Extracting the 21cm Global Signal using Artificial Neural Networks

Madhurima Choudhury,[1]* Abhirup Datta,[1]† and Arnab Chakraborty,[1]

[1]*Discipline of Astronomy, Astrophysics and Space Engineering, Indian Institute of Technology Indore, 453552, India.*

arXiv:1911.02580v2 [astro-ph.CO] 23 Dec 2019

ABSTRACT

The study of the cosmic Dark Ages, Cosmic Dawn, and Epoch of Reionization (EoR) using the all-sky averaged redshifted HI 21cm signal, are some of the key science goals of most of the ongoing or upcoming experiments, for example, EDGES, SARAS, and the SKA. This signal can be detected by averaging over the entire sky, using a single radio telescope, in the form of a Global signal as a function of only redshifted HI 21cm frequencies. One of the major challenges faced while detecting this signal is the dominating, bright foreground. The success of such detection lies in the accuracy of the foreground removal. The presence of instrumental gain fluctuations, chromatic primary beam, radio frequency interference (RFI) and the Earth's ionosphere corrupts any observation of radio signals from the Earth. Here, we propose the use of Artificial Neural Networks (ANN) to extract the faint redshifted 21cm Global signal buried in a sea of bright Galactic foregrounds and contaminated by different instrumental models. The most striking advantage of using ANN is the fact that, when the corrupted signal is fed into a trained network, we can simultaneously extract the signal as well as foreground parameters very accurately. Our results show that ANN can detect the Global signal with $\gtrsim$ 92% accuracy even in cases of mock observations where the instrument has some residual time-varying gain across the spectrum.

Key words: cosmology:dark ages, reionization, first stars, cosmology:observations, methods: statistical, cosmology: theory

1 INTRODUCTION

The redshifted 21cm line of neutral hydrogen presents a unique probe of the evolution of the neutral intergalactic medium (IGM), from the Cosmic Dark Ages through Cosmic Dawn and Cosmic Reionization (Furlanetto et al. 2006; Morales & Wyithe 2010; Pritchard & Loeb 2012). The interplay between the CMB temperature, the kinetic temperature and the spin temperature, along with radiative transfer, lead to very interesting physics of the 21cm signal evolving over a redshift range. The 21cm line does not saturate, like the $Ly\alpha$, as at these high redshifts, (Fan et al. 2002, 2006; Mortlock et al. 2011), the IGM remains somewhat translucent at large neutral fractions (Barkana & Loeb 2005a) due to the Gunn-Peterson effect. The HI 21cm observations can be used to study evolution of cosmic structure from the linear regime at high redshift (i.e., density-only evolution), and through the non-linear regime associated with luminous source formation. The 21cm line as a probe, promises to be the richest of all cosmological datasets: HI measurements are sensitive to structures ranging from very large scales down to the source scale set by the cosmological Jeans mass(Barkana & Loeb 2005b).

The 21cm cosmological signal is hidden in a bright sea of foregrounds, which are $\sim 10^4$ times higher in magnitude. The foreground needs to be characterised and removed, in order to be able to detect this faint signal. In addition to the strong foregrounds, we have ionospheric distortion, RFI and also the frequency response of instrument which cause the sky as seen by the antenna to vary with time and other factors. All these factors, make extracting of the 21cm Global signal, extremely challenging. The most traditional procedure for extracting the faint cosmological signal is by assuming that the spectrally smooth foreground is well characterized and can be removed from the total signal. What is left as the residual would contain the signatures from the early phases of the formation of the Universe. There are theoretical models which explain the evolution of the Global signature with redshift, which essentially depends on the interplay of several crucial parameters. The detection of the 21cm Global Signal would in many ways complement the detection of the 21cm

* E-mail: madhurimachoudhury811@gmail.com
† E-mail: abhirup.datta@iiti.ac.in

power spectrum, and provide meaningful insight about the physical parameters for the evolution of the IGM. Several upcoming experiments plan to observe the sky-averaged signal: examples are, the Shaped Antenna measurement of the background RAdio Spectrum, SARAS, (Patra et al. 2013; Singh et al. 2017), the Large-Aperture Experiment to Detect the Dark Ages, LEDA (Greenhill & Bernardi 2012; Price et al. 2018), SCI-HI (Voytek et al. 2014), the Broadband Instrument for Global Hydrogen Reionisation Signal, BIGHORNS (Sokolowski et al. 2015). As there is a considerable amount of contamination of the signal by man-made sources (RFI) and ionosphere for all terrestrial observations, there have been some proposals to go to the far side of the moon, Dark Ages Radio Explorer (DARE) (Burns et al. 2012; Burns et al. 2017). There has been a recent report of detection of a flattened and deep absorption trough at 78MHz, of the sky-averaged radio spectrum by EDGES, Experiment to Detect the Global Epoch of Reionization Signature (Bowman et al. 2018). The reported absorption amplitude is 0.5K, which is exceedingly high, as compared to the expected amplitude of 150-200 mK. If confirmed, this will give us new insight to the existing physics of the evolution of our Universe. Following this detection, Fialkov et al. (2018) have explored a wider range of various possible 21cm signals, varying the properties of the dark matter particles, in addition to varying the astrophysical parameters. In Ewall-Wice et al. (2018), models which produce an excess radio background, explaining the large amplitude of the absorption feature in the EDGES detection, have been explored. The 21cm Global signal can be parametrized and the parameters can be associated with the physical processes as the Universe evolved. The already existing techniques of parameter estimation, (e.g. Harker et al. (2012)) relies, to a large extent, on the choice of the statistical prior distribution and does not consider realistic scenarios. However, in the later works of Harker (2015); Harker et al. (2016); Mirocha et al. (2015), they have used broad uninformative priors, and were able to recover the signal from synthetic datasets, assuming a perfect instrument and simple foregrounds. In the work by Tauscher et al. (2017); Tauscher et al. (2018) another technique has been introduced which uses Singular Value Decomposition (SVD) and produce systematics basis functions specifically suited to different observations of the Global signal. Bernardi et al. (2016) have used a completely Bayesian formalism, HIBayes, in which the signal is modelled as a Gaussian and the foreground is parametrized as a 7th order polynomial. All these, motivate us to search for another technique for parameter extraction from this contaminated spectrum. Implementation of machine learning algorithms, have shown some encouraging results recently, as demonstrated in Shimabukuro & Semelin (2017), Schmit & Pritchard (2018) for 21cm power spectrum analyses. In Hewitt et al. (2016) they have used Fisher formalism to derive predictions of the constraints that 21cm power spectrum measurements would place on the characteristics of the X-ray sources that heated the IGM at high redshifts.
In this paper, we have explored Artificial Neural Networks (ANNs), as an alternate technique for parameter extraction for the 21cm Global signal. We produce the cosmological signal using tanh parameterization and the Accelerated Reionization Era Simulations(ARES) (Mirocha et al. 2012). We then choose a suitable foreground model to represent the
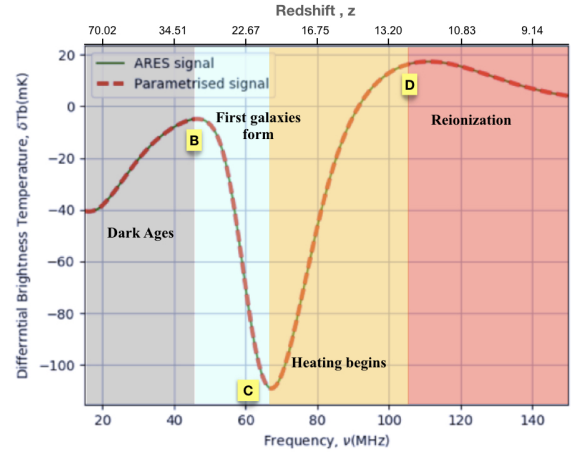
Figure 1. The signal constructed using the *tanh* model and the ARES signal. The points B,C,D mark the turning points of the turning point model. Note that the signal using the turning point model is not plotted here.

bright, dominant foregrounds and create a training dataset for the ANN. This plays a role similar to that of the priors, with the additional advantage of taking into consideration realistic datasets. This technique, might enable us to look for more complicated signals and search for a higher dimensional parameter space where parameter estimation becomes computationally costly. To the best of our knowledge, ANN is computationally cheaper and faster than MCMC for parameter extraction for large parameter spaces.

We begin with a brief review of the physical processes involved in reionization in §2. Previous reviews that discuss the HI 21cm signal from cosmic reionization in detail, include Barkana & Loeb (2001); Loeb & Barkana (2001); Furlanetto et al. (2006) and in the context of the Square Kilometer Array (SKA), Carilli et al. (2004); Koopmans et al. (2015), Furlanetto & Briggs (2004). In §3, we discuss how the 21cm Global signal can be simulated. We also give a short overview of the parameterization of the 21cm Global signal and the previous techniques of signal extraction. In §4 and §5, we describe the foregrounds concerned and the instrument dealt with in such experiments, respectively. We then explain in detail, the concept of Artificial Neural Networks, abbreviated as ANN, in §6 followed by a detailed discussion on the application of ANN in extracting the faint cosmological signal from very bright foregrounds in §7 and §8. In §9, we present the results obtained by our method. Lastly, we will discuss the advantages of using ANNs as an alternative to the traditional fitting techniques in detail in §10.

