LOCO EDGES REPORT #122

ARIZONA STATE UNIVERSITY
TEMPE, ARIZONA 85287

Foreground model selection for signal parameter estimation

Judd Bowman
July 25, 2018

REVISED August 1, 2018
REVISED March 9, 2020 – Clarify LINLOG definition (Equation 4)

1. Introduction

The goal of this memo is to evaluate the selection of foreground models to use during cosmological parameter estimation when analyzing EDGES data. We evaluate several models using simulated spectra. We assess the ability of the models to fit the simulated spectra and we measure the bias in recovered cosmological signal properties.

1.1. Astronomical foreground spectrum

The intrinsic astronomical foreground along a line of sight consists primarily of synchrotron and free-free emission (and some absorption depending on frequency) from our Milky Way Galaxy and any extragalactic sources along the line of sight. This is dominated by the $\beta \approx -2.5$ power law spectral index of the synchrotron emission of the Galaxy. Simulations of the foreground signal, such as those by Jelic et al. 2009 and Bernardi et al. 2015, yield small amounts of additional structure beyond a pure power-law. This structure varies slightly from one line of sight to another due to variations in density, composition, temperature, background radiation, and magnetic fields within the Galaxy. Because empirical knowledge of these properties is insufficient to create a foreground model sufficiently accurate to use with redshifted 21cm observations, the foreground signal is typically represented as an expansion around the -2.5 power law at each line of sight. In theoretical analyses, it is common for the expansion to be given as:

$$\ln T_f(\upsilon) = \sum_{n=0}^{N-1} a_n \left(\ln \frac{\upsilon}{\upsilon_0} \right)^n \qquad \text{LOGLOG (1)}$$

Bernardi et al. 2015 found that this expansion typically needs between N=3 to N=7 terms to account for the full spectral structure in their simulations. In general, the coefficients of the power law terms are not known for the actual sky. Oliveira de Costa et al. 2008 used observed sky maps at several frequencies to estimate that the typical spectral index $\beta = a_1 \approx -2.5$ and found evidence for an all-sky average higher order curvature of $a_2 \approx -0.1$, although with significant uncertainty. Equation 1 is equivalent to:

$$T_f(\upsilon) = e^{\sum_{n=0}^{N-1} a_n \left(\ln\frac{\upsilon}{\upsilon_0}\right)^n} \quad (2)$$

With some rearranging, this reduces to:

$$\begin{aligned} T_f(\upsilon) &= e^{a_0} e^{\sum_{n=1}^{N-1} a_n \left(\ln\frac{\upsilon}{\upsilon_0}\right)^n} \\ &= e^{a_0} e^{\ln\left(\frac{\upsilon}{\upsilon_0}\right) \sum_{n=1}^{N-1} a_n \left(\ln\frac{\upsilon}{\upsilon_0}\right)^{n-1}} \\ &= a'_0 \left(\frac{\upsilon}{\upsilon_0}\right)^{\sum_{n=1}^{N-1} a_n \left(\ln\frac{\upsilon}{\upsilon_0}\right)^{n-1}} \end{aligned} \quad \text{EXPLOG (3)}$$

using $a'_0 \equiv e^{a_0}$ and $e^{\ln\left(\frac{\upsilon}{\upsilon_0}\right)} = \frac{\upsilon}{\upsilon_0}$. Alternatively, the above expressions can be recast through Taylor series expansion to a system of equations that is linear in temperature (as opposed to linear in *ln T* as in Equation 1):

$$T_f(\upsilon) \approx \left(\frac{\upsilon}{\upsilon_0}\right)^{\beta} \sum_{m=0}^{M-1} a_m \left(\ln\frac{\upsilon}{\upsilon_0}\right)^m, \quad \text{excluding m=1} \quad \text{LINLOG (4)}$$

where we set $\beta = a_1$ from Equation 3 so that its term in Equation 4 captures the n=1 contribution from Equation 3 and we perform the expansion of Equation 3 on the remaining higher-order terms (n>1). The summation in Equation 4 should exclude the m=1 term to formally match the Taylor series expansion of Equation 3. By pulling the n=1 (m=1) term out in front with β, we center the expansion on the expected power law spectral index, β, which can be treated as a fixed parameter if desired, allowing the other terms of the expansion to pick up any deviations from the expectation. The number of terms needed in the Equation 4 summation may be different than in the previous expressions in order to achieve a desired accuracy.

1.2. Ionosphere

Neglecting refraction, the ionosphere operates on the instantaneous intrinsic astronomical signal along a line of sight with an absorption and emission contribution (see Rogers et al. 2015) given by:

$$T'_f(\upsilon) = T_f(\upsilon) e^{-\tau(\upsilon)} + T_e\left[1 - e^{-\tau(\upsilon)}\right] \quad (5)$$

Where $\tau(\upsilon) \sim \upsilon^{-2}$ is the optical depth of the ionosphere and T_e is the temperature of electrons in the ionosphere. During nighttime, $T_e \approx 800$ K (rising to 1600 K during the day) and $\tau(\upsilon) \approx 0.005 \left(\upsilon/_{75\ \mathrm{MHz}}\right)^{-2}$ (rising by an order of magnitude during the day). For EDGES low-band observations (50-100 MHz), the absorption of the incoming astronomical signal is about 0.3-1% from high to low frequencies in the band, while the electron temperature emission contribution is about 3 to 9 K.

With $\tau << 1$ for the nighttime ionosphere in the band of interest, Equation 5 can be approximated as:

$$T'_f(\upsilon) \approx T_f(\upsilon)\left[1 - \tau(\upsilon)\right] + T_e\tau(\upsilon) \quad (6)$$

Since the foreground spectrum, $T_f(\upsilon)$, is dominated by the $\upsilon^{-2.5}$ synchrotron emission and the ionosphere optical depth is proportional to υ^{-2}, we see that the ionosphere primarily adds new spectral structure to the intrinsic foreground spectrum with $\upsilon^{-4.5}$ and υ^{-2} frequency dependence.

