Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **508,** 2923–2936 (2021)
Advance Access publication 2021 September 25

https://doi.org/10.1093/mnras/stab2737

GLOBALEMU: a novel and robust approach for emulating the sky-averaged 21-cm signal from the cosmic dawn and epoch of reionization

H. T. J. Bevins ,[1]★ W. J. Handley,[1,2] A. Fialkov,[2,3] E. de Lera Acedo[1,2] and K. Javid[1,2]

[1]*Astrophysics Group, Cavendish Laboratory, J. J. Thomson Avenue, Cambridge CB3 0HE, UK*
[2]*Kavli Institute for Cosmology, Madingley Road, Cambridge CB3 0HA, UK*
[3]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

Accepted 2021 September 20. Received 2021 September 20; in original form 2021 April 16

ABSTRACT

Emulation of the Global (sky-averaged) 21-cm signal with neural networks has been shown to be an essential tool for physical signal modelling. In this paper, we present GLOBALEMU, a Global 21-cm signal emulator that uses redshift as a character-defining variable alongside a set of astrophysical parameters to estimate the signal brightness temperature. Combined with physically motivated data pre-processing, this makes for a reliable and fast emulator that is relatively insensitive to the network design. GLOBALEMU can emulate a high-resolution signal in 1.3 ms in comparison to 133 ms, a factor of 102 improvement, when using the existing public state-of-the-art 21CMGEM. We illustrate, with the standard astrophysical models used to train 21CMGEM, that GLOBALEMU is almost twice as accurate and for a test set of ≈1700 signals we achieve a mean root mean squared error of 2.52 mK across the band $z = 7$–28 [≈10 per cent the expected noise of the Radio Experiment for the Analysis of Cosmic Hydrogen (REACH)]. The models are parametrized by the star formation efficiency, f_*, minimum virial circular velocity, V_c, X-ray efficiency, f_X, cosmic microwave background optical depth, τ, the slope and low energy cut-off of the X-ray spectral energy density, α and ν_{min}, respectively, and the mean free path of ionizing photons, R_{mfp}. GLOBALEMU provides a flexible framework for easily emulating updated simulations of the Global signal and in addition the neutral fraction history. The emulator is pip installable and available at https://github.com/htjb/globalemu. GLOBALEMU will be used extensively by the REACH collaboration.

Key words: early Universe – reionization – first stars – dark ages – software: data analysis – software: simulations.

1 INTRODUCTION

The Global 21-cm signal from the cosmic dawn (CD) and epoch of reionization (EoR), if observed, will provide detailed information about the large scale properties of the early universe. The observable signal is the sky-averaged 21-cm emission from the spin flip transition in neutral hydrogen at redshifts $z = 5$–50 and redshifted to frequencies of approximately $\nu = 50$–200 MHz.

An absorption signal was reported at 78 MHz by the Experiment to Detect the Global Epoch of Reionization Signature (EDGES) collaboration in 2018 (Bowman et al. 2018). However, the reported signal is significantly larger in amplitude than that predicted by standard Lambda cold dark matter cosmology (Reis, Fialkov & Barkana 2021) and there are concerns about the data analysis used (Hills et al. 2018; Singh & Subrahmanyan 2019; Sims & Pober 2020; Bevins et al. 2021). Efforts are underway to make further observations of the signature with a variety of different radio telescopes including SARAS (Shaped Antenna measurement of the background RAdio Spectrum; Singh et al. 2018), REACH (Radio Experiment for the Analysis of Cosmic Hydrogen; de Lera Acedo 2019), PRIZM (Probing Radio Intensity at High-Z from Marion; Philip et al. 2019), LEDA (Large-aperture Experiment to Detect the Dark Ages; Price et al. 2018), DAPPER (Dark Ages Polarimeter PathfindER, https://www.colorado.edu/project/dark-ages-polarimeter-pathfinder/), and MIST (Mapper of the IGM Spin Temperature, http://www.physics.mcgill.ca/mist/) among others.

The intensity of the signal is measured against the radio background, typically assumed to be equal to the cosmic microwave background (CMB) temperature, and characterized by an absorption trough and an emission at late redshifts. The relative magnitude of the signal features is determined by various astrophysical processes including the Wouthuysen–Field effect (Wouthuysen 1952; Field 1959), Lyman α heating and CMB heating (Chuzhoy & Shapiro 2007; Venumadhav et al. 2018; Mittal & Kulkarni 2020; Villanueva-Domingo, Mena & Miralda-Escudé 2020; Reis et al. 2021), X-ray heating, and ionization of the hydrogen gas by UV emission (Madau, Meiksin & Rees 1997). A detailed discussion of the physics describing the Global 21-cm signal can be found in Furlanetto, Oh & Briggs (2006), Pritchard & Loeb (2012), Barkana (2016), and Mesinger (2019). The physical processes themselves and hence the Global signal can be characterized by a set of astrophysical parameters (see Section 3 and Cohen et al. 2020): the star formation efficiency, f_*, the minimal virial circular velocity, V_c, the X-ray efficiency, f_X, the CMB optical depth, τ, the slope of the X-ray spectral energy density (SED), α, the low energy cut-off of the X-ray SED, ν_{min}, and the mean free path of ionizing photons, R_{mfp}.

★ E-mail: htjb2@cam.ac.uk

Hybrid approaches are used to calculate realizations of the 21-cm signal, over a large cosmological volume and redshift range, which then can be averaged at every redshift separately to give the Global signal (e.g. Mesinger, Furlanetto & Cen 2011; Visbal et al. 2012; Fialkov & Barkana 2014; Cohen et al. 2017; Reis et al. 2021). Each simulation takes several hours to perform on a desktop (Monsalve et al. 2019) and though this is much faster than hydrodynamical simulations this time is too long to allow us to constrain astrophysical parameters using data. Therefore, the desire to emulate the Global signal with neural networks, trained on the results of the large-scale simulations, has arisen. The neural networks can produce a realization of the Global signal in a fraction of a second by interpolating between the simulated cosmological and astrophysical models. This means that they can be used to physically model the signal in, for example, Bayesian nested sampling loops,[1] as in the REACH data analysis pipeline (Anstey, de Lera Acedo & Handley 2020), in which millions of calculations need to be made to infer cosmological parameters [List & Lewis 2020; Liu & Shaw 2020; Chatterjee, Choudhury & Mitra 2021; Sims et al. (in preparation)].

A number of papers have considered emulation of the 21-cm power spectrum using convolutional neural networks and other techniques (e.g. Jennings et al. 2018; Schmit & Pritchard 2018; Mondal et al. 2020). At the time of writing, 21CMGEM is the only publicly available emulator used to accurately, with a maximum normalized root mean squared error (RMSE) of 10.55 per cent, and quickly (see Section 6.1) emulate the Global 21-cm signal (Cohen et al. 2020).[2] It has previously been used to provide constraints on the parameter space of the 21-cm signal using EDGES high-band data (Monsalve et al. 2019). The emulator uses principal component analysis (PCA; Pearson 1901), the seven astrophysical parameters detailed above and in Section 3, additional information about the mean collapsed fraction of haloes as a function of redshift, $f_{\rm coll}(z)$, and the fraction of X-ray energy above 1 keV, $f_{\rm XR>1\,keV}$, and 2 keV, $f_{\rm XR>2\,keV}$, and relies on a division of the signal into two or three distinct segments defined by the turning points. It involves the application of a decision tree for classification and several regression neural networks estimating PCA components, the frequencies and temperatures of turning points as well as additional parameters such as the frequency at which the neutral fraction equals 0.16, $\nu(x_{\rm H\,I} = 0.16)$.

In this paper, we present GLOBALEMU that uses a novel and robust approach with a single small-scale neural network to emulate the Global 21-cm signal given a comprehensive set of astrophysical parameters and redshift range. Where previously 21CMGEM was designed to take in astrophysical parameters and return a low dimensional representation of the Global signal as a function of redshift, GLOBALEMU takes in the same astrophysical parameters and redshift and returns the signal temperature at the corresponding redshift (see Fig. 1). This greatly simplifies the complexity of the relationship being learnt by the neural network. It means we can achieve accurate results, with the smoothness of the signal imposed by the interpolation of the neural network between signals in the training data set, using a small network and without the need for a compressed representation of the signals like PCA where there is a potential loss of information. Additionally, GLOBALEMU relies on a physically motivated pre-processing of the training data and can emulate a high-resolution, $\delta z = 0.1$ over the range $z = 5$–50, Global 21-cm signal in ≈1.3 ms. GLOBALEMU will be used extensively by the REACH collaboration and has been designed to have an average accuracy less than or equal to 10 per cent of the expected noise in the REACH system, estimated at 25 mK (REACH Collaboration, in preparation).

GLOBALEMU is written in PYTHON using `tensorflow` and the `KERAS` backend; it is pip installable via `pip install globalemu` and available at https://github.com/htjb/globalemu. It is flexible enough to be retrained on any set of Global 21-cm signal models while maintaining the novel design and physically motivated pre-processing. We provide a demonstration of its accuracy and efficiency in this paper using the same data used to train 21CMGEM and the corresponding trained models are publicly available on GitHub. We use GitHub actions to perform continuous integration.

In Section 2, we describe the novel approach used to parametrize GLOBALEMU. Section 3 describes the training and test data used to illustrate the capabilities of GLOBALEMU in this paper and the astrophysical parameters in the simulations of the Global signal. We then describe the predominantly physically motivated pre-processing of the inputs and outputs of the neural network in Section 4. A discussion of the neural network structure follows in Section 5 and the quality of emulation is assessed in Section 6. We conclude in Section 7.

