21cmEMUv3: a Hybrid Diffusion-LSTM Emulator for Joint Inference of Cosmic Dawn and Reionisation

Anonymous Authors[1]

Abstract

The cosmic dawn (CD) and epoch of reionisation (EoR) mark the first billion years of cosmic history, during which the first stars and galaxies formed and drove the final major phase transition of the Universe as their radiation ionised the neutral intergalactic medium (IGM), culminating in the present-day Universe. Over the past decade, a surge of observations has begun to illuminate these periods, with upcoming observations of the redshifted 21-cm line of neutral hydrogen — a direct tracer of the neutral IGM — expected to revolutionise our understanding of the first luminous sources. Robustly interpreting these observations with Bayesian inference requires repeatedly simulating the entire Universe across high-dimensional cosmological and astrophysical parameter spaces, making ML a critical tool for tractable inference, with emulation emerging as a powerful approach. We present 21cmEMUv3, an emulator trained on 21cmFASTv3 simulations modelling two galaxy populations, conditioned on eleven parameters (σ_8 plus ten astrophysical parameters), and designed to jointly predict seven summary observables including 21-cm power spectra, IGM thermal history, and UV luminosity functions. We emulate the 2D 21-cm PS via score-based diffusion and the remaining six summaries via long short-term memory (LSTM) networks, achieving sub-percent median accuracy across all outputs. We showcase the emulator by producing forecasts for the upcoming Square Kilometre Array (SKA), finding that SKA will enable percent-level constraints on the neutral hydrogen fraction at EoR midpoint, while further demonstrating that inference pipelines remain highly sensitive to model misspecification, an important methodological challenge for robust scientific inference with upcoming observations.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the AI4Physics Workshop at the International Conference on Machine Learning (AI4Physics@ICML 2026). Do not distribute.

1. Introduction

One of the greatest challenges in modern cosmology is understanding the cosmic dawn (CD) and the epoch of reionisation (EoR): the first billion years of cosmic history during which the first stars and galaxies formed. Over the past decade, significant progress has been made through a diverse set of complementary observational probes, including the Lyman-α forest (Fan et al., 2006; Becker et al., 2015; Bosman et al., 2018; D'Odorico et al., 2023), damping wings in quasar spectra (Bolton et al., 2011; Mortlock et al., 2011; Bañados et al., 2018; Wang et al., 2020; Yang et al., 2020), Lyman-α emission from galaxies (Ouchi et al., 2010; Clément et al., 2012; Konno et al., 2014; Drake et al., 2017; Hoag et al., 2019; Shibuya et al., 2019), large-scale cosmic microwave background (CMB) polarisation (Planck Collaboration et al., 2020; de Belsunce et al., 2021; Heinrich & Hu, 2021), and upper limits on the 21-cm power spectrum (Mertens et al., 2025; Nunhokee et al., 2025; The HERA Collaboration et al., 2025). Despite this progress, many open questions remain, and the redshifted 21-cm line of neutral hydrogen, as a direct three-dimensional tracer of the neutral IGM across cosmic time, is set to revolutionise our understanding with the upcoming Square Kilometre Array (SKA; Koopmans et al. 2015), offering an unprecedented tomographic view of the morphology and thermal state of the IGM for over half of the observable Universe.

Robustly interpreting these diverse observations requires cosmological simulations that self-consistently track the impact of galactic radiation fields on the IGM. Even the most efficient semi-numerical simulators (Davies et al., 2025; Santos et al., 2010; Hutter et al., 2021) typically require $\gtrsim$1 core hour per forward model, making full exploration of the high-dimensional parameter space of astrophysical and cosmological uncertainties computationally prohibitive. These challenges are further exacerbated by the rapid pace of new data releases, with each new dataset typically requiring a new inference.

As a result, there has been increasing interest in ML methods

that amortize the computational cost of Bayesian inference, that is, investing upfront in training a neural network so that subsequent inferences can be reused at negligible cost. Two broad approaches have emerged: emulation, which learns the mapping from parameters to simulation outputs, and simulation-based inference (SBI; Cranmer et al. 2020), which learns the posterior (Papamakarios & Murray, 2016), likelihood (Papamakarios et al., 2018), or likelihood-to-evidence ratio (Hermans et al., 2019) directly from simulated data.

SBI bypasses the need to explicitly define a likelihood, which is particularly advantageous when the likelihood is highly non-Gaussian. However, for the specific observables used in this work — the 21-cm PS, UV LFs, and Thomson optical depth τ_e — Gaussian likelihood approximations are well established, following HERA Collaboration et al. 2023. Gaussian approximations do require careful covariance modelling, and not all EoR observables share this property: Lyα statistics, for example, are known to carry non-Gaussian likelihoods. Prelogović & Mesinger 2023 characterise the regime in which Gaussianity holds for 21-cm PS statistics, finding it adequate at the signal-to-noise levels of current experiments; for these specific observables, the practical advantage of SBI is therefore reduced. Emulation, on the other hand, can be reused across multiple datasets without retraining, making it particularly well suited for keeping pace with frequent data releases, though it requires an explicit likelihood specification during inference, which requires careful specification in cases of non-Gaussian or strongly correlated data.

Emulation has a growing history in the CD/EoR community, with efforts progressively expanding from single observables and simple architectures (Kern et al., 2017; Shimabukuro & Semelin, 2017; Schmit & Pritchard, 2018; Ghara et al., 2020; Cohen et al., 2020; Bevins et al., 2021; Bye et al., 2022) to more sophisticated models (Prelogović et al., 2022; Dorigo Jones et al., 2024; Zhao et al., 2023; Moriwaki et al., 2026; Diao & Mao, 2025). Despite this progress, most existing emulators are designed to predict a single observable. Yet joint inference combining multiple complementary observables has been shown to significantly tighten constraints on astrophysical and cosmological parameters (HERA Collaboration et al., 2022; 2023; The HERA Collaboration et al., 2025; Nunhokee et al., 2025), motivating the development of emulators capable of jointly predicting multiple observables within a unified framework (Breitman et al., 2024; Cang et al., 2024). In this work, we present `21cmEMUv3`, the third addition to the `21cmEMU` emulator suite, providing a significant expansion in both the physical model and the emulation architecture.

We describe the machine learning architecture of `21cmEMUv3` in Section 3 and quantify its performance on a test set in Section 4. In Section 5, we showcase the emulator by producing parameter forecasts for the upcoming SKA, and further demonstrate that state-of-the-art inference pipelines remain highly sensitive to model misspecification. We conclude in Section 6.

2. Simulated database

Our emulator is trained on simulations from `21cmFASTv3` (Mesinger & Furlanetto, 2007; Mesinger et al., 2011; Murray et al., 2020), a publicly available semi-numerical simulator of cosmic radiation fields, including the 21-cm signal. For a given set of astrophysical and cosmological parameters, `21cmFASTv3` evolves the matter density field, assigns galaxy properties to dark matter halos via empirical relations, computes the resulting radiation fields, and outputs 3D lightcones of the 21-cm brightness temperature across cosmic time. These lightcones can then be compressed into summary statistics for direct comparison with observations.

We use `21cmFASTv3` to model two distinct populations of early galaxies: atomically-cooling galaxies (ACGs), which form in more massive halos and are thought to host primarily Population II stars, and molecularly-cooling galaxies (MCGs), which form in smaller, lower-mass halos and are thought to host the first generation of metal-free Population III stars. These two populations play complementary roles in shaping the thermal and ionisation history of the IGM: ACGs dominate the ionizing photon budget during the EoR, while MCGs are thought to drive the earliest episodes of X-ray heating and star formation during the CD. Together, their properties are described by eleven free parameters: seven for ACGs, three for MCGs, and one cosmological parameter σ_8. In Table 1, we list the `21cmEMUv3` input parameters and their limiting values in the database used to train the emulator. Interested readers are referred to Appendix A for a detailed description of the galaxy model and its parameters.

We create a database of $\sim$47k simulations, each consisting of a 400 cMpc lightcone spanning redshift $z \sim 5 - 35$ at a resolution of 1.5 cMpc/cell. Rather than sampling parameters uniformly, we populate the database by running a Bayesian inference whose prior corresponds to the limits shown in Table 1, ensuring denser sampling in regions of parameter space consistent with existing observations. The inference is informed by three datasets: (i) Thomson scattering optical depth to the CMB (Planck Collaboration et al., 2020; Qin et al., 2020) using a Gaussian likelihood around $\tau_e = 0.0569^{+0.0081}_{-0.0086}$; (ii) UV luminosity functions (Bouwens et al., 2015; 2016; Oesch et al., 2018) using a Gaussian likelihood at redshifts $z = 6, 7, 8, 10$; (iii) Lyα forest effective optical depth (Bosman et al., 2018; D'Odorico et al., 2023) using a Gaussian likelihood comparison to the measured CDF (Bosman et al., 2018).

Table 1. In 21cmEMUv3, we model ACGs and MCGs by varying a total of eleven parameters: seven ACG astrophysical parameters, three MCG astrophysical parameters, and cosmological parameter σ_8. The second column shows the upper and lower bounds of each parameter in the emulator training database [1].

Parameter	Limits	Description
$\log_{10} f_{*,10}$	$[-2, -0.5]$	SHMR normalisation for ACGs
$\alpha_{*,\rm ACG}$	$[0, 1]$	power-law index of the SHMR for ACGs
t_*	$[0.01, 1]$	characteristic SF timescale
$\log_{10} f_{\rm esc,10}$	$[-3, 0]$	normalisation of the ionizing escape fraction to halo mass relation for ACGs
$\alpha_{\rm esc}$	$[-1, 1]$	power-law index of the ionizing escape fraction to halo mass relation
$\log_{10} f_{*,7}$	$[-4, -1]$	SHMR normalisation for MCGs
$\log_{10} f_{\rm esc,7}$	$[-3, -1]$	normalisation of the ionizing escape fraction to halo mass relation for MCGs
$\log_{10} L^{\rm ACG}_{\rm X<2keV}/{\rm SFR}$	$[38, 43]\ \frac{\rm erg\ s^{-1}}{M_\odot\ {\rm yr}^{-1}}$	soft-band X-ray luminosity per unit SFR for ACGs
$\log_{10} L^{\rm MCG}_{\rm X<2keV}/{\rm SFR}$	$[39, 44]\ \frac{\rm erg\ s^{-1}}{M_\odot\ {\rm yr}^{-1}}$	soft-band X-ray luminosity per unit SFR for MCGs
E_0	$[100, 1500]$ keV	minimum energy of X-ray photons capable of escaping their host galaxy
σ_8	$[0.75, 0.85]$	matter power spectrum normalisation

2.1. Simulations and summary observables

Current CD/EoR instruments are not sensitive enough to precisely measure high-dimensional fields such as the full 3D 21-cm brightness temperature lightcone. Instead, they observe compressed summaries such as the 21-cm PS and UV LFs. As such, 21cmEMUv3 is designed to predict the following seven summary observables directly:

1. $\overline{\boldsymbol{x}}_{\rm HI}(\boldsymbol{z})$ — volume-averaged (global) neutral fraction of hydrogen and helium as a function of redshift (i.e. the EoR history).

2. $\overline{\boldsymbol{T}}_{\rm S}(\boldsymbol{z})$ — the mean neutral IGM spin temperature as a function of redshift.

3. $\overline{\boldsymbol{\delta T}}_{\rm b}(\boldsymbol{z})$ — global 21-cm brightness temperature (e.g., Madau et al. 1997; Furlanetto 2006; Pritchard & Loeb 2012):

$$\delta T_{\rm b}(\boldsymbol{x}, z) = \frac{T_{\rm S} - T_{\rm R}}{1+z}(1 - e^{\tau_{21}}) \tag{1}$$
$$\approx 27\, x_{\rm HI}(1+\delta_b)\left(\frac{\Omega_b h^2}{0.023}\right)\left(\frac{0.15}{\Omega_m h^2}\frac{1+z}{10}\right)^{1/2} \times \left(\frac{T_{\rm S} - T_{\rm R}}{T_{\rm S}}\right)\left[\frac{\partial_r v_r}{(1+z)H(z)}\right],$$

where τ_{21} is the 21-cm optical depth of the intervening gas, $\delta_b \equiv \rho/\bar{\rho} - 1$ is the baryon overdensity, with ρ being the baryon density, and $T_{\rm S}$ and $T_{\rm R}$ are the spin and background temperatures, respectively. We assume throughout that $T_{\rm R} = T_{\rm CMB}$, i.e. the radio background is provided by the CMB. We note that 21cmFAST computes the brightness temperature at each cell location, $\mathbf{x}$, using the exact expression in the first line of the equation above; the second line is a Taylor expansion in the limit of $\tau_{21} \ll 1$ that provides physical intuition.

4. $\boldsymbol{\Delta^2_{21}(k_\perp, k_\parallel, z)}$ — cylindrically-averaged 21-cm power spectrum (PS) as a function of sky-plane mode $k_\perp$ and line-of-sight (LOS) mode $k_\parallel$: $\Delta^2_{21}(k_\perp, k_\parallel, z)\,[{\rm mK}^2] \equiv k^3/(2\pi^2)\langle \tilde{T}_b \tilde{T}_b^* \rangle$, where $k = \sqrt{k_\perp^2 + k_\parallel^2}$, and $\tilde{T}_b(\boldsymbol{k}, z)$ is the Fourier dual of the brightness temperature from eq. (1). Note that in this work we calculate all PS with tuesday[2] (Breitman et al., 2025) with the interpolation feature which provides a more accurate PS. When interpolating, we obtain an estimate of the PS at the exact desired wavemode as opposed to the usual approach where an average is taken over a bin of a certain width. This usual approach may lead to biases as the signal evolves over the width of the bin.

5. $\boldsymbol{\Delta^2_{21}(k, z)}$ — spherically-averaged 21-cm PS as a function of wavemode k [3].

