

Direct parameter inference from global EoR signal with Bayesian statistics

Junhua Gu ¹★ and Jingying Wang ²★

¹National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Beijing 100101, China

²Department of Physics and Astronomy, University of the Western Cape, Cape Town 7535, South Africa

Accepted 2020 January 6. Received 2020 January 2; in original form 2019 May 16

ABSTRACT

In the observation of sky-averaged HI signal from Epoch of Reionization (EoR), model parameter inference can be a computation-intensive work, which makes it hard to perform a direct one-stage model parameter inference by using Markov Chain Monte Carlo (MCMC) sampling method in Bayesian framework. Instead, a two-stage inference is usually used, i.e. the parameters of some characteristic points on the EoR spectrum model are first estimated, which are then used as the input to estimate physical model parameters further. However, some previous works had noticed that this kind of method could bias results, and it could be meaningful to answer the question of whether it is feasible to perform direct one-stage MCMC sampling and obtain unbiased physical model parameter estimations. In this work, we studied this problem and confirmed the feasibility. We find that unbiased estimations to physical model parameters can be obtained with a one-stage direct MCMC sampling method. We also study the influence of some factors that should be considered in practical observations to model parameter inference. We find that a very tiny amplifier gain calibration error (10^{-5} relative error) with complex spectral structures can significantly bias the parameter estimation; the frequency-dependent antenna beam and geographical position can also influence the results, so that should be carefully handled.

Key words: methods: numerical – methods: statistical – cosmology: observations – dark ages, reionization, first stars – diffuse radiation.

1 INTRODUCTION

Detection to the cosmic dawn and the following Epoch of Reionization (EoR hereafter) has become the frontier of radio astronomy and observational cosmology. It is crucial to answer questions such as what sources are responsible for the reionization, how the neutral hydrogen fraction evolved during EoR, how the mass accretion on to dark haloes modulates the ionization process, etc. The redshifted neutral hydrogen 21 cm signals from the early Universe are generally regarded to bring critical information to answer the above questions. Compared with abundant observations to the last scattering surface (i.e. the farthest Universe we can see through the electromagnetic window) and low redshift (e.g. $z \lesssim 6$) Universe, much fewer observational evidence about the cosmic dawn and EoR exists by now, which is necessary to discriminate candidate models and constrain cosmological model parameters. In order to fill this gap, the new generation of radio telescopes have been built or being constructed. The very long list of these new radio telescopes includes the Low Frequency Array, LOFAR (van Haarlem et al. 2013), the Murchison Widefield Array, MWA (Tingay et al. 2013), the 21 CentiMeter Array, 21CMA (e.g. Wang et al.

2013; Zheng et al. 2016; Huang et al. 2016), the Hydrogen Epoch of Reionization Array, HERA (DeBoer et al. 2017), the Precision Array to Probe the Epoch of Reionization, PAPER (Parsons et al. 2010), the Long Wavelength Array, LWA (Taylor et al. 2012), the Square Kilometre Array, SKA (Carilli & Rawlings 2004), the Shaped Antenna measurement of the background Radio Spectrum, SARAS (Patra et al. 2013), the Broadband Instrument for Global Hydrogen Reionisation, BIGHORNS (Sokolowski et al. 2015), the Experiment to Detect the Global EoR Signature, EDGES (Mozdzen et al. 2016), the Sonda Cosmológica de las Islas para la Detección de Hidrógeno Neutro, SCI-HI (Voytek et al. 2014), etc. The detection to the HI 21 cm signal from the cosmic dawn and EoR is extremely hard. Several conflicting factors should be compromised, including angular resolution, redshift resolution, signal-to-noise ratio (S/N), survey speed, the quantity of data to be processed, and accuracy of instrument calibration. Different paradigms of detection reflecting the concerning to different factors are invented.

The three major paradigms are (1) imaging, (2) measuring power spectrum, and (3) measuring global 21 cm signal (i.e. the monopole component), which have their own problems respectively to solve and can produce different levels of details about cosmic dawn and EoR. Direct imaging to cosmic dawn and EoR can obtain the most abundant information among all the above three paradigms, but it requires to solve big problems, including direction-dependent

* E-mail: jhgu@nao.cas.cn (JG), astro.jywang@gmail.com (JW)

ionosphere correction and wide-field imaging difficulties. Interferometers can directly output power spectrum and can seemingly escape from the difficulties of direct imaging paradigm (e.g. Zheng et al. 2012). However, year scale integration time is still needed (Zheng et al. 2012) to suppress the thermal noise. Measuring global 21 cm signal needs much less integration time to reach a required S/N. Direction-dependent ionosphere correction and wide field imaging can be avoided as well. Furthermore, the instrument for measuring global 21 cm signal can be as simple as a single dipole, so that the response of antenna can be relatively precisely measured in a microwave darkroom (with near-field measuring technique). Simplicity is pre-requisite for reliability. Though losing a significant amount of details (only monopole component can be measured), especially before other paradigms obtain reliable results, global 21 cm signal can be a promising method for detecting the cosmic dawn and EoR.

The framework of global 21 cm signal studies can be divided into following three tightly related aspects as the global 21 cm signal modelling, statistical inferring based on observational data, and the obtaining of accurate enough observational data.

For the modelling part, it has been widely accepted that the evolution of observed 21 cm signal brightness temperature contrast to CMB is dominated by the competition between the influence of cosmic microwave background (CMB) and gas on the atomic hydrogen (HI) spin temperature (e.g. Furlanetto, Oh & Briggs 2006). The CMB temperature decreases as it adiabatically expands with the universe and follows a relation to redshift z as $T_\gamma \propto (1+z)$. Meanwhile, the gas kinetic temperature is simultaneously modulated by the universe expansion and heating and cooling of a variety of different mechanisms. The CMB affects spin temperature mainly through Compton scattering, and gas affects spin temperature through collision and Wouthuysen–Field effect (Wouthuysen 1952; Field 1958). The strength of above three effects are mainly controlled by gas neutral fraction, gas temperature, and background Ly α intensity, which are all tightly connected to the evolving history of potential ionizing sources, i.e. black holes and stars (see Mirocha 2014; Mirocha et al. 2012, for a quantitative discussion).

Inferring model parameters from observational data is another critical aspect of global 21 cm signal detection. Optimization-based model fitting is a traditional method of parameter inferring; however Markov Chain Monte Carlo (MCMC) based methods are gradually becoming major statistical inference tools and have been widely used in the studies of global 21 cm signal. A lot of recent works e.g. Mirocha, Harker & Burns (2013), Mirocha, Harker & Burns (2015), Harker et al. (2016), Bernardi et al. (2016), and Spinelli et al. (2019) have used MCMC-based methods as their major methods of statistical inferences. Comparing to classical optimization-based methods, MCMC-based methods cannot only solve the optimal model parameters but also inspect the probability distribution of model parameters, so that can offer the information about the structure of the model parameter space. However, there is still one shortage of MCMC-based methods, i.e. most of such kind of methods require evaluating model values for a vast number of times to ensure the Markov chains to reach equilibrium. Considering the factor that the model evaluation of the global 21 cm signal itself is computation intensive, it requires a large amount of computing resources to perform a direct model evaluation in MCMC methods. Currently, few work directly evaluate the model predicted values in the step of parameter inference. At least two methods are utilized to avoid direct model evaluation in MCMC, and the first one is to approximate the global 21 cm signal model with some simple

mathematical expressions (e.g. Bernardi et al. 2016, modelled the early reionization stage signal with a Gaussian model), another method is to split the MCMC inference into two stages (e.g. Mirocha et al. 2013, 2015; Harker et al. 2016). In these works, they first infer the frequencies and intensities of some characteristic points of the global 21 cm signal spectrum, and then in the second stage use the distribution of these characteristic points to infer the values of physical model parameters. The pitfall of the expressing the global 21 cm signal model in a simple mathematical formula is that either the formula cannot fully approximate the precise model values, or the formulae would have too many degrees of freedom, as a result, cannot be well constrained and lead to the overfitting problem. As to the two-stage method, as Harker et al. (2016) mentioned, there exists a risk to obtain biased results. In the same paper, the authors raised a question of whether it is feasible to perform direct single-stage parameter inference and whether unbiased parameter estimation can be obtained.

The efforts to obtain the well-calibrated observational data forms the last aspect of the study to the global EoR signal detection. This problem mainly falls into the category of antenna and electronic system design. The issue about how to design and calibrate electronic system will not be covered in this work, so we are not going to expand this topic here. However, there are two interesting issues in the early stages of projects that are aiming to detect global 21 cm signals. The first issue is how accurate the analogue front end should be calibrated and how the calibration errors propagate in the data reduction chain. The second one is to guide the design of the antenna by evaluating the influence of frequency-dependent antenna pattern (FDAP) through simulation. These two issues raise for both of them can cause artificial frequency structures in the global 21 cm signal spectrum. In some previous works, e.g. Bernardi et al. (2016), the amplifier calibration error is treated to be uncorrelated between different frequency channels, which, however, does not fully reflect its influence on the measured data. The frequency-dependent antenna beam issue has been mentioned by e.g. Liu et al. (2013), and is also widely considered in the antenna design of recent projects. For example, in experiments including SARAS (Patra et al. 2013), BIGHORNS (Sokolowski et al. 2015), EDGES (Mozdzen et al. 2016), and SCI-HI (Voytek et al. 2014), the antennas are specially designed to make the beam pattern less frequency-dependent. Nevertheless, the effect of frequency-dependent beam cannot be fully cancelled, which must be evaluated and considered in actual data reductions.

The motivation of this work is to inspect what information we can extract from the global HI 21 cm signal in actual observation conditions. This goal can be further divided into the following two aspects. The first aspect of this work is to find out the feasibility of inferring physical model parameters from observation data through a direct physics-based modelling method. We intend to evaluate the physics-based global 21 cm model in each step of the MCMC sampling and to check whether unbiased parameter estimations can be inferred. The second motivation is to study in practical observations, how will different factors, including thermal noise, instrument calibration errors, and frequency-dependent antenna beam patterns, affect model parameter inference, and model comparison.

So in this work, we investigate the feasibility of constraining the parameters of physical global 21 cm signal model with MCMC sampling method. We will evaluate the influence of uncorrected instrumental effects and frequency-dependent antenna beam patterns on the results and estimate how well the instruments should be calibrated in order to gain meaningful results. The

paper is organized as follows, we explain the modelling method of the reionization processes in Section 2, describe the MCMC sampling method that we use to constrain the model and the model parameters in Section 3, present the results in Section 4, and give a discussion and conclude this work in Section 5. We adopt the Λ -cold dark matter cosmological parameters released by Planck Collaboration et al. (2016) ($H_0 = 67.74 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_{\Lambda,0} = 0.3089$, $\Omega_{\text{m},0} = 0.3089$, $\Omega_{\text{b},0} = 0.0486$, $\sigma_8 = 0.8159$, and $n_s = 0.9667$).

2 MODELLING SKY-AVERAGED LOW-FREQUENCY SPECTRUM

In this section, we present the global HI 21 cm signal model of cosmic dawn and EoR used in this work. We first briefly review the algorithm for computing the global HI signal based on previous studies. Then, the parameter values for simulating the reference 21 cm signal are presented. Last, we show the foreground emission spectrum model.

2.1 Redshifted 21 cm signal from the EoR

The computing of the global HI signal involves three threads, i.e. (1) solving radiative transfer equation (RTE) to determine the background radiation intensity, (2) updating ionization, recombination, cooling and heating rate coefficients to compose the rate equations, and (3) solving the rate equations for each time slice sequentially. These threads are twisted and complicated. Even though it is not the central topic of this paper, and has been well and comprehensively studied by many other authors, we still choose to briefly review the modelling method as follows.

To model the global HI signal in the low-frequency radio band, we follow the method and algorithm that has been comprehensively described by previous works mainly including Furlanetto et al. (2006), Mirocha et al. (2012), and Mirocha (2014). Among these three works, Mirocha (2014) described the detailed algorithm to calculate the global 21 cm signal, together with some important practical instructions; Mirocha et al. (2012) described the one dimensional radiative transfer algorithm used by the former work, and the authors released a PYTHON package *ARES*¹ (Mirocha et al. 2012; Mirocha 2014), that has been being actively developed; Furlanetto et al. (2006) summarized equations to calculate the physical quantities that are required to derive the global 21 cm signal. Although the package *ARES* is well developed and is in principle suitable for solving the problem of this work, MCMC sampling that we will perform requires evaluating the model with different parameter sets for a vast number of times, and the PYTHON-based package *AREA* becomes a performance bottleneck for us. As a result, we take the *ARES* code as a reference and re-implement the algorithm in *RUST*² with some minor modifications to gain a required high computing efficiency so that in very limited time, we can draw a large enough sample with the MCMC method. As a necessary comment to the code performance, although our code is implemented with *RUST* and compiled to native code, the computational performance is improved by less than one order of magnitude. It is still practical to perform similar MCMC sampling with the *ARES* code.

