

Simulation-based inference of the sky-averaged 21-cm signal from CD-EoR with REACH

Anchal Saxena¹,^{*} P. Daniel Meerburg,¹ Christoph Weniger,² Eloy de Lera Acedo^{3,4} and Will Handley^{3,4}

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*Cavendish Astrophysics, University of Cambridge, Cambridge CB3 0HE, UK*

⁴*Kavli Institute for Cosmology, Madingley Road, Cambridge CB3 0HA, UK*

Accepted 2024 November 7. Received 2024 October 10; in original form 2024 June 21

ABSTRACT

The redshifted 21-cm signal from the cosmic dawn and epoch of reionization carries invaluable information about the cosmology and astrophysics of the early Universe. Analysing data from a sky-averaged 21-cm signal experiment requires navigating through an intricate parameter space addressing various factors such as foregrounds, beam uncertainties, ionospheric distortions, and receiver noise for the search of the 21-cm signal. The traditional likelihood-based sampling methods for modelling these effects could become computationally demanding for such complex models, which makes it infeasible to include physically motivated 21-cm signal models in the analysis. Moreover, the inference is driven by the assumed functional form of the likelihood. We demonstrate how simulation-based inference through truncated marginal neural ratio estimation (TMNRE) can naturally handle these issues at a reduced computational cost. We estimate the posterior distribution on our model parameters with TMNRE for simulated mock observations, incorporating beam-weighted foregrounds, physically motivated 21-cm signal, and radiometric noise. We find that maximizing information content by analysing data from multiple time slices and antennas significantly improves the parameter constraints and enhances the exploration of the cosmological signal. We discuss the application of TMNRE for the current configuration of the REACH experiment and demonstrate its potential for exploring new avenues.

Key words: Methods: data analysis – Methods: statistical – Data Methods – Machine Learning – Dark ages, reionization, first stars.

1 INTRODUCTION

Cosmic dawn (CD) marks the formation of the first stars and galaxies, which emitted high energy radiation in X-ray and UV bands and ionized the neutral hydrogen that filled the primordial Cosmos during the so-called epoch of reionization (EoR) (Barkana & Loeb 2001; Furlanetto, Oh & Briggs 2006; Pritchard & Loeb 2012). The current astrophysical constraints on CD-EoR are derived from the observations of high redshift quasar spectra (Becker et al. 2001; Fan et al. 2003; Boera et al. 2019), the optical depth of electrons scattering from the cosmic microwave background (CMB) (Kaplinghat et al. 2003; Komatsu et al. 2011; Planck Collaboration VI 2020), and the luminosity function and clustering properties of Lyman- α emitters (Jensen et al. 2012; Dijkstra 2014; Gangolli et al. 2020). These indirect probes fail to offer a precise understanding of the physical processes unfolding during CD-EoR. The hyperfine transition of the neutral hydrogen, characterized by a rest-frame wavelength of 21-cm, allows us to directly probe these epochs at low frequencies.

Various experiments are already underway to detect the redshifted 21-cm signal. Interferometric observations from LOFAR (Mertens

et al. 2020), GMRT (Paciga et al. 2013), MWA (Barry et al. 2019; Li et al. 2019), NenuFAR (Mertens, Semelin & Koopmans 2021; Munshi et al. 2024), HERA (DeBoer et al. 2017; Berkhout et al. 2024), and the upcoming SKA (Koopmans et al. 2015; Mellema et al. 2015) characterize the spatial fluctuations in the 21-cm brightness temperature using Fourier statistics. In parallel, there is an alternative approach, which is conceptually simpler but more challenging to calibrate, to detect the sky-averaged 21-cm signal with a wide beam antenna system. These experiments include EDGES (Bowman et al. 2018), SARAS (Singh et al. 2018, 2021), PRIZM (Philip et al. 2018), SCI-HI (Voytek et al. 2014), LEDA (Price et al. 2018), DAPPER (Burns et al. 2021), MIST (Monsalve et al. 2024), and REACH (de Lera Acedo et al. 2022).

A common challenge for all these experiments is the characterization and isolation of the foregrounds, which are 4–6 orders of magnitude brighter than the cosmological 21-cm signal. The key to separating the foregrounds and the 21-cm signal is to exploit the spectral differences between the two. Foregrounds can be considered spectrally smooth in contrast to the 21-cm signal (Shaver et al. 1999). However, the beam pattern of the observing instrument carries a spatial and spectral dependence, which corrupts the smoothness of the foregrounds. The resulting beam-weighted foregrounds have a significant co-variance with the 21-cm signal, which makes the

* E-mail: anchal.saxena.009@gmail.com (AS); p.d.meerburg@rug.nl (PDM)

extraction of the 21-cm signal quite challenging (Vedantham et al. 2013; Bernardi, McQuinn & Greenhill 2015; Tauscher et al. 2018; Anstey, de Lera Acedo & Handley 2021; Spinelli et al. 2021; Saxena et al. 2023a). In addition, distortions due to ionosphere (Vedantham et al. 2013; Shen et al. 2021), Radio Frequency Interference (RFI) (Anstey & Leeney 2023; Leeney, Handley & de Lera Acedo 2023), and receiver uncertainties (Tauscher et al. 2021; Razavi-Ghods et al. 2023) need to be carefully modelled in the data analysis. To enhance the separation of the 21-cm signal from these systematics, one can leverage drift-scan time information (Tauscher, Rapetti & Burns 2020b; Saxena et al. 2023a), measurements of all four Stokes parameters (Nhan, Bradley & Burns 2017; Nhan et al. 2019; Tauscher et al. 2020b), and the dipole of the signal (Slosar 2017; Deshpande 2018; Ignatov, Pritchard & Wu 2023).

In 2018, the experiment to detect the global EoR signature (EDGES) reported the observation of an absorption trough of depth -500_{-500}^{+200} mK centred at 78 ± 1 MHz (Bowman et al. 2018). However, the characteristics of the identified signal diverge significantly from the anticipated features outlined in standard astrophysical models (Cohen et al. 2017, 2020; Reis, Fialkov & Barkana 2021). This unexpected discrepancy prompts the consideration of either an exotic explanation to enhance the contrast between the radio background temperature and the spin temperature of the 21-cm signal or an unaccounted systematic effect within the data. Two theoretically plausible scenarios for the former explanation involve the excessive cooling of hydrogen gas due to interactions with dark matter (Barkana 2018; Berlin et al. 2018; Liu et al. 2019) and the existence of an amplified radio background distinct from the CMB (Ewall-Wice et al. 2018; Feng & Holder 2018; Fialkov & Barkana 2019; Ewall-Wice, Chang & Lazio 2020; Reis, Fialkov & Barkana 2020). Nevertheless, a re-examination of the EDGES data indicates that the presence of unaccounted systematics may distort or obscure the signal (Hills et al. 2018; Singh & Subrahmanyan 2019; Sims & Pober 2020; Bevins et al. 2021a). Additionally, potential deficiencies in the foreground model used to fit the observations could contribute to the observed trough (Tauscher et al. 2020a). An independent measurement of the radio sky spectrum by SARAS in the 55–85 MHz band, as reported in Singh et al. (2021), suggests at a confidence level of $\sim 2\sigma$ that the signal detected by EDGES may not be entirely of astrophysical origin.

This work is based in the context of the radio experiment for the analysis of cosmic hydrogen (REACH).¹ REACH is a wide band experiment encompassing both the CD and EoR to measure the sky-averaged 21-cm signal in 50–170 MHz frequency range. Deployed in RFI-quiet Karoo radio reserve in South Africa, REACH, in its phase I, is observing the Southern hemisphere sky with a hexagonal dipole antenna operating between 50 and 130 MHz (Cumner et al. 2022; Lera Acedo et al. 2022). Phase II is expected to feature more antennas, including a conical log-spiral antenna operating in 50–170 MHz, and systems sensitive to the polarization of the sky radiation.

In order to achieve a successful and convincing detection, the data must be jointly modelled in a Bayesian framework to account for the 21-cm signal, foregrounds, and instrumental effects, with the goal of identifying and separating these components as much as possible, as they are not inherently orthogonal. Physically motivated models of the sky-averaged 21-cm signal incorporate several astrophysical and cosmological parameters which define the shape of the signal (Thomas et al. 2009; Mesinger, Furlanetto & Cen 2010; Santos et al. 2010; Fialkov & Barkana 2014). These models can generate a

realization of the signal on the order of seconds to minutes. However, fitting such 21-cm signal models jointly with the foregrounds in a Bayesian framework requires $\sim 10^7$ samples, which is computationally expensive. Therefore, existing data analysis and interpretation methods use the shape of a Gaussian (Bernardi et al. 2016; Anstey et al. 2021) or a flattened Gaussian (Bowman et al. 2018) to define the 21-cm signal, which cannot describe both the absorption and emission part of the signal. The development of neural network based emulators such as `globalemu` (Bevins et al. 2021b), 21CMVAE (Bye, Portillo & Fialkov 2022), and 21CMEMU (Breitman et al. 2024) enable fast and efficient parameter estimation. However, such emulators must be very accurate for an unbiased parameter inference (Jones et al. 2023).

Moreover, the statistical analysis with the conventional likelihood-based approach relies on the functional form of the likelihood, which is generally assumed to be Gaussian. This assumption leads to inaccurate posterior inferences if the noise structure is a priori unknown (Scheutwinkel, Handley & de Lera Acedo 2023). These methods sample the full joint posterior distribution of parameters and consequently require more samples (and computation time) to converge for high dimensional parameter spaces (Handley, Hobson & Lasenby 2015). In the context of the data analysis of any sky-averaged 21-cm signal experiment, such methods could become a bottleneck because of a large number of degrees of freedom, including the foreground parameters, beam uncertainties, 21-cm signal parameters, ionospheric distortions, and receiver effects.