6. $\boldsymbol{\phi(\mathrm{M}_{1500}, z)}$ — the non-ionizing UV luminosity function (UV LF), defined as the number density of galaxies per UV magnitude, M_{1500}, as a function of redshift. The $\sim$1500 Å rest frame luminosity is assumed to be proportional to the SFR: $L_{\rm UV} = \dot{M}_*(M_h, z)/\mathcal{K}_{\rm UV}$, where $\mathcal{K}_{\rm UV} = 1.15 \cdot 10^{-28} {\rm M}_\odot {\rm yr}^{-1}$ Hz s erg^{-1} assumes a Salpeter initial mass function (e.g., Madau & Dickinson 2014; Sun & Furlanetto 2016). The UV luminosity is related to the AB magnitude using (Oke & Gunn, 1983): $\log\left(\frac{L_{\rm UV}}{\rm erg\ s^{-1}\ Hz^{-1}}\right) = 0.4 \times (51.63 - M_{\rm UV})$.

[2] https://github.com/21cmfast/tuesday

[3] We emulate the 1D PS separately rather than averaging over the cylindrical 21-cm PS to minimise the uncertainty and speed up the emulation. Averaging the 2D PS uses score-based diffusion and is therefore slower. Since the 2D PS is a much more complex quantity than the 1D PS, averaging it down to 1D yields a larger error than emulating the 1D PS directly. Users may use either method depending on the desired application.

7. $\boldsymbol{\tau_e}$ — the CMB optical depth: $\tau_e = \sigma_T \int_0^{z_{\rm LSS}} dz \left|\frac{cdt}{dz}\right| n_e$, where σ_T is the Thomson scattering cross section and n_e is the electron number density calculated assuming hydrogen and helium are singly ionized at a fraction $(1 - \overline{x}_{\rm HI})$ and that helium is doubly ionized at $z < 3$.

3. 21cmEMUv3 architecture and training

Figure 1 shows the architecture of `21cmEMUv3`. It consists of two independent components: a score-based diffusion model for the cylindrical 21-cm PS, and an LSTM-based network for the remaining six summary observables.

3.1. Generative score-based diffusion emulator of the cylindrical 21-cm PS

State-of-the-art CD/EoR simulations of the 21 cm signal typically span volumes that are orders of magnitude smaller than those probed by observations, primarily because such simulations become computationally impractical as their size increases. As a result, the number and overlap of Fourier modes common to both simulation and observation is limited, particularly in the cylindrical (2D) k-space that is natural for 21 cm interferometry. This mismatch introduces a key challenge for interpreting upcoming 21-cm observations at larger scales: sample variance, the deviation of the simulated realisation from the true population mean due to the finite simulation volume. Following Breitman et al. 2025, we treat sample variance as mildly non-Gaussian noise (Mondal et al., 2015; Shaw et al., 2019) added onto the target mean PS, and use score-based diffusion (Song et al., 2020; Ho et al., 2020) to marginalise over it. Conditioning the reverse SDE on the eleven input parameters θ allows us to generate new mean PS samples for any parameter combination. We adopt the variance-preserving (VP) SDE (Ho et al., 2020) and a U-Net (Ronneberger et al., 2015) as the score network, finding this combination outperforms the subVP and variance-exploding (VE) alternatives. Since we plan to integrate this emulator into an inference pipeline, we solve the reverse SDE via the probability-flow ODE (Song et al., 2020), emulating a mean 2D PS averaged over 200 samples in ∼2 s on a GPU. The network is trained on 44k 2D PS at 32 redshifts logarithmically spaced between $z = 5.5$–29, for ∼500 epochs using the `AdamW` (Loshchilov & Hutter, 2017) optimiser with an initial learning rate of 10^{-4}, a cosine annealing schedule with $\eta_{\rm min} = 10^{-6}$, and exponential moving average (EMA) checkpointing. Since each simulation uses different initial conditions, we find that `21cmEMUv3` naturally learns to marginalise over sample variance, producing PS predictions closer to the IC-averaged mean than a single realisation.

3.2. LSTM emulator for timeseries data

We adopt LSTMs following Dorigo Jones et al. 2024, who found them well-suited for emulating the global 21-cm signal, and building on the earlier finding of Prelogović et al. 2022 that RNNs perform well on 21-cm lightcone regression. The 2D outputs (1D PS and UV LFs) use an LSTM-based head designed to capture correlations across both redshift and wavenumber/magnitude axes, which we find outperforms both a standard LSTM head and a CNN head for these outputs. The scalar Thomson optical depth is emulated with a simple feed-forward network. All LSTM branches use two layers, as additional layers do not improve performance, consistent with Dorigo Jones et al. 2024.

Training this multi-output network presents a multi-task learning challenge, as the PS is significantly more complex than the other summaries and benefits from being trained jointly rather than in isolation. We explored several strategies including gradient surgery methods such as CAGrad (Liu et al., 2021) and various loss weighting schemes, finding only marginal improvements over a simpler baseline. Moreover, we find that weight decay is catastrophic for LSTM performance and do not use it. We therefore adopt a straightforward training procedure: the `Adam` optimiser (Kingma & Ba, 2017) with a `ReduceLROnPlateau` scheduler, decaying from 10^{-3} to 10^{-7}, with EMA checkpointing. Training converges by ∼400 epochs on ∼45k training samples. Our code is publicly available[4].

4. 21cmEMUv3 performance

We assess the performance of `21cmEMUv3` by evaluating it on a test set and calculating the error using the following metrics:

1. absolute difference (Abs Diff):

$$\text{Abs Diff} \equiv |y_{\rm true} - y_{\rm pred}| \tag{2}$$

2. fractional error (FE):

$$\text{FE}(\%) = \frac{\text{Abs Diff}}{\max(|y_{\rm true}|, y_{\rm floor})} \times 100, \tag{3}$$

where the floor is set to $y_{\rm floor} = 10^{-2}$ mK^2 for the 21-cm PS in linear space[5], and neutral fraction, $y_{\rm floor} = 5$ mK for the global 21-cm signal. No floor is applied to other summaries.

In Figure 2, we plot true (solid) and emulated (dashed) outputs for 10 random test set samples for the 1D 21-cm PS on the left, and show one random example for the 2D PS on

[4] https://github.com/21cmfast/21cmEMU

[5] The floor for the FE calculation on $\log_{10}$PS is 1.0.

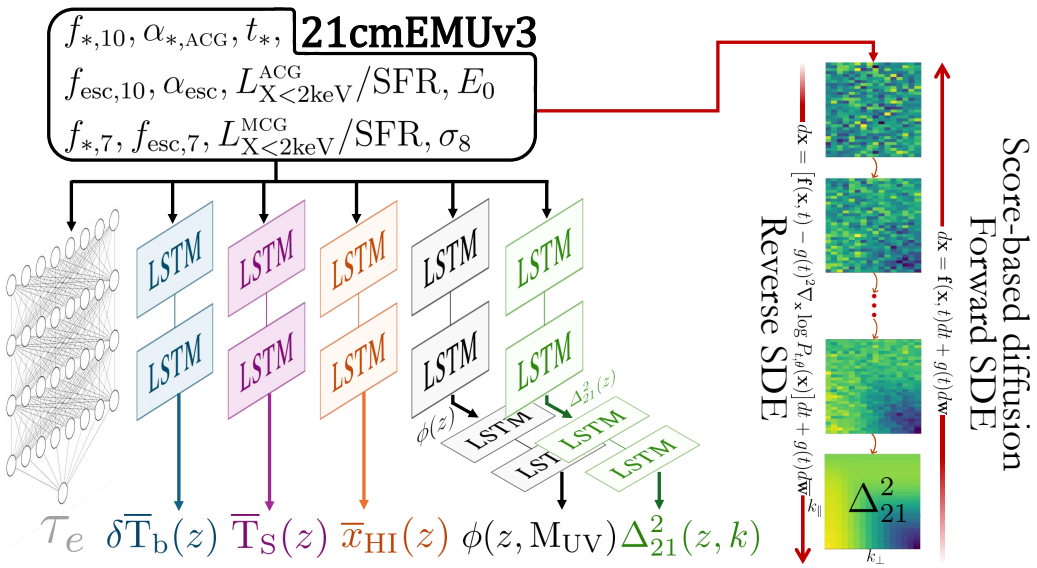

Figure 1. Diagram of the architecture of `21cmEMUv3`. The input consists of a total of eleven parameters: σ_8 and ten astrophysical parameters, three of which affect Pop III stars. `21cmEMUv3` consists of two networks: timeseries LSTM network on the left and the score-based diffusion network for the 2D 21-cm PS on the right. The network on the left emulates the globally-averaged 21-cm signal, spin temperature, and neutral fraction all as a function of redshift z, as well as UV LFs at seven redshifts and spherically-averaged PS with 2D LSTM networks. This network also includes a branch with a feed-forward neural network that emulates the Thomson scattering optical depth. The eleven parameters along with the desired redshift are also passed to the score-based diffusion network that produces one 2D 21-cm PS in about 2 seconds on a GPU.

the right. The complete figure showing true and emulated across all summaries is in Appendix B in Figure 5. The 10 samples are identical across panels in Figure 5. True and emulated curves are typically indistinguishable by eye, consistent with the sub-percent accuracy reported in Table 2. In the 1D PS panel on the left in Figure 2, we additionally show individual IC realisations (thin solid lines) alongside the mean (thick solid line), confirming that the emulated 1D PS is closer to the mean than any single realisation. In the right panel of Figure 2, we show one 2D 21-cm PS realisation on the top left and the corresponding emulated output on the top right. The bottom row then compares the top row to the mean PS estimated from averaging 76 2D PS realisations, highlighting that `21cmEMUv3` emulates the 2D 21-cm PS mean with sub-percent accuracy across most scales, while a single realisation may deviate by over $\sim$10%.

For both the 2D and 1D PS, the median FE is $\sim$7% and $\sim$5% lower when evaluated against the mean rather than individual PS realisations, consistent with the expectation that sample variance is larger for the 2D PS, which averaged over fewer Fourier modes than the spherically-averaged 1D PS, thus confirming that both the score-based diffusion model and the LSTM naturally learn to predict the mean PS. The remaining summaries are all emulated to sub-percent median accuracy, matching or exceeding `21cmEMUv1` despite a training set nearly two orders of magnitude smaller, a higher-dimensional input space, and the added challenge of sample variance.

We summarise performance in Table 2, where statistics are evaluated over 4k test samples for all summaries except the 1D and 2D PS, for which we use the 90-parameter mean PS test set. The median (68% confidence limit (CL)) error is 0.45% (0.81%) for the log 2D PS[6] and 0.73% (2.57%) for the log 1D PS[7]. This is particularly noteworthy in comparison to `21cmEMUv1`, which achieved a median (68% CL) 1D PS error of 1.4% (3.8%) despite being trained on a database nearly two orders of magnitude larger, on a lower-dimensional summary statistic, and on a fixed IC realisation with no sample variance. `21cmEMUv3` matches or exceeds this performance across all summaries despite these additional challenges.

5. Applications to inference

We apply `21cmEMUv3` to two inference problems using the `21cmMC` driver (Greig & Mesinger, 2015) with the `nautilus` sampler (Lange, 2023), which we find provides a good balance between posterior accuracy and efficiency. In both cases, we use the 1D 21-cm PS likelihood, which allows inference to be completed in about an hour on a GPU, highlighting the computational advantage of ML-based emu-

[6]The median (68% CL) error on the linear 21-cm 2D PS is 1.74% (2.05%).

[7]The median (68% CL) error on the linear 21-cm 1D PS is 2.17% (14.09%).

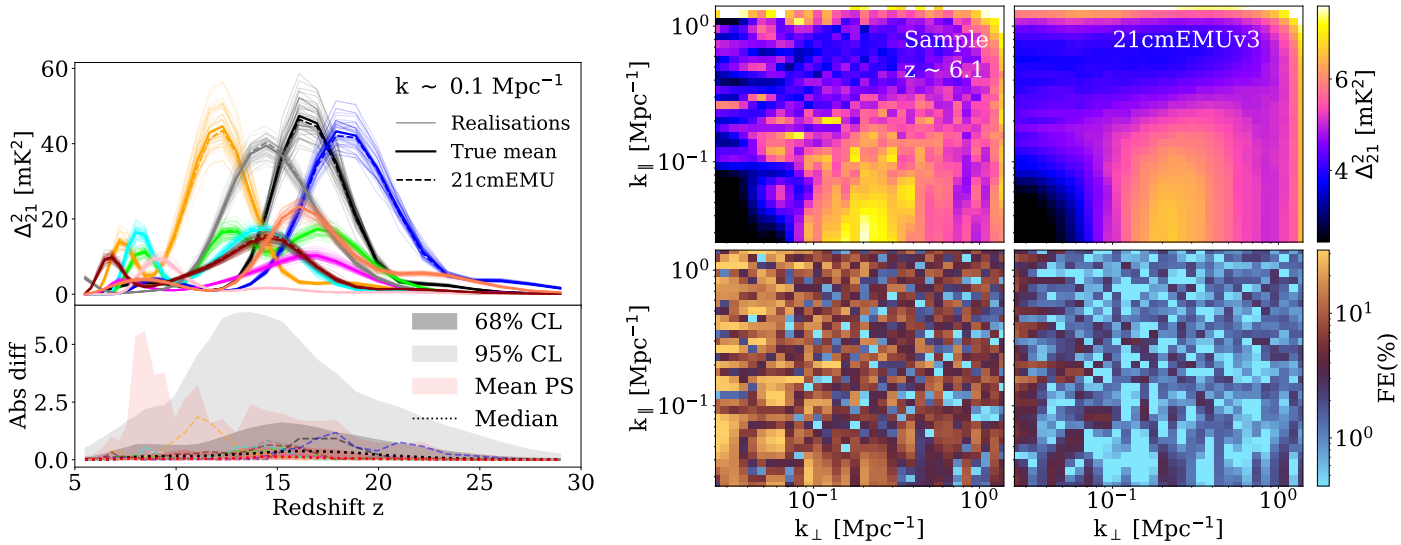

Figure 2. Left panel: True (solid) and emulated (dashed) outputs from 21cmEMUv3 for ten random test set samples of the 1D 21-cm PS at $k = 0.1$ Mpc^{-1}. The red shaded regions for the PS are evaluated against PS means rather than individual realisations. The full plot including the other summaries is in the appendix in Figure 5. Right panel, top row: a cylindrical 21-cm PS realisation from the test set (left) and the corresponding 21cmEMUv3 output (right). Right panel, bottom row: fractional error comparing each PS in the top row against the mean PS estimated by averaging 76 IC realisations at fixed astrophysical and cosmological parameters. The emulator produces a 2D PS much closer to the mean than the single realisation.

