microwave Background (CMB) (Kaplinghat et al. 2003; Komatsu et al. 2011; Planck Collaboration 2020), and the luminosity function and clustering properties of Ly-α emitters (Jensen et al. 2012; Dijkstra 2014; Bouwens 2016; Gangolli et al. 2020) currently provide some constraints on the astrophysical evolution of the CD and EoR. The 21-cm line associated with the spin-flip hyperfine transition of the hydrogen atom offers the most promising probe to study these eras.

There are a large number of ongoing radio interferometric experiments, including GMRT (Paciga et al. 2013), HERA (DeBoer et al. 2017), LOFAR (Ghara et al. 2020; Mertens et al. 2020), LWA (Eastwood et al. 2019), MWA (Barry et al. 2019; Li et al. 2019), and PAPER (Kolopanis et al. 2019). These experiments target the detection of the 21-cm signal by quantifying its spatial fluctuations using various Fourier statistics. We get increasingly interesting upper limits on the 21-cm power spectra from these experiments, some of which already enable us to rule out certain astrophysical models (Ghara et al. 2020; 2021; Mondal et al. 2020; Greig et al. 2021; Abdurashidova et al. 2022; The HERA Collaboration 2022). The upcoming Square Kilometre Array (SKA) (Koopmans et al. 2015; Mellema et al. 2015) is expected to detect the 21-cm power spectrum and owing to its high sensitivity, it is likely that SKA will also be able to do the full tomography of the 21-cm signal.

Once the signal is detected, the next goal would be to constrain the parameters of the CD–EoR models to pin down the astrophysics of the early Universe. Modelling the 21-cm signal from CD–EoR using full radiative transfer simulations (Mellema et al. 2006; Ghara, Choudhury & Datta 2015) is computationally expensive

★ E-mail: anchal.saxena.009@gmail.com (AS); daanmeerburg@gmail.com (PDM)

and unfeasible to perform parameter inference. To overcome this challenge, various approximate and efficient semi-numerical models are used to model the signal accurately at scales ≥ 1 Mpc (Zahn et al. 2011). The traditional framework that is used to explore parameter space is 21CMMC [1] (Greig & Mesinger 2015, 2017, 2018; Park et al. 2019), which uses a semi-numerical framework of the 21-cm signal simulator `21cmFAST`[2] (Mesinger, Furlanetto & Cen 2010) and embed this code in a Markov Chain Monte Carlo (MCMC) sampler. While 21CMMC is quite powerful in systematically performing parameter inference, it becomes computationally quite expensive once we take into account the inhomogeneous X-ray heating in the simulations. Alternatively, one can use analytical models of the 21-cm signal during CD–EoR (Qin et al. 2022; Muñoz 2023). Quite generally, as the dimensionality of parameter space increases, it takes longer for an MCMC, which samples the full joint posterior, to converge.

To circumnavigate these problems, machine learning techniques have been explored in various astrophysical and cosmological problems. In the context of 21-cm cosmology, one common approach is to use emulators, which are trained using artificial neural networks to replace actual simulations. This makes the likelihood evaluations and, consequently, the parameter inference significantly faster (Kern et al. 2017; Schmit & Pritchard 2017; Shimabukuro & Semelin 2017; Tiwari et al. 2022). However, the application of emulators is currently limited to low-order summary statistics. The likelihood could become intractable for higher-order information such as the full 3D 21-cm images. For a tractable likelihood function, the traditional MCMC algorithm can be used to sample from the posterior distribution. However, when the likelihood itself is intractable, techniques such as the Approximate Bayesian Computation (ABC) (Toni et al. 2008) can be used to sample from the approximate posterior. This approach uses simulated data sets to avoid the likelihood evaluations; however, it requires the introduction of summary statistics, which can significantly affect the quality of the approximation.

These issues can be resolved by performing a Simulation-Based Inference (SBI) (Alsing et al. 2019; Papamakarios, Sterratt & Murray 2019; Cranmer, Brehmer & Louppe 2020), where deep learning algorithms along with the ABC are used to estimate the posterior distribution. In this work, we will apply the Marginal Neural Ratio Estimation (MNRE) algorithm (Miller et al. 2021) using SWYFT[3] (Miller et al. 2022). It directly estimates the marginal likelihood-to-evidence ratios through neural networks, which makes it much more efficient than sampling the full joint posterior with an MCMC. In addition, MNRE offers the flexibility to ignore large numbers of nuisance parameters, learning only the parameters of interest. This has already been applied for the cosmological parameter inference from the CMB power spectra (Cole et al. 2022), reconstructing the halo clustering and halo mass function from N-body simulations (Dimitriou, Weniger & Correa 2022), and gravitational lensing analyses (Coogan et al. 2022).

In this work, we use this framework for the astrophysical parameter inference with the 21-cm power spectrum from the CD–EoR. In a recent study, Zhao, Mao & Wandelt (2022b) have performed the reionization parameter inference from the EoR using the density estimation likelihood free inference. Their analysis, however, was limited to a 2D parameter space to model the 21-cm signal during the EoR. Here, we extend the parameter space to six dimensions to also include the parameters that govern the inhomogeneous X-ray heating during the CD. In this case, a single 21-cm power spectrum simulation is $\sim$5 times slower than the former 2D parameter space. This implies that an MCMC for the six-dimensional parameter space would be even slower because of the typical exponential scaling of the required samples as a function of the number of parameters. Moreover, such analysis with conventional methods while also co-varying the cosmic seed for the forward models is pushed out of the realm of feasibility.

Generating the simulated data set and performing the inference with MNRE are two independent processes within SWYFT. This allows us to utilize the same training data set for various applications. To highlight this aspect of SWYFT, in a worked-out example, we will let the neural network determine which set of parameters are sensitive at which redshifts at no extra cost of 21-cm simulations. The distribution of integration time over different redshifts can be considered a proxy to determine which part of the data each parameter is most sensitive to. This could be indicative of the possible degeneracies between parameters for more complex astrophysical models of the 21-cm signal.

This paper is organized as follows: In Section 2, we briefly outline the implementation of MNRE using SWYFT. In Section 3, we describe the 21-cm signal modelling and the parameters of interest. In Section 4, we present the posterior inference and investigate the sensitivity of model parameters in different redshift ranges. We conclude in Section 5. Throughout this work, we assumed a Λ cold dark matter Universe with cosmological parameters $\Omega_{\rm m} = 0.308$, $\Omega_{\rm b} = 0.048$, $\Omega_\Lambda = 0.692$, $h = 0.678$, and $\sigma_8 = 0.81$ (Planck Collaboration 2016a).

2 IMPLEMENTATION OF MNRE USING SWYFT

The probability distribution of model parameters $\boldsymbol{\theta}$ for a given observation $\boldsymbol{x}$ follows from Bayes' theorem

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x})}\, p(\boldsymbol{\theta})\,, \tag{1}$$

where $p(\boldsymbol{x}|\boldsymbol{\theta})$ is the likelihood of the data $\boldsymbol{x}$ for given parameters $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$ is the prior probability distribution over the parameters, and $p(\boldsymbol{x})$ is the evidence of the data.

In SBI, the information about the likelihood is implicitly accessed via a stochastic simulator, which maps from input parameters $\boldsymbol{\theta}$ to data $\boldsymbol{x}$. We generate sample-parameter pairs from this simulator $\{(\boldsymbol{x}^1, \boldsymbol{\theta}^1)\,(\boldsymbol{x}^2, \boldsymbol{\theta}^2), \cdots\}$. Here $\boldsymbol{\theta}^i$ is typically drawn from the prior, so these pairs are drawn from the joint distribution $p(\boldsymbol{x}, \boldsymbol{\theta})$. These pairs are used to train a neural network to approximate the likelihood-to-evidence ratio, a procedure known as Neural Ratio Estimation (Durkan, Murray & Papamakarios 2020; Hermans, Begy & Louppe 2020a, b). Following equation (1), this ratio (which we denote $r(\boldsymbol{x}, \boldsymbol{\theta})$) can be expressed as:

$$r(\boldsymbol{x}, \boldsymbol{\theta}) \equiv \frac{p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x})} = \frac{p(\boldsymbol{\theta}|\boldsymbol{x})}{p(\boldsymbol{\theta})} = \frac{p(\boldsymbol{x}, \boldsymbol{\theta})}{p(\boldsymbol{x})\, p(\boldsymbol{\theta})}\,. \tag{2}$$

In other words, $r(\boldsymbol{x}, \boldsymbol{\theta})$ is equal to the ratio of the joint probability density $p(\boldsymbol{x}, \boldsymbol{\theta})$ to the product of marginal probability densities $p(\boldsymbol{x})\, p(\boldsymbol{\theta})$. A binary classifier $d_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{\theta})$ is then trained to distinguish between jointly drawn and marginally drawn pairs. Here $\boldsymbol{\phi}$ denotes the learnable parameters of the model, which are updated as the model is trained.

More precisely, we introduce a binary label y to denote whether a pair was drawn jointly ($y = 1$) or marginally ($y = 0$). Strictly speaking, y is a random variable. The output of the classifier (assuming it is trained well) approximates the probability that a sample-parameter

[1] https://github.com/BradGreig/21cmMC
[2] https://github.com/21-cmfast/21cmFAST
[3] https://github.com/undark-lab/swyft

pair $(\boldsymbol{x}, \boldsymbol{\theta})$ is drawn jointly $(y = 1)$, i.e.

$$\begin{aligned} d_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{\theta}) &\approx p(y = 1|\boldsymbol{x}, \boldsymbol{\theta}) \\ &= \frac{p(\boldsymbol{x}, \boldsymbol{\theta}|y = 1)p(y = 1)}{p(\boldsymbol{x}, \boldsymbol{\theta}|y = 1)p(y = 1) + p(\boldsymbol{x}, \boldsymbol{\theta}|y = 0)p(y = 0)} \\ &= \frac{p(\boldsymbol{x}, \boldsymbol{\theta})}{p(\boldsymbol{x}, \boldsymbol{\theta}) + p(\boldsymbol{x})p(\boldsymbol{\theta})}, \end{aligned}$$

where we assumed $p(y = 0) = p(y = 1) = \frac{1}{2}$. This learning problem is associated with a binary cross-entropy loss function

$$-\int \left[p(\boldsymbol{x}, \boldsymbol{\theta}) \ln d_{\phi}(\boldsymbol{x}, \boldsymbol{\theta}) + p(\boldsymbol{x})\, p(\boldsymbol{\theta}) \ln\{1 - d_{\phi}(\boldsymbol{x}, \boldsymbol{\theta})\}\right] \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{\theta}, \tag{3}$$

which is minimized using stochastic gradient descent to find the optimal parameters $\boldsymbol{\phi}$ of the network. The binary classifier is simply a dense neural network with a few hidden layers. Once the network is trained, it results in

$$d_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{\theta}) \approx \frac{p(\boldsymbol{x}, \boldsymbol{\theta})}{p(\boldsymbol{x}, \boldsymbol{\theta}) + p(\boldsymbol{x})p(\boldsymbol{\theta})} = \frac{r(\boldsymbol{x}, \boldsymbol{\theta})}{r(\boldsymbol{x}, \boldsymbol{\theta}) + 1}, \tag{4}$$

which can be re-written as

$$r(\boldsymbol{x}, \boldsymbol{\theta}) \approx \frac{d_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{\theta})}{d_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{\theta}) - 1} \Rightarrow p(\boldsymbol{\theta}|\boldsymbol{x}) \approx \frac{d_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{\theta})}{d_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{\theta}) - 1}\, p(\boldsymbol{\theta}) \tag{5}$$

to estimate the posterior probability distribution. This procedure can directly estimate marginal posteriors by omitting model parameters from the network's input, a variant called MNRE. In this work, we use MNRE as implemented in the software package SWYFT (Miller et al. 2022).

