Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **533,** 2530–2545 (2024) https://doi.org/10.1093/mnras/stae1984
Advance Access publication 2024 August 17

Inferring astrophysical parameters using the 2D cylindrical power spectrum from reionization

Bradley Greig [1,2,3★] David Prelogović,[4] Yuxiang Qin [2,3] Yuan-Sen Ting(丁源森) [1,5,6,7,8] and Andrei Mesinger [4]

[1]*Research School of Astronomy & Astrophysics, Australian National University, Canberra, ACT 2611, Australia*
[2]*School of Physics, University of Melbourne, Parkville, VIC 3010, Australia*
[3]*ARC Centre of Excellence for All-Sky Astrophysics in 3 Dimensions (ASTRO 3D)*
[4]*Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56125 Pisa, Italy*
[5]*School of Computing, Australian National University, Acton, ACT 2601, Australia*
[6]*Department of Astronomy, The Ohio State University, Columbus, OH 45701, USA*
[7]*Center for Cosmology and AstroParticle Physics (CCAPP), The Ohio State University, Columbus, OH 43210, USA*
[8]*Faculty of Science, Department of Physics, Universiti Malaya, Kuala Lumpur 50603, Malaysia*

Accepted 2024 August 12. Received 2024 July 16; in original form 2024 March 20

ABSTRACT

Enlightening our understanding of the first galaxies responsible for driving reionization requires detecting the 21-cm signal from neutral hydrogen. Interpreting the wealth of information embedded in this signal requires Bayesian inference. Parameter inference from the 21-cm signal is primarily restricted to the spherically averaged power spectrum (1D PS) owing to its relatively straightforward derivation of an analytic likelihood function enabling traditional Monte Carlo Markov Chain approaches. However, in recent years, simulation-based inference (SBI) has become feasible which removes the necessity of having an analytic likelihood, enabling more complex summary statistics of the 21-cm signal to be used for Bayesian inference. In this work, we use SBI, specifically marginal neural ratio estimation to learn the likelihood-to-evidence ratio with SWYFT, to explore parameter inference using the cylindrically averaged 2D PS. Since the 21-cm signal is anisotropic, the 2D PS should yield more constraining information compared to the 1D PS which isotropically averages the signal. For this, we consider a mock 1000 h observation of the 21-cm signal using the Square Kilometre Array and compare the performance of the 2D PS relative to the 1D PS. Additionally, we explore two separate foreground mitigation strategies, perfect foreground removal and wedge avoidance. We find the 2D PS outperforms the 1D PS by improving the marginalized uncertainties on individual astrophysical parameters by up to $\sim$ 30–40 per cent irrespective of the foreground mitigation strategy. Primarily, these improvements stem from how the 2D PS distinguishes between the transverse, $k_\perp$, and redshift-dependent, $k_\parallel$, information which enables greater sensitivity to the complex reionization morphology.

Key words: galaxies: high-redshift – intergalactic medium – dark ages, reionization, first stars – diffuse radiation – early Universe – cosmology: theory.

1 INTRODUCTION

Roughly 400 000 yr after the big bang recombination occurs, whereby the photons and baryons of the primordial plasma decouple and the baryons subsequently combine to form neutral hydrogen. After this point the omnipresence of neutral hydrogen enshrouds the Universe in a fog rendering it opaque to most forms of radiation. This fog persists until the ignition of the first star formation episodes within the primordial galaxies, referred to as the cosmic dawn, which emit copious amounts of ultraviolet (UV) photons into the intergalactic medium (IGM) and ionize their local neighbourhood. Over time, as these galaxies grow and become more abundant their cumulative UV output accelerates the eradication of this fog, rendering the IGM essentially completely ionized. This process is referred to as the epoch of reionization (EoR).

Directly observing these primordial galaxies is near to impossible. Their vast distance from us makes them extremely faint, and this is prior to the extinction of their radiation by the neutral IGM. All however is not lost. The primordial galaxies imprint their signal on the neutral IGM, enabling us to indirectly infer their presence by measuring the 21-cm hyperfine spin-flip transition of the neutral hydrogen. This signal is observable by detecting the differential intensity of radiation emitted by the neutral hydrogen relative to a uniform background source, for example, the cosmic microwave background (see e.g. Gnedin & Ostriker 1997; Madau, Meiksin & Rees 1997; Shaver et al. 1999; Tozzi et al. 2000; Gnedin & Shaver 2004; Furlanetto, Oh & Briggs 2006; Morales & Wyithe 2010; Pritchard & Loeb 2012). As this frequency (redshift) dependent signal originates from the IGM, detecting it yields a three-dimensional,

★ E-mail: brad.s.greig@gmail.com

time-evolving picture of the thermal and ionization state of the IGM throughout reionization.

Accessing the wealth of information embedded in the 3D cosmic 21-cm signal requires large-scale radio interferometer experiments to pick up the spatially varying signal. Specifically, we measure the complex visibilities of the interference fringes from the arriving signal which are naturally represented by a Fourier transform. This signal can then be split into two components, $k_{\parallel}$ which corresponds to the line-of-sight (frequency) dependent aspect of the signal and $k_{\perp}$ which describes the 2D spatial variation of the signal across the sky. Typically, given the cosmic 21-cm signal is incredibly faint relative to the bright foregrounds, in order to boost its signal to noise we compress the available information by measuring the 1D spherically averaged power spectrum (PS). This characterizes the statistical properties of the 21-cm signal by describing the variance as a function of spatial scale. It is this quantity that is sought by the first-generation interferometer experiments such as the Low-Frequency Array (van Haarlem et al. 2013), the Murchison Wide Field Array (Tingay et al. 2013; Wayth et al. 2018), the Precision Array for Probing the Epoch of Reionization (Parsons et al. 2010), the Owens Valley Radio Observatory Long Wavelength Array (Eastwood et al. 2019), and the upgraded Giant Metrewave Radio Telescope (Gupta et al. 2017).

Importantly, the 21-cm signal is non-Gaussian owing to the complex 3D morphology of the ionized regions. Therefore, computing the 1D PS performs suboptimal compression as we disregard valuable non-Gaussian information. The lower sensitivity of these first generation of experiments necessitates the trade-off in information loss in order to boost the overall signal to noise with the 1D PS. However, for the forthcoming Square Kilometre Array (SKA; Mellema et al. 2013; Koopmans et al. 2015) this should no longer be a concern with the increased sensitivity theoretically enabling more complex summary statistics of the 21-cm signal beyond the 1D PS. In fact, the SKA has been specifically designed with tomographic imaging in mind, opening up a wealth of possibilities for analysing the 21-cm signal.

This has lead to the explosion of studies exploring alternative probes of the 21-cm signal beyond the 1D PS. For example, with the bispectrum (e.g. Yoshiura et al. 2015; Shimabukuro et al. 2016; Majumdar et al. 2018, 2020; Watkinson et al. 2019; Hutter et al. 2021; Kamran et al. 2021), position-dependent PS (Giri et al. 2019a), one-point statistics (Watkinson & Pritchard 2014; Shimabukuro et al. 2015; Kubota et al. 2016; Banet et al. 2021; Gorce, Hutter & Pritchard 2021), morphological and topographical coefficients extracted from 21-cm images (e.g. Yoshiura et al. 2017; Bag et al. 2019; Chen et al. 2019; Elbers & van de Weygaert 2019; Kapahtia, Chingangbam & Appleby 2019; Gazagnes, Koopmans & Wilkinson 2021; Giri & Mellema 2021; Kapahtia et al. 2021), the ionized bubble size distribution (Kakiichi et al. 2017; Giri et al. 2018a, 2019b; Giri, Mellema & Ghara 2018b; Bianco et al. 2021), and the compression of 21-cm images using the wavelet scattering transform (e.g. Greig, Ting & Kaurov 2022; Hothi et al. 2024).

Importantly, in order to extract the astrophysical properties of these first galaxies we must perform Bayesian inference. Typically, this demands generating 3D reionization simulations on-the-fly within a Monte Carlo Markov Chain (MCMC) framework (e.g. 21CMMC Greig & Mesinger 2015, 2017, 2018; Park et al. 2019) to compare against an observation of the 21-cm signal. However, this approach is extremely restrictive as it requires defining an analytic expression to compute the likelihood of the 21-cm signal given the model astrophysical parameter set. As a result, essentially none of the aforementioned alternative statistics to the 1D PS have been explored rigorously within a Bayesian inference context. Instead, most resort to the Fisher information matrix (Fisher 1935), which still imposes an implicit Gaussian assumption, to provide simple forecasts (e.g. Shimabukuro et al. 2017; Greig et al. 2022; Hothi et al. 2024) or adopt several simplifying assumptions regarding the covariance and likelihood form (Tiwari et al. 2022; Watkinson, Greig & Mesinger 2022). Alternatively, instead of performing direct inference one can apply regression in an attempt to extract astrophysical information from the 21-cm signal by bypassing the need for a summary statistic entirely through the application of convolutional neural networks (CNNs) trained directly on 2D or 3D images of the 21-cm signal to extract astrophysical information (e.g. Gillet et al. 2019; Hassan et al. 2019; La Plante & Ntampaka 2019; Hassan, Andrianomena & Doughty 2020; Kwon, Hong & Park 2020; Mangena, Hassan & Santos 2020; Prelogović et al. 2022).

However, we can completely forego this restrictive requirement of an analytic likelihood through the concept of likelihood-free or simulation-based inference (SBI; see e.g. Cranmer, Brehmer & Louppe 2020 for a recent review). Essentially, we apply machine learning on a training set of simulated data to learn our likelihood function (neural likelihood estimation) or the likelihood-to-evidence ratio (neural ratio estimation; NRE) after which we can perform an MCMC to obtain our posteriors or bypass the likelihood estimation entirely to directly obtain our posterior distribution given our data (neural posterior estimation). The power of such an approach is that it enables us to explore any complex or non-Gaussian summary statistic or feature extraction method applied to the 21-cm signal. Our only requirement is the generation of the simulated training set. Realizing this potential, in recent years SBI has been gaining traction for tackling astrophysical inference from the cosmic 21-cm signal (e.g. Zhao et al. 2022a; Zhao, Mao & Wandelt 2022b; Prelogović & Mesinger 2023; Saxena et al. 2023; Greig et al. 2024).

In this work we choose to explore the often overlooked 2D cylindrically averaged (2D PS) for astrophysical parameter inference.[1] Its previous omission stemmed from our inability to define a robust likelihood function along with the additional computational overheads required for estimating the 21-cm signal covariance. Specifically, we follow the approach of our companion work (Greig et al. 2024) and perform our SBI using marginal neural ratio estimation (MNRE; Miller et al. 2021) using the publicly available PYTHON package SWYFT[2] (Miller et al. 2022) whose goal is to learn the marginal likelihood-to-evidence ratios for each individual parameter. Rather than spherically averaging over the $k_{\parallel}$ and $k_{\perp}$ components of the signal into a single k and losing morphological information, the 2D PS keeps these components distinct, better separating out the 2D structural information from the spatial fluctuations across the sky ($k_{\perp}$) from the frequency varying component along the line of sight, $k_{\parallel}$. Further, the 2D PS more naturally follows the signal characteristics obtained from radio interferometry allowing us to more cleanly deal with foreground contamination (e.g. the 'wedge'). Although it is still a Gaussian statistic, and therefore still suboptimal, it should lose less information than the 1D PS. In fact, in a complimentary study by Prelogović & Mesinger (2024) exploring the information content of a variety of 21-cm summaries using Fisher Matrices these authors predict improvements in the variance on the astrophysical

[1]Mondal et al. (2022) performed an initial exploratory analysis of the multifrequency angular power spectrum (MAPS), which has some analogies to the 2D PS. For a basic three-parameter astrophysical model the MAPS was found to outperform the 1D PS.

[2]https://github.com/undark-lab/swyft

parameters of ~ 15 per cent. Further, the 2D PS is considerably more straightforward to measure observationally and should require less integration time to achieve sufficient sensitivity in comparison to many of the aforementioned non-Gaussian approaches. Therefore, it is an important and valuable summary statistic to explore.

The remainder of this paper is organized as follows. In Section 2 we summarize our 21-cm simulations using 21CMFAST and in Section 3 we describe our SBI set-up with SWYFT including the generation of our data base of 21-cm simulations and our mock observation. In Section 4 we then perform our comparison of the 2D PS to the 1D PS for different foreground mitigation strategies before concluding with our closing remarks in Section 5. Unless stated otherwise, all quantities are in co-moving units and we adopt the cosmological parameters: (Ω_Λ, Ω_M, Ω_b, n, σ_8, H_0) = (0.69, 0.31, 0.048, 0.97, 0.81, 68 km s^{-1} Mpc^{-1}), consistent with recent results from the Planck mission (Planck Collaboration VI 2020).

2 SIMULATING THE 21-CM SIGNAL

To simulate the 3D cosmic 21-cm signal emanating from the neutral hydrogen during reionization we use the seminumerical simulation code 21CMFAST[3] (Mesinger & Furlanetto 2007; Mesinger, Furlanetto & Cen 2011). In particular, we use the latest public release, v3 (Murray et al. 2020), and adopt the Park et al. (2019) flexible galaxy parametrization to describe the UV and X-ray properties of the galaxy population. In this section we outline the main ingredients of 21CMFAST, in particular focussing on the astrophysical parameters within the model we seek to constrain using parameter inference. For additional details and discussions we refer the reader to these earlier publications.