Table 2. Performance of the emulator over the test set for all summaries. Note that the test set for the PS consists of 90 mean PS, while for all other summaries, the test set contains 4k samples.

Summary	Median FE (%)	68% CL (%)
$\log_{10}\Delta^2_{21}(k_\perp, k_\parallel)$	0.45	0.81
$\log_{10}\Delta^2_{21}(k)$	0.73	2.57
$\overline{T}_{\rm b}$	0.46	1.49
$\log_{10}\overline{T}_{\rm S}$	0.47	3.48
$\overline{x}_{\rm HI}$	0.02	0.29
$\tau_{\rm e}$	0.29	0.47
$\log\phi$	0.21	1.33

lation. We also check that 21cmEMUv3 is able to reproduce previous results with upper limits on the 21-cm PS from the HERA instrument (HERA Collaboration et al., 2023), see Appendix C.

5.1. SKA forecast

The Square Kilometre Array Observatory (SKAO)[8] is an upcoming radio telescope currently under construction, designed to be the world's largest radio observatory. Its low-frequency component, SKA-Low, is being built in Australia and will observe the redshifted 21-cm line across redshifts $z \sim 5.5$–27, directly probing the CD and EoR. With its unprecedented sensitivity and collecting area, SKA-Low is expected to not only deliver high-significance detections of the 21-cm PS but to produce the first tomographic images of the neutral IGM during the first billion years of cosmic history: a transformative step in our understanding of the first galaxies. Construction is proceeding in staged Array Assembly (AA) increments; here we consider two layouts representative of different deployment stages: AA*, comprising 307 stations, and the fully deployed AA4 with all 512 stations, allowing us to assess how constraining power scales with array size.

To simulate a realistic observation, we select a parameter set from the emulator test set as our fiducial mock signal and use 21cmSense (Murray et al., 2024), modified to interface with the SKA antenna configuration package ska-ost-array-config[9], to forecast thermal noise sensitivities for both layouts. The SKA forecasting pipeline is publicly available in the 21cmSense documentation[10]. The mock observation and corresponding sensitivities are shown in the left panel of Figure 3 at $k \sim 0.48$ Mpc^{-1}, the most sensitive wavemode in the forecast, where the 21-cm PS is detected to a significance of over 10σ at $z \sim 6.5$–10.

We adopt a joint likelihood combining four complementary CD/EoR observables, following HERA Collaboration et al. 2023:

1. UV luminosity functions (Bouwens et al., 2015; 2016;

[8] https://www.skao.int/en

[9] https://gitlab.com/ska-telescope/ost/ska-ost-array-config

[10] https://21cmsense.readthedocs.io/en/latest/tutorials/SKA_forecast.html

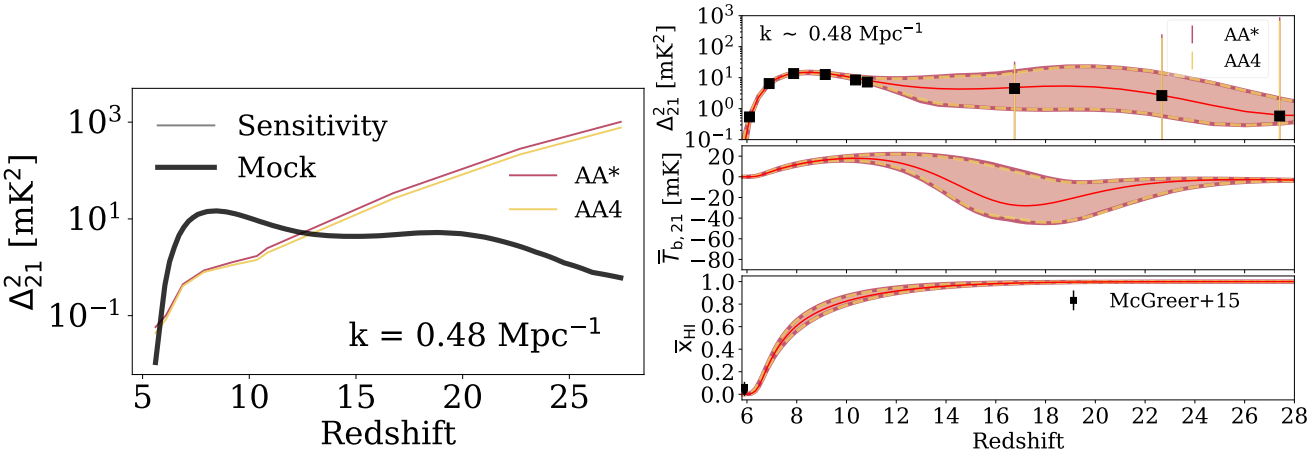

Figure 3. Left panel: Mock SKA observation of the 21-cm 1D PS at wavemode $k \sim 0.48$ Mpc^{-1} (black solid), together with the corresponding thermal noise sensitivity for the AA4 (yellow) and AA* (purple) antenna layouts simulated with `21cmSense`. Right panel: Marginalised 95% credible intervals on the 1D 21-cm PS (top), global 21-cm signal (middle), and neutral fraction (bottom) as a function of redshift, for the AA4 and AA* layouts. The ground truth from the mock is shown in red in each panel.

Oesch et al., 2018) with a Gaussian likelihood;

2. Thomson scattering optical depth to the CMB (Planck Collaboration et al., 2020; Qin et al., 2020) which uses a Gaussian likelihood around $\tau_e = 0.0569^{+0.0081}_{-0.0086}$;

3. upper limit $\overline{x}_{\rm HI} < 0.06 \pm 0.05$ on the EoR history at $z = 5.9$ obtained with the Lyα dark fraction method (McGreer et al., 2015). The likelihood function is unity at $\overline{x}_{\rm HI}(z = 5.9) < 0.06$, decreasing as a one-sided Gaussian for higher neutral fraction values;

4. SKA 21-cm PS mock observation assuming a Gaussian likelihood with a diagonal, parameter-independent covariance whose entries are the thermal noise and cosmic variances computed by `21cmSense` for each (k, z) bin.

We use `21cmEMUv3` with `nautilus` to perform 11-parameter Bayesian inference on the simulated mock 21-cm PS observation with a physically motivated prior on the astrophysical parameters. The resulting posteriors on derived summaries are shown in the right panel of Figure 3, and the full corner plot is provided in Appendix D.

The EoR history is accurately recovered with a 95% credible interval (CI) of width $\Delta\bar{x}_{\rm HI} \lesssim 0.09$ around the midpoint of reionisation. The global 21-cm signal is likewise well recovered. The Thomson scattering optical depth τ_e is constrained to $\sim$6 times the precision of the Planck18 measurement (Planck Collaboration et al., 2020).

5.2. Model misspecification

Model misspecification occurs when the forward model used for inference does not accurately represent the true data-generating process, analogous to distribution shift in ML, where a well-trained model performs poorly when deployed on data from a different distribution. In the context of 21-cm cosmology, misspecification arises naturally since all forward models are necessarily approximations of a complex underlying reality. This is distinct from, but complementary to, Bayesian model selection: while model selection uses the evidence to identify the best model among a set of candidates, it still assumes that one candidate is correct. Model misspecification analysis asks the harder question: what if none of the candidates is correct? A model can win a model comparison yet still produce significantly biased posteriors. As observations become more precise, previously negligible modelling assumptions can lead to such failures, even when standard diagnostics such as posterior predictive checks appear satisfactory.

To demonstrate this, we repeat the SKA inference from the previous section but replace `21cmEMUv3` with `21cmEMUv1` (Breitman et al., 2024) as the forward model. `21cmEMUv1` models only a single population of ACGs, omitting the MCG population present in `21cmEMUv3`. Of its nine astrophysical parameters, seven are shared with `21cmEMUv3`; the remaining two differ in their treatment: the X-ray spectral index is a free parameter in `21cmEMUv1` but fixed to $\alpha_X = 1$ in `21cmEMUv3`, while the minimum halo mass for star formation is a free parameter in `21cmEMUv1` but computed dynamically from feedback processes in `21cmEMUv3`.

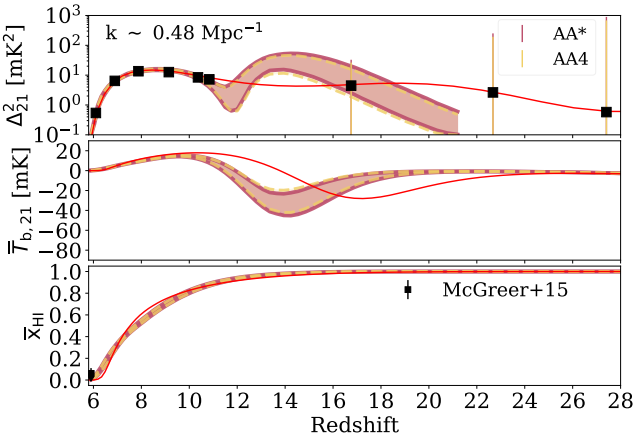

Figure 4. Marginalised 95% CI on the 1D 21-cm PS (top), global 21-cm signal (middle), and neutral fraction (bottom) as a function of redshift for the two SKA layouts, AA* and AA4. Note that the PS posterior appears cut short because 21cmEMUv1 only produces PS up to $z \sim 21$. The ground truth is shown in red in each panel.

Despite modelling different astrophysics, Figure 4 shows that inference with 21cmEMUv1 can still broadly reproduce the true summaries of the 21cmEMUv3 mock (red). However, the full corner plot in Figure 8 in Appendix D reveals that most parameters are significantly biased. This is a particularly insidious form of misspecification: the misspecified model successfully fits the observed 21-cm PS at the data points thus satisfying the likelihood (black squares in Figure 4, top panel), yet predicts very different behaviour between them ($z \sim 12$–16 in Figure 4, top panel). Such behaviour may lead to inferring wrong physics from observations. Furthermore, the derived global 21-cm signal is significantly biased despite the agreement of the 1D PS with the mock (Figure 4, middle panel). The neutral fraction, on the other hand, is recovered well with a 95% CI of width $\Delta\bar{x}_{\rm HI} \lesssim 0.03$ around the midpoint of reionisation since it is less sensitive to the presence or absence of the additional MCG population (Figure 4, bottom panel). The Thomson scattering optical depth τ_e is constrained to $\sim$15 times the precision of the Planck18 measurement (Planck Collaboration et al., 2020)[11]. Crucially, as shown in Figure 8 in Appendix D, the recovered astrophysical parameters are so severely biased that the ground truth lies outside the posterior range entirely for several parameters. This illustrates that successfully fitting the observed data is not sufficient to guarantee unbiased parameter inference when the forward model is misspecified.

[11]The tighter τ_e constraint obtained with 21cmEMUv1 is not a consequence of model misspecification but rather reflects the simpler parameter space of 21cmEMUv1, which has nine rather than eleven parameters and does not include MCGs. We verify this by running the same inference on a 21cmEMUv1 mock and find similarly tight constraints with correct parameter recovery.

These results caution that the choice of forward model can have a decisive impact on scientific conclusions, even when the inferred model appears consistent with observations. We note that adopting SBI rather than a parametric Gaussian likelihood would not resolve this issue: a misspecified forward model produces biased posteriors regardless of whether the likelihood is parametric or learned from simulations, since the simulated training data itself is generated by the wrong model. SBI changes the form of the bias, not its existence, and a Gaussian likelihood is therefore a perfectly valid framework for demonstrating model misspecification. Jointly constraining multiple complementary observables as enabled by 21cmEMU is a natural step toward making inference pipelines more robust to such failures, since a misspecified model is less likely to simultaneously reproduce all independent observational probes. Developing dedicated tools to formalise this intuition, for example by estimating the Bayesian evidence as a goodness-of-fit diagnostic across different observing scales (Akhmetzhanova et al., 2025) or through field-level robustness analysis (Bayer et al., 2025), represents a promising research direction. Nevertheless, model misspecification remains an important open challenge for inference in astrophysics and cosmology, whether ML-accelerated or not, and mitigating it is one of the greatest current challenges, as it directly impacts the reliability of scientific conclusions drawn from upcoming surveys.

6. Conclusion

We have presented 21cmEMUv3, an emulator of 21cmFASTv3 modelling two galaxy populations: atomically and molecularly cooling galaxies, conditioned on eleven astrophysical and cosmological parameters. 21cmEMUv3 jointly predicts seven summary observables using a score-based diffusion model for the cylindrical 21-cm PS and LSTM networks for the remaining time-evolving summaries, achieving sub-percent median fractional error across all outputs. Notably, the 2D PS is emulated at a better accuracy than the 1D PS from 21cmEMUv1, despite being a significantly more complex summary statistic and being trained on a database nearly two orders of magnitude smaller.