2 PARAMETRIZING THE PROBLEM

There are several approaches that can be used to emulate the Global 21-cm signal with a neural network. The ultimate goal of the emulator is to take in a set of astrophysical parameters and return an estimate of the signal brightness temperature as a function of redshift, $\delta T(z)$, where the relationship has been learned from detailed numerical simulations. This can be done directly with a neural network that returns a value of δT for each redshift data point it has been trained on. However, assuming that the network is trained on high-resolution signals this would result in a large number of outputs, making it hard to train, and would be limited in predictive power to specific values of redshift. The process can also be achieved by estimating, via a neural network, coefficients of a compressed representation of the signal space. For example, using PCA as with 21CMGEM (Cohen et al. 2020) or learning coefficients of basis functions for polynomials or wavelets that when combined return the Global signal. However, while this approach reduces the number of outputs compared to a direct emulation, if incorrectly designed this can result in information loss and is equally limited in predictive power.

We take the novel approach of using redshift as an input to the network alongside the astrophysical parameters and returning from the network a single temperature corresponding to the given redshift. This is beneficial for two reasons; the small number of inputs and outputs means that the network can retain a simple structure and secondly the network will be able to interpolate between the values of redshift that it has been trained on. The smooth structure of the output from the network is guaranteed by the smooth interpolation performed by the neural network and by the smooth structure of the signals it is learning. Vectorized calls to the network are used to emulate the temperature as a function of redshift.

In GLOBALEMU, we also provide the ability to emulate the evolution of the neutral fraction, $x_{\rm H\,I}$, of hydrogen as a function of redshift. We use an identical framework as when emulating the Global signal to do this with a set of astrophysical parameters and a redshift as inputs to a neural network and a corresponding value of $x_{\rm H\,I}$ as an output.

[1] Note that here we are not referring to a Bayesian Neural Network (see Javid et al. 2020) but rather parameter optimization algorithms such as POLYCHORD (see Handley, Hobson & Lasenby 2015a, b).

[2] During review of this work, a pre-print describing the global signal emulator 21CMVAE has been published (Hellum Bye, Portillo & Fialkov 2021) in which the comparative performance of GLOBALEMU is discussed.

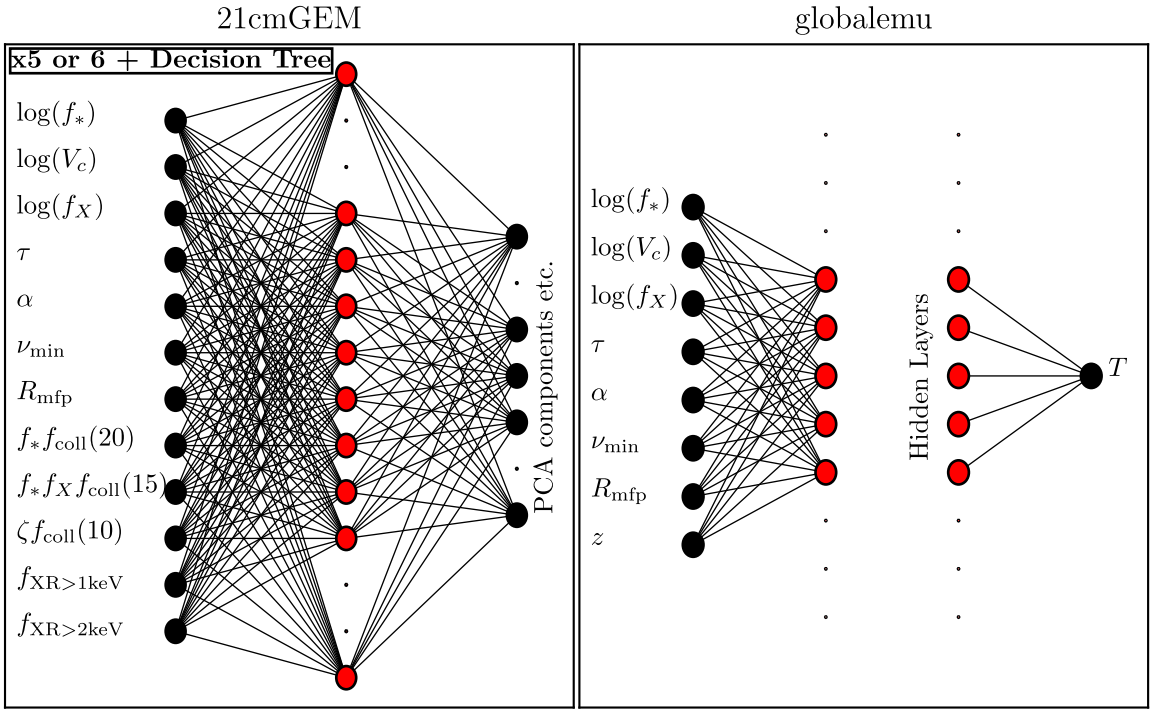

Figure 1. *Left-hand panel:* There currently exists only one other global signal emulator, 21CMGEM, and we provide here an illustration of the regression neural networks used there. Note that 21CMGEM uses either five or six of these networks and a decision tree when making predictions. For one of the regression networks used, the only input parameters are the seven astrophysical parameters detailed in Section 3 (excluding redshift). However, for the others there are an additional five derived parameters (see the text and Cohen et al. 2020) and we illustrate the full set of 12 inputs. The number of output nodes depends on the specific application of the network and they can correspond to either PCA components (four nodes), additional parameters such as $\nu(x_{\rm H\,I} = 0.16)$ (one node), and the frequencies and brightness temperatures of turning points in the signal (seven or five nodes). The hidden layer in all of the 21CMGEM regression networks has 40 nodes. For a full illustration of the 21CMGEM algorithm and detailed description, see fig. 11 and section 4.2 in Cohen et al. (2020). *Right-hand panel:* An illustration of the GLOBALEMU neural network. Note the use of only one network to emulate the global signal in comparison to the five or six used for 21CMGEM. Here, the input layer has eight nodes (seven astrophysical parameters plus redshift) and the output layer is a single node returning the brightness temperature corresponding to the input redshift. We show a sizable hidden layer structure here with the red nodes and '...'; however, we note that the reduced number of inputs and outputs implies that a small architecture will be sufficient to achieve a high level of accuracy in the emulation (see Section 5).

We have built a network that can emulate the Global 21-cm signal to a high degree of accuracy without the need for L1 and L2 regularization, dropout (Srivastava et al. 2014), batch normalization (Ioffe & Szegedy 2015), or other similar concepts. We have achieved this by focusing on the pre-processing of the network inputs and outputs in the desire to set our problem up in a way that is simple to solve with a basic neural network of a 'reasonable' size.

3 THE TRAINING AND TEST DATA

In this paper, we use the same model signals and corresponding astrophysical parameters used to train and test 21CMGEM (available at https://doi.org/10.5281/zenodo.4541500; Cohen et al. 2021). Examples of the Global signals from the training set are shown in Fig. 2. In total, the data set contains 27 292 training models and 2174 test models with each model being dependent on seven astrophysical parameters. Each Global signal in the data set has 451 redshift data points and this means that each signal corresponds to 451 training points with the same astrophysical parameters and different redshifts. Therefore, for GLOBALEMU the 27 292 training models become 12 308 692 training data points. However, we continue throughout this paper to talk generally of training signals rather than data points because the emulator will be used to determine the signal structure over a redshift range (see Section 6 for more details).

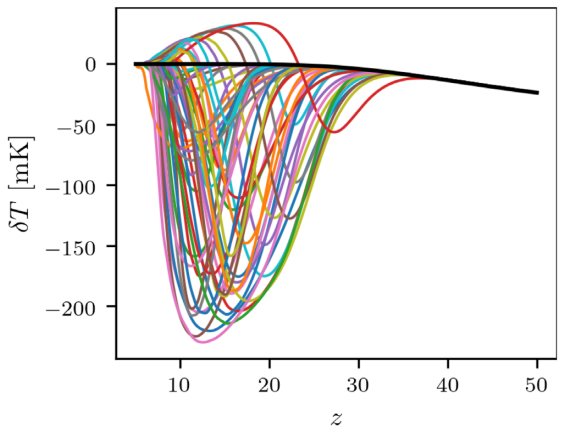

Figure 2. A subsample of 50 Global 21-cm signals from the 21CMGEM training set used here to demonstrate the efficiency of GLOBALEMU. The signals show the expected variety of structure with deep and shallow absorption troughs caused by Lyman α coupling and terminated by X-ray heating. We also see emission against the CMB background at low redshifts in some of the models where there has been sufficient heating. Also shown in black is the Astrophysics Free Baseline (AFB; Section 4.1) that we model and remove from the training signals before we pass them through the neural network. Subtraction of the AFB prevents our network from attempting to learn a steadily decreasing temperature at high redshifts prior to star formation.

The data pre-processing is explained in detail in Section 4. Explicitly, the inputs (Cohen et al. 2017, 2020; Monsalve et al. 2019) to the neural network when training on the 21CMGEM data are as follows (ranges are based on those in the 21CMGEM training data set; see section 2.5 of Cohen et al. 2020):

(i) f_*: The star formation efficiency takes values in the range of 0.0001–0.5 and characterizes the amount of gas converted to stars in the dark matter haloes. A low star formation rate results in a low Lyman α flux and late onset of X-ray heating, leading to a shallower absorption trough and weak or non-existent emission above 0 mK in the signal.

(ii) V_c: The minimal virial circular velocity has a value in the range of 4.2–100 km s^{-1} and is proportional to the cube root of the minimum threshold mass for star formation. A low value of V_c corresponds to a small minimum mass threshold that in turn leads to an earlier onset of Lyman α coupling, responsible for the absorption feature in the Global 21-cm signal, and shifts the minimum of the signal to higher redshifts.