3 SIMULATIONS AND TRAINING DATA

3.1 21cmFAST

To model the 21-cm signal and the underlying astrophysics of heating and reionization, we use the publicly available semi-numerical formalism, `21cmFAST` (Mesinger et al. 2010). We first generate the initial density perturbation at $z = 300$ on a high-resolution 1024^3 grid. These perturbations are evolved using the Zel'dovich approximation (Zel'dovich 1970) at later redshifts. To produce the ionization map, the high-resolution density field is first mapped on a coarser grid. Then, `21cmFAST` uses an excursion-set based formalism (Furlanetto, Zaldarriaga & Hernquist 2004) to identify the ionized regions by comparing the number of ionizing photons with the number of baryons within the spheres of decreasing radius $R_{\rm min} \leq R \leq R_{\rm max}$. Here, $R_{\rm min}$ depends on the spatial resolution of the simulation, and $R_{\rm max}$ is the maximum horizon for ionizing photons (see Section 3.1.3). A grid point located at $(\mathbf{x}, z)$ is considered fully ionized if for any $R_{\rm min} \leq R \leq R_{\rm max}$

$$\zeta\, f_{\rm coll}(\mathbf{x}, z, R, M_{\rm min}) \geq 1, \tag{6}$$

where ζ represents the ionizing efficiency (see Section 3.1.1) and $f_{\rm coll}(\mathbf{x}, z, R, M_{\rm min})$ is the fraction of collapsed matter within a spherical region of radius R centred at $(\mathbf{x}, z)$, which depends on the minimum mass of the halo formation $M_{\rm min}$ (Press & Schechter 1974; Sheth & Tormen 1999). The cells that do not satisfy equation (6) are assigned a partial ionization fraction, $\zeta f_{\rm coll}(\mathbf{x}, z, R_{\rm min})$. The resulting ionization map is then converted into the 21-cm brightness temperature map using (Furlanetto, Oh & Briggs)

$$\begin{aligned} \delta T_{\rm b} &= 27(1 - x_{\rm H\,II})(1 + \delta_b) \left(\frac{\Omega_b h^2}{0.023}\right) \left(\frac{0.15}{\Omega_m h^2} \frac{1+z}{10}\right)^{1/2} \\ &\times \left(\frac{T_{\rm S} - T_{\rm CMB}}{T_{\rm CMB}}\right) \left[\frac{\partial_r v_r}{(1+z)H(z)}\right], \end{aligned} \tag{7}$$

where $x_{\rm H\,II}$ is the ionization fraction, $\delta_{\rm b}$ is the baryon overdensity, $\Omega_{\rm m}$ is the matter density, $\Omega_{\rm b}$ is the baryon density, h is the Hubble parameter, $T_{\rm S}$ and $T_{\rm CMB}$ are the spin temperature and CMB temperature, respectively, and the last term takes into account the velocity gradient along the line of sight.

The spin temperature $T_{\rm S}$ can couple to (i) the CMB temperature $T_{\rm CMB}$, in which case $\delta T_{\rm b} = 0$, (ii) the kinetic gas temperature $T_{\rm K}$ through collisional coupling, and (iii) the Ly-α colour temperature $T_{\rm C}$ through the Wouthuysen–Field coupling (Wouthuysen 1952), where $T_{\rm C} \approx T_{\rm K}$. To track the evolution of the gas temperature, `21cmFAST` simulates the inhomogeneous heating of the IGM by X-rays by integrating the angle-averaged specific X-ray emissivity ($\epsilon_{\rm X}$) along the lightcone for each cell. The specific X-ray emissivity is given as (Mesinger et al. 2010; Greig & Mesinger 2017)

$$\epsilon_{\rm X}(\boldsymbol{x}, E, z) = \frac{L_{\rm X}}{\rm SFR} \left[\rho_{\rm crit,0} \Omega_b f_{\star} (1 + \delta_{\rm nl}) \frac{\mathrm{d} f_{\rm coll}(z)}{\mathrm{d}t}\right], \tag{8}$$

where $\rho_{\rm crit,0}$ is the current critical density, $f_{\star}$ is the fraction of baryons in stars, $\delta_{\rm nl}$ is the evolved density. The term enclosed in square brackets is the star formation rate (SFR) density along the lightcone. $L_{\rm X}$ is the specific X-ray luminosity which is assumed to follow a power law, $L_{\rm X} \propto E^{-\alpha_{\rm X}}$. The photons below an energy threshold E_0 are absorbed by the interstellar medium. The X-ray efficiency is normalized by quantifying an integrated soft-band (<2 keV) luminosity per SFR

$$\frac{L_{\rm X<2\,keV}}{\rm SFR} = \int_{E_0}^{2\,\rm keV} \left(\frac{L_{\rm X}}{\rm SFR}\right) \mathrm{d}E. \tag{9}$$

The semi-numerical model adopted in this work consists of six astrophysical parameters which govern the evolution of the 21-cm signal during the CD–EoR. We briefly describe each of these parameters and the adopted priors below.

3.1.1 *Ionizing efficiency, (ζ)*

The UV ionizing efficiency of high-redshift galaxies can be expressed in terms of various factors as (Barkana & Loeb 2001; Mesinger et al. 2010)

$$\zeta = 30 \left(\frac{f_{\rm esc}}{30}\right) \left(\frac{f_{\star}}{0.05}\right) \left(\frac{N_{\gamma/b}}{4000}\right) \left(\frac{2}{1 + n_{\rm rec}}\right), \tag{10}$$

where $f_{\rm esc}$ is the fraction of ionizing photons that escape into the IGM, $f_{\star}$ is the fraction of galactic gas in stars, $N_{\gamma/b}$ is the number of ionizing photons produced per baryon in stars, and $n_{\rm rec}$ is the average number of times a hydrogen atom recombines. We assume a single population of efficient star-forming galaxies (a constant ionizing efficiency for all the galaxies) hosted by haloes with a sufficient mass.

The timing and duration of reionization strongly depend on ζ. Large values of ζ will speed up the ionization process if we keep the other parameters fixed. We adopt a flat prior $\zeta \in (10, 100)$, although an extended range with the upper limit of $\zeta = 250$ has also been studied in Greig & Mesinger (2017) to explore the models where the EoR is driven by rare, very bright galaxies.

3.1.2 Minimum virial temperature of haloes, $T_{\rm vir}^{\rm min}$

The minimum threshold for a halo to host a star-forming galaxy is defined in terms of its virial temperature, $T_{\rm vir}^{\rm min}$. It is related to the mass of the halo (Barkana & Loeb 2001) as:

$$M_{\rm vir}^{\rm min} = \frac{10^8}{h}\left[\frac{0.6}{\mu}\frac{10}{1+z}\frac{T_{\rm vir}^{\rm min}}{1.98\times10^4}\right]^{3/2}\left[\frac{\Omega_m}{\Omega_m^z}\frac{\Delta_c}{18\pi^2}\right]^{-1/2} M_\odot, \tag{11}$$

where μ is the mean molecular weight, $\Omega_m^z = \Omega_m(z)$, and $\Delta_c = 18\pi^2 + 82d - 39d^2$ where $d = \Omega_m^z - 1$. The choice of $T_{\rm vir}^{\rm min}$ determines the cut-off in the UV luminosity function. Galaxies that are hosted within a halo with $T_{\rm vir} < T_{\rm vir}^{\rm min}$ have no contribution to star formation due to internal feedback processes. We note that $T_{\rm vir}^{\rm min}$ has a significant impact on both the EoR and the Epoch of Heating (EoH) because, within the `21cmFAST` framework, the physics of star formation drives both the X-ray heating and ionization fields.

We adopt a flat prior on $T_{\rm vir}^{\rm min} \in (10^4, 10^6)$ K. The minimum temperature required for efficient atomic cooling defines our lower limit of $T_{\rm vir}^{\rm min} = 10^4$ K, and the upper limit is consistent with the observation of Lyman break galaxies at high redshifts (Kuhlen & Faucher-Giguère 2012; Barone-Nugent et al. 2014).

3.1.3 Mean free path of the ionizing photons, R_{mfp}

The physical size of the ionized region is governed by the distance ionizing photons propagate through the IGM, which depends on the population of the photon absorption systems where recombinations take place. To take into account this effect, we define $R_{\rm mfp}$ as the maximum horizon for the ionizing photons.

It has been shown by Greig & Mesinger (2017) that this parameter is only sensitive during the later stages of reionization when the typical size of the H II regions approaches $R_{\rm mfp}$. We use a flat prior on $R_{\rm mfp} \in (5, 25)$ cMpc similar to Greig & Mesinger (2015), which is consistent with the sub-grid recombination model of Sobacchi & Mesinger (2014).

3.1.4 Integrated soft-band luminosity, $L_{\rm X<2keV}/{\rm SFR}$

The total integrated soft-band (<2 keV) luminosity per SFR escaping the host galaxies ($L_{\rm X<2keV}/{\rm SFR}$) controls the efficiency with which X-rays heat the IGM. It decides the timing and duration of the EoH in a manner similar to ζ for the EoR.

For sufficiently large values of $L_{\rm X<2keV}/{\rm SFR}$, the X-rays can also ionize the IGM at $\sim 10-20$ per cent level, in addition to heating. We use a flat prior on $\log_{10}(L_{\rm X<2keV}/{\rm SFR}) \in (38, 42)$. This range is motivated by population synthesis models at high redshifts (Fragos et al. 2013) and the observations of the local population of galaxies (Mineo, Gilfanov & Sunyaev 2012; Sazonov & Khabibullin 2017).

3.1.5 X-ray energy threshold for self-absorption by the host galaxies, E_0

The soft X-rays produced by the host galaxies can be absorbed by the interstellar medium, in which case they can no longer contribute to the heating of the IGM. From the simulations of high z galaxies, it has been shown by Das et al. (2017) that the attenuation of the X-ray profile can be approximated by a step function below an energy threshold E_0.