We showcase 21cmEMUv3 by producing SKA parameter forecasts from 21-cm PS observations in $\sim$1 GPU-hour, over 6 orders of magnitude faster than with direct simulation on CPUs. SKA is forecast to detect the 21-cm PS at high significance of over 10σ at $z \sim 6.5$–10 and $k \sim 0.48$ Mpc^{-1}. A joint 11-parameter inference combining the SKA mock with UV luminosity functions, the Thomson optical depth, and an upper limit on the neutral fraction at $z = 5.9$ recovers the EoR history with a 95% CI width of $\Delta\bar{x}_{\rm HI} \lesssim 0.09$ around the midpoint of reionisation, and constrains τ_e to $\sim$6 times the precision of Planck18 (Planck Collaboration

et al., 2020).

We then demonstrate model misspecification by repeating the inference with 21cmEMUv1, which models only ACGs, on the SKA mock generated with the 21cmFASTv3 model of 21cmEMUv3. Despite broadly reproducing the observed summaries and satisfying the likelihood, the recovered astrophysical parameters are severely biased. This illustrates that a good fit to the observations is not sufficient to guarantee unbiased inference under model misspecification.

Looking ahead, a natural near-term step is to perform inference directly on the cylindrical 21-cm PS, which 21cmEMUv3 already emulates at sub-percent accuracy, preserving anisotropic information lost by spherical averaging. Future emulator generations will extend to field-level SKA image summaries and additional JWST galaxy observables, broadening the set of jointly constrained probes and thereby reducing sensitivity to model misspecification.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning and its applications to fundamental astrophysics and cosmology. By enabling Bayesian inference that is several orders of magnitude faster than direct simulation, this work also contributes to reducing the computational and energy costs associated with cosmological parameter inference. There are no harmful societal consequences of our work that we feel must be specifically highlighted here.

References

Akhmetzhanova, A., Cuesta-Lazaro, C., and Mishra-Sharma, S. Detecting model misspecification in cosmology with scale-dependent normalizing flows. *Machine Learning: Science and Technology*, 6(4):045020, December 2025. doi: 10.1088/2632-2153/ae107d.

Bañados, E., Venemans, B. P., Mazzucchelli, C., Farina, E. P., Walter, F., Wang, F., Decarli, R., Stern, D., Fan, X., Davies, F. B., Hennawi, J. F., Simcoe, R. A., Turner, M. L., Rix, H.-W., Yang, J., Kelson, D. D., Rudie, G. C., and Winters, J. M. An 800-million-solar-mass black hole in a significantly neutral Universe at a redshift of 7.5. *Nature*, 553(7689):473–476, January 2018. doi: 10.1038/nature25180.

Barkana, R. and Loeb, A. In the beginning: the first sources of light and the reionization of the universe. *Physics Reports*, 349(2):125–238, July 2001. doi: 10.1016/S0370-1573(01)00019-9.

Bayer, A. E., Villaescusa-Navarro, F., Sharief, S., Teyssier, R., Garrison, L. H., Perreault-Levasseur, L., Bryan, G. L., Gatti, M., and Visbal, E. Field-level Comparison and Robustness Analysis of Cosmological N-body Simulations. *ApJ*, 989(2):207, August 2025. doi: 10.3847/1538-4357/adef4e.

Becker, G. D., Bolton, J. S., Madau, P., Pettini, M., Ryan-Weber, E. V., and Venemans, B. P. Evidence of patchy hydrogen reionization from an extreme Lyα trough below redshift six. *MNRAS*, 447(4):3402–3419, March 2015. doi: 10.1093/mnras/stu2646.

Behroozi, P. S. and Silk, J. A Simple Technique for Predicting High-redshift Galaxy Evolution. *ApJ*, 799(1):32, January 2015. doi: 10.1088/0004-637X/799/1/32.

Bevins, H. T. J., Handley, W. J., Fialkov, A., de Lera Acedo, E., and Javid, K. GLOBALEMU: a novel and robust approach for emulating the sky-averaged 21-cm signal from the cosmic dawn and epoch of reionization. *MNRAS*, 508(2):2923–2936, December 2021. doi: 10.1093/mnras/stab2737.

Bolton, J. S., Haehnelt, M. G., Warren, S. J., Hewett, P. C., Mortlock, D. J., Venemans, B. P., McMahon, R. G., and Simpson, C. How neutral is the intergalactic medium surrounding the redshift z = 7.085 quasar ULAS J1120+0641? *MNRAS*, 416(1):L70–L74, September 2011. doi: 10.1111/j.1745-3933.2011.01100.x.

Bosman, S. E. I., Fan, X., Jiang, L., Reed, S., Matsuoka, Y., Becker, G., and Haehnelt, M. New constraints on Lyman-α opacity with a sample of 62 quasarsat z ¿ 5.7. *MNRAS*, 479(1):1055–1076, September 2018. doi: 10.1093/mnras/sty1344.

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., Trenti, M., Labbé, I., Bradley, L., Carollo, M., van Dokkum, P. G., Gonzalez, V., Holwerda, B., Franx, M., Spitler, L., Smit, R., and Magee, D. UV Luminosity Functions at Redshifts z $\sim$ 4 to z $\sim$ 10: 10,000 Galaxies from HST Legacy Fields. *ApJ*, 803(1):34, April 2015. doi: 10.1088/0004-637X/803/1/34.

Bouwens, R. J., Oesch, P. A., Labbé, I., Illingworth, G. D., Fazio, G. G., Coe, D., Holwerda, B., Smit, R., Stefanon, M., van Dokkum, P. G., Trenti, M., Ashby, M. L. N., Huang, J. S., Spitler, L., Straatman, C., Bradley, L., and Magee, D. The Bright End of the z $\sim$ 9 and z $\sim$ 10 UV Luminosity Functions Using All Five CANDELS Fields*. *ApJ*, 830(2):67, October 2016. doi: 10.3847/0004-637X/830/2/67.

Breitman, D., Mesinger, A., Murray, S. G., Prelogović, D., Qin, Y., and Trotta, R. 21CMEMU: an emulator of 21CMFAST summary observables. *MNRAS*, 527(4):9833–9852, February 2024. doi: 10.1093/mnras/stad3849.

Breitman, D., Mesinger, A., Murray, S. G., and Acharya, A. Sample variance denoising in cylindrical 21 cm power spectra. *A&A*, 703:A130, 2025. doi: 10.1051/0004-6361/202556449. URL https://doi.org/10.1051/0004-6361/202556449.

Bye, C. H., Portillo, S. K. N., and Fialkov, A. 21cmVAE: A Very Accurate Emulator of the 21 cm Global Signal. *ApJ*, 930(1):79, May 2022. doi: 10.3847/1538-4357/ac6424.

Cang, J., Mesinger, A., Murray, S. G., Breitman, D., Qin, Y., and Trotta, R. The EDGES measurement disfavors an excess radio background during the cosmic dawn. *arXiv e-prints*, art. arXiv:2411.08134, November 2024. doi: 10.48550/arXiv.2411.08134.

Clément, B., Cuby, J. G., Courbin, F., Fontana, A., Freudling, W., Fynbo, J., Gallego, J., Hibon, P., Kneib, J. P., Le Fèvre, O., Lidman, C., McMahon, R., Milvang-Jensen, B., Moller, P., Moorwood, A., Nilsson, K. K., Pentericci, L., Venemans, B., Villar, V., and Willis, J. Evolution of the observed Lyα luminosity function from z = 6.5 to z = 7.7: evidence for the epoch of reionization? *A&A*, 538:A66, February 2012. doi: 10.1051/0004-6361/201117312.

Cohen, A., Fialkov, A., Barkana, R., and Monsalve, R. A. Emulating the global 21-cm signal from Cosmic Dawn and Reionization. *MNRAS*, 495(4):4845–4859, July 2020. doi: 10.1093/mnras/staa1530.

Cranmer, K., Brehmer, J., and Louppe, G. The frontier of simulation-based inference. *Proceedings of the National Academy of Science*, 117(48):30055–30062, December 2020. doi: 10.1073/pnas.1912789117.

Dalla Vecchia, C. and Schaye, J. Simulating galactic outflows with kinetic supernova feedback. *MNRAS*, 387(4): 1431–1444, July 2008. doi: 10.1111/j.1365-2966.2008.13322.x.

Das, A., Mesinger, A., Pallottini, A., Ferrara, A., and Wise, J. H. High-mass X-ray binaries and the cosmic 21-cm signal: impact of host galaxy absorption. *MNRAS*, 469 (1):1166–1174, July 2017. doi: 10.1093/mnras/stx943.

Davies, J. E., Mesinger, A., and Murray, S. Efficient simulation of discrete galaxy populations and associated radiation fields during the first billion years. *arXiv e-prints*, art. arXiv:2504.17254, April 2025. doi: 10.48550/arXiv.2504.17254.

de Belsunce, R., Gratton, S., Coulton, W., and Efstathiou, G. Inference of the optical depth to reionization from low multipole temperature and polarization Planck data. *MNRAS*, 507(1):1072–1091, October 2021. doi: 10.1093/mnras/stab2215.

Diao, K. and Mao, Y. Multifidelity Emulator for Large-scale 21 cm Lightcone Images: A Few-shot Transfer Learning Approach with Generative Adversarial Network. *ApJ*, 992 (2):173, October 2025. doi: 10.3847/1538-4357/ae0325.

D'Odorico, V., Bañados, E., Becker, G. D., Bischetti, M., Bosman, S. E. I., Cupani, G., Davies, R., Farina, E. P., Ferrara, A., Feruglio, C., Mazzucchelli, C., Ryan-Weber, E., Schindler, J. T., Sodini, A., Venemans, B. P., Walter, F., Chen, H., Lai, S., Zhu, Y., Bian, F., Campo, S., Carniani, S., Cristiani, S., Davies, F., Decarli, R., Drake, A., Eilers, A. C., Fan, X., Gaikwad, P., Gallerani, S., Greig, B., Haehnelt, M. G., Hennawi, J., Keating, L., Kulkarni, G., Mesinger, A., Meyer, R. A., Neeleman, M., Onoue, M., Pallottini, A., Qin, Y., Rojas-Ruiz, S., Satyavolu, S., Sebastian, A., Tripodi, R., Wang, F., Wolfson, M., Yang, J., and Zanchettin, M. V. XQR-30: The ultimate XSHOOTER quasar sample at the reionization epoch. *MNRAS*, 523(1):1399–1420, July 2023. doi: 10.1093/mnras/stad1468.

Dorigo Jones, J., Bahauddin, S. M., Rapetti, D., Mirocha, J., and Burns, J. O. 21CMLSTM: A Fast Memory-based Emulator of the Global 21 cm Signal with Unprecedented Accuracy. *ApJ*, 977(1):19, December 2024. doi: 10.3847/1538-4357/ad8b20.

Draine, B. T. and Bertoldi, F. Structure of Stationary Photodissociation Fronts. *ApJ*, 468:269, September 1996. doi: 10.1086/177689.

Drake, A. B., Guiderdoni, B., Blaizot, J., Wisotzki, L., Herenz, E. C., Garel, T., Richard, J., Bacon, R., Bina, D., Cantalupo, S., Contini, T., den Brok, M., Hashimoto, T., Marino, R. A., Pelló, R., Schaye, J., and Schmidt, K. B. MUSE deep-fields: the Ly α luminosity function in the Hubble Deep Field-South at 2.91 ¡ z ¡ 6.64. *MNRAS*, 471 (1):267–278, October 2017. doi: 10.1093/mnras/stx1515.

Fan, X., Strauss, M. A., Becker, R. H., White, R. L., Gunn, J. E., Knapp, G. R., Richards, G. T., Schneider, D. P., Brinkmann, J., and Fukugita, M. Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with z~6 Quasars. II. A Sample of 19 Quasars. *AJ*, 132(1):117–136, July 2006. doi: 10.1086/504836.

Furlanetto, S. R. The global 21-centimeter background from high redshifts. *MNRAS*, 371(2):867–878, September 2006. doi: 10.1111/j.1365-2966.2006.10725.x.

Ghara, R., Giri, S. K., Mellema, G., Ciardi, B., Zaroubi, S., Iliev, I. T., Koopmans, L. V. E., Chapman, E., Gazagnes, S., Gehlot, B. K., Ghosh, A., Jelić, V., Mertens, F. G., Mondal, R., Schaye, J., Silva, M. B., Asad, K. M. B., Kooistra, R., Mevius, M., Offringa, A. R., Pandey, V. N., and Yatawatta, S. Constraining the intergalactic medium

at z = 9.1 using LOFAR Epoch of Reionization observations. *MNRAS*, 493(4):4728–4747, February 2020. doi: 10.1093/mnras/staa487.

Greig, B. and Mesinger, A. 21CMMC: an MCMC analysis tool enabling astrophysical parameter studies of the cosmic 21 cm signal. *MNRAS*, 449(4):4246–4263, June 2015. doi: 10.1093/mnras/stv571.

Haiman, Z., Abel, T., and Rees, M. J. The Radiative Feedback of the First Cosmological Objects. *ApJ*, 534(1): 11–24, May 2000. doi: 10.1086/308723.

Heinrich, C. and Hu, W. Reionization effective likelihood from Planck 2018 data. *PRD*, 104(6):063505, September 2021. doi: 10.1103/PhysRevD.104.063505.