2.1 Galaxy UV properties

First, it is assumed that the stellar mass, M_*, of a galaxy depends on its host halo mass, M_h (e.g. Kuhlen & Faucher-Giguère 2012; Dayal et al. 2014; Behroozi & Silk 2015; Mitra, Choudhury & Ferrara 2015; Mutch et al. 2016; Ocvirk et al. 2016; Sun & Furlanetto 2016; Yue, Ferrara & Xu 2016; Hutter et al. 2021) via the following relation:

$$M_*(M_h) = f_* \left(\frac{\Omega_b}{\Omega_m}\right) M_h, \tag{1}$$

with f_* being the fraction of galactic gas in stars and Ω_b and Ω_m being the baryonic and total matter content of the Universe. f_* also depends on its host halo mass,

$$f_* = f_{*,10} \left(\frac{M_h}{10^{10}\,M_\odot}\right)^{\alpha_*}, \tag{2}$$

dependent on the two free parameters, α_* and its normalization, $f_{*,10}$, for a dark matter halo mass of $10^{10}M_\odot$. This power-law behaviour directly follows from semi-empirical fits to observations (e.g. Harikane et al. 2016; Tacchella et al. 2018; Behroozi et al. 2019; Stefanon et al. 2021) and semi-analytic model predictions (e.g Mutch et al. 2016; Yung et al. 2019; Hutter et al. 2021).

The stellar mass is then converted into a star formation rate (SFR) by dividing by a characteristic time-scale, t_*, which is a free parameter of the model and is defined to be a fraction, $t_* \in [0.05, 1]$, of the Hubble time, $H^{-1}(z)$:

$$\dot{M}_*(M_h, z) = \frac{M_*}{t_* H^{-1}(z)}. \tag{3}$$

[3]https://github.com/21cmfast/21cmFAST

Similarly as above, the fraction of UV photons that escape their host galaxy and enter into the IGM, f_{esc}, also depend on their host halo mass,

$$f_{esc} = f_{esc,10} \left(\frac{M_h}{10^{10}\,M_\odot}\right)^{\alpha_{esc}}, \tag{4}$$

giving rise to an additional two free parameters, α_{esc} and $f_{esc,10}$.

Not all dark matter haloes can contribute to reionization. Internal feedback mechanisms and/or inefficient gas cooling can suppress star formation in low-mass haloes. This behaviour is parametrized via an effective duty-cycle:

$$f_{duty} = \exp\left(-\frac{M_{turn}}{M_h}\right), \tag{5}$$

with $(1 - f_{duty})$ defining the fraction of star-forming galaxies that are suppressed below a characteristic mass scale M_{turn} (e.g. Giroux, Sutherland & Shull 1994; Shapiro, Giroux & Babul 1994; Hui & Gnedin 1997; Barkana & Loeb 2001; Springel & Hernquist 2003; Mesinger & Dijkstra 2008; Okamoto, Gao & Theuns 2008; Sobacchi & Mesinger 2013a, b).

2.2 Galaxy X-ray properties

In addition to contributing the UV photons responsible for driving reionization, the first galaxies also emit X-ray photons which escape and heat the cold IGM gas. The origin of the X-ray photons is thought to be stellar remnants left over from earlier star formation episodes. To model the X-ray heating caused by these energetic photons 21CMFAST computes a cell-by-cell angle-averaged specific X-ray intensity, $J(\boldsymbol{x}, E, z)$ (in erg s^{-1} keV^{-1} cm^{-2} sr^{-1}),

$$J(\boldsymbol{x}, E, z) = \frac{(1+z)^3}{4\pi} \int_z^\infty \mathrm{d}z' \frac{c\mathrm{d}t}{\mathrm{d}z'} \epsilon_X e^{-\tau}, \tag{6}$$

by integrating the co-moving X-ray specific emissivity, $\epsilon_X(\boldsymbol{x}, E_e, z')$, back along the light-cone accounting for IGM attenuation, $e^{-\tau}$. The specific emitted emissivity, $E_e = E(1+z')/(1+z)$, is then

$$\epsilon_X(\boldsymbol{x}, E_e, z') = \frac{L_X}{\mathrm{SFR}} \left[(1+\bar{\delta}_{nl}) \int_0^\infty \mathrm{d}M_h \frac{\mathrm{d}n}{\mathrm{d}M_h} f_{duty} \dot{M}_* \right], \tag{7}$$

where $\bar{\delta}_{nl}$ is the mean, non-linear overdensity in a shell centred on the simulation cell $(\boldsymbol{x}, z)$ and the quantity in square brackets is the SFR density along the light-cone with $\frac{\mathrm{d}n}{\mathrm{d}M_h}$ corresponding to the halo mass function (HMF).[4] The quantity L_X/SFR (erg s^{-1} keV$^{-1}M_\odot^{-1}$ yr) is the specific X-ray luminosity per unit star formation escaping the host galaxies which depends on the spectral energy distribution describing the source of X-rays, $L_X \propto E^{-\alpha_X}$. Throughout, we adopt $\alpha_X = 1$, consistent with local Universe observations of high-mass X-ray binaries (e.g. Mineo, Gilfanov & Sunyaev 2012; Fragos et al. 2013; Pacucci et al. 2014).

Finally, we normalize L_X/SFR by its integrated soft-band (< 2 keV) luminosity per SFR (in erg s$^{-1}M_\odot^{-1}$ yr),

$$L_{X<2\,\mathrm{keV}}/\mathrm{SFR} = \int_{E_0}^{2\,\mathrm{keV}} \mathrm{d}E_e\, L_X/\mathrm{SFR}, \tag{8}$$

with E_0 denoting the minimum X-ray photon energy capable of escaping the host galaxy into the IGM.

[4]Throughout this work we adopt the Sheth–Tormen HMF (Sheth, Mo & Tormen 2001) as our fiducial HMF.

2.3 Ionization and thermal state of the IGM

The thermal state of the IGM is computed via the IGM spin temperature, $T_{\rm S}$, which is determined by self-consistently computing the heating and ionization rates owing to structure formation, Compton scattering off CMB photons, heating following partial ionizations, and X-ray heating and ionizations. To calculate $T_{\rm S}$ we determine its weighted mean,

$$T_{\rm S}^{-1} = \frac{T_{\rm CMB}^{-1} + x_\alpha T_\alpha^{-1} + x_{\rm c} T_{\rm K}^{-1}}{1 + x_\alpha + x_{\rm c}}, \tag{9}$$

where $T_{\rm K}$, T_α, and $T_{\rm CMB}$ are the gas, Lyman α (Ly α) colour, and CMB temperatures. T_S depends on the local gas density and Ly α radiation intensity, with the Ly α background sourced by the cumulative sum of X-ray excitations of neutral hydrogen atoms and direct stellar emission of Lyman band photons by the first galaxies. The quantities x_α and $x_{\rm c}$ are the coupling coefficients for the Wouthuysen–Field mechanism (Wouthuysen 1952; Field 1958) and between the free elections and CMB photons, respectively.

Calculating the 3D ionization of the IGM requires the application of excursion-set theory (Furlanetto, Zaldarriaga & Hernquist 2004) on the evolved density field. This compares the cumulative number of ionizing photons, $n_{\rm ion}$, to the total number of neutral hydrogen atoms plus cumulative recombinations, $\bar{n}_{\rm rec}$ (Sobacchi & Mesinger 2014) within spheres of decreasing radii, R, and corresponding overdensity, δ_R. Evaluated within each individual simulation voxel, a voxel is deemed to be ionized when

$$\bar{n}_{\rm ion}(\boldsymbol{x}, z|R, \delta_R) \geq (1 + \bar{n}_{\rm rec})(1 - \bar{x}_e), \tag{10}$$

where the $(1 - \bar{x}_e)$ factor includes the contribution of X-rays to ionizations and

$$n_{\rm ion} = \bar{\rho}_b^{-1} \int_0^\infty {\rm d}M_{\rm h} \frac{{\rm d}n(M_h, z|R, \delta_R)}{{\rm d}M_{\rm h}} f_{\rm duty} \dot{M}_* f_{\rm esc} N_{\gamma/b}. \tag{11}$$

Here, $\bar{\rho}_b$ is the mean baryon density and $N_{\gamma/b}$ is the total number of ionizing photons produced per stellar baryon.[5]

2.4 21-cm brightness temperature

The quantity we measure observationally is the brightness temperature, $\delta T_{\rm b}(\nu)$, the differential intensity of the neutral hydrogen illuminated by the CMB (Furlanetto et al. 2006),

$$\delta T_{\rm b}(\nu) = \frac{T_{\rm S} - T_{\rm CMB}(z)}{1 + z} \left(1 - {\rm e}^{-\tau_{\nu_0}}\right) \ {\rm mK}, \tag{12}$$

and

$$\tau_{\nu_0} \propto (1 + \delta_{\rm nl})(1 + z)^{3/2} \frac{x_{\rm H\,I}}{T_{\rm S}} \left(\frac{H}{{\rm d}v_{\rm r}/{\rm d}r + H}\right), \tag{13}$$

where τ_{ν_0} is the optical depth of the neutral hydrogen which depends on the local gas overdensity, $\delta_{\rm nl} \equiv \rho/\bar{\rho} - 1$, the neutral hydrogen fraction, $x_{\rm H\,I}$, the Hubble parameter, $H(z)$, and the line-of-sight gradient of the peculiar velocity. For simplicity the spatial dependence of the quantities have been omitted and it is evaluated at the redshift $z = \nu_0/\nu - 1$.

[5]By default this is assumed to be $N_{\gamma/b} = 5000$ consistent with a Salpeter initial mass function (Salpeter 1955).

3 SBI SET-UP

3.1 Parameter inference with SWYFT

In parameter inference, the quantity of interest is the posterior, $p(\boldsymbol{\theta}|\boldsymbol{x})$, which describes the probability distribution of obtaining our model parameters, $\boldsymbol{\theta}$, given an observation, $\boldsymbol{x}$. This characterizes the best set of model parameters for describing the given data. This posterior is computed following Bayes' theorem,

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x})} p(\boldsymbol{\theta}), \tag{14}$$

where $p(\boldsymbol{x}|\boldsymbol{\theta})$ is the likelihood to obtain our observation given our set of model parameters, $p(\boldsymbol{\theta})$ characterizes our prior knowledge of reasonable values for our model parameters, and $p(\boldsymbol{x})$ is the evidence of the data.

The basic idea of SBI is to replace the explicit likelihood evaluation with a stochastic simulator of the signal. With this, we generate a training set of data–parameter pairs, $[(\boldsymbol{x}_1, \boldsymbol{\theta}_1), ...(\boldsymbol{x}_N, \boldsymbol{\theta}_N)]$, which are drawn from our prior distribution and connects our model parameters to the observed data. We then train a neural network on these data to estimate either the posterior, the likelihood, or the likelihood-to-evidence ratio. The advantage of these approaches is that we no longer require any assumptions on the form of the likelihood enabling any complex summary statistic to be explored, provided we can compute it in our forward-modelled simulations.

In this work, we perform SBI using SWYFT (Miller et al. 2022). Specifically, it performs MNRE (e.g. Durkan, Murray & Papamakarios 2020; Hermans et al. 2021) to approximate the marginal likelihood-to-evidence ratio for any individual parameter or 2D parameter pair [denoted $\tilde{\boldsymbol{\theta}}$ to signify any parameter pair; i.e (θ_i, θ_j) rather than the likelihood-to-evidence ratio of the entire parameter set]. Denoting $r(\boldsymbol{x}, \tilde{\boldsymbol{\theta}})$ to be this marginal likelihood-to-evidence ratio:

$$r(\boldsymbol{x}, \tilde{\boldsymbol{\theta}}) \equiv \frac{p(\boldsymbol{x}|\tilde{\boldsymbol{\theta}})}{p(\boldsymbol{x})} = \frac{p(\tilde{\boldsymbol{\theta}}|\boldsymbol{x})}{p(\tilde{\boldsymbol{\theta}})} = \frac{p(\boldsymbol{x}, \tilde{\boldsymbol{\theta}})}{p(\boldsymbol{x})p(\tilde{\boldsymbol{\theta}})}, \tag{15}$$

which is the ratio of the probability density for a jointly drawn sample–parameter pair, $\boldsymbol{x}, \tilde{\boldsymbol{\theta}} \sim p(\boldsymbol{x}, \tilde{\boldsymbol{\theta}})$ and a marginally pair $\boldsymbol{x}, \tilde{\boldsymbol{\theta}} \sim p(\boldsymbol{x})p(\tilde{\boldsymbol{\theta}})$. This ratio is estimated by training a binary classification network, $d_\phi(\boldsymbol{x}, \tilde{\boldsymbol{\theta}})$, where ϕ describes the network parameters, which distinguishes between two hypotheses: whether the sample–parameter pairs are jointly ($C = 1$) or marginally ($C = 0$) drawn. The binary classifier is trained using a binary-cross entropy loss function:

$$L[d_\phi(\boldsymbol{x}, \tilde{\boldsymbol{\theta}})] = -\int {\rm d}\boldsymbol{x}{\rm d}\tilde{\boldsymbol{\theta}} \left\{ p(\boldsymbol{x}, \tilde{\boldsymbol{\theta}}) \log d_\phi(\boldsymbol{x}, \tilde{\boldsymbol{\theta}}) + p(\boldsymbol{x})p(\tilde{\boldsymbol{\theta}}) \log\left[1 - d_\phi(\boldsymbol{x}, \tilde{\boldsymbol{\theta}})\right] \right\}, \tag{16}$$

which is minimized when $d_\phi(\boldsymbol{x}, \tilde{\boldsymbol{\theta}})$ approximates the probability density of the jointly drawn sample–parameter pair (e.g. $C = 1$). This returns

$$d_\phi(\boldsymbol{x}, \tilde{\boldsymbol{\theta}}) = p(C = 1|\boldsymbol{x}, \tilde{\boldsymbol{\theta}}) = \frac{p(\boldsymbol{x}, \tilde{\boldsymbol{\theta}})}{p(\boldsymbol{x}, \tilde{\boldsymbol{\theta}}) + p(\boldsymbol{x})p(\tilde{\boldsymbol{\theta}})} \equiv \sigma[\log r(\boldsymbol{x}, \tilde{\boldsymbol{\theta}})], \tag{17}$$

where the last equality connects the likelihood-to-evidence ratio, r, to the binary classifier, d_ϕ, using the sigmoid function, $\sigma(y) = [1 + {\rm e}^{-y}]^{-1}$.