The small values of E_0 lead to very efficient and inhomogeneous heating. It has been shown by Pacucci et al. (2014) that the amplitude of the power spectra for such softer spectral energy distributions (SEDs) is larger by up to an order of magnitude. We adopt a flat prior on $E_0 \in (0.1, 1.5)$ keV.

Table 1. Observation parameters for SKA1 low configuration used in this work to simulate the thermal noise.

Parameter	Value
$N_{\rm ant}$	512
$\Delta\nu$	195.3 kHz
Δt	10 s
$t_{\rm obs}^{\rm day}$	6 h
$t_{\rm obs}^{\rm tot}$	1000 h

3.1.6 X-ray spectral index, α_X

The spectral index governs the spectrum that emerges from the X-ray sources and depends on the dominant physical process emitting the X-ray photons. We take a flat prior on $\alpha_{\rm X} \in (-0.5, 2.5)$ similar to Greig & Mesinger (2017) to take into account various relevant X-ray SEDs such as high-mass X-ray binaries, mini-quasars, host interstellar medium, supernovae remnants.

Our simulations are performed within a $[250\ {\rm cMpc}]^3$ box on a $[128]^3$ grid. The training data is composed of 20 000 power spectra samples evaluated at ten different redshifts in range (25, 6). These samples are drawn randomly from the priors. We use 80 per cent of the samples for training, 10 per cent for validation, and 10 per cent for the test data set. We also vary the cosmic seed in our forward models. The impact of the size of the training data is investigated in Appendix D.

3.2 Telescope noise profile

To simulate the thermal noise, we first estimate the uv coverage for SKA1 low, assuming 1000 h of observations. Thermal noise is simulated using `ps_eor`[4] by creating a system equivalent flux density of 2500 Jy at the central frequency of the observation. The current configuration of SKA1-Low has 512 stations, 224 of which are placed randomly in a circular core of radius 350 m. The remaining 288 stations are distributed among 36 clusters in three spiral arms extending up to a radius of 35 km from the central core. We integrate for 10 s per visibility and observe for 6 h each day with a frequency resolution of 195.3 kHz. These parameters are tabulated in Table 1. This results in the thermal noise ($\sigma_{\rm therm}$). Note that in Greig & Mesinger (2017), the authors also include a 20 per cent modelling uncertainty on the sampled power spectra to take into account the differences with various semi-numerical and radiative transfer simulations. This can be easily incorporated without running any additional 21-cm signal simulations in our analysis. The impact of including the modelling uncertainty is discussed in Appendix C.

3.3 Mock observation

To form our mock observation, we consider a model with $\{\zeta, \log_{10}(T_{\rm vir}^{\rm min}), R_{\rm mfp}, \log_{10}(L_{\rm X}), E_0, \alpha_{\rm X}\} = \{30, 4.70, 15, 40.5, 0.5, 1\}$. It corresponds to the FAINT GALAXIES model from Mesinger,

[4] https://gitlab.com/flomertens/ps_eor

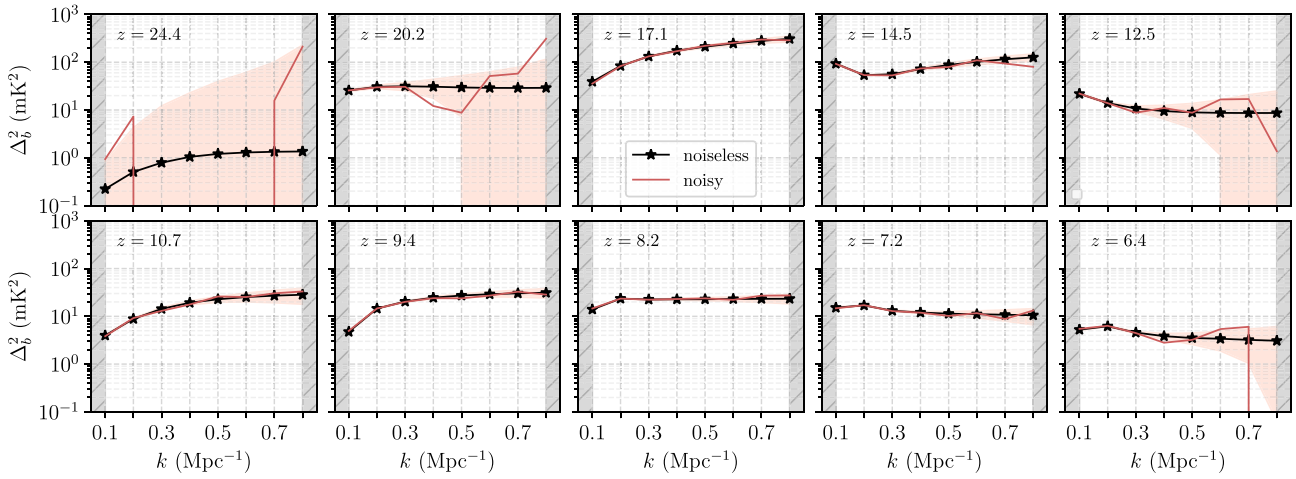

Figure 1. The cosmological (black) and noisy (orange) mock power spectrum for the FAINT GALAXIES model at different redshifts, where $k \in (0.1, 0.8)$ Mpc^{-1}. The shaded region represents the power spectrum uncertainty level.

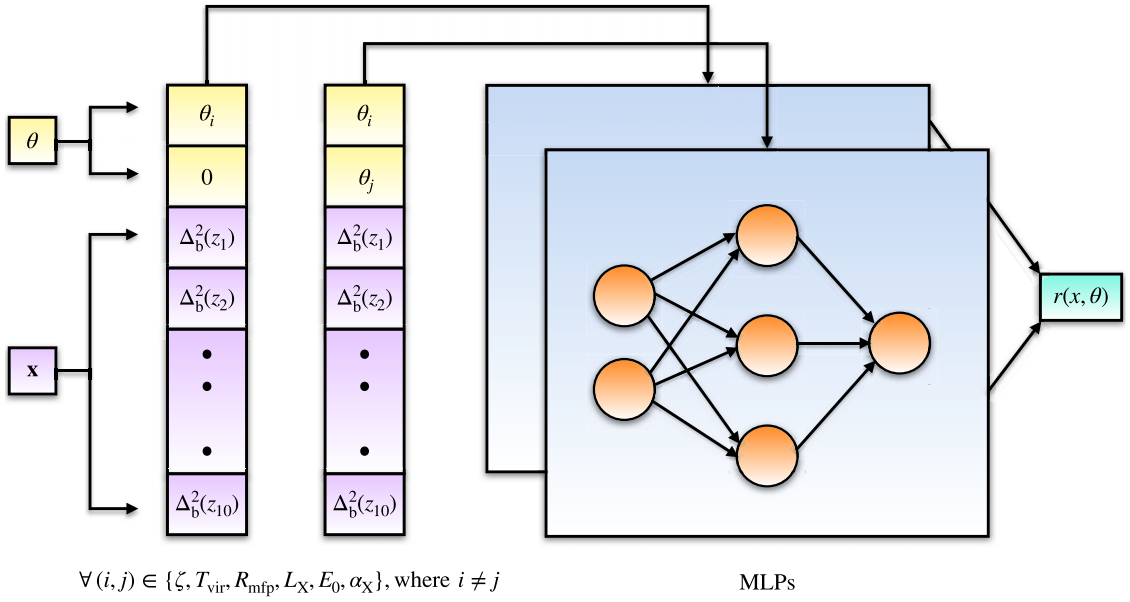

Figure 2. Illustration of the network architecture. Input data $\boldsymbol{x}$ and parameters $\boldsymbol{\theta}$ are mapped to the marginal parameter combinations. The individual ratio estimators are trained with an MLP. The outputs are estimated ratios $r(\boldsymbol{x}, \boldsymbol{\theta})$ for the marginal posteriors of interest.

Greig & Sobacchi (2016) and Greig & Mesinger (2017) in which reionization is driven by numerous sources with low ionizing efficiency. This set of parameter values results in the reionization history and the Thomson optical depth τ consistent with Planck data (Planck Collaboration 2016b). The mock observation is simulated within a [500 cMpc]^3 box on a [256]^3 grid.

In Fig. 1, we show the cosmological 21-cm power spectra (black line) from our mock observation at different redshifts. The shaded region represents the 21-cm power spectrum uncertainty. We then draw a random realization from the normal distribution $\sim \mathcal{N}(0, \sigma^2_{\rm therm}(k, z))$, and add it to the cosmological 21-cm power spectrum to form the noisy mock observation (orange line). We restrict our analysis to the k-modes in the range $k \in (0.1, 0.8)$ Mpc^{-1} to avoid the impact of foreground contamination on large scales and thermal noise on small scales (Greig & Mesinger 2015, 2017).

4 RESULTS

In this section, we discuss the application of SWYFT to obtain the posterior probability distributions for our six-dimensional 21-cm power spectra model for the simulated mock observation and explore the constraints on different astrophysical parameters as a function of redshift in Section 4.1. In Section 4.2, we show how the distribution of integration time over different redshifts can be used as a proxy to find which part of the data each model parameter is sensitive to. These examples emphasize the flexibility of our framework.

4.1 Posterior inference with SWYFT

To obtain the posterior distribution from MNRE, we first concatenate the power spectra from different redshifts into a 1D array. This is then fed as the input for the multi-layer perceptron (MLP) with three