¹<https://bitbucket.org/mirochaj/ares>

²<https://www.rust-lang.org>

Followings are the details about the global 21 cm signal modelling algorithm. The redshifted global 21 cm signal from cosmic dawn and EoR is calculated as

$$\delta T_{\text{b}} = 0.027(1 - \bar{x}_i) \left(\frac{\Omega_{\text{b},0} h^2}{0.023} \right) \left(\frac{0.15}{\Omega_{\text{m},0} h^2} \frac{1+z}{10} \right)^{1/2} \left(1 - \frac{T_{\text{y}}}{T_{\text{S}}} \right) \text{K}, \quad (1)$$

where \bar{x}_i is the mean ionized fraction and T_{S} is the spin temperature. Both \bar{x}_i and T_{S} are to be solved as follows.

We follow the ‘two-component’ scenario utilized by e.g. Mirocha (2014) and Furlanetto et al. (2006). The two-component scenario divides the intergalactic medium (IGM) into the ‘bulk IGM’ that is mostly neutral, and H II regions that is fully ionized. The filling factor of H II region is denoted by x_i and the ionization fraction in the bulk IGM regions is denoted by x_{HII} , so that in above equation (1), the average ionization fraction is calculated as $\bar{x}_i = x_i + (1 - x_i)x_{\text{HII}}$. For the bulk IGM regions, the fraction of H II can be solved from following rate equation.

$$\frac{d}{dt} x_{\text{HII}} = (\Gamma_{\text{HI}} + \gamma_{\text{HI}} + \beta_{\text{HI}} n_e) x_{\text{HII}} - \alpha_{\text{HII}}^{\text{B}} n_e x_{\text{HII}}, \quad (2)$$

where n_e is the electron density and the meanings of other symbols are listed in Table 1. The temperature of bulk IGM is solved from the equation

$$\frac{d}{dt} T_{\text{K}} = -2H(z)T_{\text{K}} + \frac{2}{3} \frac{\epsilon_{\text{X}} + \epsilon_{\text{comp}} - \mathcal{C}}{k_{\text{B}} n_{\text{tot}}}, \quad (3)$$

where n_{tot} is the total number density of particles, and $H(z)$ is the Hubble constant at redshift z . \mathcal{C} represents the contribution of all cooling mechanisms, including Hubble cooling, collisional ionization cooling, collisional excitation cooling, and recombination cooling. The coefficients of the latter three cooling mechanisms are calculated with the equations in Fukugita & Kawasaki (1994). The rate equation for the filling factors of the H II region surrounding galaxies is

$$\frac{dx_i}{dt} = f_* f_{\text{esc}} N_{\text{ion}} \bar{n}_{\text{b}}^0 \frac{df_{\text{coll}}}{dt} (1 - x_{\text{HII}}) - \alpha^{\text{A}} n_e x_i C(z) \quad (4)$$

where f_{coll} is the fraction of gas in collapsed haloes more massive than some minimum mass (see below), f_* is the star formation efficiency, and \bar{n}_{b}^0 is the baryon number density today. We set the clumping factor $C(z) \equiv 1$, same as Mirocha (2014). Ionization on the boundary of H II regions is assumed to be mainly driven by photons with energy between 13.6 and 200 eV, the escape fraction of which is denoted by f_{esc} . N_{ion} is the number of photons per baryon particle in star formation and is set to be 4000 (Same as Mirocha 2014). The photon ionization coefficient Γ_{HI} , ionization rate due to fast secondary electrons γ_{HI} , and X-ray heating rate ϵ_{X} in equation (3) are determined by background radiation intensity $\hat{J}_v(z)$ (see equations 9–11 of Mirocha 2014, \hat{J}_v means the intensity in the unit of photon numbers), which is obtained from the solution of the cosmological RTE (see the equation 3 in Mirocha 2014) as

$$\hat{J}_v(z) = \frac{c}{4\pi} (1+z)^2 \int_z^{z_{\text{f}}} \frac{\hat{\epsilon}_{\text{X},v}(z')}{H(z')} e^{-\bar{\tau}_v} dz', \quad (5)$$

where the photon depth $\bar{\tau}_v$ is the mean optical depth (see below), and z_{f} is some ‘first light redshift’, when the ionizing sources turn on. The emissivity $\hat{\epsilon}_{\text{X},v}(z)$ (again in the unit of photon numbers) depends on the accretion rate and model of first generation of luminous celestial objects as

$$\hat{\epsilon}_{\text{X},v}(z) = \rho_{\text{b}}^0 \frac{C_{\text{X}}}{E_{\text{v}}} f_* f_{\text{X}} \frac{df_{\text{coll}}}{dt} I_v, \quad (6)$$

Table 1. The references of how following quantities are calculated.

Quantity	Meaning	Related equation in this work	Ref.
α^A, α^B	Cases-A and -B recombination coefficients	2, 4	Equations (B5)–(B9) of Fukugita & Kawasaki (1994)
β	Collisional ionization coefficients	2	Equations (B1)–(B4) of Fukugita & Kawasaki (1994)
ϵ_{comp}	Compton heating coefficient	3	Equation (54) of Seager, Sasselov & Scott (2000)
ϵ_X	X-ray heating coefficient	3	Equation (11) of Mirocha (2014)
γ	Photon ionization coefficients	2	Equation (10) of Mirocha (2014)
Γ	Secondary photon ionization coefficients	2	Equation (9) of Mirocha (2014)
J_ν	Background intensity	5	Equations (4) or (25) of Mirocha (2014)
J_α	Ly α background intensity	14	Equations (7) or (25) of Mirocha (2014)
τ_ν	Optical depth	5, 8	Equation (6) of Mirocha (2014)
T_K	Gas kinetic temperature	3, 13	Equation (14) of Mirocha (2014)
T_S	H I spin temperature	1, 13	Equation (20) of Furlanetto et al. (2006)
$x_{\text{HI}}, x_{\text{HII}}$	Fraction of H I and H II in bulk IGM region, respectively	2	Equation (12) of Mirocha (2014)
x_i	Volume filling fraction of H II regions	4	Equation (13) of Mirocha (2014)
x_c	Collisional coupling coefficient	13	Equation (24) of Furlanetto et al. (2006)
x_α	Wouthuysen–Field coupling coefficient	13	Equation (40) of Furlanetto et al. (2006)

where $\bar{\rho}_b^0$ is today’s baryon matter density, I_ν is the source spectrum that is normalized as $\int I_\nu d\nu = 1$, and $C_X = 2.6 \times 10^{39} \text{ erg s}^{-1} M_\odot^{-1} \text{ yr}$ is the energy of X-ray emission produced when unit mass of matter is accreted \bar{E}_γ is the mean photon energy within the total band, and f_X reflects the uncertainty of C_X . In the process of heating and ionization in bulk IGM regions, we follow the treatment of Mirocha (2014) that only X-ray radiation within 0.2–30 keV is considered and assume the first generation of luminous celestial objects to be accreting black holes, with a power-law spectrum with a fixed slope of -1.5 . The energy range of the X-ray responsible for reionization and the slope of the power-law spectrum are both fixed because allowing these parameters to vary will introduce degeneracy in parameter space and make the Markov Chain hard to reach an equilibrium. Studying which kind of source is responsible for the reionization is not the major topic of this paper. In our future work, we will study the feasibility of discriminating different ionizing sources by using the global H I signal. We note that more sophisticated methods have been developed recently (e.g. Mirocha et al. 2018; Park et al. 2019, based on luminosity functions) to model ionizing sources more accurately. These new modelling methods offer the possibility of inferring more detailed information about the cosmic dawn and EoR, and we will consider including them in our future work. In this work, we still use the method that we have described above.

In equation (5), the emitting frequency ν' of a photon emitted at redshift z' that is observed at frequency ν at redshift z is

$$\nu' = \nu \left(\frac{1+z'}{1+z} \right). \quad (7)$$

The optical depth $\bar{\tau}_\nu$ is calculated as

$$\bar{\tau}_\nu(z, z') = \sum_j \int_z^{z'} n_j(z'') \sigma_{j,\nu''} \frac{dl}{dz''} dz'', \quad (8)$$

where $dl/dz = cH(z)/(1+z)$ is the proper cosmological line element, and $\sigma_{j,\nu}$ is the bound-free absorption cross-section of species $j = \text{HI}, \text{HeI}, \text{HeII}$ with number density n_j . Same as Mirocha (2014), we approximate x_{HeIII} to be 0 and assume $x_{\text{HeII}} = x_{\text{HII}}$. Note that n_j is a function of redshift, which is to be solved in equation (2). Fortunately, n_j at any redshift z_i relies only on previous redshift slices with $z > z_j$. Nevertheless, a strict calculation requires solving equations (2) and (8) alternatively. This issue has been addressed in section 5.1 of Mirocha (2014). In their work and the code ARES, a priori ionization history was assumed and used to generate tabulated

optical depth $\bar{\tau}$. In our implementation, we do not use the priori ionization history; instead, we solve equations (2) and (8) in turn. With the radiation intensity \hat{J}_ν obtained, we compute the ionization and heating coefficients by using equations (9)–(11) in Mirocha (2014).

In equations (4) and (6), the coefficients are related to the fraction of gas in collapsed haloes more massive than a redshift-dependent minimum mass $m_{\text{min}}(z)$

$$f_{\text{coll}} = \rho_m^{-1} \int_{m_{\text{min}}}^{\infty} mn(m) dm, \quad (9)$$

where ρ_m is the mean comoving density of the universe, $n(m)dm$ is the comoving number density of haloes with mass between m and $m + dm$. In Mirocha (2014), f_{coll} is computed with the HMF-CALC code (Murray, Power & Robotham 2013). In this work, we use an analytic equation

$$f_{\text{coll}} = \frac{1}{2} \left[1 + \text{erf} \left(\frac{z_0 - z}{z_w \sqrt{2}} \right) \right], \quad (10)$$

where the two parameters z_0 and z_w can vary to represent different evolving histories. So that

$$\frac{df_{\text{coll}}}{dt} = \frac{df_{\text{coll}}}{dz} \frac{dz}{dt}, \quad (11)$$

and

$$\frac{df_{\text{coll}}}{dz} = -\frac{1}{z_w \sqrt{2\pi}} e^{-\frac{(z-z_0)^2}{2z_w^2}}. \quad (12)$$

In order to check how well this formula can approximate the numerically computed $f_{\text{coll}}(z)$ with HMF-CALC, we perform a series of ordinary least-squares model fittings by using the above equation (10) to the data calculated with HMF-CALC of different T_{min} ’s. The model-fitting results are shown in Figs 1(a) and (b). The fitting results show that our analytic $f_{\text{coll}}(z)$ model is a good approximation to the data computed with HMF-CALC. We choose $z_0 = -0.64$ and $z_w = 6.4$ as the parameter of our reference model, which roughly corresponds to $T_{\text{min}} = 10^4 \text{ K}$.

If necessary, f_{coll} can be replaced by other analytic expressions in the future. This significantly enhances flexibility and computing performance in MCMC sampling.

On the other hand, the spin temperature T_S is determined as

$$T_S^{-1} \approx \frac{T_\gamma^{-1} + x_c T_K^{-1} + x_\alpha T_c^{-1}}{1 + x_c + x_\alpha}, \quad (13)$$

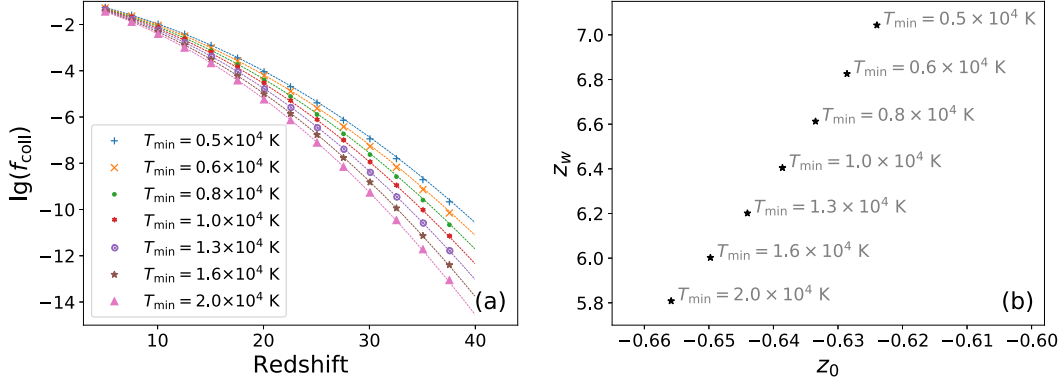


Figure 1. (a) The data points denote the $f_{\text{coll}}(z)$ data computed with HMF-CALC and the dashed lines denote $f_{\text{coll}}(z)$ values computed with equation (10) by using corresponding z_w 's and z_0 's obtained with model fittings. (b) Values of z_w 's and z_0 's corresponding to different T_{min} 's.