As this approach only learns the marginal likelihood-to-evidence ratio for any parameter pair, $\tilde{\boldsymbol{\theta}}$, for an M-dimensional model we are required to train M 1D and $M(M-1)/2$ 2D networks to fully

describe the marginal posterior distribution given an observation. This is because the simulated training set generated by the stochastic simulator inherently contains the variance due to all sampled model parameters. Therefore, the marginalization over the remaining (nuisance) model parameters is always implicitly performed and thus the training of the binary classifier is limited to at most two dimensions.

3.2 Simulated data

Our only requirement for SBI is that our stochastic simulator models the complexities of the cosmic 21-cm signal including the observational characteristics of realistic data and that our data set contains sufficient samples. Below, we summarize the main steps adopted for pipeline based on our previous work (e.g. Greig et al. 2022, 2023, 2024).

We generate 3D realizations of the cosmic 21-cm signal using 21CMFAST, simulated within 250^3 Mpc^3 comoving volumes on a 150^3 grid. The final evolved density fields are downsampled from an initially higher resolution grid, 450^3, after applying second-order Lagrange perturbation theory (e.g Scoccimarro 1998). We track the evolution of the 21-cm signal from $z = 25$ down to $z = 5.2$ and stitch together the comoving simulation cubes via linear interpolation to generate a 21-cm light-cone. Using the same training set as constructed in Greig et al. (2024) we have 150 000 independent realizations of the cosmic 21-cm signal for our forward-modelled training set.[6] Note, for this work, non-linear redshift-space distortions (RSDs) were not included in the simulated 21-cm signal (e.g. Mao et al. 2012; Jensen et al. 2013). Primarily, these RSDs serve to elongate the 21-cm power along the line of sight amplifying the anisotropy of the 21-cm signal. As a result, differences between the 2D PS and 1D PS are likely to be underestimated in this work.

Radio interferometers are only sensitive to the spatial fluctuations in the signal and thus the observed data are zero mean distributed. To mimic this, we first split our 21-cm light-cones into equal comoving distance (250 Mpc) chunks. This choice is adopted in order to measure our 1D and 2D PS using a 3D cubic volume for computational ease. For each of these chunks we then remove the mean signal before adding in the instrumental effects as outlined below. Following Greig et al. (2024), we split our simulated 3D 21-cm light-cone into 10 equal co-moving chunks spanning from $z = 5.7$ to $z = 18.1$, from which we measure either the 1D or 2D PS. Although the SKA is designed to be sensitive down to 50 MHz ($z \sim 27.8$), at these redshifts the thermal noise dominates over our fiducial model therefore we limit the redshift dimension for our data to $z \leq 18$.

3.2.1 *Instrumental noise*

To add interferometric noise along with the finite resolution of the instrument to our simulated 21-cm data we use a modified version of the publicly available PYTHON module 21CMSENSE[7] (Pober et al. 2013, 2014). Provided any antenna configuration 21CMSENSE first generates the corresponding uv-visibility tracks for each sampled baseline before gridding for computational efficiency. Specifically for this work, we use the SKA configuration System Baseline Design document[8] which includes 512 37.5m antennae stations distributed within a 500m core radius. These stations are modelled assuming a system temperature, $T_{\rm sys} = 1.1T_{\rm sky} + 40$ K and a corresponding sky temperature of $T_{\rm sky} = 60\left(\frac{\nu}{300\ \rm MHz}\right)^{-2.55}$ K (Thompson, Moran & Swenson 2007). For our set-up, we assume a total observing time of 1000 h based on a single 6-h phase-tracked scan of the sky per night.

Taking the gridded uv- visibilities as input, 21CMSENSE then computes the total thermal noise power, $P_{\rm N}(k)$;

$$P_{\rm N}(k) \approx X^2Y\frac{\Omega'}{2t}T_{\rm sys}^2, \tag{18}$$

where X^2Y performs the cosmological conversions between observing bandwidth, frequency, and co-moving distance, Ω' is a beam-dependent factor derived by Parsons et al. (2014), and t is the total observing time.

As we are interested in 3D noise realizations rather than the 1D total noise power we perform the following modifications:

(i) We first 3D Fourier transform the input (simulated) mean removed 21-cm data cube.

(ii) We then filter this cube using the gridded uv- visibilities for the SKA computed by 21CMSENSE. Cells with finite uv-coverage are multiplied by unity, all others are set to zero.

(iii) At each cell we then determine the amplitude of the thermal noise, $P_{\rm N}(k_x, k_y, k_z)$, using equation (18) where k_x and k_y correspond to the two transverse (on-sky) directions and k_z is the line-of-sight direction.

(iv) We then add random noise (zero mean with variance based on the PS amplitude in the cell) to each cell to mimic the effect of thermal noise.

(v) Finally, we then 3D inverse Fourier transform back to obtain our noisy 21-cm data.

3.2.2 *The foreground wedge*

Unfortunately, individual uv visibilities from a radio interferometer baseline are frequency-dependent. This means that the line-of-sight (frequency-dependent) power can leak into the transverse (frequency-independent) Fourier modes resulting in a well-defined contaminated 'wedge' in cylindrical 2D Fourier space (Datta, Bowman & Carilli 2010; Morales et al. 2012; Parsons et al. 2012; Trott, Wayth & Tingay 2012; Vedantham, Udaya Shankar & Subrahmanyan 2012; Thyagarajan et al. 2013, 2015a, b; Liu, Parsons & Trott 2014a, b; Pober et al. 2016; Murray & Trott 2018). This gives rise to two separate philosophies for dealing with this wedge contamination: foreground removal and foreground avoidance.

In the first case, we assume that we can mitigate or 'clean' these contaminated modes (see e.g. Chapman & Jelić 2019 for a review, or using machine learning Gagnon-Hartman et al. 2021) enabling us to recover and use the entire 21-cm signal. In the latter case, we conservatively avoid this wedge contaminated region of Fourier space and only use the 'clean' Fourier modes located above this 'wedge'. In this work, we shall consider both scenarios when exploring the 2D PS for parameter inference. Note, in both of these instances, we assume that the impact of foregrounds can be perfectly dealt with, that is, after following the removal or avoidance of the foregrounds we are left with a clean cosmological signal. However,

[6]Note this choice was deliberately chosen to be conservative and in fact we found we could recover comparable posteriors as our main results when using training sets up to 25–50 per cent smaller. Overall, computing the training data used exclusively for this work (i.e. 21-cm PS and 2D PS data using only the Sheth-Tormen HMF) took roughly 150 000 CPU hours.

[7]https://github.com/jpober/21cmSense

[8]http://astronomers.skatelescope.org/wp-content/uploads/2016/09/SKA-TEL-SKO-0000422_02_SKA1_LowConfigurationCoordinates-1.pdf

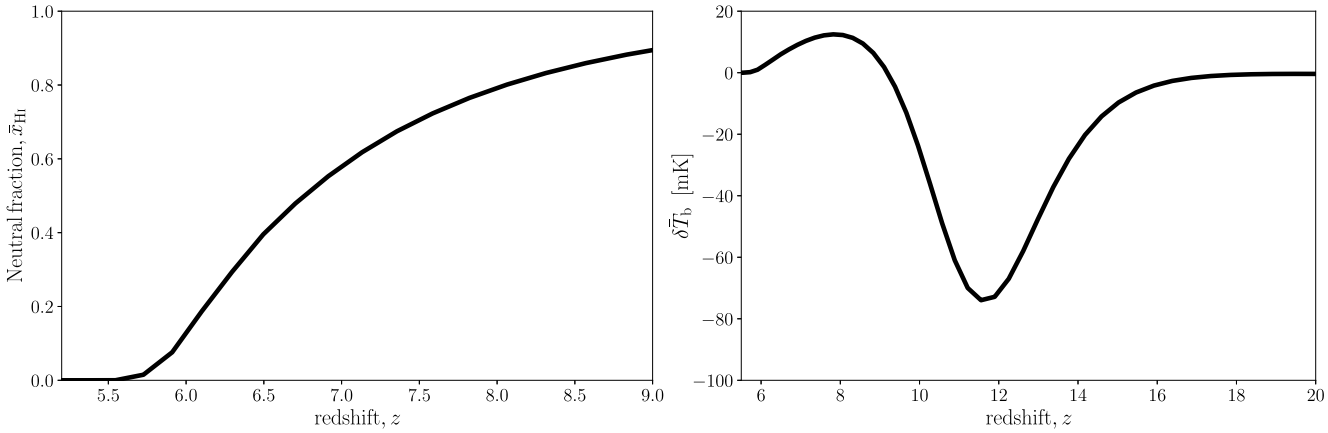

Figure 1. The volume-averaged IGM neutral fraction (left panel) and the mean brightness temperature signal (right panel) for our fiducial astrophysical parameter set used for constructing our mock observation.

this will not be the case for realistic observations which will contain some residual artefacts from the various cleaning algorithms within the data. Currently, there has been few efforts to investigate the consequences of imperfect foreground removal. Nasirudin et al. (2020) demonstrated that performing astrophysical inference using the 21-cm PS in the presence of realistic foregrounds could result in some parameters being strongly biased. More recently, Bianco et al. (2024) explored the ability to correctly identify ionized regions with synthetic 2D images of 21-cm data in the presence of realistic foreground residuals. Using a U-Net, these authors achieved high levels of precision for the ionized region classification when the contrast of the data relative to the noise was fairly high. Nevertheless, considerably more work is required to study the impact of imperfect foreground removal.

While the foreground removal case utilizes the full simulated 21-cm data, performing wedge avoidance requires an additional step to those discussed in the previous section. The boundary defining this foreground 'wedge' in 2D Fourier space is given by

$$k_{\parallel} = mk_{\perp} + b, \tag{19}$$

where $k_{\parallel}$ and $k_{\perp}$ are the line-of-sight and transverse Fourier modes, b is a additive buffer which we assume to be $\Delta k_{\parallel} = 0.1\,h$ Mpc^{-1} which accounts for bleeding of noise extending beyond the horizon limit, and m is the gradient of this boundary given by

$$m = \frac{D_{\rm C} H_0 E(z) \sin(\theta)}{c(1+z)}. \tag{20}$$

This boundary depends on the comoving distance, $D_{\rm C}$, the Hubble constant, H_0, cosmological factor $E(z) = \sqrt{\Omega_{\rm m}(1+z)^3 + \Omega_{\Lambda}}$, and $\sin(\theta)$ denotes the observed viewing angle for our observation, for which we assume as $\theta = \pi/2$ (i.e. a zenith pointing observation).

In order to account for the foreground wedge, we must remove the Fourier modes from below the wedge. Therefore, after 3D Fourier transforming our input 3D 21-cm data cube, we first zero all modes that fall below this foreground 'wedge' before adding the thermal noise for all modes above the wedge.

3.3 Mock 21-cm observation

Exploring the 2D PS for parameter inference requires the construction of a mock observation of the 21-cm signal. For this, we assume the same fiducial parameter set as in Greig et al. (2024). Namely, we select parameters for our UV galaxies in line with the recovered model of Qin et al. (2021) based on Ly α forest observations by Bosman et al. (2018). Below we summarize the individual model parameter values along with their associated flat prior ranges and in Fig. 1 we provide the volume-averaged IGM neutral fraction and mean brightness temperature as a function of redshift:

(i) $\log_{10}(f_{*,10}) = -1.10; \in [-3.0, 0.0]$
(ii) $\alpha_* = 0.5; \in [-0.5, 1.0]$
(iii) $\log_{10}(f_{\rm esc,10}) = -1.30; \in [-3.0, 0.0]$
(iv) $\alpha_* = -0.35; \in [-1.0, 0.5]$
(v) $\log_{10}(M_{\rm turn}) = 8.55; \in [8.0, 10.0]$
(vi) $t_* = 0.5; \in [0.05, 1.0]$
(vii) $L_{\rm X<2\,keV}/{\rm SFR} = 40.50; \in [38.0, 42.0]$
(viii) $E_0 = 0.5; \in [0.1, 1.5]$

In Fig. 2 we compare the 1D and 2D PS for the first four redshift ranges extracted from our mock observation of the 21-cm light-cone. Additionally, we demonstrate the differences in measured PS as a result of the two distinct treatments of the astrophysical foregrounds. For the 1D PS demonstrated in the top row, we distinguish between perfect foreground wedge removal (black) and wedge-avoidance (red). The black vertical dashed lines correspond to the region of the 1D PS between which we use for performing astrophysical parameter inference, namely $k = 0.1$ and 1.0 Mpc^{-1}. The impact of ignoring the foreground contaminated wedge region is clearly evident here, restricting the range of Fourier modes accessible for performing our inference. Beyond the visible removal of modes, there will also be less spherically averaged modes per k-bin for the wedge avoidance case, which will also lead to an increase in the corresponding sample variance uncertainty resulting in broadened inferred astrophysical posteriors.