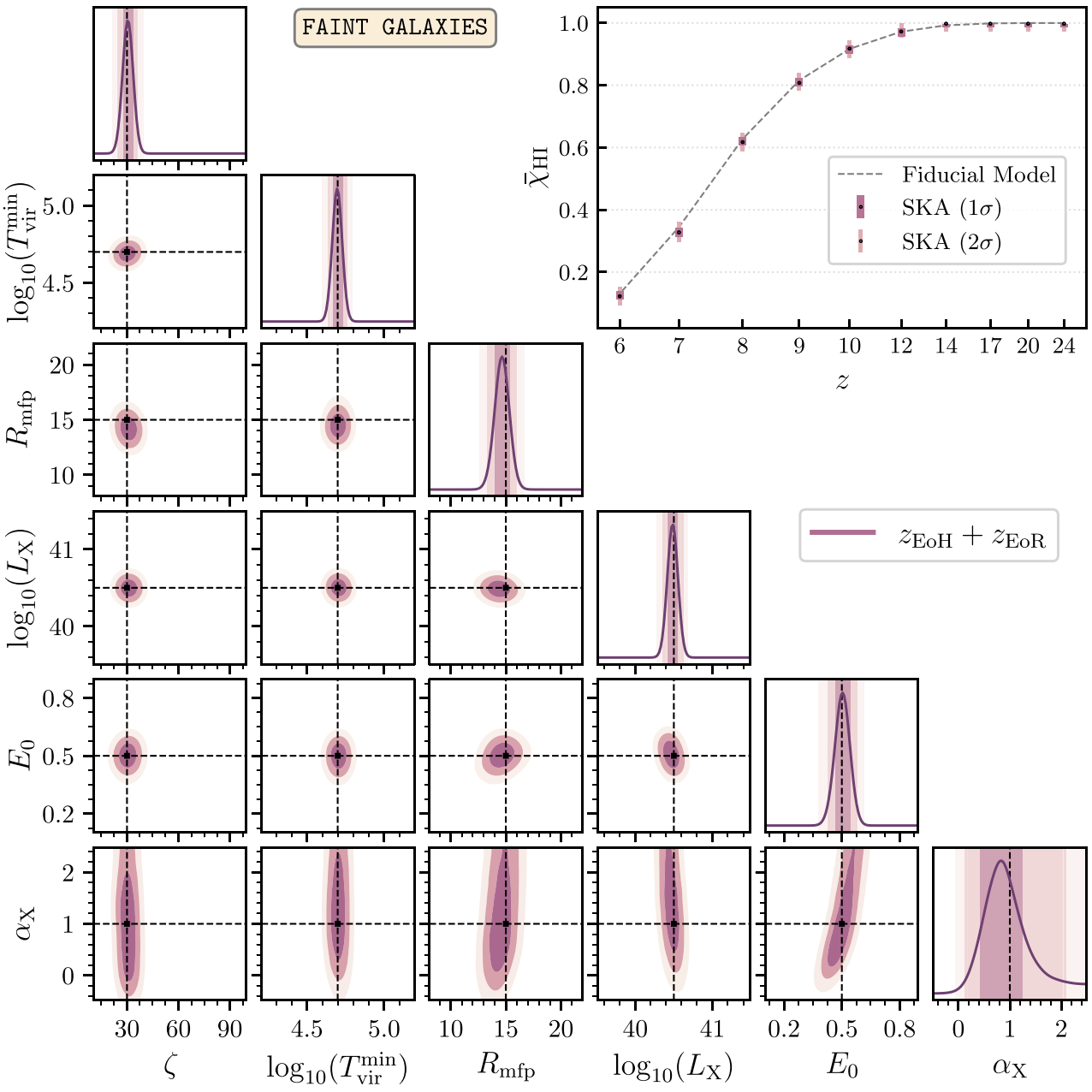

Figure 3. Recovered 1D and 2D marginals for the six-dimensional FAINT GALAXIES model assuming 1000 h observation from SKA. The dashed lines denote the input parameters $\{\zeta, \log_{10}(T_{\rm vir}^{\rm min}), R_{\rm mfp}, \log_{10}(L_{\rm X}), E_0, \alpha_{\rm X}\} = \{30, 4.70, 15, 40.5, 0.5, 1\}$. Inset: Recovered 1σ and 2σ constraints on the reionization history. The dashed line shows the evolution of $\bar{\chi}_{\rm H\,I}(z)$ for the fiducial model.

layers, each containing 256 neurons. The network is trained with a batch size of 64, and we decay the initial learning rate of 10^{-3} by 0.95 after every epoch. The output of the trained network is the estimated ratios for the parameters of interest. In Fig. 2, we show a schematic diagram of the network architecture. Once the network is trained, the ratio estimator allows for very fast MCMC sampling from the approximate posterior.

In Fig. 3, we present the posteriors on the astrophysical parameters obtained from SWYFT for the FAINT GALAXIES model assuming 1000 h observation from the SKA. The diagonal panels show the 1D marginalized posterior for each parameter, and 2D marginals are shown in the lower off-diagonal panels. The dashed lines represent the true value of the parameters. The inferred model parameters and the corresponding 16th and 84th percentiles are tabulated in Table 2.

Consistent with Greig & Mesinger (2017), we are able to tightly constrain all our model parameters except $\alpha_{\rm X}$, which has a relatively small impact on the amplitude of the 21-cm power spectra. The small degeneracies between $\zeta - \log_{10}(T_{\rm vir}^{\rm min})$ and $E_0 - \alpha_{\rm X}$ are in agreement with Greig & Mesinger (2017), and the findings of Ewall-Wice et al.

Table 2. The inferred parameter values and the associated 16th and 84th percentiles for the posteriors from (i) $z_{\rm EoH} + z_{\rm EoR}$ shown in Fig. 3, and (ii) $z_{\rm EoH}$ and $z_{\rm EoR}$ shown in Fig. 4.

Model	ζ	$\log_{10}(T_{\rm vir}^{\rm min})$	$R_{\rm mfp}$	$\log_{10}(L_{\rm X})$	E_0	$\alpha_{\rm X}$
$z_{\rm EoH} + z_{\rm EoR}$	$30.25^{+2.70}_{-1.80}$	$4.70^{+0.03}_{-0.02}$	$14.65^{+0.56}_{-0.56}$	$40.49^{+0.04}_{-0.06}$	$0.50^{+0.03}_{-0.03}$	$0.84^{+0.39}_{-0.39}$
$z_{\rm EoH}$	$22.15^{+5.40}_{-5.40}$	$4.70^{+0.03}_{-0.02}$	–	$40.49^{+0.06}_{-0.06}$	$0.49^{+0.03}_{-0.04}$	$0.68^{+0.51}_{-0.45}$
$z_{\rm EoR}$	$29.35^{+2.70}_{-3.60}$	$4.66^{+0.04}_{-0.05}$	$14.65^{+0.56}_{-0.56}$	$40.47^{+0.12}_{-0.12}$	$0.32^{+0.07}_{-0.10}$	–

(2016) and Kern et al. (2017). In the top right panel of Fig. 3, we show the 1σ and 2σ constraints on the mean neutral fraction ($\bar{\chi}_{\rm HI}$) as a function of redshift z, where the dashed line represents the true ionization history of the model. For this analysis, we use $\bar{\chi}_{\rm HI}(z)$ in place of parameters $\boldsymbol{\theta}$ of the network architecture shown in Fig. 2. We find tight constraints on the ionization history from the SKA.

Next, we investigate the sensitivity of our model parameters during different redshifts. We perform the parameter inference by dividing the entire redshift range into two bins: (i) $z_{\rm EoH} \in (25, 12)$, which corresponds to the X-ray heating, and (ii) $z_{\rm EoR} \in (11, 6)$ that corresponds to reionization. Note that this analysis does not require any extra 21-cm power spectra simulations. The same training data can be re-used with a minimal change in the network's architecture, which is not possible for an MCMC analysis. In this case, the MLP takes the power spectra from $z_{\rm EoH}$ (or $z_{\rm EoR}$) as the input and estimates the ratios for the parameters of interest.

In Fig. 4, we show the resulting 1D and 2D marginal posteriors from $z_{\rm EoH}$ (red) and $z_{\rm EoR}$ (blue). The inferred parameters and the corresponding 16th and 84th percentiles are tabulated in Table 2. We find that ζ, $\log_{10}(T_{\rm vir}^{\rm min})$ and $R_{\rm mfp}$ are well constrained with the 21-cm power spectra from $z_{\rm EoR}$, which is expected as these parameters play a significant role during reionization. On the other hand, $\log_{10}(T_{\rm vir}^{\rm min})$, $\log_{10}(L_{\rm X})$ and E_0 are constrained with the power spectra from $z_{\rm EoH}$. Note that the minimum virial temperature of a halo to host the star-forming galaxies, $T_{\rm vir}^{\rm min}$ can be well constrained with either redshift bin because, within `21cmFAST`, the galaxies that host the ionizing sources are the same galaxies that are responsible for X-ray heating. So, this parameter impacts both the EoH and EoR.

In the top right panel of the Fig. 4, we show the constraints on reionization history from the 21-cm power spectra during $z_{\rm EoH}$ (red) and $z_{\rm EoR}$ (blue). We find that from the 21-cm power spectra at $z_{\rm EoH}$, we can constrain the neutral fraction reasonably well at $z \geq 12$ ($z_{\rm EoH}$), but it does not provide tight constraints during the intermediate and late stages of reionization. However, with the 21-cm power spectra at $z_{\rm EoR}$, we can infer the entire reionization history of the FAINT GALAXIES model. The constraints from the 21-cm power spectra at $z_{\rm EoR}$ on the neutral fraction at $z_{\rm EoH}$ comes from the fact that throughout our models, the neutral fraction $\bar{\chi}_{\rm HI} \approx 1$ at $z \geq 12$.

4.2 Information from different redshifts

In order to study the information content from different redshifts, we consider a toy scenario where we use the distribution of integration time over different redshifts as a proxy to find which part of the data each parameter is most sensitive to. So far, in our analysis, we considered the distribution of integration time to be uniform over redshifts. However, for a fixed total integration time, this distribution can be optimized since the thermal noise level at redshift z depends on the integration time t_z allocated for that redshift.

The optimization is achieved via gradient descent by maximizing the information the network learns about any given parameter from different redshifts. For a fixed total integration time $T_{\rm tot}$, we optimize the integration time for each redshift t_z, such that $T_{\rm tot} = \sum t_z$. We parametrize t_z as:

$$t_z = (T_{\rm tot} \times softmax(\boldsymbol{v}))_z \ , \tag{12}$$

where $\boldsymbol{v}$ is a vector that corresponds to the number of redshift bins. Larger components in $\boldsymbol{v}$ correspond to more integration time for that redshift bin. Next, we consider $\boldsymbol{v}$ to be one of the network parameters that is optimized during the training. As we train the classifier (MLP) to learn the 1D posterior for any given parameter, at the same time, it learns the optimal way of distributing the integration time for that parameter. We further obtain the uncertainties on the optimal time distribution through Monte Carlo Dropout (MCD) (Gal & Ghahramani 2016).

In Fig. 5, we show the results from this information content analysis. For each parameter, the top panel shows the uniform time distribution (orange dashed line) and the optimized time distribution (violins), where the uncertainties follow from the MCD. The bottom panel shows the histogram of the 1σ uncertainty interval on the posterior distribution of each parameter from 500 different mock observations drawn randomly from the test data set assuming the uniform (orange) and optimized (black) time distribution.

We find that the parameters $\log_{10}(T_{\rm vir}^{\rm min})$, $\log_{10}(L_{\rm X})$, E_0, and $\alpha_{\rm X}$ are assigned a larger integration time at high redshifts $z \geq 12$ after the optimization of the network. This implies that the information for these parameters is contained at high redshifts. These findings are consistent with the posteriors from $z_{\rm EoH}$ and $z_{\rm EoR}$ shown in Fig. 4. On the contrary, for the mean free path $R_{\rm mfp}$, the network allocates large integration time at redshift $z = 9$. This is also in agreement with the analysis in Fig. 4, where we found a flat posterior on $R_{\rm mfp}$ from $z_{\rm EoH}$, and the constraints only came from $z_{\rm EoR}$. For each parameter, the histogram of the 1σ uncertainty interval from the optimized integration time distribution tends towards lower 1σ uncertainty on the posterior distribution, which indicates that the network learns more information about a given parameter from the optimized time distribution in comparison to the uniform time distribution.