HERA Collaboration, Abdurashidova, Z., Aguirre, J. E., Alexander, P., Ali, Z. S., Balfour, Y., Barkana, R., Beardsley, A. P., Bernardi, G., Billings, T. S., Bowman, J. D., Bradley, R. F., Bull, P., Burba, J., Carey, S., Carilli, C. L., Cheng, C., DeBoer, D. R., Dexter, M., de Lera Acedo, E., Dillon, J. S., Ely, J., Ewall-Wice, A., Fagnoni, N., Fialkov, A., Fritz, R., Furlanetto, S. R., Gale-Sides, K., Glendenning, B., Gorthi, D., Greig, B., Grobbelaar, J., Halday, Z., Hazelton, B. J., Heimersheim, S., Hewitt, J. N., Hickish, J., Jacobs, D. C., Julius, A., Kern, N. S., Kerrigan, J., Kittiwisit, P., Kohn, S. A., Kolopanis, M., Lanman, A., La Plante, P., Lekalake, T., Lewis, D., Liu, A., Ma, Y.-Z., MacMahon, D., Malan, L., Malgas, C., Maree, M., Martinot, Z. E., Matsetela, E., Mesinger, A., Mirocha, J., Molewa, M., Morales, M. F., Mosiane, T., Muñoz, J. B., Murray, S. G., Neben, A. R., Nikolic, B., Nunhokee, C. D., Parsons, A. R., Patra, N., Pieterse, S., Pober, J. C., Qin, Y., Razavi-Ghods, N., Reis, I., Ringuette, J., Robnett, J., Rosie, K., Santos, M. G., Sikder, S., Sims, P., Smith, C., Syce, A., Thyagarajan, N., Williams, P. K. G., and Zheng, H. HERA Phase I Limits on the Cosmic 21 cm Signal: Constraints on Astrophysics and Cosmology during the Epoch of Reionization. *ApJ*, 924(2):51, January 2022. doi: 10.3847/1538-4357/ac2ffc.

HERA Collaboration, Abdurashidova, Z., Adams, T., Aguirre, J. E., Alexander, P., Ali, Z. S., Baartman, R., Balfour, Y., Barkana, R., Beardsley, A. P., Bernardi, G., Billings, T. S., Bowman, J. D., Bradley, R. F., Breitman, D., Bull, P., Burba, J., Carey, S., Carilli, C. L., Cheng, C., Choudhuri, S., DeBoer, D. R., de Lera Acedo, E., Dexter, M., Dillon, J. S., Ely, J., Ewall-Wice, A., Fagnoni, N., Fialkov, A., Fritz, R., Furlanetto, S. R., Gale-Sides, K., Garsden, H., Glendenning, B., Gorce, A., Gorthi, D., Greig, B., Grobbelaar, J., Halday, Z., Hazelton, B. J., Heimersheim, S., Hewitt, J. N., Hickish, J., Jacobs, D. C., Julius, A., Kern, N. S., Kerrigan, J., Kittiwisit, P., Kohn, S. A., Kolopanis, M., Lanman, A., La Plante, P., Lewis, D., Liu, A., Loots, A., Ma, Y.-Z., MacMahon, D. H. E., Malan, L., Malgas, K., Malgas, C., Maree, M., Marero, B., Martinot, Z. E., McBride, L., Mesinger, A., Mirocha, J., Molewa, M., Morales, M. F., Mosiane, T., Muñoz, J. B., Murray, S. G., Nagpal, V., Neben, A. R., Nikolic, B., Nunhokee, C. D., Nuwegeld, H., Parsons, A. R., Pascua, R., Patra, N., Pieterse, S., Qin, Y., Razavi-Ghods, N., Robnett, J., Rosie, K., Santos, M. G., Sims, P., Singh, S., Smith, C., Swarts, H., Tan, J., Thyagarajan, N., Wilensky, M. J., Williams, P. K. G., van Wyngaarden, P., and Zheng, H. Improved Constraints on the 21 cm EoR Power Spectrum and the X-Ray Heating of the IGM with HERA Phase I Observations. *ApJ*, 945(2):124, March 2023. doi: 10.3847/1538-4357/acaf50.

Hermans, J., Begy, V., and Louppe, G. Likelihood-free MCMC with Amortized Approximate Ratio Estimators. *arXiv e-prints*, art. arXiv:1903.04057, March 2019. doi: 10.48550/arXiv.1903.04057.

Ho, J., Jain, A., and Abbeel, P. Denoising Diffusion Probabilistic Models. *arXiv e-prints*, art. arXiv:2006.11239, June 2020. doi: 10.48550/arXiv.2006.11239.

Hoag, A., Bradač, M., Huang, K., Mason, C., Treu, T., Schmidt, K. B., Trenti, M., Strait, V., Lemaux, B. C., Finney, E. Q., and Paddock, M. Constraining the Neutral Fraction of Hydrogen in the IGM at Redshift 7.5. *ApJ*, 878(1):12, June 2019. doi: 10.3847/1538-4357/ab1de7.

Hopkins, P. F., Kereš, D., Oñorbe, J., Faucher-Giguère, C.-A., Quataert, E., Murray, N., and Bullock, J. S. Galaxies on FIRE (Feedback In Realistic Environments): stellar feedback explains cosmologically inefficient star formation. *MNRAS*, 445(1):581–603, November 2014. doi: 10.1093/mnras/stu1738.

Hui, L. and Gnedin, N. Y. Equation of state of the photoionized intergalactic medium. *MNRAS*, 292(1):27–42, November 1997. doi: 10.1093/mnras/292.1.27.

Hutter, A., Dayal, P., Yepes, G., Gottlöber, S., Legrand, L., and Ucci, G. Astraeus I: the interplay between galaxy formation and reionization. *MNRAS*, 503(3):3698–3723, May 2021. doi: 10.1093/mnras/stab602.

Kern, N. S., Liu, A., Parsons, A. R., Mesinger, A., and Greig, B. Emulating Simulations of Cosmic Dawn for 21 cm Power Spectrum Constraints on Cosmology, Reionization, and X-Ray Heating. *ApJ*, 848(1):23, October 2017. doi: 10.3847/1538-4357/aa8bb4.

Kimm, T., Katz, H., Haehnelt, M., Rosdahl, J., Devriendt, J., and Slyz, A. Feedback-regulated star formation and escape of LyC photons from mini-haloes during reionization. *MNRAS*, 466(4):4826–4846, April 2017. doi: 10.1093/mnras/stx052.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization, 2017. URL https://arxiv.org/abs/1412.6980.

Konno, A., Ouchi, M., Ono, Y., Shimasaku, K., Shibuya, T., Furusawa, H., Nakajima, K., Naito, Y., Momose, R., Yuma, S., and Iye, M. Accelerated Evolution of the Lyα Luminosity Function at z ¿~7 Revealed by the Subaru Ultra-deep Survey for Lyα Emitters at z = 7.3. *ApJ*, 797 (1):16, December 2014. doi: 10.1088/0004-637X/797/1/16.

Koopmans, L., Pritchard, J., Mellema, G., Aguirre, J., Ahn, K., Barkana, R., van Bemmel, I., Bernardi, G., Bonaldi, A., Briggs, F., de Bruyn, A. G., Chang, T. C., Chapman, E., Chen, X., Ciardi, B., Dayal, P., Ferrara, A., Fialkov, A., Fiore, F., Ichiki, K., Illiev, I. T., Inoue, S., Jelic, V., Jones, M., Lazio, J., Maio, U., Majumdar, S., Mack, K. J., Mesinger, A., Morales, M. F., Parsons, A., Pen, U. L., Santos, M., Schneider, R., Semelin, B., de Souza, R. S., Subrahmanyan, R., Takeuchi, T., Vedantham, H., Wagg, J., Webster, R., Wyithe, S., Datta, K. K., and Trott, C. The Cosmic Dawn and Epoch of Reionisation with SKA. In *Advancing Astrophysics with the Square Kilometre Array (AAemu4)*, pp. 1, April 2015. doi: 10.22323/1.215.0001.

Lange, J. U. NAUTILUS: boosting Bayesian importance nested sampling with deep learning. *MNRAS*, 525(2): 3181–3194, October 2023. doi: 10.1093/mnras/stad2441.

Lazare, H., Sarkar, D., and Kovetz, E. D. HERA bound on x-ray luminosity when accounting for population III stars. *PRD*, 109(4):043523, February 2024. doi: 10.1103/PhysRevD.109.043523.

Lewis, J. S. W., Ocvirk, P., Aubert, D., Sorce, J. G., Shapiro, P. R., Deparis, N., Dawoodbhoy, T., Teyssier, R., Yepes, G., Gottlöber, S., Ahn, K., Iliev, I. T., and Chardin, J. Galactic ionizing photon budget during the epoch of reionization in the Cosmic Dawn II simulation. *MNRAS*, 496(4):4342–4357, August 2020. doi: 10.1093/mnras/staa1748.

Liu, B., Liu, X., Jin, X., Stone, P., and Liu, Q. Conflict-Averse Gradient Descent for Multi-task Learning. *arXiv e-prints*, art. arXiv:2110.14048, October 2021. doi: 10.48550/arXiv.2110.14048.

Loshchilov, I. and Hutter, F. Decoupled Weight Decay Regularization. *arXiv e-prints*, art. arXiv:1711.05101, November 2017. doi: 10.48550/arXiv.1711.05101.

Ma, X., Quataert, E., Wetzel, A., Hopkins, P. F., Faucher-Giguère, C.-A., and Kereš, D. No missing photons for reionization: moderate ionizing photon escape fractions from the FIRE-2 simulations. *MNRAS*, 498(2):2001–2017, October 2020. doi: 10.1093/mnras/staa2404.

Madau, P. and Dickinson, M. Cosmic Star-Formation History. *ARAA*, 52:415–486, August 2014. doi: 10.1146/annurev-astro-081811-125615.

Madau, P., Meiksin, A., and Rees, M. J. 21 Centimeter Tomography of the Intergalactic Medium at High Redshift. *ApJ*, 475(2):429–444, February 1997. doi: 10.1086/303549.

McGreer, I. D., Mesinger, A., and D'Odorico, V. Model-independent evidence in favour of an end to reionization by z $\approx$ 6. *MNRAS*, 447(1):499–505, February 2015. doi: 10.1093/mnras/stu2449.

Mertens, F. G., Mevius, M., Koopmans, L. V. E., Offringa, A. R., Zaroubi, S., Acharya, A., Brackenhoff, S. A., Ceccotti, E., Chapman, E., Chege, K., Ciardi, B., Ghara, R., Ghosh, S., Giri, S. K., Hothi, I., Höfer, C., Iliev, I. T., Jelić, V., Ma, Q., Mellema, G., Munshi, S., Pandey, V. N., and Yatawatta, S. Deeper multi-redshift upper limits on the Epoch of Reionization 21-cm signal power spectrum from LOFAR between z=8.3 and z=10.1. *arXiv e-prints*, art. arXiv:2503.05576, March 2025. doi: 10.48550/arXiv.2503.05576.

Mesinger, A. and Furlanetto, S. Efficient Simulations of Early Structure Formation and Reionization. *ApJ*, 669(2): 663–675, November 2007. doi: 10.1086/521806.

Mesinger, A., Furlanetto, S., and Cen, R. 21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal. *MNRAS*, 411(2):955–972, February 2011. doi: 10.1111/j.1365-2966.2010.17731.x.

Mondal, R., Bharadwaj, S., Majumdar, S., Bera, A., and Acharyya, A. The effect of non-Gaussianity on error predictions for the Epoch of Reionization (EoR) 21-cm power spectrum. *MNRAS*, 449:L41–L45, April 2015. doi: 10.1093/mnrasl/slv015.

Moriwaki, K., Jun, R. L., Osato, K., and Yoshida, N. CosmoGLINT: cosmological generative model for line intensity mapping with transformer. *MNRAS*, 545(3):staf2124, January 2026. doi: 10.1093/mnras/staf2124.

Mortlock, D. J., Warren, S. J., Venemans, B. P., Patel, M., Hewett, P. C., McMahon, R. G., Simpson, C., Theuns, T., Gonzáles-Solares, E. A., Adamson, A., Dye, S., Hambly, N. C., Hirst, P., Irwin, M. J., Kuiper, E., Lawrence, A., and Röttgering, H. J. A. A luminous quasar at a redshift of z = 7.085. *Nature*, 474(7353):616–619, June 2011. doi: 10.1038/nature10159.

Muñoz, J. B., Qin, Y., Mesinger, A., Murray, S. G., Greig, B., and Mason, C. The impact of the first galaxies on cosmic dawn and reionization. *MNRAS*, 511(3):3657–3681, April 2022. doi: 10.1093/mnras/stac185.

Murray, S., Greig, B., Mesinger, A., Muñoz, J., Qin, Y., Park, J., and Watkinson, C. 21cmFAST v3: A Python-integrated C code for generating 3D realizations of the cosmic 21cm signal. *The Journal of Open Source Software*, 5(54):2582, October 2020. doi: 10.21105/joss.02582.

Murray, S., Pober, J., and Kolopanis, M. 21cmSense v2: A modular, open-source 21 cm sensitivity calculator. *The Journal of Open Source Software*, 9(97):6501, May 2024. doi: 10.21105/joss.06501.

Mutch, S. J., Geil, P. M., Poole, G. B., Angel, P. W., Duffy, A. R., Mesinger, A., and Wyithe, J. S. B. Dark-ages reionization and galaxy formation simulation - III. Modelling galaxy formation and the epoch of reionization. *MNRAS*, 462(1):250–276, October 2016. doi: 10.1093/mnras/stw1506.

Nunhokee, C. D., Null, D., Trott, C. M., Barry, N., Qin, Y., Wayth, R. B., Line, J. L. B., Jordan, C. H., Pindor, B., Cook, J. H., Bowman, J., Chokshi, A., Ducharme, J., Elder, K., Guo, Q., Hazelton, B., Hidayat, W., Ito, T., Jacobs, D., Jong, E., Kolopanis, M., Kunicki, T., Lilleskov, E., Morales, M. F., Pober, J. C., Selvaraj, A., Shi, R., Takahashi, K., Tingay, S. J., Webster, R. L., Yoshiura, S., and Zheng, Q. Limits on the 21 cm power spectrum at z=6.5-7.0 from MWA observations. *arXiv e-prints*, art. arXiv:2505.09097, May 2025. doi: 10.48550/arXiv.2505.09097.