In the middle panel of Fig. 2 we provide the 2D PS assuming perfect foreground removal, whereas the bottom panel corresponds to the wedge-avoidance scenario. The vertical and horizontal dashed lines correspond to the boundaries for the Fourier modes we shall consider for our parameter inference from the 2D PS. These are selected to roughly balance modes that are adequately sampled by our simulation volume and also to correspond to the same scales as used in our inference pipeline for the 1D PS. Namely, we consider $k_{\perp} = 0.05$ and 0.9 Mpc^{-1} and $k_{\parallel} = 0.08$ and 0.9 Mpc^{-1}. Note, although with these boundaries it is possible to sample modes at $k > 1$ Mpc^{-1} (where $k = \sqrt{(k_{\perp}^2 + k_{\parallel}^2)}$) these modes are the most

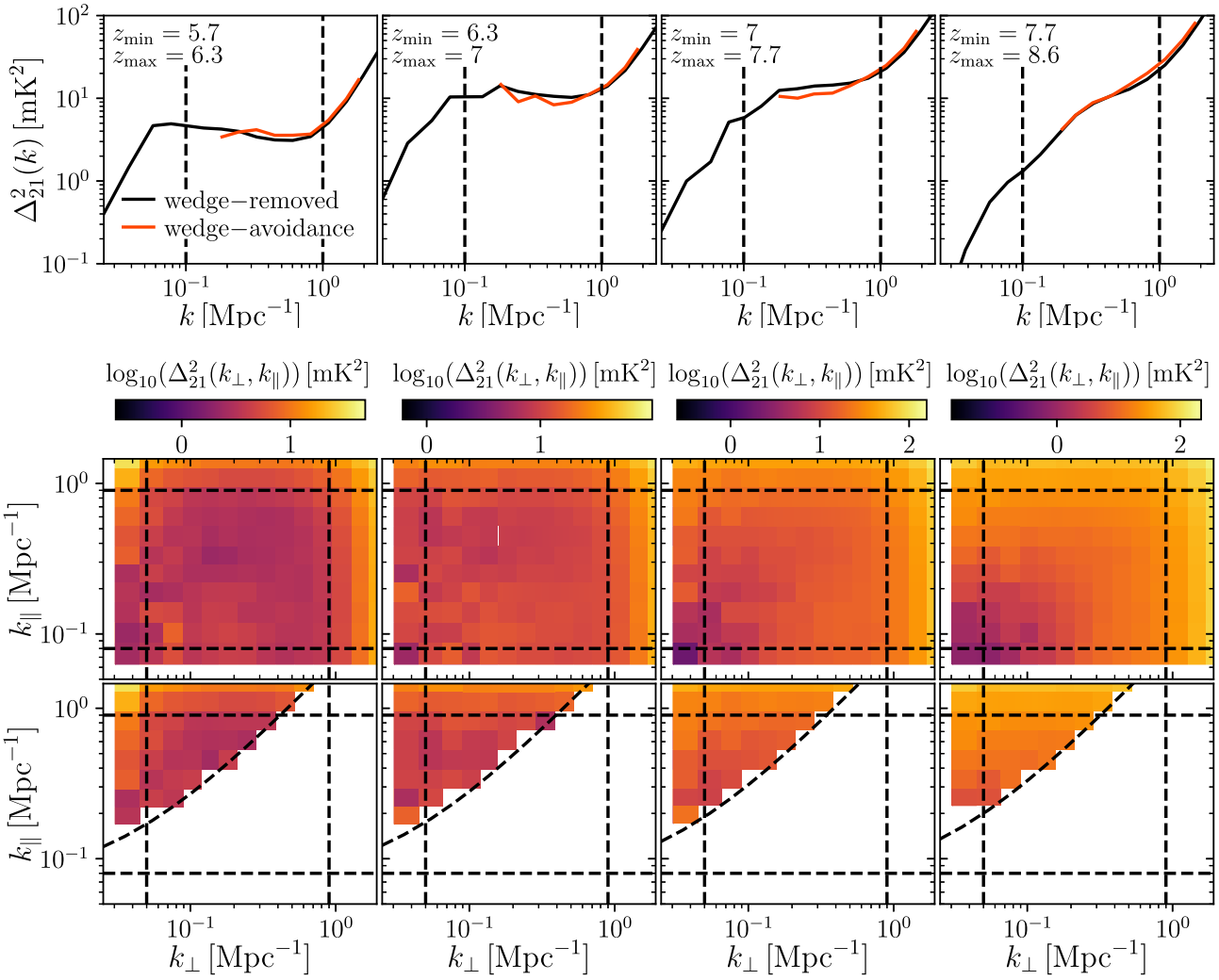

Figure 2. Comparison of the 1D PS and the 2D PS from four different redshift ranges obtained from our simulated 21-cm light-cone. *Top row:* the 1D PS assuming perfect foreground removal (black curve) and after wedge avoidance (red curve). The vertical dashed lines at $k = 0.1$ and 1.0 Mpc^{-1} correspond to the region within which we use for parameter inference. *Middle row:* the 2D PS assuming perfect foreground removal. The vertical dashed lines correspond to $k_\perp = 0.05$ and 0.9 Mpc^{-1} and the horizontal dashed lines correspond to $k_\parallel = 0.08$ and 0.9 Mpc^{-1}. We use all 2D PS information bounded within these regions for our parameter inference. *Bottom row:* the 2D PS after performing wedge avoidance (observing only modes above the wedge denoted by the diagonal black dashed line).

severely affected by instrumental thermal noise and thus will not provide much additional constraining power.

From these two panels it is immediately clear the significant impact that the foreground wedge has on our ability to measure the 21-cm signal. In terms of the 2D PS, the foreground wedge contaminates well over 60 per cent of the Fourier modes contained within our boundaries to be used for parameter inference. Further, this only gets worse for increasing redshifts as the wedge boundary is redshift-dependent. However, what matters is where the information is lost. The vast majority of the information is lost for large $k_\perp$, which is less sensitive to the astrophysical parameters. In the case of the 1D PS, most of the constraining power comes from the 'knee' like feature around $k \sim 0.1$ Mpc^{-1} (see Greig & Mesinger 2015) which corresponds to the typical sizes of the ionized regions. Since the 2D PS still has reasonable sampling of these modes, that is for low $k_\perp$, we should not see such a drastic reduction in constraining power between the two foreground mitigation scenarios as we are still sensitive to the morphological information both during the EoR and in the epoch of X-ray heating (EoH). It will be instructive to quantify the relative difference between the two mitigation strategies.

Note there are several competing effects leading to our inability to access information below $k \sim 0.15$ Mpc^{-1} for the 1D PS. Simply increasing the simulation size does not immediately alleviate the issue. First, as we aim to mimic realistic observations of the 21-cm signal from a radio interferometer, when computing the 1D PS we do not consider the case when $k_\perp = 0$ (i.e. $k = k_\parallel$). These modes are not visible to radio interferometers as the minimum available baseline is the diameter of the receiving element (dish or antennae station). This, coupled with the definition of the foreground wedge, severely limits the spherically binned Fourier modes below $k \sim 0.15$ Mpc^{-1}. For instance, we have our horizon buffer at $0.1h$ Mpc^{-1}, which sets the minimum allowed for $k_\parallel$ (equation 19). Therefore, to obtain $k \sim 0.1$ Mpc^{-1} for wedge-avoidance, we require $k_\perp \lesssim 0.015$ Mpc^{-1} corresponding to a simulation with side-length of at least ∼420 Mpc. However, this is for one single mode, to have a reasonable statistical sampling of $k_\perp \lesssim 0.015$ Mpc^{-1} we would require at least $2-3$

times larger side-lengths, corresponding to $\gtrsim 1$ Gpc. Generating such large simulation volumes for parameter inference is infeasible. Of course, this would be less severe if we removed this additive horizon buffer. Importantly, since we can obtain a reasonable statistical sampling of Fourier modes for our 2D PS in the case of perfect foreground removal, our simulation volumes are sufficient for this analysis.

4 PARAMETER INFERENCE WITH THE 2D PS

4.1 Estimating the posteriors with SBI

In SWYFT, to obtain our desired marginal posterior distributions for our astrophysical parameters we need to construct neural networks to learn the likelihood-to-evidence ratios given our mock observation of the 21-cm signal. Within each of the 10 250 Mpc co-moving chunks of the 21-cm light-cone used to mimic an observation with the SKA we also restrict the number of Fourier modes to be used for our inference pipeline to those between $k = 0.1$ Mpc^{-1} and $k = 1.0$ Mpc^{-1} for both the 1D and 2D PS. This results in a total of 60 1D PS data points when considering wedge avoidance. These data points are then simply taken as the input layer to a three-layered fully connected neural network consisting of 256 neurons. That is, we do not use an embedding network to reduce the dimensionality of the input data prior to the fully connected neural network. For the 1D PS under perfect foreground removal, we instead recover eight Fourier modes within our defined boundary, resulting in a total of 80 1D PS data points. For this, we use the same network architecture, with these 80 data points being the input layer.

For the 2D PS, as one would expect we have considerably more input data. Considering wedge-avoidance, and our corresponding Fourier cuts, we obtain 161 data points for the 2D PS. This is not overly restrictive computationally and thus we retain these data as a linear input layer for our three-layered neural network. On the other hand, assuming perfect foreground removal, we have a total of 900 data points. In this work, we choose to apply a linear transformation to these data to reduce them down to 256 features with which we take as our 1D input to our three-layered neural network, with 512 neurons per layer. That is, we adopt this linear transformation as an embedding network. However, given that the 2D PS naturally lends itself to a 2D image representation of the data, one could instead apply a CNN as our embedding network to more optimally extract the features within the data. For example, Breitman et al. (2024) found that applying a CNN to the 1D PS data represented as a 2D image (k, z) improved the overall performance of their emulator, implying better relative performance at feature extraction. Nevertheless, after exploring several different network architectures, we found this linearization of the data to be sufficient to extract the data, given the number of Fourier modes that are dominated by thermal noise. However, in future we will return to this to perform a more rigorous exploration of optimal network architectures for extracting the relevant features of our data.

For all the above MNRE networks, the training was performed in batches of size 64 along with an initial learning rate of 10^{-3} that is decayed by 0.95 after each epoch. The total training time for each network varied between ~ 1 and 2 h using a single Nvidia A100. Considering more aggressive learning rates, increasing the number of neurons within each neural network layer or more advanced embedding networks (for foreground removal with the 2D PS) did not substantially improve the performance of the resultant MNRE networks.

4.2 Perfect foreground removal

First, we consider the somewhat optimistic case of perfectly removing foreground contamination enabling the full use of the Fourier information (see e.g. Fig. 2). In Fig. 3, we present the 1D and 2D marginalized posteriors following our SBI approach with SWYFT for our mock 21-cm observation. For this we demonstrate the resultant posteriors for the 2D (1D) PS by the black (red) curves, respectively. Below the marginalized 1D PDFs along the diagonal we demonstrate the 95th percentile joint 2D posteriors. In Table 1 we summarize the recovered constraints and marginalized 68th percentile uncertainties for our eight astrophysical model parameters. Additionally, in Appendix A we demonstrate our trained network coverage demonstrating its convergence.

The 2D PS outperforms the 1D PS as evident by the narrower marginalized posteriors between the two summary statistics. However, the relative improvements are relatively modest. Based on the 68th percentile marginalized uncertainties we see on average improvements of approximately (5, 15, 30, 30, 40, 20) per cent for ($f_{*,10}$, α_*, $f_{\rm esc,10}$, $\alpha_{\rm esc}$, $M_{\rm turn}$, $L_{X<2\,\rm keV}$/SFR) with no improvement for t_* or E_0. Recall, in this work we do not include non-linear RDSs in our simulations of the 21-cm signal, thus these differences likely underestimate the actual differences when RSDs are included which serve to amplify the anisotropy of the 21-cm signal along the line of sight, $k_\parallel$. Nevertheless, these modest improvements are consistent with the Fisher Matrix expectations of Prelogović & Mesinger (2024) who predict relative improvements of 15 per cent on the variance of the individual parameters based on the factor of ~ 2 improvement in the total Fisher information. Note, we find little to no improvement in the X-ray parameters between the 2D PS and 1D PS. Likely, this is due to the selection of only two free X-ray parameters, $L_{X<2\,\rm keV}$/SFR and E_0, in our model. $L_{X<2\,\rm keV}$/SFR is constrained by the PS amplitude and is relatively independent of the EoH morphology. By only having one morphological X-ray parameter, E_0, combined with the increasing thermal noise to higher redshifts we limit the ability for the 2D PS to outperform the 1D PS. If we were to additionally consider the spectral index of the X-ray photons, $\alpha_{\rm X}$, as a free parameter, which is degenerate with E_0, then we would anticipate the 2D PS outperforming the 1D PS for this parameter combination due to the additional 2D spatial information provided by the 2D PS, although the relative improvement would still depend on the thermal noise amplitude.

These improvements in the constraining power arise due to the distinction of the Fourier information into their transverse ($k_\perp$) and redshift evolving ($k_\parallel$) components. Although the relative noise in each individual $k_\perp, k_\parallel$ bin increases due to the larger sample variance relative to the spherically averaged k-bins of the 1D PS, the anisotropic nature of the 21-cm signal yields additional information (see e.g. Fig. 2). By having the transverse spatial information independent of redshift we are more sensitive to the spatial morphology during the EoR and EoH. That is, sampling $k_\perp$ for a specific $k_\parallel$ provides unique information about the relative amplitudes of the spatial fluctuations as a function of redshift, providing more fine-grained detail than the 1D PS, which averages the anisotropic signal into the 'knee'-like feature at $k \sim 0.1$ Mpc^{-1}.