These issues can be resolved by performing a simulation-based inference (SBI), where deep learning algorithms are used to estimate the posterior distribution of parameters (Alsing et al. 2019; Cranmer, Brehmer & Louppe 2020; Lueckmann et al. 2021). In SBI methods, the likelihood is implicitly defined through forward simulations. There are a variety of methods to perform SBI; however, in this work, we apply truncated marginal neural ratio estimation (TMNRE) (Miller et al. 2021) algorithm for parameter inference. This approach can directly estimate the marginal posteriors for the parameters of interest instead of sampling the full joint posterior, making it significantly more efficient for exploring high-dimensional parameter spaces. This has already been applied in several astrophysical and cosmological parameter inference applications (Cole et al. 2022; Coogan et al. 2022; Montel & Weniger 2022; Alvey et al. 2023a, b; Bhardwaj et al. 2023; Karchev, Trotta & Weniger 2023; Saxena et al. 2023b; Karchev et al. 2024).

In this work, we perform SBI of the foregrounds and 21-cm signal for simulated mock observations for the REACH experiment requiring up to an order of magnitude fewer simulations than the likelihood-based sampling methods. We demonstrate the flexibility of our framework by applying it, for the first time, on a physically motivated 21-cm signal model for the current setup of REACH. We further discuss how this algorithm can be used to explore potential avenues for REACH phase II. This article is organized as follows: In Section 2, we give a brief overview of SBI and its implementation through TMNRE. In Section 3, we describe the modelling of the foregrounds and 21-cm signal. In Section 4, we present the inference of foregrounds and cosmological signal for two different models of the 21-cm signal for the REACH experiment. In Section 5, we summarize our findings and discuss the scope for future work.

2 SIMULATION-BASED INFERENCE

In this section, we give a brief overview of SBI (Alsing et al. 2019; Cranmer et al. 2020; Lueckmann et al. 2021), and its implementation as adopted in this work. In the past 4–5 yr, there has been a

¹<https://www.astro.phy.cam.ac.uk/research/research-projects/reach>

notable rise in the advancement and utilization of SBI approaches for analysing data (Cranmer et al. 2020). Specifically, the central question guiding various algorithms is *whether a robust Bayesian inference is achievable for a given generative model*.

To be more precise, consider a forward generative model, $p(\mathbf{x}, \boldsymbol{\theta})$, which takes some underlying set of parameters $\boldsymbol{\theta}$ and generates simulated data \mathbf{x} . In a Bayesian sense, the forward model takes the form $p(\mathbf{x}, \boldsymbol{\theta}) = p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is sampled from a prior distribution $p(\boldsymbol{\theta})$. This expression represents the key functionality of SBI-based methods since running a forward generative model is equivalent to sampling from the simulated data likelihood $p(\mathbf{x}|\boldsymbol{\theta})$. Scientifically, we are interested in the posterior distribution $p(\boldsymbol{\theta}|\mathbf{x})$ of the parameters $\boldsymbol{\theta}$ for an observation \mathbf{x} , which follows from Bayes' theorem

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

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

Given this setup and the ability to sample from the forward model, there are three main approaches in which SBI methods construct posterior densities. The first approach is *neural posterior estimation* (Papamakarios & Murray 2018; Tejero-Cantero et al. 2020; Zeghal et al. 2022; Dax et al. 2023), where the goal is to represent the posterior density $p(\boldsymbol{\theta}|\mathbf{x})$ with some flexible density estimators such as normalizing flow. The second class of methods falls under the category of *neural likelihood estimation* (Papamakarios & Murray 2018; Alsing et al. 2019; Lin et al. 2023), where the goal is to have an estimator for the simulated data likelihood $p(\mathbf{x}|\boldsymbol{\theta})$ to be embedded later in traditional inference techniques such as Markov chain Monte Carlo (MCMC) or nested sampling. The final class of methods, which is also the focus of this work, is *neural ratio estimation* (Miller et al. 2021), which approximates the likelihood-to-evidence ratio $p(\mathbf{x}|\boldsymbol{\theta})/p(\mathbf{x})$. This is discussed in detail in the following section.

2.1 Truncated marginal neural ratio estimation

In this work, we use a specific algorithm of SBI known as truncated marginal neural ratio estimation (TMNRE). The two critical features of this method, which make it highly efficient compared to traditional algorithms, are truncation and marginalization. We will first discuss the neural ratio estimation part of this approach and later discuss how these features come into play.

2.2 Neural ratio estimation

In SBI, the information about the likelihood is implicitly accessed through a stochastic simulator, which defines a mapping from input parameters $\boldsymbol{\theta}$ to a sample $\mathbf{x} \sim p(\mathbf{x}|\boldsymbol{\theta})$. We sample from this simulator to generate sample-parameter pairs $\{(x^1, \boldsymbol{\theta}^1)(x^2, \boldsymbol{\theta}^2), \dots\}$. Here, $\boldsymbol{\theta}^i$ is typically drawn from the prior, so the pairs are drawn from the joint distribution $p(\mathbf{x}, \boldsymbol{\theta})$. These pairs are then employed to train a neural network to approximate the likelihood-to-evidence ratio, a procedure known as neural ratio estimation (NRE). This ratio can be expressed as

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

In other words, $r(\mathbf{x}, \boldsymbol{\theta})$ is equal to the ratio of the joint probability density $p(\mathbf{x}, \boldsymbol{\theta})$ to the product of marginal probability densities $p(\mathbf{x})p(\boldsymbol{\theta})$. This method then takes the form of a binary classification

task to distinguish between jointly drawn and marginally drawn pairs. A binary classifier $d_\phi(\mathbf{x}, \boldsymbol{\theta})$ with some learnable parameters ϕ is then optimized to perform this classification. The parameters ϕ of the model are updated as the model is trained.

For optimizing the model $d_\phi(\mathbf{x}, \boldsymbol{\theta})$, we introduce a binary label y to differentiate between the jointly drawn ($y = 1$) and marginally drawn ($y = 0$) sample-parameter pairs. To be precise, y is a random variable. Once the classifier is trained, it approximates the probability that a sample-parameter pair $(\mathbf{x}, \boldsymbol{\theta})$ is drawn jointly ($y = 1$), i.e.

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

where we assumed $p(y = 0) = p(y = 1) = 0.5$. This learning problem is associated with a binary cross-entropy loss function

$$-\int [p(\mathbf{x}, \boldsymbol{\theta}) \ln d_\phi(\mathbf{x}, \boldsymbol{\theta}) + p(\mathbf{x})p(\boldsymbol{\theta}) \ln\{1 - d_\phi(\mathbf{x}, \boldsymbol{\theta})\}] d\mathbf{x} d\boldsymbol{\theta}, \quad (3)$$

which is minimized using stochastic gradient descent to find the optimal parameters ϕ of the network. Once the network is trained, it results in

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

which can be re-written as

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

to estimate the posterior probability distribution. In practice, this classifier is simply a dense neural network with a few hidden layers.

The two key features of this method are the following.

(i) It can directly estimate marginal posteriors by omitting model parameters from the network's input instead of first sampling the full joint posterior and then performing the marginalization and,

(ii) The analysis is done sequentially in multiple rounds. The posterior densities are estimated at the end of each round for the given observation. The prior for the next round is then truncated based on the inference in the current round, hence the name truncated marginal (NRE). This leads to a targeted inference where extremely low posterior density regions are discarded after each round.

Both these features significantly reduce the simulation budget compared to likelihood-based sampling methods. In this work, we use TMNRE as implemented in the software package `swyft`.² This algorithm is summarized in Fig. 1. The four basic steps involved in this framework are as follows.

(i) **(Re-)Simulate:** Generate samples $\mathbf{x} \sim p(\mathbf{x}|\boldsymbol{\theta})$ based on the current prior distribution $p(\boldsymbol{\theta})$ of model parameters.³

(ii) **Ratio estimation:** Train ratio estimators $r(\mathbf{x}, \boldsymbol{\theta})$ for each parameter of interest θ_i using samples from step (i) to approximate the posterior-to-prior ratio $p(\theta_i|\mathbf{x})/p(\theta_i)$.

(iii) **Inference:** For a given mock observation x_0 , estimate the ratios $r(x_0, \theta_i)$ across the prior samples $p(\theta_i)$, and weight the samples based on this ratio.

²<https://github.com/undark-lab/swyft>

³This step is fully parallelized in our implementation which makes it significantly fast with access to appropriate hardware.

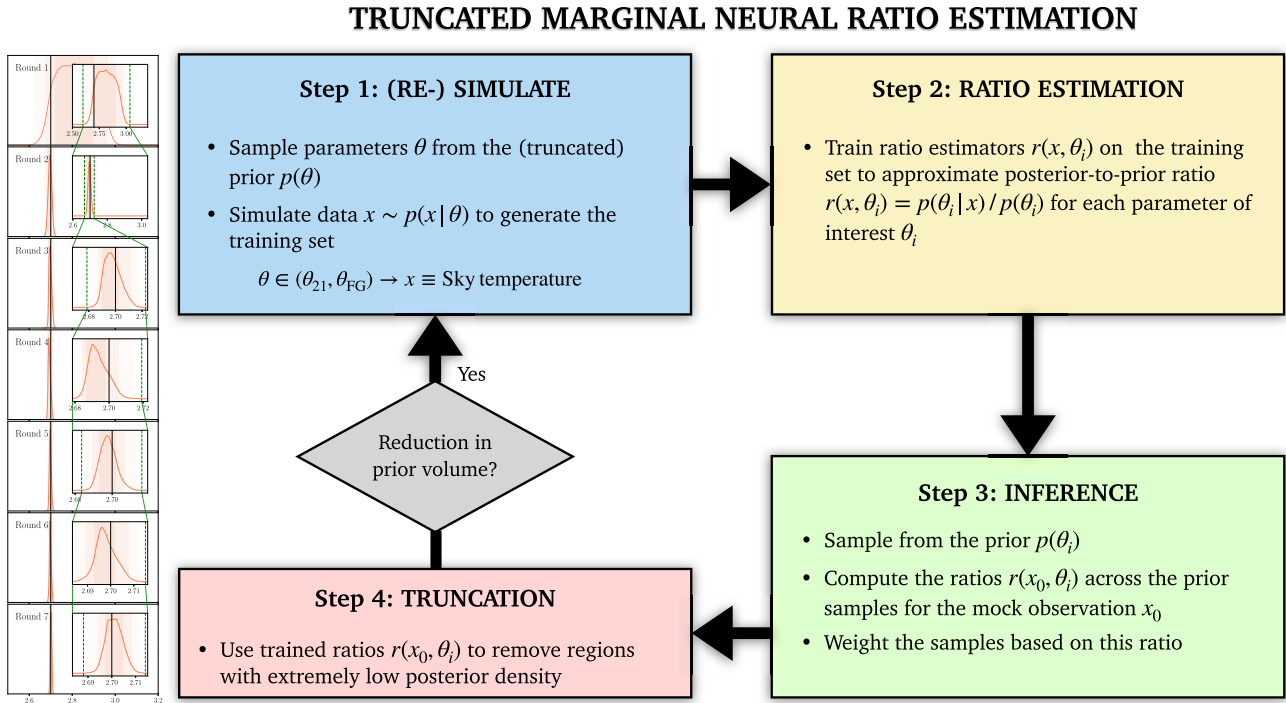


Figure 1. A schematic view of the TMNRE algorithm as adopted in our SBI framework.