Ocvirk, P., Aubert, D., Sorce, J. G., Shapiro, P. R., Deparis, N., Dawoodbhoy, T., Lewis, J., Teyssier, R., Yepes, G., Gottlöber, S., Ahn, K., Iliev, I. T., and Hoffman, Y. Cosmic dawn II (CoDa II): a new radiation-hydrodynamics simulation of the self-consistent coupling of galaxy formation and reionization. *Monthly Notices of the Royal Astronomical Society*, 496(4):4087–4107, may 2020. doi: 10.1093/mnras/staa1266. URL https://doi.org/10.1093%2Fmnras%2Fstaa1266.

Oesch, P. A., Bouwens, R. J., Illingworth, G. D., Labbé, I., and Stefanon, M. The Dearth of z $\sim$ 10 Galaxies in All HST Legacy Fields—The Rapid Evolution of the Galaxy Population in the First 500 Myr. *ApJ*, 855(2):105, March 2018. doi: 10.3847/1538-4357/aab03f.

Okamoto, T., Gao, L., and Theuns, T. Mass loss of galaxies due to an ultraviolet background. *MNRAS*, 390(3):920–928, November 2008. doi: 10.1111/j.1365-2966.2008.13830.x.

Oke, J. B. and Gunn, J. E. Secondary standard stars for absolute spectrophotometry. *ApJ*, 266:713–717, March 1983. doi: 10.1086/160817.

Ouchi, M., Shimasaku, K., Furusawa, H., Saito, T., Yoshida, M., Akiyama, M., Ono, Y., Yamada, T., Ota, K., Kashikawa, N., Iye, M., Kodama, T., Okamura, S., Simpson, C., and Yoshida, M. Statistics of 207 Lyα Emitters at a Redshift Near 7: Constraints on Reionization and Galaxy Formation Models. *ApJ*, 723(1):869–894, November 2010. doi: 10.1088/0004-637X/723/1/869.

Paardekooper, J.-P., Khochfar, S., and Dalla Vecchia, C. The First Billion Years project: the escape fraction of ionizing photons in the epoch of reionization. *MNRAS*, 451(3):2544–2563, August 2015. doi: 10.1093/mnras/stv1114.

Papamakarios, G. and Murray, I. Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation. *arXiv e-prints*, art. arXiv:1605.06376, May 2016. doi: 10.48550/arXiv.1605.06376.

Papamakarios, G., Sterratt, D. C., and Murray, I. Sequential Neural Likelihood: Fast Likelihood-free Inference with Autoregressive Flows. *arXiv e-prints*, art. arXiv:1805.07226, May 2018. doi: 10.48550/arXiv.1805.07226.

Planck Collaboration, Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A. J., Barreiro, R. B., Bartolo, N., Basak, S., Battye, R., Benabed, K., Bernard, J. P., Bersanelli, M., Bielewicz, P., Bock, J. J., Bond, J. R., Borrill, J., Bouchet, F. R., Boulanger, F., Bucher, M., Burigana, C., Butler, R. C., Calabrese, E., Cardoso, J. F., Carron, J., Challinor, A., Chiang, H. C., Chluba, J., Colombo, L. P. L., Combet, C., Contreras, D., Crill, B. P., Cuttaia, F., de Bernardis, P., de Zotti, G., Delabrouille, J., Delouis, J. M., Di Valentino, E., Diego, J. M., Doré, O., Douspis, M., Ducout, A., Dupac, X., Dusini, S., Efstathiou, G., Elsner, F., Enßlin, T. A., Eriksen, H. K., Fantaye, Y., Farhang, M., Fergusson, J., Fernandez-Cobos, R., Finelli, F., Forastieri, F., Frailis, M., Fraisse, A. A., Franceschi, E., Frolov, A., Galeotta, S., Galli, S., Ganga, K., Génova-Santos, R. T., Gerbino, M., Ghosh, T., González-Nuevo, J., Górski, K. M., Gratton, S., Gruppuso, A., Gudmundsson, J. E., Hamann, J., Handley, W., Hansen, F. K., Herranz, D., Hildebrandt, S. R., Hivon, E., Huang, Z., Jaffe, A. H., Jones, W. C., Karakci, A., Keihänen, E., Keskitalo, R., Kiiveri, K., Kim, J., Kisner, T. S., Knox, L., Krachmalnicoff, N., Kunz, M., Kurki-Suonio, H., Lagache, G., Lamarre, J. M., Lasenby, A., Lattanzi, M., Lawrence, C. R., Le Jeune, M., Lemos, P., Lesgourgues, J., Levrier, F., Lewis, A., Liguori, M., Lilje, P. B., Lilley, M., Lindholm, V., López-Caniego, M., Lubin, P. M., Ma, Y. Z., Macías-Pérez, J. F., Maggio, G., Maino, D., Mandolesi, N., Mangilli, A., Marcos-Caballero, A., Maris, M., Martin, P. G., Martinelli, M., Martínez-González, E., Matarrese, S., Mauri, N., McEwen, J. D., Meinhold, P. R., Melchiorri, A., Mennella, A., Migliaccio, M., Millea, M., Mitra, S., Miville-

Deschênes, M. A., Molinari, D., Montier, L., Morgante, G., Moss, A., Natoli, P., Nørgaard-Nielsen, H. U., Pagano, L., Paoletti, D., Partridge, B., Patanchon, G., Peiris, H. V., Perrotta, F., Pettorino, V., Piacentini, F., Polastri, L., Polenta, G., Puget, J. L., Rachen, J. P., Reinecke, M., Remazeilles, M., Renzi, A., Rocha, G., Rosset, C., Roudier, G., Rubiño-Martín, J. A., Ruiz-Granados, B., Salvati, L., Sandri, M., Savelainen, M., Scott, D., Shellard, E. P. S., Sirignano, C., Sirri, G., Spencer, L. D., Sunyaev, R., Suur-Uski, A. S., Tauber, J. A., Tavagnacco, D., Tenti, M., Toffolatti, L., Tomasi, M., Trombetti, T., Valenziano, L., Valiviita, J., Van Tent, B., Vibert, L., Vielva, P., Villa, F., Vittorio, N., Wandelt, B. D., Wehus, I. K., White, M., White, S. D. M., Zacchei, A., and Zonca, A. Planck 2018 results. VI. Cosmological parameters. *A&A*, 641:A6, September 2020. doi: 10.1051/0004-6361/201833910.

Prelogović, D. and Mesinger, A. Exploring the likelihood of the 21-cm power spectrum with simulation-based inference. *MNRAS*, 524(3):4239–4255, September 2023. doi: 10.1093/mnras/stad2027.

Prelogović, D., Mesinger, A., Murray, S., Fiameni, G., and Gillet, N. Machine learning astrophysics from 21 cm lightcones: impact of network architectures and signal contamination. *MNRAS*, 509(3):3852–3867, January 2022. doi: 10.1093/mnras/stab3215.

Pritchard, J. R. and Loeb, A. 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, 75(8):086901, August 2012. doi: 10.1088/0034-4885/75/8/086901.

Qin, Y., Poulin, V., Mesinger, A., Greig, B., Murray, S., and Park, J. Reionization inference from the CMB optical depth and E-mode polarization power spectra. *MNRAS*, 499(1):550–558, November 2020. doi: 10.1093/mnras/staa2797.

Qin, Y., Mesinger, A., Prelogović, D., Becker, G., Bischetti, M., Bosman, S., Davies, F., D'Odorico, V., Gaikwad, P., Haehnelt, M., Keating, L., Lai, S., Ryan-Weber, E., Satyavolu, S., Walter, F., and Zhu, Y. Percent-level timing of reionisation: Self-consistent, implicit-likelihood inference from XQR-30+ Lyα forest data. *PASA*, 42:e049, April 2025. doi: 10.1017/pasa.2025.35.

Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv e-prints*, art. arXiv:1505.04597, May 2015. doi: 10.48550/arXiv.1505.04597.

Santos, M. G., Ferramacho, L., Silva, M. B., Amblard, A., and Cooray, A. Fast large volume simulations of the 21-cm signal from the reionization and pre-reionization epochs. *MNRAS*, 406(4):2421–2432, August 2010. doi: 10.1111/j.1365-2966.2010.16898.x.

Schmit, C. J. and Pritchard, J. R. Emulation of reionization simulations for Bayesian inference of astrophysics parameters using neural networks. *MNRAS*, 475(1):1213–1223, March 2018. doi: 10.1093/mnras/stx3292.

Shaw, A. K., Bharadwaj, S., and Mondal, R. The impact of non-Gaussianity on the error covariance for observations of the Epoch of Reionization 21-cm power spectrum. *MNRAS*, 487(4):4951–4964, August 2019. doi: 10.1093/mnras/stz1561.

Shibuya, T., Ouchi, M., Harikane, Y., and Nakajima, K. Morphologies of $\sim$190,000 Galaxies at z = 0-10 Revealed with HST Legacy Data. III. Continuum Profile and Size Evolution of Lyα Emitters. *ApJ*, 871(2):164, February 2019. doi: 10.3847/1538-4357/aaf64b.

Shimabukuro, H. and Semelin, B. Analysing the 21 cm signal from the epoch of reionization with artificial neural networks. *MNRAS*, 468(4):3869–3877, July 2017. doi: 10.1093/mnras/stx734.

Sobacchi, E. and Mesinger, A. The depletion of gas in high-redshift dwarf galaxies from an inhomogeneous reionization. *MNRAS*, 432:L51–L55, May 2013. doi: 10.1093/mnrasl/slt035.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-Based Generative Modeling through Stochastic Differential Equations. *arXiv e-prints*, art. arXiv:2011.13456, November 2020. doi: 10.48550/arXiv.2011.13456.

Springel, V. and Hernquist, L. The history of star formation in a Λ cold dark matter universe. *MNRAS*, 339(2):312–334, February 2003. doi: 10.1046/j.1365-8711.2003.06207.x.

Sun, G. and Furlanetto, S. R. Constraints on the star formation efficiency of galaxies during the epoch of reionization. *MNRAS*, 460(1):417–433, July 2016. doi: 10.1093/mnras/stw980.

The HERA Collaboration, Abdurashidova, Z., Adams, T., Aguirre, J. E., Baartman, R., Barkana, R., Berkhout, L. M., Bernardi, G., Billings, T. S., Bizarria, B. B., Bowman, J. D., Breitman, D., Bull, P., Burba, J., Byrne, R., Carey, S., Sushovan Chandra, R., Chen, K.-F., Choudhuri, S., Cox, T., DeBoer, D. R., de Lera Acedo, E., Dexter, M., Dhandha, J., Dillon, J. S., Dynes, S., Eksteen, N., Ely, J., Ewall-Wice, A., Fagnoni, N., Fialkov, A., Furlanetto, S. R., Gale-Sides, K., Garsden, H., Gorce, A., Gorthi, D., Halday, Z., Hazelton, B. J., Hewitt, J. N., Hickish, J., Huang, T., Jacobs, D. C., Josaitis, A., Kern, N. S., Kerrigan, J., Kittiwisit, P., Kolopanis, M., Lanman, A., La Plante, P., Liu, A., Ma, Y.-Z., MacMahon, D. H. E., Malan, L., Malgas, C., Malgas, K., Marero, B., Martinot,

Z. E., McBride, L., Mesinger, A., Mirocha, J., Mohamed-Hinds, N., Molewa, M., Morales, M. F., Muñoz, J. B., Murray, S. G., Nikolic, B., Nuwegeld, H., Parsons, A. R., Pascua, R., Patra, N., Pochinda, S., Qin, Y., Rath, E., Razavi-Ghods, N., Riley, D., Rosie, K., Santos, M. G., Singh, S., Storer, D., Swarts, H., Tan, J., Thélie, E., van Wyngaarden, P., Wilensky, M. J., Williams, P. K. G., and Zheng, H. First Results from HERA Phase II. *arXiv e-prints*, art. arXiv:2511.21289, November 2025.

Visbal, E., Haiman, Z., and Bryan, G. L. Limits on Population III star formation in minihaloes implied by Planck. *MNRAS*, 453(4):4456–4466, November 2015. doi: 10.1093/mnras/stv1941.

Wang, F., Davies, F. B., Yang, J., Hennawi, J. F., Fan, X., Barth, A. J., Jiang, L., Wu, X.-B., Mudd, D. M., Banados, E., Bian, F., Decarli, R., Eilers, A.-C., Farina, E. P., Venemans, B., Walter, F., and Yue, M. A Significantly Neutral Intergalactic Medium Around the Luminous z=7 Quasar J0252-0503. *arXiv e-prints*, art. arXiv:2004.10877, April 2020. doi: 10.48550/arXiv.2004.10877.

Wise, J. H. and Abel, T. Suppression of H_2 Cooling in the Ultraviolet Background. *ApJ*, 671(2):1559–1567, December 2007. doi: 10.1086/522876.

Xu, H., Wise, J. H., Norman, M. L., Ahn, K., and O'Shea, B. W. Galaxy Properties and UV Escape Fractions during the Epoch of Reionization: Results from the Renaissance Simulations. *ApJ*, 833(1):84, December 2016. doi: 10.3847/1538-4357/833/1/84.

Yang, J., Wang, F., Fan, X., Hennawi, J. F., Davies, F. B., Yue, M., Banados, E., Wu, X.-B., Venemans, B., Barth, A. J., Bian, F., Boutsia, K., Decarli, R., Farina, E. P., Green, R., Jiang, L., Li, J.-T., Mazzucchelli, C., and Walter, F. Pōniuā'ena: A Luminous z = 7.5 Quasar Hosting a 1.5 Billion Solar Mass Black Hole. *ApJ*, 897(1):L14, July 2020. doi: 10.3847/2041-8213/ab9c26.

Yue, B., Ferrara, A., and Xu, Y. On the faint-end of the high-z galaxy luminosity function. *MNRAS*, 463(2):1968–1979, December 2016. doi: 10.1093/mnras/stw2145.