For example, since we are more sensitive to the spatial morphology (i.e. distribution of the ionized regions as a function of scale and redshift), we recover improved constraints on the parameters that control the typical sizes of the ionized regions. Namely $M_{\rm turn}$ which defines the characteristic masses of the star-forming galaxies and both the normalizations and mass-dependent power-law indices of

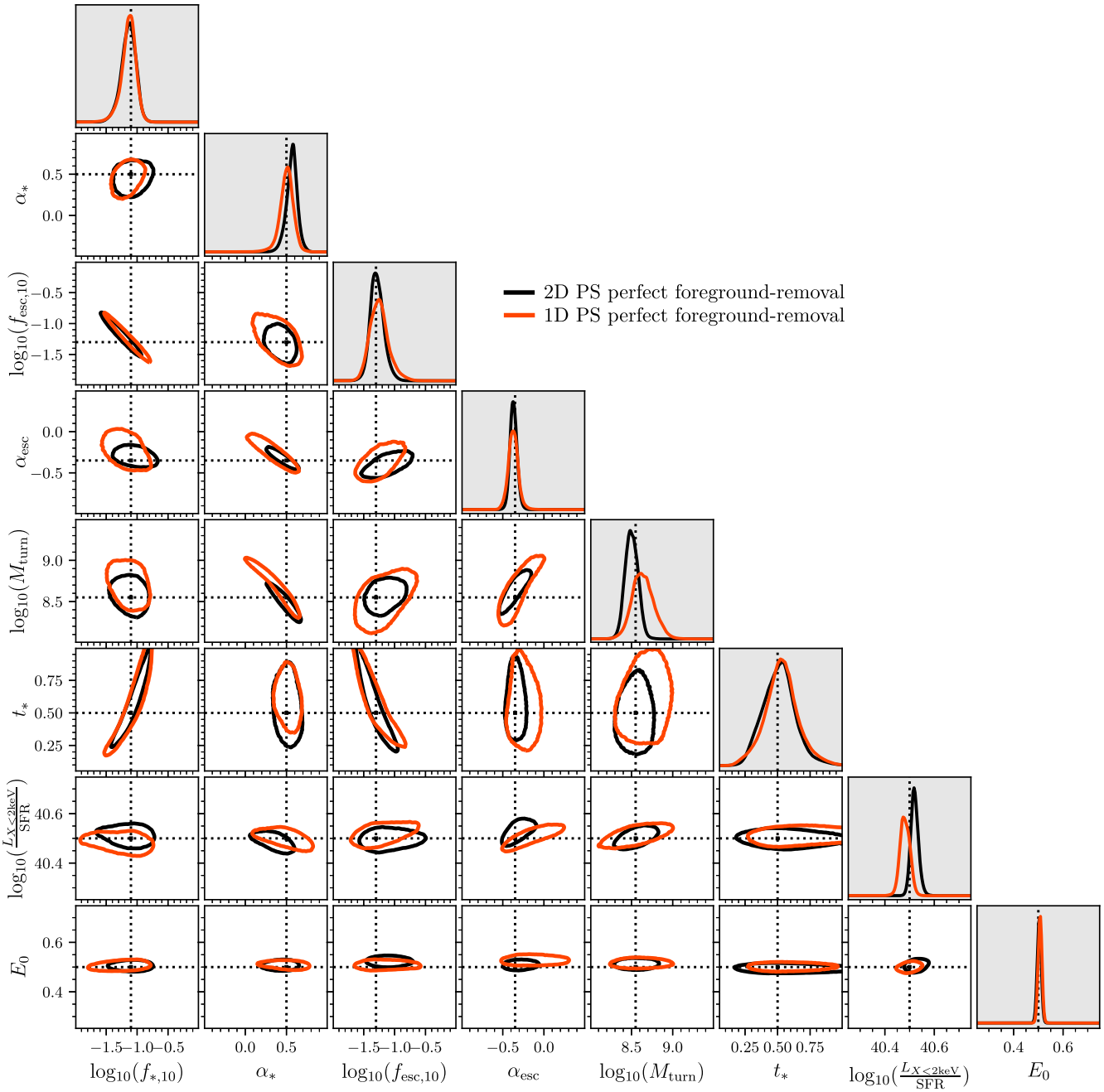

Figure 3. The recovered one- and two-dimensional marginalized posteriors on our astrophysical parameters assuming a mock 1000 h observation of the 21-cm signal assuming perfect foreground removal with the SKA. The black curves correspond to using the 2D cylindrically averaged PS (2D PS), whereas the red curves correspond to the 3D spherically averaged PS (1D PS). The 2D contours below the diagonal correspond to the 95th percentiles. The vertical and horizontal black dashed lines denote our fiducial astrophysical parameter set.

f_* and $f_{\rm esc}$ which control the production of UV ionizing photons and their escape into the IGM to drive ionizations.

4.3 Foreground avoidance

In Fig. 4, we now show the marginalized 1D and 2D posteriors for the same mock 21-cm observation of the 1D (teal dashed) and 2D PS (magenta dashed) except when assuming foreground wedge avoidance for a 1000 h observation with the SKA. For comparison, we also show the posteriors for the 2D PS assuming perfect foreground removal. Again, we provide the recovered constraints and 68th percentile uncertainties in Table 1.

As one would expect, considering foreground avoidance results in reduced constraining power relative to perfect foreground removal. However, the 2D PS still outperforms the 1D PS, albeit to a slightly lesser extent. Averaging over the marginalized 68th percentile uncertainties, we recover improvements of approximately (40, 10, 25, 15, 15) per cent for (α_*, $\alpha_{\rm esc}$, $M_{\rm turn}$, t_*, $L_{X<2\,\rm keV}$/SFR), with little to no improvement for the remaining parameters. In short, the 2D PS always outperform the 1D PS irrespective of the foreground removal strategy. Note, the slightly different selection of parameters

Table 1. A summary of the recovered astrophysical parameter constraints plus 68th marginalized uncertainties obtained following SBI on our mock observation of the 21-cm signal using either the 1D or 2D PS. We also consider two foreground mitigation strategies: (i) perfect foreground removal and (ii) foreground avoidance for a 1000 h observation using the SKA. Finally, we consider the improvements in the constraining power following the inclusion of UV LFs (see the text for further details).

	$\log_{10}(f_{*,10})$	α_*	$\log_{10}(f_{\rm esc,10})$	$\alpha_{\rm esc}$	$\log_{10}(M_{\rm turn})$ ($M_\odot$)	t_*	$\log_{10}\frac{L_{X<2\,\rm keV}}{\rm SFR}$ (erg s^{-1} $M_\odot^{-1}$ yr^{-1})	E_0 (keV)
Mock observation	-1.1	0.5	-1.30	-0.35	8.55	0.5	40.5	0.5
Foreground removal								
1D PS	$-1.12^{+0.10}_{-0.11}$	$0.51^{+0.07}_{-0.08}$	$-1.26^{+0.12}_{-0.16}$	$-0.37^{+0.06}_{-0.06}$	$8.61^{+0.14}_{-0.11}$	$0.53^{+0.12}_{-0.12}$	$40.48^{+0.02}_{-0.02}$	$0.51^{+0.01}_{-0.01}$
2D PS	$-1.13^{+0.11}_{-0.11}$	$0.55^{+0.06}_{-0.07}$	$-1.31^{+0.11}_{-0.09}$	$-0.37^{+0.04}_{-0.04}$	$8.49^{+0.08}_{-0.07}$	$0.54^{+0.09}_{-0.14}$	$40.51^{+0.02}_{-0.01}$	$0.50^{+0.01}_{-0.01}$
1D PS + UV LFs	$-1.13^{+0.10}_{-0.11}$	$0.49^{+0.05}_{-0.06}$	$-1.29^{+0.08}_{-0.10}$	$-0.31^{+0.04}_{-0.04}$	$8.57^{+0.07}_{-0.07}$	$0.60^{+0.13}_{-0.09}$	$40.50^{+0.01}_{-0.01}$	$0.51^{+0.01}_{-0.01}$
2D PS + UV LFs	$-1.09^{+0.10}_{-0.10}$	$0.46^{+0.06}_{-0.05}$	$-1.32^{+0.10}_{-0.12}$	$-0.34^{+0.03}_{-0.04}$	$8.60^{+0.07}_{-0.07}$	$0.46^{+0.13}_{-0.13}$	$40.50^{+0.01}_{-0.01}$	$0.51^{+0.01}_{-0.01}$
Foreground avoidance								
1D PS	$-1.14^{+0.10}_{-0.13}$	$0.48^{+0.13}_{-0.19}$	$-1.25^{+0.12}_{-0.14}$	$-0.29^{+0.14}_{-0.08}$	$8.60^{+0.27}_{-0.15}$	$0.48^{+0.14}_{-0.15}$	$40.47^{+0.03}_{-0.03}$	$0.49^{+0.03}_{-0.02}$
2D PS	$-1.10^{+0.11}_{-0.12}$	$0.51^{+0.08}_{-0.11}$	$-1.24^{+0.11}_{-0.15}$	$-0.27^{+0.10}_{-0.11}$	$8.43^{+0.18}_{-0.15}$	$0.53^{+0.16}_{-0.10}$	$40.48^{+0.03}_{-0.02}$	$0.48^{+0.04}_{-0.03}$
1D PS + UV LFs	$-1.11^{+0.13}_{-0.13}$	$0.45^{+0.08}_{-0.09}$	$-1.20^{+0.13}_{-0.11}$	$-0.30^{+0.05}_{-0.05}$	$8.47^{+0.16}_{-0.12}$	$0.58^{+0.16}_{-0.12}$	$40.49^{+0.03}_{-0.02}$	$0.49^{+0.04}_{-0.03}$
2D PS + UV LFs	$-1.13^{+0.11}_{-0.13}$	$0.36^{+0.09}_{-0.06}$	$-1.30^{+0.12}_{-0.11}$	$-0.30^{+0.05}_{-0.05}$	$8.49^{+0.15}_{-0.13}$	$0.57^{+0.15}_{-0.11}$	$40.51^{+0.03}_{-0.03}$	$0.45^{+0.03}_{-0.03}$

that recover slight improvements (e.g. $f_{*,10}$, $f_{\rm esc,10}$, and t_*) between the two foreground strategies are due to the strength of the complex parameter degeneracies and the fairly modest actual improvements in these specific parameters.

After considering foreground avoidance, we do not see an improvement in $f_{*,10}$ or $f_{\rm esc,10}$, with the only improvements in f_* and $f_{\rm esc}$ coming from their power-law mass dependence (e.g. α_* and $\alpha_{\rm esc}$). Nevertheless, we still recover improvements of $\sim 10-40$ per cent for $M_{\rm turn}$ and these power-law indices which highlights that even when applying foreground avoidance, we still pick up additional information from the spatial morphology of the 21-cm signal due to how the 2D PS distinguishes the structural information from that of redshift evolution. This is despite the fact that after applying foreground avoidance we loose more than 60–80 per cent of the 2D PS Fourier modes (see e.g. Fig. 2). However, predominately this information loss is for larger $k_\perp$'s, with still relatively decent sampling of $k_\perp$ at $\lesssim 0.1$ Mpc^{-1} where we predominately extract most of the constraining information (Greig & Mesinger 2015). At these scales, we are still recovering the redshift evolution of the 21-cm signal (e.g. $k_\parallel$), therefore we remain sensitive to how the EoR morphology evolves with redshift. This enables the still relatively strong constraints on our astrophysical parameters.

Relative to perfect foreground removal, for the 2D PS we determine increases in the marginalized 68th percentiles of approximately (5, 50, 30, 275, 200, 5, 70, 300) per cent for ($f_{*,10}$, α_*, $f_{\rm esc,10}$, $\alpha_{\rm esc}$, $M_{\rm turn}$, t_*, $L_{X<2\,\rm keV}$/SFR, E_0) by considering foreground avoidance. Over our entire mock 21-cm observation with the 2D PS, foreground avoidance results in a loss of $\sim$ 60–80 per cent of the 2D Fourier modes owing to the redshift dependence of the foreground wedge. Therefore, despite the loss in over approximately five times the amount of information, we do not exhibit such severe losses in constraining power. What is important is not the total amount of information lost, but rather where this information is lost. As highlighted above, we still access the redshift evolution of the 21-cm PS on those scales most sensitive to the EoR (e.g. $k_\perp \sim 0.1$ Mpc^{-1}). As a result, we recover relatively more modest losses in constraining power on our EoR parameters. Nevertheless, the loss in information below the wedge at moderate scales, $k_\perp \sim 0.5$ does limit our ability to constrain $M_{\rm turn}$ which drives the increased uncertainties in the power-law indices. For the X-ray parameters, the relative losses are more significant, and this is due to the increasing amplitude of the wedge during the heating epoch removing more spatial information (the wedge moves vertically upward in Fig. 2 for increasing redshift). Thus, we have considerably less spatial information during the EoH heating. However, despite these uncertainties increasing by up to a factor of ~ 3 the X-ray parameters are still very strongly constrained, highlighting how sensitive the X-ray parameters are tied to the amplitude of the 21-cm signal.

Repeating this analysis for the 1D PS, we recover increases in the marginalized 68th percentiles of approximately (10, 200, 10, 200, 70, 25, 60, 300) per cent for ($f_{*,10}$, α_*, $f_{\rm esc,10}$, $\alpha_{\rm esc}$, $M_{\rm turn}$, t_*, $L_{X<2\,\rm keV}$/SFR, E_0) by considering foreground avoidance instead of foreground removal. These relative increases are comparable in amplitude to those for the 2D PS, as one would expect. Again, this highlights that it is not the amount of information lost, rather where the information is lost relative to where the 21-cm signal is most sensitive.

4.4 Mock 2D PS observation with UV LFs

Thus far, we have only considered the relative improvements in our astrophysical parameter constraints when considering the 2D PS compared to the 1D PS. However, one can also include additional constraining information from alternative probes of the reionization epoch, such as observed UV luminosity functions (LFs). This simply requires concatenating the UV LF data to the existing PS data and passing this information into SWYFT and retraining the ratio networks. In the case of the 1D PS, the role of the UV LFs is to break the degeneracy between f_* and $f_{\rm esc}$ improving the constraining power on these parameters and consequently also on $M_{\rm turn}$ (Park et al. 2019). Since these same parameters are more strongly constrained with the 2D PS relative to the 1D PS, it will thus be illustrative to consider a joint observation of the 2D PS and UV LFs.

