globalemu: A novel and robust approach for emulating the sky-averaged 21-cm signal from the cosmic dawn and epoch of reionisation

Harry Bevins (htjb2@cam.ac.uk), Will Handley, Anastasia Fialkov, Eloy de Lera Acedo
On behalf of the REACH[1] collaboration.
University of Cambridge, Cambridge, UK

Emulating the sky-averaged 21-cm signal

Global (sky-averaged) 21-cm signal emulators have been shown to be a useful tool for physical signal modelling and parameter estimation in global 21-cm experiments[2].

The existing state of the art emulator, `21cmGEM`[3], takes a set of astrophysical parameters and estimates a principle component decomposition of hybrid simulations of the global signal. It uses multiple regression and classifier neural networks. Compressed representations can lead to a loss of information and the large number of hyper-parameters that the emulator is dependent on can result in significant errors.

We present here **`globalemu` which uses the novel approach of taking redshift as an input along with the astrophysical parameters and returning a corresponding brightness temperature**. We demonstrate that with this approach and a physically motivated pre-processing we can emulate a high resolution signal in **8 ms, compared to 500 ms for `21cmGEM`**, to an accuracy of **≤ 6.32% compared to 10.55%** using a single small neural network and avoiding 'neural network magic'.

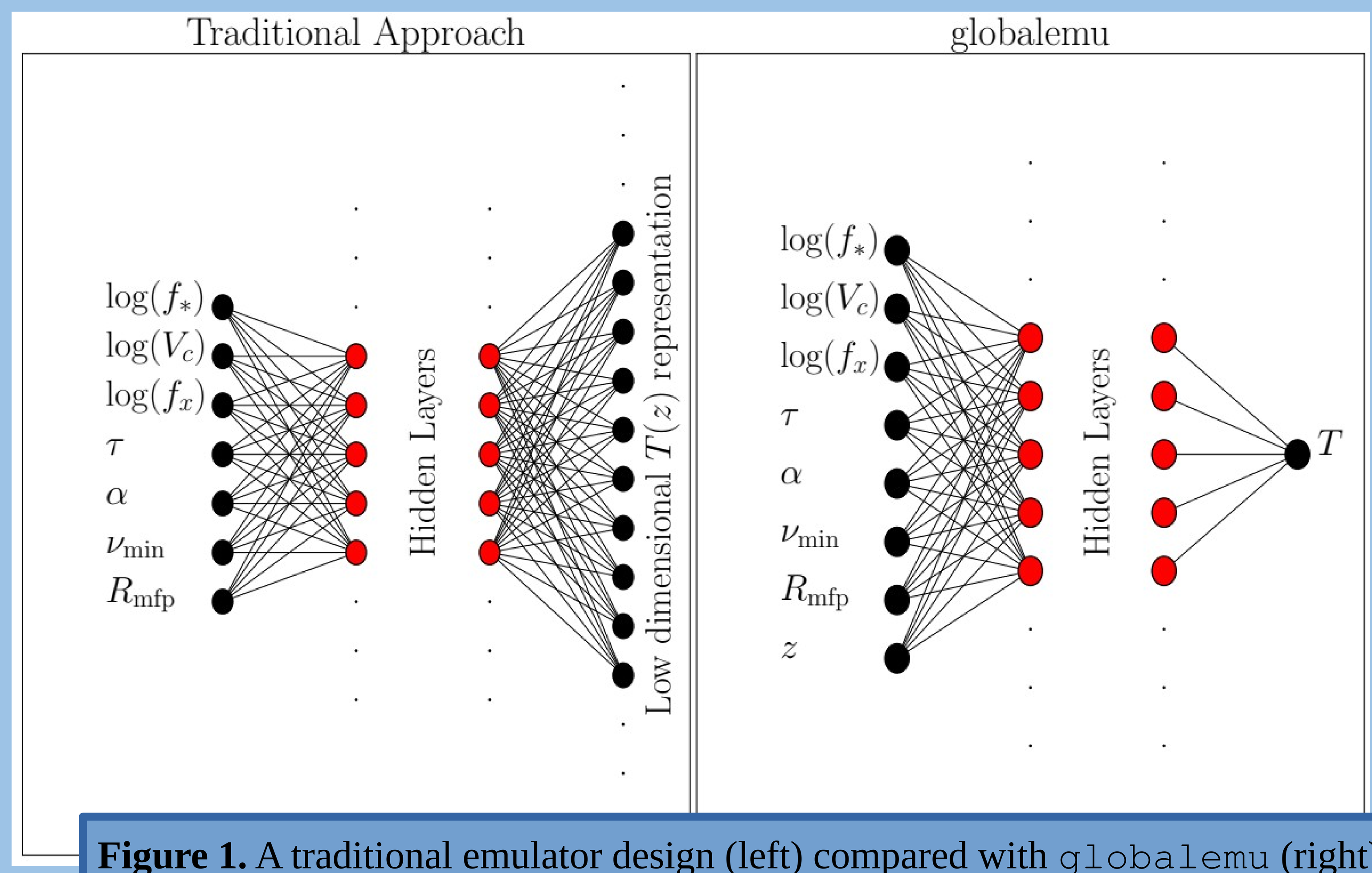

Figure 1. A traditional emulator design (left) compared with `globalemu` (right).

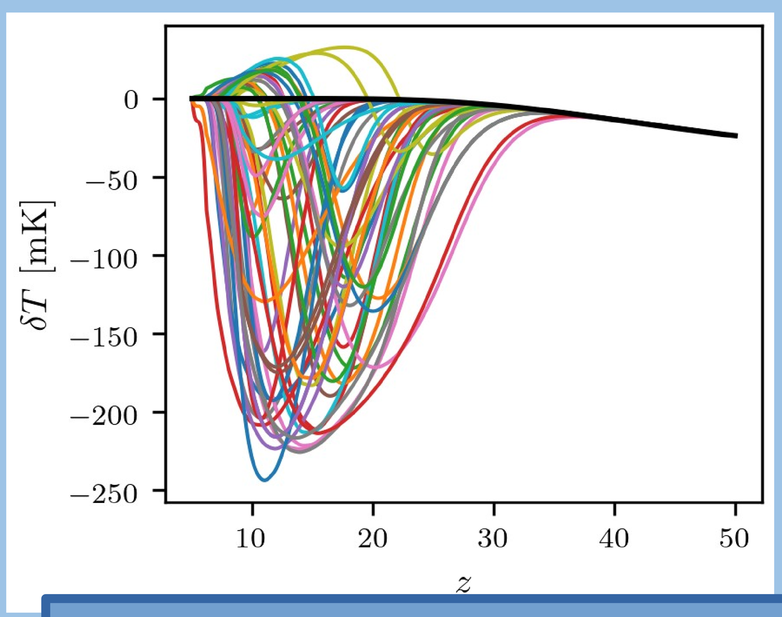

Figure 2. Example Global signals from the `21cmGEM` training data. Also shown in black is the astrophysics free baseline.

Data pre-processing

We assess the accuracy of `globalemu` using the same training and testing data sets used to assess `21cmGEM`. There are **≈30,000 signals** dependent on **seven astrophysical inputs**[2]: 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 and low energy cut off of the X-ray SED, α and $\nu_{\min}$ and the mean free path of ionizing photons, $R_{\rm mfp}$.

We use a physically motivated pre-processing of the training data in an attempt to simplify the complexity of the learnt relationship. We subtract off an **astrophysics free baseline** that dominates the structure at high z as it is a consistent feature in all models. We **resample** the signals with a higher rate where the signal variation is greatest in order for it to be properly emphasised. We then **divide the signals by the standard deviation** of the training set so that they are of order unity.

The input variables are normalised to be between 0 and 1 and we normalise the logarithm of f_*, V_c and f_x.

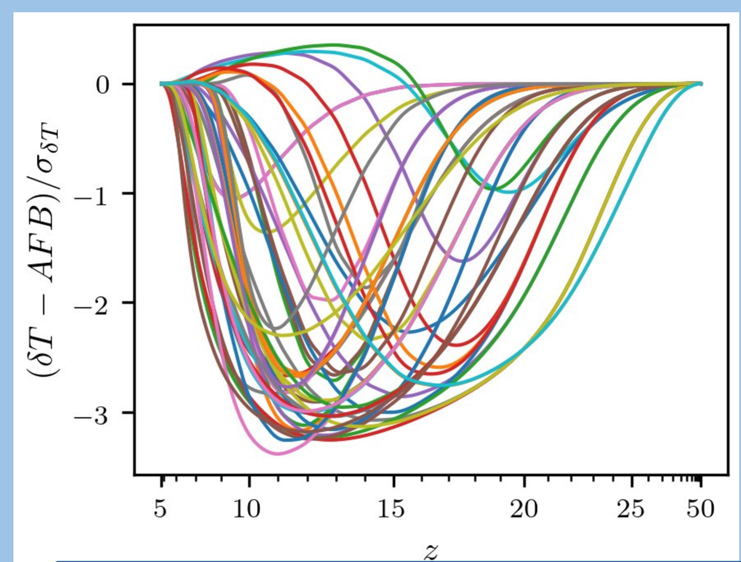

Figure 3. The pre-processed signals as seen by `globalemu`.