2   THE COSMOLOGICAL 21CM SIGNAL

The hyperfine transition line of atomic hydrogen (in the ground state) arises due to the interaction between the electron and proton spins. The quantity that we measure is the brightness temperature or more accurately called the differential brightness temperature, δT_b, measured relative to the

CMB (Pritchard et al. 2015),

$$\delta T_b \equiv T_b - T_\gamma \tag{1}$$

$$\delta T_b(\nu) = \frac{T_s - T_\gamma}{1+z}(1 - \exp^{-\tau_{\nu_0}})$$
$$\approx 27 x_{HI}(1 + \delta_b)\left(\frac{\Omega_b h^2}{0.023}\right)\left(\frac{0.15}{\Omega_{m,0} h^2}\frac{1+z}{10}\right)^{1/2} \tag{2}$$
$$\left(1 - \frac{T_\gamma(z)}{T_s}\right)\left[\frac{\partial_r v_r}{(1+z)H(z)}\right]^{-1}$$

where, x_{HI} denotes the neutral fraction of hydrogen, δ_b is the fractional over-density of baryons, Ω_b and Ω_M are the baryon and total matter density respectively, in units of the critical density, $H(z)$ is the Hubble parameter and $T_\gamma(z)$ is the CMB temperature at redshift z, T_s is the spin temperature of neutral hydrogen, and $\partial_r v_r$ is the velocity gradient along the line of sight. The 21cm Global signal is the sky averaged signal, whose characteristic shape contains information about global cosmic events. The differential brightness temperature can tell us about the ionizing radiation, which destroys neutral hydrogen, the X-rays, which can heat the gas and raise T_k, and $Ly-\alpha$, which causes Wouthuysen-Field coupling (Wouthuysen 1952; Field 1959). Fig. 1 shows the typical evolution of the signal with frequency. In our calculations, we neglect the peculiar velocity term and the density fluctuation term in the Global signal (Eqn. 2), as it averages out to a linear order and adds to a very small correction. So, the shape of the Global signal broadly depends on the density, neutral fraction and the spin temperature.

$$\delta T_b \approx 27(1 - x_i)\left(\frac{\Omega_{b,0} h^2}{0.023}\right)\left(\frac{0.15}{\Omega_{m,0} h^2}\frac{1+z}{10}\right)^{1/2}\left(1 - \frac{T_\gamma}{T_s}\right) \tag{3}$$

This equation is primarily used to construct the Global signal and would be the working equation for our work throughout the paper. The important epochs in the evolution of the signal are labelled in Fig. 1.

The most interesting quantity in the expression for the differential brightness temperature is the spin temperature, T_s, which primarily determines the intensity of the 21cm radiation [See Eqn. 3]. There are three competing processes that determine T_s. They are: (1) absorption of CMB photons (as well as stimulated emission); (2) collisions with other hydrogen atoms, free electrons, and protons; and (3) scattering of Lyman alpha photons through excitation and de-excitation. The spin temperature is defined as (Field 1959; Pritchard & Loeb 2012):

$$T_s^{-1} = \frac{T_\gamma^{-1} + x_k T_k^{-1} + x_\alpha T_\alpha^{-1}}{1 + x_k + x_\alpha}, \tag{4}$$

where, T_γ is the CMB temperature, T_k is the kinetic gas temperature, T_α is the temperature related to the existence of ambient Lyman-alpha ($Ly\alpha$) photons and x_c, x_α are respectively the collisional coupling and the $Ly\alpha$ coupling terms. In Barkana & Loeb (2005a), the evolution of the Global 21cm signal is discussed in great detail. We briefly explain the evolution of the Global 21cm signal in the following paragraph. At redshifts $z > 200$, free electrons couple T_γ and T_k through Thomson scattering and gas collisions, while the density is high enough to keep T_k and T_s in equilibrium. Here, $T_s = T_\gamma$

and there is no 21cm signal. This period is called the *Dark Ages*. At $z \approx 30 - 200$, the ionization fraction and density is too low to couple T_k to T_γ. Thus, the gas cools adiabatically, with the temperature falling as $(1 + z)^2$, which is faster than the rate at which the CMB temperature falls, i.e, as $(1 + z)$. Thus the gas becomes colder than the CMB but the mean density is still high enough to provide coupling between T_s and T_k through collisions. Around this redshift range, the 21cm signal might be seen in absorption against the CMB. At $z \approx 20 - 30$, collisions can no longer couple T_k to T_s, and thus T_s begins to approach CMB. We might expect the first luminous structures near the end of this redshift range. This period is called the *Cosmic Dawn* : the birth of the first luminous sources in the Universe. The $Ly\alpha$ photons from these objects would induce local coupling of T_k and T_s, which might lead to some absorption regions of the 21cm signal. These photons, as well as X-rays from the first luminous sources, could lead to warming of the IGM above the CMB temperature, well before reionization. Hence, one might expect some patches of regions with no signal, absorption and maybe emission,in the 21cm line. At $z \approx 6 - 20$ more physical processes come into play. The scattering of the $Ly\alpha$ photons, X-rays from the first galaxies and black holes and weak shocks associated with structure formation warms up the IGM, so that T_k is larger than T_γ (Furlanetto et al. 2004). These objects are reionizing the Universe, from the evolution of large scale structure, to a bubble dominated era of HII regions. In this regime, the expected 21cm signal is rich. After reionization, i.e., $z \approx 6$, IGM is completely ionized and the 21cm signal is gone again.

There are several experiments either proposed or already operational which are designed to detect this Global 21cm signal either in some specific redshift ranges or over the entire range of redshifts ranging from the cosmic *Dark Ages* through *Cosmic Dawn* to *Epoch of Reionization*. In this paper, we consider a large range of redshifts covering most of these three important epochs.

3 SIMULATING THE GLOBAL SIGNAL

Based on several theoretical models the cosmological Global 21cm signal changes its shape and amplitude as a function of redshifted frequencies. Fialkov et al. (2014, 2015); Cohen et al. (2017) have used a semi-numerical approach to model possible 21cm Global signals in the redshift range, $z \sim 6 - 40$, and is flexible to explore the large dynamical range of astrophysical parameters. Parameterization of the cosmological signal has been done using physical parameters which are directly related to the IGM properties, from which we can infer the physics of the earliest sources. Pritchard & Loeb (2010) proposed a turning point model, where the parameters of the model are the positions of the turning points (marked by points B, C, D in Fig. 1) - these are positions in redshift or frequency, and in brightness temperature. The signal between these turning points is modelled as a cubic spline. This parameterization is very flexible and can describe a wide range of 21cm signals, but the turning point positions require further interpretation in order to relate them to the physics of the first sources (Mirocha et al. 2013). The tanh parameterization (Harker 2015; Mirocha et al. 2015), uses a set of tanh functions to model the Global signal. In Bernardi
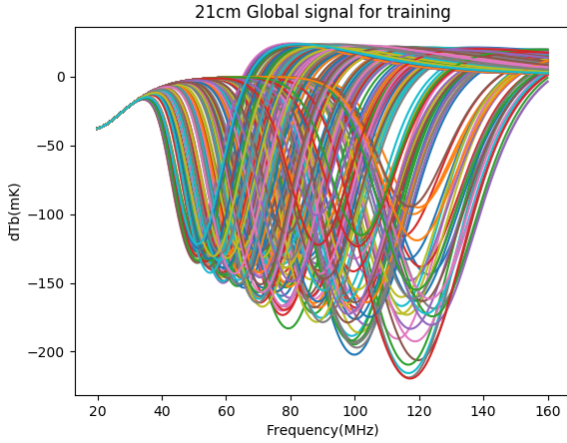
Figure 2. The set of 21cm Global signals generated to construct the training datasets, by varying the signal parameters.

et al. (2015, 2016), the absorption feature is modelled as a Gaussian.

In this paper, we have used tanh parameterization to simulate the Global 21cm signal as proposed in Mirocha et al. (2015). This method employs modelling the $Ly\alpha$ background, IGM temperature and ionization fraction as simple tanh functions (Harker et al. 2016). The $Ly\alpha$ background determines the strength of Wouthuysen-Field coupling, T the temperature of the IGM, and the $\overline{X}$ is the ionized fraction of hydrogen. We let each quantity evolve as a tanh function (Mirocha et al. 2015) given by:

$$A(z) = \frac{A_{\text{ref}}}{2}\{1 + tanh[(z_0 - z)/\Delta z]\}, \qquad (5)$$

where $A(z)$ represents the parameters $J_\alpha(z)$, $T(z)$ and $\overline{X}_i(z)$. The free parameters of the tanh model are the step height, A_{ref}, pivot redshift, z_0 and a width or duration, Δz. $J_\alpha(z)$ represents the $Ly\alpha$ background which determines the strength of the Wouthuysen-Field coupling. $T(z)$ represents the temperature of the IGM and $\overline{X}_i$ is the ionization fraction. These parameters are directly linked with IGM properties but not to the source properties. So, in some sense this is an intermediate between the physical modeling and phenomenological models, like the spline or Gaussian models.

In this paper, we have used three tanh functions representing the parameters $J_\alpha(z)$, $T(z)$ and $\overline{X}_i(z)$ which in turn are functions of three parameters each as listed below:

$$J(z) = \frac{J_{\text{ref}}}{2}\{1 + tanh[(J_{z0} - z)/J_{dz}]\}$$

$$\overline{X}_i(z) = \frac{X_{\text{ref}}}{2}\{1 + tanh[(X_{z0} - z)/X_{dz}]\} \qquad (6)$$

$$T(z) = \frac{T_{\text{ref}}}{2}\{1 + tanh[(T_{z0} - z)/T_{dz}]\}$$

The normalization A_{ref}, (Eqn. 5) of the tanh function corresponding to the $Ly\alpha$ flux is J_{ref}, in the units of 10^{-21} erg s^{-1} cm^{-2} Hz^{-1} sr^{-1}. The redshift interval (Δz) and the central redshift (z_0) over which the $Ly\alpha$ background turns on, are represented by J_{dz} and J_{z0} respectively. For X-ray heating, $T(z)$ is in units of K, Δz and z_0 are denoted by T_{dz} and T_{z0}. The step height corresponding to $T(z)$, i.e.,
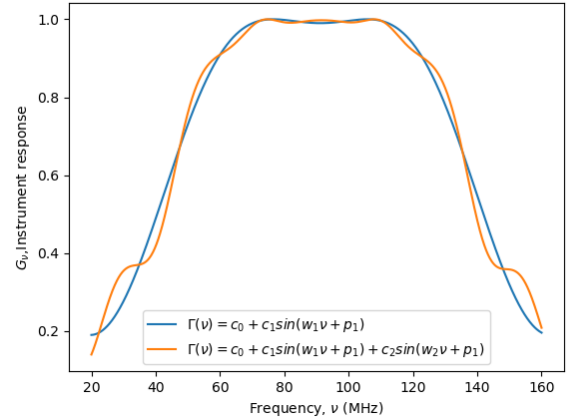
Figure 3. Here we have plotted the instrument response represented by $G(\nu) = 1 - |\Gamma|^2$, where $\Gamma(\nu)$ is the antenna reflection coefficient for a simple instrument and a moderate instrument. The parameters were chosen so that $G(\nu)$ would resemble a band-shape.