Following Park et al. (2019), we consider a limited selection of observed UV LFs at $z = 6$ (Bouwens et al. 2017), $z = 7$ and 8 (Bouwens et al. 2015), and $z = 10$ (Oesch et al. 2018). This choice is motivated by limiting the systematic differences across the various groups within the literature and how each deals with their observational and statistical uncertainties. Ideally, to be truly robust (and conservative) one should average across all results in the literature to obtain a mean UV LF with a scatter encompassing all the differences across the various results. In future, we shall return

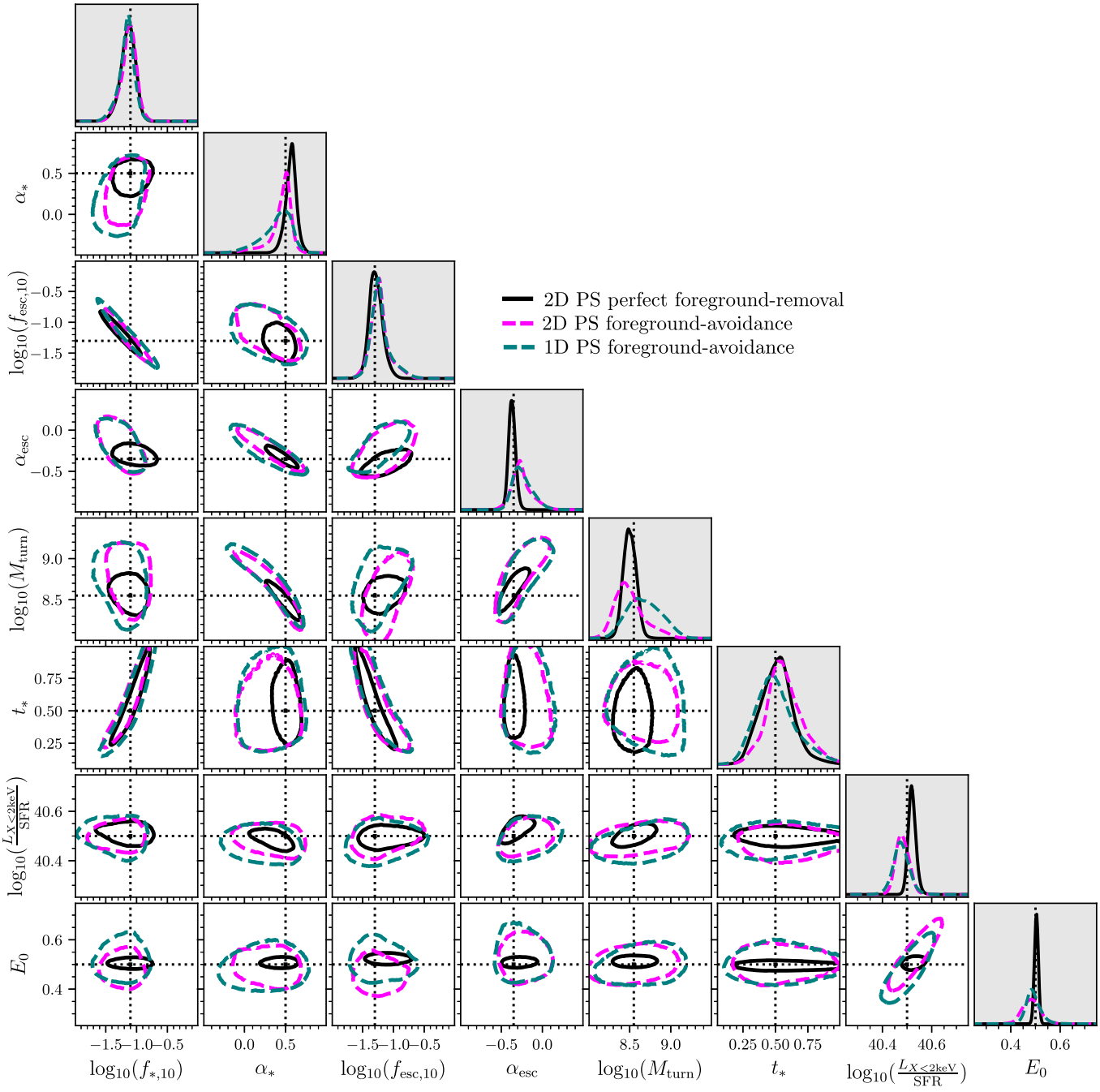

Figure 4. The same as Fig. 3 except now considering a 1000 h observation with the SKA assuming foreground wedge avoidance. The magenta (teal) dashed contours correspond to the 2D (1D) PS, whereas the black contours represent the 2D PS assuming foreground removal for comparison.

to this while also extending our UV LF sampling to higher redshifts as observed by the *JWST* (e.g. Castellano et al. 2022; Harikane et al. 2022; Naidu et al. 2022; Atek et al. 2023; Bouwens et al. 2023; Donnan et al. 2023; Labbé et al. 2023; Willott et al. 2024). Importantly, when including UV LFs into our inference pipeline, we only consider UV magnitudes fainter that $M_{\rm UV} < -20$, for which it is argued that these are relatively dust-free (see Park et al. 2019).

In Fig. 5 we demonstrate the 1D and 2D marginalized posteriors for a mock 1000 h observation of the 2D PS in addition to UV LFs at $z = 6$, 7, 8, and 10. Below the diagonal the orange (black) contours correspond to the 95th percentile joint marginalized posteriors when considering wedge removal with (without) UV LFs. Above the diagonal, we provide the equivalent following foreground avoidance with the magenta (purple) dashed contours denoting with (without) UV LFs. Finally, in Table 1 we summarize the constraints and 68th percentile marginalized uncertainties.

In both cases, it is clear that the complimentary constraining power from the UV LFs improves the overall constraints on our astrophysical parameters using the 2D PS. In general, we find the amplitude of the improvements are larger for wedge avoidance relative to wedge removal. For example, we recover improvements of ~ 20 and ~ 30 per cent for α_* and $\alpha_{\rm esc}$ along with ~ 15 per cent improvements for $M_{\rm turn}$, whereas for wedge removal, at most we see improvements of ~ 10 per cent for these same parameters. These

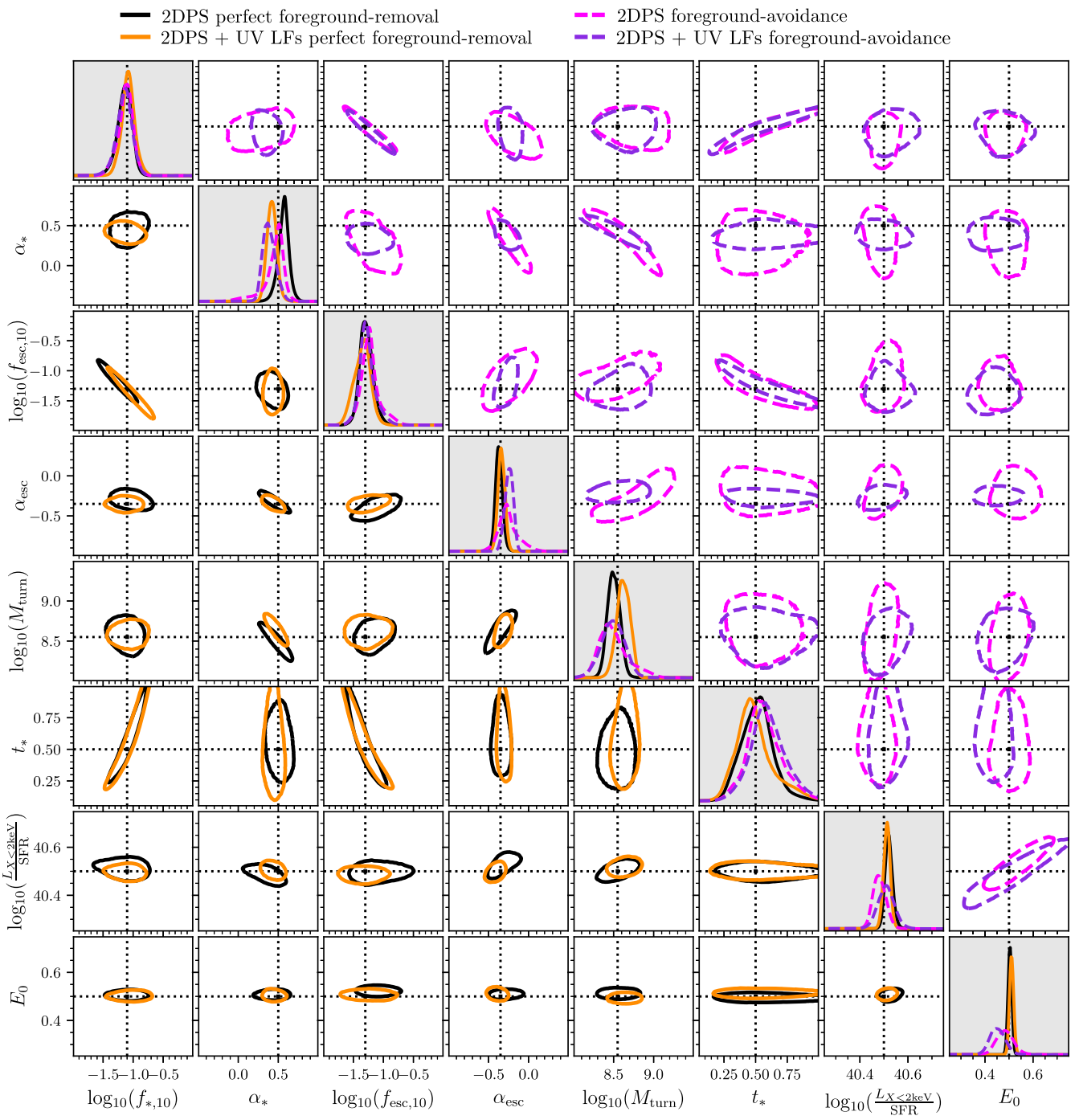

Figure 5. The recovered one- and two-dimensional marginalized posteriors on our astrophysical parameters assuming a mock 1000 h observation of the 21-cm signal using the 2D PS from the SKA combined with UV LFs at $z = 6$–10. Below the diagonal, we consider the case of perfect foreground removal, with the orange (black) curves corresponding to with (without) the UV LFs, respectively. Above the diagonal, we present the results instead assuming foreground avoidance, with the magenta (purple) dashed curves denoting observations with (without) the UV LFs. In all cases, the 2D contours represent the joint 95th percentile marginalized uncertainties. The vertical and horizontal black dashed lines denote our fiducial astrophysical parameter set.

relatively larger gains for foreground wedge avoidance following the inclusion of the UV LF information are due to the originally broader constraints and stronger degeneracies on the UV galaxy parameters, namely α_* and $f_{\rm esc,10}$. For the 2D PS with perfect foreground removal, since we have additional structural information on the EoR morphology through $k_\perp$, we are able to limit the degeneracy between $\alpha_* - f_{\rm esc,10}$ (see Fig. 4). Therefore, since this degeneracy is already reduced, the relative gains for the 2D PS with perfect foreground removal with UV LFs are also reduced.

Interestingly, once we include UV LFs with the 1D and 2D PS, the 68th percentile uncertainties become similar. That is, the relative difference between the reported uncertainties for the 2D PS + UV LFs and the 1D PS + UV LFs has been reduced compared to the differences between just the 2D PS and the 1D

PS. Now, we recover at most improvements of ~ 10 per cent when considering the 2D PS + UV LFs relative to the 1D PS + UV LFs (compared to up to 30 per cent in the absence of the UV LF data for the 2D PS compared to the 1D PS). This holds true for either foreground mitigation strategy, with foreground removal still notably outperforming foreground avoidance. The origin of this smaller difference between the 2D PS and the 1D PS following the inclusion of the UV LF data is due to where the additional information is provided. As highlighted earlier, compared to the 1D PS, the 2D PS is more sensitive to the EoR morphology as the structural information, $k_\perp$, is kept distinct from the redshift evolving component of the 21-cm signal ($k_\parallel$). This enables the 2D PS to improve over the 1D PS at constraining the EoR parameters, notably reducing the degeneracy between α_* and $f_{\rm esc,10}$.

However, including UV LF data serves a similar purpose. The additional constraining power provided by the UV LFs adds unique information on f_* which leads to a reduction of the same degeneracy between α_* and $f_{\rm esc,10}$. However, the reduction in this degeneracy is stronger with the unique information from the UV LF data than it is for the additional information that the 2D PS provides relative to the 1D PS. Thus, the inclusion of the constraining power from the UV LFs somewhat minimizes the contribution to the overall improvements in constraining power that come from just the 2D PS relative to the 1D PS. However, this behaviour is likely dependent on the underlying astrophysical model parametrization. A model with either additional astrophysical parameters or more complex scalings with mass or redshift that are more sensitive to the EoR morphology (e.g. the 2D PS) would more significantly benefit from the increased information that arises from the 2D PS compared to the 1D PS. Thus, for astrophysical models containing more discriminating power in the 2D PS relative to the 1D PS, we would expect that the constraints obtained from 2D PS + UV LFs would more strongly outperform those from the 1D PS. + UV LFs as the UV LFs likely would add little additional information to more complex parametrizations.