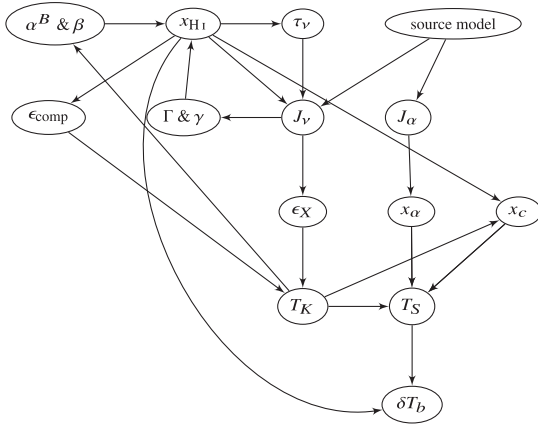


Figure 2. The dependency graph of the quantities needed to calculate the global H I 21 cm spectrum. Any one arrow points from one quantity to another one, the computing of which depends on the former one.

where the colour temperature of UV radiation field T_c is approximated to be $\simeq T_K$. The coupling coefficient x_c and x_α are calculated by using equations (24) and (40) of Furlanetto et al. (2006), respectively. The x_α depends on the background intensity of Ly α , which is calculated as

$$\hat{J}_\alpha = \frac{c}{4\pi} (1+z)^2 \sum_{n=2}^{n_{\text{max}}} f_{\text{rec}}^{(n)} \int_z^{z_{\text{max}}} \frac{\hat{\epsilon}_v(z')}{H(z')} dz', \quad (14)$$

and

$$\hat{\epsilon}_v = \bar{\rho}_b^0 C_{\text{LW}} f_* f_{\text{esc,LW}} \frac{df_{\text{coll}}}{dt} I_v, \quad (15)$$

where C_{LW} is fixed to be $4.72 \times 10^{27} \text{ g}^{-1}$, corresponding to 9690 photons per baryon particle, and $f_{\text{esc,LW}}$ is the corresponding escape fraction. Same as Mirocha (2014), n_{max} is truncated to 23, and the $f_{\text{rec}}^{(n)}$ is the ‘recycling fraction’ (Pritchard & Furlanetto 2006).

Finally, by assembling above computing together, we can obtain the global H I 21 cm signal of high redshifts. Looking back above still very incompletely summarized procedures, it is obvious that the rate equations (2) and (4), temperature evolving equation (3), and RTE (5) twist with each other. This can be shown a bit more clearly in Fig. 2. The dependence relations between different quantities form two loops, i.e. $x_{\text{HI}} \rightarrow \tau_v \rightarrow J_v \rightarrow \Gamma \& \gamma \rightarrow x_{\text{HI}}$ and $x_{\text{HI}} \rightarrow \epsilon_{\text{comp}} \rightarrow T_K \rightarrow \alpha^B \& \beta \rightarrow x_{\text{HI}}$. The most natural way is to solve them by turns from a high enough initial redshift to lower redshifts

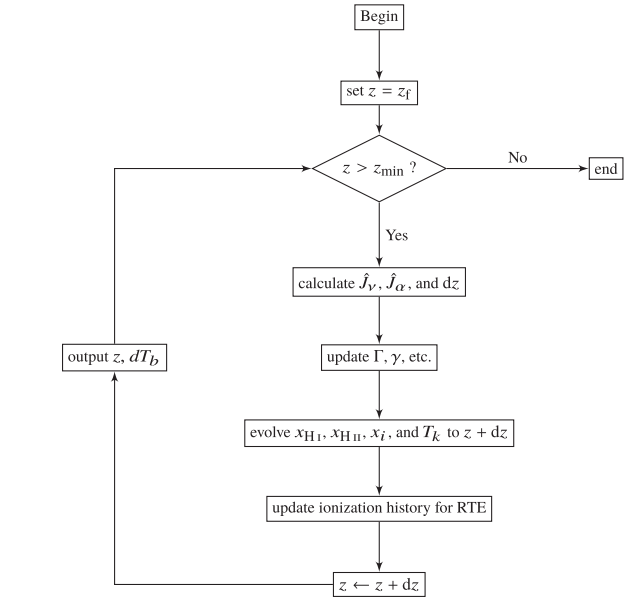


Figure 3. The flowchart of the algorithm for computing the global H I 21 cm spectrum.

Table 2. Reference Model Parameters.

Parameter	Value
f_*	0.05
f_{esc}	0.1
$f_{\text{esc,LW}}$	1.0
f_X	1.0
z_0	-0.64
z_w	6.4
N_{poly}	1
T_{100}	806 K
p_1	-2.27

iteratively. We summarize the algorithm of computing the global H I 21 cm signal in Fig. 3.

2.2 Reference model and frequency range to be used

We choose the parameter values listed in Table 2, and assume the ionizing sources to be black holes with a power-law spectrum to

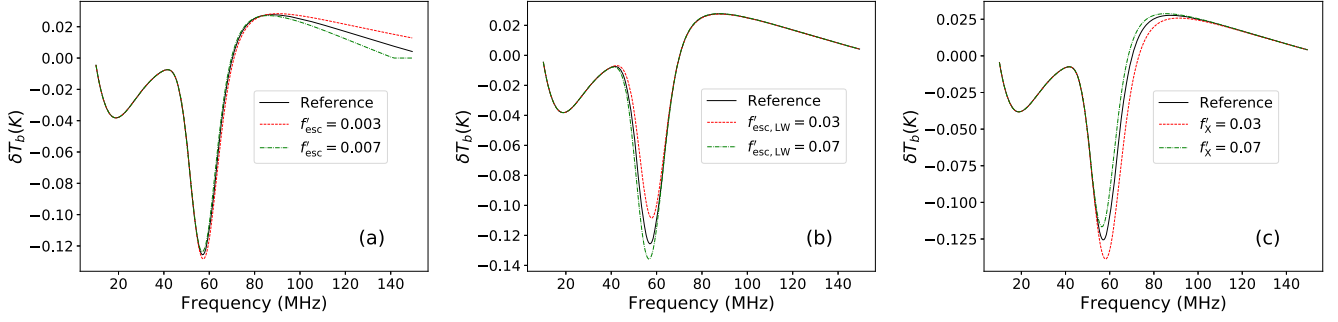


Figure 4. Brightness temperature contrast to CMB temperature calculated with the reference and varied model parameters. In (a)–(c), we show the influence of varying f'_{esc} , $f'_{\text{esc,LW}}$, and f'_X , respectively.

calculate the reference model values. The power-law index α is set to be -1.5 . Note that among f_* , f_{esc} , $f_{\text{esc,LW}}$, and f_X , only three can be determined independently, because f_* always appears together with one of the other three factors in above equations (4), (6), and (15). Consequently, it becomes an issue that needs to be discussed how to choose the free parameters to be inferred. The difference in selecting different sets of free parameters can be nontrivial, given that the corresponding priori distributions are not adjusted accordingly. We regard it more natural and intrinsic to use

$$f'_{\text{esc}} \equiv f_* f_{\text{esc}}, \quad (16)$$

$$f'_{\text{esc,LW}} \equiv f_* f_{\text{esc,LW}}, \quad (17)$$

and

$$f'_X \equiv f_* f_X \quad (18)$$

as the free parameters to infer. Note that there can be better choice when more priori knowledge is available. In Fig. 4, we compare $\delta T_b(\nu)$'s calculated with the reference model parameters and different f'_{esc} , $f'_{\text{esc,LW}}$, and f'_X values. It is not surprising that below 40 MHz, the signal changes little no matter how the parameters vary. At redshift higher than $z \approx 35$ (corresponding to 40 MHz), most of the current models predict that ionizing sources had not yet appeared, the brightness temperature of HI signal is mainly modulated by the adiabatic expansion progress of CMB radiation and gas. Because the ionosphere will significantly block electromagnetic wave with frequency $\lesssim 30$ MHz, ground-based experiments cannot arbitrarily extend the frequency range to lower frequencies. So it is reasonable to set the lower limit of the frequency range to be 40 MHz. On the other hand, as we describe above, HI brightness temperature model assumes that the IGM is divided into the 'bulk IGM' and H II regions, and this assumption can be violated in the late stage. As a result, we set the upper limit of the frequency to be 120 MHz, corresponding to $z = 10.8$. At this redshift, the reference model predicts an average HI fraction still higher than 50 per cent, which means it is not in the very late stage of reionization. As a conclusion, we will assume the data to be acquired within the frequency range of 40–120 MHz in most of the following analyses.

2.3 Foreground emission

Because the foreground is dominated by the synchrotron radiation from the Milky Way, we naturally model the foreground as a simple power-law spectrum when simulating the signals. Then in the stage of parameter inference, i.e. MCMC sampling, we treat

the foreground spectral model as a seventh-degree log-polynomial, to emulate any uncorrected instrumental effects on foreground spectrum, just like the treatment in Bernardi et al. (2016), though we do not think it can adequately compensate for the instrumental effects. We express the foreground spectrum as

$$T_{\text{fg}}(\nu) = T_{100} \exp \left[\sum_{n=1}^{N_{\text{poly}}} p_n \left(\log \frac{\nu}{100 \text{ MHz}} \right)^n \right]. \quad (19)$$

When simulating the foreground, we set the polynomial degree N_{poly} to be 1, and $p_1 = -2.27$. The brightness temperature at 100 MHz T_{100} is set to be 806 K (see Table 2), which is in agreement with the value used in Bernardi et al. (2016). In the parameter inference stage, we set N_{poly} to be 7, and allow all the coefficients in equation (19) to vary.

3 MCMC SAMPLING

3.1 Posterior probability distribution of model parameters

According to the Bayes' theorem, given a set of measurements that follow an underlying model, the posterior distribution of the model's parameters can be derived with

$$p(\theta | \mathbf{T}_b^{\text{obs}}) = p(\theta) p(\mathbf{T}_b^{\text{obs}} | \theta), \quad (20)$$

where $p(\theta)$ is the priori distribution of the model parameters, and $p(\mathbf{T}_b^{\text{obs}} | \theta)$ is the likelihood. We assume that there is no intrinsic fluctuation in the model; all deviation of the measurements from the model value is caused by thermal noise. In this condition, we can model the distribution of the noise value as normal distribution and safely regard the noise in different frequency channels to be uncorrelated. So that the likelihood can be written as

$$p(\mathbf{T}_b^{\text{obs}} | \theta) = \prod_i \frac{1}{\sqrt{2\pi\sigma_{v_i}^2}} e^{-\frac{(T_b^{\text{obs}}(v_i) - T_b^{\text{m}}(v_i))^2}{2\sigma_{v_i}^2}}, \quad (21)$$

where $\mathbf{T}_b^{\text{obs}}$ is a vector of measured antenna temperature in all used frequency channels, \mathbf{T}_b^{m} is the corresponding model value, T_b^{obs} and T_b^{m} are respectively the value for a certain frequency channel, and σ_{v_i} is the measurement uncertainty caused by thermal noise in the i th channel. According to the Gaussian white noise assumption, the noise is

$$\sigma_{v_i} = \frac{T_b^{\text{obs}}}{\sqrt{\Delta\nu\tau}}, \quad (22)$$

where $\Delta\nu$ is the channel width and τ is the integration time.

According to the foreground spectrum model equation (19), the antenna temperature at the lowest frequency 40 MHz is 6451 K. If we assume a channel bandwidth of 1 MHz³ and an integration time of 10 d, the noise level at 40 MHz is $\sigma_{40 \text{ MHz}} \approx 6.94 \text{ mK}$. Apparently, the S/N of global H I signal detections can reach a rather high value with very limited integration time.

The rest part of the posterior distributions is the priori distribution of unknown parameters. We assume parameters f'_{esc} , $f'_{\text{esc,LW}}$, and f'_X to independently follow a uniform distribution in the interval (0,100). Note that although these parameters have the meaning of ‘escape fraction’ and in principle should be limited between 0 and 1, we are using them to compensate for the uncertainties of corresponding physical parameters (i.e. N_{ion} , C_{LW} , and C_X), so that we do not impose the limit between 0 and 1.

The priori distribution of z_0 is

$$\frac{dp(z_0)}{dz_0} = \frac{1}{\sigma_{z_0}\sqrt{2\pi}} e^{-\frac{(z_0 - \hat{z}_0)^2}{2\sigma_{z_0}^2}}, \quad (23)$$

and the priori distribution of z_w is

$$\frac{dp(z_w)}{dz_w} = \frac{1}{\sigma_{z_w}\sqrt{2\pi}} e^{-\frac{(z_w - \hat{z}_w)^2}{2\sigma_{z_w}^2}}, \quad (24)$$

where the hyperparameters $\hat{z}_0 = -0.64$ and $\hat{z}_w = 6.4$. In order to set the values of hyperparameters σ_{z_0} and σ_{z_w} as a part of the priori knowledge, we examine Fig. 1(b). A reasonable range of z_0 corresponding to $T_{\text{min}} \in [0.5 \times 10^4, 2 \times 10^4]$ can be $z_0 \in [-0.66, -0.62]$, so that it is safe enough to set σ_{z_0} to be 0.2, which allows z_0 to vary within a slightly larger range. Similarly, z_w values corresponding $T_{\text{min}} \in [0.5 \times 10^4, 2 \times 10^4]$ can be safely enclosed within the interval [5.8, 7.1], so that we set σ_{z_w} to be 1. It is also possible that the $f_{\text{coll}}(z)$ calculated with other parameters or other codes can lead to larger reasonable ranges, but they will not differ by order of magnitudes. If necessary, in actual parameter inference, σ_{z_0} and σ_{z_w} can be replaced by other values.

3.2 Sampling algorithm

With the posterior distribution of the model parameters in hand, we then move on to sampling realizations from the posterior distribution. There have been a dozen of different sampling method available. We decided to test the affine invariant sampling method proposed by Goodman & Weare (2010) and its parallel tempering version proposed by Liang, Liu & Carroll (2011). The most important advantage of the affine invariant sampling method is that it does not need to be finely tuned to fit into a concrete problem, and is efficient enough to sample from very complex posterior distribution function. The computing of the global H I signal spectra in this work happens to be such kind of problem. Even after being reimplemented with RUST, the program still needs about 4 s for one CPU core to compute one spectrum. To reach a convergence, we set a ‘burn-in’ of 30 000 steps, the total length of the Markov Chain is set to be 60 000. In the beginning, we use the origin affine invariant sampling method and find that the chains are very hard to converge when the initial parameter is not close enough to the optimal value. So that we finally use the parallel tempering method to accelerate the convergence.