5 SUMMARY

In this paper, we performed Simulation-Based Inference through an MNRE algorithm, SWYFT, to constrain the astrophysical parameters that govern the X-ray heating and reionization during the CD–EoR. We used `21cmFAST` to model the 21-cm power spectra during CD–EoR with a six-dimensional astrophysical parameter space. We showed that this framework is significantly more efficient as it directly learns the marginal posteriors of interest through neural networks than the conventional likelihood-based methods such as MCMC, which samples the full joint posterior.

With the training data composed of 20 000 21-cm power spectra simulations and the expected thermal noise level from the SKA, we were able to constrain the parameters of our model. The 1D and 2D marginal posteriors obtained through MNRE look consistent

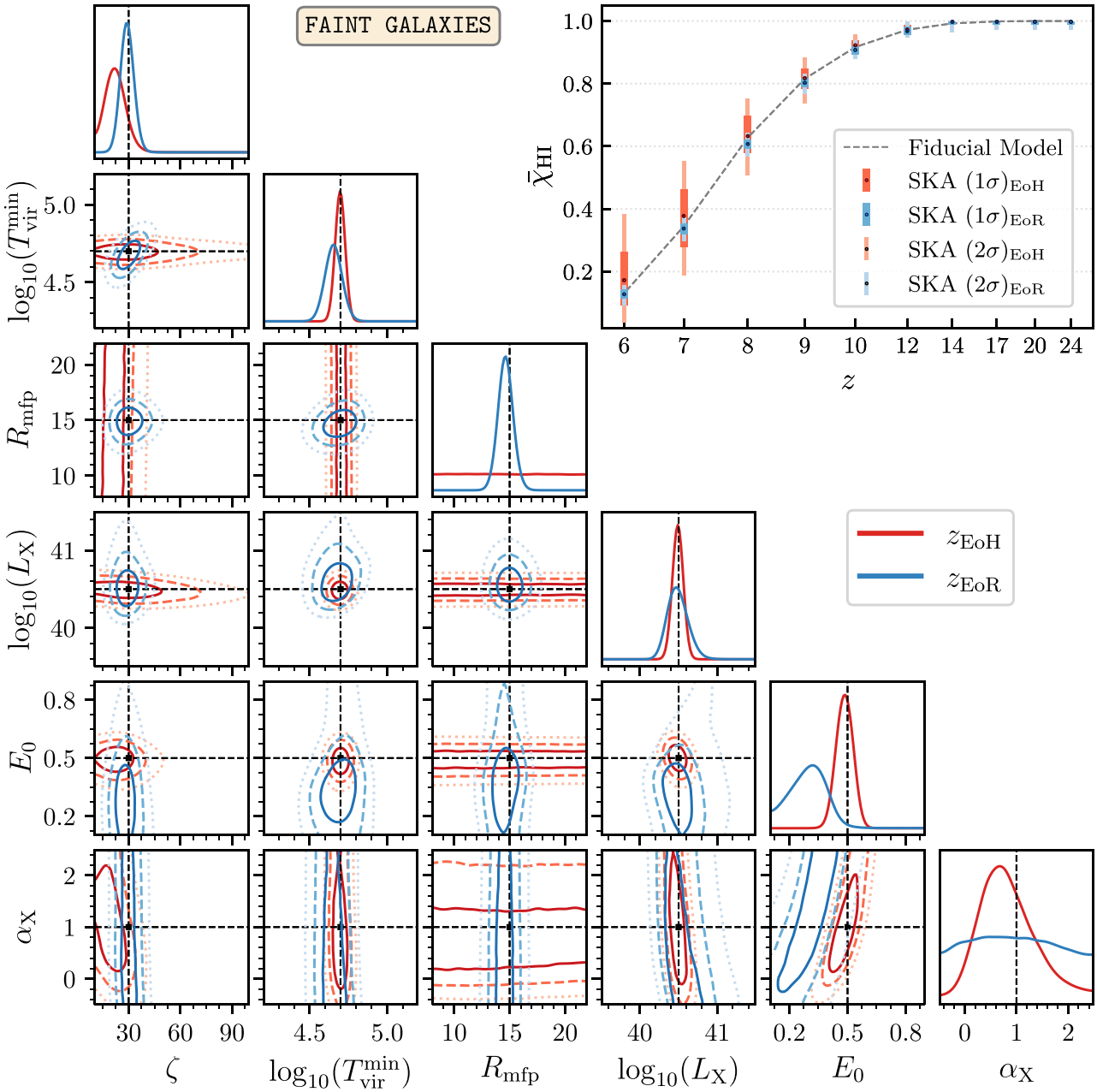

Figure 4. Recovered 1D and 2D marginals from $z_{\rm EoH}$ (red) and $z_{\rm EoR}$ (blue) for the six-dimensional FAINT GALAXIES model assuming 1000 h observation with SKA. The dashed lines denote the input parameters $\{\zeta, \log_{10}(T_{\rm vir}^{\rm min}), R_{\rm mfp}, \log_{10}(L_{\rm X}), E_0, \alpha_{\rm X}\} = \{30, 4.70, 15, 40.5, 0.5, 1\}$. The inset plot shows the recovered reionization history from the 21-cm power spectra during $z_{\rm EoH}$ (red) and $z_{\rm EoR}$ (blue).

with the earlier studies performed with an MCMC, which required an order of magnitude more samples to converge. We further checked the statistical consistency of the trained network by evaluating the nominal and empirical expected coverage probabilities.

Within SWYFT, generating the training data set and MNRE are two independent processes. This feature gives us the flexibility to reuse the simulations and utilize the same training data set for various applications. To demonstrate this aspect of SWYFT, we investigated the sensitivity of different parameters over two different redshift ranges that correspond to the EoH ($z_{\rm EoH}$) and EoR ($z_{\rm EoR}$). We obtained the posterior probability distribution on the model parameters from $z_{\rm EoH}$ and $z_{\rm EoR}$ at no extra cost of 21-cm power spectra simulation. An MCMC analysis in this scenario would otherwise require a new chain, and the simulations can not be used efficiently.

We further studied the information content for each parameter from different redshifts by considering a toy scenario where we consider the distribution of the integration time to be part of the network parameters, which is optimized during the training. We found the optimized time distribution to be consistent with the posterior probability distribution of model parameters from $z_{\rm EoH}$ and $z_{\rm EoR}$. This could be used as an indicator of the possible degeneracies for more complex astrophysical 21-cm signal models without running

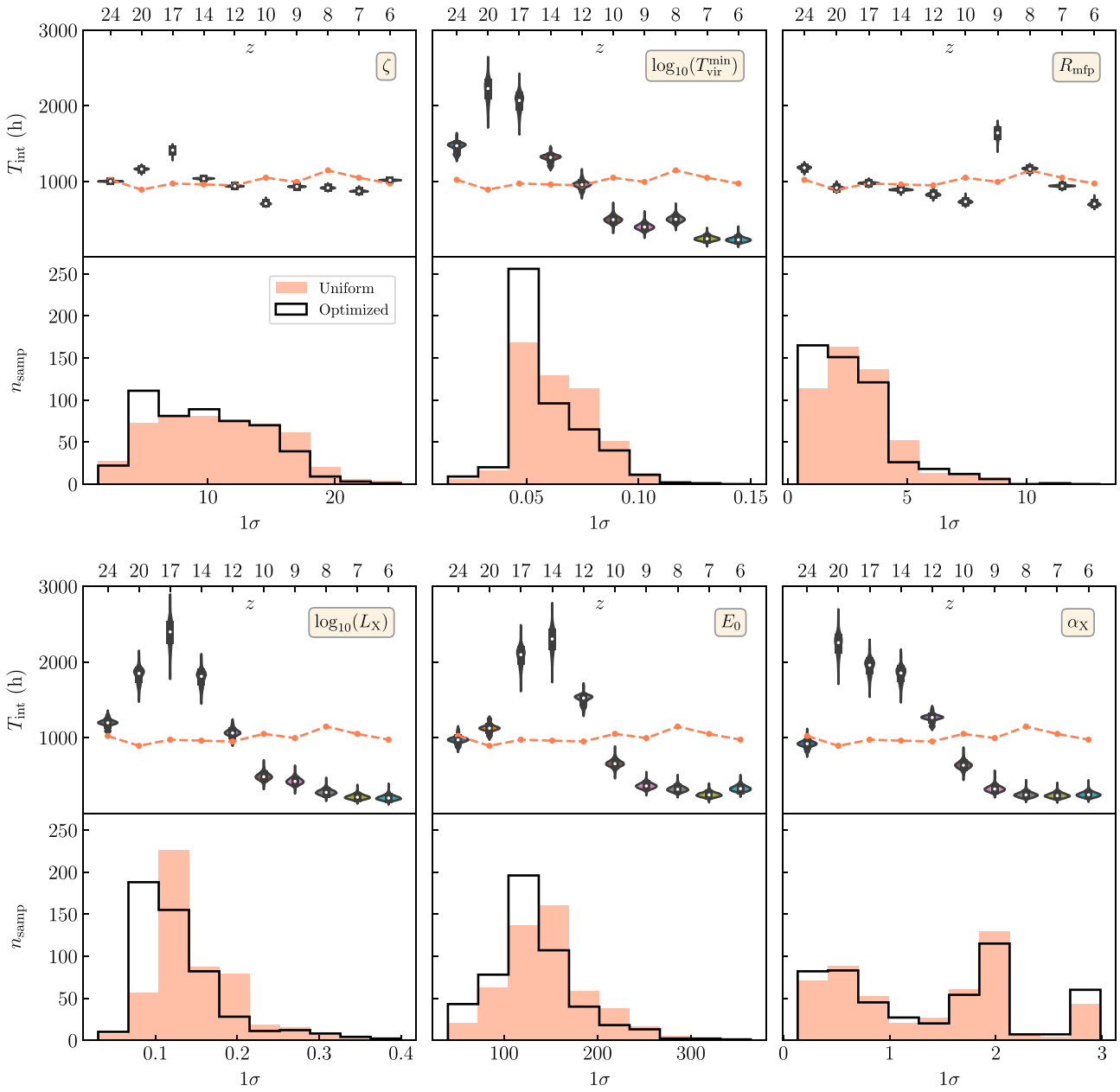

Figure 5. Result of integration time optimization as a proxy for information content for ζ, $\log_{10}(T_{\rm vir}^{\rm min})$, $R_{\rm mfp}$, $\log_{10}(L_{\rm X})$, E_0, $\alpha_{\rm X}$. *Top panel:* Violins represent the optimized time distribution, and the orange line shows the initial time distribution. *Bottom panel:* The histogram of 1σ uncertainty interval on the posterior distribution of each parameter for 500 different mock observations drawn randomly from the test data set assuming a uniform (orange) and optimized (black) time distribution.

additional simulations. This establishes that with such efficient and scalable inference techniques, one can increase the complexity of the 21-cm model even further, which could otherwise be impractical for the likelihood-based approaches.