Table 1. Settings for the TMNRE algorithm used in this work.

Setting	Value
Initial learning rate	5×10^{-4}
Max training epochs	70
Early stopping	20
Number of rounds	7
Simulation schedule	30 k, 30 k, 60 k, 90 k, 150 k, 300 k, 300 k
Bounds threshold	10^{-3}

(iv) **Truncation:** Apply truncation by removing regions with extremely low posterior density based on the weighted samples from step (iii). If the prior volume is reduced, then re-simulate from step (i) with the truncated prior.

Our TMNRE setup is implemented with the settings listed in Table 1. These include the initial learning rate, the maximum number of epochs to train the network, the number of epochs to wait for the validation loss to increase (early stopping), the number of rounds to train the network, simulation schedule i.e. the number of samples to be used in each round and bounds threshold for applying truncation.

3 SIMULATIONS AND TRAINING DATA

This section describes our forward modelling of the foregrounds and the 21-cm signal. We assume an observation (\mathbf{T}_{obs}) to be composed of the sky-averaged 21-cm signal (\mathbf{T}_{21}), beam-weighted foregrounds (\mathbf{T}_{FG}), and a radiometric noise component given as $\sigma = (\mathbf{T}_{21} + \mathbf{T}_{\text{FG}})/\sqrt{\Delta\nu\Delta t}$, where $\Delta\nu$ is the channel width and Δt is the integration time. The forward model then takes the form

$$\mathbf{T}_{\text{obs}}(\nu) = \mathbf{T}_{21}(\nu) + \mathbf{T}_{\text{FG}}(\nu) + \sigma(\nu). \quad (6)$$

There are also other sources of error in the data due to imperfect calibration of the instrument, additive, and multiplicative biases through the receiver's gain. In this work, we assume that the

instrument is well-calibrated and do not take into account these sources of uncertainty. These effects will be considered in a future work.

3.1 Modelling the foregrounds

A beam-weighted foreground model can be constructed from the following features: (i) a spatial brightness temperature distribution, (ii) a spatial distribution of spectral index, and (iii) the spatial and spectral dependence of the observing antenna.

We generate our sky model as

$$T_{\text{sky}}(\theta, \phi, \nu) = [T_{230}(\theta, \phi) - T_{\text{CMB}}] \left(\frac{\nu}{230} \right)^{-\beta(\theta, \phi)} + T_{\text{CMB}}, \quad (7)$$

where $T_{230}(\theta, \phi)$ is the spatial brightness temperature distribution derived from the global sky map (GSM) at 230 MHz (de Oliveira-Costa et al. 2008), which we use as our base foreground map, and $T_{\text{CMB}} = 2.725$ K. Our simulated observations start at 2019 October 1 00:00:00 UTC and are integrated for 6 h. In the left panel of Fig. 2, we show the overhead sky at 230 MHz on 2019 October 1 00:00:00 UTC for an antenna located in the Karoo radio reserve in South Africa.

Next, for deriving a spectral index variation map, we apply a pixel-wise tracing between an instance of GSM at 230 and 408 MHz as

$$\beta(\theta, \phi) = \frac{\log \left(\frac{T_{230}(\theta, \phi) - T_{\text{CMB}}}{T_{408}(\theta, \phi) - T_{\text{CMB}}} \right)}{\log \left(\frac{230}{408} \right)}. \quad (8)$$

The resulting spectral index map is shown in the middle panel of Fig. 2. To model the foregrounds in a parametrized manner, we divide the sky into nine regions and assign a constant spectral index to each region. This physically motivated modelling of the foregrounds follows from Anstey et al. (2021). Our choice of using nine regions is based on the findings of Anstey et al. (2021) for the

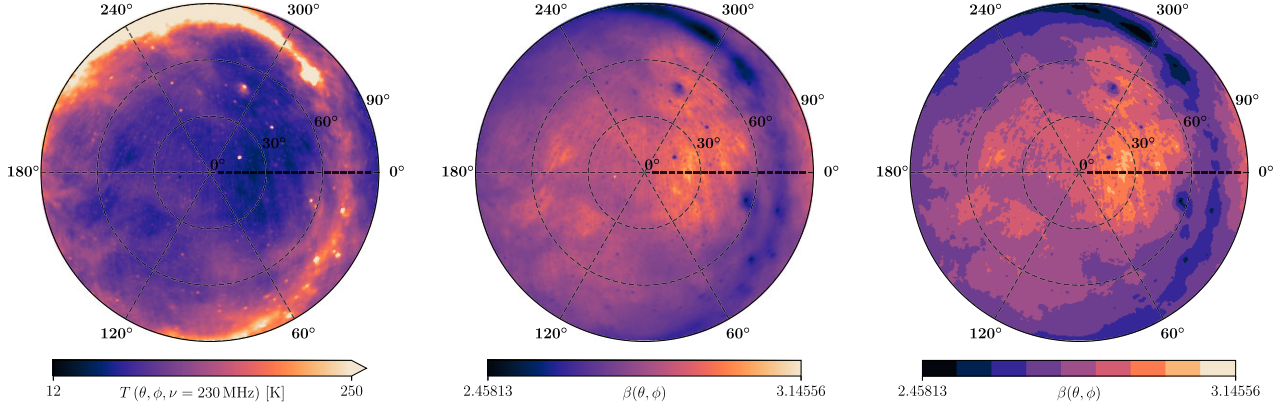


Figure 2. Radio sky GSM at 230 MHz (*left panel*), spectral index distribution (*middle panel*), and the division of the sky into nine regions (*right panel*). These projections correspond to a simulated observation on 2019 October 1 at 00:00:00 UTC when the galactic disc remains below the horizon ($\theta > 90^\circ$) for an antenna located in the Karoo radio reserve. These are shown in co-latitude (θ) and azimuth (ϕ) coordinates of the antenna’s frame with the zenith in the centre.

simulated observations of REACH. However, the number of regions required to adequately model the data also depends on the number of time slices and the chromaticity of the antenna beam. Therefore, the selection of the number of regions should be data-driven, based on Bayesian evidence (Anstey et al. 2021, 2023; Pagano et al. 2023). For modern SBI algorithms, the recently proposed learned harmonic mean estimator (McEwen et al. 2023) offers a promising method for estimating Bayesian evidence (Spurio Mancini et al. 2023), which we plan to investigate in future work. These regions are formed by dividing the full range of spectral indices as shown in the middle panel of Fig. 2 into nine equal intervals, and each region is defined as the patch of the sky with spectral indices within these intervals. The parametrized foreground model $T_{\text{sky}}(\theta, \phi, \nu)$ then takes the form

$$\left[\sum_{i=1}^9 M_i(\theta, \phi) (T_{230}(\theta, \phi) - T_{\text{CMB}}) \left(\frac{\nu}{230} \right)^{-\beta_i} \right] + T_{\text{CMB}}, \quad (9)$$

where $M_i(\theta, \phi)$ represents a mask for the sky region i , which has a spectral index β_i . $M_i(\theta, \phi) = 1$ for every pixel within the region and 0 elsewhere. The division of the sky is shown in the right panel of Fig. 2, where the spectral index for each region is randomly sampled from $\beta_i \in (2.4, 3.2)$ for $1 \leq i \leq 9$.

This parametrized sky model is convolved with the beam of the antenna to form beam-weighted foregrounds and then averaged over the sky

$$T_{\text{FG}}(\nu) = \frac{1}{4\pi} \int_0^{4\pi} G(\theta, \phi, \nu) T_{\text{sky}}(\theta, \phi, \nu) d\Omega, \quad (10)$$

where $G(\theta, \phi, \nu)$ describes the spatial and spectral dependence of the antenna beam. In this work, we consider two different antenna designs: a hexagonal dipole and a conical log-spiral antenna. Our simulations are run in 50–200 MHz frequency range. The beam patterns of these antennas are shown in Fig. 3 at three different frequencies $\nu = 50, 125,$ and 200 MHz in the antenna’s frame of reference, where the green line marks the horizon.

To generate a realization of the beam-weighted foregrounds, we sample the spectral index for each region $\beta_i \in (2.4, 3.2)$ for $1 \leq i \leq 9$, and use equations (9) and (10). In the left panel of Fig. 4, we show a thin slice of the beam-weighted foreground modelling set (\mathbf{M}_{FG}) for a dipole antenna, generated by uniformly sampling over the prior on β_i . The bold line represents a random realization used in the mock observation ($T_{\text{FG}}^{\text{mock}}$) with $\beta = \{2.54, 2.8, 2.63, 2.85, 2.7, 2.92, 2.72, 3.05, 2.91\}$.

3.2 Modelling the 21-cm signal

In this work, we consider two different types of models (i) a physically motivated model of the sky-averaged 21-cm signal, and (ii) a Gaussian profile, which is generally used as an approximation of the cosmological signal (Anstey et al. 2021, 2023). The details of these two models are discussed in the following subsections.

Gaussian 21-cm signal model

For simplicity, our first model of the sky-averaged 21-cm signal takes the form of a Gaussian, which is described as

$$T_{21}(\nu) = -A \exp \left[-\frac{(\nu - \nu_0)^2}{2\sigma^2} \right], \quad (11)$$

where A is the amplitude of the absorption feature, ν_0 is the central frequency, and σ is the standard deviation. We adopt uniform priors on each of these parameters as follows: $A \in (0 \text{ K}, 0.25 \text{ K})$, $\nu_0 \in (50 \text{ MHz}, 200 \text{ MHz})$, and $\sigma \in (10 \text{ MHz}, 20 \text{ MHz})$.