³Most systems have a much higher spectral resolution to excise radio frequency interference. Because the predicted global H I signal varies slowly with frequency, it is feasible to combine fine raw frequency channels to coarse channels to improve the computational performance.

We reimplement the PYTHON package EMCEE⁴ (Foreman-Mackey et al. 2013) that performs both the affine invariant sampling method and its parallel tempering version with RUST and run the sampling program on a high-performance computing cluster, utilizing parallel computing technique. For the parallel tempering method, we set six ‘temperature’s and for each temperature, 28 walkers (hence totally 168 chains) are used. The β values for the six temperatures are set to be 2^{-i} where $i = 0, 1, \dots, 5$. Totally 168 CPU cores are allocated for the MCMC program. Using more CPUs will not help reduce the time needed to compute one spectrum, but will allow us to set more chains, thus may help speed up the convergence of each chain.

3.3 The standard we use to diagnose if the inference is biased

Strictly speaking, in order to diagnose if the inference is biased, an analytical expression of the posterior distribution probability of interesting parameters is needed to deduce the expectations and compare them with input values. However, it is hard or even impossible in the condition of detection to global H I signal, because H I signal spectra are computed through massive numerical methods. We have to choose an alternative (not so strict) way – compare the input values with the mean sampled values and their dispersions. In practical parameter inference, when the distribution of sampled parameters is not heavily multimodal, we can express the inference result as a central value x together with its standard deviation Δx (for normal distribution, it is the 68 per cent confidence level error) in the form of $x \pm \Delta x$. If the interval $[x - \Delta x, x + \Delta x]$ encloses the true parameter values (i.e. input values here), we can practically regard the inference method to be good enough, so that practically unbiased. Similar tests can also be performed based on the plots of joint distributions of pairs of parameters. The above standard may not be appropriate, especially when one trying to examine it from a strict mathematics-based point of view. Some readers may not agree with our above standard, so in the following sections, we present objective results, including inferred parameter values and some of the plots of the parameter distributions we obtained, and readers can make their own decisions.

4 STUDIED CASES AND RESULTS

In this section, we study the influence of a variety of factors on the inference of model parameters. The factors include thermal noise, amplifier gain calibration error, and frequency-dependent antenna beam pattern. As is mentioned above, we choose f'_{esc} , $f'_{\text{esc,LW}}$, f'_X , z_0 , and z_w as our interesting parameters, which can reflect the detailed physical images of the reionization process. We choose 1 MHz spectral resolution, which is achievable with most practical data acquisition systems, and simultaneously sufficiently high for inferring most of the interesting parameters. In order to rule out the effect of the fluctuation from the random noise itself, we use the same random number generator seed to simulate the noise in the following sections.

4.1 The property of parameter space when ignoring foreground and noise

Before testing the conditions that emulate practical observations, we first inspect the property of parameter space with the foreground

⁴<https://github.com/dfm/emcee>

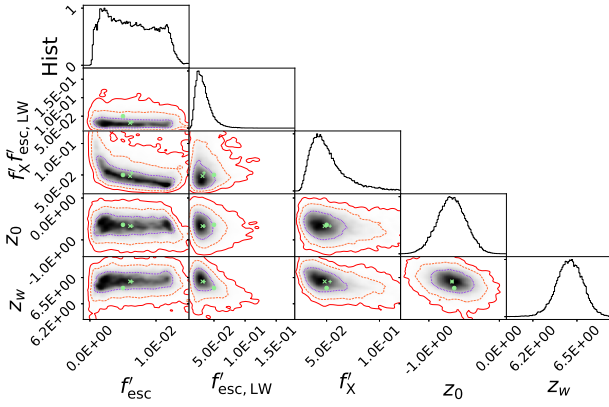


Figure 5. Probability distributions of single parameters (histograms, panels on the diagonal) and of joint parameters (68 per cent, 95 per cent, and 99.7 per cent probability contours, remaining panels) for the test case BL7 in Section 4.2 with the thermal noise calculated by assuming an integration time of 10 d, $\sigma_{z_0} = 0.2$, and $\sigma_{z_w} = 1.0$. The instruments are assumed to be perfectly calibrated. The ‘+’ symbols denote the estimation to the expected values by averaging the sample. The ‘x’ symbols denote the estimations by using median values. The dots denote the parameter values used to simulate the data.

component absent. The purpose is to discover potential intrinsic multimodal features and degeneracy in parameter space and separate them from the influence of the existence of foreground components. We simulate the total signal that is composed of only the global HI signal by using the parameters listed in Table 2 and thermal noise by using equation (22). When simulating the noise, the foreground signal is included in the T_b^{obs} of equation (22) temporarily and observation time of 10 d is assumed so that the realization of the noise component is identical to what will be used in the following sections. In the parameter inference stage, we test two models, one is the pure global HI signal model, and the other is the sum of the global HI signal model and a seventh-degree log-polynomial foreground component (equation 19). Because the nominal value of T_{100} is 0 K here, other foreground coefficients p_i ’s degenerates in the parameter space. If no constraint is applied to p_i ’s,

they can be an arbitrary value and cause the chain hard to converge. Hence, we assume all p_i ’s to independently follow uniform distributions between -1 and $+1$. The hyperparameters σ_{z_0} and σ_{z_w} here are both set to be 10 temporarily, which are fairly loose constraints, in following sections they will be set to $\sigma_{z_0} = 0.2$ and $\sigma_{z_w} = 1.0$. The sampling of both models reach equilibrium very fast, and the joint distribution of parameters show neither multimodal feature nor degeneracy. This phenomenon implies the feasibility of inferring parameters of global HI models with the MCMC sampling algorithm. The detailed inference result is unimportant here and cannot be compared to following testing conditions, so we omit them.

4.2 The influence of thermal noise

As has been mentioned in Section 3.1, an observation with very limited integration time can already obtain a high enough S/N. Nevertheless, it is still interesting to study how well the parameters can be constrained in an ideal observation condition (i.e. only inevitable thermal noise is considered). By using the above equation (22), we generate the noise signal for each frequency channel and add it to the total signal.

With the posterior distribution we set in the previous Section 3.1, we perform the sampling. We plot the parameter distribution in Fig. 5 (i.e. the ‘corner plot’, only the HI signal model parameters, foreground model parameters are not shown here). The numeric results are listed in Table 3, labelled as BL7 (‘BL’ means baseline, and the ‘7’ denotes the degree of the polynomial for the foreground component). By using the averaged values as an unbiased estimator, we find that the estimated values are all enclosed by 90 per cent probability contours. The parameter $f'_{\text{esc,LW}}$ is slightly underestimated, and 68 per cent probability contours cannot enclose the input value, so that seems to be biased. We also recover the global HI signal by using two methods: (1) directly computing the signal with sampled HI signal parameters, (2) subtracting the sampled foreground component (computed from the sampled foreground parameters) from the total signal. The comparison between the input and recovered signals are shown in Fig. 6, from which we find that the input signal does not pass through the centre of the

Table 3. Estimation to the model parameters obtained with the MCMC sampling method. Symbols \hat{f}_* , \hat{f}_{esc} , and $\hat{f}_{\text{esc,LW}}$ denote the parameter values used to simulate the signal. The signal is simulated with the reference model (Section 2.2), and the parameters are sampled assume the ionization sources to be black holes with power-law spectra.

Label	$\hat{f}'_{\text{esc}} / \hat{f}'_{\text{esc}}$	$\hat{f}'_{\text{esc,LW}} / \hat{f}'_{\text{esc,LW}}$	\hat{f}'_X / \hat{f}'_X	z_0	z_w	$\log Z$	Related figures	Description
Only thermal noise is considered, no other instrumental effect, see Sections 4.2 and 4.3								
BL7	1.27 ± 0.72	0.68 ± 0.25	1.07 ± 0.41	-0.67 ± 0.20	6.46 ± 0.08	402.09	5, 6	$N_{\text{poly}} = 7$ for foreground
BL7NN	1.18 ± 0.76	1.27 ± 0.54	1.23 ± 0.47	-0.67 ± 0.20	6.36 ± 0.08	437.08	7, 8	No noise; $N_{\text{poly}} = 7$ for foreground
BL3NN	1.01 ± 0.30	1.03 ± 0.12	1.05 ± 0.15	-0.67 ± 0.20	6.40 ± 0.05	439.64	9, 10	No noise; $N_{\text{poly}} = 3$ for foreground
BL4NN	1.13 ± 0.74	1.11 ± 0.20	1.23 ± 0.54	-0.68 ± 0.20	6.38 ± 0.07	439.55	11	No noise; $N_{\text{poly}} = 7$ for foreground
BL3	0.82 ± 0.28	0.97 ± 0.11	0.97 ± 0.11	-0.67 ± 0.20	6.44 ± 0.05	403.26		$N_{\text{poly}} = 3$ for foreground
Frontend gain calibration error considered, $N_{\text{poly}} = 7$ for foreground, see Section 4.4								
G7	1.21 ± 0.74	0.70 ± 0.26	1.10 ± 0.43	-0.68 ± 0.20	6.46 ± 0.08	402.08	12, 16	$n = 7$
G8	4.63 ± 0.74	30.75 ± 14.02	5.22 ± 1.43	-0.65 ± 0.24	5.73 ± 0.11	376.20	13, 16	$n = 8$
FDAP considered, in free space, $N_{\text{poly}} = 7$ for foreground, see Section 4.5.1								
L100	1.28 ± 0.73	0.73 ± 0.27	1.05 ± 0.33	-0.69 ± 0.20	6.46 ± 0.07	402.35	14, 15	$L = 100$ cm
L200	1.80 ± 0.88	2.08 ± 0.93	1.32 ± 0.41	-0.65 ± 0.20	6.26 ± 0.08	400.84	14, 15	$L = 200$ cm
FDAP considered, above the ground, $N_{\text{poly}} = 7$ for foreground; see Section 4.5.2								
L100GN	1.14 ± 0.74	0.52 ± 0.18	1.21 ± 0.60	-0.66 ± 0.20	6.48 ± 0.09	401.92	14, 15	$L = 100$ cm, $H = 50$ cm, latitude = $45^\circ N$
L200GN	0.04 ± 0.17	9.97 ± 2.34	10.34 ± 1.82	0.03 ± 0.37	5.89 ± 0.07		14, 15	$L = 200$ cm, $H = 100$ cm, latitude = $45^\circ N$
L100GS	2.10 ± 0.86	1.77 ± 0.86	1.17 ± 0.37	-0.68 ± 0.20	6.29 ± 0.09	400.25	14, 15	$L = 100$ cm, $H = 50$ cm, latitude = $45^\circ S$
L200GS	16.10 ± 1.82	0.08 ± 0.12	0.07 ± 0.002	0.36 ± 0.21	3.53 ± 0.03		14, 15	$L = 200$ cm, $H = 100$ cm, latitude = $45^\circ S$

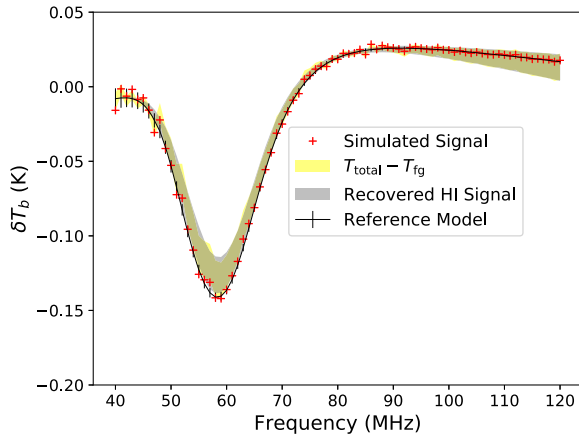


Figure 6. Comparisons between input H I 21 cm signal and the recovered signal for test case BL7 described in Section 4.2. Solid line denotes the input global H I signal, the error bars denote the noise level calculated with equation (22). Red crosses denote the sum of input signal and the actually simulated noise. The grey band denotes the 68 per cent confidence interval of the recovered H I signal. The yellow band denote the 68 per cent confidence interval of the difference between the total signal and the recovered foreground.

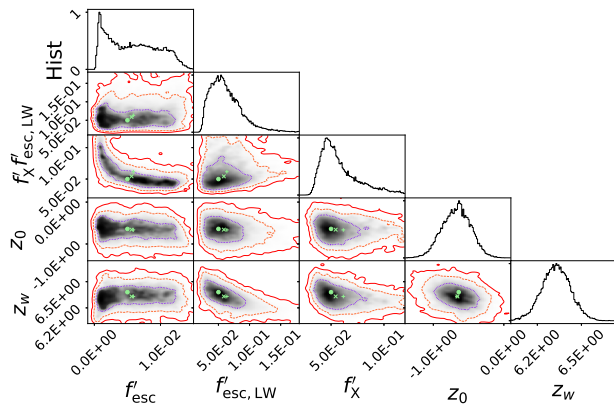


Figure 7. Same as Fig. 5, except that the noise component is removed from the total signal, corresponding to test case BL7NN.

bands that marks 68 per cent confidence range. This phenomenon also suggests that the inference seems to be biased. We suspect that this can be caused either by the thermal noise or by the overfitting problem. Also, we note there seems to exist a multimodal feature in the corner plot, especially in the histogram of the f'_{esc} . So that we further do two more tests: (1) removing the noise component from total signal (but still taking into account the noise when calculating the likelihood function, i.e. the parameters σ_{ν_i}); and (2) using lower degree polynomials to model the foreground in the sampling (the noise component is removed as well).