While our analysis has shown that MNRE is a powerful framework to analyse the 21-cm power spectrum, in reality, the 21-cm signal during CD–EoR is highly non-Gaussian (Shimabukuro et al. 2016; Majumdar et al. 2018; Watkinson et al. 2018), so the 21-cm power spectrum is probably not the most optimal summary statistics to use for parameter inference. In future work, we plan to explore the higher-order summary statistics such as the 21-cm bispectrum (Tiwari et al. 2022), the morphology of the ionized regions (Gazagnes, Koopmans & Wilkinson 2021; Kapahtia et al. 2021) and convolutional neural networks on the 21-cm tomographic images (Gillet et al. 2019; Zhao et al. 2022a) for parameter inference through MNRE.

ACKNOWLEDGEMENTS

We thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster. PDM acknowledges support from the Netherlands organization for scientific research (NWO) VIDI grant (dossier 639.042.730). SG acknowledges support from the Harlan J. Smith McDonald fellowship. SJW and CW were sup-

ported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement no. 864035 – Undark). The project was partially funded by the Netherlands eScience Center, grant number ETEC.2019.018.

DATA AVAILABILITY

Accompanying code is available at https://github.com/anchal-009/swyft21cm. The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Abdurashidova Z. et al., 2022, ApJ, 924, 51
Alsing J., Charnock T., Feeney S., Wandelt B., 2019, MNRAS, 488, 4440
Barkana R., Loeb A., 2001, Phys. Rep., 349, 125
Barone-Nugent R. L. et al., 2014, ApJ, 793, 17
Barry N. et al., 2019, ApJ, 884, 1
Becker R. H. et al., 2001, AJ, 122, 2850
Boera E., Becker G. D., Bolton J. S., Nasir F., 2019, ApJ, 872, 101
Bouwens R., 2016, Understanding the Epoch of Cosmic Reionization. Springer International Publishing, Switzerland, p. 111
Cole A., Miller B. K., Witte S. J., Cai M. X., Grootes M. W., Nattino F., Weniger C., 2022, J. Cosmol. Astropart. Phys., 2022, 004
Coogan A., Montel N. A., Karchev K., Grootes M. W., Nattino F., Weniger C., 2022,
Cranmer K., Brehmer J., Louppe G., 2020, Proc. Natl. Acad. Sci. USA, 117, 30055
Das A., Mesinger A., Pallottini A., Ferrara A., Wise J. H., 2017, MNRAS, 469, 1166
DeBoer D. R. et al., 2017, PASP, 129, 045001
Dijkstra M., 2014, Publ. Astron. Soc. Aust., 31
Dimitriou A., Weniger C., Correa C. A., 2022,
Durkan C., Murray I., Papamakarios G., 2020,
Eastwood M. W. et al., 2019, AJ, 158, 84
Ewall-Wice A., Hewitt J., Mesinger A., Dillon J. S., Liu A., Pober J., 2016, MNRAS, 458, 2710
Fan X. et al., 2003, AJ, 125, 1649
Fragos T., Lehmer B. D., Naoz S., Zezas A., Basu-Zych A., 2013, ApJ, 776, L31
Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, ApJ, 613, 1
Furlanetto S. R., Oh S. P., Briggs F. H., 2006a, Phys. Rep., 433, 181
Gal Y., Ghahramani Z., 2016, Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning, Vol: 48, Proceedings of the 33rd International Conference on Machine Learning, New York
Gangolli N., D'Aloisio A., Nasir F., Zheng Z., 2020, MNRAS, 501, 5294
Gazagnes S., Koopmans L. V. E., Wilkinson M. H. F., 2021, MNRAS, 502, 1816
Ghara R., Choudhury T. R., Datta K. K., 2015, MNRAS, 447, 1806
Ghara R. et al., 2020, MNRAS, 493, 4728
Ghara R., Giri S. K., Ciardi B., Mellema G., Zaroubi S., 2021, MNRAS, 503, 4551
Gillet N., Mesinger A., Greig B., Liu A., Ucci G., 2019, MNRAS, 484, 282,
Greig B., Mesinger A., 2015, MNRAS, 449, 4246
Greig B., Mesinger A., 2017, Proc. IAU Symp. 12, Cambridge Univ. Press, Cambridge, p. 18
Greig B., Mesinger A., 2018, MNRAS, 477, 3217
Greig B. et al., 2021, MNRAS, 501, 1
Hermans J., Begy V., Louppe G., 2020a, Proceedings of the 37th International Conference on Machine Learning, Vol: 119, PMLR
Hermans J., Begy V., Louppe G., 2020b, in DaumeIII H., Singh A., eds, Proc. Machine Learning Research, Proc. 37th Int. Conf. Machine Learning, Vol. 119. PMLR, p. 4239,
Hermans J., Delaunoy A., Rozet F., Wehenkel A., Begy V., Louppe G., 2022, Conf. Machine Learning and the Phy. Sci. workshop, New Orleans Convention Center, New Orleans, USA
Jensen H., Laursen P., Mellema G., Iliev I. T., Sommer-Larsen J., Shapiro P. R., 2012, MNRAS, 428, 1366
Kapahtia A., Chingangbam P., Ghara R., Appleby S., Choudhury T. R., 2021, J. Cosmol. Astropart. Phys., 2021, 026
Kaplinghat M., Chu M., Haiman Z., Holder G. P., Knox L., Skordis C., 2003, ApJ, 583, 24
Kern N. S., Liu A., Parsons A. R., Mesinger A., Greig B., 2017, ApJ, 848, 23
Kolopanis M. et al., 2019, ApJ, 883, 133
Komatsu E. et al., 2011, ApJS, 192, 18
Koopmans L. et al., 2015, Proc. of Advancing Astrophysics with the Square Kilometre Array (AASKA14), Bourke Tyler L, Giardini Naxos, Italy, p. 1
Kuhlen M., Faucher-Giguère C.-A., 2012, MNRAS, 423, 862
Li W. et al., 2019, ApJ, 887, 141
Majumdar S., Pritchard J. R., Mondal R., Watkinson C. A., Bharadwaj S., Mellema G., 2018, MNRAS, 476, 4007
Mellema G., Iliev I. T., Pen U.-L., Shapiro P. R., 2006, MNRAS, 372, 679
Mellema G., Koopmans L., Shukla H., Datta K. K., Mesinger A., Majumdar S., 2015, Proc. of Advancing Astrophysics with the Square Kilometre Array (AASKA14), Bourke Tyler L, Giardini Naxos, Italy
Mertens F. G. et al., 2020, MNRAS, 493, 1662
Mesinger A., Furlanetto S., Cen R., 2010, MNRAS, 411, 955
Mesinger A., Greig B., Sobacchi E., 2016, MNRAS, 459, 2342
Miller B. K., Cole A., Forré P., Louppe G., Weniger C., 2021, 35th Conf. Neural Information Processing Systems.
Miller B. K., Cole A., Weniger C., Nattino F., Ku O., Grootes M. W., 2022, J. Open Source Softw., 7, 4205
Mineo S., Gilfanov M., Sunyaev R., 2012, MNRAS, 419, 2095
Mondal R. et al., 2020, MNRAS, 498, 4178
Muñoz J. B., 2023, MNRAS, 523, 2587
Paciga G. et al., 2013, MNRAS, 433, 639
Pacucci F., Mesinger A., Mineo S., Ferrara A., 2014, MNRAS, 443, 678
Papamakarios G., Sterratt D., Murray I., 2019, in Chaudhuri K., Sugiyama M., eds, Proc. Machine Learning Research, Vol. 89, Proc. Twenty-Second Int. Conf. Artificial Intelligence and Statistics. PMLR, Naha, Okinawa, Japan, p. 837,
Park J., Mesinger A., Greig B., Gillet N., 2019, MNRAS, 484, 933
Planck Collaboration 2016a, A&A, 594, A13
Planck Collaboration 2016b, A&A, 596, A108
Planck Collaboration 2020, A&A, 641, A6
Press W. H., Schechter P., 1974, ApJ, 187, 425
Pritchard J. R., Loeb A., 2012, Rep. Prog. Phys., 75, 086901
Qin W., Schutz K., Smith A., Garaldi E., Kannan R., Slatyer T. R., Vogelsberger M., 2022, Phys. Rev. D, 106,123506
Sazonov S., Khabibullin I., 2017, MNRAS, 466, 1019
Schmit C. J., Pritchard J. R., 2017, MNRAS, 475, 1213
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Shimabukuro H., Semelin B., 2017, MNRAS, 468, 3869
Shimabukuro H., Yoshiura S., Takahashi K., Yokoyama S., Ichiki K., 2016, MNRAS, 458, 3003
Sobacchi E., Mesinger A., 2014, MNRAS, 440, 1662
The HERA Collaboration 2022, Improved Constraints on the 21 cm EoR Power Spectrum and the X-Ray Heating of the IGM with HERA Phase I Observations.
Tiwari H., Shaw A. K., Majumdar S., Kamran M., Choudhury M., 2022, J. Cosmol. Astropart. Phys., 2022, 045
Toni T., Welch D., Strelkowa N., Ipsen A., Stumpf M. P., 2008, J. R. Soc. Interface, 6, 187
Watkinson C. A., Giri S. K., Ross H. E., Dixon K. L., Iliev I. T., Mellema G., Pritchard J. R., 2018, MNRAS, 482, 2653
Wouthuysen S. A., 1952, AJ, 57, 31
Zahn O., Mesinger A., McQuinn M., Trac H., Cen R., Hernquist L. E., 2011, MNRAS, 414, 727
Zel'dovich Y. B., 1970, A&A, 5, 84
Zhao X., Mao Y., Cheng C., Wandelt B. D., 2022a, ApJ, 926, 151
Zhao X., Mao Y., Wandelt B. D., 2022b, ApJ, 933, 236