(iii) f_X: The X-ray efficiency of sources has a range between 0 and 1000 and a high value corresponds to a high total X-ray luminosity. This leads to an earlier onset of X-ray heating that also shifts the minimum of the signal to higher redshifts, contributes to a shallower absorption, and results in a significant emission feature during reionization at low redshifts.

(iv) τ: The CMB optical depth in the 21CMGEM data sets takes a value in the range of 0.04–0.2 and a higher value of τ corresponds to a higher value of the ionizing efficiency of sources, ζ. For high τ, we would see an earlier reionization of the hydrogen gas. We note that τ is given as 0.054 ± 0.007 by Planck Collaboration VI (2020) and that this falls at the lower end of the range in our training and testing data sets. More recent parameter studies have explored lower values of τ in greater detail (Reis et al. 2021). However, the 21CMGEM data are sufficient to demonstrate the abilities of GLOBALEMU.

(v) α: The power defining the slope of the X-ray SED with a range given by 1–1.5. The Global 21-cm signal is expected to have a very weak dependence on α with the largest effect happening at low redshifts.

(vi) $\nu_{\rm min}$: The low energy cut-off of the X-ray SED has a range of 0.1–3 keV. Low values of $\nu_{\rm min}$ correspond to a soft X-ray SED, efficient X-ray heating, and a weak absorption feature in the 21-cm signal.

(vii) $R_{\rm mfp}$: The mean free path of ionizing photons, with a range of 10–50 Mpc. $R_{\rm mfp}$ is expected to have a very weak effect and only at low redshifts (see e.g. Monsalve et al. 2019). A low $R_{\rm mfp}$ corresponds to a slower ionization of the neutral hydrogen gas.

(viii) z: The redshift of the 21-cm brightness temperature is a measure of time and provides details about when each feature of the signal occurred. For example, the brightness temperature is expected to reach 0 mK, corresponding to the end of the EoR, at low redshifts or more recent times. It is interchangeable with frequency given that the rest frequency, ν_r, of the 21-cm line is 1420 MHz

$$z + 1 = \frac{\nu_r}{\nu}. \tag{1}$$

To ensure that we make a fair comparison of our results with those found when using 21CMGEM, we make the same physically motivated cuts to the test data as are detailed in section 2.4 of Cohen et al. (2020). This equates to limits on the ionizing efficiency of sources, $\zeta < \zeta_{\rm max} = 40\,000 f_*$, and on the neutral fraction history at $z = 5.9$, $x_{\rm H\,I}(z = 5.9) < 0.16$. Respectively the limits are motivated by stellar models (Bromm, Kudritzki & Loeb 2001) and quasar absorption troughs (McGreer, Mesinger & D'Odorico 2014). We also note that some of the parameters in the testing data have different ranges and the ranges are as follows; f_*: 0.0001–0.5, V_c: 4.2–76.5 km s^{-1}, f_X: 0–10, τ: 0.055–0.1, α: 1–1.5, $\nu_{\rm min}$: 0.1–3 keV, and $R_{\rm mfp}$: 10–50 Mpc. In total, the final testing data set is comprised of 1703 models.

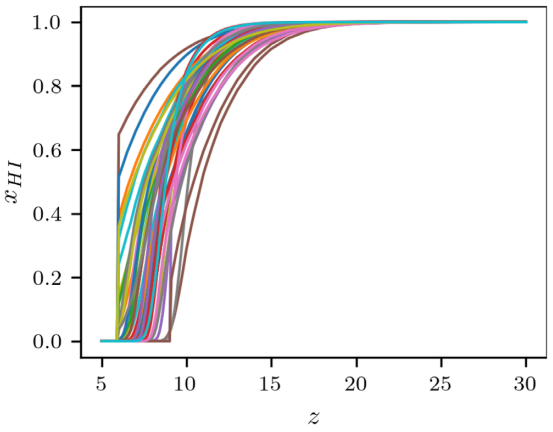

Figure 3. A subsample of 50 neutral fraction histories from the training set used in this paper. At high redshift, the hydrogen in the universe is predominantly neutral and consequently $x_{\rm H\,I} = 1$. As the gas is ionized by UV emission from the first stars that form, the neutral fraction decreases until $x_{\rm H\,I} = 0$ at the end of the EoR.

GLOBALEMU includes a simple to use PYTHON graphical user interface (GUI)[3] in which the variation of the signal with each of the astrophysical parameters listed above can be explored in more detail. We note that the GUI is a feature made possible by the speed of emulation when using GLOBALEMU (see Section 6). There is an equivalent GUI for the neutral fraction history emulation.

As previously stated, GLOBALEMU is not limited to emulating signals modelled with the above astrophysical parameters. It is flexible enough that more complicated astrophysical relationships can also be emulated. For example, one explanation for the unexpected depth of the EDGES absorption trough is the presence of a higher than expected radio background that can be characterized with a quantity $f_{\rm radio}$ determining the normalization of the radio emissivity (assuming that the source of the excess radio background is stellar; Reis, Fialkov & Barkana 2020). GLOBALEMU is in principle capable of being trained on models that consider $f_{\rm radio}$ in addition to the above seven astrophysical parameters and redshift as inputs since it assumes nothing about the astrophysical parameters themselves. Equally, GLOBALEMU could be trained on less complex models.

For the neutral fraction, $x_{\rm H\,I}$, we use a set of models produced as a by-product of the detailed 21CMGEM Global signal simulations. The data set is smaller with 10 047 training models and 791 test models; however, the relationship between the astrophysical parameters and $x_{\rm H\,I}$ is expected to be (and shown to be, Section 6) simpler. We note that for $z \gtrsim 30$ the neutral fraction is expected to be always 1 and so we only emulate the neutral fraction over the range $z = 5$–30. The models have not been released publicly but the parameter ranges are the same for this data set as detailed above. A subsample of the training models is shown in Fig. 3.

We note that a non-uniform coverage of the parameter space in the training data set, as with the 21CMGEM Global signal and neutral

[3]After installation via PIP or from source, the GUI can be called from the terminal using the command *globalemu*. See the documentation at https://globalemu.readthedocs.io/ for more details.

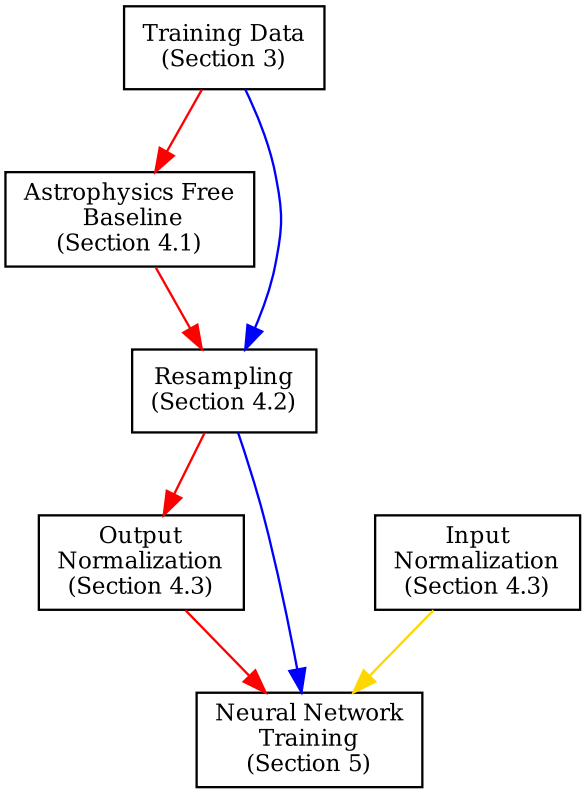

Figure 4. The pre-processing applied to the training data in GLOBALEMU. Each box is outlined in more detail in the corresponding sections. The red path is the pre-processing steps used for the Global 21-cm signals, the blue path is for the neutral fraction histories, and the gold path is steps that occur when training both neural networks.

fraction data, may introduce bias in the neural network. The network will tend to learn regions of the astrophysical parameter space where the sampling is heavier better than others. For the purposes of illustrating the accuracy of the emulation in this paper, this is not an issue. However, it can become an issue when using an emulator to physically model a signal in a data set where parameter estimation may be biased towards a false set of parameters. Training a network on a more uniform data set can alleviate this issue and we leave further exploration of this for future work.

4 DATA PRE-PROCESSING

The details in the following discussion[4] outline the pre-processing for the network predicting the Global 21-cm signal. In Section 4.4, we briefly discuss the pre-processing for the neutral fraction networks that is a largely similar process. The pre-processing is summarized as a flow chart in Fig. 4.

4.1 Astrophysics free baseline subtraction

In the region where the structure of the Global 21-cm signal is expected to be dominated by collisional coupling, it is independent on the seven astrophysical parameters used here as inputs to the emulator. This means that, in the corresponding redshift range, each of the signals in our training and testing data sets has the same brightness temperatures. To prevent our network unnecessarily attempting to learn a non-trivial structure in this region, we can treat it as an astrophysics free baseline (AFB), and model and remove it from the signals before they are passed to the network for training. By doing this, our network will learn a simpler relationship at high redshift between the parameters and $\delta T(z)$ than the existing steadily decreasing trend (see Fig. 2).

In Appendix A, we give an approximate calculation of the AFB for the simulated signals that comprise the training data sets. The calculation is approximate because it follows the mean evolution of the signal and in contrast the simulations are produced over large-scale cosmological volumes evolved over cosmological time and then averaged. We therefore normalize our result to the temperature of the signals in the training set at $z = 50$ and find that this is sufficient to represent the astrophysics free component of the models.

As stated, the AFB is then subtracted from the models before training the network and added back in after making predictions.