In the centre-left panel of Fig. 4, we show a thin slice of the 21-cm signal modelling set ($\mathbf{M}_{21}^{\text{GAUSS}}$) generated with the aforementioned priors. The bold line represents the realization used in the mock observation (T_{21}^{mock}) with $A = 0.155 \text{ K}$, $\nu_0 = 85 \text{ MHz}$, and $\sigma = 15 \text{ MHz}$.

Physical 21-cm signal model

To simulate the physically motivated sky-averaged 21-cm signal, we use accelerated reionization era simulations (ARES⁴), which is a semi-analytical code based on 1D radiative transfer (Mirocha et al. 2012; Mirocha 2014; Mirocha, Furlanetto & Sun 2017). Most sky-averaged 21-cm signal models are seminumerical and can generate a realization of the 21-cm signal within minutes to hours (Mesinger et al. 2010; Fialkov & Barkana 2014; Ghara, Choudhury & Datta 2015; Murray et al. 2020; Hutter et al. 2023; Schaeffer, Giri & Schneider 2023); however, this is computationally infeasible for an analysis that requires millions of samples. Conversely, the semi-analytical approach of ARES produces a realization of the 21-cm signal within a few seconds. Its efficiency stems from evolving the mean radiation background directly, rather than averaging over large cosmological volumes. For a detailed description of how ARES

⁴<https://ares.readthedocs.io/en/ares-dev/>

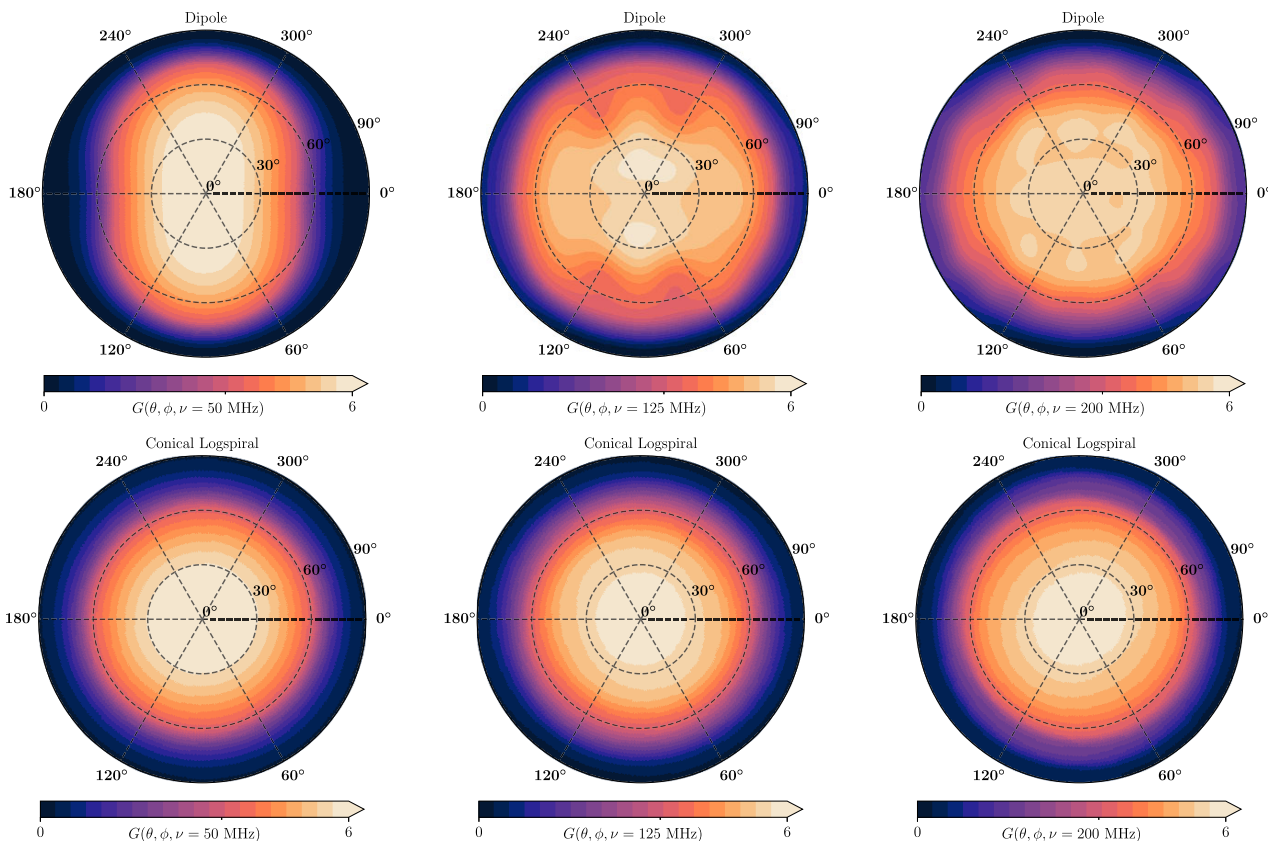


Figure 3. Beam patterns $G(\theta, \phi, \nu)$ for the dipole (top row) and conical log spiral (bottom row) antenna at $\nu = 50$ MHz (first column), 125 MHz (second column), and 200 MHz (third column) shown in the co-latitude (θ) and azimuth (ϕ) coordinates of the antenna's frame.

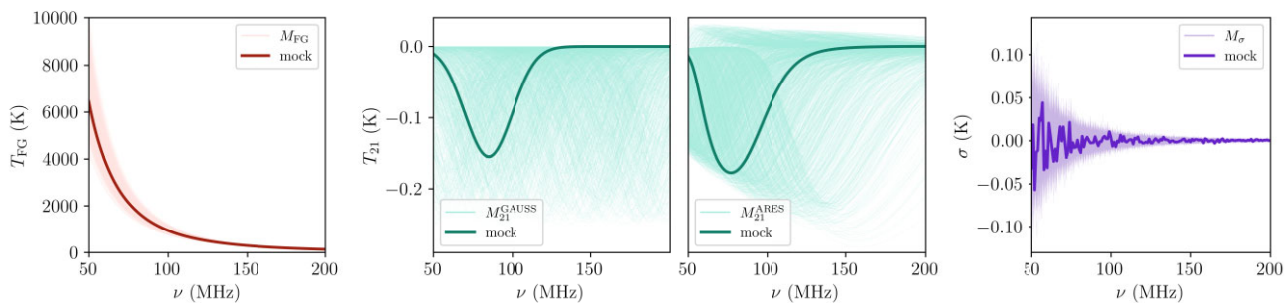


Figure 4. Left panel: The beam-weighted foreground modeling set (M_{FG}) for a dipole antenna. Centre panel: The sky-averaged 21-cm signal training set for a Gaussian profile (left) and a physical model (right). Right panel: Radiometric noise realizations (M_{σ}) for each component of the training set. The bold lines in all panels represent the realization used in the mock observation.

models the 21-cm signal, we refer the interested reader to section 2 of Mirocha et al. (2017).

In our ARES model, we consider three astrophysical parameters that define the shape and amplitude of the cosmological signal: (i) c_X , which is the normalization factor of X-ray luminosity–SFR relation and governs the efficiency of galaxies in producing X-ray photons, (ii) f_{esc} , which is the escape fraction of ionizing photons, and (iii) T_{vir} , the minimum virial temperature of a collapsed halo to form star-forming galaxies. We adopt a log-uniform prior on $c_X \in (10^{36}, 10^{41}) \text{ erg s}^{-1} (\text{M}_{\odot} \text{ yr}^{-1})^{-1}$, a uniform prior on $f_{\text{esc}} \in (0.1, 1)$, and a log-uniform prior on $T_{\text{vir}} \in (3 \times 10^2 \text{ K}, 5 \times 10^5 \text{ K})$. The 21-cm signal modelling set (M_{21}^{ARES}) generated with these priors is shown in the centre-right panel of Fig. 4. The dark green line shows the realiza-

tion used in mock observation with $c_X = 10^{38} \text{ erg s}^{-1} (\text{M}_{\odot} \text{ yr}^{-1})^{-1}$, $f_{\text{esc}} = 0.5$, and $T_{\text{vir}} = 10^4 \text{ K}$.

4 RESULTS

In this section, we discuss the application of the TMNRE algorithm to simultaneously constrain the beam-weighted foregrounds and 21-cm signal for our simulated mock observations. We first delve into its application for the current setup of the REACH experiment in Section 4.1 and subsequently explore potential avenues to further improve the constraints in Section 4.2.