Zhao, X., Ting, Y.-S., Diao, K., and Mao, Y. Can diffusion model conditionally generate astrophysical images? *MNRAS*, 526(2):1699–1712, December 2023. doi: 10.1093/mnras/stad2778.

A. Galaxy model

Our simulations use the code 21cmFASTv3, with the galaxy model that includes two distinct galaxy populations: ACGs and MCGs. For more details on the galaxy model, users are encouraged to consult (Qin et al., 2020) and (Muñoz et al., 2022). Below we summarize the most relevant points, focusing on the free parameters that will serve as inputs to our emulator.

We characterize the stellar-to-halo mass relations (SHMRs) of the faint galaxies that dominate the EoR photon budget (e.g. Qin et al. 2025) via power laws (e.g., Behroozi & Silk 2015; Mutch et al. 2016; Sun & Furlanetto 2016; Yue et al. 2016). Specifically, the stellar mass $M_{*,\text{ACG (MCG)}}$ of an ACG (MCG) galaxy hosted by a halo of mass M_h can be written as:

$$M_{*,\text{ACG (MCG)}}/M_h = \min\left[1, f_{*,10(7)}\left(\frac{M_h}{M_{10(7)}}\right)^{\alpha_{*,\text{ACG (MCG)}}}\right] \times \left(\frac{\Omega_b}{\Omega_m}\right), \tag{4}$$

where Ω_b and Ω_m are the baryon and total matter energy densities, respectively. We may write the stellar fraction in ACGs (MCGs) as $f_{*,\text{ACG (MCG)}} \equiv f_{*,10(7)}\left(\frac{M_h}{M_{10(7)}}\right)^{\alpha_{*,\text{ACG (MCG)}}} \in [0,1]$, with $f_{*,10(7)}$ corresponding to the fraction of galactic gas in stars normalized to the amount in a halo of mass $M_{10(7)} \equiv 10^{10(7)} M_\odot$. These SHMRs introduce a total of four free parameters: the normalisation and power-law index of each of the two relations. We keep $f_{*,10(7)}$ and $\alpha_{*,\text{ACG}}$ as three free parameters, and fix $\alpha_{*,\text{MCG}} = \alpha_{*,\text{ACG}}$.

We assume that stellar mass grows exponentially over time. We write the corresponding star formation rate (SFR) as a function of the Hubble time, $H^{-1}(z)$, (or analogously the dynamical time, which also scales with the Hubble time during matter domination):

$$\dot{M}_{*,\text{ACG (MCG)}} = \frac{M_{*,\text{ACG (MCG)}}}{t_* H^{-1}(z)}, \tag{5}$$

where the characteristic star formation timescale, $t_* \in [0,1]$, is another free parameter.

Star formation is suppressed in low mass halos due to inefficient gas cooling and/or feedback (e.g., Hui & Gnedin 1997; Springel & Hernquist 2003; Okamoto et al. 2008; Sobacchi & Mesinger 2013; Xu et al. 2016; Ocvirk et al. 2020; Ma et al. 2020). We account for this suppression by including an exponential duty cycle to the mass function of halos (HMF) that host star-forming ACGs (MCGs):

$$\phi^{\text{ACG (MCG)}} = \frac{\mathrm{d}n}{\mathrm{d}M_h} \times \begin{cases} \exp\left(-\frac{M_{\text{crit}}^{\text{ACG}}}{M_h}\right) \\ \exp\left(-\frac{M_{\text{crit}}^{\text{MCG}}}{M_h}\right)\exp\left(-\frac{M_h}{M_{\text{crit}}^{\text{cool}}}\right) \end{cases}. \tag{6}$$

$M_{\text{crit}}^{\text{ACG (MCG)}}$ is the turnover mass below which star formation is suppressed, which we write as:

$$M_{\text{crit}}^{\text{ACG (MCG)}} = \max\left[M_{\text{crit}}^{\text{cool (diss)}}, M_{\text{crit}}^{\text{ion}}, M_{\text{crit}}^{\text{SN}}\right]. \tag{7}$$

Here the efficiency of star formation is limited by three physical processes: (i) inefficient cooling $M_{\text{crit}}^{\text{cool}}$ below the atomic cooling threshold for ACGs (e.g. Barkana & Loeb 2001) and Lyman-Werner feedback that photodissociates molecular hydrogen via the two-step Solomon process, $M_{\text{crit}}^{\text{diss}}$, for MCGs (e.g. Draine & Bertoldi 1996; Visbal et al. 2015); (ii) photoheating feedback, $M_{\text{crit}}^{\text{ion}}$, inside the ionized IGM (e.g. Hui & Gnedin 1997; Okamoto et al. 2008; Sobacchi & Mesinger 2013); (iii) supernova feedback, $M_{\text{crit}}^{\text{SN}}$ (e.g. Haiman et al. 2000; Wise & Abel 2007; Dalla Vecchia & Schaye 2008; Hopkins et al. 2014; Kimm et al. 2017). Moreover, for MCGs, in addition to the lower mass threshold, $M_{\text{crit}}^{\text{MCG}}$, we also include a second exponential term with an upper mass threshold, $M_{\text{crit}}^{\text{cool}}$, to smoothly transition between MCGs and ACGs around the atomic cooling threshold $T_{\text{vir}} \sim 10^4$K (c.f. Qin et al. 2020).

The typical ionizing escape fraction, $f_{\text{esc,ACG (MCG)}} \in [0,1]$ is described by a power-law (e.g., Paardekooper et al. 2015; Kimm et al. 2017; Lewis et al. 2020):

$$f_{\text{esc,ACG (MCG)}} = \min\left[1, f_{\text{esc},10(7)}\left(\frac{M_h}{M_{10(7)}}\right)^{\alpha_{\text{esc}}}\right], \tag{8}$$

where we have a total of three free parameters for both relations: the normalizations $f_{\text{esc},10(7)}$, and the power-law index, α_{esc}.

Table 3. Comparison of the priors used in this work (two rightmost columns) with the priors used in Lazare et al. 2024 (second column). The table only lists the parameters whose prior in this work is different from Lazare et al. 2024. The third column is the closest prior possible to Lazare et al. 2024 within the constraints of our emulator training database. The fourth column is an informed prior based on Xu et al. 2016 and ensures continuity around 10^8 $M_\odot$.

Parameter	Lazare+24	Lazare-like
$\log_{10} f_{*,7}$	$[-3.5, -1]$	$[-3.5, -1]$
$\alpha_{*,\mathrm{MCG}}$	$[-0.5, 0.5]$	$\alpha_{*,\mathrm{ACG}}$
$\log_{10} f_{*,10}$	$[-3, 0]$	$[-2, 0.5]$
$\alpha_{*,\mathrm{ACG}}$	$[-0.5, 1]$	$[0, 1]$
$\log_{10} f_{\mathrm{esc},7}$	$[-3, 0]$	$[-3, -1]$
α_X	$[-1, 3]$	1

The specific X-ray luminosity escaping the galaxies is also taken to be a power-law in energy (e.g., Das et al. 2017), $L_{X,\mathrm{ACG\,(MCG)}} \propto E^{-\alpha_X}$. We normalise it via the soft-band (i.e. < 2 keV) X-ray luminosity per unit SFR for each galaxy population:

$$L^{\mathrm{ACG\,(MCG)}}_{X<2\mathrm{keV}/}/\dot{M}_{*,\mathrm{ACG\,(MCG)}} = \int_{E_0}^{2\mathrm{keV}} \mathrm{dE}\; L_{\mathrm{X}}/\dot{M}_{*,\mathrm{ACG\,(MCG)}}, \tag{9}$$

where E_0 is the minimum energy of X-ray photons capable of escaping their host galaxy, for a total of three additional free parameters.

B. 21cmEMUv3 performance

In Figure 5, we plot true (solid) and emulated (dashed) outputs for 10 random test set samples across all summaries. The 10 samples are identical across panels. True and emulated curves are typically indistinguishable by eye, consistent with the sub-percent accuracy reported in Table 2. In the 1D PS panel, we additionally show individual IC realisations (thin solid lines) alongside the mean (thick solid line), confirming that the emulated 1D PS is closer to the mean than any single realisation.

C. HERA23 inference with MCGs

HERA Collaboration et al. 2023 (henceforth HERA23) reported the deepest upper limits on the 21-cm PS to date and interpreted them with Bayesian inference using 21cmFASTv3 (Mesinger & Furlanetto, 2007; Mesinger et al., 2011; Murray et al., 2020), where the likelihood was a synergy of four complementary CD/EoR observables:

1. UV luminosity functions (Bouwens et al., 2015; 2016; Oesch et al., 2018) with a Gaussian likelihood;
2. Thomson scattering optical depth to the CMB (Planck Collaboration et al., 2020; Qin et al., 2020) which uses a Gaussian likelihood around $\tau_e = 0.0569^{+0.0081}_{-0.0086}$;
3. upper limit $\overline{x}_{\mathrm{HI}} < 0.06 \pm 0.05$ on the EoR history at $z = 5.9$ obtained with the Lyα dark fraction method (McGreer et al., 2015). The likelihood function is unity at $\overline{x}_{\mathrm{HI}}(z = 5.9) < 0.06$, decreasing as a one-sided Gaussian for higher neutral fraction values;
4. HERA23 upper limits on 21-cm PS with a positive unbounded prior on systematics.

The analysis in HERA Collaboration et al. 2022 and HERA23 reveals that current upper limits on the 21-cm PS favour early galaxies that were more X-ray luminous per unit star formation rate, compared to present-day galaxies. The X-ray luminosity of present day star-forming galaxies is dominated by high mass X-ray binary stars (HMXBs), and the inferred luminosities from HERA23 are consistent with HMXB models and extrapolations of empirical trends to the low metallicity regime expected in the first galaxies.

The HERA23 analysis, however, only accounted for HMXBs hosted in ACGs. As pointed out in HERA23, if the first MCGs hosted efficient HMXBs, they would add a second, previously unaccounted for population contributing to the X-ray background. This would weaken the inferred constraints on ACG X-ray efficiency, allowing for lower values than those

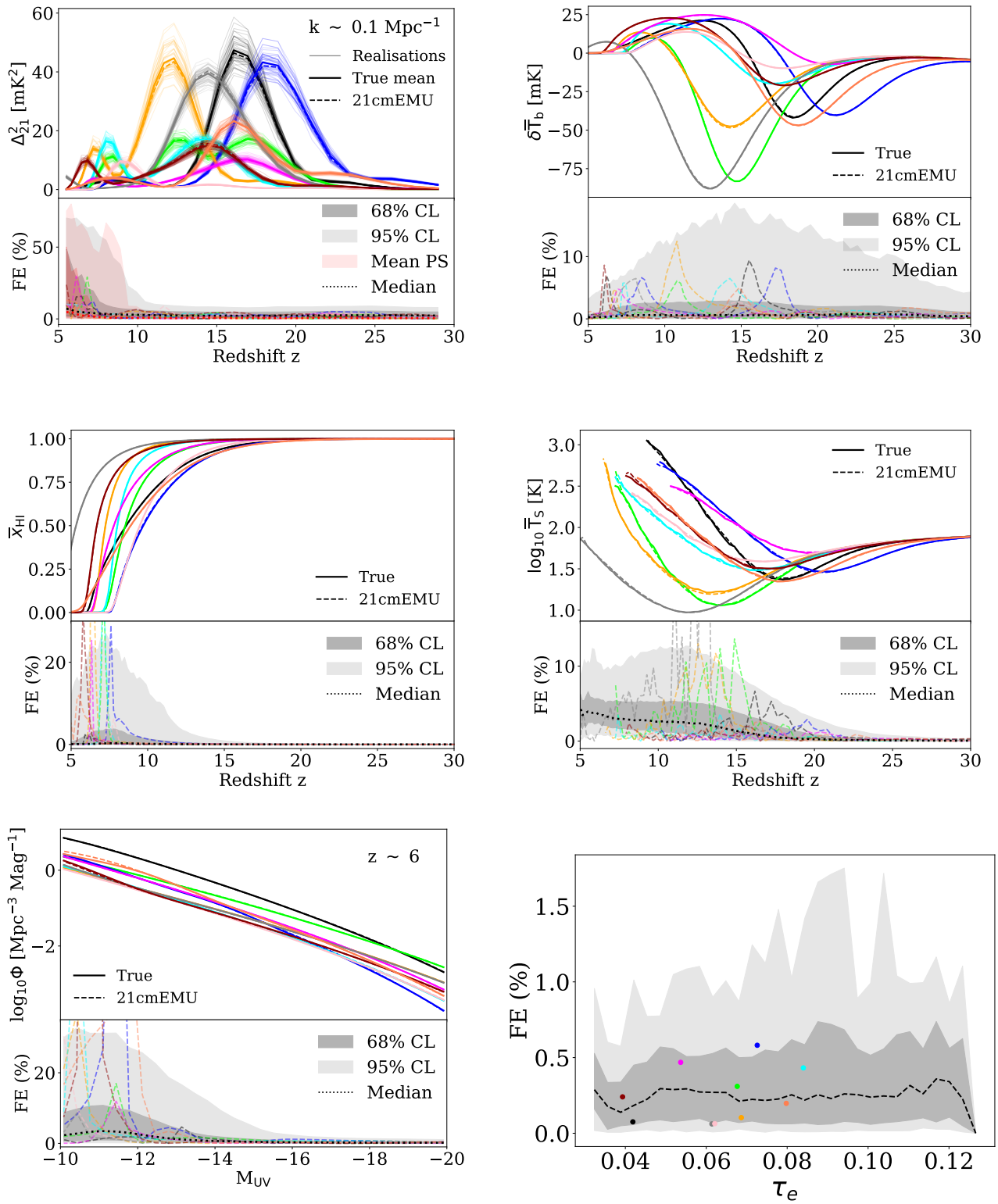

Figure 5. True (solid) and emulated (dashed) outputs from `21cmEMUv3` for ten random test set samples. From top to bottom, left to right: 1D 21-cm PS at $k = 0.1$ Mpc^{-1}, mean 21-cm brightness temperature, neutral hydrogen fraction, mean spin temperature (all as a function of redshift), UV LF at $z = 6$, and CMB optical depth. Bottom sub-panels show fractional errors between true and emulated quantities; individual sample errors are shown in colour and the median over the full test set in dashed black. Dark (light) shaded regions enclose 68% (95%) CL. The red shaded regions for the PS are evaluated against PS means rather than individual realisations.