21CMGEM uses five extra parameters in addition to the seven astrophysical parameters used here. In principle, these parameters could be passed to the neural network during training due to the flexible nature of GLOBALEMU. However, three of these additional parameters rely on the fraction of mass contained in haloes above the minimum cooling threshold, $f_{\rm coll}(z)$, and help the network learn the signal structure at high redshift where collisional coupling (cosmology) dominates. Here, we do not consider these parameters, which are derived from the seven astrophysical parameters described above, as we are instead removing the AFB. The final two parameters, for reference, are the fractions of X-ray energy above 1 and 2 keV. These parameters are added to further characterize the X-ray SED, but we find with GLOBALEMU that we do not need to consider them to achieve accurate results.

4.2 Resampling of signals

The turning points, and gradients between them, of the Global 21-cm signal encode all of the information about the efficiency of Lyman α coupling, X-ray heating, and reionization. They are therefore highly dependent on the relevant astrophysical parameters and in the region where the features typically occur variation in the signals is significant. The original signal models are sampled uniformly in redshift and there is no particular physical motivation for this. However, to improve the quality of modelling we resample the signals at a higher rate across the redshift ranges that typically correspond to the locations of the turning points and at a lower rate where the signal structure deviates less from the 'average' signal (e.g. above $z \approx 30$ where the signal is free of astrophysics).

To do this, we look at the variation in the signal amplitudes, after subtracting the AFB, across the training data set

$$\Delta(\delta T(z)) = \delta T_{\rm max}(z) - \delta T_{\rm min}(z), \tag{2}$$

and we treat this as a probability distribution

$$P(\Delta(\delta T(z))) = \frac{\Delta(\delta T(z))}{\sum_z \Delta(\delta T(z))}. \tag{3}$$

Where the variation in the signal at a given redshift across the training data set is large, the probability distribution is also large (see Fig. 5). We then calculate the corresponding cumulative distribution function (CDF) and use inverse transform sampling to produce a new redshift distribution with a high sampling rate in regions of high

[4]Specifically, the discussion details the steps used in this paper when training with the 21CMGEM data. However, the various pre-processing steps outlined can be switched on and off by a user when training an emulator with GLOBALEMU.

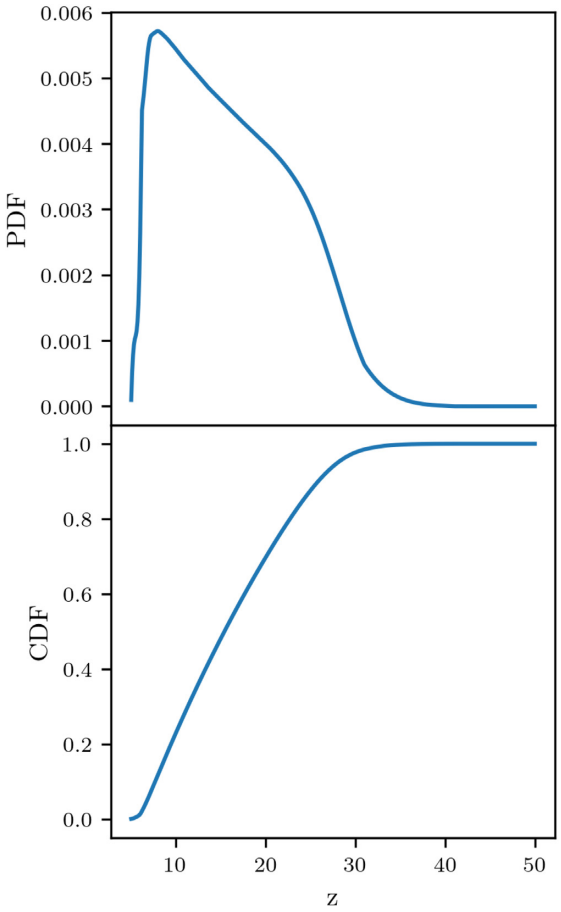

Figure 5. *Top panel:* The probability distribution calculated from the difference between the maximum and minimum signal temperatures in the 21CMGEM training data set using equations (2) and (3). *Bottom panel:* The CDF corresponding to the probability distribution in the top panel. We use this CDF to resample the training Global 21-cm signals in order to capture the variation at low redshifts across the distribution and allow the emulator to better learn this behaviour.

variation. For each signal, we can then perform an interpolation to get the corresponding δT values.

4.3 Output and input normalization

Neural networks typically perform better when the outputs and inputs are of order unity and uniformly distributed. Hence, it is typical to manipulate the data sets via logarithms, normalization, and/or standardization to improve performance.

After subtracting the AFB and resampling our signals, we also proceed to divide the signals by the standard deviation across the training data set. This type of scaling was motivated by the typically used standardization technique; however, we wanted to ensure that when scaling our signals a value of $\delta T = 0$ remained as 0 because it holds physical meaning (an equivalence between the spin temperature and radio background, $T_{\rm s} = T_{\rm r}$). The signals shown in Fig. 2, as seen by the neural network, after pre-processing are shown in Fig. 6.

For the input redshift distribution, we transform our resampled redshifts back on to a uniform distribution between 0 and 1, before they are input into the network, using the CDF detailed in the previous section. It is the combination of resampling and uniform redshift

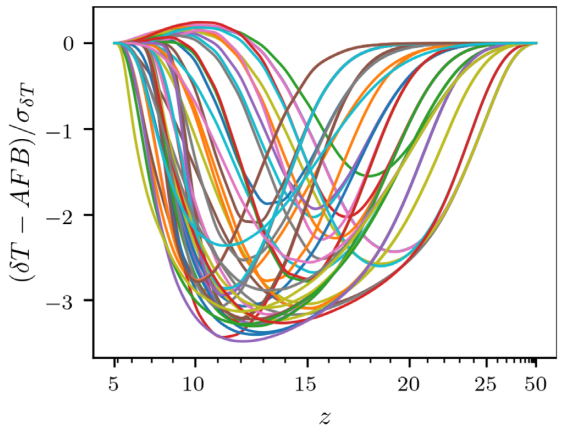

Figure 6. The equivalent signals from Fig. 2 after pre-processing. Subtraction of the AFB and the subsequent resampling mean that the important information encoding the dependence on the astrophysical parameters is retained and appropriately emphasized in the training data. Here, we have plotted the resampled redshift data points as being uniformly distributed since this is how the network is set up to interpret the input. The following division by the standard deviation across the training data set scales the signals to order unity without changing the physically significant value of $\delta T(z) = 0$ where the spin temperature of the neutral hydrogen is equivalent to the radio background temperature. Minor ticks are at intervals of one on the x-axis.

input that ensures the neural network 'sees' 'stretched' signals as in Fig. 6. This technique allows the neural network to interpolate the signal at redshifts it has not been trained on to a higher degree of accuracy where the signals vary greatly than if we had used uniform sampling.

For the other input astrophysical parameters, we use a Min–Max normalization scaling each feature between 0 and 1. For example, considering the distribution of the CMB optical depth, $\boldsymbol{\tau}$ in our training data as a vector we normalize it such that

$$\tilde{\boldsymbol{\tau}} = \frac{\boldsymbol{\tau} - \boldsymbol{\tau}_{\rm min}}{\boldsymbol{\tau}_{\rm max} - \boldsymbol{\tau}_{\rm min}}. \tag{4}$$

The decision to use this type of normalization was arrived at after testing standardization, Min–Max normalization, and division by the max values for the input parameters while maintaining the physically motivated pre-processing for the signal temperatures detailed in the above subsections.

For $f_{\rm X}$, f_*, and $V_{\rm c}$, the distributions are uniform in log-space and so we perform the Min–Max normalization on the logarithm of these variables and use these as our inputs.

4.4 $x_{\rm H\,I}$ pre-processing

As discussed, we provide provision in GLOBALEMU to emulate the neutral fraction of hydrogen as a function of redshift. For this network, the pre-processing just involves resampling of the signals since

(i) The equivalent AFB for the neutral fraction has a value of 1 at all redshifts and to subtract this from our training data set would invert our signals providing no benefit to training.

(ii) The neutral fraction has a value between 0 and 1 by definition and so we do not need to normalize the output of the network to be of order unity.

The benefits of performing resampling for the neutral fraction histories are the same as for the Global signal network. It allows the

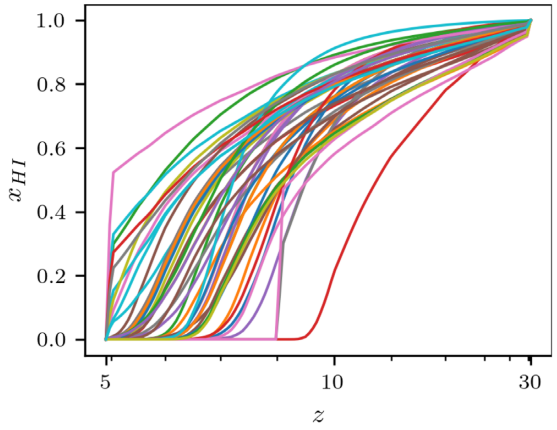

Figure 7. The equivalent neutral fraction histories from Fig. 3 after pre-processing. For the neutral fraction histories, since the signals are already of order unity and subtraction of the equivalent AFB would not be beneficial, the pre-processing just involves resampling of the signals around regions of high variation. As with Fig. 6, the minor ticks are at intervals of one on the x-axis.

network to learn the variation in the training models and interpolation across redshift with a higher degree of accuracy. We perform the resampling with the equivalent of equations (2) and (3). Fig. 7 shows the same set of neutral fraction histories as in Fig. 3 after pre-processing.

5 NEURAL NETWORK STRUCTURE

As stated, the goal with GLOBALEMU is to maintain a simple network that is highly accurate without having to use dropout, regularization, batch normalization, etc. However, in the design of any neural network the optimizer, the architecture, loss function, activation function, and learning rate are core considerations.