1.3. Models for fitting data

Because there is generally insufficient knowledge of the true foreground spectral and angular structure on the sky to subtract it directly from global 21cm observations to the needed accuracy, when analyzing EDGES observations a foreground model is fit to the data simultaneously with a 21cm signal model. There are several foreground models that may be used, including:

- EXPLOG: The model typically used by theorists in any of the forms shown in Equations 1-3. It is linear in *ln T* space, but non-linear in T:

$$T_f(\upsilon) = a_0 \left(\frac{\upsilon}{\upsilon_0}\right)^{\sum_{n=1}^{N-1} a_n \ln\left(\frac{\upsilon}{\upsilon_0}\right)^{n-1}}$$

  where the coefficients a_n are found by fitting the model to data.
- LINPHYS: A linearized approximation to the EXPLOG model and ionspheric contributions. It can capture up to third-order structure of the intrinsic astronomical foreground and first-order structure introduced by the ionosphere:

$$T_F(\upsilon) \approx a_0\ (\upsilon/\upsilon_0)^{\beta} + a_1\ (\upsilon/\upsilon_0)^{\beta} \log(\upsilon/\upsilon_0) + a_2\ (\upsilon/\upsilon_0)^{\beta} [\log(\upsilon/\upsilon_0)]^2 + a_3\ (\upsilon/\upsilon_0)^{\beta-2} + a_4\ (\upsilon/\upsilon_0)^{-2} \quad (7)$$

  In our typical use of this equation, the astronomical spectrum is assumed to have a dominant power law spectral index of β=-2.5.
- LINLOG: Approximation to the EXPLOG model to make it linear in *T* as given by Equation 4:

$$T_f(\upsilon) = \left(\frac{\upsilon}{\upsilon_0}\right)^{\beta} \sum_{n=0}^{N-1} a_n \left(\ln \frac{\upsilon}{\upsilon_0}\right)^n$$

  where N is usually chosen to be five in our analysis and β is typically fixed to -2.5.
- LINPOLY: A linear polynomial with a spectral index offset (also called the EDGES polynomial):

$$T_f(\upsilon) = \left(\frac{\upsilon}{\upsilon_0}\right)^{\beta} \sum_{n=0}^{N-1} a_n \left(\frac{\upsilon}{\upsilon_0}\right)^n \quad (8)$$

  where N is usually chosen to be five in our analysis and β is typically fixed to -2.5. When β=0, this reduces to a common polynomial expansion.

When using any of the models above, υ_0 is chosen to be the middle of the observed band (75 MHz for low-band data; 150 for high-band data).

Because ionopsheric effects primarily introduce spectral structures generally similar in form to the astrophysical foreground power laws in the models above (see Equation 6 and discussion immediately following), any of the above models, when fit to an observed foreground spectrum as seen from the ground, may be able to absorb some of the ionospheric contributions, even if the model does not explicitly include ionospheric physics. Alternatively, any of the models (other than LINPHYS) could be used as the intrinsic astrophysical foreground model in conjunction with the ionosphere physics model of Equation 5 or 6. In the latter case, we can expect that there will be significant covariance between the intrinsic foreground model parameters and the ionosphere physics parameters due to their similar spectral shapes. For this reason, here we investigate only cases where the foreground

models above are used to account for all foreground and ionosphere structure without introducing separate parameters directly for the ionosphere contributions.

We note that the EXPLOG and derived models originated to account for structure in the foreground spectrum along individual lines of sight, whereas EDGES observations integrate over a large solid area on the sky. Because the foreground spectral structure is reasonably similar across lines of sight in simulations, we expect the models will function well when applied to spectra averaged over large areas. Similarly, the ionospheric physics model is accurate for instantaneous observations (or for non-time varying ionospheric conditions) along a single line of sight, whereas EDGES spectra are integrated over time and solid angle. A similar argument applies here and we expect the ionosphere physics model (when applied) will generally be able to capture the structure in EDGES observations. However, we note that direction-dependent effects of the ionosphere due to refraction (e.g. Vedantham et al. 2014) are excluded here and warrant further investigation.

2. Evaluation of models using simulations

We begin by simulating possible foreground spectra observed by EDGES at high Galactic latitudes using Equation 5 to model the ionosphere contribution and Equation 3 (EXPLOG) with N=5 to model the intrinsic foreground spectrum (T_F) upon which the ionosphere acts. Our aim is model the performance of the foreground models under expected sky conditions. We generate artificial ionosphere-corrupted spectra (T'_F) over a grid of realistic input parameters that span a probable range of foreground and ionosphere properties:

a_0 = [1500, 1750, 2000]
a_1 = [-2.6, -2.55, -2.5, -2.45, -2.4]
a_2 = [-0.1, -0.05, 0, 0.05, 0.1]
a_3 = [-0.1, -0.05, 0, 0.05, 0.1]
a_4 = [-0.1, -0.05, 0, 0.05, 0.1]
τ = [0, 0.005, 0.05]
T_e = [0, 500, 750, 1000]

This yields 22,500 simulated spectra. Figure 1 shows a sample of the foreground spectra from this set. We use the same frequency channel bin size of 0.391 MHz as used in analysis of Bowman et al. 2018.

As an initial assessment of the performance of the other foreground models, we fit the LINPHYS, LINLOG, and LINPOLY models to the simulated spectra using υ_0=75 MHz and N=5 in all models and β=-2.5. We perform the fits over two frequency ranges: 51-99 MHz and 63-99 MHz, which were typical choices in the data analysis performed in Bowman et al. 2018. The residual RMS for each model and each frequency case is plotted in Figure 2. Because all three of the test models have the same number of degrees of freedom in this analysis, comparison of the RMS within a given frequency range provides a direct comparison of the ability of the model to fit the simulated spectra over the frequency range. LINLOG performs the best, as might be expected since it is directly derived from the EXPLOG model used to create the intrinsic foreground spectra in the simulations, even though it contains no terms to directly capture the ionosphere physics. LINPHYS is worse that LINLOG by about a factor of three. LINPOLY performs relatively poorly by more than an order of magnitude compared to LINLOG and LINPHYS. In general, residuals increased by an order of magnitude for each model going from 63-99 MHz to 51-99 MHz.

Next, we repeat the fits for LINLOG and LINPOLY test models with varying number of terms, N = {3,4,5,6,7}. Table 1 summarizes the median residual RMS across all simulated spectra for each test case. We see that LINPOLY performs significantly worse than LINLOG, regardless of N.