T_{ref} is fixed at 1000 K. As the signal saturates at low redshifts, so the precise height of the step is not important. The step height corresponding to the ionization fraction, $\overline{X}_i$ is fixed to unity, since it represents a fraction. So, we have 7 signal parameters to construct the signal using the tanh parameterization. We compute the coupling coefficients using $ARES$, and plug in the values of the parameters into Eqn. 2 to obtain the simulated Global 21cm signal. Hence, our parametrized Global 21cm signal depends on total 7 parameters (with 2 parameters, $X_{\text{ref}} = 1.0, T_{ref} = 1e3$, fixed). We take the inferred values of these parameters from Harker (2015): $J_{\text{ref}} = 11.69, Jz_0 = 18.54, Xz_0 = 8.68, T_{z_0} = 9.77, J_{dz} = 3.31, T_{dz} = 2.82, X_{dz} = 2.83$. Each of these parameters are varied by ±50% to generate our training sets. The parameter range explored here is sufficient to represent a wide range of shapes of the Global signal. A typical set of such signals generated for this work is shown in Fig: 2. The reason why we choose the tanh model is because, it can mimic the shape of the Global 21cm signal very well, and can be related to the physical properties of the IGM to a great extent. The tanh parameters are directly related to the IGM properties, but cannot give us information about the source properties directly. Thus, it can be considered to be somewhere between the 'turning point' model, which is completely phenomenological and other entirely physical models.

4 FOREGROUNDS

The faint Global 21cm signal from Cosmic Dawn/Epoch of Reionization has to be detected in presence of bright foreground sources in the sky and instrumental/atmospheric corruptions. Radio emission from our Galaxy and other extragalactic sources are many orders of magnitude brighter than the cosmological signal of interest. The expected Global 21cm signal is about 10^{-4} times weaker than the foreground emission. The bright foregrounds, Radio Frequency Interference (RFI), instrumental calibration errors poise signifi-

cant challenges to Global 21cm experiments. Hence, several sophisticated simulations are necessary to understand the effect of these corruption terms on the possible signal extraction method. That makes it critical to have an accurate model for the foregrounds at these radio frequencies.

Following Pritchard & Loeb (2010); Bernardi et al. (2015), the foreground spectrum can be modelled as a polynomial in $\log(\nu)$-$\log(T)$. While Harker (2015) showed that use of 3rd or 4th order polynomial is sufficient to represent the sky spectrum, Bernardi et al. (2015) showed that a 7th order polynomial is necessary when incorporating the chromatic primary beam of the antenna. In our case, we restrict our foreground model to 3rd order polynomial in $\ln(T)$-$\ln(\nu)$ (Harker 2015), representing diffuse foregrounds:

$$\ln T_{FG} = \sum_{i=0}^{n} a_i [\ln(\nu/\nu_0)]^i, \tag{7}$$

where, ν_0 is taken to be 80MHz, which is an arbitrary reference frequency and $T_0, a_1, a_2, ... a_n$ are the parameters of the model, where $a_0 = \ln T_0$, as the zeroeth order coefficient has units of temperature. Increasing n implies more complex, less smooth foregrounds. Here we haven taken the values of $\{T_0, a1, a2, a3\} = \{2039.611, -2.42096, -0.08062, 0.02898\}$ (Harker 2015; de Oliveira-Costa et al. 2008). For the simulations, we varied them as $\{15\%, 10\%, 1\%, 1\%\}$ respectively. As the foregrounds are spectrally smooth as compared to the cosmological signal, this feature is taken advantage of in signal extraction.

5 THE INSTRUMENT

In this section, we introduce the model for the instrument or the radio telescope using which the detection of the cosmological signal can be attempted. Although it appears to be a very simple case to observe with a total power radiometer, it is actually quite complex to expect the stability of the instrument over such a large bandwidth of frequencies and over large hours of observing time (Burns et al. 2012; Burns et al. 2017). The calibration of such an instrument becomes quite a challenge as any residual calibration error might mimic the cosmological signal. Hence, understanding of the instrument stability and robustness of our proposed signal extraction technique with instrumental calibration error is critical.

In Harker et al. (2012); Harker (2015) or other related works, the analyses mostly consider a perfect instrument, with no residual calibration error or impact on the foreground removal. Frequency and angular dependence of real antenna response complicates the foreground modelling, commonly based on the assumption of spectral smoothness. The coupling between the foregrounds and antenna gain pattern generates spectral structure in the foregrounds that is required to be represented by additional higher order terms in the foreground.

Here, we have followed a model of the instrument as described for DARE (Dark Ages Radio explorer) (Bradley 2012; Burns et al. 2017), where a set of sinusoids is fitted to

the modeled antenna reflection coefficient, $\Gamma(\nu)$;

$$\Gamma(\nu) = c_0 + c_1 \ \sin(\omega_1\nu + p_1) + c_2 \ \sin(\omega_2\nu + p_2)$$
$$+c_3 \ \sin(\omega_3\nu + p_3) + c_4 \ \sin(\omega_4\nu + p_4) \tag{8}$$

And the instrument response or the gain function is represented by,

$$G(\nu) = 1 - |\Gamma(\nu)|^2 \tag{9}$$

We have considered two cases for the instrument: (a) simple and (b) moderate. We call the (a) *simple* instrument case, where Γ is represented by a single sinusoid only. We chose the parameters of Γ such that, the response mimics a bandpass shape. We have chosen, $c_0 = 0.4; c_1 = 0.5; w_1 = 0.044; p_1 = 0.7$ [See Fig. 3].

$$\Gamma(\nu) = c_0 \ + \ c_1 \ \sin(\omega_1\nu + p_1) \tag{10}$$

We call the other one as (b) *moderate* instrument, where we have assumed that the instrument response can be represented as a sum of two sinusoids-one with a large amplitude and a low frequency and another with a very small amplitude and a high frequency .

$$\Gamma(\nu) = c_0 \ + \ c_1 \ \sin(\omega_1\nu + p_1) + c_2 \ \sin(\omega_2\nu) \tag{11}$$

The chosen value of the parameters for this case were: $c_0 = 0.4, c_1 = 0.5, \omega_1 = 0.044, p_1 = 0.7, c_2 = -0.05, \omega_2 = 0.25$ [See Fig. 3]. We have varied each of the instrument parameters by $\sim 5\%$ for all our simulations in this work.

6 ARTIFICIAL NEURAL NETWORKS

6.1 Introduction: Basic architecture

Artificial Neural Networks are information processing systems, whose performance characteristics are inspired by the structure and functioning of the human brain and the nervous system. ANNs simulate the learning process from examples mimicking the way humans tend to learn. There can be different kinds of supervised and unsupervised learning algorithms, which can be implemented to suitably teach and train a network as per requirement.

The basic building blocks of the ANNs are the *neurons*. The most simple ANN usually is a three-layer network comprising of the input layer, a hidden layer and the output layer [see Fig. 4]. The number of neurons in the hidden layers decide how *wide* the network is and the number of hidden layers decide how *deep* the network is. The final structure of the network is decided after taking into consideration several factors, which will be dealt in detail in the following sections. As briefly outlined in Choudhury & Datta (2017), in this section, we introduce in detail how a neural network works. ANNs construct functions, which associate the input with the output data. The basic neural network model can be described as a series of functional transformations. The *input layer*, is the first layer of the network. We consider that there are D inputs, $(x_1, x_2, ..., x_D)$. Each neuron in the input layer (which will be denoted by an index i) is connected to each neuron in the next layer, which is the hidden layer (denoted by an index j). A weight w_{ji} and a bias w_{j0} are associated with each connection. The input to a single neuron in the hidden layer is a linear combination of the
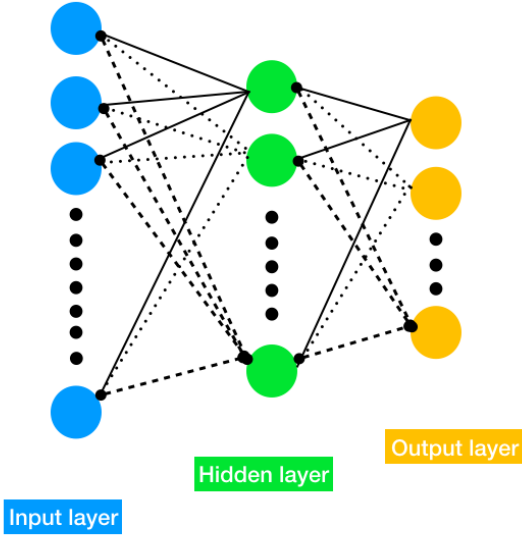
Figure 4. Schematic representation of a typical neural network. Each coloured circle represents a neuron, which form the basic units of an ANN. The lines represent the connections from neurons of one layer to the next. The blue, green and the orange layers represent the input, hidden and the output layers of the network. It should be noted that there can be more than one hidden layer in an ANN.

input neurons with their respective weights and biases, and is written as:

$$a_j^{\mathrm{H}} = \sum_{i=1}^{D} w_{ji}^{\mathrm{I}\rightarrow\mathrm{H}} x_i + w_{j0}^{\mathrm{I}\rightarrow\mathrm{H}} \tag{12}$$

where, w_{ji} 's are the weights and w_{j0} 's are the biases associated with each connection between the neurons of the input and hidden layer respectively, denoted by the superscript I $\rightarrow$ H. These a_j's, are known as *activation* [Fig. 5] and the superscript H represents a hidden layer neuron. In the hidden layer, each of the activations is transformed using a non-linear activation function h, such that,

$$z_j = h(a_j^{\mathrm{H}}) \tag{13}$$

These quantities z_j's correspond to the outputs of each neuron in the hidden layer. These activation functions are very interesting and can be chosen to be non-linear functions, such as the logistic sigmoid function or tanh function. These functions are smooth and differentiable, and it saturates and returns a constant output when the absolute value of the input is sufficiently large. It is because of the activation function that a non-linearity is introduced in the network and the trained ANN performs appropriately. Otherwise, the entire algorithm would just be solving some linear set of functional transformations. These values (z_j's) are again linearly combined along with the corresponding weights and biases associated with the respective connections to give the inputs to each neuron in the output layer, denoted by the index k. These are also called the output

unit activations, a_k^{O}, given by:

$$a_k^{\mathrm{O}} = \sum_{j=1}^{M} w_{kj}^{\mathrm{H}\rightarrow\mathrm{O}} z_j + w_{k0}^{\mathrm{H}\rightarrow\mathrm{O}} \tag{14}$$

here, w_{kj}'s and the w_{k0}'s are the weights and biases respectively, connecting the hidden and the output layer, denoted by the superscript H $\rightarrow$ O. The output unit activations are then transformed using an appropriate activation function, to give a a set of network outputs y_k. [See Fig. 5]

$$y_k = h(a_k^{\mathrm{O}}) \tag{15}$$

These y_k's are the end outputs of the *feed-forward process* and h is the activation function. It must be noted that the activation function is usually not applied to the output layer, or in most cases it is a linear function. So far, we have given a input vector X, of dimension D and obtained an output vector Y, of our desired dimension, chosen as per requirement.