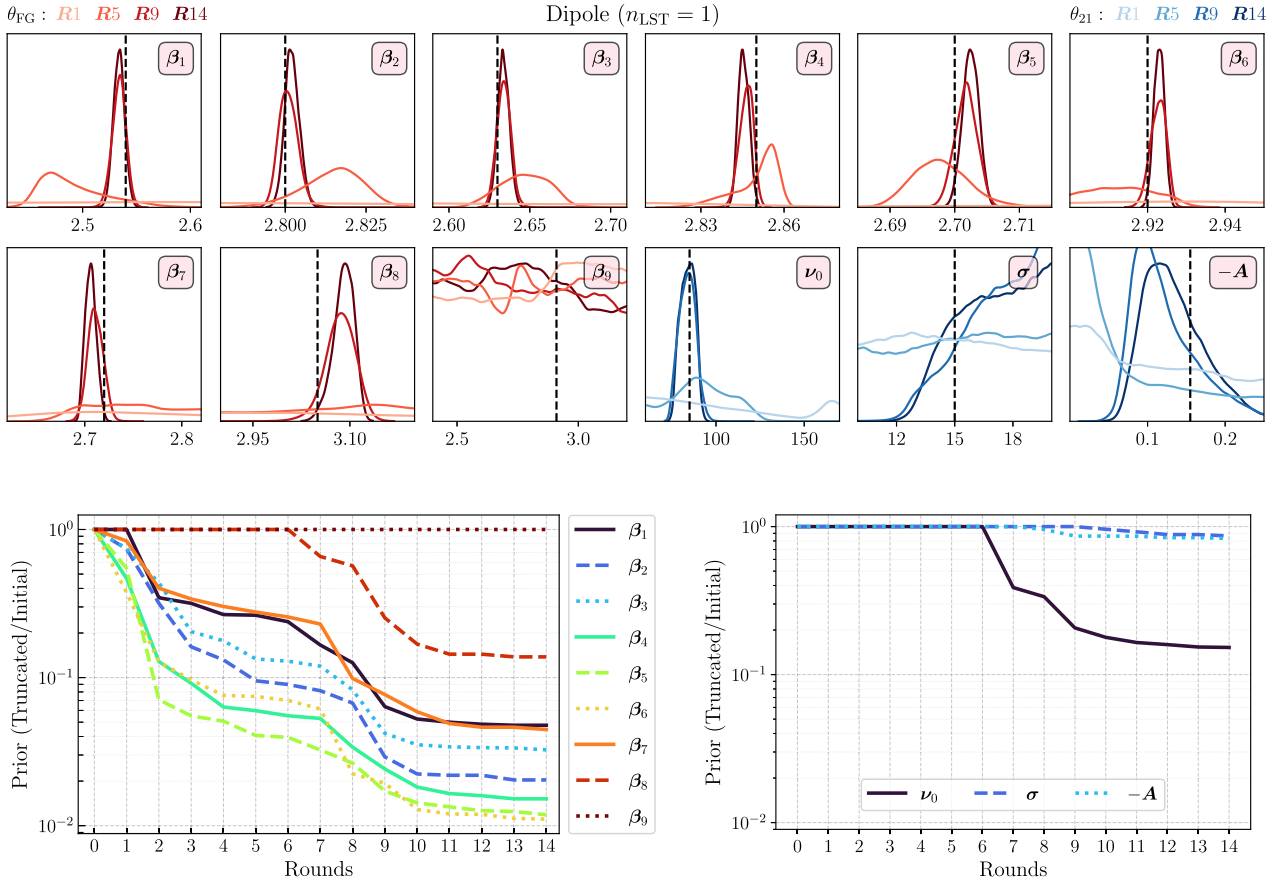


Figure 5. Application of TMNRE on the simulated mock observation for a dipole antenna. Top panel: 1D marginal posterior distribution of foreground and 21-cm signal parameters evaluated after rounds 1, 5, 9, and 14 (shown by different shades). The dashed line represents the true values of the parameters. Bottom panel: The evolution of truncated prior to the initial prior ratio of the foreground (left) and 21-cm parameters (right) as the network learns over multiple rounds.

4.1 The existing setup of REACH

REACH, in its current configuration, is observing the sky with a hexagonal dipole antenna located in the Karoo Radio Reserve in South Africa. In this section, we explore the robustness of this setup to simultaneously constrain the foregrounds and 21-cm signal from the mock data.

Gaussian 21-cm signal model

We first train our neural ratio estimator, assuming a Gaussian profile for the 21-cm signal. In Fig. 5, we show the truncation of prior for the foreground and 21-cm signal parameters. The top panel shows the 1D marginal posteriors after rounds 1, 5, 9, and 14 (shown by different shades from light to dark). This figure demonstrates how the TMNRE algorithm performs a targeted inference through truncation. After the first round of training, only a few foreground parameters prior distributions get truncated significantly, for example, β_4 , β_5 , and β_6 . This is expected because the sky regions corresponding to these spectral indices cover the most significant portion of the sky (25.5 per cent, 31.2 per cent, 20.5 per cent, respectively) and substantially impact the observed sky-averaged brightness temperature. However, after this truncation, the network is able to learn other parameters in the following rounds. This can be seen in the bottom panel of Fig. 5, where we show the truncated prior to initial prior ratio for each foreground (left) and 21-cm signal parameter (right).

The 21-cm signal parameters are learned by the network after the proposal distributions of all foreground parameters undergo a significant truncation. After seven rounds of training, the network is able to truncate the proposal distribution of the central frequency ν_0 of the 21-cm signal. The algorithm converges once we see no further truncation of priors in the following rounds. After fourteen rounds of training, the network can correctly identify the central frequency ν_0 of the cosmological signal along with relatively weak constraints on the absorption amplitude $-A$. However, the width of the signal σ remains unconstrained. Once the algorithm converges, the ratio of truncated prior volume to initial prior volume becomes $\sim 10^{-14}$.

In order to validate our inference results, one way is to compare them with traditional sampling-based methods, which are extremely costly. Therefore, it becomes crucial to have additional methods to cross-check and validate our inference (Hermans et al. 2022; Lemos et al. 2023). In SBI, this is commonly achieved by estimating the statistical coverage of the trained network. We test the statistical properties of our inference by quantifying the nominal and empirical expected coverage probabilities. The reasoning behind this test is that given a set of n independent and identically distributed samples $(x_i, \theta_i^*) \sim p(x, \theta)$, the x per cent credible interval for the posterior should contain the ground truth θ_i^* in x per cent of the n samples for a well-calibrated posterior distribution. In practice, once the algorithm is converged, we take a large number of samples $n = 2000$ from the truncated prior and perform inference on each sample. The local amortization in `swyft` (Miller et al. 2021) allows us to

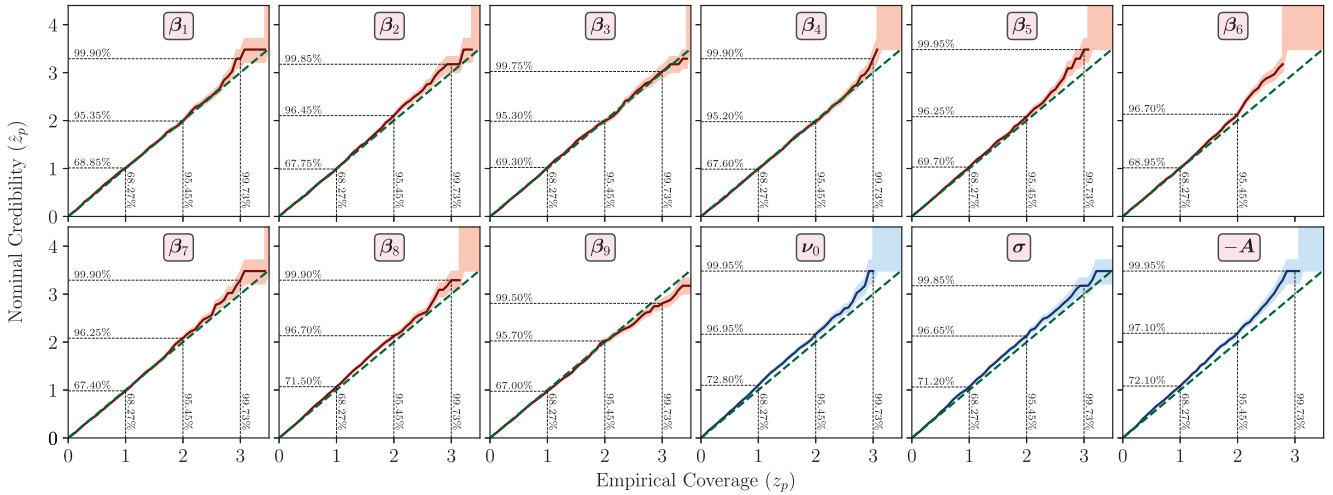


Figure 6. Nominal credibility of the converged network from Fig. 5 as a function of empirical coverage for all foreground and 21-cm signal parameters. The diagonal dashed line represents a perfect coverage.

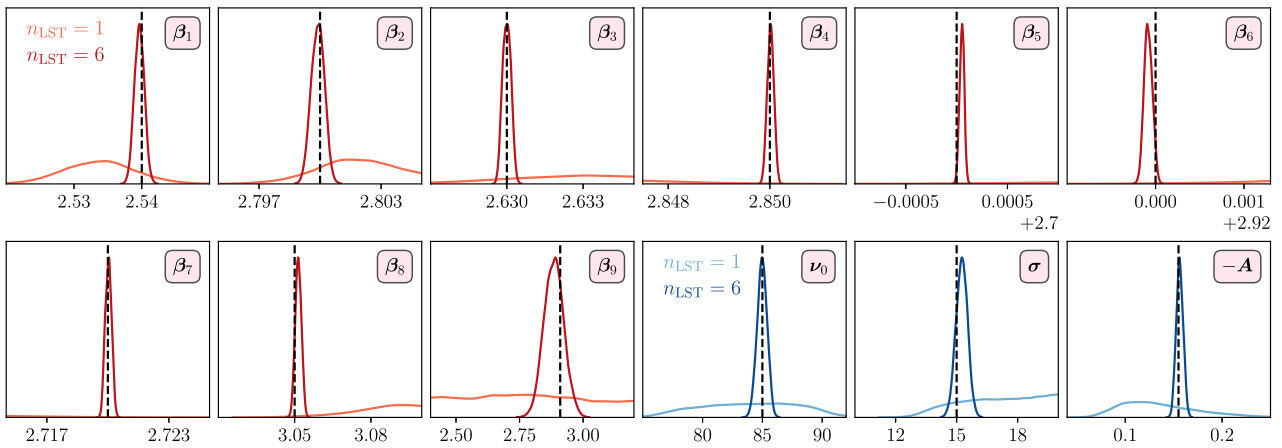


Figure 7. Utilization of drift scan spectra: 1D marginal posteriors of foreground and 21-cm signal parameters for a dipole antenna when the data are analysed from a single time slice of 12 h ($n_{\text{LST}}=1$) and six time slices of 2 h each ($n_{\text{LST}}=6$). These posteriors are generated once the algorithm is converged for each case. The dashed lines represent the ground truth.

quickly generate posteriors for any given sample. In Fig. 6, we show the nominal credibility as a function of the empirical coverage for all model parameters. The uncertainty on the nominal credibility follows from the finite number of samples (n) and is estimated using the Jeffreys interval (Cole et al. 2022). The diagonal dashed line indicates a perfect coverage. Across all parameters, we find an excellent posterior coverage.