5.1 Architecture

Dropout (Srivastava et al. 2014) and the commonly used L1 and L2 regularization are typically employed to prevent overfitting where the network learns the training data to such a high degree of accuracy that it is unable to generalize. Overfitting is generally a result of using a neural network that is too big and has an excessive number of layers and nodes. On the other hand, a network that is too small often produces poor-quality predictions and consequently the aim is to produce a 'reasonably' sized network. The scope of what constitutes a reasonably sized network is dependent on the number of input/output nodes, the variation in the training data, and the complexity of the relationship between the inputs and outputs.

By using the novel approach of having redshift as an input to the network, both our Global signal and neutral fraction emulators have, in the case of the 21CMGEM data, eight input nodes and one output node, meaning that our network can remain small in size. Additionally, we have made a significant effort to simplify the problem with physically motivated pre-processing, which also helps us to reduce what constitutes as a 'reasonable' size for our networks.

GLOBALEMU is set up in such a way that the number of layers and layer sizes can be adjusted by the user. As a result, we do not provide a prescription of what constitutes a 'reasonable' size as this may not be pragmatic. We note, however, that a significant effort can be undertaken to determine the optimum 'reasonable' architecture that

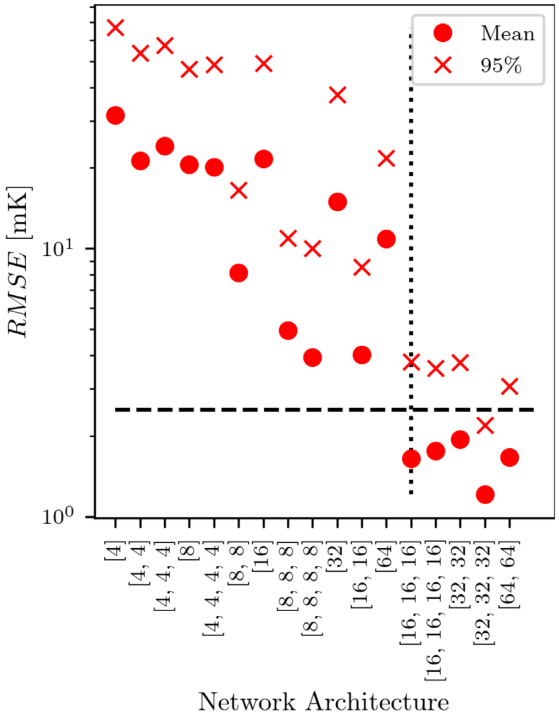

Figure 8. The mean and 95 percentile RMSE (see Section 6.2) for a set of different network architectures trained for 12 h (approximately 250 epochs) on an HPC with the 21CMGEM Global signal training data and assessed with the corresponding test data. The architectures have between 1 and 4 layers of varying sizes between 4 and 64 nodes. They are ordered based on the number of weights in the network (equivalent to the number of connections) as this is a useful measure of network size and an indication of predictive power. The graph is used to determine a 'reasonable' architecture considering the practical target accuracy of on average 10 per cent the expected noise of a Global 21-cm experiment (here illustrated by the black dashed line at 2.5 mK). Throughout the rest of the paper, we use a network with 3 layers each consisting of 16 nodes, which is the first to produce a mean value within our target accuracy. Our choice is highlighted with a dotted vertical line.

maximizes accuracy and that this can also be impractical. Instead, we suggest that as a minimum requirement a 'reasonable' architecture for a trained GLOBALEMU model should meet the following criteria:

(i) The network should not overfit the training data; otherwise, the predictive power will be lost.
(ii) The network should have an average accuracy $\lesssim$10 per cent the noise of a typical Global 21-cm experiment (see Section 6.2 for a further discussion).

Based on the above criteria, the size of our input and output layers and a minimal exploration starting from a small network, trained with the pre-processed signals, and increasing the size until our accuracy criteria were met without overfitting (see Appendix B), we use a network with three hidden layers all of size 16 for both the Global signal and neutral fraction emulation in this paper.

Fig. 8 illustrates the processes used to determine our architecture for the Global network. We consider a set of different network sizes with one to four layers and 4, 8, 16, 32, or 64 nodes in each layer. For each of the tested networks, we run a 'full', 12 h on an HPC equating to approximately 250 epochs using the full 21CMGEM training data, training of GLOBALEMU. We then assess

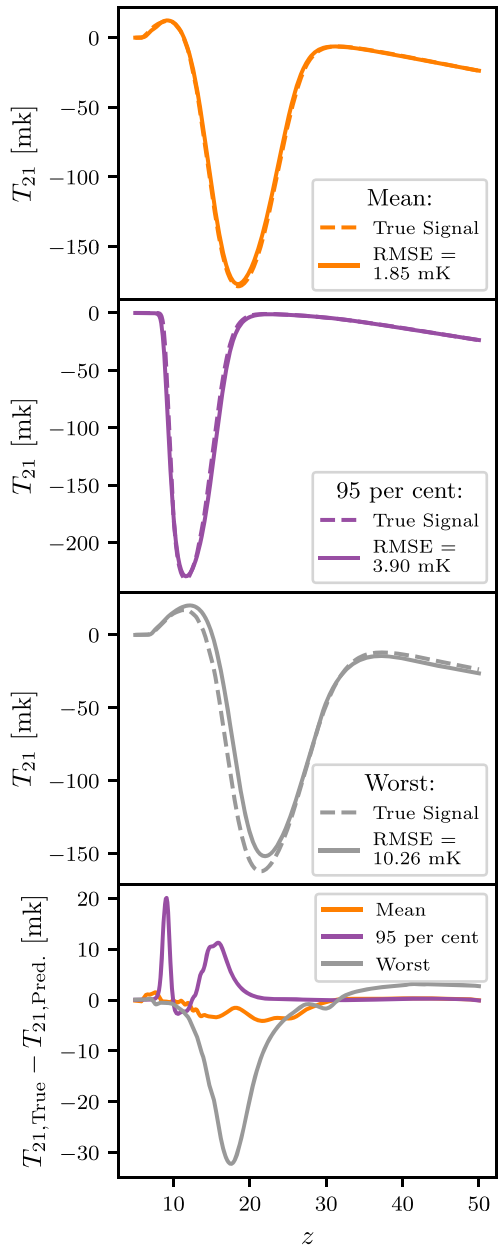

Figure 9. The top three panels show the mean, 95 percentile, and the worst emulations, respectively, based on the RMSE, for the Global 21-cm signal across the entire test set of 1703 models. The bottom panel shows the difference between the simulations and predictions as a function of redshift. Full details of the accuracy of the emulation can be found in Table 1 and a discussion can be found in the text.

the accuracy of the trained models using the ≈1700 testing models in the 21CMGEM data set. We compare the mean and 95 percentile RMSE (see Section 6.2) for each architecture. We find that a network of size [16, 16, 16] is the first to meet our target accuracy of on average 10 per cent the expected noise in a Global 21-cm experiment.

While we may be able to achieve a better accuracy with a larger network, this pragmatic approach leads to a sufficiently accurate network for physical signal modelling in the data analysis pipeline of a Global experiment like REACH. We also note that a smaller network can be evaluated faster than a larger architecture and that this is important when we are making multiple evaluations inside a nested sampling loop.

5.2 Loss function and learning rate

GLOBALEMU uses the mean squared error, typical for a regression network, as the loss function. In the case of the Global signal, network is given by

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=0}^{N}(\delta T_{\mathrm{sim,i}}(z_i) - \delta T_{\mathrm{pred,i}}(z_i))^2, \tag{5}$$

where N is a batch size equivalent to the number of redshift data points in each signal. $\delta T_{\mathrm{sim}}(z)$ is the simulated signal temperature at a given redshift and $\delta T_{\mathrm{pred}}(z)$ is the emulated equivalent. GLOBALEMU trains the neural networks in batches primarily to prevent memory related issues since the training data can be large (≈27 000 models times 451 redshift points ≈12 million data points for the 21CMGEM data). We find that a reasonable batch size is equal to the number of redshift data points in each model.

For the 21CMGEM data and the GLOBALEMU framework, we determine an effective learning rate to be 0.001. As with the architecture, the learning rate can be adjusted by the user of GLOBALEMU to meet the requirements of the data that they are training on.

5.3 Optimizer

The neural network optimizer is used to change the network hyperparameters to minimize the loss function. There are a number of different optimizers available (Ruder 2016) and the choice can be dependent on the complexity of the problem and loss surface. A more robust optimizer is less likely to fall into and get stuck in local minima when training the network, resulting in more accurate emulation. Therefore, the choice of optimizer is important in designing an effective emulator. However, since GLOBALEMU is designed to minimize the complexity of the relationship between the inputs and outputs and an MSE loss surface is relatively smooth[5] our choice is less consequential. We use therefore the commonly applied ADAM (Kingma & Ba 2014) optimizer that is a momentum-based modified stochastic gradient descent algorithm.

5.4 Activation functions

For both the Global signal and neutral fraction history network, we use a *tanh* activation function in the hidden layers that can range between (−1, 1). However, for our final layer in the Global signal network we use a linear activation since the pre-processed temperature can be positive or negative and range between approximately −4 and 0.5. Similar consideration is given to the output layer in the neutral fraction network where we use a ReLU (Rectified Linear Unit) activation that ensures that the output is always positive. Again, the activation functions can be changed by the user of GLOBALEMU to meet the requirements of their data. We note that the above output layer activations are designed to prevent unphysical outputs and that this is a crucial consideration for any user.

[5]This can be assessed with a plot of the loss versus epoch number during training. We find that for the results presented in Section 6 the surface is smooth up until the loss has plateaued and training is complete at which point we see noise-like behaviour.