6.2   The Training Process

ANNs are viewed as a general class of parametric non-linear functions from an input variables vector X to a vector Y of output variables. The parameters in question are the respective weights and biases. These parameters are determined usually by minimizing a sum-of-squares error function. We denote this error function by $E(w)$. Error functions for a set of independent observations comprise of sum of terms, one for each data point,

$$E(w) = \sum_{n=1}^{N} E_n(w) \tag{16}$$

This is the total *error function* or the total *cost function*. The goal is to find a suitable weight vector W which minimizes the chosen function $E(w)$. A very small step in the weight space from w to $w+\delta w$, changes the error function by a quantity, $\delta E \simeq \delta w^T \nabla E(w)$, where the vector $\nabla E(w)$ points in the direction of the greatest rate of increase in the error function. As the error $E(w)$ is a smooth continuous function of w, its smallest value will occur at a point in the weight space such that the gradient of the error function vanishes, so that $\nabla E(w) = 0$. Else, we could make a small step in the direction of $-\nabla E(w)$ and further reduce the error. The error function typically has a nonlinear dependence on the weights and bias parameters, so there will be many points where the gradient vanishes or becomes very small. As we cannot directly find the solution to the equation $\nabla E(w) = 0$, numerical procedures are used to find the solution. Some initial value, $w^{(0)}$ is chosen for the weight vector and then we move through the weight space in succession of small steps till error function is minimized. This optimizing is done using a stochastic gradient decent, or *adam* (Kingma & Ba 2014). Adam deals with outliers very well and has been proven to be a better optimizer than the standard stochastic gradient descent, which is discussed in detail in Kingma & Ba (2014). So, the outputs obtained are compared with the inputs with which the network was fed and an error function is computed. We find out an optimum set of weights and biases, that ensures that the output produced by the network is sufficiently close to the desired output values.
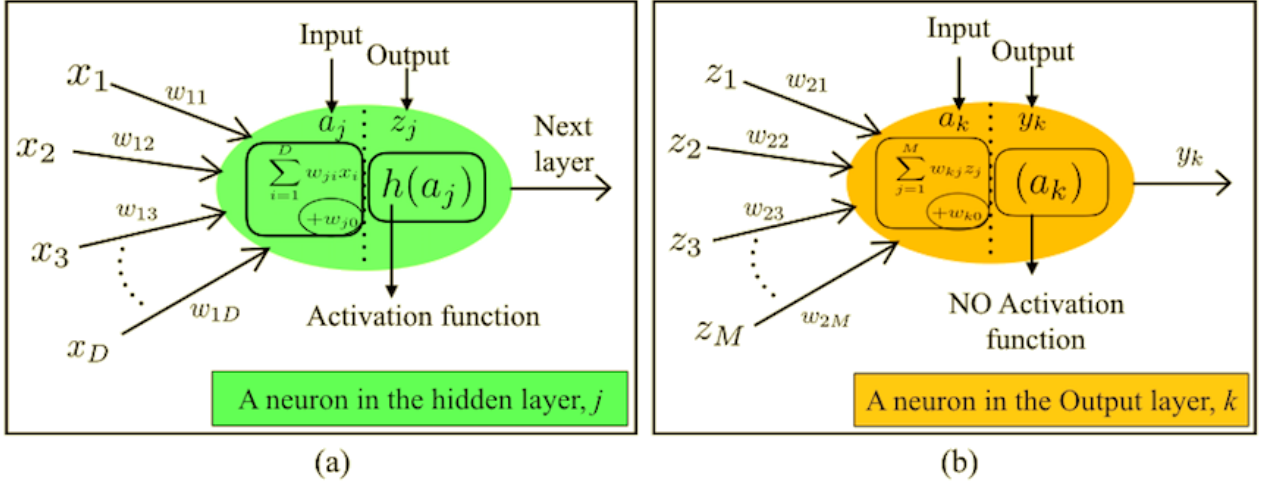
Figure 5. Typical neuron structures explaining what exactly happens inside the neurons of the hidden layer (left) and the output layer (right). Each neuron receives the input from all the neurons in the previous layer, along with their respective weights and biases.

6.3 Back-propagation

The error function E is a function of the weights. The error computed at the end of the output layer, at each node (output of the neurons) is the difference between the target and the actual value obtained. The node error can be defined as:

$$e_k = t_k - o_k \tag{17}$$

where, t_k is the target and o_k is the output obtained at node k, in the output layer.

The *back-propagation algorithm*, is referred to as *back-prop* or BPP. The errors are 'back-propagated' to adjust the weights such that the error function is minimized. The total error is proportionately propagated backwards. The higher weights are considered to be more responsible for the errors, thus are proportionately split to the connected links [See Fig. 6].

The errors associated with the internal links are not so straight-forward. The output layer errors are split in proportion to the connection weights and recombine the bits of error at each node in the hidden layer. The slope of the error function is calculated, so that the weights can be accordingly adjusted.

$$\frac{\partial E}{\partial w_{kj}^{\text{H}\rightarrow\text{O}}} = -(e_k)h\left(\sum_j w_{kj}^{\text{H}\rightarrow\text{O}}.o_j\right)\cdot\left[1 - h\left(\sum_j w_{kj}^{\text{H}\rightarrow\text{O}}.o_j\right)\right].o_j \tag{18}$$

Here, h is the activation function and the superscripts represents the connections to which the weights are associated. Similarly, the weights w_{ji} are adjusted as,

$$\frac{\partial E}{\partial w_{ji}^{\text{I}\rightarrow\text{H}}} = -(e_j)h\left(\sum_i w_{ji}^{\text{I}\rightarrow\text{H}}.o_i\right)\cdot\left[1 - h\left(\sum_i w_{ji}^{\text{I}\rightarrow\text{H}}.o_i\right)\right].o_i \tag{19}$$

Here, e_j is the recombined back-propagated error from the hidden nodes [See Fig. 6].

The weights are updated using the above relations. It is to be noted that the weights are changed in a direction opposite to the gradient. A quantity α is introduced which
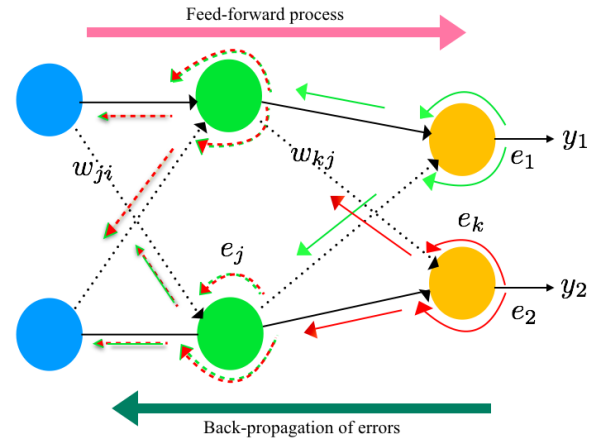
Figure 6. The back propagation algorithm schematically explained. The coloured arrows show how the error is propagated backwards, while the solid black arrows demonstrate the feed forward process.

is called the learning rate.

$$\text{new } w_{kj}^{\text{H}\rightarrow\text{O}} = \text{old } w_{kj}^{\text{H}\rightarrow\text{O}} - \alpha.\frac{\partial \text{E}}{\partial w_{kj}^{\text{H}\rightarrow\text{O}}} \tag{20}$$

After each update the gradient is re-evaluated for the new weight vector and this process is repeated. An important point to be noted is that the error function is defined with respect to a training set, and the entire training set is processed at each step, in order to compute the gradient. Such techniques which use the entire dataset at once are called batch methods. At each step the weight vector is moved in the direction of the greatest rate of decrease of the error function. This update is repeated by cycling through the data either in sequence or by selecting points at random with replacement.

7   ARTIFICIAL NEURAL NETWORKS - ANOTHER TECHNIQUE TO EXTRACT THE GLOBAL 21CM SIGNAL

The technique of ANN can be implemented to extract a faint cosmological signal from a significantly bright, dominant foreground. The most important requirement for a good model for predicting the parameters is a rich training dataset, which would contain several possible realizations of the Global signal, along with foreground added to it. We can build a good network by training, testing and validating it, the process has been described in detail in the above section. Here we have built an ANN which would estimate the signal and the foreground parameters. In the following sections we will describe in detail how we have formulated the problem and built the network.

8   SIMULATION

8.1   Constructing the training dataset

Our training dataset consists of 270 samples. Each sample in this training dataset is a total signal (T_{tot}), which contains one realization of the cosmological signal that has been generated using the tanh parameterization (§3) T_{21}, along with the foreground, which is generated by the log-polynomial model (§4) T_{FG}. This constitutes the training dataset, in the simplest case.

$$T_{\text{train}}(\nu) = T_{21}(\nu) + T_{\text{FG}}(\nu) \tag{21}$$

In the cases where we add the instrument response $G(\nu)$, the training dataset is constructed as:

$$T_{\text{train}}(\nu) = (T_{21}(\nu) + T_{\text{FG}}(\nu)) \cdot G(\nu) \tag{22}$$

A block diagram illustrating the entire process is shown in Fig. 7.