In the analysis of the data from a single time slice ($n_{\text{LST}}=1$), as shown in Fig. 5, the constraints on the 21-cm signal parameters are weak because of a significant co-variance between the beam-weighted foregrounds and the cosmological 21-cm signal. The overlap between these two components can be reduced by utilizing the time dependence of the foregrounds (Liu et al. 2013; Switzer & Liu 2014; Tauscher et al. 2020b; Anstey et al. 2023; Saxena et al. 2023a). This can be achieved by fitting the data from multiple time slices, and utilizing the fact that the foregrounds change from one time slice to another, but the 21-cm signal appears identically in each time slice. For this purpose, we simultaneously fit the data $T_{\text{obs}}(\nu)$ from six time slices by giving it as the input to the multilayer perceptron.

In Fig. 7, we compare the posterior distribution of the model parameters from a single time slice (light) and six time slices (dark). We note that the integration time is 12 h for both scenarios. We find that simultaneously analysing multiple time slices improves the parameter constraints significantly. By utilizing the time dependence of the foregrounds, the network is able to constrain β_9 , which is relatively hard to infer since the region corresponding to this parameter covers a tiny portion (0.04 per cent) of the sky. Also, the algorithm can now correctly identify the width (σ) of the 21-cm signal. In this case, the ratio of truncated prior to initial prior volume goes down to $\sim 10^{-26}$ compared to $\sim 10^{-14}$ for a single time slice. Moreover, by increasing the information content, the algorithm converges after seven rounds of training, which is computationally more efficient than the former case.

Physical 21-cm signal model

Next, we focus on a physically motivated model of the sky-averaged 21-cm signal, ARES. As discussed earlier, our 21-cm signal model is composed of three astrophysical parameters (c_X , f_{esc} , T_{vir}), which is

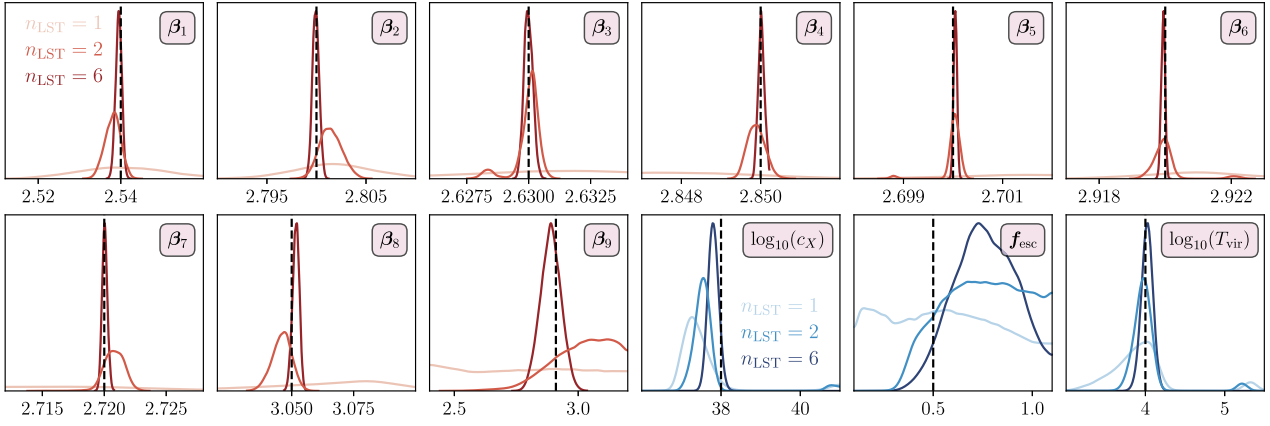


Figure 8. Application of TMNRE to a physically motivated model of the 21-cm signal: 1D marginal posteriors of foreground and 21-cm signal parameters for a dipole antenna when the data is analysed using $n_{\text{LST}} = 1, 2,$ and 6 . The total integration time is set at 12 h, which is then evenly divided across multiple time slices. These posteriors are generated once the algorithm is converged for each case.

then used in the simulated data analysis. Such models are computationally quite expensive to be explored in the traditional likelihood-based methods because the fitting procedure requires millions of samples. However, this problem can be circumvented within an SBI-based framework, thanks to the efficiency of such algorithms. To increase the efficiency of our algorithm in this case, we first generate training samples of the 21-cm signal from the initial prior range. As we perform the TMNRE, we do not apply truncation on the 21-cm signal parameters during the initial rounds of training until all the foreground parameters undergo a significant truncation. In this way, we re-use the samples of 21-cm signal during the initial rounds of training. Once the foreground parameters begin to converge, we apply the truncation on 21-cm signal parameters. This approach used 500 000 21-cm signal samples to estimate the posteriors with the data from a single time slice, which is an order of magnitude fewer compared to a likelihood-based framework (Anstey et al. 2021).

In Fig. 8, we present the posterior distribution of our model parameters for the ARES 21-cm signal model when the data are analysed for $n_{\text{LST}} = 1, 2,$ and 6 (from light to dark). Applying TMNRE on the data from a single time slice, the network is able to constrain the efficiency of X-ray photon production c_X and minimum virial temperature of a halo to host star-forming galaxies T_{vir} , with no constraints on the escape fraction of ionizing photons f_{esc} . The constraints on c_X and T_{vir} are further improved with the inclusion of multiple time slices with relatively weak constraints on f_{esc} . This is expected because the escape fraction of UV photons f_{esc} weakly affects the emission part of the cosmological signal. Moreover, for $n_{\text{LST}} = 1$ and 2 , the posteriors of c_X and T_{vir} are bi-modal due to degeneracy between the foreground and 21-cm signal parameters (as shown in Appendix A). This degeneracy is mitigated with the addition of more time slices, particularly when $n_{\text{LST}} = 6$.

4.2 Future possibilities

In this section, we discuss potential avenues for the REACH experiment. As mentioned in Lera Acedo et al. (2022), REACH in Phase II will feature more antennas to observe the southern hemisphere sky for a better understanding and isolation of the signal components associated with hardware systematics. The follow-up configurations might include, for example, two completely different antennas or one antenna rotated at some angle (θ_{rot}) with respect to the other. We analyse the following four cases in terms of constraining the

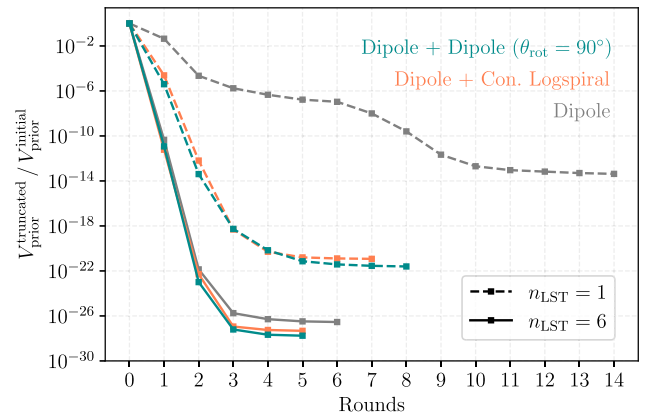


Figure 9. Ratio of truncated prior to initial prior volume for different possible antenna configurations for the upcoming phases of REACH. These are compared with the current configuration of a single dipole antenna with $n_{\text{LST}} = 1$ and $n_{\text{LST}} = 6$.

foregrounds and the cosmological 21-cm signal, and compare with the current configuration of REACH.

- (i) Dipole + Dipole ($\theta_{\text{rot}} = 90^\circ$) with $n_{\text{LST}} = 1$
- (ii) Dipole + Dipole ($\theta_{\text{rot}} = 90^\circ$) with $n_{\text{LST}} = 6$
- (iii) Dipole + Conical Logspiral with $n_{\text{LST}} = 1$
- (iv) Dipole + Conical Logspiral with $n_{\text{LST}} = 6$

In each case, we apply the TMNRE algorithm, where we concatenate the data from different antennas and time slices, which is given as input to the MLP. In a more traditional approach, one has to work with a beam and time-dependent likelihood function. In Fig. 9, we show the evolution of the ratio of truncated prior volume to initial prior volume for each case until the algorithm converges. These possible configurations are compared with the current configuration of REACH (a single dipole antenna). We note that for a fair comparison, the total integration time is 12 h for each case, i.e. for multiple antennae (and time) analysis, it is uniformly distributed for each antenna (and time). From the figure, we notice that for a single time slice ($n_{\text{LST}} = 1$), adding another dipole antenna rotated at 90° or a conical log spiral antenna to the existing setup of REACH reduces the truncated prior to initial prior volume by seven orders of magnitude resulting in improved parameter constraints.

This improvement comes from the fact that the beam-weighted foregrounds are different for different antennas; however, the 21-cm signal remains the same (Anstey et al. 2023; Saxena et al. 2023a).

In the case of multiple time slices ($n_{\text{LST}} = 6$), we notice an order of magnitude reduction in the truncated prior volume by adding a different antenna in the analysis. This improvement is not as significant compared to a single time slice because a significant covariance between the foregrounds and the 21-cm signal has already been reduced by simultaneously fitting multiple time slices. Another interesting thing to note is that for multiple antennas (and time slices), the algorithm converges faster than a single antenna (and a single time slice). This is achieved because the network sees distinct information corresponding to the same set of parameters.

5 SUMMARY

The data analysis of any sky-averaged 21-cm signal experiment generally involves high dimensional parameter spaces to properly take into account the foregrounds, antenna beam uncertainties, ionospheric effects, receiver noise, and the cosmological signal itself. For such highly complex models, the traditional likelihood-based methods of fitting the observed data could become computationally quite expensive. Moreover, given that the traditional Bayesian fitting, i.e. sampling the full joint posterior, typically requires millions of samples, it is not feasible to include physically motivated models of the 21-cm signal in the data analysis.

In this article, we have shown how SBI provides a step towards evading these issues. We considered simulated mock observations composed of foregrounds, the 21-cm signal, and a radiometric noise component. We then performed SBI through a technique known as TMNRE in which neural networks approximate the likelihood-to-evidence ratio. The two key aspects of this approach (i) *marginalization* i.e. directly estimating the marginal posteriors and (ii) *truncation* of wide priors in multiple rounds based on the posterior densities at the end of each round, significantly reduce the simulation budget and leads to a targeted inference. These two features make our approach computationally more efficient and scalable to high-dimensional parameter spaces.