5 CONCLUSIONS

In recent years, SBI has begun to gain traction for performing Bayesian inference from the 21-cm signal to gain insights into the galaxies responsible for reionization (e.g. Zhao et al. 2022a, 2022b; Prelogović & Mesinger 2023; Saxena et al. 2023; Greig et al. 2024). The significant advantage of SBI is that it applies machine learning principles to bypass the requirement to have an analytic expression to describe the likelihood function to accurately describe our 21-cm summary statistics. By removing this crucial bottleneck we are now able to rigorously explore more complex summary statistics than the simple, but extensively explored 1D spherically averaged power spectrum (1D PS). As a demonstration of the power of SBI, in this work we explore using the 2D cylindrically averaged PS (2D PS), which has previously been overlooked owing to the complexities in computing its likelihood.

For exploring the 2D PS we consider a mock 1000 h observation of the 21-cm signal using the SKA. Throughout, we simulate the 21-cm signal using 21CMFAST (Mesinger & Furlanetto 2007; Mesinger et al. 2011; Murray et al. 2020), in particular the flexible UV galaxy parametrization introduced in Park et al. (2019). As a result we have an eight-parameter astrophysical model to describe the UV and X-ray properties of the first galaxies responsible for driving reionization. Further, we consider two foreground mitigation strategies: (i) perfect foreground removal whereby we have access to the whole 2D information and (ii) foreground avoidance where we only use the pristine cosmological signal above the foreground contaminated wedge. Throughout, we perform SBI using MNRE to learn the likelihood-to-evidence for performing parameter inference using SWYFT (Miller et al. 2022).

When considering perfect foreground removal, we find the 2D PS outperforms the 1D PS by reducing the 68th percentile uncertainties on individual parameters by up to $\sim$ 30–40 per cent. These relative improvements in the 2D PS over the 1D PS are consistent with recent predictions using the amplitude of the Fisher Information (Prelogović & Mesinger 2024). Primarily, the most significant gains are in $M_{\rm turn}$ which effectively describes the minimum mass for star-forming galaxies along with α_* and $\alpha_{\rm esc}$ which describe the mass dependence of star formation efficiency, f_* and IGM escape fraction, $f_{\rm esc}$. These improvements are achieved due to the 2D PS cleanly separating the transverse information, $k_\perp$, from the redshift-evolving component of the signal, $k_\parallel$. In this way, we are more sensitive to the redshift evolution of the ionization morphology allowing for improved constraints on the UV galaxy parameters. Unlike the 1D PS which combines and averages the anisotropic information into a single k when spherically averaging.

Even when performing foreground avoidance, when we lose a large fraction of the 2D PS information relative to the case of perfect foreground removal, the 2D PS still outperforms the 1D PS. However, the relative boosts in performance are reduced, with only 20–30 per cent improvements on our individual model parameters. Nevertheless, the largest gains remain for $M_{\rm turn}$, α_*, and $\alpha_{\rm esc}$. This implies that despite the loss of a large fraction of information due to foreground contamination, distinguishing between the spatial ($k_\perp$) and frequency-dependent ($k_\parallel$) Fourier modes still yields additional constraining power over the 1D PS for constraining the UV galaxy parameters during the EoR.

Comparing the two foreground mitigation strategies directly, we find foreground avoidance results in increased 68th percentile uncertainties of at worst $\sim$ 2–3 compared to foreground removal. In general, the largest increases are for the X-ray parameters, which are due to the growth of the foreground contaminated region towards larger redshifts, where the 21-cm signal is more sensitive to the X-ray contribution. However, we also see increases in the uncertainties at a similar level for $\alpha_{\rm esc}$ and $M_{\rm turn}$, owing to the loss of a significant fraction of spatial ($k_\perp$) information due to foreground wedge contamination. For the remainder, the 68th marginalized uncertainties increase by ≤ 70 per cent.

Finally, we also include independent astrophysical information by considering UV galaxy LFs at $z = 6 - 10$. Doing so, we find improvements of ~ 10 per cent primarily on α_*, $\alpha_{\rm esc}$, and $M_{\rm turn}$ for foreground removal. For foreground avoidance, we find improvements of up to $\sim$ 20–30 per cent for these same parameters. Generally speaking, for the 1D PS the addition of UV LFs is to break the degeneracy between f_* and $f_{\rm esc}$. However, for the 2D PS, as it is more sensitive to the EoR morphology through the distinct spatial information, the f_*-$f_{\rm esc}$ degeneracy is not nearly as strong. Therefore, the UV LFs have reduced benefit for foreground removal over foreground avoidance as we have additional 2D spatial information to reduce this otherwise strong degeneracy.

The power of SBI is that it enables the study of complex and non-Gaussian summary statistics of the 21-cm signal to be explored in the context of astrophysical parameter inference. Here, we have demonstrated the value of SBI with the first study of the more complex 2D PS. In future, to maximize the wealth of information expected to be available from the 21-cm signal we will explore alternative non-Gaussian statistics with SBI.

ACKNOWLEDGEMENTS

We would like to thank the anonymous referees whose comments improved this manuscript. Parts of this research were supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. YST acknowledges financial support from the Australian Research Council through DECRA Fellowship DE220101520. AM acknowledges support from the Ministry of Universities and Research (MUR) through the PRIN project'Optimal inference from radio images of the epoch of reionization' as well as the PNRR project'Centro Nazionale di Ricerca in High Performance Computing, Big Data e Quantum Computing'. This work was performed on the OzSTAR national facility at Swinburne University of Technology. The OzSTAR program receives funding in part from the Astronomy National Collaborative Research Infrastructure Strategy (NCRIS) allocation provided by the Australian Government, and from the Victorian Higher Education State Investment Fund (VHESIF) provided by the Victorian Government.

*Software:*NUMPY (Harris et al. 2020), SCIPY (Virtanen et al. 2020), H5PY (Collette 2013), MATPLOTLIB (Hunter 2007), CORNER (Foreman-Mackey 2016), and SEABORN (Waskom et al. 2017).

DATA AVAILABILITY

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

REFERENCES

Atek H. et al., 2023, MNRAS, 519, 1201
Bag S., Mondal R., Sarkar P., Bharadwaj S., Choudhury T. R., Sahni V., 2019, MNRAS, 485, 2235
Banet A., Barkana R., Fialkov A., Guttman O., 2021, MNRAS, 503, 1221
Barkana R., Loeb A., 2001, Phys. Rep., 349, 125
Behroozi P. S., Silk J., 2015, ApJ, 799, 32
Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143
Betancourt M., 2019, Ann. Phys., 531, 1700214
Bianco M., Giri S. K., Iliev I. T., Mellema G., 2021, MNRAS, 505, 3982
Bianco M., Giri S. K., Prelogović D., Chen T., Mertens F. G., Tolley E., Mesinger A., Kneib J.-P., 2024, MNRAS, 528, 5212
Bosman S. E. I., Fan X., Jiang L., Reed S., Matsuoka Y., Becker G., Haehnelt M., 2018, MNRAS, 479, 1055
Bouwens R. J. et al., 2015, ApJ, 803, 34
Bouwens R. J., Oesch P. A., Illingworth G. D., Ellis R. S., Stefanon M., 2017, ApJ, 843, 129
Bouwens R. J. et al., 2023, MNRAS, 523, 1036
Breitman D., Mesinger A., Murray S. G., Prelogović D., Qin Y., Trotta R., 2024, MNRAS, 527, 9833
Castellano M. et al., 2022, ApJ, 938, L15
Chapman E., Jelić V., 2019, preprint (arXiv:1909.12369)
Chen Z., Xu Y., Wang Y., Chen X., 2019, ApJ, 885, 23
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
Collette A., 2013, Python and HDF5. O'Reilly
Cranmer K., Brehmer J., Louppe G., 2020, Proc. Natl. Acad. Sci., 117, 30055
Datta A., Bowman J. D., Carilli C. L., 2010, ApJ, 724, 526
Dayal P., Ferrara A., Dunlop J. S., Pacucci F., 2014, MNRAS, 445, 2545
Donnan C. T. et al., 2023, MNRAS, 518, 6011
Durkan C., Murray I., Papamakarios G., 2020, in III H. D., Singh A.eds, Proceedings of Machine Learning Research Vol. 119, Proceedings of the 37th International Conference on Machine Learning. PMLR, p. 2771, https://proceedings.mlr.press/v119/durkan20a.html
Eastwood M. W. et al., 2019, AJ, 158, 84
Elbers W., van de Weygaert R., 2019, MNRAS, 486, 1523
Field G. B., 1958, Proc. Inst. Radio Eng., 46, 240
Fisher R. A., 1935, J. R. Stat. Soc., 98, 39
Foreman-Mackey D., 2016, J. Open Source Softw., 1, 24
Fragos T. et al., 2013, ApJ, 764, 41
Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, ApJ, 613, 1
Furlanetto S. R., Oh S. P., Briggs F. H., 2006, Phys. Rep., 433, 181
Gagnon-Hartman S., Cui Y., Liu A., Ravanbakhsh S., 2021, MNRAS, 504, 4716
Gazagnes S., Koopmans L. V. E., Wilkinson M. H. F., 2021, MNRAS, 502, 1816
Gillet N., Mesinger A., Greig B., Liu A., Ucci G., 2019, MNRAS, 484, 282
Giri S. K., Mellema G., 2021, MNRAS, 505, 1863
Giri S. K., Mellema G., Dixon K. L., Iliev I. T., 2018a, MNRAS, 473, 2949
Giri S. K., Mellema G., Ghara R., 2018b, MNRAS, 479, 5596
Giri S. K., Mellema G., Aldheimer T., Dixon K. L., Iliev I. T., 2019a, MNRAS, 489, 1590
Giri S. K., D'Aloisio A., Mellema G., Komatsu E., Ghara R., Majumdar S., 2019b, J. Cosmol. Astropart. Phys., 2019, 058
Giroux M. L., Sutherland R. S., Shull J. M., 1994, ApJ, 435, L97
Gnedin N. Y., Ostriker J. P., 1997, ApJ, 486, 581
Gnedin N. Y., Shaver P. A., 2004, ApJ, 608, 611
Gorce A., Hutter A., Pritchard J. R., 2021, A&A, 653, A58
Greig B., Mesinger A., 2015, MNRAS, 449, 4246
Greig B., Mesinger A., 2017, MNRAS, 472, 2651
Greig B., Mesinger A., 2018, MNRAS, 477, 3217
Greig B., Ting Y.-S., Kaurov A. A., 2022, MNRAS, 513, 1719
Greig B., Ting Y.-S., Kaurov A. A., 2023, MNRAS, 519, 5288
Greig B., Prelogović D., Mirocha J., Qin Y., Ting Y.-S., Mesinger A., 2024, preprint (arXiv:2403.14061)
Gupta Y. et al., 2017, Curr. Sci., 113, 707
van Haarlem M. P. et al., 2013, A&A, 556, 2
Harikane Y. et al., 2016, ApJ, 821, 123
Harikane Y. et al., 2022, ApJ, 929, 1
Harris C. R. et al., 2020, Nature, 585, 357
Hassan S., Liu A., Kohn S., La Plante P., 2019, MNRAS, 483, 2524
Hassan S., Andrianomena S., Doughty C., 2020, MNRAS, 494, 5761
Hermans J., Delaunoy A., Rozet F., Wehenkel A., Begy V., Louppe G., 2021, preprint (arXiv:2110.06581)
Hothi I., Allys E., Semelin B., Boulanger F., 2024, A&A, 686, A212
Hui L., Gnedin N. Y., 1997, MNRAS, 292, 27
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Hutter A., Dayal P., Yepes G., Gottlöber S., Legrand L., Ucci G., 2021, MNRAS, 503, 3698
Jensen H. et al., 2013, MNRAS, 435, 460
Kakiichi K. et al., 2017, MNRAS, 471, 1936
Kamran M., Ghara R., Majumdar S., Mondal R., Mellema G., Bharadwaj S., Pritchard J. R., Iliev I. T., 2021, MNRAS, 502, 3800
Kapahtia A., Chingangbam P., Appleby S., 2019, J. Cosmol. Astropart. Phys., 2019, 053
Kapahtia A., Chingangbam P., Ghara R., Appleby S., Choudhury T. R., 2021, J. Cosmol. Astropart. Phys., 2021, 026
Koopmans L. et al., 2015, Proc. Sci., The Cosmic Dawn and Epoch of Reionisation with SKA. SISSA, Trieste, PoS#1
Kubota K., Yoshiura S., Shimabukuro H., Takahashi K., 2016, PASJ, 68, 61
Kuhlen M., Faucher-Giguère C.-A., 2012, MNRAS, 423, 862
Kwon Y., Hong S. E., Park I., 2020, J. Korean Phys. Soc., 77, 49
La Plante P., Ntampaka M., 2019, ApJ, 880, 110
Labbé I. et al., 2023, Nature, 616, 266
Liu A., Parsons A. R., Trott C. M., 2014a, Phys. Rev. D, 90, 023018
Liu A., Parsons A. R., Trott C. M., 2014b, Phys. Rev. D, 90, 023019
Madau P., Meiksin A., Rees M. J., 1997, ApJ, 475, 429
Majumdar S., Pritchard J. R., Mondal R., Watkinson C. A., Bharadwaj S., Mellema G., 2018, MNRAS, 476, 4007
Majumdar S., Kamran M., Pritchard J. R., Mondal R., Mazumdar A., Bharadwaj S., Mellema G., 2020, MNRAS, 499, 5090
Mangena T., Hassan S., Santos M. G., 2020, MNRAS, 494, 600