We use a feed-forward network, with two hidden layers. We have 1024 neurons in the input layer, corresponding to the 1024 frequency channels we choose to work with, for a bandwidth of $20 - 160$ MHz, and 18 neurons in the first hidden layer, which is connected to the next hidden layer with 14 neurons. The choice of the number of neurons in the hidden layer is determined after checking how the error is evolving. We begin with an initial guess, and vary the number of neurons in the hidden layers and the number of hidden layers, and note the numbers around which the RMSE is the least. We then decide upon the final structure of the network. Then, we vary the number of iterations to look at how the RMSE of the predicted parameters would change (see Fig: 8). We also look at the model loss, for those number of neurons, which should ideally decrease with increasing number of iterations and saturate after a certain point (see Fig. 9). Taking all these into consideration, we decide upon a number. The hidden layer neurons are activated by a sigmoid function. The number of neurons in the output layer is easy to choose. We have 11 output neurons corresponding to each parameter we are interested to find out, i.e, 7 signal parameters and 4 foreground parameters.

Our main aim is to extract the signal parameters and reconstruct the 21cm signal. For each case that we have studied, we trained and used different networks to predict the parameters. We have used packages SCIKITLEARN (Pedregosa et al. 2011) and KERAS to build simple neural networks. We construct a simple neural network with two Dense-layers, using the *Sequential* model available in *keras*.

8.2   Training, validation and testing the network

The training process is an iterative procedure that begins by collecting the data and preprocessing it to make training more efficient. This stage involves normalizing the data, and dividing it into training, validation and testing sets. In this type of network, the information flow is unidirectional that takes place with the help of an activation function between each layer. In our network, the data is normalized using min-max normalization. The entire dataset is split into 2 parts: training and the testing set, in a 7:3 ratio. The training set is further split into chunks of validation sets, iteratively, during the training process. Once the training is complete, the network is tested on the testing set.

The network uses back-propagation [see Fig. 6] as the learning technique. We choose the mean-squared error as the error-function. After repeating this process for a sufficiently large number of training cycles, the network will usually converge to some state where the error function is the minimum. In this case, one would say that the network has learned a certain target function. To adjust weights properly, we use the 'adam' (Adaptive Moment Estimation) optimizer for non-linear optimization. The main advantage of choosing *adam* over the commonly used *stochastic gradient descent* is that, it uses the combined averages of previous gradients at different moments to give it more expressive power to better update the parameters, in a more adaptive manner.

After training the network, we analyse the performance of the network. This analysis may lead us to discover problems with the data, network architecture, or the training algorithm. The entire procedure is repeated until the network performance is satisfactory. As we run a particular model for training, at each step we get a validation score, in the form of validation accuracy and model loss. In our models we had obtained a validation accuracy of around 80 percent.

### 8.3   Constructing the prediction datasets

The prediction set consists of samples which are unknown to the network and are used to test the robustness of our "trained network". In order to construct the prediction dataset, we use the models of 21cm signal and foreground, and add instrumental thermal noise to these datasets. We have produced a set of 90 prediction datasets following the equation:

$$T_{\text{pred}}(\nu) = T_{21}(\nu) + T_{\text{FG}}(\nu) + n(\nu) \tag{23}$$
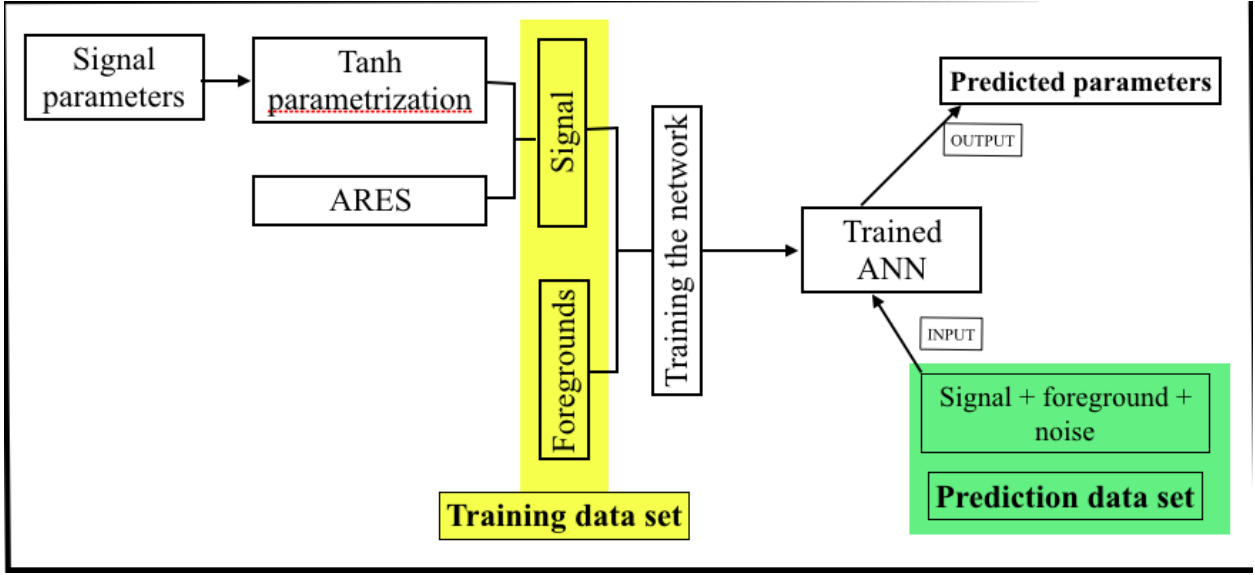
Figure 7. This shows how we have constructed the training datasets and built the network. Once the network is ready, we have fed the network with the prediction datasets, unknown to the network.
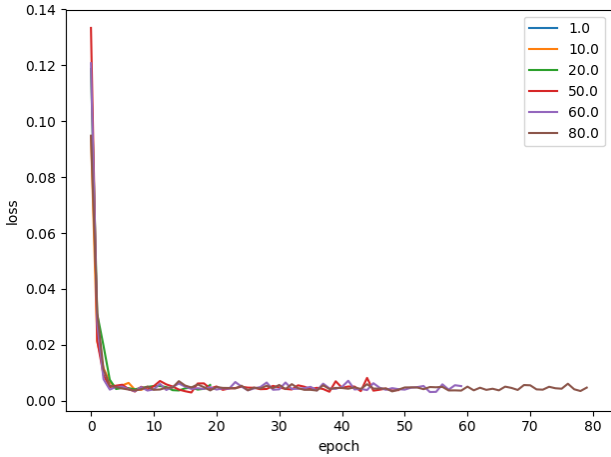

Figure 8. The RMSE calculated for the different parameters were plotted for different networks with varying number of epochs. One *epoch* is one complete pass of all the training samples, while an iteration is the total number of passes. Say, if we have 500 training samples, and a batch size of 100, then it will take 5 iterations to complete one epoch. We can see that for number of epochs greater than 50, all the parameters have a sufficiently low RMSE, which almost saturates.

In the cases where we have added the effect of the instrument, Eqn: 23 modifies as,

$$T_{\text{pred}}(\nu) = (T_{21}(\nu) + T_{\text{FG}}(\nu)) \cdot G(\nu) \; + n(\nu) \tag{24}$$

We use $T'_{sig}(\nu)$ to represent the prediction datasets in the following sections. We introduce the instrumental thermal noise corresponding to a certain hours of observations (N_t) to the prediction dataset.

From the ideal radiometer equation, the noise, $n(\nu)$, in the
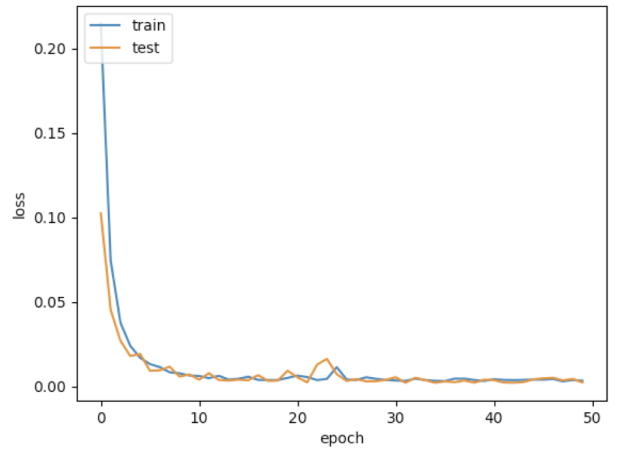
Figure 9. This is the model loss function for the perfect instrument case. We observe that for this case, the model loss saturates after around 50 epochs.

observed spectrum can be written as:

$$n(\nu) \approx \frac{T_{\text{sys}}(\nu)}{\sqrt{\Delta\nu \cdot \tau}} \tag{25}$$

where, $\Delta\nu$ is the bandwidth of the observation and τ is the observation time. At these low radio frequencies, T_{sys} is dominated by the brightness temperature due to the foregrounds. Hence, the above equation can be re-written as:

$$n(\nu) = \frac{T_{\text{FG}}(\nu)}{\sqrt{\Delta\nu \cdot 10^6 \cdot 3600 \cdot N_t}} \tag{26}$$

where, bandwidth $\Delta\nu$ is in MHz and N_t is in hours. The factors 10^6 and 3600 are the respective conversion factors to the standard units.