We first demonstrated the foregrounds and 21-cm signal reconstruction for the current setup of the REACH experiment, which is observing the sky with a hexagonal dipole antenna. We first considered a Gaussian 21-cm signal model, which is generally used as an approximation for the cosmological 21-cm signal. We found that the network first learns the foreground parameters, truncates their prior distribution, and later identifies the 21-cm signal. However, the constraints on some of the 21-cm signal parameters are found to be weak when we analysed the data from a single time slice ($n_{\text{LST}} = 1$) of 12 h. This was the consequence of a large co-variance between the beam-weighted foregrounds and the 21-cm signal. This overlap is significantly reduced by analysing the data from multiple time slices and utilizing the fact that the beam-weighted foregrounds but not the 21-cm signal would be different for different time slices. A simultaneous fitting of multiple time slices ($n_{\text{LST}} = 6$) reduces the truncated prior volume by ~ 12 orders of magnitude in comparison to $n_{\text{LST}} = 1$.

We further presented a validation of our inference results through a statistical coverage test by quantifying the nominal and expected coverage probabilities of our trained network. We found excellent posterior coverage, indicating that the network is neither conservative nor overconfident and that we have a well-calibrated posterior distribution of model parameters. We then applied our algorithm to a physically motivated 21-cm signal model, which is computationally rather expensive to explore within a likelihood-

based framework. However, the efficiency of the SBI approaches allows one to use such complicated models. This is crucial because the resulting parameter constraints can be directly linked to the astrophysics of the early Universe.

Finally, we discussed potential paths for the REACH experiment in Phase II, where we applied our algorithm to several different configurations involving a second observing antenna in addition to the current hexagonal dipole. This also leads to an improvement in the parameter constraints due to the beam-weighted foregrounds being different for different instruments but the 21-cm signal being the same. We found that simultaneously analysing the data from multiple antennas and time slices would maximize the information content and thus lead to a better search for the cosmological signal.

The signal extraction, as presented in this work, could be further improved with the inclusion of instruments sensitive to the polarization of the sky radiation to reduce further the co-variance between the foregrounds and cosmological signal (Tauscher et al. 2018). Our approach is flexible enough to marginalize over nuisance parameters, which describe the ionospheric distortions, RFI, errors in the foreground map (Pagano et al. 2023), and beam uncertainties (Cumner et al. 2023). These effects will be considered in future work.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

ACKNOWLEDGEMENTS

We thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Hábrók high performance computing cluster. PDM acknowledges support from the Netherlands Organization for Scientific Research (NWO) VIDI grant (dossier 639.042.730). CW is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 864035 – Undark). The project has been partially funded by the Netherlands eScience Center, grant number ETEC.2019.018.

REFERENCES

- Alsing J., Charnock T., Feeney S., Wandelt B., 2019, *MNRAS*, 488, 4440
 Alvey J., Bhardwaj U., Domcke V., Pieroni M., Weniger C., 2023a, *Phys. Rev. D*, 109, 083008
 Alvey J., Bhardwaj U., Nissanke S., Weniger C., 2023b, preprint (arXiv:2308.06318)
 Anstey D., Leeney S. A. K., 2023, *RASTI*, 3, 372
 Anstey D., de Lera Acedo E., Handley W., 2021, *MNRAS*, 506, 2041
 Anstey D., de Lera Acedo E., Handley W., 2023, *MNRAS*, 520, 850
 Barkana R., 2018, *Nature*, 555, 71
 Barkana R., Loeb A., 2001, *Phys. Rep.*, 349, 125
 Barry N. et al., 2019, *ApJ*, 884, 1
 Becker R. H. et al., 2001, *AJ*, 122, 2850
 Berkhout L. M. et al., 2024, *PASP*, 136, 045002
 Berlin A., Hooper D., Krnjaic G., McDermott S. D., 2018, *Phys. Rev. Lett.*, 121, 011102
 Bernardi G., McQuinn M., Greenhill L. J., 2015, *ApJ*, 799, 90
 Bernardi G. et al., 2016, *MNRAS*, 461, 2847
 Bevins H. T. J., Handley W. J., Fialkov A., de Lera Acedo E., Greenhill L. J., Price D. C., 2021a, *MNRAS*, 502, 4405
 Bevins H. T. J., Handley W. J., Fialkov A., de Lera Acedo E., Javid K., 2021b, *MNRAS*, 508, 2923
 Bhardwaj U., Alvey J., Miller B. K., Nissanke S., Weniger C., 2023, *Phys. Rev. D*, 108, 042004

- Boera E., Becker G. D., Bolton J. S., Nasir F., 2019, *ApJ*, 872, 101
- Bowman J. D., Rogers A. E. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, *Nature*, 555, 67
- Breitman D., Mesinger A., Murray S. G., Prelogović D., Qin Y., Trotta R., 2024, *MNRAS*, 527, 9833
- Burns J. et al., 2021, preprint (arXiv:2103.05085)
- Bye C. H., Portillo S. K. N., Fialkov A., 2022, *ApJ*, 930, 79
- Cohen A., Fialkov A., Barkana R., Lotem M., 2017, *MNRAS*, 472, 1915
- Cohen A., Fialkov A., Barkana R., Monsalve R. A., 2020, *MNRAS*, 495, 4845
- 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., 2023, *MNRAS*, 527, 66
- Cranmer K., Brehmer J., Louppe G., 2020, *Proc. Natl Acad. Sci.*, 117, 30055
- Cumner J. et al., 2022, *J. Astron. Instrum.*, 11, 2250001
- Cumner J., Pieterse C., Villiers D. D., de Lera Acedo E., 2024, *MNRAS*, 531, 4734
- Dax M., Green S. R., Gair J., Deistler M., Schölkopf B., Macke J. H., 2023, preprint (arXiv:2111.13139)
- DeBoer D. R. et al., 2017, *PASP*, 129, 045001
- Deshpande A. A., 2018, *ApJ*, 866, L7
- Dijkstra M., 2014, *Publ. Astron. Soc. Aust.*, 31, e040
- Ewall-Wice A., Chang T. C., Lazio J., Doré O., Seiffert M., Monsalve R. A., 2018, *ApJ*, 868, 63
- Ewall-Wice A., Chang T.-C., Lazio T. J. W., 2020, *MNRAS*, 492, 6086
- Fan X. et al., 2003, *AJ*, 125, 1649
- Feng C., Holder G., 2018, *ApJ*, 858, L17
- Fialkov A., Barkana R., 2014, *MNRAS*, 445, 213
- Fialkov A., Barkana R., 2019, *MNRAS*, 486, 1763
- Furlanetto S. R., Oh S. P., Briggs F. H., 2006, *Phys. Rep.*, 433, 181
- Gangolli N., D'Aloisio A., Nasir F., Zheng Z., 2020, *MNRAS*, 501, 5294
- Ghara R., Choudhury T. R., Datta K. K., 2015, *MNRAS*, 447, 1806
- Handley W. J., Hobson M. P., Lasenby A. N., 2015, *MNRAS*, 450, L61
- 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
- Hills R., Kulkarni G., Meerburg P. D., Puchwein E., 2018, *Nature*, 564, E32
- Hutter A., Trebitsch M., Dayal P., Gottlöber S., Yepes G., Legrand L., 2023, *MNRAS*, 524, 6124
- Ignatov Y. D., Pritchard J. R., Wu Y., 2023, *MNRAS*, 527, 11206
- Jensen H., Laursen P., Mellema G., Iliev I. T., Sommer-Larsen J., Shapiro P. R., 2012, *MNRAS*, 428, 1366
- Jones J. D., Rapetti D., Mirocha J., Hibbard J. J., Burns J. O., Bassett N., 2023, *ApJ*, 959, 49
- Kaplinghat M., Chu M., Haiman Z., Holder G. P., Knox L., Skordis C., 2003, *ApJ*, 583, 24
- Karchev K., Trotta R., Weniger C., 2023, preprint (arXiv:2311.15650)
- Karchev K., Grayling M., Boyd B. M., Trotta R., Mandel K. S., Weniger C., 2024, *MNRAS*, 530, 3881
- Komatsu E. et al., 2011, *ApJS*, 192, 18
- Koopmans L. et al., 2015, in Bourke T. L., eds, Proc. Sci., Advancing Astrophysics with the SquareKilometre Array (AASKA14). SISSA, Trieste, PoS#1
- Leeney S. A. K., Handley W. J., de Lera Acedo E., 2023, *Phys. Rev. D*, 108, 062006
- Lemos P., Coogan A., Hezaveh Y., Perreault-Levasseur L., 2023, Proc. 40th International Conference on Machine Learning, 202, 19256
- de Lera Acedo E. et al., 2022, *Nat. Astron.*, 6, 984
- Li W. et al., 2019, *ApJ*, 887, 141
- Lin K., von wietersheim Kramsta M., Joachimi B., Feeney S., 2023, *MNRAS*, 524, 6167
- Liu A., Pritchard J. R., Tegmark M., Loeb A., 2013, *Phys. Rev. D*, 87, 043002
- Liu H., Outmezguine N. J., Redigolo D., Volansky T., 2019, *Phys. Rev. D*, 100, 123011
- Lueckmann J.-M., Boelts J., Greenberg D. S., Gonçalves P. J., Macke J. H., 2021, preprint (arXiv:2101.04653)
- McEwen J. D., Wallis C. G. R., Price M. A., Mancini A. S., 2023, preprint (arXiv:2111.12720)
- Mellema G., Koopmans L., Shukla H., Datta K. K., Mesinger A., Majumdar S., 2015, in Bourke T. L., ed., Proc. Sci., Advancing Astrophysics with the Square Kilometre Array (AASKA14). SISSA, Trieste
- Mertens F. G. et al., 2020, *MNRAS*, 493, 1662
- Mertens F. G., Semelin B., Koopmans L. V. E., 2021, in Siebert A., ed., SF2A-2021: Proc. Annual meeting of the French Society of Astronomy and Astrophysics. p. 211
- Mesinger A., Furlanetto S., Cen R., 2010, *MNRAS*, 411, 955
- Miller B. K., Cole A., Forré P., Louppe G., Weniger C., 2021, *Advances in Neural Information Processing Systems*, 34, 129
- Mirocha J., 2014, *MNRAS*, 443, 1211
- Mirocha J., Skory S., Burns J. O., Wise J. H., 2012, *ApJ*, 756, 94
- Mirocha J., Furlanetto S. R., Sun G., 2017, *MNRAS*, 464, 1365
- Monsalve R. A. et al., 2024, *MNRAS*, 530, 4125
- Montel N. A., Weniger C., 2022, preprint (arXiv:2211.04291)
- Munshi S. et al., 2024, *A&A*, 681, A62
- Murray S. G., Greig B., Mesinger A., noz J. B. M., Qin Y., Park J., Watkinson C. A., 2020, *J. Open Source Softw.*, 5, 2582
- Nhan B. D., Bradley R. F., Burns J. O., 2017, *ApJ*, 836, 90
- Nhan B. D., Bordenave D. D., Bradley R. F., Burns J. O., Tauscher K., Rapetti D., Klima P. J., 2019, *ApJ*, 883, 126
- de Oliveira-Costa A., Tegmark M., Gaensler B. M., Jonas J., Landecker T. L., Reich P., 2008, *MNRAS*, 388, 247
- Paciga G. et al., 2013, *MNRAS*, 433, 639
- Pagano M., Sims P., Liu A., Anstey D., Handley W., de Lera Acedo E., 2023, *MNRAS*, 527, 5649
- Papamakarios G., Murray I., 2018, preprint (arXiv:1605.06376)
- Philip L. et al., 2018, *J. Astron. Instrument.*, 8, 1950004
- Planck Collaboration VI, 2020, *A&A*, 641, A6
- Price D. C. et al., 2018, *MNRAS*, 478, 4193
- Pritchard J. R., Loeb A., 2012, *Rep. Prog. Phys.*, 75, 086901
- Razavi-Ghods N. et al., 2023, preprint (arXiv:2307.00099)
- Reis I., Fialkov A., Barkana R., 2020, *MNRAS*, 499, 5993
- Reis I., Fialkov A., Barkana R., 2021, *MNRAS*, 506, 5479
- Santos M. G., Ferramacho L., Silva M. B., Amblard A., Cooray A., 2010, *MNRAS*, 406, 2421
- Saxena A., Meerburg P. D., de Lera Acedo E., Handley W., Koopmans L. V. E., 2023a, *MNRAS*, 522, 1022
- Saxena A., Cole A., Gazagnes S., Meerburg P. D., Weniger C., Witte S. J., 2023b, *MNRAS*, 525, 6097
- Schaeffer T., Giri S. K., Schneider A., 2023, *MNRAS*, 526, 2942
- Scheutwinkel K. H., Handley W., de Lera Acedo E., 2023, *Publ. Astron. Soc. Aust.*, 40, e016
- Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, *A&A*, 345, 380
- Shen E., Anstey D., de Lera Acedo E., Fialkov A., Handley W., 2021, *MNRAS*, 503, 344
- Sims P. H., Pober J. C., 2020, *MNRAS*, 492, 22
- Singh S., Subrahmanyan R., 2019, *ApJ*, 880, 26
- Singh S. et al., 2018, *ApJ*, 858, 54
- Singh S. et al., 2021, *Nat. Astron.*, 6, 607
- Slosar A., 2017, *Phys. Rev. Lett.*, 118, 151301
- Spinelli M., Bernardi G., Garsden H., Greenhill L. J., Fialkov A., Dowell J., Price D. C., 2021, *MNRAS*, 505, 1575
- Spurio Mancini A., Docherty M. M., Price M. A., McEwen J. D., 2023, *RAS Techn. Instrum.*, 2, 710
- Switzer E. R., Liu A., 2014, *ApJ*, 793, 102
- Tauscher K., Rapetti D., Burns J. O., Switzer E., 2018, *ApJ*, 853, 187
- Tauscher K., Rapetti D., Burns J. O., 2020a, *ApJ*, 897, 132
- Tauscher K., Rapetti D., Burns J. O., 2020b, *ApJ*, 897, 175
- Tauscher K. et al., 2021, *ApJ*, 915, 66
- Tejero-Cantero A., Boelts J., Deistler M., Lueckmann J.-M., Durkan C., Gonçalves P. J., Greenberg D. S., Macke J. H., 2020, *J. Open Source Softw.*, 5, 2505
- Thomas R. M. et al., 2009, *MNRAS*, 393, 32