Mao Y., Shapiro P. R., Mellema G., Iliev I. T., Koda J., Ahn K., 2012, MNRAS, 422, 926
Mellema G. et al., 2013, Exp. Astron., 36, 235
Mesinger A., Dijkstra M., 2008, MNRAS, 390, 1071
Mesinger A., Furlanetto S., 2007, ApJ, 669, 663
Mesinger A., Furlanetto S., Cen R., 2011, MNRAS, 411, 955
Miller B., Cole A., Forré P., Louppe G., Weniger C., 2021, Adv. Neural Inf. Process. Syst., 34, 129
Miller B., Cole A., Weniger C., Nattino F., Ku O., Grootes M., 2022, J. Open Source Softw., 7, 4205
Mineo S., Gilfanov M., Sunyaev R., 2012, MNRAS, 419, 2095
Mitra S., Choudhury T. R., Ferrara A., 2015, MNRAS, 454, L76
Mondal R., Mellema G., Murray S. G., Greig B., 2022, MNRAS, 514, L31
Morales M. F., Wyithe J. S. B., 2010, ARA&A, 48, 127
Morales M. F., Hazelton B., Sullivan I., Beardsley A., 2012, ApJ, 752, 137
Murray S. G., Trott C. M., 2018, ApJ, 869, 25
Murray S., Greig B., Mesinger A., Muñoz J., Qin Y., Park J., Watkinson C., 2020, J. Open Source Softw., 5, 2582
Mutch S. J., Geil P. M., Poole G. B., Angel P. W., Duffy A. R., Mesinger A., Wyithe J. S. B., 2016, MNRAS, 462, 250
Naidu R. P. et al., 2022, ApJ, 940, L14
Nasirudin A., Murray S. G., Trott C. M., Greig B., Joseph R. C., Power C., 2020, ApJ, 893, 118
Ocvirk P. et al., 2016, MNRAS, 463, 1462
Oesch P. A., Bouwens R. J., Illingworth G. D., Labbé I., Stefanon M., 2018, ApJ, 855, 105
Okamoto T., Gao L., Theuns T., 2008, MNRAS, 390, 920
Pacucci F., Mesinger A., Mineo S., Ferrara A., 2014, MNRAS, 443, 678
Park J., Mesinger A., Greig B., Gillet N., 2019, MNRAS, 484, 933
Parsons A. R. et al., 2010, AJ, 139, 1468
Parsons A. R., Pober J. C., Aguirre J. E., Carilli C. L., Jacobs D. C., Moore D. F., 2012, ApJ, 756, 165
Parsons A. R. et al., 2014, ApJ, 788, 106
Planck Collaboration VI, 2020, A&A, 641, A6
Pober J. C. et al., 2013, AJ, 145, 65
Pober J. C. et al., 2014, ApJ, 782, 66
Pober J. C. et al., 2016, ApJ, 819, 8
Prelogović D., Mesinger A., 2023, MNRAS, 524, 4239
Prelogović D., Mesinger A., 2024, preprint (arXiv:2401.12277)
Prelogović D., Mesinger A., Murray S., Fiameni G., Gillet N., 2022, MNRAS, 509, 3852
Pritchard J. R., Loeb A., 2012, Rep. Prog. Phys., 75, 086901
Qin Y., Mesinger A., Bosman S. E. I., Viel M., 2021, MNRAS, 506, 2390
Roy V., 2020, Annu. Rev. Stat. Appl., 7, 387
Salpeter E. E., 1955, ApJ, 121, 161
Saxena A., Cole A., Gazagnes S., Meerburg P. D., Weniger C., Witte S. J., 2023, MNRAS, 525, 6097
Scoccimarro R., 1998, MNRAS, 299, 1097
Shapiro P. R., Giroux M. L., Babul A., 1994, ApJ, 427, 25
Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, A&A, 345, 380
Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1
Shimabukuro H., Yoshiura S., Takahashi K., Yokoyama S., Ichiki K., 2015, MNRAS, 451, 467
Shimabukuro H., Yoshiura S., Takahashi K., Yokoyama S., Ichiki K., 2016, MNRAS, 458, 3003
Shimabukuro H., Yoshiura S., Takahashi K., Yokoyama S., Ichiki K., 2017, MNRAS, 468, 1542
Sobacchi E., Mesinger A., 2013a, MNRAS, 432, L51
Sobacchi E., Mesinger A., 2013b, MNRAS, 432, 3340
Sobacchi E., Mesinger A., 2014, MNRAS, 440, 1662
Springel V., Hernquist L., 2003, MNRAS, 339, 312
Stefanon M., Bouwens R. J., Labbé I., Illingworth G. D., Gonzalez V., Oesch P. A., 2021, ApJ, 922, 29
Sun G., Furlanetto S. R., 2016, MNRAS, 460, 417
Tacchella S., Bose S., Conroy C., Eisenstein D. J., Johnson B. D., 2018, ApJ, 868, 92
Thompson A. R., Moran J. M., Swenson G. W., 2007, Interferometry and Synthesis in Radio Astronomy. Wiley, New York
Thyagarajan N. et al., 2013, ApJ, 776, 6
Thyagarajan N. et al., 2015a, ApJ, 804, 14
Thyagarajan N. et al., 2015b, ApJ, 807, L28
Tingay S. J. et al., 2013, Publ. Astron. Soc. Aust., 30, 7
Tiwari H., Shaw A. K., Majumdar S., Kamran M., Choudhury M., 2022, J. Cosmol. Astropart. Phys., 2022, 045
Tozzi P., Madau P., Meiksin A., Rees M. J., 2000, ApJ, 528, 597
Trott C. M., Wayth R. B., Tingay S. J., 2012, ApJ, 757, 101
Vedantham H., Udaya Shankar N., Subrahmanyan R., 2012, ApJ, 745, 176
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Waskom M. et al., 2017, mwaskom/seaborn: v0.8.1 (September 2017), Zenodo, https://doi.org/10.5281/zenodo.883859
Watkinson C. A., Pritchard J. R., 2014, MNRAS, 443, 3090
Watkinson C. A., Giri S. K., Ross H. E., Dixon K. L., Iliev I. T., Mellema G., Pritchard J. R., 2019, MNRAS, 482, 2653
Watkinson C. A., Greig B., Mesinger A., 2022, MNRAS, 510, 3838
Wayth R. et al., 2018, Publ. Astron. Soc. Aust., 35, 33
Willott C. J. et al., 2024, ApJ, 966, 74
Wouthuysen S. A., 1952, AJ, 57, 31
Yoshiura S., Shimabukuro H., Takahashi K., Momose R., Nakanishi H., Imai H., 2015, MNRAS, 451, 266
Yoshiura S., Shimabukuro H., Takahashi K., Matsubara T., 2017, MNRAS, 465, 394
Yue B., Ferrara A., Xu Y., 2016, MNRAS, 463, 1968
Yung L. Y. A., Somerville R. S., Popping G., Finkelstein S. L., Ferguson H. C., Davé R., 2019, MNRAS, 490, 2855
Zhao X., Mao Y., Cheng C., Wandelt B. D., 2022a, ApJ, 926, 151
Zhao X., Mao Y., Wandelt B. D., 2022b, ApJ, 933, 236

APPENDIX A: ASSESSING NETWORK COVERAGE

One of the key defining features of SBI approaches such as MNRE is that once the network is trained they allow for the rapid recovery of the posteriors for any new realization of the input data. In this case, we can perform parameter inference for a large number of mock observations drawn from within our prior range to determine how frequently they fall within their predicted posteriors. Measuring this frequency for a sufficiently large number of models enables the computation of the network coverage (e.g. Cole et al. 2022). This large number of direct posterior evaluations provides a much more robust quantity to indicate network convergence than those typically adopted by direct MCMC approaches (e.g. Betancourt 2019; Roy 2020).

Following Cole et al. (2022), we define $\Theta_{\hat{p}(\hat{\boldsymbol{\theta}}|\boldsymbol{x}_i)}(1-\alpha)$ to be a function which determines the $(1-\alpha)$ highest probability density region (HPDR) for our estimated posterior, $\hat{p}(\hat{\boldsymbol{\theta}}|\boldsymbol{x}_i)$, given the input model-parameter pair, $\boldsymbol{x}_i, \boldsymbol{\theta}_i^*$. To demonstrate, a 95 per cent HPDR would correspond to $\alpha = 0.05$. For a set of n independently drawn model-parameter pairs we can then determine the actual error rate $1-\hat{\alpha}$ of the HPDR given our estimated posterior:

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

The quantities α ($\hat{\alpha}$) are redefined in terms of a new variable, z, corresponding to the $1-\alpha/2$ ($1-\hat{\alpha}/2$) quantile of the standard normal distribution. By definition this implies the $1\sigma, 2\sigma, 3\sigma$ regions correspond to $z = 1, 2, 3$ with $1-\alpha = 0.6827, 0.9545, 0.9997$. The uncertainties on the error rate $\hat{\alpha}$ are determined by the Jeffreys interval (Cole et al. 2022).[9] In Fig. A1 we present the empirical

[9]Specifically, this interval is obtained from the 68.27 per cent central interval of a Beta distribution defined by the parameters $n-k+1/2, k+1/2$, where

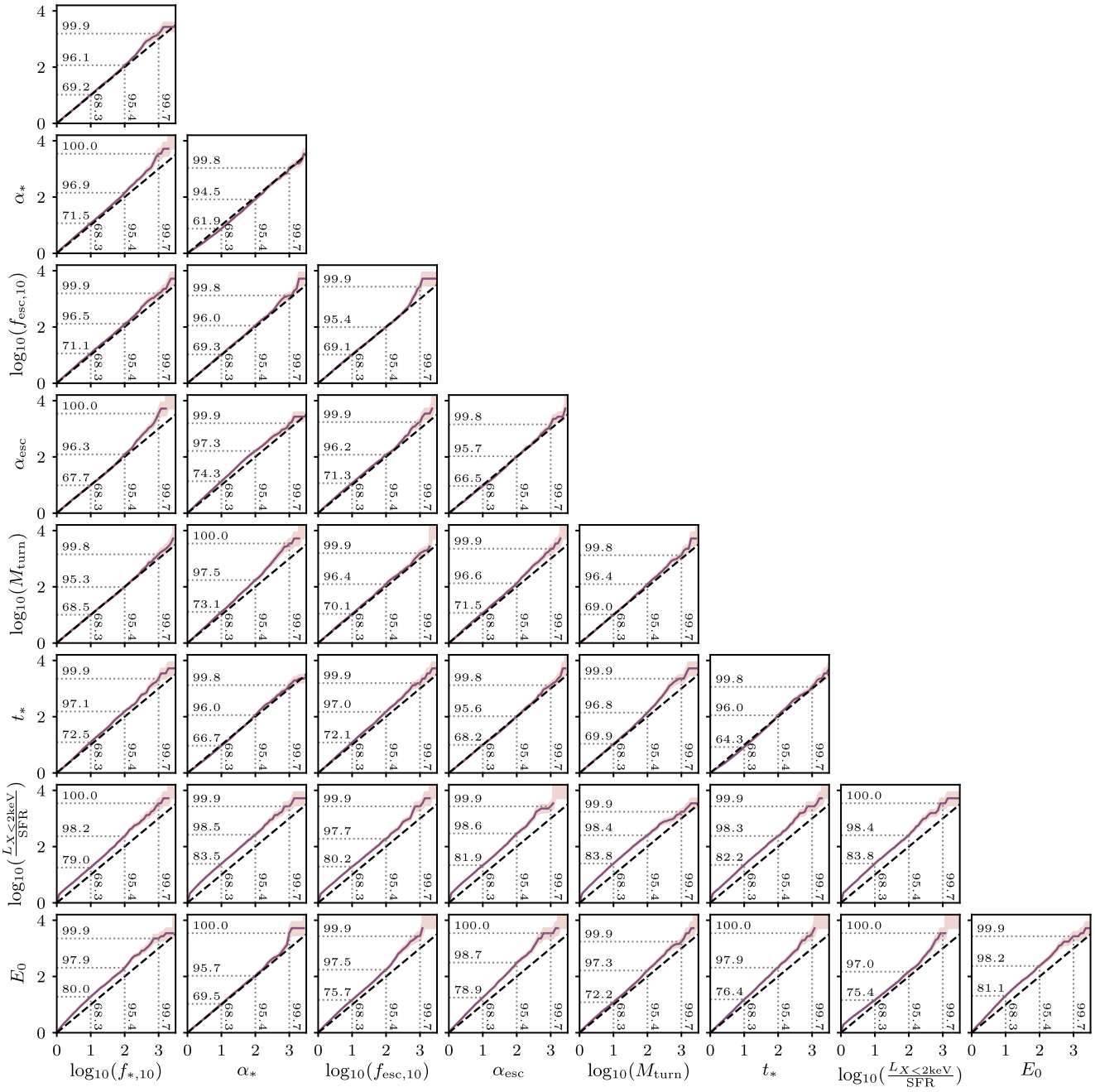

Figure A1. The empirical expected coverage probability of our trained MNRE network with SWYFT (vertical axis) as a function of the confidence level (horizontal axis). The purple line demonstrates the coverage of our network, with the goal of perfect coverage denoted by the diagonal black dashed line. The dotted lines indicate the coverage for the 68th, 95th, and 99.7th percentiles, whereas the shaded region corresponds to the Jeffrey's interval (see the text for further details).

expected coverage probability of our trained network as a function of confidence levels for all 1D and 2D marginalized posteriors using 5000 unique realizations drawn randomly from our posterior volume. Optimal network performance is demonstrated by the black dashed curves. If the coverage probability resides above the black dashed line, the network coverage is deemed conservative (i.e. our actual error rate is lower than the theoretical error implying larger than expected posteriors uncertainties), whereas if it is below the diagonal it is considered overconfident. For the vast majority of our astrophysical parameters, our coverage probability is aligned or above the black dashed line indicating strong coverage performance. Following the performance of our coverage tests we are confident that our results are robust.

n is the total number of samples from the joint model and k is the number of times the HPDR predicted by the network does not contain the true astrophysical parameters.

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