Table 1. Median of residual RMS in Kelvin across all 22500 trials for the specified model and number of terms

	LINLOG		LINPHYS		LINPOLY	
N	**51-99**	**63-99**	**51-99**	**63-99**	**51-99**	**63-99**
3	0.699166	0.158705	--	--	1.321293	0.249821
4	0.102653	0.017158	--	--	0.289572	0.035771
5	0.001949	0.000204	0.006209	0.000629	0.058986	0.005127
6	0.000241	0.000017	--	--	0.011409	0.000684
7	0.000025	0.000001	--	--	0.002077	0.000088

We can estimate the goodness of fit of each test model using the Bayesian information criterion (BIC):

$$\mathrm{BIC} = n\ \ln\left(\frac{RSS}{n}\right) + k\ \ln(n)$$

Where n is the number data points (123 for 51-99 MHz and 92 for 63-99 MHz), RSS is the residual sum squares to a model fit, and k is the number of parameters in the model (equal to N in our models). We note that RSS/n is equal to the square of the residual RMS for a given model fit. We use the median residual RMS in our calculations to represent the ensemble of possible foregrounds simulated.

The BIC is useful because it is an absolute quantity that enables direct comparisons between different models, numbers of data points, and numbers of parameters. It is typically used in model selection. A model with a lower BIC should be selected over a model with a higher BIC.

Table 2 summarizes the BIC values derived for our models. Overall, the LINLOG model yields lower BIC. Increasing the number of terms in any model lowers the BIC, suggesting in this simple scenario that selecting higher N is better. For a given N, the smaller frequency range is preferred for LINLOG for N<=5, but we see relatively little difference between the 51-99 and 63-99 frequency ranges for the LINLOG model above N>5, with a slight preference for the larger frequency range. For LINPOLY on the other hand, the 63-99 MHz frequency range consistently outperforms the wider range for a given N.

Table 2. Bayesian information criterion derived from values in Table 1.

	LINLOG		LINPHYS		LINPOLY	
N	**51-99**	**63-99**	**51-99**	**63-99**	**51-99**	**63-99**
3	-32	-141	--	--	36	-105
4	-235	-317	--	--	-124	-258
5	-656	-669	-532	-579	-292	-412
6	-877	-866	--	--	-465	-571
7	-1117	-1090	--	--	-645	-732

In practice, residual RMS for a model fit to actual data will be limited by thermal noise. We can estimate the effects of thermal noise in the BIC calculation by adding 25 mK RMS in quadrature with each of the median residual RMS values in Table 1 and recalculating the BIC. This is equivalent to having 25 mK thermal noise per channel in the data, which matches the thermal noise in the Bowman et al. 2018 observations. We show the resulting BIC values in Table 3. For the LINLOG case in both frequency ranges, the intrinsic model performance produces residuals well below the thermal noise by N=5. Further, the BIC value yields a minimum at N=5, indicating that N=5 should be selected if using this model on the EDGES data over either frequency range. For the LINPOLY model, the BIC does not reach a clear minimum for the larger frequency range in our trials, but it does start to level off for N=6 and at N=7 it matches the BIC of the LINLOG model, which is rising with increasing N compared to the LINLOG minimum at N=5 due to the increasing number of parameters, indicating that N=7 will be the LINPOLY minimum, as well. For the smaller bandwidth, the BIC effectively saturates at N=5 or N=6. With the inclusion of the thermal noise, LINPHYS performs essentially as well as LINLOG at N=5.

Table 3. Same as Table 2, but calculated after adding 25 mK thermal noise per channel in quadrature with the residual RMS in Table 1.

	LINLOG		LINPHYS		LINPOLY	
N	**51-99**	**63-99**	**51-99**	**63-99**	**51-99**	**63-99**
3	-32	-140	--	--	36	-105
4	-232	-272	--	--	-124	-242
5	-383	-285	-380	-285	-283	-283
6	-382	-283	--	--	-371	-283
7	-379	-281	--	--	-379	-281

Lastly, we can estimate the BIC for a combined foreground and signal model fit. If we assume that the flattened Gaussian absorption profile model is able to describe the 21cm signal perfectly and is being fit simultaneously with the foreground model, we can calculate BIC values for the combined model by adding the four additional parameters to the BIC calculation (setting k = N + 4) and using again the median residual RMS values from Table 1 along with the 25 mK thermal noise additional RMS added in quadrature. Table 4 shows the resulting BIC values for this representation of the full model. Here we see that both LINLOG and LINPHYS continue to be the best selections for N=5. LINPOLY is a valid selection for the smaller bandwidth at N=5, but would need N=7 to be a competitive selection for the larger bandwidth and even then fails to reach the best-case of -375 with LINLOG at N=5.

Table 4. Same as Table 3, but setting k=N+4 to include approximate the effect of a perfect fit of the flattened Gaussian absorption profile model on the BIC.

	LINLOG		LINPHYS		LINPOLY	
N+4=k	**51-99**	**63-99**	**51-99**	**63-99**	**51-99**	**63-99**
3+4=7	-24	-132	--	--	44	-97
4+4=8	-223	-264	--	--	-115	-235
5+4=9	-375	-277	-372	-277	-275	-275
6+4=10	-373	-275	--	--	-363	-275
7+4=11	-371	-273	--	--	-371	-273

Based on the above analysis, we find that for analyzing the larger bandwidth, the best model selection is LINLOG using N=5, with LINPHYS nearly equally valid. When analyzing the smaller bandwidth, any of the three models at N=5 should be essentially comparable selections. For (nearly) equal BIC, the model with fewer terms is preferred and we find that LINPOLY should not be used on the larger bandwidth.

2.1. Parameter estimation with simulated spectra

Next, we assess the ability of the models to recover a simulated 21cm signal. We add a simulated 21cm absorption profile to the foreground/ionosphere spectra from the previous section. For the simulated 21cm absorption profile, we use the flattened Gaussian profile model described in Bowman et al. 2018 with A=-0.5, w=20 MHz, v_c=80 MHz, and τ_{21}=7. To reduce the computational time, we use only a subset of 648 of the original 22,500 foreground/ionosphere trials, equivalent to a grid at the following foreground/ionosphere model coordinates:

a_0 = [1500, 1750]
a_1 = [-2.6, -2.5, -2.4]
a_2 = [-0.1, 0, 0.1]
a_3 = [-0.1, 0, 0.1]
a_4 = [-0.1, 0, 0.1]
τ = [0, 0.005]
T_e = [0, 1000]

For each simulated spectrum, we perform a grid search over absorption profile parameters to recover the best-fit profile. We search over:

v_0 = [60 to 90 in steps of 1 MHz]
w = [1 to 40 in steps of 1 MHz]
τ_{21} = [1 to 10 in steps of 1]
A = [-2 to 2 in steps of 0.05 K]

At each point in the grid search, we subtract the absorption model specified by the current location in the grid from the simulated spectrum, then perform a least-squares fit of a foreground model (using N=5) to that spectrum and record the residual RMS. The grid point with the lowest residual RMS yields the best-fit recovered signal profile. We repeat this procedure for each of the LINLOG, LINPHYS, and LINPOLY foreground models and the two frequency ranges used earlier.