- Vedantham H. K., Koopmans L. V. E., de Bruyn A. G., Wijnholds S. J., Ciardi B., Brentjens M. A., 2013, *MNRAS*, 437, 1056
- Voytek T. C., Natarajan A., Jáuregui García J. M., Peterson J. B., López-Cruz O., 2014, *ApJ*, 782, L9
- Zeghal J., Lanusse F., Boucaud A., Remy B., Aubourg E., 2022, Proc. 39th International Conference on Machine Learning (ICML 2022), Machine Learning for Astrophysics. p. 52

APPENDIX A: 2D MARGINAL POSTERiors

In this section, we explore the application of TMNRE to investigate correlations among model parameters through the estimation of 2D marginal posteriors. Specifically, we input pairs of parameters

along with the data into a multi-layer perceptron (MLP) to estimate the likelihood-to-evidence ratio for each parameter pair. Fig. A1 presents the 1D and 2D marginal posteriors for the simulated mock observation for a hexagonal dipole antenna and a single time slice ($n_{\text{LST}} = 1$). The results reveal significant correlations among the foreground parameters i.e. spectral indices. Furthermore, several spectral indices are also correlated with the 21-cm signal parameters in our physically motivated model. Notably, the parameters $\log_{10}(c_X)$ and $\log_{10}(T_{\text{vir}})$ exhibit a bi-modal distribution. These parameter degeneracies could be resolved by leveraging the time and beam dependence of the foregrounds, which reduces the correlations between the foregrounds and the 21-cm signal, as shown in Fig. 9.

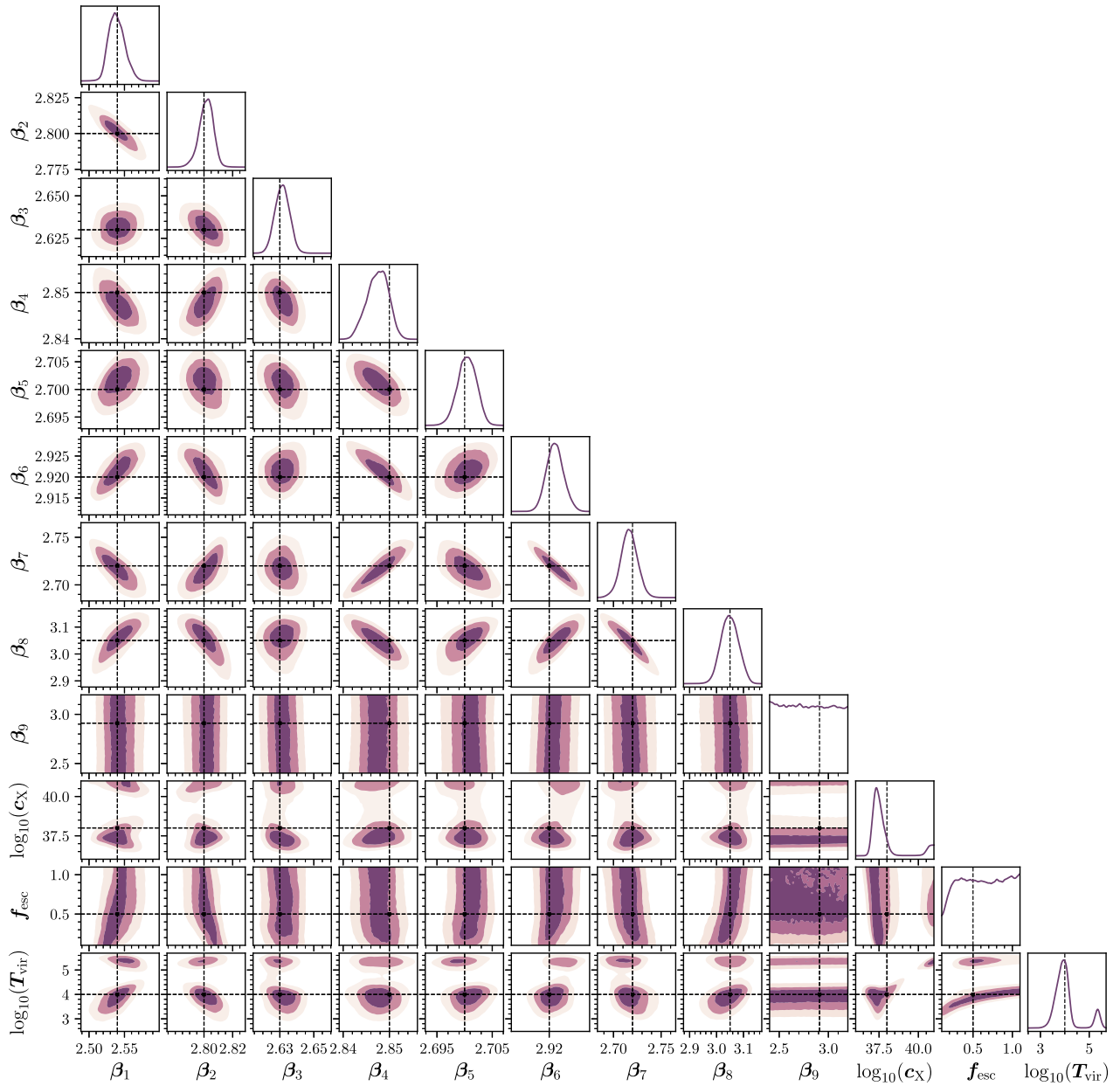


Figure A1. Recovered 1D and 2D marginals on the foreground and the 21-cm signal parameters for the simulated observation for a hexagonal dipole antenna and a single time slice integration. The dashed lines represent the ground truth.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.