The distributions of parameters sampled with noise component removed (labelled as BL7NN) are shown in Fig. 7, and the numeric results are listed in Table 3. Obviously, the input parameter values can be all enclosed within 68 per cent contours. In Fig. 8, we note that the input signal right passes through the centre of the bands that marks 68 per cent confidence range of the recovered signals. Hence we can conclude that the underestimation of parameter $f'_{\text{esc,LW}}$ is caused by the certain realization of the noise component, and statistically speaking, the inference is unbiased. About the slightly

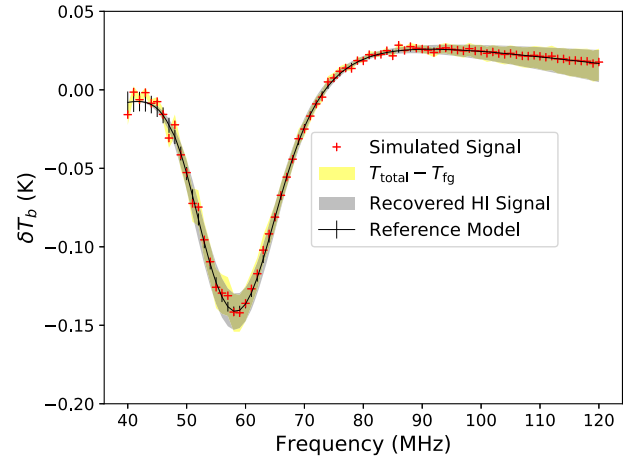


Figure 8. Same as Fig. 6, except that the noise component is removed from the total signal, corresponding to test case BL7NN.

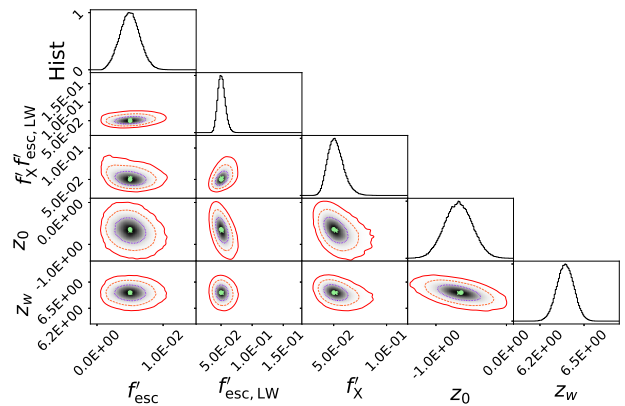


Figure 9. Same as Fig. 5, except that the noise component is removed from the total signal and a third-degree polynomial is used to model the foreground component, corresponding to test case BL3NN.

multimodal feature, we suspect it is caused by the foreground model with too many degrees of freedom. We leave the test of using lower degree polynomials to model the foreground in Section 4.3.

4.3 The influence of overfitting

It is meaningful to inspect if this choice of using seventh-degree polynomials to model the foreground component introduces overfitting problems so that we perform the sampling with different degrees of polynomials and check the distribution of sampled parameters.

We first simulate the total signal, including both the H I signal and the foreground emission by using the parameters listed in Table 2, but without the thermal noise component. Then, we run the sampling program and draw the parameter distributions.

We find that when using a third-degree polynomial to represent the foreground component (labelled as BL3NN), the distributions of parameters are close to the no-foreground condition test as performed in Section 4.1, i.e. close to a multivariate normal distribution, which can be seen in the plot of parameter distributions shown in Fig. 9. Fig. 10 shows that the H I signal can also be recovered unbiasedly. When the degree of the polynomial is increased to fourth (labelled as BL4NN), the distributions of parameters (Fig. 11)

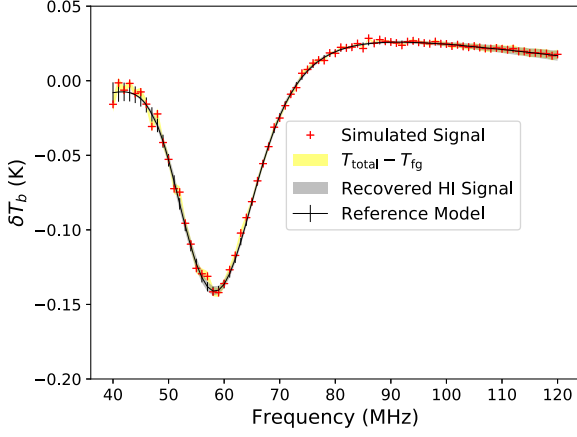


Figure 10. Same as Fig. 6, except that the noise component is removed from the total signal and a third-degree polynomial is used to model the foreground component, corresponding to test case BL3NN.

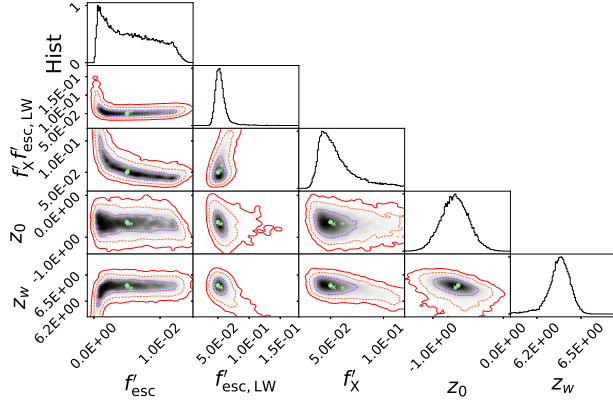


Figure 11. Same as Fig. 5, except that the noise component is removed from the total signal and a fourth-degree polynomial is used to model the foreground component, corresponding to test case BL4NN.

begin to deviate from multivariate normal-like distributions. The recovered HI signal resembles the result of test case BL7NN, which is still unbiased and but constrained better. So we do not show it here. We also test the case when noise is included, and $N_{\text{poly}} = 3$ (labelled as BL3), and find that both the distribution of parameters and the recovered HI signal are similar to the test case BL3NN.

So we may reach the conclusion that the overfitting problem does exist when the degree of freedom of the foreground model reaches a certain value ($N_{\text{poly}} = 4$ in our test cases). A foreground component model with too many degrees of freedom may cause the parameter inference too sensitive to thermal noise, and reducing the degree of freedom of foreground model can suppress the overfitting and the influence of thermal noise. However, in actual observation data reductions, it is infeasible to use a too simple foreground model. In the following sections, we will show that the foreground model with proper degrees of freedom is essential to compensate for instrumental gain calibration errors. As we will point out in the following Sections 4.4 and 4.5, a properly modelled foreground will not only help represent the foreground component itself but can also compensate for inaccurately calibrated amplifier gain and FDAPs. So we still use a $N_{\text{poly}} = 7$ polynomial to model the foreground in the following sections.

4.4 The influence of front end gain calibration error

Instrument calibration is crucial to precisely deduce the underlying physical parameters from the observed sky-averaged brightness temperature $T_b^{\text{obs}}(\nu)$. The brightness temperature of foreground emission is around 10^3 K level, while the interesting HI signal is around 10^{-1} K level. The deviation of *apparent* foreground spectra from a log-polynomial form can easily exceed 10^{-1} K level when the precision of amplifier gain flatness calibration is worse than 10^{-4} . However, there is still opportunity to recover HI signal if the foreground model (here we use the seventh-degree log-polynomial) can fit the spectral structure induced by the gain calibration error, given that no significant overfitting to the global HI signal happens. So in this section, we study how the parameter inference can be affected by instrument calibration errors. Here, we are not going to limit our work to any concrete instrument, but try to propose a universal method to model the frequency-dependent instrument gain calibration error.

Suppose after calibration, the actual frontend gain is

$$g(\nu) \equiv 1 + \Delta g(\nu), \quad (25)$$

and the observed total spectrum is

$$T_b^{\text{obs}} = T_b^{\text{true}} g(\nu) \quad (26)$$

$$= \delta T_b(\nu)(1 + \Delta g(\nu)) + T_{\text{fg}}(1 + \Delta g(\nu)). \quad (27)$$

We assume $\Delta g(\nu) \ll 1$, and $\delta T_b(\nu) \ll T_b^{\text{fg}}(\nu)$, so that $\delta T_b(\nu)\Delta g(\nu)$ can be ignored. As a result,

$$T_b^{\text{obs}} \approx \delta T_b(\nu) + T_{\text{fg}}(1 + \Delta g(\nu)). \quad (28)$$

We can define the gain error distorted foreground as

$$T'_{\text{fg}} \equiv T_{\text{fg}}(1 + \Delta g(\nu)). \quad (29)$$

We substitute equation (22) into above equation and obtain

$$T'_{\text{fg}} = T_{100} \exp \left[\sum_{n=1}^{N_{\text{poly}}} p_n \left(\log \frac{\nu}{100 \text{ MHz}} \right)^n \right] \exp[\log(1 + \Delta g(\nu))] \quad (30)$$

$$\approx T_{100} \exp \left[\sum_{n=1}^{N_{\text{poly}}} p_n \left(\log \frac{\nu}{100 \text{ MHz}} \right)^n \right] \exp(\Delta g(\nu)) \quad (31)$$

$$= T_{100} \exp \left[\sum_{n=1}^{N_{\text{poly}}} p_n \left(\log \frac{\nu}{100 \text{ MHz}} \right)^n + \Delta g(\nu) \right]. \quad (32)$$

Theoretically, if $\Delta g(\nu)$ can be expanded with a polynomial about $\log(\nu)$ with a degree not higher than N_{poly} , then the gain error distorted foreground T'_{fg} can be still fitted by the equation (19), otherwise, a larger N_{poly} is required. However, a too large N_{poly} may lead to the overfitting problem. According to our tests in Sections 4.2 and 4.3, an $N_{\text{poly}} = 7$ though higher than 1, which we use to simulate the foreground signal, will cause slight overfitting, but still acceptable. Using $N_{\text{poly}} > 7$ to model the foreground can cause more severe overfitting. Practical gain calibration error can have rather complicated spectral structures. For example, considering Chebyshev filter is used in the front-end amplifier module and this kind of filter has ripple-like frequency response. If a frequency-independent relative gain calibration error exists, the ripple-like feature will be kept after calibration. Such kinds of calibration error have complicated spectral structure and it is easy to verify that they cannot be well represented by a polynomial

about $\log(\nu)$ with a degree not higher than 7. So the question now is how precise the instrument needs to be calibrated in order to perform a meaningful detection to the global HI signal, and how serious the gain calibration error can interfere with the parameter inference.

A model is required to describe the gain calibration error as a function of frequency. The calibration error as a function of frequency can be expressed as the linear combination of a series of Legendre Polynomials as

$$\Delta g(\nu) = \sum_n^{\infty} a_n P_n(x) \quad (33)$$

where

$$x = \frac{2 \log(\nu/\nu_{\min})}{\log(\nu_{\max}/\nu_{\min})} - 1, \quad (34)$$

$$P_n(x) = \frac{1}{2^n} \sum_{k=0}^n \binom{n}{k} (x+1)^k \quad (35)$$

is the Legendre polynomial of order n , a_n is the corresponding combination coefficient. The most important reason why we do not simply use a general form of polynomial $\sum_{n=1}^N a_n x^n$ is that the Legendre polynomials are quasi-periodic on the definition domain, and the ‘period’ decreases as the order increases. This property makes the Legendre polynomials a suitable formula to express the spectral structures induced by gain errors at different complexity levels. If we want to find some actual examples with similar features from practical devices, one may recall the ripple structure in the frequency response of Chebyshev filters, if the frequency response were not properly calibrated such kind of quasi-periodic oscillating spectral structure will appear in the final data.

Two conditions with $a_n = 10^{-5} \delta_{nm}$ (δ is the Kronecker delta), where $m = 7$ (labelled as G7) or $m = 8$ (labelled as G8) are tested, respectively. For the G7 test case

$$\Delta g_7(\nu) = \frac{10^{-5}}{16} (429x^7 - 693x^5 + 315x^3 - 35x), \quad (36)$$

and for the G8 test case

$$\Delta g_8(\nu) = \frac{10^{-5}}{128} (6435x^8 - 12012x^6 + 6930x^4 - 1260x^2 + 35), \quad (37)$$

where x is defined as above equation (34). For both test cases, an $N_{\text{poly}} = 7$ polynomial is used to model the foreground.