Figures 3-6 show histograms of the best-fit values recovered for the four signal parameters. In general, all models with N=5 for both frequency ranges recover a feature centered at the input 80 MHz and with width close to the input 20 MHz. However, the recovery of tau (flattening) and the profile amplitude are more complicated. LINLOG and LINPHYS generally recover the input absorption profile reliably with the correct amplitude and tau, but LINPOLY performs much worse. It fails to recover tau over either frequency range. LINPOLY fails to recover the input amplitude when using the larger frequency range. Over the smaller frequency range it recovers the amplitude less robustly (larger spread) than the other models and with a slight bias to weaker amplitudes.

Finally, we add thermal noise to the simulated spectra with 25 mK RMS per 0.391 MHz channel to match the EDGES data in Bowman et al. 2018. We repeat the signal recovery grid search and show the results of the best-fit amplitude recovery in Figure 7. The outcome is generally unchanged from Figure 6. LINLOG and LINPHYS continue to recover the input signal amplitude well and LINPOLY continues to fail over the larger frequency range.

In Figure 8 we investigate whether LINLOG with N=4 is a viable model selection for the smaller 63-99 MHz frequency range. Table 3 indicates the BIC for N=4 model is -275, compared to -285 at N=5. However, Figure 8 shows that the recovered best-fit amplitude shows more wider scatter around the true value for LINLOG at N=4 than at N=5 in noise-less simulations, suggesting this is not a good choice.

In Figure 9 we investigate whether LINPOLY with N>5 is a viable model selection for the larger 51-99 MHz frequency range. Table 2 s the BIC for LINPOLY continues to improve with increasing N from -268 to -371 to -379 for N=5, 6, 7, respectively. Figure 9 shows that the recovered best-fit amplitude still shows considerable scatter around the true value for LINPOLY at N=6, but converges more closely for N=7. Nevertheless, the scatter at N=7 is still larger than for LINLOG at N=5 and comparable to LINPHYS at N=5

3. Conclusions

We conclude from the analysis presented above that the LINPOLY foreground model with N=5 and β=-2.5 should not be used over the 51-99 MHz range. It is not selected over the LINLOG and LINPHYS models in a Bayesian information criterion analysis. It fails to recover the amplitude of a simulated absorption profile similar to that reported in Bowman et al. 2018. LINLOG emerges as the best model investigated, warranting further analysis and preferred use during data anlysis, although LINPHYS performed nearly as well.

Further work is needed to:

- Validate performance of models found through simulation with actual data
- Investigate models constructed from SVD-derived basis sets.
- Investigate models that combine intrinsic foreground and ionosphere parameters.
- Investigate robustness of models to systematic effects (e.g. a sinusoid in the spectrum)

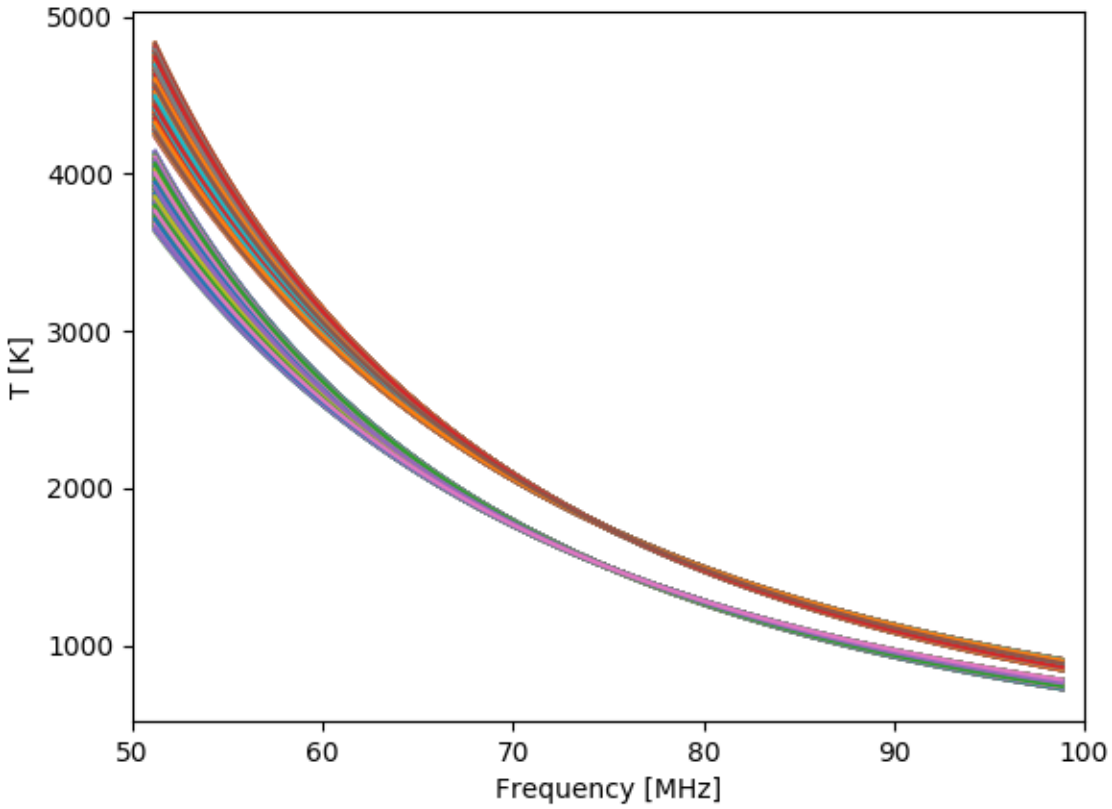


Figure 1. *The 648 simulated foreground spectra used in the signal grid search analysis. These are a subset of the 22,500 models used to generate Tables 1-4, but they span the input parameter range to the same extremes as the larger set for most parameters.*