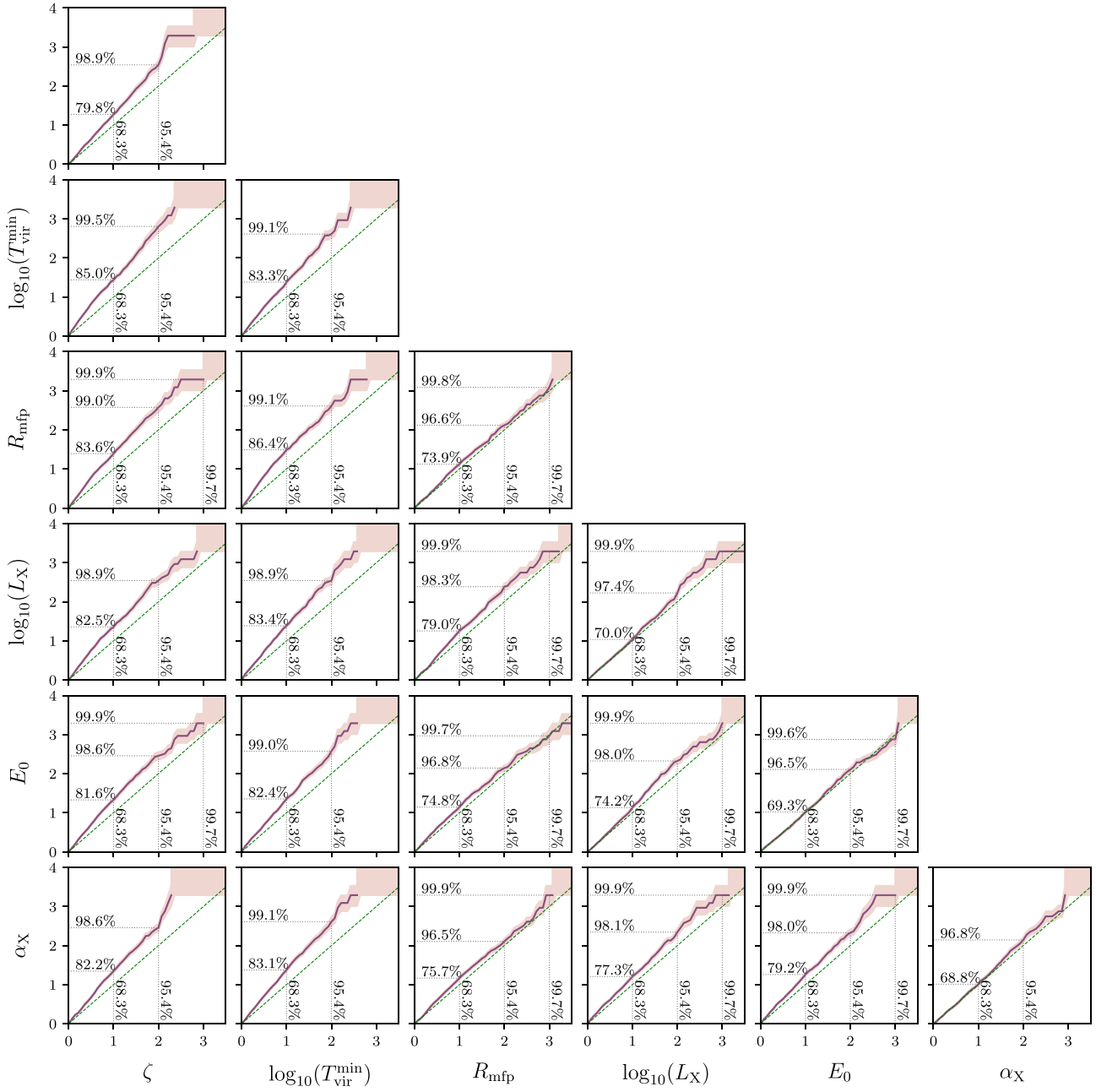

Figure A1. Empirical expected coverage probability $(1-\hat{\alpha})$ of the trained network as a function of confidence level $(1-\alpha)$ for all 1D and 2D marginal posteriors. In case of a perfect coverage, the purple line coincides with the green dashed line.

APPENDIX A: COVERAGE OF THE NETWORK

With the trained network $d_\phi(\boldsymbol{x}, \boldsymbol{\theta})$, we can quickly estimate the posterior $p(\boldsymbol{\theta}|\boldsymbol{x})$ for any mock observation $\boldsymbol{x}$. This allows us to test the statistical properties of the Bayesian inference. We evaluate the nominal and empirical expected coverage probabilities to check the consistency of the trained network. Given a set of n i.i.d. samples $(\boldsymbol{x}_i, \boldsymbol{\theta}_i^*) \sim p(\boldsymbol{x}, \boldsymbol{\theta})$, the empirical expected coverage probability of the $(1-\alpha)$ highest posterior density regions (HPDR) for the posterior estimator $\hat{p}(\boldsymbol{\theta}|\boldsymbol{x})$ is given as (Hermans et al. 2022)

$$1-\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left[\boldsymbol{\theta}_i^* \in \Theta_{\hat{p}(\boldsymbol{\theta}|\boldsymbol{x}_i)}(1-\alpha)\right] , \tag{A1}$$

where $\Theta_{\hat{p}(\boldsymbol{\theta}|\boldsymbol{x}_i)}(1-\alpha)$ function gives the $(1-\alpha)$ HPDR of $\hat{p}(\boldsymbol{\theta}|\boldsymbol{x})$ for the mock data $\boldsymbol{x}_i$ with the ground truth $\boldsymbol{\theta}_i^*$. We then compare it with the nominal expected coverage probability, which is equal to the confidence level $(1-\alpha)$. For an estimator with perfect coverage, the empirical coverage probability is equal to the nominal coverage probability, so when we randomly generate n samples $(\boldsymbol{x}_i, \boldsymbol{\theta}_i^*) \sim$

$p(\boldsymbol{x}, \boldsymbol{\theta})$, the ground truth $\boldsymbol{\theta}_i^*$ lies outside the $(1-\alpha)$ HPDR in α of the cases (Cole et al. 2022).

We re-parametrize α ($\hat{\alpha}$) in terms of a new variable z which is $1-\alpha/2$ ($1-\hat{\alpha}/2$) quantile of the standard normal distribution. This means that the (1, 2, 3)σ regions correspond to z = (1, 2, 3) with $(1-\alpha)$ = (0.6827, 0.9545, 0.9997). The uncertainties on the empirical expected coverage probability follow from the finite number of samples (n) and are estimated using the Jeffreys interval (Cole et al. 2022). In Fig. A1, we show the empirical expected coverage probability of the network as a function of confidence levels for all 1D and 2D marginal posteriors. We find that in all cases, they match to good precision.

APPENDIX B: COMPARISON WITH 21CMMC

In this section, we compare the posteriors obtained from MNRE with an MCMC sampling-based method, 21CMMC. As sampling our joint six-dimensional parameter space with the likelihood-based approach is computationally very demanding, we shrink our parameter space to two dimensions, including ζ and $\log_{10}(T_{\rm vir}^{\rm min})$. We keep the other parameters fixed to $\{R_{\rm mfp}, \log_{10}(L_{\rm X}), E_0, \alpha_{\rm X}\}$ = $\{15, 40.5, 0.5, 1\}$ and target redshifts z = 10, 9, and 8.

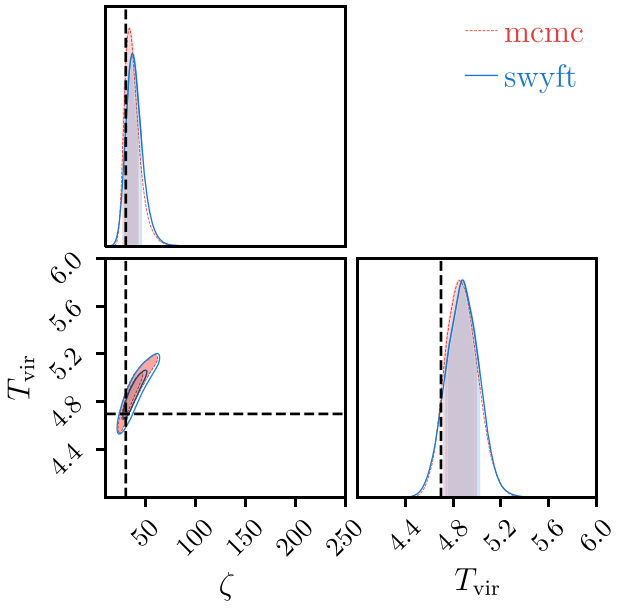

Figure B1. Comparison of the 1D and 2D marginal posteriors estimated from 21CMMC (red) with MNRE (blue) for the FAINT GALAXIES reionization model (ζ, $\log_{10}(T_{\rm vir}^{\rm min})$) = (30, 4.70).

To set up 21CMMC, we use 48 random walkers with 2000 iterations each, generating $\sim 10^5$ samples. On the other hand, the training data for MNRE consists of 10^4 simulations. The mock observation with (ζ, $\log_{10}(T_{\rm vir}^{\rm min})$) = (30, 4.70) is generated using a different realization of the density field from the one used in sampling. In Fig. B1, we show the posteriors obtained from 21CMMC (red) and MNRE (blue). The 1D and 2D marginal posteriors obtained from MNRE are in good agreement with 21CMMC at a significantly reduced computational cost. These results are consistent with the findings of Zhao et al. (2022b).

APPENDIX C: IMPACT OF INCLUDING MODELLING UNCERTAINTY

In this section, in addition to thermal noise, we consider an additional source of uncertainty due to the EoR modelling. We assume a constant multiplicative error of 10 per cent to take into account the errors in semi-numerical approximations. This is added in quadrature to the thermal noise uncertainty to get the 21-cm power spectrum uncertainty

$$\sigma(k_i, z) = \sqrt{\sigma_{\rm therm}^2(k_i, z) + \sigma_{\rm mod}^2(k_i, z)}\,. \tag{C1}$$

In Fig. C1, we show the recovered 1D and 2D marginal posteriors assuming a 10 per cent modelling uncertainty (red), and compare it with the constraints derived by excluding this error (green). We find that including the modelling uncertainty results in wider posteriors, which is consistent with Greig & Mesinger (2015). The inferred model parameters and the corresponding 16th and 84th percentiles for both scenarios are tabulated in Table C1. We note that this analysis does not require re-running any 21-cm signal simulations. The existing training data with the modified noise model given by equation (C1), which is sampled on-the-fly during the training of the network, can be re-used. This example demonstrates the flexibility and efficiency of our approach.