About the determination of the value of a_n , we do not have any nominate parameter of actual amplifiers used in global HI signal detection experiments, and can only give an estimation based on the performance of commercial products. Our estimations are based on following facts: (1) typical gain error of commercial radio frequency amplifier chips (e.g. Analog Devices® AD8079) is around 10^{-3} ; (2) the tolerance of high accuracy resistors can reach 5×10^{-5} ; (3) the tolerance of high accuracy capacitor can reach 1 per cent; and (4) typical measurement error of general-purpose commercial measuring receivers (e.g. Agilent® 8902A, R&S FSMR series) is around 1 per cent. Considering there can be a significant potential of improving the measurement accuracy for specially designed receivers, we set a_n to be $10^{-5} \delta_{nm}$, which is higher than above commercial measuring receivers by about 3 orders of magnitude. We expect the $N_{\text{poly}} = 7$ foreground model can handle the G7 test case, while the G8 test case will significantly bias the results. The results are summarized in Table 3. Same as Section 4.2, we plot the recovered HI signals together with the input signals in Figs 12 and 13.

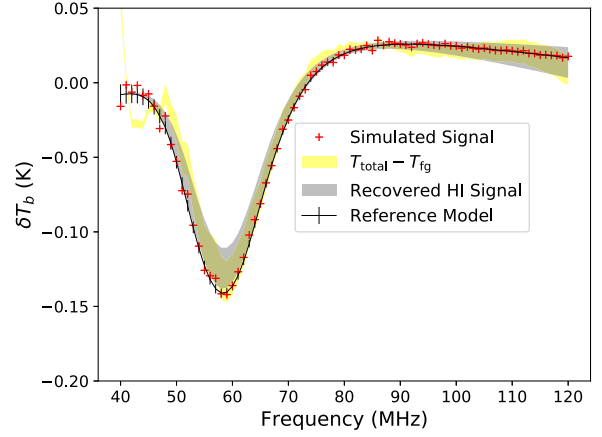


Figure 12. Same as Fig. 6, except that the noise component is removed from the total signal, corresponding to test case G7.

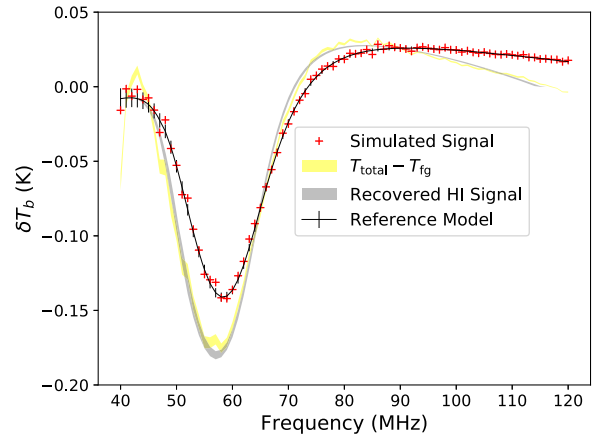


Figure 13. Same as Fig. 6, except that the noise component is removed from the total signal, corresponding to test case G8.

4.5 The influence of frequency-dependent antenna pattern

In most of the current work about detecting global averaged HI signal, it is assumed that a simple dipole-like antenna (not necessarily to be exactly dipole antenna) is used to receive the signal from very large sky area. According to e.g. Wilson, Rohlfs & Hüttemeister (2009), given the normalized power pattern of an antenna $P_n(\mathbf{n})$ and the sky brightness temperature distribution $T_b(\mathbf{n})$, both of which are the functions of direction \mathbf{n} , the instant antenna temperature

$$T_A(\nu) = \frac{\int T_b(\mathbf{n}, \nu) P_n(\mathbf{n}, \nu) d\Omega}{\int P_n(\mathbf{n}, \nu) d\Omega}, \quad (38)$$

which is actually the weighted mean value of the sky brightness temperature. When the weight changes, the output antenna temperature changes accordingly. The power patterns for most antennas are frequency-dependent so that it will create artificial spectral structures. When ground effect cannot be ignored, more significant frequency-dependent feature will appear. As a result, the effect of FDAP is potentially non-trivial. This issue has been pointed out by previous works (e.g. Liu et al. 2013), here we perform some numeric computations, in order to offer practical guides about the antenna design for such kinds of experiments. In more general conditions, when the orientation of the antenna changes with time,

mainly because of the diurnal motion, the time-averaged antenna temperature

$$\bar{T}_A(\nu) = \frac{1}{\tau} \int \frac{\int T_b(\mathbf{M}(t)\mathbf{n}, \nu) P_n(\mathbf{n}, \nu) d\Omega}{\int P_n(\mathbf{n}, \nu) d\Omega} dt \quad (39)$$

$$= \frac{\tau^{-1} \iint T_b(\mathbf{M}(t)\mathbf{n}, \nu) P_n(\mathbf{n}, \nu) d\Omega dt}{\int P_n(\mathbf{n}, \nu) d\Omega} \quad (40)$$

is actually used, where $\mathbf{M}(t)$ is the rotation matrix. When the antenna is fixed in space, \mathbf{M} is a constant matrix, so that equation (40) degenerates into equation (38).

Note that as the HI signal is regarded to be isotropic on large scales, the influence of FDAP to the global HI spectrum can be ignored, so we only need to calculate its influence to the foreground signal. Some all-sky survey map is required to evaluate equations (38) and (40). We use the map $T_{408}^{\text{Haslam}}(\mathbf{n})$ produced by Haslam et al. (1981) as a sky template. In order to ensure that in the condition of a perfect omnidirectional antenna, the calculated global averaged foreground spectrum degenerates to equation (19), we renormalize it to

$$T_{\text{fg}}(\mathbf{n}, 408 \text{ MHz}) = 4\pi T_{100} \left(\frac{408 \text{ MHz}}{100 \text{ MHz}} \right)^{p_1} \frac{T_{408}^{\text{Haslam}}(\mathbf{n})}{\int T_{408}^{\text{Haslam}}(\mathbf{n}) d\Omega}, \quad (41)$$

then

$$T_{\text{fg}}(\mathbf{n}, \nu) = 4\pi T_{100} \left(\frac{\nu}{100 \text{ MHz}} \right)^{p_1} \frac{T_{408}^{\text{Haslam}}(\mathbf{n})}{\int T_{408}^{\text{Haslam}}(\mathbf{n}) d\Omega}, \quad (42)$$

so that when the antenna beam is frequency-independent and perfectly isotropic, i.e. $P_n(\mathbf{n}, \nu) \equiv 1$, the sky-averaged foreground spectrum

$$T_{\text{fg}}(\nu) = \frac{\int T_{\text{fg}}(\mathbf{n}, \nu) d\Omega}{\int d\Omega} \equiv T_{100} \left(\frac{\nu}{100 \text{ MHz}} \right)^{p_1}, \quad (43)$$

which is actually equation (19) with $N_{\text{poly}} = 1$.

With equation (42) in hand, let us evaluate the above equations (38) for an antenna with fixed orientation in free space and (40) for an antenna on the ground. For an antenna with fixed orientation in free space,

$$T_A(\nu) = T_{100} \left(\frac{\nu}{100 \text{ MHz}} \right)^{p_1} 4\pi \frac{\int P_n T_{408}^{\text{Haslam}}(\mathbf{n}) d\Omega}{\int T_{408}^{\text{Haslam}}(\mathbf{n}) d\Omega \int P_n d\Omega} \quad (44)$$

$$= T_{100} \left(\frac{\nu}{100 \text{ MHz}} \right)^{p_1} \frac{\int P_n T_{408}^{\text{Haslam}}(\mathbf{n}) d\Omega}{\int P_n d\Omega} \times \left(\frac{\int T_{408}^{\text{Haslam}}(\mathbf{n}) d\Omega}{4\pi} \right)^{-1} \quad (45)$$

$$\equiv T_{100} \left(\frac{\nu}{100 \text{ MHz}} \right)^{p_1} g'_A(\nu), \quad (46)$$

where the equivalent gain $g'_A(\nu)$ is defined as

$$g'_A(\nu) \equiv \frac{\int P_n(\mathbf{n}, \nu) T_{408}^{\text{Haslam}}(\mathbf{n}) d\Omega}{\int P_n(\mathbf{n}, \nu) d\Omega} \left(\frac{\int T_{408}^{\text{Haslam}}(\mathbf{n}) d\Omega}{4\pi} \right)^{-1}. \quad (47)$$

For an antenna on the ground,

$$\bar{T}_A(\nu) = T_{100} \left(\frac{\nu}{100 \text{ MHz}} \right)^{p_1} \frac{\tau^{-1} \iint T_{408}^{\text{Haslam}}(\mathbf{M}(t)\mathbf{n}) P_n d\Omega dt}{(4\pi)^{-1} \int T_{408}^{\text{Haslam}}(\mathbf{n}) d\Omega \int P_n d\Omega} \quad (48)$$

$$= T_{100} \left(\frac{\nu}{100 \text{ MHz}} \right)^{p_1} \times \frac{\tau^{-1} \iint T_{408}^{\text{Haslam}}(\mathbf{M}(t)\mathbf{n}) P_n d\Omega dt}{\int P_n d\Omega} \times \left(\frac{\int T_{408}^{\text{Haslam}}(\mathbf{n}) d\Omega}{4\pi} \right)^{-1} \quad (49)$$

$$\equiv T_{100} \left(\frac{\nu}{100 \text{ MHz}} \right)^{p_1} g'_A(\nu), \quad (50)$$

where the equivalent gain $g'_A(\nu)$ is defined as

$$g'_A(\nu) \equiv \frac{\tau^{-1} \iint P_n(\mathbf{n}, \nu) T_{408}^{\text{Haslam}}(\mathbf{M}(t)\mathbf{n}) d\Omega dt}{\int P_n(\mathbf{n}, \nu) d\Omega} \times \left(\frac{\int T_{408}^{\text{Haslam}}(\mathbf{n}) d\Omega}{4\pi} \right)^{-1}. \quad (51)$$

As a result, the influence of FDAP to foreground spectra can be uniformly expressed as an equivalent gain and dealt within the mathematical framework that we described in Section 4.4. It is also notable that when calculating above equivalent gain, we do not need the sky template in every single frequency, and its normalization is not important, either.

Before applying the equivalent gain to the foreground spectra, we further renormalize it as

$$g_A(\nu) = \frac{g'_A(\nu)}{B^{-1} \int g'_A(\nu) d\nu}, \quad (52)$$

where B is the total bandwidth. The purpose of this step is to shift the mean equivalent gain over the whole bandpass to 1 so that the produced foreground spectra is comparable to our baseline test cases (e.g. BL7). In other words, although the normalization of the foreground spectrum can be affected by FDAP, here we decide to eliminate this factor to ensure that the differences in results are caused by FDAP itself purely, rather than the change of relative strength of foreground and HI signals. Following the above steps, the corresponding equivalent $g_A(\nu)$ is calculated for each test case and multiplied to the foreground spectrum.

4.5.1 A dipole in free space with fixed orientation

As a relatively simple beginning, we assume a dipole in free space (e.g. on a satellite) with the orientation fixed. The normalized power pattern of a dipole antenna is

$$P_n(\theta) = \frac{1}{P_{\text{max}}} \left[\frac{\cos(\frac{kL}{2} \cos \theta) - \cos(\frac{kL}{2})}{\sin \theta} \right]^2, \quad (53)$$

where the polar angle θ is 0 or π on the axis of the dipole, k is the wavenumber, and L is the length of the dipole. We substitute above equation into equation (38), then use equation (52) to calculate the equivalent gain and apply to the foreground signal spectra.

We study dipole antennas with $L = 100, 200,$ and 300 cm, respectively. We first plot the equivalent gain of dipole antennas with different length calculated with equation (52) in Fig. 14(a). Apparently, we can directly rule out the $L = 300$ cm dipole. For $L = 100$ and 200 cm dipole antennas, though the deviation of their equivalent gain from 1 has reached 10^{-3} to 10^{-2} level, it is possible that they can still be handled by the $N_{\text{poly}} = 7$ foreground model. So we perform following tests for these two dipole antennas.

We generate the new total sky-averaged signal by using the above equivalent gain caused by FDAP, which is then fed into the MCMC

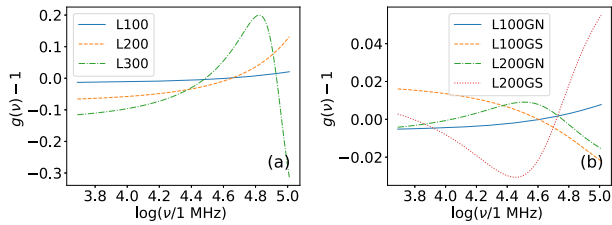


Figure 14. Equivalent gain of dipole antennas with different length fixed in (a) free space, and (b) fixed on the earth. The results are derived with equation (52).

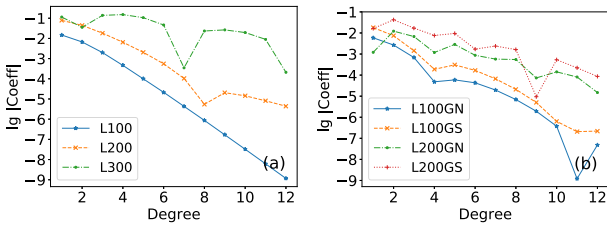


Figure 15. The common logarithm of the absolute values of the coefficients calculated by expanding the equivalent gain of dipole antennas with different lengths that are (a) fixed in free space or (b) fixed on the earth with Legendre Polynomials.

sampling program. We find that for both $L = 100$ and 200 cm conditions (denoted as L100 and L200, respectively), the MCMC can quickly reach a convergence. The inferred parameters are listed in Table 3. It is not surprising that the inferred parameters of the $L = 100$ cm condition are slightly closer to the input values than the $L = 200$ cm condition. As we stated above, by expressing FDAP into equivalent gain makes it easier to attribute the bias of result to the artificial spectral structures and compare it to the amplifier gain caused deviation that is described in the previous section. So we expand the equivalent gain with Legendre polynomials and inspect the amplitude of coefficients of different degrees. The plots of coefficients of different degrees are shown in Fig. 15(a). Obviously, the coefficients of L300 condition for degrees larger than 7 is significantly higher than that of L100 and L200. The coefficients of L200 is only slightly higher than that of L100. This agrees well with our conclusion above that the required order of polynomials to approximate the equivalent gain can strongly affect whether the parameter inference is biased.