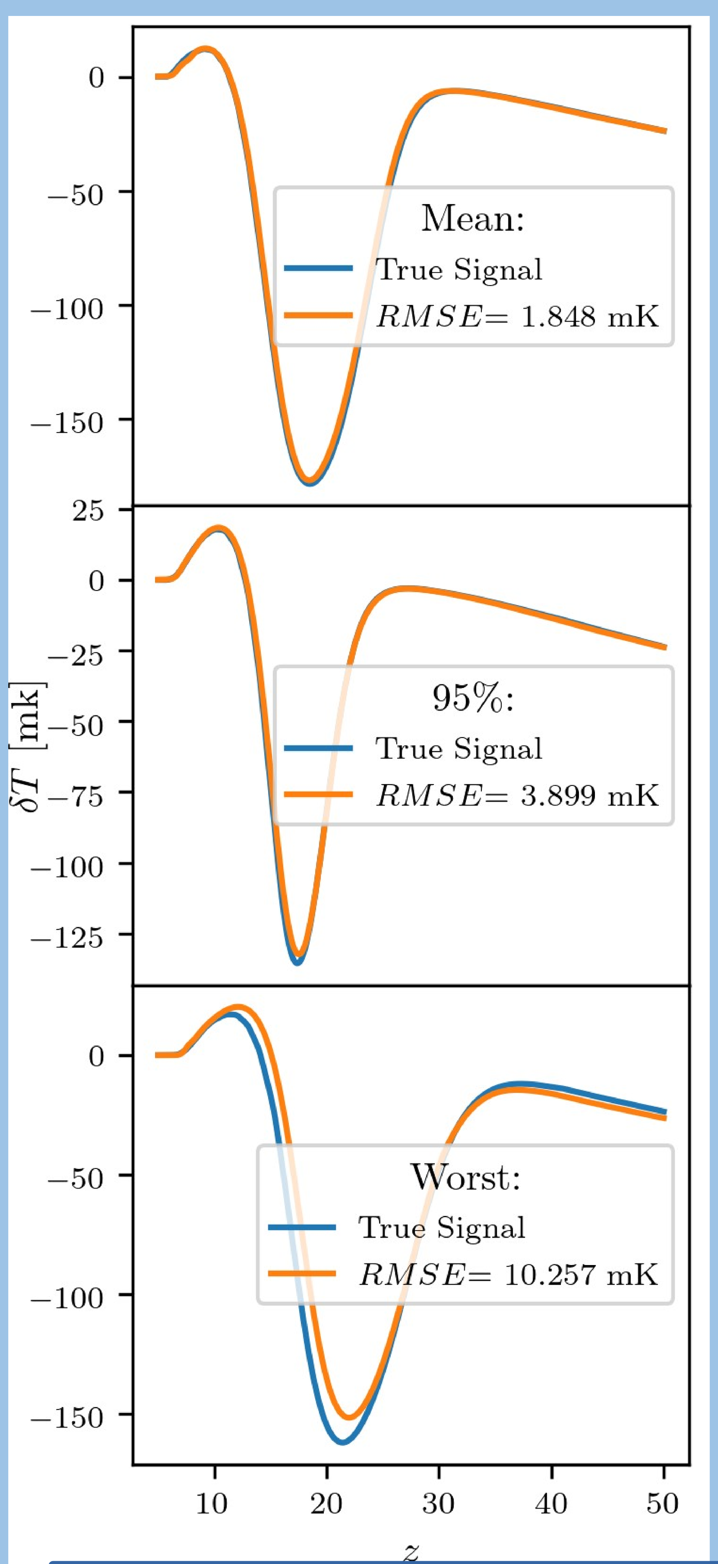

Figure 4. The mean, 95th percentile and worst emulated signals using `globalemu`.