Table 1. Detailed results of the emulation using GLOBALEMU and the 21CMGEM training and test data for both the Global signal and the neutral fraction history. We find that GLOBALEMU achieves the desired accuracy of on average ≈10 per cent the expected noise of a typical Global 21-cm experiment (equating to ≈2.5 mK in the REACH band of $z = 7$–28). Of note are the recorded 95 per cent percentiles, the RMSE for which 95 per cent of the models have values smaller than or equal to, which are significantly lower than the maximum RMSE values. A discussion comparing the results of 21CMGEM and GLOBALEMU, in terms of $\widetilde{\text{RMSE}}$, can be found in the text. Briefly we find that our Global signal emulator has a maximum $\widetilde{\text{RMSE}}$ approximately half that achieved with 21CMGEM. For the neutral fraction, RMSE = $\widetilde{\text{RMSE}}$ and so we only report one set of results. We find a higher degree of accuracy here with an identical network and similar pre-processing indicating a simpler relationship.

		Global signal		Neutral fraction	
		$z = 5$–50	$z = 7$–28	$z = 5$–30	$z = 7$–28
RMSE	Minimum	0.30 mK	0.31 mK	0.09 per cent	0.08 per cent
	Mean	1.85 mK	2.52 mK	0.29 per cent	0.26 per cent
	95th percentile	3.90 mK	5.37 mK	0.47 per cent	0.44 per cent
	Maximum	10.26 mK	15.10 mK	1.12 per cent	0.65 per cent
$\widetilde{\text{RMSE}}$	Minimum	0.21 per cent	0.26 per cent	–	–
	Mean	1.12 per cent	1.53 per cent	–	–
	95th percentile	2.41 per cent	3.22 per cent	–	–
	Maximum	6.32 per cent	9.31 per cent	–	–

6 RESULTS

6.1 Emulation time

In Cohen et al. (2020), the reported average time taken per signal with 21CMGEM is 160 ms when emulating a set of signals in a vectorized call. Here, we compare the speed of 21CMGEM and GLOBALEMU by emulating the 1703 test models and taking an average time per signal. The tests are performed with MATLAB and PYTHON, respectively, on the same computer with the following processors: Intel® Core™ i3-10110U CPU @ 2.10 GHz × 4. For 21CMGEM, we make a vectorized call to the emulator as this results in a quicker performance than repeated single calls. We use a for loop to repeatedly call GLOBALEMU that currently does not support such vectorized calls as they are not needed for physical signal modelling in a nested sampling loop.

For GLOBALEMU, we record a total time of 2.29 s and a corresponding average time per signal of 1.3 ± 0.01 mK. In comparison when emulating the same signals in a vectorized call with 21CMGEM we record a total time of 227.18 s and an average time per signal of 133 ms. We therefore achieve a factor of 102 improvement in emulation time with GLOBALEMU. We note that when using the PYMATBRIDGE (https://github.com/arokem/python-matlab-bridge) PYTHON wrapper for MATLAB the average time taken to run a single prediction with 21CMGEM using a vectorized call is comparable to a direct call in MATLAB.

6.2 Measuring accuracy

In this section, we primarily consider the accuracy of the Global signal emulator because the neutral fraction network has a similar design. We note, as previously stated, that the relationship between the neutral fraction and the astrophysical parameters is expected to be simpler and therefore easier to learn.

To assess the accuracy of GLOBALEMU when emulating a Global 21-cm signal simulation, we use a combination of two metrics: the RMSE and the normalized RMSE given in Cohen et al. (2020) as

$$\widetilde{\text{RMSE}} = \frac{\text{RMSE}}{\max|\delta T_{\text{sim}}(z)|}, \tag{6}$$

where

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=0}^{N}(\delta T_{\text{sim,i}}(z_i) - \delta T_{\text{pred,i}}(z_i))^2}. \tag{7}$$

For the neutral fraction network, $\widetilde{\text{RMSE}}$ and RMSE [with $x_{\text{sim}}^{\text{H1}}(z)$ and the equivalent for the emulation in place of temperature] are equal since $\max|x_{\text{sim}}^{\text{H1}}(z)| = 1$. We assess the accuracy in the uniform redshift space and consequently our assessment is independent of the loss function used for training.

$\widetilde{\text{RMSE}}$ is a dimensionless quantity used by 21CMGEM and by assessing the quality of our network with this metric we can make direct comparisons between the two emulators. We also want our emulator to have an accuracy significantly lower than the expected noise floor of Global 21-cm experiments. This is required if the emulator is to be used to confidently model the Global 21-cm signal and draw conclusions about the astrophysics during the CD and EoR. To assess this accuracy requirement, we can use the dimensionful RMSE metric.

As highlighted in the previous section, we suggest an average accuracy of $\lesssim$10 per cent the expected noise of a Global 21-cm experiment such as REACH, equivalent to an RMSE $\lesssim$ 2.5 mK, to be a sufficient limit. Since the accuracy of emulation is a function of the bandwidth, we report the accuracy across the entire range of the simulations $z = 5$–50 ($z = 5$–30 for the neutral fraction network) and across the expected REACH Phase I bandwidth of $z = 7$–28 (REACH Collaboration, in preparation). The target is demonstratively achievable (see the following section). It is also a practical target, if we want to use GLOBALEMU for physical signal modelling, given that the noise in a 21-cm experiment has a fundamental effect on our confidence in any astrophysical parameter values inferred from the data and that this will likely be larger than the uncertainty introduced from GLOBALEMU.

6.3 Global 21-cm signal

Fig. 9 shows that the mean RMSE value across the redshift range $z = 5$–50 is 1.85 mK and that the maximum value is 10.26 mK. Further, Table 1 shows that performing the same calculation of the RMSE

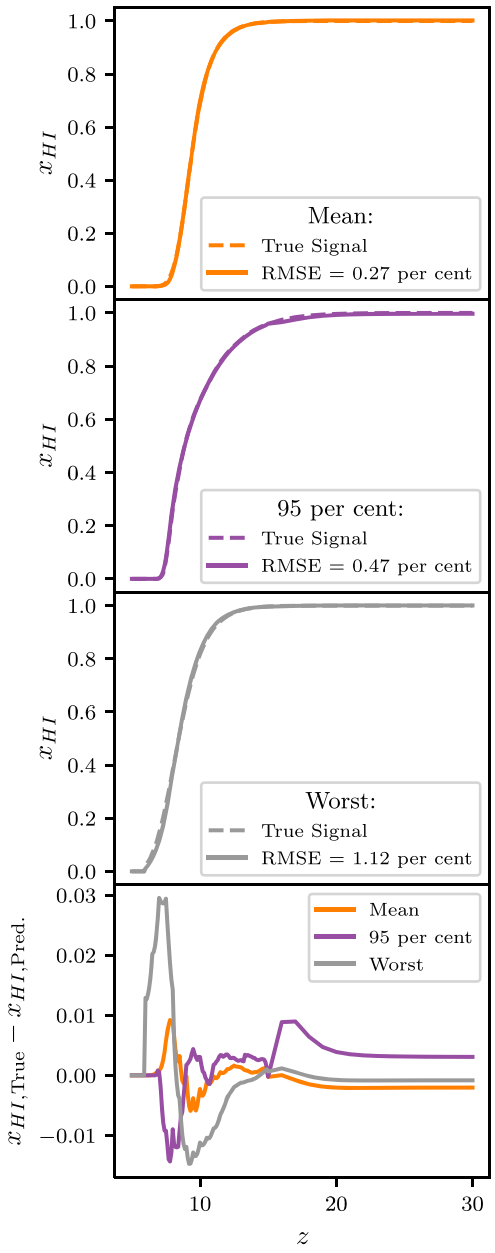

Figure 10. The top three panels show the mean, 95 percentile, and the worst emulations, respectively, for the neutral history across the entire test set of 791 models and the difference between the simulations and predictions is shown in the bottom panel. The level of accuracy here is higher than that for the Global signal, despite using a similar pre-processing and identical network, demonstrating that the relationship between the astrophysical parameters, redshift, and the network output is simpler here.

inside the REACH band, $z = 7$–28 gives a mean value of 2.52 mK that is very close to the desired 2.5 mK limit. We also report in the table the RMSE value for which 95% of the models have a value smaller than or equal to. In the REACH band, this equates to 5.37 mK. For all of the reported results, the 95 percentile is significantly lower than the maximum values (a factor of 3 for the Global signal and a factor of approximately 1.5–2 depending on bandwidth for the neutral fraction histories). This means that out of a set of 1703 Global signals only 85 have RMSE values above 3.90 mK across the band $z = 5$–50 for example. We note that the values reported, averages across redshift ranges, in the REACH band are generally higher than across the whole redshift range because the REACH band excludes redshifts $\gtrsim$30 where the emulation is expected to be very precise.

Section C shows the explored parameter space for the Global 21-cm signal in the 21CMGEM test data set and the corresponding error when emulating the signals with GLOBALEMU.

Finally, in Table 1 we also report the $\widetilde{\mathrm{RMSE}}$ values in both the REACH band and across the whole redshift range. Cohen et al. (2020) report similar results for 21CMGEM and particularly we note that, when training and testing on the same data sets, we recover a mean $\widetilde{\mathrm{RMSE}}$ of 1.12 per cent compared to 1.59 per cent when using 21CMGEM. Similarly, we report a maximum value of 6.32% in comparison to the value of 10.55% reported by Cohen et al. for 21CMGEM. This further demonstrates that GLOBALEMU can achieve a high degree of accuracy in its emulation.