Here, we are dealing with mock observations which are in turn simulated using some assumed instrument response and
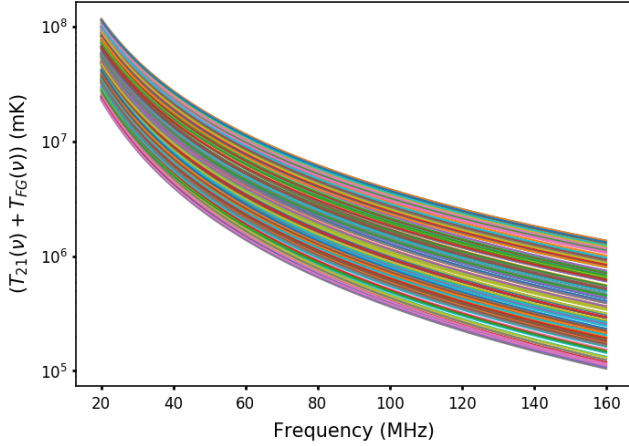
Figure 10. Training dataset for the perfect instrument (Case 1). Note how the signal is entirely dominated by the foregrounds.

observational strategy. In future, similar network will be used for real observations to predict the redshifted Global 21cm signal. For real observations, the prediction data will be replaced by the actual data. The error estimation in case of real observations will then be computed from the error we get during training the network for 'test' dataset.

9   RESULTS

In this section, we will discuss results from simulations depicting different signal extraction scenarios: no instrumental effect, time independent instrumental effect and time dependant instrumental effect.

9.1   Case 1: A Perfect Instrument

For the simplest case, we take the 21cm signal along with the model foreground, assuming that the instrument is perfect. This is an ideal case where the instrument does not modify the signal in any way. The training dataset is constructed using the following relation:

$$T_{\text{sig}}(\nu) = T_{21}(\nu) + T_{\text{FG}}(\nu) \tag{27}$$

where, T_{21} is the Global 21cm signal and T_{FG} is the model foreground (Eqn: 7) as described in §4. In Fig. 10 we show the training dataset, which is the total signal, T_{sig} in mK units, including the cosmological signal and the foreground. The network is trained using this dataset and saved. We construct a prediction dataset containing 90 samples, using the following equation:

$$T'_{\text{sig}}(\nu) = T_{21}(\nu) + T_{\text{FG}}(\nu) + n(\nu) \tag{28}$$

where, $n(\nu)$ is given by Eqn: 26. We use the saved network to predict the parameters from the prediction dataset. So we obtain 90 sets of predictions, for each sample in the prediction dataset. Following Shimabukuro & Semelin (2017), we calculate a normalized RMSE for the predictions:

$$\text{RMSE} = \sqrt{\frac{1}{N_{\text{pred}}} \sum_{i=1}^{N_{\text{pred}}} \left(\frac{Y_{\text{ori}} - Y_{\text{pred}}}{Y_{\text{ori}}}\right)^2} \tag{29}$$

where, Y_{ori} and Y_{pred} are the original and predicted values respectively, of one of the parameters, $Y \in [J_{ref}, J_{z0}, J_{dz}, X_{z0}, X_{dz}, T_{z0}, T_{dz}]$. N_{pred} denotes the total number of samples in the prediction dataset. The original and the predicted values of each of the parameters, along with their RMSE are plotted in Fig. 11. Lower value of RMSE implies more accurate prediction of the parameters. When the instrument is considered to be perfect, we can see that all the parameters are predicted with RMSE $\approx$ 0.02 [See Fig: 11].

In order to check the effect of the instrumental thermal noise on detection of Global 21cm signal, we have taken one sample 21cm Global signal and have varied the thermal noise, for observation times, $N_t \in (100, 500, 1000)$ hours [see Eqn: 26]. We have computed the corresponding residuals, which is the difference between the input signal and the reconstructed signals, and plotted them in Fig. 12. The reconstructed signal appears to have overlapped for all the N_t. But in the zoomed section, it can be clearly seen that the residual is least for the signal with 1000 hours of observation (blue curve, see Fig. 12).

9.2   Case 2: Imperfect instrument (fixed response)

The instrument response modifies the signal considerably. By a fixed instrument response we mean that the exact behaviour of the instrument is known, it is very well calibrated, and it does not vary with changes in the surroundings, through the observation time.

9.2.1   Simple instrument, fixed response

We have introduced the effect of an instrument response in §5 (Eqn. 9), where the antenna reflection coefficient (Γ) is represented by a set of sinusoidal harmonic basis functions. For a simple instrument, we consider only one sinusoidal term.

$$\Gamma(\nu) = c_0 + c_1 \ \sin(\omega_1 \nu) \tag{30}$$

The training dataset is constructed following the relation:

$$T_{\text{sig}}(\nu) = (T_{21}(\nu) + T_{\text{FG}}(\nu)) \cdot G(\nu, c_0, c_1, \omega_1) \tag{31}$$

Here, the function $G(\nu)$ is the instrument response and c_0, c_1, ω_1 are the parameters for the simple fixed instrument model. The network is trained with this dataset (see Fig. 13) and saved. The prediction dataset consists of a noisy data, which we feed into the network. The plots of the original versus the predicted values of the parameters, for the prediction dataset are shown in Fig. 14. The RMSE for the parameters in this case is $\lesssim 7\%$, and has clearly deteriorated as compared to the perfect instrument case.

We then take a fixed signal and vary the total observation time, to see how the percentage error varies with increasing N_t. The reconstructed signals from the predicted parameters are shown in Fig. 15. From the plot of the residual signal, it is observed that the residual is least for the observation time, $N_t = 1000$ hours shown by the green curve.

9.2.2   Moderate instrument(fixed response)

We train the network with a slightly more sophisticated instrument model, which we call a moderate instrument. We
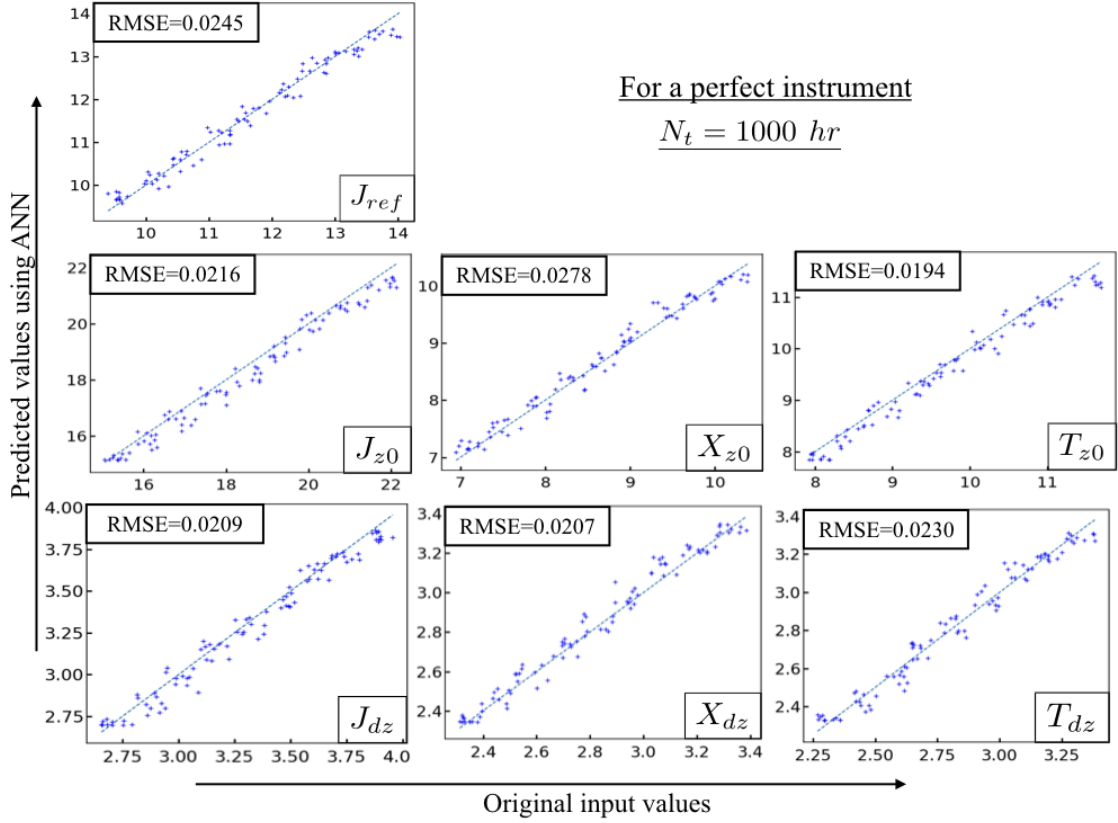
Figure 11. Case 1-Perfect Instrument. The true values vs predicted values of the signal parameters are plotted. The dashed straight line in each plot represents the true values of the parameters.
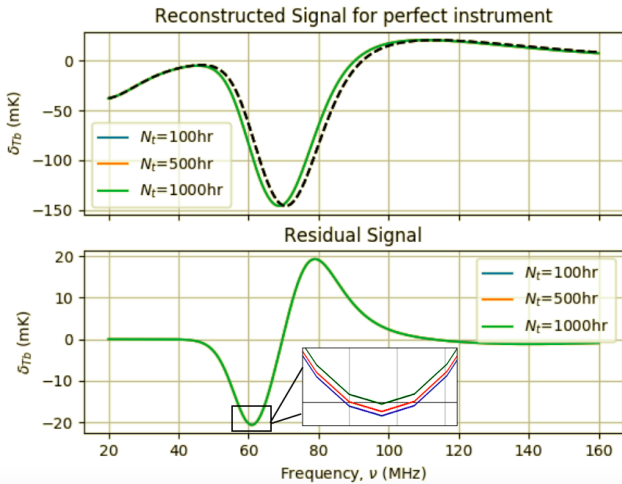
Figure 12. Case 1: Reconstructed signal for the different N_t assuming a perfect instrument (top panel). The residual signal, is plotted in the bottom panel, for increasing observation time, N_t. The zoomed section shows clearly how the residual decreases with increasing observation time.

introduce an instrument response, whose Γ consists of a combination of two sinusoids.

$$\Gamma(\nu) = c_0 + c_1 \ \sin(\omega_1 \nu) + c_2 \ \sin(\omega_2 \nu) \qquad (32)$$
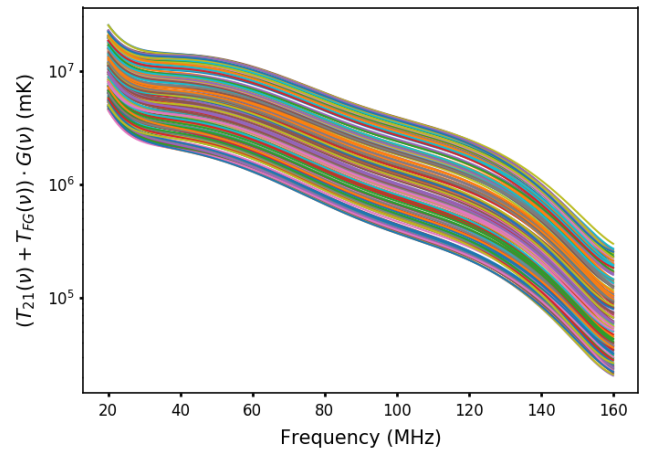
Figure 13. Case 2A: Training dataset for a simple instrument with a fixed response whose Γ is represented by one sinusoid. Note how the total signal is entirely dominated by the foregrounds and contaminated by the instrument effect.

