MIST Memo 66
Testing the Accuracy of the Beam Interpolation Code

Xinze Guo
UC Berkeley
sunnyguo@berkeley.edu

May 12, 2022

1 Background

The antenna is a delicate device to detect and measure radio signals and it is an important measurement in the 21cm cosmology. The current antennas' beam gain depends on the geometric shape of the antenna, frequency, elevation, soil permittivity and resistivity, and other complicated elements. To represent the sensitivity of the antenna, beam gain cube, a 3D NumPy array in dimensions of frequency, elevation, and azimuth, is created.

The Antenna-Beam-Model-Interpolation-Package is a fast way to produce antenna beam models using interpolation between preexisting beam models computed using EM simulations. Here is the link for more information about the package. This code was written by Ian Hendricksen, and it can take in FEKO data and process it much quicker than the traditional full EM simulations of antenna beam models with the sacrifice of reasonable accuracy.

The package is designed to interpolate FEKO data generated from simulations. In this demo, I tested the accuracy of the interpolation package for theoretical models generated through ideal formulas. To reach the goal, I created theoretical cubes according to two formulas and used the package to interpolate. I used 12 different metrics to measure the accuracy of the interpolation and found the pattern between orders and accuracy.

2 Theory and Research Method

To test the codes' ability to interpolate beam cubes created by formula instead of simulated FEKO data, I wrote a few functions to let it accept another input format. Specifically, I wrote a function to generate beam cubes according to a formula of amplitude A and parameter σ_x. Also, I wrote a new pg-collector that could match the format with the abm-interpolator. Finally, a function to reformat the 2D table into a 3D beam cube was written for further analysis.

The theoretical beam cubes used in this memo are 3D NumPy arrays in the form of $51 \times 181 \times 360$. It represents the sensitivity of the antenna at the frequency of 50 to 100 Hz, the elevation of -90 to 90 degrees, and the azimuth of 0 to 359 degrees. The beam cube depends on frequency, elevation (el), and σ in this Gaussian formula:

$$G(\nu, el, \sigma) = \frac{A_0}{75} \times \nu \times e^{-(el-el_0)^2/2\sigma^2} \tag{1}$$

where ν is the frequency, and A_0 is the amplitude at 75 MHz. In this memo, I chose the parameters el_0 to be 90, σ to be [35,45,55], and A_0 was chosen to be [8,9,10]. Using my cube generation function and the above parameters, 9 beam cubes with the combination of parameters were generated.

I generated a reference cube with Formula 1 and parameters of $A_0 = 9.5$ and $\sigma = 40$. Figure 1 and Figure 2 show the reference beam cube's beam gain. Figure 1 shows the beam gain map for all elevation and azimuth at 50 MHz and we can see that the beam gain increases as the elevation increases and it is

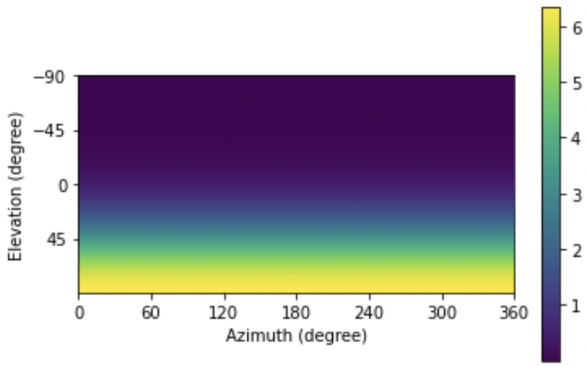

Figure 1: Reference cube's beam gain map at 50 MHz.

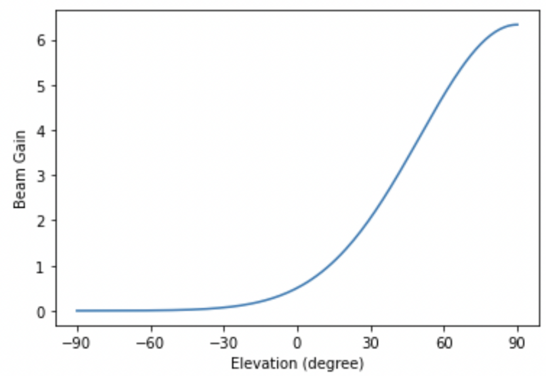

Figure 2: Beam gain at 50 MHz, azimuth at 0.

independent of azimuth. Figure 2 is a plot of beam gain VS elevation at 50 MHz and azimuth at 0 degrees, and we can see a Gaussian distribution of the beam gain.