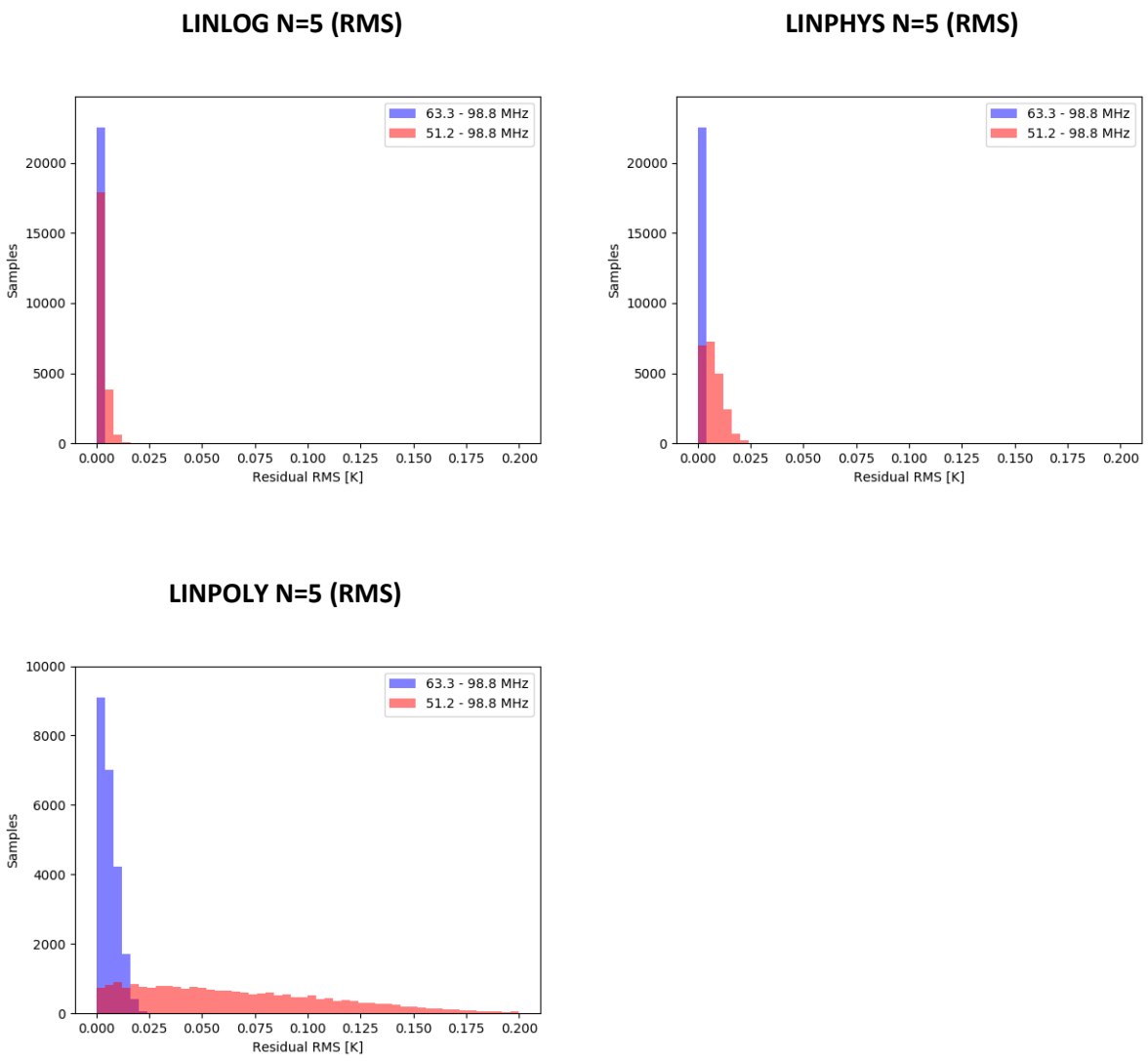
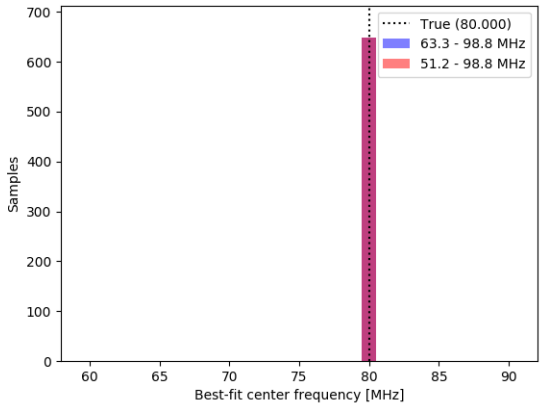
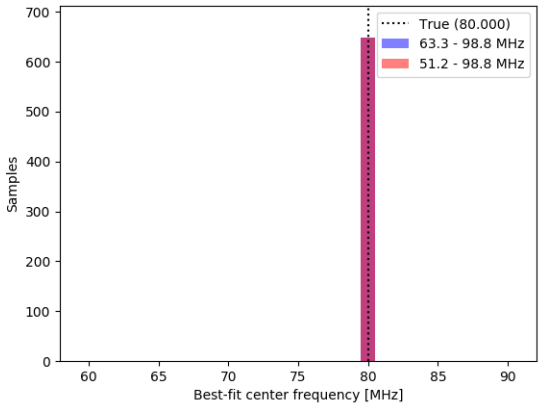


Figure 2. *Histograms of residual RMS for LINLOG (top left), LINPHYS (top right), and LINPOLY (bottom left) fit over two frequency ranges to the 22,500 simulated foreground/ionosphere models using N=5 and v_0=75 MHz and β=-2.5. LINLOG performs best, as might be expected since it is most closely derived from the EXPLOG model used in part to create the simulated spectra. LINPHYS performs nearly as well while LINPOLY performs much worse (note the change in horizontal scale by nearly an order of magnitude) by nearly an order of magnitude. Despite not explicitly including parameters for the ionospheric physics, LINLOG is able to capture the structure in the simulated spectra at <10 mK in nearly all cases and <1 mK in the narrower band cases.*

LINLOG N=5 (v_0)



LINPHYS N=5 (v_0)



LINPOLY N=5 (v_0)