4.5.2 A dipole antenna on the ground

Then let us consider another more practical condition – ground-based experiments. We assume a dipole antenna horizontally placed at some certain height above the ground. This brings two changes compared to an antenna in free space: (1) the earth will block the signal of the hemisphere below the horizon, and (2) ground effects involve. The first change can be handled rather straightforward, while for the ground effect, a ground model is needed. Here, we assume an infinitely large perfect conductive ground plane so that the antenna pattern can be calculated by simply assuming a mirror antenna. Because of the diurnal motion, the sky above the horizon changes with time, and equation (40) should be used. The effective area of the antenna is zero in directions below the horizon. Obviously, the received signal is related to the latitude of the point where the antenna is placed, so two representative latitudes are considered: $45^{\circ}00'00''N$ and $45^{\circ}00'00''S$. The orientation of the

dipole antennas are assumed to be along the meridian, and two lengths $L = 100$ and 200 cm are assumed. The height H is assumed to be $L/2$ correspondingly. The equivalent gains of the four conditions are calculated and shown in Fig. 14(b). We list the results in Table 3, labelled as L100GN, L200GN, L100GS, and L200GS, respectively.

From Fig. 14(b), we can conclude that the equivalent gain error of conditions L100GS and L200GS are larger than conditions L100GN and L200GN. On the other hand, we find that the result of L100GN is best among all the ground-based conditions, L100GS is worse, and L200GN and L200GS conditions give result significantly biased from the input values. This phenomenon implies that the latitude of the survey site can significantly affect the results. Then just like what we have done above, we expand the equivalent gain with Legendre polynomials and plot the coefficients verse the polynomial degree in Fig. 15(b). By checking the amplitudes of coefficients of degree higher than 7, we find the relative amplitudes of different conditions agree well with the qualities of parameter inference: better inference quality corresponds to smaller coefficients. This again confirms our above hypothesis that the spectral complexity is a more critical factor than the amplitude of gain calibration error that influences the quality of parameter inference.

5 DISCUSSION

5.1 Are directly inferred parameter unbiased?

As is mentioned in e.g. Harker et al. (2016), the two-stage fitting method can obtain biased results, and it would be interesting to see if biases in IGM properties would persist even if the signal were fit with the exact model used to generate it. According to our tests shown in Sections 4.2 and 4.3, the answer to this question is not as straightforward as it looks. As we have presented in Section 3.3, without the analytical expression of posterior distribution probability, it is hard to answer this question strictly. We have to make some compromises by comparing input parameters with the inferred confidence range (represented as standard deviations for single parameters and contours for joint distributions of pairs of parameters). In Section 4.2, we note that the parameter $f'_{\text{esc,LW}}$ seems to be underestimated when thermal noise is involved (test case BL7). Then by removing thermal noise from the total signal (but still keep σ_{v_i} 's when computing posterior probability) to form test case BL7NN, we find that the inference is actually unbiased, and the seemingly biased results are caused by certain noise realization. We also note that there are slightly multimodal features in the corner plots (Figs 5 and 7). From test cases BL3NN and BL4NN, we find that reducing N_{poly} from 4 to 3 will significantly change inferred parameter distributions from slightly multimodal to multivariate normal-like distribution. Furthermore, when $N_{\text{poly}} = 3$, even if noise is brought back to the total signal, the inferred parameters are still very close to input values, which means the degrees of freedom of foreground model will affect the sensitivity to thermal noise.

As a result, we can conclude that using seventh-degree polynomials to model foreground and directly inferring parameters with the exact model used to generate the signal can obtain unbiased results if only thermal noise is involved.

5.2 Factors that should be considered in practical global HI signal detection

5.2.1 Instrumental gain calibration

Through the tests described in Section 4.4, we have shown when the instrumental gain error as a function of frequency cannot be

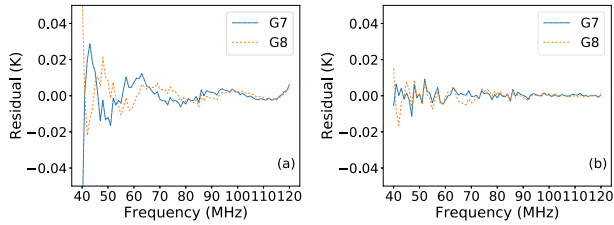


Figure 16. Residuals calculated as the total signal and the model predictions for test cases G7 and G8. In (a), parameters for simulating the signals are used, while in (b), optimal parameters (i.e. with the highest posterior probability) are used.

well approximated by a low-degree polynomial, very tiny gain calibration error (e.g. 10^{-5} as we have tested) can cause both the underlying physical model parameter and the recovered H I signal to significantly bias from the true values. Higher degree polynomials will be needed, but a too high polynomial degree can lead to the overfitting problem (see Section 4.3). The reason why this issue must be considered seriously is that we cannot simply treat the instrumental calibration error same as thermal noise, which is independent in each frequency channel and can be added to the σ^2 term in equation (21). Our result still cannot clarify the requirement of instrumental calibration precision, but we have shown that even at the level of 10^{-5} , the instrumental calibration error can be nontrivial.

We are curious about whether the bias is caused by the calibration error, or caused by the malfunction of the sampling method. In order to discriminate the above two possibilities, we plot the residual calculated with the parameters that are used to simulate the signal and the optimal parameters (i.e. parameters with the maximum posterior probability) in Figs 16(a) and (b), respectively. Apparently, for both G7 and G8 conditions, the residuals corresponding to optimal parameters are smaller than the residual corresponding to input parameters. Meanwhile, for the optimal parameters, the residual of G8 is larger than the G7 condition; this is in agreement with our above results that the G7 condition can lead to an acceptable result, while G8 condition cannot. Also, after examining the parameter distribution in parameter space and the plot of parameter values versus sample numbers, we conclude that for both G7 and G8 conditions, the Markov chains have reached convergences. So we could safely rule out the possibility of the malfunction of the sampling method.

Although we use only mathematics-based models to represent calibration problems, instead of actual front-end electronic devices (mainly a series of amplifiers, filters, and in some conditions, frequency mixers), we suggest that there are at least two potential sources could cause frequency-dependent gain calibration error. One is the broad-band amplifier gain flatness calibration error, and the other is the complexity of broad-band impedance matching between analogue devices. The latter factor has been addressed in e.g. Harker et al. (2012). We choose to not calculate the result with the equations in their work because besides the impedance mismatch, and there are other complex factors that will make the result hard to be represented in any analytical form.

5.2.2 Frequency-dependent antenna beam pattern

If in principle, analogue front-end gain can always be calibrated with sufficient accuracy, FDAP is inevitable. According to antenna theory, when the antenna has a size to wavelength ratio $L/\lambda \ll 1$, it would become the so-called electrically small antenna, the beam pattern of which is nearly omnidirectional. So it is plausible that

in order to suppress effects caused by the FDAP, smaller antennas are preferable. Besides the antenna size, we also show that the ground effect can cause even more severe FDAP (see Section 4.5.2). The reason is straightforward – the interval between the antenna and the earth can form a structure with a size comparable to the wavelength. This could be a big challenge that prevents ground-based instruments from detecting the global H I signal. In order to make ground-based instruments suitable for global H I signal detection, a smaller distance between the antenna and the ground plane can be a good choice. Actually among several currently running experiments that are devoted to the detection of global 21 cm signal, the antenna design of EDGES (Mozzden et al. 2016), PAPER (Pober et al. 2011), and BIGHORNS (Sokolowski et al. 2015) seems to fulfill the above requirements about the antenna size. It could be helpful for us to get more realistic results if we can take into account the actual mechanical models of the above antennas in our simulation.

There should be other ways to solve the FDAP problem. One method is to use digitized beamforming arrays because digitized beamforming arrays can adjust the complex gain (both amplitude and phase) of each antenna element to form the desired antenna beam that is independent of frequency. However, instrument calibration will become a more significant challenge for such kinds of complex systems. Another method is to scan as large as possible area of the sky so that the FDAP induced effects could be smeared out, either by putting the instrument in sites with different latitudes or for space-based detectors, pointing the antenna uniformly to different directions.

5.3 Which set of free parameters to be used?

As we have mentioned in Section 2.2, the difference between selecting different sets of free parameters can be nontrivial. In our above tests, we choose f'_{esc} , $f'_{\text{esc,LW}}$, and f'_X to be free parameters and assume their priori distributions to be a uniform distribution. On the other hand, there is another possible choice: directly sampling f_* , f_{esc} , and $f_{\text{esc,LW}}$ and fix $f_X = 1$, still assuming their priori distributions to be uniform, then the obtained corresponding distribution of $(f'_{\text{esc}}, f'_{\text{esc,LW}}, f'_X)$ will be definitely not same as the results we have presented above. The reason is obvious: if f_* and f_{esc} (or $f_{\text{esc,LW}}$) both independently follow the uniform distribution, their product f'_{esc} (or $f'_{\text{esc,LW}}$) will not follow the uniform distribution.

If we do not have enough priori knowledge, assuming the chosen parameters to follow uniform distributions independently seems to be a rather natural decision. However, this decision can cause a potential problem. Different choices of free parameters can lead to different estimation results. In this work, we are not able to answer the question of how to choose the free parameters.

5.4 Using priori knowledge from other observations

As as we have pointed out in Section 4.3, when $N_{\text{poly}} \geq 4$ is used in the foreground component model, the parameters cannot be tightly constrained. This phenomenon is most apparent for parameter f'_{esc} . From Figs 5, 7, and 11, it can be seen that parameter distributions deviate from normal-like distributions. And by examining the joint distributions of f'_{esc} and f'_X in the above figures, one will easily find a strong correlation between these two parameters. This phenomenon suggests that these two parameters can play some equivalent roles in forming the global H I spectrum, especially for the absorption trough feature. It could be possible to use the priori knowledge from other observations as further constraints to decouple these two

parameters. Recalling that f'_{esc} mainly affect the process of ionizing, the CMB optical depth

$$\tau_{\text{CMB}}(z) = n_{\text{H}}(0)c\sigma_{\text{T}} \int_z^{z_{\text{max}}} dz \bar{x}_i(z) \frac{(1+z)^2}{H(z)}, \quad (54)$$

where $n_{\text{H}}(0)$ is the number density of hydrogen atoms at $z = 0$, c is the speed of light, σ_{T} is the Thomson scattering cross-section, and $H(z)$ is the Hubble parameter at redshift z , can be used to impose further constraint on this parameter. This can be verified by examining corresponding τ_{CMB} values of different f'_{esc} . For example, the reference model value $f'_{\text{esc}} = 0.005$ corresponds to $\tau_{\text{CMB}}(10.8) = 0.012$, which does not conflict with the *Planck* result $\tau_{\text{CMB}}^{\text{Planck}}(0) = 0.054 \pm 0.007$ (Planck Collaboration et al. 2018). If we increase f'_{esc} by one order of magnitude to 0.05, then a $\tau_{\text{CMB}}(10.8) = 0.071 > \tau_{\text{CMB}}^{\text{Planck}}(0)$ is obtained, which can immediately be ruled out. Increasing f'_{esc} further by another one order of magnitude to 0.5 will result in a $\tau_{\text{CMB}}(10.8) = 0.133$. In this condition, the global H I signal reaches zero shortly after the absorption peak. Though we only use the CMB optical depth at $z = 0$ from the *Planck* data as an example here, a refined optical depth at higher redshifts can be calculated by properly modelling the electron number density at lower redshifts and subtracting their contribution to the total value, which can be used as a tighter constraint to the global H I signal model parameters.

5.5 The performance of simplified Gaussian model

As was introduced in Bernardi et al. (2016), a simplified model that is based on Gaussian function can be a possible method to extract information from global 21 cm signal from the cosmic dawn. We test the Gaussian model and compare the results with the previous physical model here.

We use the reference model that we have simulated above, with the parameters listed in Table 2. Same as Bernardi et al. (2016), the Gaussian model is expressed as

$$\delta T_{\text{b}}(\nu) = A_{\text{H I}} e^{-\frac{(\nu - \nu_{\text{H I}})^2}{2\sigma_{\text{H I}}^2}}, \quad (55)$$

where the amplitude $A_{\text{H I}}$, peak position $\nu_{\text{H I}}$, and standard deviation $\sigma_{\text{H I}}$ are to be inferred through the MCMC sampling. Foreground model is still the seventh-degree log-polynomial same as we used in previous sections. For the Gaussian model is only suitable to approximate the 21 cm signal from pre-reionization epoch, it is necessary to set an upper frequency limit ν_{max} . We set ν_{max} to be 100, 105, 110, and 120 MHz according to our knowledge about the simulated signal, and perform the test.