The training dataset (see Fig. 16) is constructed as:

$$T_{sig}(\nu) = (T_{21}(\nu) + T_{FG}(\nu)) \cdot G(\nu, c_0, c_1, c_2, \omega_1, \omega_2) \qquad (33)$$

Here, $c_0, c_1, c_2, \omega_1, \omega_2$ are the parameters of the antenna reflection coefficient, and G has it's usual meaning. The noisy
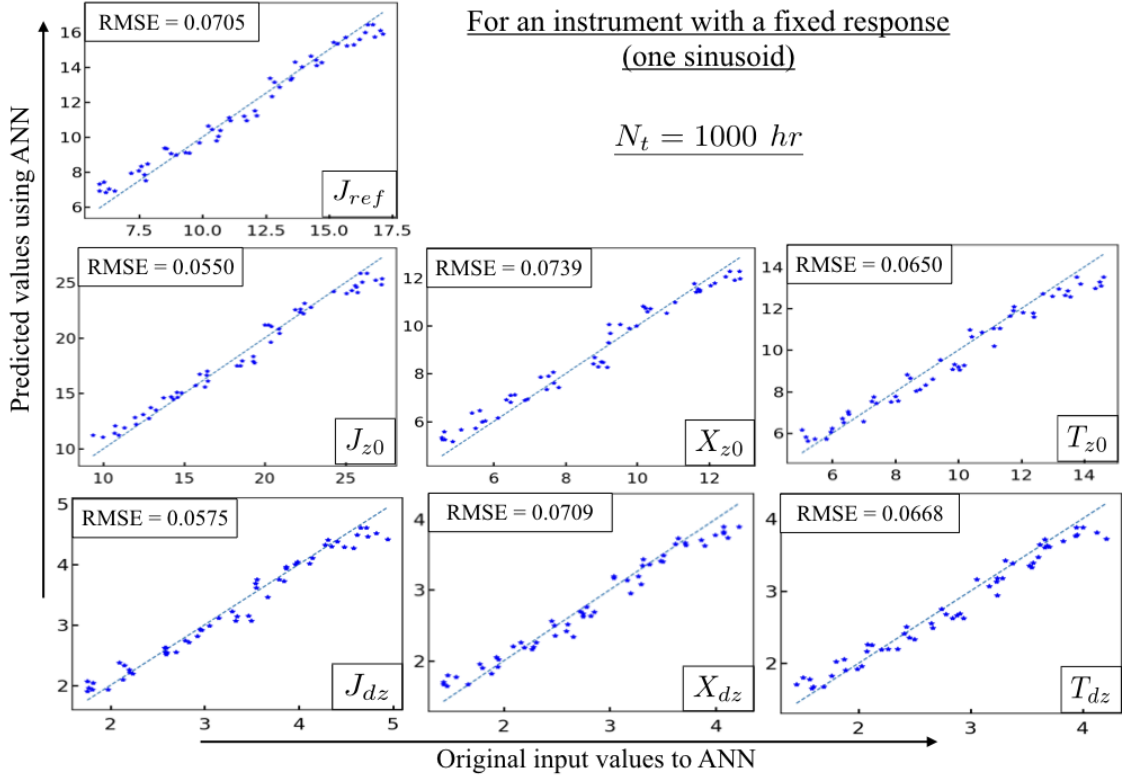
Figure 14. Case 2A: Fixed response, simple instrument.The dashed straight

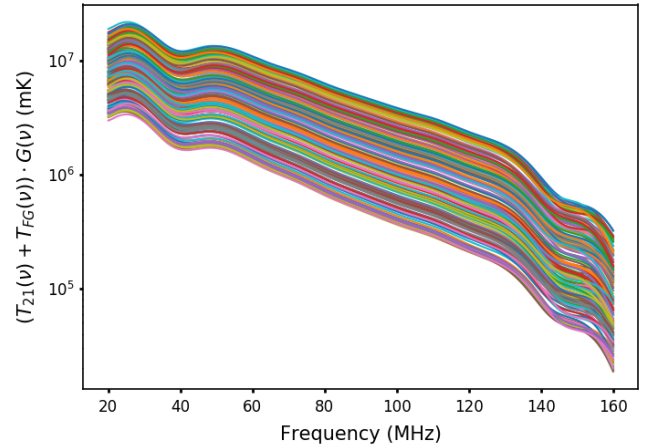line in each plot represents the true values of the parameters, the dots represent the predicted values.
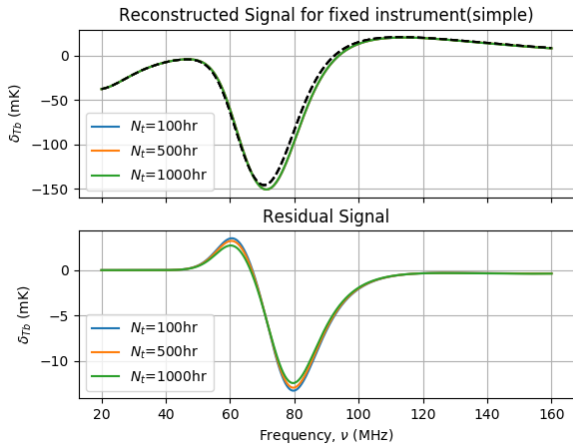
Figure 15. Case 2A: The reconstructed signal (top panel) and the residual signal, with increasing observation time, N_t. The residual signal gives an idea of the quality of the reconstructed Global 21cm signal with the change in integration time.

Figure 16. Case 2B: Fixed response, moderate instrument (Γ consists of 2 sinusoids): Training dataset for the network

prediction dataset is then fed into the network.

$$T'_{sig}(\nu) = (T_{21}(\nu) + T_{FG}(\nu)) \cdot G(\nu) + n(\nu) \qquad (34)$$

where, noise, n is given by Eqn. 26.
The plots of the input versus the predicted values are shown

in Fig. 17. The reconstructed signals from the predicted parameters and the residuals are shown in Fig. 18. The RMSEs for the parameters in this case is $\lesssim 6\%$, and has also deteriorated as compared to the perfect instrument case.
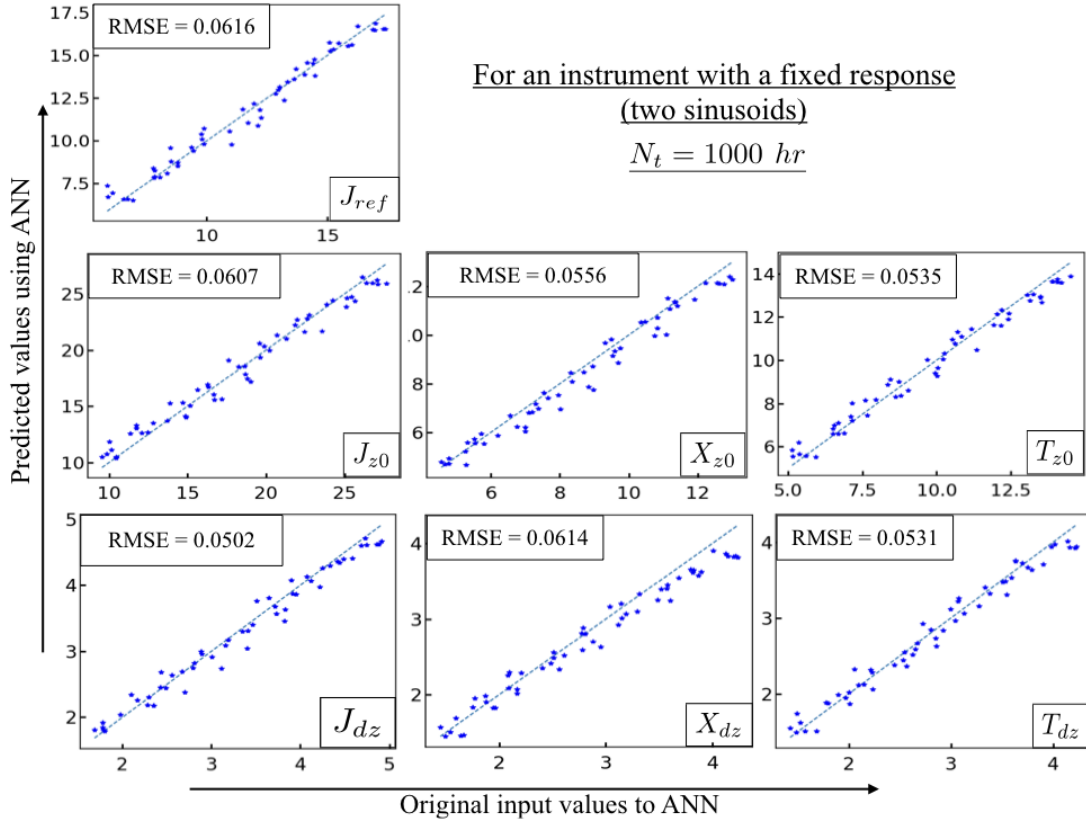
Figure 17. Case 2B:Fixed response, moderate instrument.The true values vs predicted values of parameters are plotted. The dashed straight line in each plot represents the true values of the parameters.
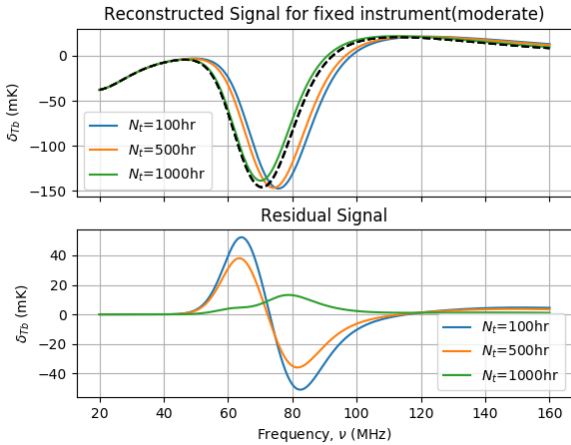
Figure 18. Case 2B: The reconstructed signal (top panel) and the residual signal(bottom panel), with increasing observation time, N_t. The residual signal gives an idea of the quality of the reconstructed Global 21cm signal with the change in integration time.