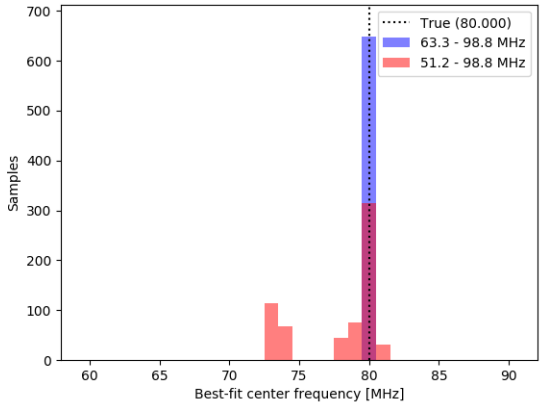
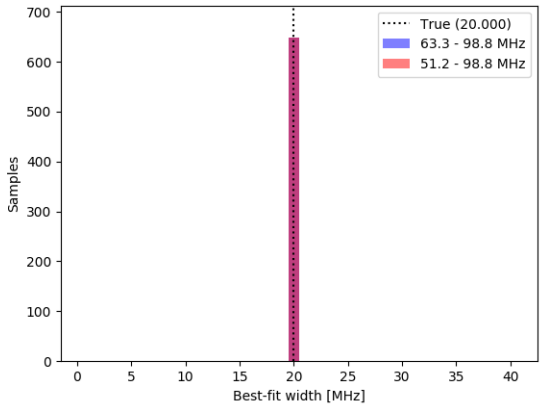
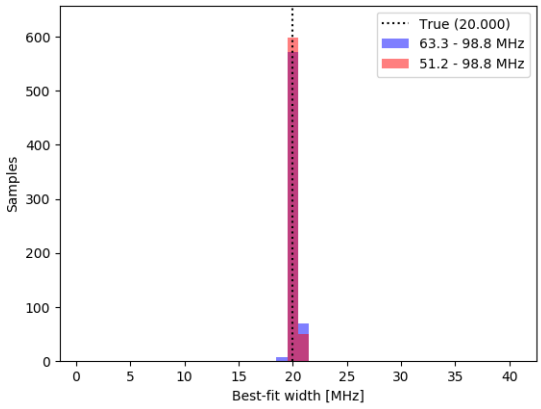


Figure 3. *Histograms of best-fit profile center frequency recovered by each model with N=5 from noiseless simulations. The input profile center frequency was 80 MHz. All three models recover the profile width well, although LINPOLY is somewhat less robust compared to the perfect recovered values of LINLOG and LINPHYS.*

LINLOG N=5 (w)



LINPHYS N=5 (w)



LINPOLY N=5 (w)

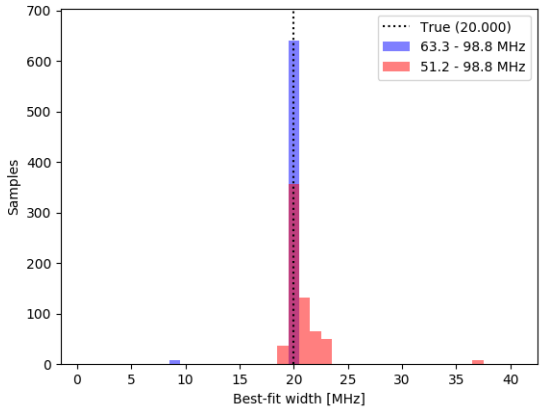


Figure 4. *Histograms of best-fit profile width recovered by each model with N=5 from noiseless simulations. The input profile width was 20 MHz. All three models recover the profile width well, although LINPOLY does have a minor tendency to find a narrower feature.*

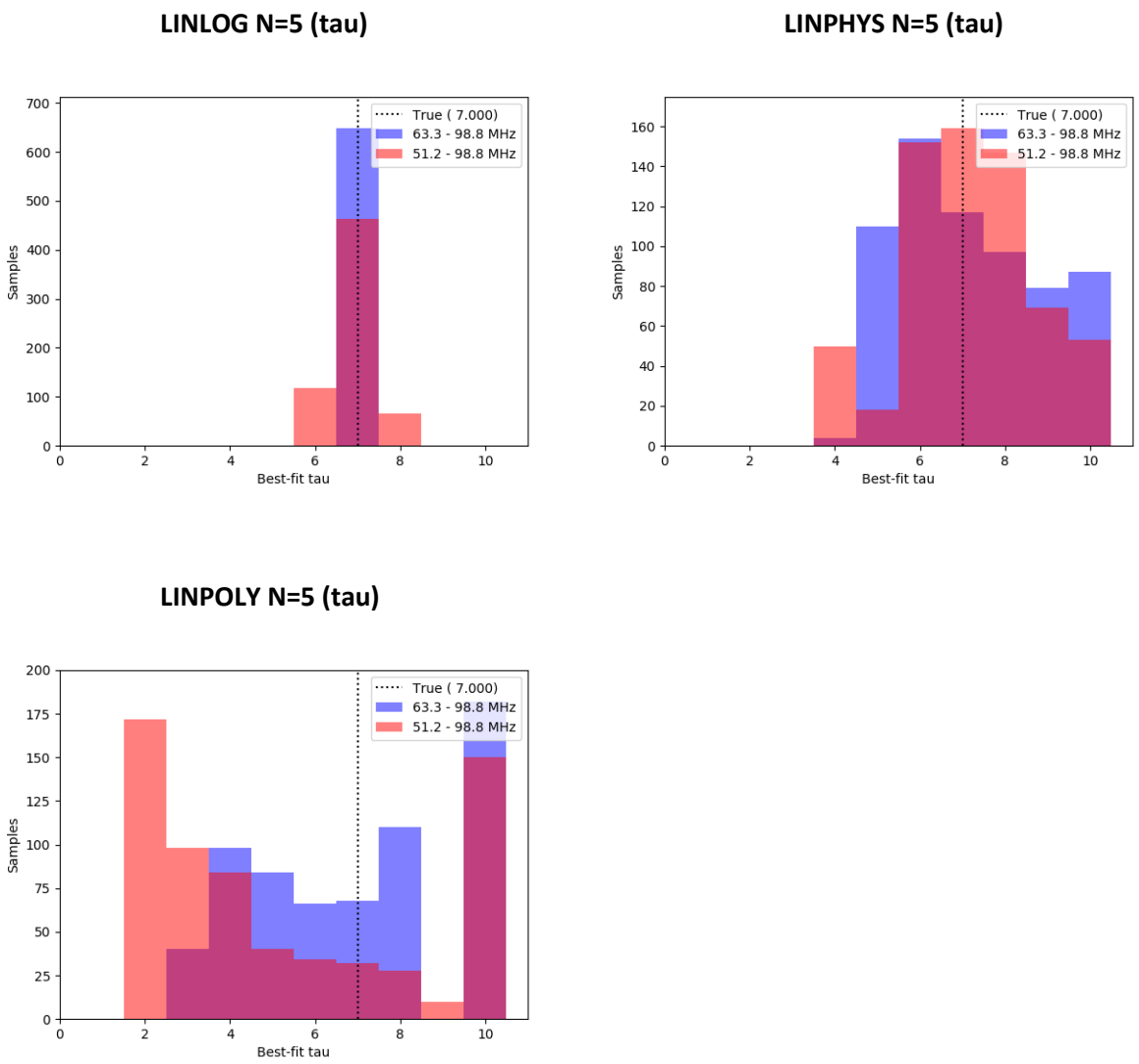
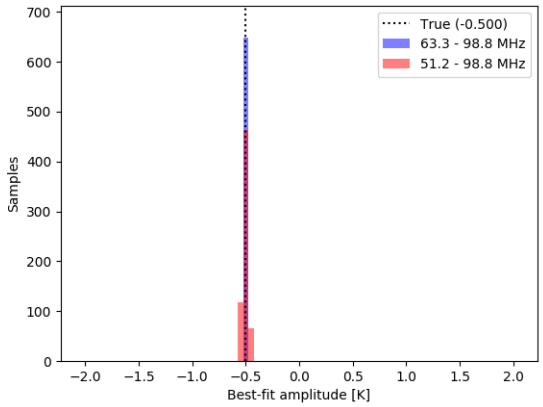
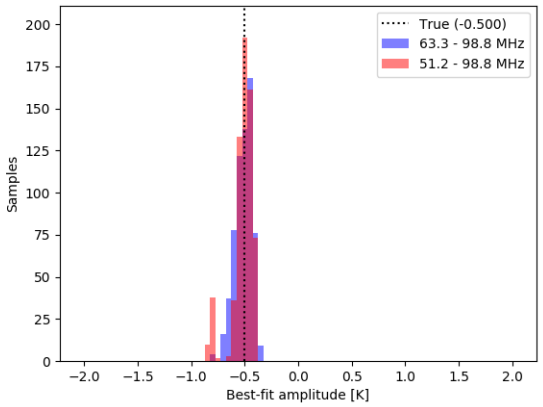