The interpolation package requires the given 9 beam cubes and their corresponding parameters, the parameter needs to interpolate, and also the order. The order shows the degree of interpolation and it consists of an array of numbers whose length matches the number of parameters. In this demo, since there are 2 parameters, A_0, and σ_0, the order contains two numbers. The choice of order will affect the interpolation result, so I also tried to find the pattern between the interpolation accuracy and the order in this memo.

I used the modified code to interpolate with nine different orders (11, 12, 13, 21, 22, 23, 31, 32, 33) and got 9 interpolated beam cubes. To determine the accuracy of the interpolation, I had 12 metrics to calculate the difference. As shown in Figure 4, I did data analysis in two groups: one for the absolute difference and one for the absolute percentage difference. The absolute difference was calculated by the following formula:

$$d = |x - x_0| \tag{2}$$

The percentage difference was calculated using the following formula:

$$d = 100 \times \frac{|x - x_0|}{x_0} \tag{3}$$

where x is the interpolated cube data and x_0 is the corresponding pixel's reference cube data for Formula 2 and Formula 3.

Since the beam gain in 45 to 90 degrees elevation is bigger and is more important to study, there are two subgroups for the absolute difference metric and percentage difference metric: the whole data group and the partial data group that contains data from 45 to 90 degrees elevation. For each of the four subgroups, there are mean, median, and maximum metrics to compare the accuracy. Take the median metric in the partial data group as an example, for each order, The median value of the 360 × 45 pixels for each frequency was calculated and drawn as a line in the median beam gain difference versus frequency plot, as shown in Figure 6. We can see nine lines in the plots, and the lowest line is the most accurate while the top line is the least accurate. I drew 12 plots and ranked the orders from the most accurate to the least accurate for each one. Figure 4 summarizes the max value in each order calculated using the metrics.

3 Results

From Figure 4, the summary chart and the 12 plots, we can conclude the following:

1. The interpolation package is very accurate for interpolating theoretical model.
2. For absolute difference, the difference calculated by partial data mean metric is bigger than that by the whole data mean metric, while for percentage difference, it's the opposite: partial data mean metric calculates much smaller differences.

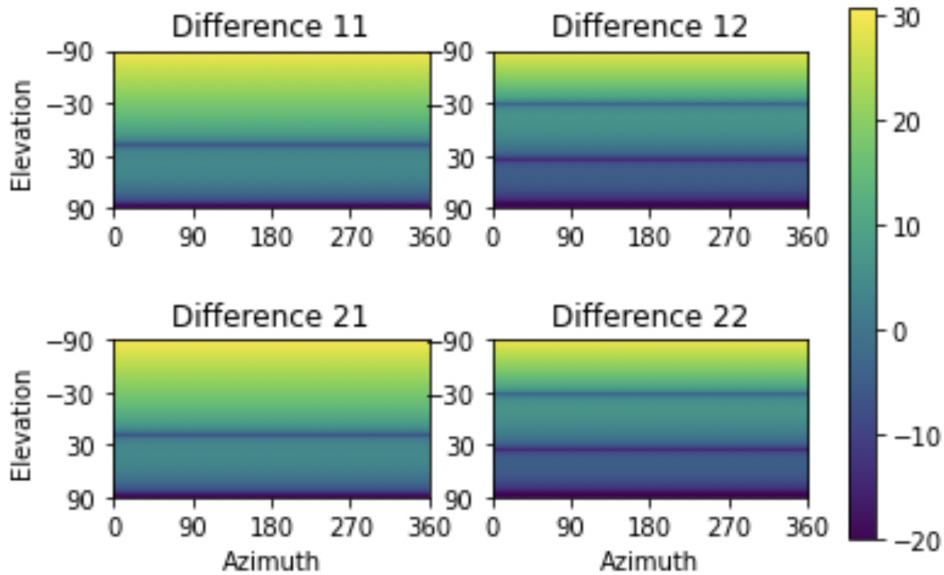

Figure 3: $10 \times log_{10}$ the percentage difference of orders 11, 12, 21, 22 at 75 MHz.