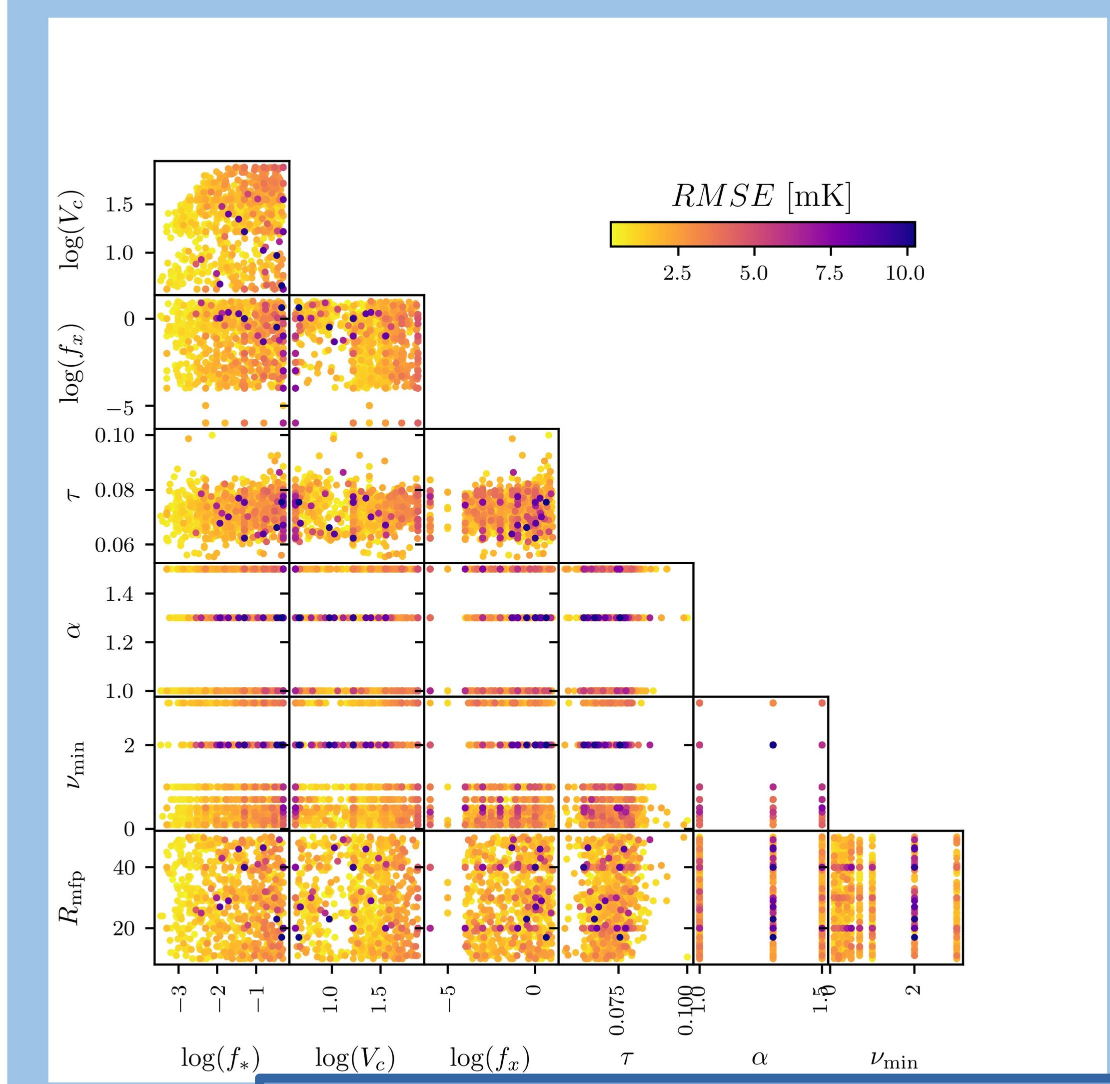

Figure 5. The emulator accuracy as a function of the parameters space.

Results

The results found when emulating, with `globalemu`, **1,703 test models** are shown below in the table. We use a fully connected network with 3 layers each of 16 nodes. We set a target RMSE of on average 2.5 mK, approximately 10% the expected noise of the Radio Experiment for the Analysis of Cosmic Hydrogen (REACH)[1]. We achieve this target in both the full band z = 5 - 50 and the REACH band z = 7 - 28.