Figure 5. *Histograms of best-fit profile tau (flattening) recovered by each model with N=5 from noiseless simulations. The input profile tau was 7. Compared to the center frequency and width, tau is recovered less robustly by all three models. LINLOG recovers tau most robustly, with LINPHYS performing relatively well. LINPOLY fails to recover the injected tau.*

LINLOG N=5 (A)



LINPHYS N=5 (A)



LINPOLY N=5 (A)

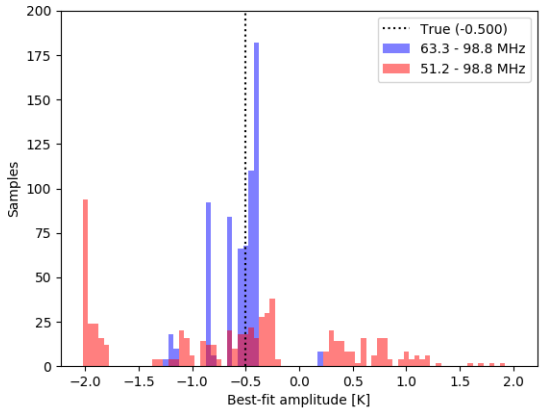
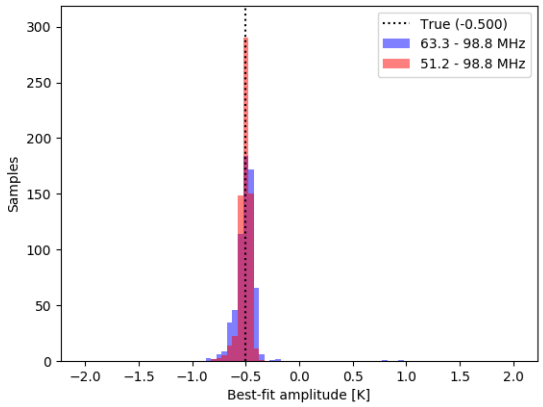
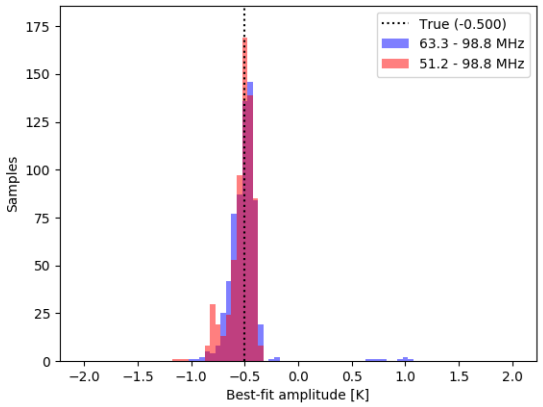


Figure 6. *Histograms of best-fit profile amplitude recovered by each model with N=5 from noiseless simulations. The input profile amplitude was -0.5 K. LINLOG performs best as was the case in Figures 2 through 5. LINPHYS performs nearly as well while LINPOLY performs much worse and general fails to recover the correct amplitude completely over the larger frequency range, but does tend to find an amplitude close to the input value (although biased a bit low) for the smaller frequency range.*

LINLOG N=5 (NOISE)



LINPHYS N=5 (NOISE)



LINPOLY N=5 (NOISE)

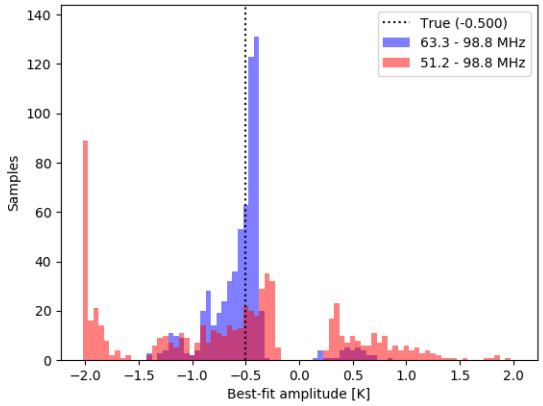


Figure 7. *Same as Figure 6 but with thermal noise of 25 mK RMS per 0.396 MHz channel included in the simulated spectra. LINLOG and LINPHYS continue to perform well.*