10 SUMMARY AND DISCUSSIONS

We have presented the detailed results obtained by using ANN to extract the cosmological Global signal, from a combined spectrum, which included bright dominant fore-grounds, model instrumental effects, and thermal noise. We have considered two types of instrument models for this - a simple and a moderate instrument [see §5]. We have demonstrated how a simple ANN can easily handle up to 16 parameters (7 signal parameters, 4 foreground parameters, 5 instrument parameters, in the most complicated scenario which we have considered). The number of neurons in the hidden layer was optimized for each case for the best results. We summarize the calculated normalized RMSE, for each parameter, in each of the cases studied, in Table: 1.

If we look at the RMSE's in Table: 1, we see the effect of increasing the complexity of the input data on the accuracy levels. If the network is trained sufficiently well, then the accuracy levels continue to be high. All errors are less than 10%, the worst prediction being around $\sim 8\%$. The accuracy levels obtained are between $\sim 92 - 98\%$ for the signal parameters. The RMSEs for the cases which includes the fixed response (*2A* & *2B*), for the simple and moderate instruments, are not very different. This is because we are considering simple instrument models, where the moderate instrument model just has an additional sinusoidal term. However, if we compare to the perfect instrument case, we can see that the contamination affects the errors in predictions.

To the best of our knowledge, most authors have used MCMC, nested sampling or similar methods to sample the parameter space. We cannot directly compare the speed of ANN with an efficient MCMC approach, as the methodologies are different. But, we bypass the task of computing the

Parameters	Perfect Instrument Case 1	Fixed(simple) Instrument Case 2A	Fixed(moderate) Instrument Case 2B
J_{ref}	0.0245	0.0705	0.0616
dz_J	0.0209	0.0575	0.0502
dz_T	0.0230	0.0668	0.0531
dz_X	0.0207	0.0709	0.0614
$z0_J$	0.0216	0.0550	0.0607
$z0_T$	0.0194	0.0650	0.0535
$z0_X$	0.0278	0.0739	0.0556

Table 1. The RMSE values for the signal parameters, computed for the prediction dataset for each of the cases studied, are summarized here. Considering the perfect instrument case to be the reference, we can see how the errors increase in the cases where the instrument effect is considered.

likelihood function, a large number of times, to arrive at the inferred values of the parameters, while using ANN. So, when dealing with a higher dimensional parameter space, ANN is faster and computationally more efficient. In the case of neural networks, we can think of the training sets as a more realistic equivalent to the prior in MCMC.

In Harker et al. (2016), Gaussian priors for the foreground and uniform priors for the signal parameters have been used. The authors have inferred tanh parameters from a fit to the reference ARES model. They have constructed synthetic datasets, by assuming an ideal instrument and observing four different sky regions. In later works, (Mirocha et al. 2015) have used broad, uninformative priors, but with simple foregrounds and ideal instruments. Bernardi et al. (2016), used uniform priors on all parameters. We have not made any such assumptions in our work, and have predicted the signal parameters with good accuracy. We do not require any specific model for fitting, neither do we require extensive prior information. So, we can consider this technique of signal extraction using ANN to be an alternative to the already existing techniques like MCMC explorations.

Our simple network has been able to detect the Global signal with an accuracy $\gtrsim 92\%$, in cases of mock observations where the instrument has some residual time-varying gain across the spectrum. This is just a proof of concept paper, and we will deal with more complex scenarios in future work. The computations and results presented in this work has been performed on a personal computer (Macbook Air, 8GB RAM) only and does not require too much computational power. This shows how conveniently these algorithms can be used on any PC. However, in order to establish the robustness of the ANN framework for the Global 21cm signal extraction, we need to test this algorithm for more realistic instrument responses like that of EDGES, SARAS, etc. We also plan to address the issue of effect of chromatic primary beam and changes in the foreground model spectrum due to the same. In future, we would like to use ionospheric models to corrupt the datasets and extract 21cm signal in presence of the atmosphere. Moreoever, the 21cm signal model is now dependent on 'tanh' parametrization. But ANN can handle an non-parametrized 21cm signal. This will be used in future publications. At the end, we wish to apply the developed ANN framework on some real data for validation.

11   ACKNOWLEDGEMENTS

REFERENCES

Barkana R., Loeb A., 2001, Phys. Rep., 349, 125
Barkana R., Loeb A., 2005a, ApJ, 624, L65
Barkana R., Loeb A., 2005b, ApJ, 626, 1
Bernardi G., McQuinn M., Greenhill L. J., 2015, ApJ, 799, 90
Bernardi G., et al., 2016, MNRAS, 461, 2847
Bowman J. D., Rogers A. E. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, Nature, 555, 67
Bradley R., 2012, DARE Technical Memo, p. 34
Burns J. O., et al., 2012, Advances in Space Research, 49, 433
Burns J. O., et al., 2017, The Astrophysical Journal, 844, 33
Carilli C. L., Gnedin N., Furlanetto S., Owen F., 2004, New Astron. Rev., 48, 1053
Choudhury M., Datta A., 2017, Proceedings of the International Astronomical Union, 12, 280
Cohen A., Fialkov A., Barkana R., Lotem M., 2017, MNRAS, 472, 1915
Ewall-Wice A., Chang T.-C., Lazio J., Doré O., Seiffert M., Monsalve R. A., 2018, ApJ, 868, 63
Fan X., Narayanan V. K., Strauss M. A., White R. L., Becker R. H., Pentericci L., Rix H.-W., 2002, AJ, 123, 1247
Fan X., Carilli C. L., Keating B., 2006, ARA&A, 44, 415
Fialkov A., Barkana R., Pinhas A., Visbal E., 2014, MNRAS, 437, L36
Fialkov A., Barkana R., Cohen A., 2015, Physical Review Letters, 114, 101303
Fialkov A., Barkana R., Cohen A., 2018, Phys. Rev. Lett., 121, 011101
Field G. B., 1959, ApJ, 129, 536
Furlanetto S. R., Briggs F. H., 2004, New Astron. Rev., 48, 1039
Furlanetto S. R., Sokasian A., Hernquist L., 2004, MNRAS, 347, 187
Furlanetto S. R., Oh S. P., Briggs F. H., 2006, Phys. Rep., 433, 181
Greenhill L. J., Bernardi G., 2012, arXiv e-prints, p. arXiv:1201.1700
Harker G. J. A., 2015, MNRAS, 449, L21

Harker G. J. A., Pritchard J. R., Burns J. O., Bowman J. D., 2012, MNRAS, 419, 1070

Harker G. J. A., Mirocha J., Burns J. O., Pritchard J. R., 2016, MNRAS, 455, 3829

Hewitt J., Ewall-Wice A., Dillon J. S., Mesinger A., Liu A., Pober J., 2016, Monthly Notices of the Royal Astronomical Society, 458, 2710

Kingma D. P., Ba J., 2014, preprint, (arXiv:1412.6980)

Koopmans L., et al., 2015, Advancing Astrophysics with the Square Kilometre Array (AASKA14), p. 1

Loeb A., Barkana R., 2001, ARA&A, 39, 19

Mirocha J., Skory S., Burns J. O., Wise J. H., 2012, ApJ, 756, 94

Mirocha J., Harker G. J. A., Burns J. O., 2013, ApJ, 777, 118

Mirocha J., Harker G. J. A., Burns J. O., 2015, ApJ, 813, 11

Morales M. F., Wyithe J. S. B., 2010, ARA&A, 48, 127

Mortlock D. J., et al., 2011, Nature, 474, 616

Patra N., Subrahmanyan R., Raghunathan A., Udaya Shankar N., 2013, Experimental Astronomy, 36, 319

Pedregosa F., et al., 2011, Journal of Machine Learning Research, 12, 2825

Price D. C., et al., 2018, MNRAS, 478, 4193

Pritchard J. R., Loeb A., 2010, Phys. Rev. D, 82, 023006

Pritchard J. R., Loeb A., 2012, Reports on Progress in Physics, 75, 086901

Pritchard J., et al., 2015, Advancing Astrophysics with the Square Kilometre Array (AASKA14), p. 12

Schmit C. J., Pritchard J. R., 2018, MNRAS, 475, 1213

Shimabukuro H., Semelin B., 2017, MNRAS, 468, 3869

Singh S., et al., 2017, ApJ, 845, L12

Sokolowski M., et al., 2015, Publ. Astron. Soc. Australia, 32, e004

Tauscher K. A., Burns J. O., Rapetti D., Mirocha J., Monsalve R. A., 2017, in American Astronomical Society Meeting Abstracts #229. p. 347.01

Tauscher K., Rapetti D., Burns J. O., Switzer E., 2018, The Astrophysical Journal, 853, 187

Voytek T. C., Natarajan A., Jáuregui García J. M., Peterson J. B., López-Cruz O., 2014, ApJ, 782, L9

Wouthuysen S. A., 1952, AJ, 57, 31

de Oliveira-Costa A., Tegmark M., Gaensler B. M., Jonas J., Landecker T. L., Reich P., 2008, MNRAS, 388, 247

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