6.4 Neutral fraction

For the neutral fraction history network, we show similar results. Fig. 10 demonstrates the quality of the emulation with the mean, 95 percentile, and the worst results when emulating the neutral fraction and these values are detailed in Table 1. The results generally are of higher quality than those for the Global signal and, noting that the pre-processing for the two networks is near identical and the networks themselves are of the same size, this supports the understanding that the relationship between the inputs and outputs is simpler here. In the band $z = 5$–50, only 39 of 791 test models have $\widetilde{\mathrm{RMSE}} \geq$ 0.47 per cent.

7 CONCLUSIONS

GLOBALEMU uses a novel approach to emulate, with neural networks, the Global 21-cm signal and the evolution of the neutral fraction during the CD and EoR by considering redshift as an input to the neural networks alongside the astrophysical parameters. In tandem with this reparametrization of the problem, we use a predominantly physically motivated pre-processing for both the Global signal and neutral fraction. We subtract from the Global signals an AFB that obviates the need for the network learning a non-trivial but well-understood relationship at high redshift. We then resample both the Global signals and neutral fractions so that the regions that vary significantly across the training data sets can be better characterized by the networks.

The above framework allows the complex relationships between the astrophysical parameters and the Global signal or neutral fraction history as functions of redshift to be effectively learnt with small neural networks. Each Global signal of 451 redshift data points can be emulated in on average 1.3 ms. We note that this is a factor of approximately 102 improvement on the 133 ms we record with MATLAB when predicting the same signals on the same computer with 21CMGEM.

We demonstrate the effectiveness of GLOBALEMU by using the 21CMGEM training and testing data. This allows for a direct comparison between our results and the results of 21CMGEM. We find that GLOBALEMU can emulate to a higher degree of accuracy the Global 21-cm signal than 21CMGEM with a maximum normalized RMSE of 6.32 per cent in comparison to 10.55 per cent over the range $z = 5 - 50$. We also demonstrate that GLOBALEMU can emulate a Global 21-cm signal to, on average, less than 10 per cent the expected noise of a Global 21-cm experiment like REACH.

Finally, GLOBALEMU is a flexible PYTHON package that can be easily retrained on updated models with new astrophysical dependences. For example, additional astrophysical phenomena such as Lyman α heating (Reis et al. 2021) or additional radio background produced by galaxies (or an indeterminate synchrotron-like source; Fialkov & Barkana 2019; Reis et al. 2020) can be incorporated and easily trained upon. While the results achieved with the 21CMGEM data are impressive, the novelty of GLOBALEMU is in its flexibility, incorporation of redshift as an input, and physically motivated pre-processing. Particularly, the final two points allow for an accurate mapping from parameters to temperature with a single neural network reducing the points of failure and need for excessive fine-tuning.

ACKNOWLEDGEMENTS

HB acknowledges the support of the Science and Technology Facilities Council (STFC) through grant number ST/T505997/1. WH and AF were supported by Royal Society University Research Fellowships. EA was supported by the STFC through the Square Kilometer Array grant G100521.

DATA AVAILABILITY

The Global 21-cm signals used in this paper are publicly available at https://doi.org/10.5281/zenodo.4541500 and provided by Cohen et al. (2021).

REFERENCES

Anstey D., de Lera Acedo E., Handley W., 2020, MNRAS, 506, 2041
Barkana R., 2016, Phys. Rep., 645, 1
Bevins H. T. J., Handley W. J., Fialkov A., Acedo E. d. L., Greenhill L. J., Price D. C., 2021, MNRAS, 502, 4405
Bowman J. D., Rogers A. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, Nature, 555, 67
Bromm V., Kudritzki R. P., Loeb A., 2001, ApJ, 552, 464
Chatterjee A., Choudhury T. R., Mitra S., 2021, MNRAS, 507, 2405
Chuzhoy L., Shapiro P. R., 2007, ApJ, 655, 843
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
Cohen A., Fialkov A., Barkana R., Monsalve R. A., 2021, 21cmGEM Training and Testing Data Sets, available at https://zenodo.org/record/4541500
de Lera Acedo E., 2019, in 2019 Int. Conf. Electromagn. Adv. Appl. (ICEAA), REACH: Radio Experiment for the Analysis of Cosmic Hydrogen. IEEE, Piscataway, NJ, p. 0626
Fialkov A., Barkana R., 2014, MNRAS, 445, 213
Fialkov A., Barkana R., 2019, MNRAS, 486, 1763
Field G. B., 1959, ApJ, 129, 536
Furlanetto S. R., Oh S. P., Briggs F. H., 2006, Phys. Rep., 433, 181
Handley W. J., Hobson M. P., Lasenby A. N., 2015a, MNRAS, 450, L61
Handley W. J., Hobson M. P., Lasenby A. N., 2015b, MNRAS, 453, 4384
Hellum Bye C., Portillo S. K. N., Fialkov A., 2021, preprint (arXiv:2107.05581)
Hills R., Kulkarni G., Meerburg P. D., Puchwein E., 2018, Nature, 564, E32
Ioffe S., Szegedy C., 2015, Proceedings of Machine Learning Research, 448, 37
Javid K., Handley W., Hobson M., Lasenby A., 2020, preprint (arXiv:2004.12211)
Jennings W. D., Watkinson C. A., Abdalla F. B., McEwen J. D., 2018, MNRAS, 483, 2907
Kingma D. P., Ba J., 2015, in Bengio Y.,LeCun Y., eds, 3rd International Conference on Learning Representations, San Diego, CA, USA
List F., Lewis G. F., 2020, MNRAS, 493, 5913
Liu A., Shaw J. R., 2020, PASP, 132, 062001
McGreer I. D., Mesinger A., D'Odorico V., 2014, MNRAS, 447, 499
Madau P., Meiksin A., Rees M. J., 1997, ApJ, 475, 429
Mesinger A., 2019, The Cosmic 21-cm Revolution. IOP Publ., Bristol
Mesinger A., Furlanetto S., Cen R., 2011, Astrophysics Source Code Library, record ascl:1102.023
Mittal S., Kulkarni G., 2020, MNRAS, 503, 4264
Mondal R. et al., 2020, MNRAS, 498, 4178
Monsalve R. A., Fialkov A., Bowman J. D., Rogers A. E. E., Mozdzen T. J., Cohen A., Barkana R., Mahesh N., 2019, ApJ, 875, 67
Pearson K., 1901, London Edinburgh Dublin Phil. Mag. J. Sci., 2, 559
Philip L. et al., 2019, J. Astron. Instrum., 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
Reis I., Fialkov A., Barkana R., 2020, MNRAS, 499, 5993
Reis I., Fialkov A., Barkana R., 2021, MNRAS, 506, 5479
Ruder S., 2016, preprint (arXiv:1609.04747)
Schmit C. J., Pritchard J. R., 2018, MNRAS, 475, 1213
Sims P. H., Pober J. C., 2020, MNRAS, 492, 22
Singh S., Subrahmanyan R., 2019, ApJ, 880, 26
Singh S., Subrahmanyan R., Udaya Shankar N., Sathyanarayana Rao M., Girish B. S., Raghunathan A., Somashekar R., Srivani K. S., 2018, Exp. Astron., 45, 269
Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R., 2014, J. Mach. Learn. Res., 15, 1929
Venumadhav T., Dai L., Kaurov A., Zaldarriaga M., 2018, Phys. Rev. D, 98, 103513
Villanueva-Domingo P., Mena O., Miralda-Escudé J., 2020, Phys. Rev. D, 101, 083502
Visbal E., Barkana R., Fialkov A., Tseliakhovich D., Hirata C. M., 2012, Nature, 487, 70
Wouthuysen S. A., 1952, AJ, 57, 31

APPENDIX A: CALCULATING THE AFB

To approximate the AFB, we need to consider the physics defining the signal structure during the period dominated by collision coupling. During this period, neutral hydrogen atoms collide with other neutral hydrogen atoms, protons, and electrons. The spin temperature of hydrogen is coupled to the gas temperature, $T_{\rm K}$, via the collisions and that temperature cools adiabatically at a faster rate than the background radiation, $T_{\rm r}$. The spin temperature, $T_{\rm s}$, that encodes the number of hydrogen atoms in the two hyperfine levels of the ground state (Furlanetto et al. 2006) during this period is given by

$$\frac{1}{T_{\rm s}} = \frac{1/T_{\rm r} + x_{\rm c}/T_{\rm K}}{1 + x_{\rm c}}, \tag{A1}$$

where $x_{\rm c}$ is the collisional coupling coefficient. For our approximation of the AFB calculated here, we use a reference value for the gas temperature of $T_{\rm K,ref} = 33.7340$ K at $z_{\rm ref} = 40$ from the simulations used to produce the training and test data sets. We then scale $T_{\rm K,ref}$ adiabatically using

$$T_{\rm K} = T_{\rm K,ref}\frac{(1+z)^2}{(1+z_{\rm ref})^2} \tag{A2}$$

to get $T_{\rm K}$ as a function of redshift.

The coupling is dominated by H–H collisions and so we only consider these in our simulation. The coupling coefficient for this interaction is given by Furlanetto et al. (2006)

$$x_{\rm c}^{\rm HH} = \frac{n_{\rm H}\kappa_{10}^{\rm HH}T_*}{A_{10}T_{\rm r}}, \tag{A3}$$

where $\kappa_{10}^{\rm HH}$ is the rate coefficient for the spin deactivation of neutral hydrogen, T_* is the energy defect, and A_{10} is the spontaneous emission coefficient of the 21-cm transition. $n_{\rm H}$ is the relative number density

of neutral hydrogen given by

$$n_{\rm H} = 3.403\,68 \times 10^{68}\,\frac{\rho_{\rm c}}{m_{\rm p}}(1-Y)\Omega_{\rm b}(1+z)^3, \tag{A4}$$

where $Y = 0.274$ and is the Helium abundance by mass, $\rho_{\rm c}$ is the critical mass density of the universe in $M_{\rm sol}$/cMpc^3, $m_{\rm p}$ is the proton mass in $M_{\rm sol}$, and $\Omega_{\rm b}$ is the baryon density parameter.