LINLOG N=4 (63-99 MHz) **LINLOG N=5 (63-99 MHz)**

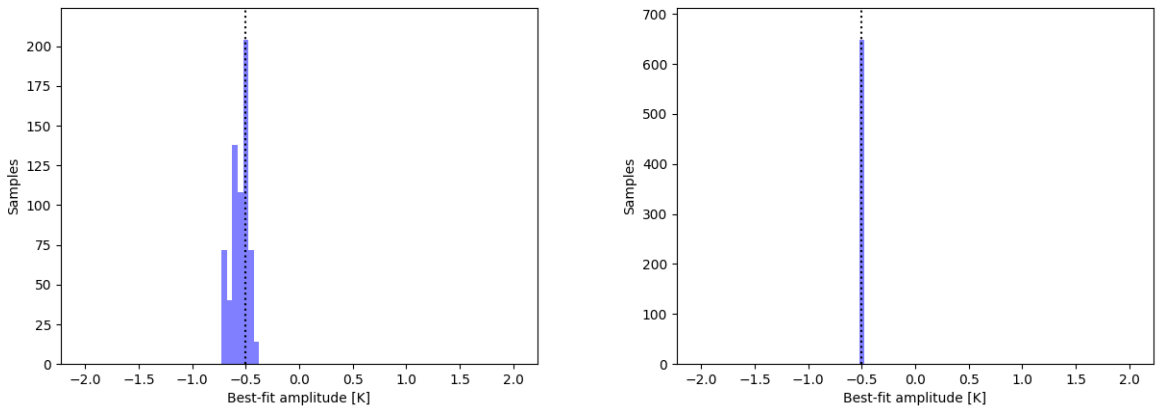


Figure 8. *Is LINLOG with N=4 a viable model choice over the smaller 63-99 MHz frequency range? The left panel shows the recovered best-fit amplitude of the injected signal in the foreground trial set for LINLOG with N=4 and omitting thermal noise in the simulated spectra. The right panel shows for N=5. While N=4 case does recover amplitudes centered on the correct value, it is easy to see that the N=5 case on the right more robustly recovers the exact input amplitude.*

LINPOLY N=6 (51-99 MHz) **LINPOLY N=7 (51-99 MHz)**

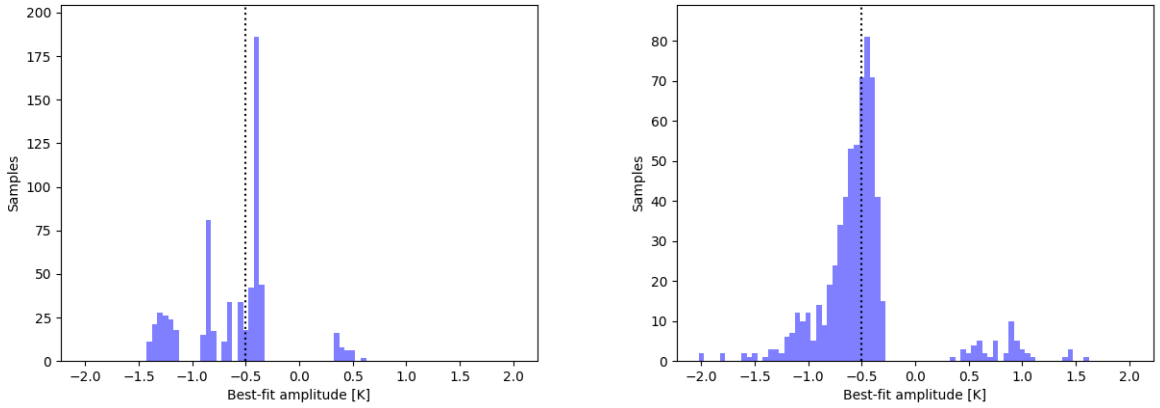


Figure 9. *Is LINPOLY with N>5 a viable model choice over the larger 51-99 MHz frequency range? The left panel shows the recovered best-fit amplitude of the injected signal in the foreground trial set for LINPOLY with N=6 and omitting thermal noise in the simulated spectra. The right panel shows for N=7. At N=6, LINPOLY continues to poorly recover the injected signal amplitude. By N=7 its performance has degraded, presumably due to increased covariance between the foreground model and the signal profile.*

REFERENCES

1. Jelić, V.; Zaroubi, S.; Labropoulos, P.; Thomas, R. M.; Bernardi, G.; Brentjens, M. A.; de Bruyn, A. G.; Ciardi, B.; Harker, G.; Koopmans, L. V. E.; Pandey, V. N.; Schaye, J.; Yatawatta, S., *Foreground simulations for the LOFAR-epoch of reionization experiment*, Monthly Notices of the Royal Astronomical Society, Volume 389, Issue 3, pp. 1319-1335, 2008
2. Bernardi, G.; McQuinn, M.; Greenhill, L. J., *Foreground Model and Antenna Calibration Errors in the Measurement of the Sky-averaged λ21 cm Signal at z~20*, The Astrophysical Journal, Volume 799, Issue 1, article id. 90, 11 pp., 2015
3. de Oliveira-Costa, Angélica; Tegmark, Max; Gaensler, B. M.; Jonas, Justin; Landecker, T. L.; Reich, Patricia, A model of diffuse Galactic radio emission from 10 MHz to 100 GHz, *Monthly Notices of the Royal Astronomical Society*, Volume 388, Issue 1, pp. 247-260, 2008
4. Rogers, A. E. E.; Bowman, J. D.; Vierinen, J.; Monsalve, R.; Mozdzen, T., *Radiometric measurements of electron temperature and opacity of ionospheric perturbations*, Radio Science, Volume 50, Issue 2, pp. 130-137, 2015
5. Vedantham, H. K.; Koopmans, L. V. E.; de Bruyn, A. G.; Wijnholds, S. J.; Ciardi, B.; Brentjens, M. A., *Chromatic effects in the 21 cm global signal from the cosmic dawn*, Monthly Notices of the Royal Astronomical Society, Volume 437, Issue 2, p.1056-1069, 2014
6. Bowman, Judd D.; Rogers, Alan E. E.; Monsalve, Raul A.; Mozdzen, Thomas J.; Mahesh, Nivedita, *An absorption profile centred at 78 megahertz in the sky-averaged spectrum*, Nature, Volume 555, Issue 7694, pp. 67-70, 2018