We find a **mean normalised RMSE of 1.12% in comparison to 1.59% for `21cmGEM`** and note that the **95th percentile values are significantly lower than the maximum values**. **Only 85 models have an RMSE > 3.90 mK** in the band z = 5 - 50. The graph on the far left shows the emulated results in the band z = 5 - 50 and we also show the RMSE as a function of the astrophysical parameters.

		Global Signal	
		$z=5-50$	$z=7-28$
$RMSE$	Minimum	0.30 mK	0.31 mK
	Mean	1.85 mK	2.52 mK
	95^{th} percentile	3.90 mK	5.37 mK
	Maximum	10.26 mK	15.10 mK
$\widetilde{RMSE}$	Minimum	0.21%	0.26%
	Mean	1.12%	1.53%
	95^{th} percentile	2.41%	3.22%
	Maximum	6.32%	9.31%

Table 1. The results of emulating the `21cmGEM` test models with `globalemu`.

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

where

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

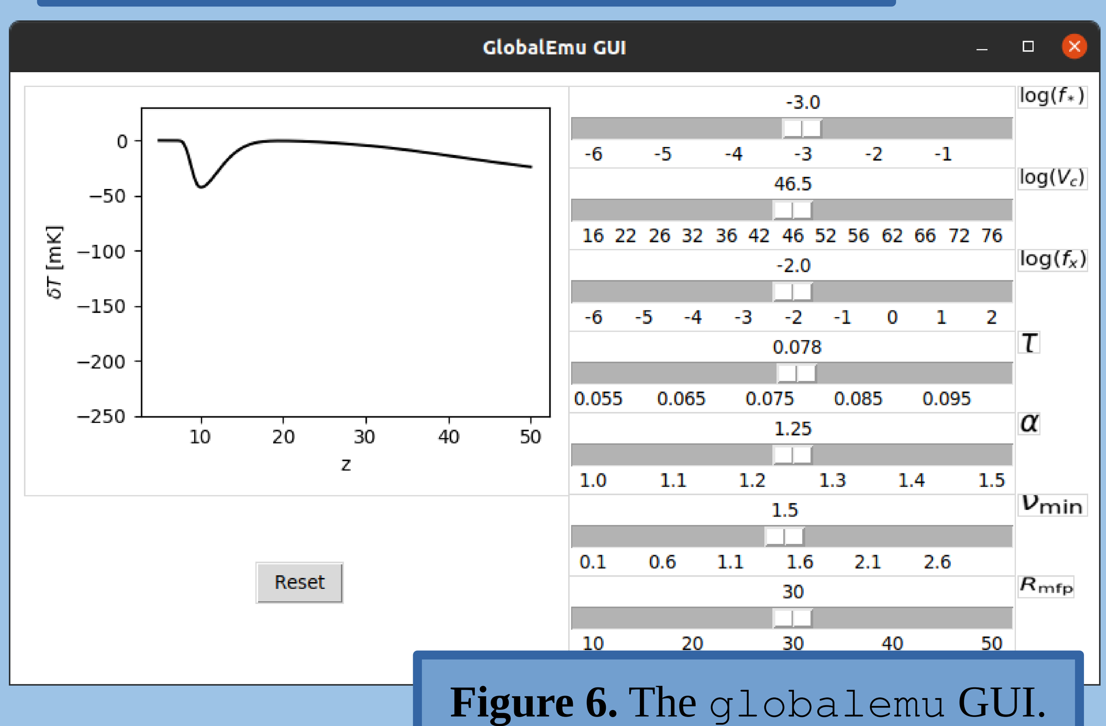

Figure 6. The `globalemu` GUI.

An adaptable emulator

`globalemu` is written in **`Python`** using **`tensorflow` and the `keras` backend**. It is **pip installable** and available for download at https://github.com/htjb/globalemu. It is a **flexible emulator** that can easily be **retained on any set of simulated 21-cm signals** whilst maintaining the novel approach, with redshift as an input, and physically motivated preprocessing.

We provide with the emulator an **easy to use GUI** (shown on the left) which is made possible by the speed of emulation. We also **release with version 1.0.0 the trained Global signal model demonstrated here**.

Two papers are in preparation to be submitted to **MNRAS** and the **Journal of Open Source Software** for a review of the code.

References
[1] - de Lera Acedo E., 2019, in 2019 International Conference on Elec-tromagnetics in Advanced Applications (ICEAA). pp 0626–0629,doi:10.1109/ICEAA.2019.8879199
[2] - 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
[3] - Cohen A., Fialkov A., Barkana R., Monsalve R. A., 2020, MNRAS, 495,4845