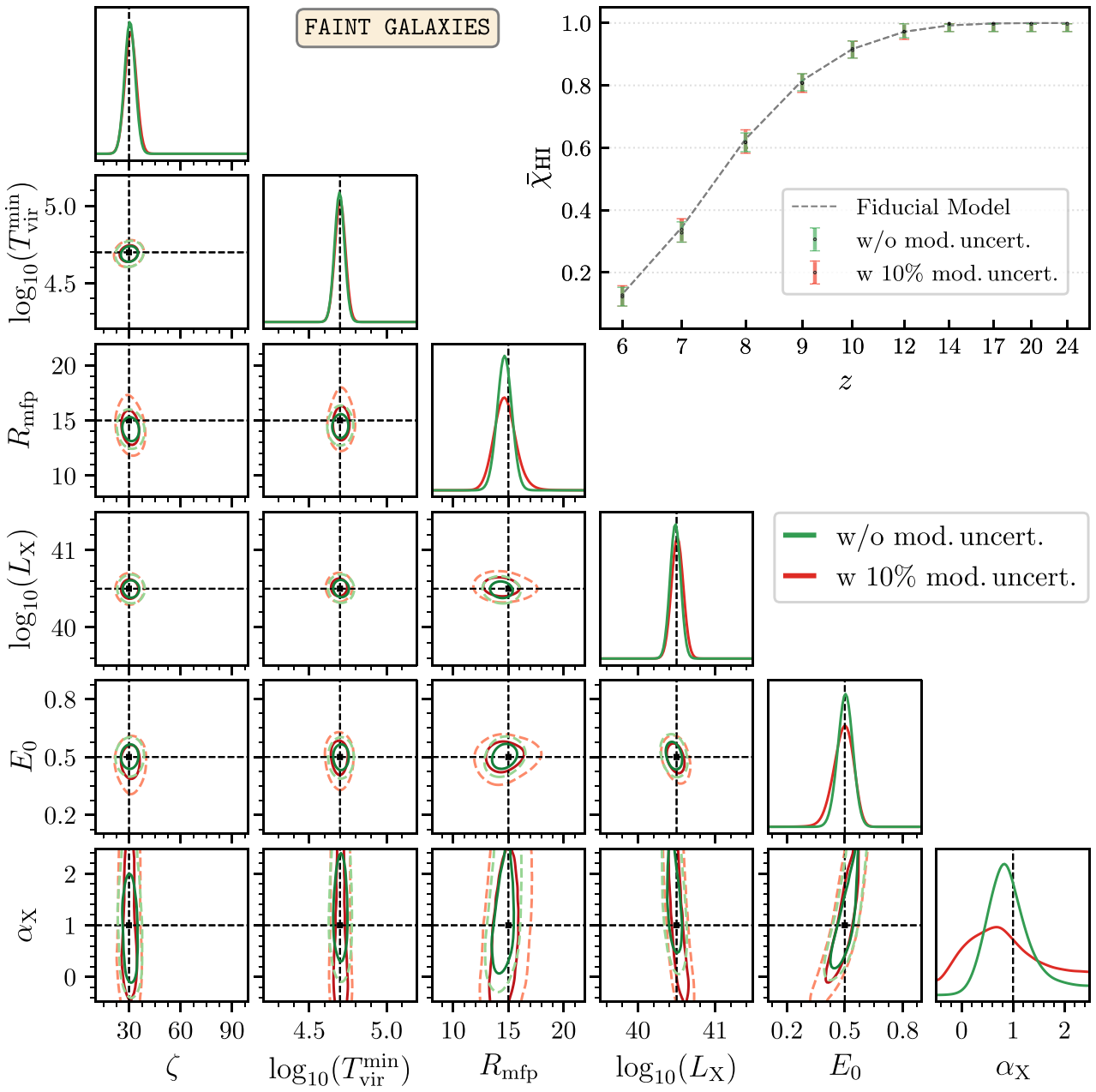

Figure C1. Recovered 1D and 2D marginals with (red) and without (green) including 10 per cent modelling uncertainty on the 21-cm power spectra. The dashed lines denote the input parameters $\{\zeta, \log_{10}(T_{\rm vir}^{\rm min}), R_{\rm mfp}, \log_{10}(L_{\rm X}), E_0, \alpha_{\rm X}\} = \{30, 4.70, 15, 40.5, 0.5, 1\}$. The inset plot shows the recovered reionization history.

Table C1. The inferred parameter values and the associated 16th and 84th percentiles for the posteriors shown in Fig. C1 with and without including 10 per cent modelling uncertainty.

Model	ζ	$\log_{10}(T_{\rm vir}^{\rm min})$	$R_{\rm mfp}$	$\log_{10}(L_{\rm X})$	E_0	$\alpha_{\rm X}$
w/o 10 per cent mod. uncert.	$30.25^{+2.70}_{-1.80}$	$4.70^{+0.03}_{-0.02}$	$14.65^{+0.56}_{-0.56}$	$40.49^{+0.04}_{-0.06}$	$0.50^{+0.03}_{-0.03}$	$0.84^{+0.39}_{-0.39}$
w 10 per cent mod. uncert.	$31.15^{+2.70}_{-2.70}$	$4.70^{+0.02}_{-0.03}$	$14.65^{+0.84}_{-0.84}$	$40.51^{+0.06}_{-0.06}$	$0.50^{+0.04}_{-0.05}$	$0.69^{+0.69}_{-0.78}$

APPENDIX D: IMPACT OF THE SIZE OF TRAINING SET ON THE POSTERIORS

In this section, we investigate the size of the training set needed to achieve the convergence for MNRE. Our default training set contains 2×10^4 ($n_{\rm samp}$) 21-cm power spectra samples. We re-train the neural ratio estimator with a subset of training set with $n_{\rm samp} = 10^4$ to estimate the posterior distribution of model parameters.

In Fig. D1, we present and compare the recovered 1D and 2D marginal posteriors generated from $n_{\rm samp} = 2 \times 10^4$ (green) and 10^4 (red) samples. The inset plot shows the 2σ constraints on reionization history. The inferred model parameters and the corresponding 16th and 84th percentiles for both scenarios are tabulated in Table D1. The posteriors on model parameters for both cases match to excellent precision, which indicates the convergence of MNRE. Therefore, $\sim 10^4$ simulations are sufficient to preserve accuracy in our SBI framework which makes it 3−10 times more computationally efficient than the classical methods of inference.

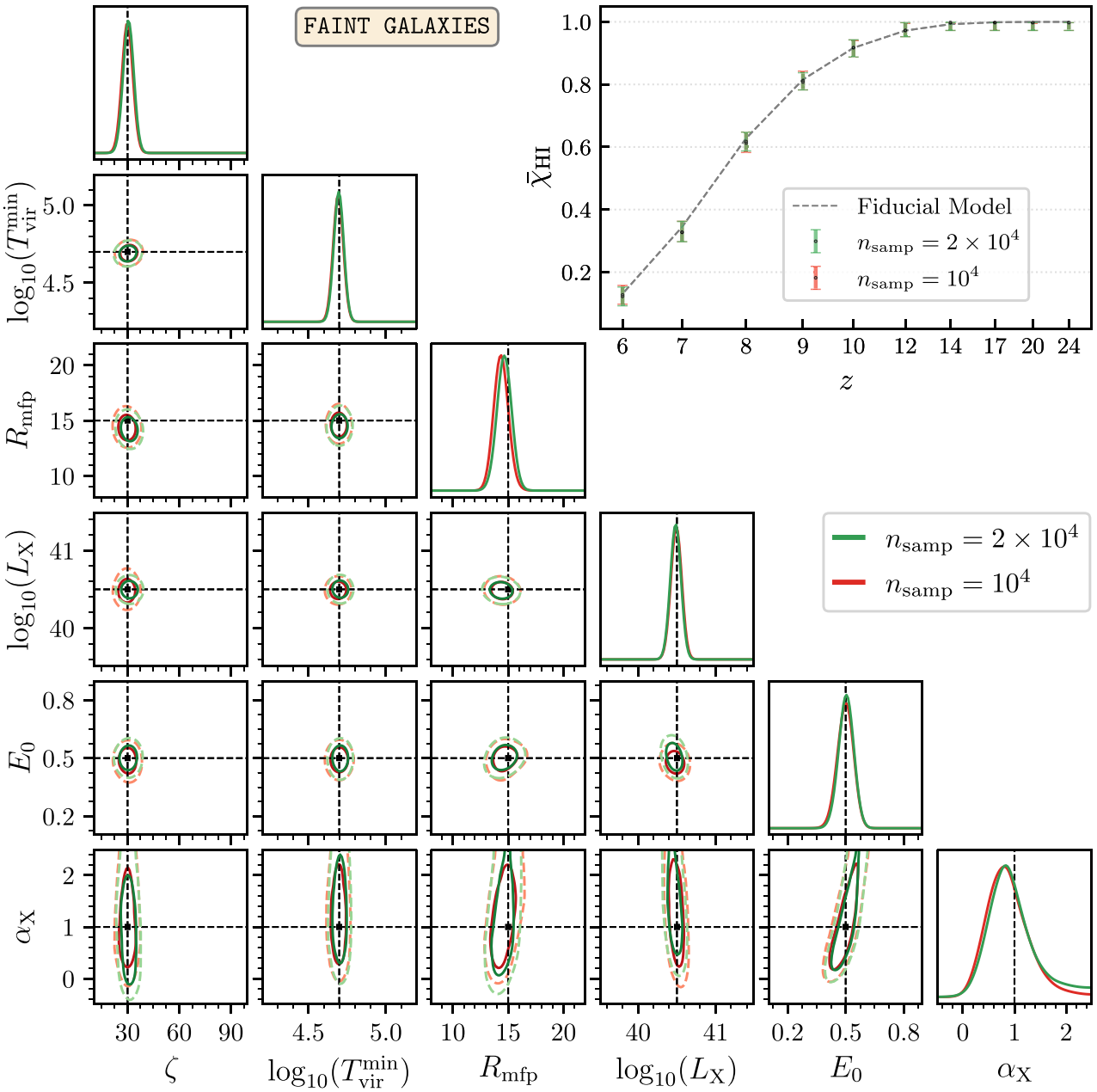

Figure D1. Recovered 1D and 2D marginals with the number of samples in the training data $n_{\rm samp} = 2 \times 10^4$ (green) and $n_{\rm samp} = 10^4$ (red). The dashed lines denote the input parameters $\{\zeta, \log_{10}(T_{\rm vir}^{\rm min}), R_{\rm mfp}, \log_{10}(L_{\rm X}), E_0, \alpha_{\rm X}\} = \{30, 4.70, 15, 40.5, 0.5, 1\}$. The inset plot shows the recovered reionization history.

Table D1. The inferred parameter values and the associated 16th and 84th percentiles for the posteriors shown in Fig. D1.

Model	ζ	$\log_{10}(T_{\rm vir}^{\rm min})$	$R_{\rm mfp}$	$\log_{10}(L_{\rm X})$	E_0	$\alpha_{\rm X}$
$n_{\rm samp} = 2 \times 10^4$	$30.25^{+2.70}_{-1.80}$	$4.70^{+0.03}_{-0.02}$	$14.65^{+0.56}_{-0.56}$	$40.49^{+0.04}_{-0.06}$	$0.50^{+0.03}_{-0.03}$	$0.84^{+0.39}_{-0.39}$
$n_{\rm samp} = 10^4$	$30.25^{+2.70}_{-2.70}$	$4.70^{+0.02}_{-0.03}$	$14.37^{+0.56}_{-0.56}$	$40.49^{+0.06}_{-0.04}$	$0.50^{+0.03}_{-0.04}$	$0.81^{+0.36}_{-0.36}$

This paper has been typeset from a TEX/LATEX file prepared by the author.