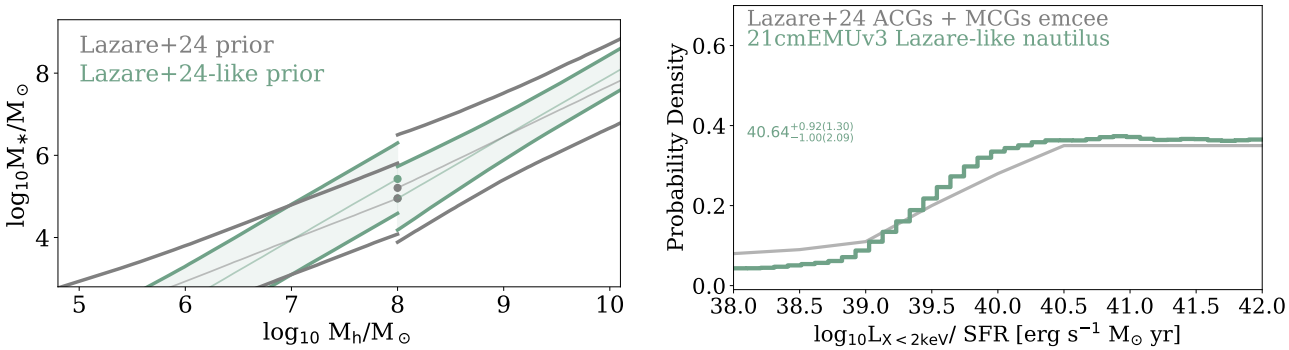

Figure 6. Left plot: In grey, we show the prior on the SHMR from L24 (Lazare et al., 2024). In green, we show the prior that we use in this work that is similar to L24 but not exactly the same due to differences in the simulation database. Note that our L24-like prior favours lower stellar masses in comparison to L24. Right plot: Posteriors of the X-ray luminosity per unit SFR parameter for the two priors shown in the left plot.

obtained under the single-population assumption. This was confirmed by Lazare et al. 2024 (henceforth L24), who repeated the HERA analysis using an emulator of a `21cmFASTv3` model that includes both ACGs and MCGs. L24 assumed the same X-ray efficiency per unit SFR ($L_{\mathrm{X<2keV}}/\mathrm{SFR}$) for both ACGs and MCGs, finding that the additional population of HMXBs hosted by MCGs resulted in weaker limits on $L_{\mathrm{X<2keV}}/\mathrm{SFR}$.

We repeat the analysis in L24 using `21cmEMUv3` with the `nautilus` sampler. Like L24, we assume $\log_{10} L_{\mathrm{X<2keV}}/\mathrm{SFR} \equiv \log_{10} L^{\mathrm{MCG}}_{\mathrm{X<2keV}}/\mathrm{SFR} = \log_{10} L^{\mathrm{ACG}}_{\mathrm{X<2keV}}/\mathrm{SFR}$. Our prior differs slightly from theirs because `21cmEMUv3` was not trained on the same simulation database as their emulator. Consequently, we adopt a "L24-like" prior that is as close as possible to theirs while ensuring that it remains reasonably within the parameter space on which our emulator was trained. Table 3 compares the two priors: the second column is the prior in L24, while the third column is the L24-like prior used in this work. We can compare visually the difference between the two priors in Figure 6, where we plot the 68% C.L. of the SHMR for their prior in grey and for our prior in green. Looking at lower halo masses at the bottom left of the plot, we can see that the L24-like prior in green excludes the higher stellar masses between $\mathrm{M}_h \sim 10^5 - 10^6 M_\odot$ allowed in the original prior.

In the right plot of Figure 6, we show in grey a rough outline of the result from L24 and in black the result from Breitman et al. 2024 repeating the HERA23 analysis. In green, we show the result obtained using `21cmEMUv3` and the L24-like prior. The two posteriors are similar qualitatively above $\log_{10} L_{\mathrm{X<2keV}}/\mathrm{SFR} \gtrsim 39$. At the lower tail of the distribution, however, we can see that our L24-like prior rules out X-ray faint galaxies much more than the original prior. Looking at the grey and green SHMRs on the left plot, this occurs because our L24-like prior excludes higher stellar masses between $\mathrm{M}_h \sim 10^5 - 10^6 M_\odot$ allowed in the original prior.

D. Full posterior distributions

In this section, we include the full corner plots for the inferences presented in Section 5. In Figure 7, we shown the full corner plot for the well-specified inference from Section 5.1. Then, in Figure 8, we show the full corner plot for the misspecified inference from Section 5.2. The derived summaries in the upper right panels in Figure 8 show that the summaries are generally well recovered and physical. The corner plot itself on the bottom left of Figure 8, however, shows that the recovery of most parameters is severely biased, with the ground truth laying outside the posterior range entirely for a few parameters.

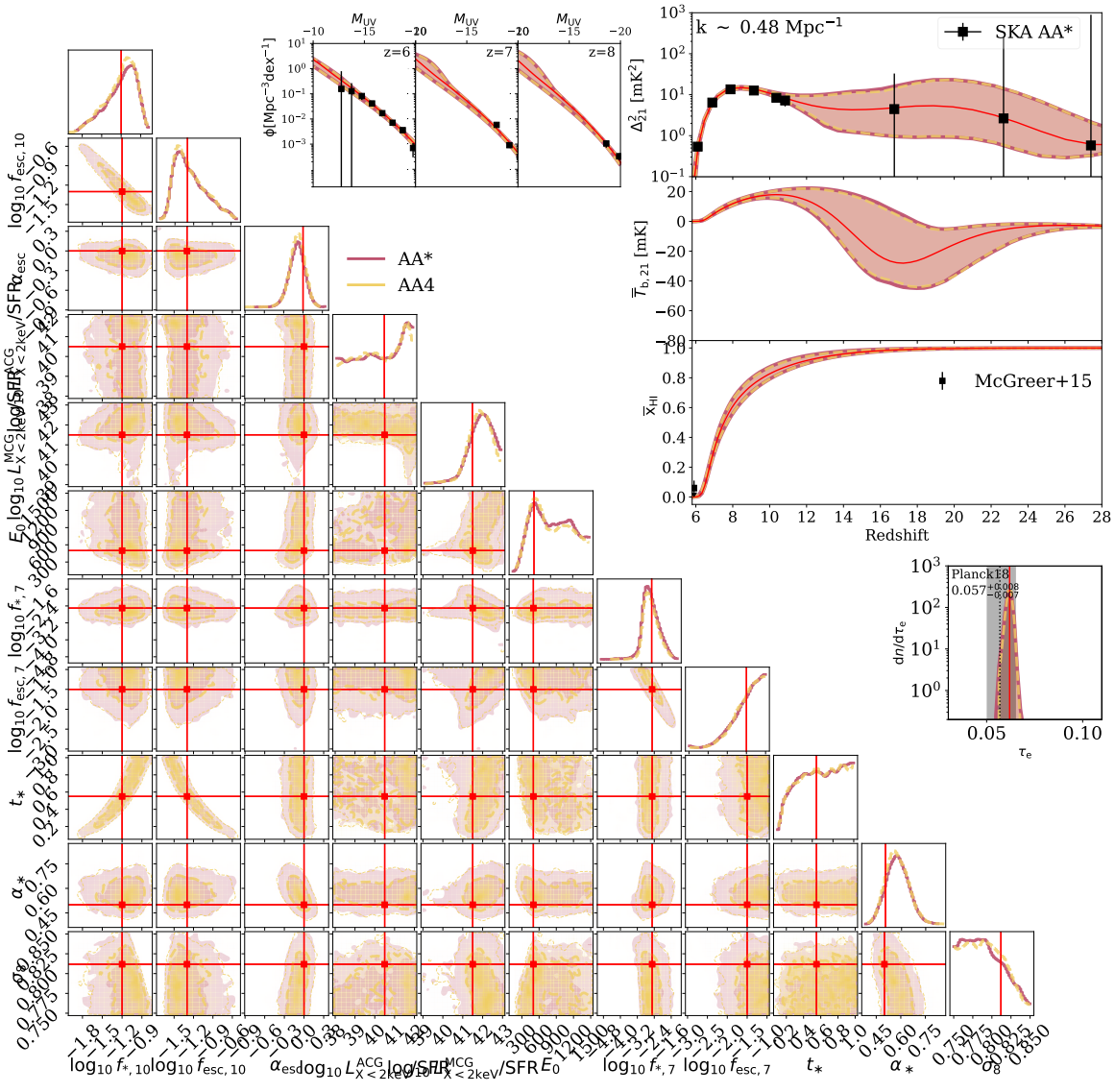

Figure 7. Posterior distributions from the well-specified SKA inference using `21cmEMUv3`. All shaded regions are 95% CI except for the Thomson scattering optical depth where we plot the 68% CI to compare directly with Planck18.

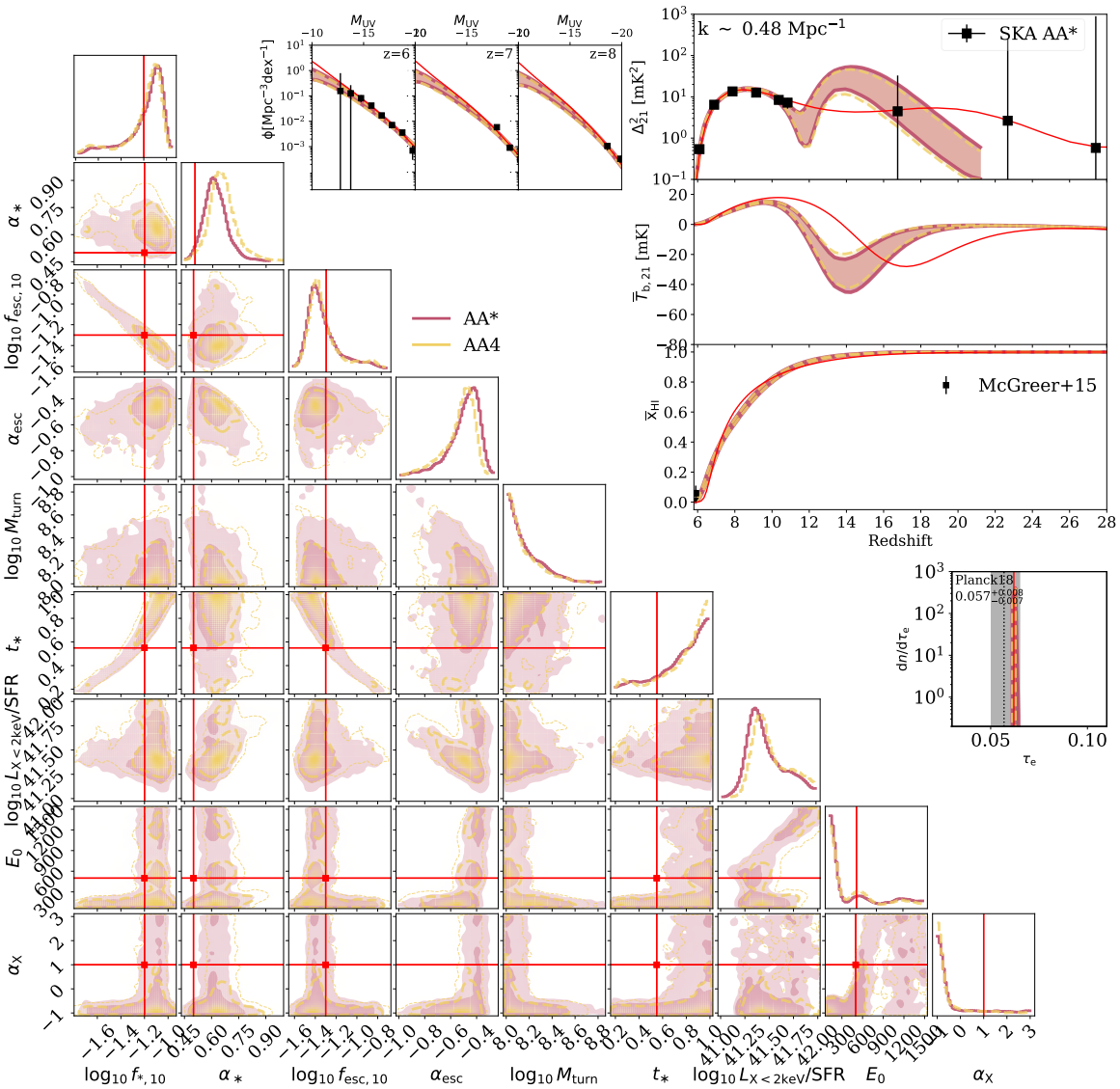

Figure 8. Corner plot of inference with different model from the mock. All shaded regions are 95% CI except for the Thomson scattering optical depth where we plot the 68% CI to compare directly with Planck18. Note that the ground truth $\alpha_{\rm esc} = 0.0$, $\log_{10} L_{\rm X<2keV}/{\rm SFR} \equiv \log_{10} {\rm L}^{\rm ACG}_{\rm X<2keV}/{\rm SFR} = 40.5$, and $\log_{10} L^{\rm MCG}_{\rm X<2keV}/{\rm SFR} = 41.5$. The posterior is biased so far away that the truth in red is outside the plot range.