We show the histograms of $A_{\text{H I}}$ in Fig. 17. It can be seen clearly that the distribution of $A_{\text{H I}}$ varies significantly with ν_{max} . In other words, the upper frequency limit can heavily affect the level of the absorption feature of H I signal. Then we study the recovered H I signal, which is plotted in Fig. 18. Obviously, if correctly choosing the upper frequency limit, we can recover the H I signal with a relatively small error. We show the residual between the total signal and the sum of H I model component and the foreground model component in Fig. 19, to check whether we can diagnose the results in actual observations. From the figure, we find that within the used frequency ranges, the residuals distribute almost symmetrically, and do not show obvious difference between test cases with different ν_{max} . The residuals of all four ν_{max} values also show oscillating features.

It would be interesting to check the residual between the recovered foreground component and the input model, which we show in Fig. 20. Obviously for $\nu_{\text{max}} = 100$ and 120 biased foreground

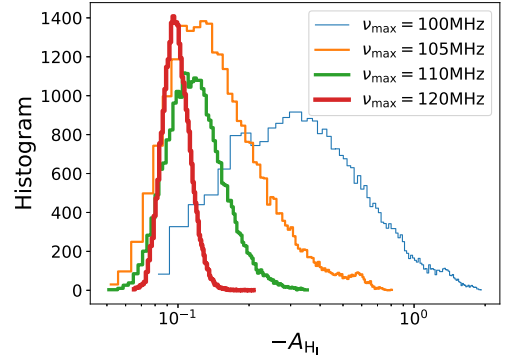


Figure 17. The histograms of $A_{\text{H I}}$ in the Gaussian model sampled assuming $\nu_{\text{max}} = 100, 105, 110,$ and 120 MHz, respectively.

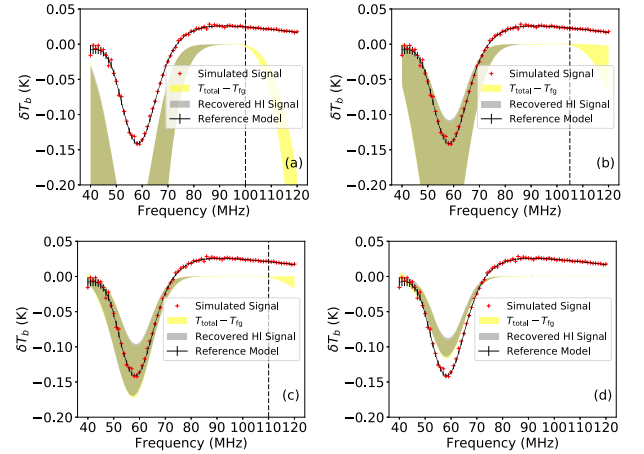


Figure 18. Recovered H I signals by using Gaussian model assuming $\nu_{\text{max}} = 100, 105, 110,$ and 120 MHz, respectively. The vertical dashed lines denote the ν_{max} . Grey bands are the 68 percent confidence ranges of the signals calculated with equation (55) by using the sampled model parameters. Yellow bands are the signals calculated as the differences between total signal and the foreground model values. The black solid lines denote the input global H I signal model. The red crosses denote the sum of the input signal and the actually simulated noise.

estimations are obtained. The most significant deviation from the reference value appears right at the frequency of absorption feature of the H I signal, which happens to compensate for the biased H I signal estimation and lead to the symmetric residual distributions as are shown in Fig. 19. Hence, although both the recovered H I and foreground components are biased, the total recovered model component show no apparent biasing evidence, except for some oscillation features, and the oscillation can also be caused by the instrument calibration inaccuracy. So we have to conclude that although modelling the H I signal in the early stage of EoR can significantly simplify the parameter inference and quickly obtain the results, it is still risky and mostly suitable only for qualitative analyses.

5.6 A caveat about this work

In the previous sections, we do confirm that with a direct modelling method and proper sampling algorithm, it is possible to obtain unbiased parameter estimation from the global H I signal. This also partially answers the question raised by Harker et al. (2016) that whether the inference result would be biased if the signal were fitted

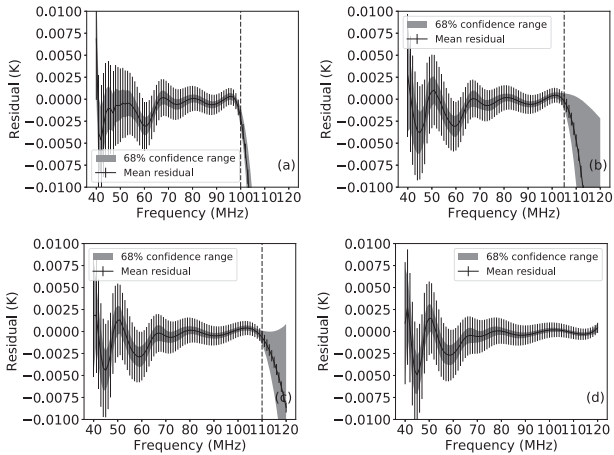


Figure 19. The distributions of the residuals between the total signal and the model predicted values assuming $\nu_{\max} = 100, 105, 110,$ and 120 MHz, respectively. The vertical dashed lines denote the ν_{\max} . The H I signal is approximated with a Gaussian model. The residuals defined as $T_{\text{total}} - T_{\text{fg}} - \delta T_{\text{b}}$, where T_{total} is the total signal and δT_{b} is calculated with equation (55). The solid lines are the mean residual calculated over all the sampled parameters. The error bars represent the standard deviation of noise. The grey bands denote the 68 per cent confidence ranges of the residuals.

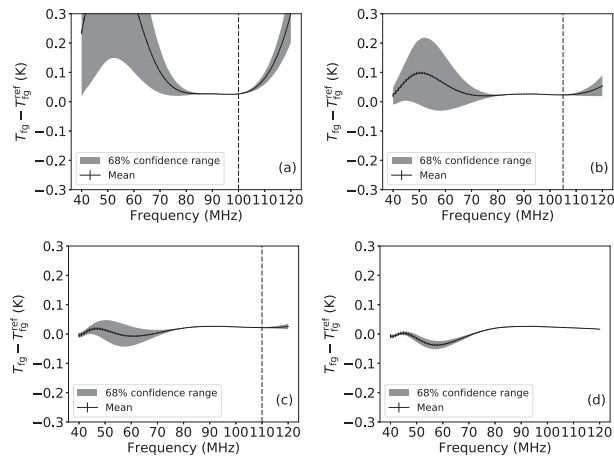


Figure 20. The errors of estimations to the foreground component in the conditions of $\nu_{\max} = 100, 105, 110,$ and 120 MHz, respectively. The vertical dashed lines denote the ν_{\max} . The solid lines represent the mean residual and the grey bands denote the 68 per cent confidence range.

with the exact the model used to generate it. However, we shall point out that we cannot test all possible reionization models and there may still exist some reionization models that could not well constrained by the observations to global H I signals, and imaging detection or power spectra measuring are required to constrain the parameters of those models. The result of this work should be treated as a test to some specific reionization scenarios. Nevertheless, the method used in this paper is still useful to test if a direct one-stage MCMC sampling works in the condition of other models that are not covered in this work.

5.7 Some practical suggestions of performing MCMC sampling

Last but not least, we would like to give some practical suggestions for MCMC sampling. It is true that in an ideal condition when

the parameter space is not multimodal, any casually chosen initial parameter can finally lead to convergence. However, our experience shows that in practical conditions, choosing a ‘good’ initial parameter is still essential. We find that when instrumental gain calibration error or FDAP involves, an arbitrarily chosen initial parameter may require a rather large number of sampling before convergence. A practical method is to find the foreground polynomial coefficients through an ordinary least square fitting with polynomial fitting to the total signal and use the fitting result as the foreground parameters’ initial values.

6 CONCLUSIONS

We survey the feasibility of directly inferring physical model parameters from global EoR signal data and the influences of some factors that should be considered in actual observations to the model parameter inference. We find that if only thermal noise is considered, unbiased model parameters can be inferred. The inaccuracy of analogue front-end and frequency-dependent antenna beam can bias the results on different levels. Even a relative amplifier gain calibration error of 10^{-5} can significantly influence the inferred model parameters given that a low-degree polynomial cannot well fitted the calibration error as a function of frequency. The FDAP should be carefully considered to prevent it from creating artificial spectral structures. Geographical position can also cause differences in the difficulties of obtaining correct model parameters. The Northern hemisphere seems to be a better choice than the Southern hemisphere because of the distribution of the brightness of the Milkyway. Finally, we give some practical suggestions about performing MCMC sampling and parameter inference in global EoR signal detections.

ACKNOWLEDGEMENTS

We are sincerely grateful to the referee for constructive and valuable suggestions and comments, which help us improve the manuscript.

We thank the Square Kilometre Array Regional Center prototype developed by Shanghai Astronomical Observatory for providing us with computing resources. We acknowledge the support of Astronomical Information Technology Group and Public Technology Service Center, National Astronomical Observatories, Chinese Academy of Sciences (NAOC); this work was supported by High Performance Computing Cluster of NAOC.

This work was supported by the National Key R&D Programme of China (grant no. 2018YFA0404600), the Strategic Priority Research Program of Chinese Academy of Sciences (grant no. XDB23010300), and the Key Projects of Frontier Science of Chinese Academy of Sciences (grant no. QYZDY-SSW-SLH022).

REFERENCES

- Bernardi G. et al., 2016, *MNRAS*, 461, 2847
- Carilli C. L., Rawlings S., 2004, *New Astron. Rev.*, 48, 979
- DeBoer D. R. et al., 2017, *PASP*, 129, 045001
- Field G. B., 1958, *Proc. IRE*, 46, 240
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Fukugita M., Kawasaki M., 1994, *MNRAS*, 269, 563
- Furlanetto S. R., Oh S. P., Briggs F. H., 2006, *Phys. Rep.*, 433, 181
- Goodman J., Weare J., 2010, *Commun. Appl. Math. Comput. Sci.*, 5, 65
- Harker G. J. A., Mirocha J., Burns J. O., Pritchard J. R., 2016, *MNRAS*, 455, 3829

- Harker G. J. A., Pritchard J. R., Burns J. O., Bowman J. D., 2012, *MNRAS*, 419, 1070
- Haslam C. G. T., Klein U., Salter C. J., Stoffel H., Wilson W. E., Cleary M. N., Cooke D. J., Thomasson P., 1981, *A&A*, 100, 209
- Huang Y., Wu X.-P., Zheng Q., Gu J.-H., Xu H., 2016, *Res. Astron. Astrophys.*, 16, 36
- Liang F., Liu C., Carroll R., 2011, *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples*, Vol. 714. John Wiley & Sons, New York
- Liu A., Pritchard J. R., Tegmark M., Loeb A., 2013, *Phys. Rev. D*, 87, 043002
- Mirocha J., 2014, *MNRAS*, 443, 1211
- 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
- Mirocha J., Mebane R. H., Furlanetto S. R., Singal K., Trinh D., 2018, *MNRAS*, 478, 5591
- Mirocha J., Skory S., Burns J. O., Wise J. H., 2012, *ApJ*, 756, 94
- Mozdzen T. J., Bowman J. D., Monsalve R. A., Rogers A. E. E., 2016, *MNRAS*, 455, 3890
- Murray S. G., Power C., Robotham A. S. G., 2013, *Astron. Comput.*, 3, 23
- Park J., Mesinger A., Greig B., Gillet N., 2019, *MNRAS*, 484, 933
- Parsons A. R. et al., 2010, *AJ*, 139, 1468
- Patra N., Subrahmanyan R., Raghunathan A., Udaya Shankar N., 2013, *Exp. Astron.*, 36, 319
- Planck Collaboration et al., 2016, *A&A*, 594, A13
- Planck Collaboration et al., 2018, preprint ([arXiv:1807.06209](https://arxiv.org/abs/1807.06209))
- Pober J. et al., 2011, *American Astronomical Society Meeting Abstracts*, 217, p. 432.06
- Pritchard J. R., Furlanetto S. R., 2006, *MNRAS*, 367, 1057
- Seager S., Sasselov D. D., Scott D., 2000, *ApJS*, 128, 407
- Sokolowski M. et al., 2015, *PASA*, 32, e004
- Spinelli M., Bernardi G., Santos M. G., 2019, *MNRAS*, 489, 4007
- Taylor G. B. et al., 2012, *J. Astron. Instrum.*, 1, 1250004
- Tingay S. J. et al., 2013, *PASA*, 30, e007
- van Haarlem M. P. et al., 2013, *A&A*, 556, A2
- Voytek T. C., Natarajan A., Jáuregui García J. M., Peterson J. B., López-Cruz O., 2014, *ApJ*, 782, L9
- Wang J. et al., 2013, *ApJ*, 763, 90
- Wilson T. L., Rohlf K., Hüttemeister S., 2009, *Tools of Radio Astronomy*. Springer-Verlag, Berlin
- Wouthuysen S. A., 1952, *AJ*, 57, 31
- Zheng Q., Wu X.-P., Gu J.-H., Wang J., Xu H., 2012, *ApJ*, 758, L24
- Zheng Q., Wu X.-P., Johnston-Hollitt M., Gu J.-h., Xu H., 2016, *ApJ*, 832, 190

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