From the above, we can then calculate δT as

$$\delta T = \frac{T_{\rm s} - T_{\rm r}}{1+z}(1 - \exp(-\tau_{\nu_0})), \tag{A5}$$

where τ_{ν_0} is the 21-cm optical depth of the diffuse IGM

$$\tau_{\nu_0} = \frac{3hc^3 A_{10} x_{\rm H_I} n_{\rm H}}{32\pi k_{\rm b} T_{\rm s} \nu_0^2 H(z)}, \tag{A6}$$

where ν_0 is the rest frequency of the 21-cm emission and $H(z)$ is the Hubble rate. In our calculation, we use the same cosmological parameters that were used to generate the signals (see Cohen et al. 2020). Here, the neutral fraction, $x_{\rm H_I}$, has a value of 1 since there is no astrophysics involved in the AFB.

APPENDIX B: TESTING THE NEURAL NETWORKS FOR OVERFITTING

We can demonstrate that, for both the Global signal and neutral fraction, the chosen network sizes of 3 hidden layers of 16 nodes do not overfit the training data by comparing the distribution of loss values across the training and test data sets. This is shown in Figs B1 and B2 for the Global signal and neutral fraction, respectively. Again,

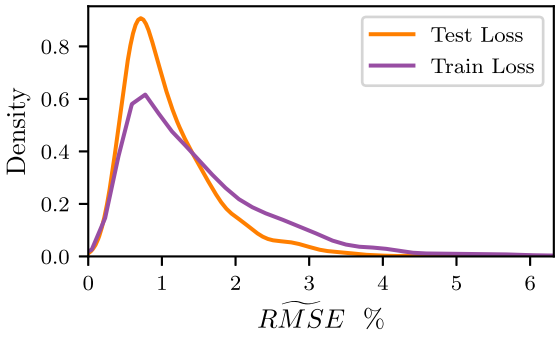

Figure B1. The probability density for the loss distribution found when emulating the training and test data sets for the Global 21-cm signal with GLOBALEMU.

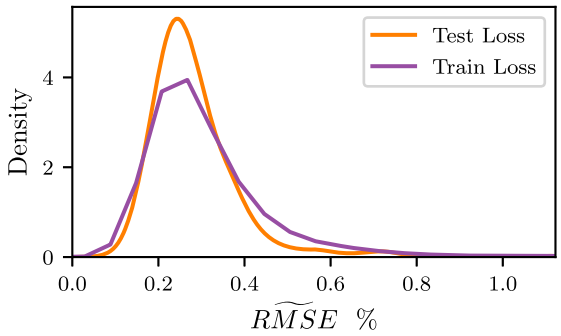

Figure B2. The probability density for the loss distribution found when emulating the training and test data sets for the neutral fraction with GLOBALEMU.

we have used a Gaussian kernel density estimation to calculate continuous probability density curves from the discrete histograms of losses. We can see in both cases that the losses, evaluated with equation (6), for the testing and training data sets, when emulated with the trained neural networks, have similar distributions and can consequently conclude that the neural networks are not overfitting the training data. In the event that the training data were being overfitted, the purple distribution, showing the training data losses, would peak to the left of the orange distribution, showing the test data losses, because the network would have learnt the training data to such a high degree of accuracy that it cannot generalize well to the testing data.

APPENDIX C: ERROR VERSUS PARAMETER

Fig. C1 shows the parameter space explored in the 21CMGEM test data set as a scatter plot. The data points are coloured based on the RMSE value calculated when comparing the corresponding true signal with the emulation, over the range $z = 5$–50, from GLOBALEMU.

Fig. C2 shows the equivalent graph with the colours determined using the dimensionless $\widetilde{\rm RMSE}$ metric across the band $z = 5$–50.

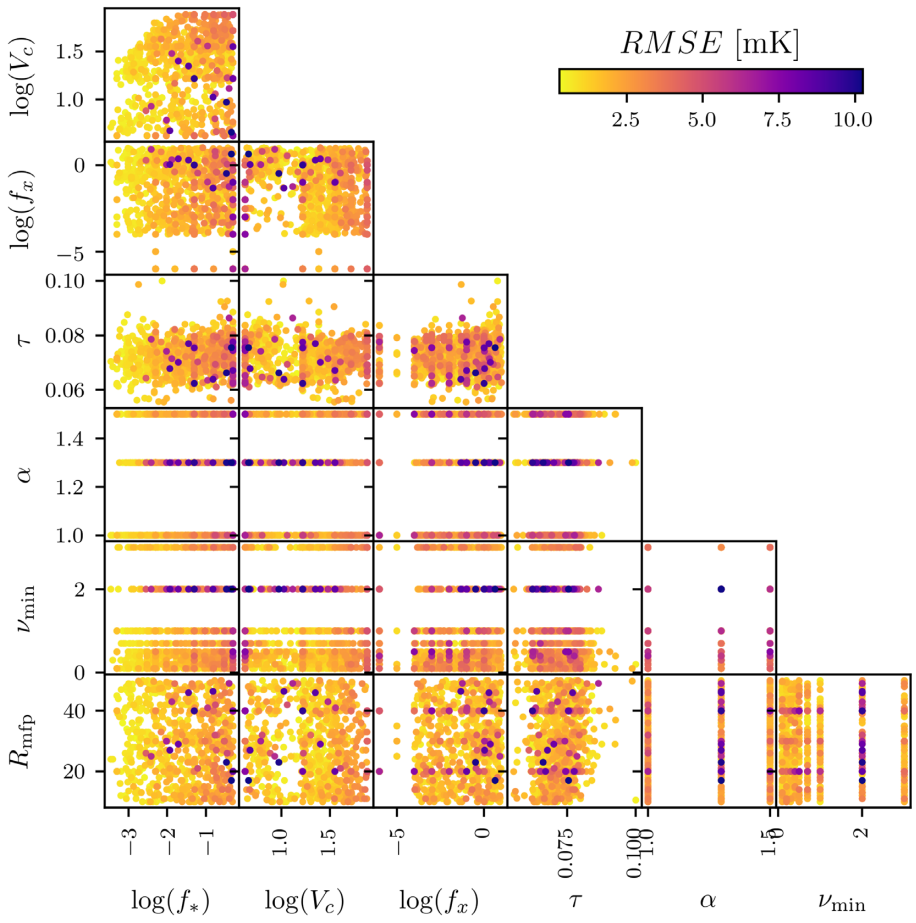

Figure C1. The parameter space explored by the 21CMGEM test data set. Each panel shows the 1703 models plotted as data points based on the corresponding astrophysical parameter values. They are coloured according to the RMSE calculated when comparing the true signals to the emulation from GLOBALEMU across the range $z = 5$–50.

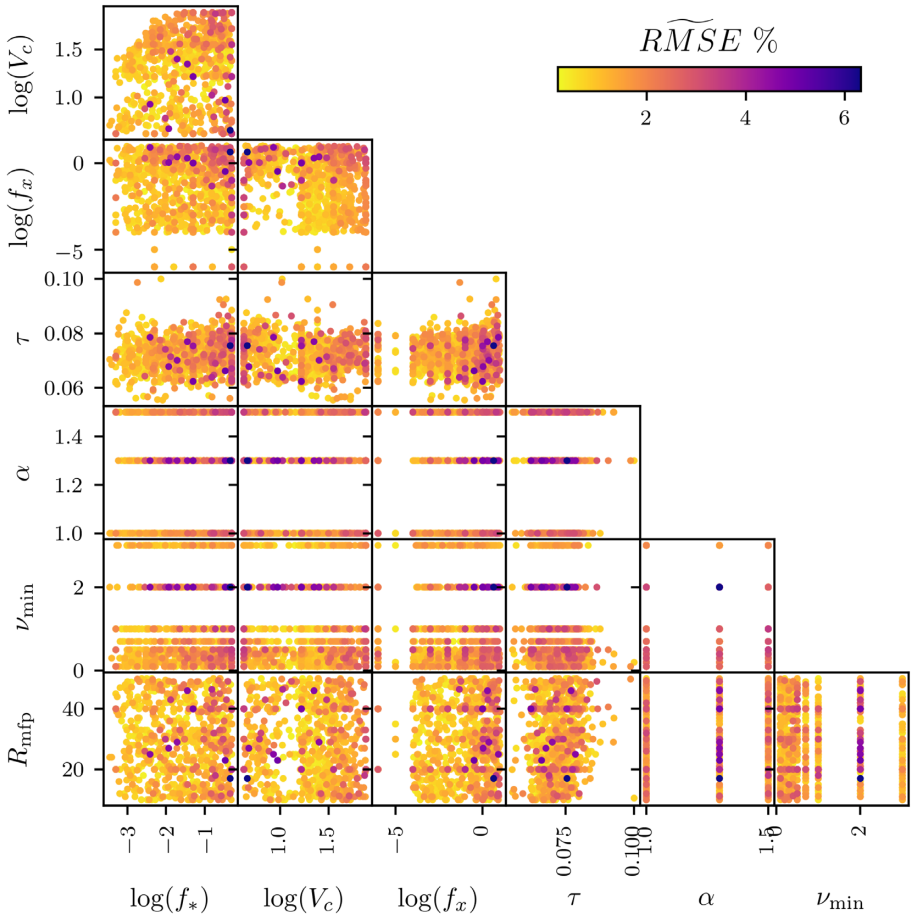

Figure C2. The equivalent of Fig. C1 with the data points coloured based on the dimensionless $\widetilde{\rm RMSE}$ error calculated across the band $z = 5$–50.

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