3. For all plots, they follow the same order from the most accuracy to the least accuracy. The most accurate interpolation orders are 12 and 22. The order from the most accurate to the least accurate is: 12≈22, 11≈21, 31, 32, 13, 23, 33.

4. The interpolations are relatively inaccurate when containing at least one 3. They are very inaccurate when the second place of order is 3.

5. The second position of the order should be interpolated with a higher order of 2, while the first position only requires an order of 1.

6. The mean percentage difference is larger than the median percentage difference.

1D Gaussian												
	Difference											
	Whole Data						45-90 Elevation Data					
	Mean		Median		Max		Mean		Median		Max	
1	12&22	0.017	12&22	0.012	12&22	0.4	12&22	0.018	12&22	0.02	12&22	0.03
2	11&21	0.08	11&21	0.075	11&21	0.18	11&21	0.1	11&21	0.1	11&21	0.18
3	31	2.5	31	1.1	31	7.5	31	6.4	31	6.6	31	7.8
4	32	100	32	45	32	300	32	272	32	281	32	321
	Percentage Difference											
	Whole Data						45-90 Elevation Data					
	Mean		Median		Max		Mean		Median		Max	
1	12&22	7.60E+01	12&22	2.64	12&22	940	12&22	0.19	12&22	0.2	12&22	0.35
2	11&21	1.37E+02	11&21	11	11&21	1164	11&21	1.12	11&21	0.95	11&21	2.6
3	31	5.00E+02	31	110	31	4250	31	61	31	60.9	31	61.6
4	32	1.40E+04	32	4500	32	100000	32	2607	32	2580	32	2760

Figure 4: Order of accuracy summary chart using 12 metrics. From top to bottom is the order from the most accurate to the least accurate, and the number on the right shows the maximum in all frequency using that metric and order.

4 Discussion

The accuracy is measured by how small the percentage difference is. From Figure 4, we can see that the maximum mean of percentage differences is 7.6% in the better interpolation with orders 12 or 22. In a worse chosen order, it's 13.7%. However, the beam gain at small elevation angles (-90 to 45 degrees) in the reference cube is very low (around 0.00025), so the percentage difference is very high in that region according to Formula 3. Moreover, we pay more attention to the region near the peak, since the beam gain is higher and contributes more when calculating the sky temperature. Thus, when I calculated the mean of the percentage difference near the peak, I found that the maximum mean is 0.19% for the orders 12 and 22 and in the worse case, it's 1.12%. This indicates that the interpolation data has an average percentage difference of less than 0.19%, which is a very small number, indicating that the interpolation is accurate. Thus, we can conclude that the interpolation package for interpolating the theoretical model is very accurate.

As we discussed previously, Formula 3 divides the absolute difference by the reference cube, so small data in the reference cube will affect the percentage difference severely. The very small beam gain in the low elevation region results in a big percentage difference, and it leads to a bigger mean. Around the peak region, the extreme big data is eliminated so the partial mean metric in percentage difference results in smaller data compared to the whole data mean metric.

For the absolute difference, since the value around the peak is much larger than the low elevation area, a similar percentage error will result in a bigger absolute difference. The absolute difference near the peak is higher than others, so its mean is also higher than the whole data mean.

The interpolation is worse when the orders are 31 and 32, and exploded (with the difference of 10^{13}) when the second position is three. From the accuracy of using the 12$\approx$22 orders are better than using 11$\approx$21 orders, we can see that two is better than one in the dimension's second place. However, 31 is more accurate than 32. The reason might be that three exceeds the limit of degrees of freedom, and two has greater degrees of freedom than one and will make the interpolation worse. Similar reason for 13 is more accurate than 23 and 23 is better than 33.

Order 11 and order 21 are similar, while order 12 and order 22 are also similar. From this, we can conclude that for the first position in the order, the amplitude, a one degree will give an accurate result, so we don't need to make it two. However, as discussed in the last paragraph, there is a difference between one and two in the second place. This shows that the second position, σ, requires a higher order to interpolate.

The percentage difference has extreme high values when the elevation is low, since the mean is more sensitive to extreme values.

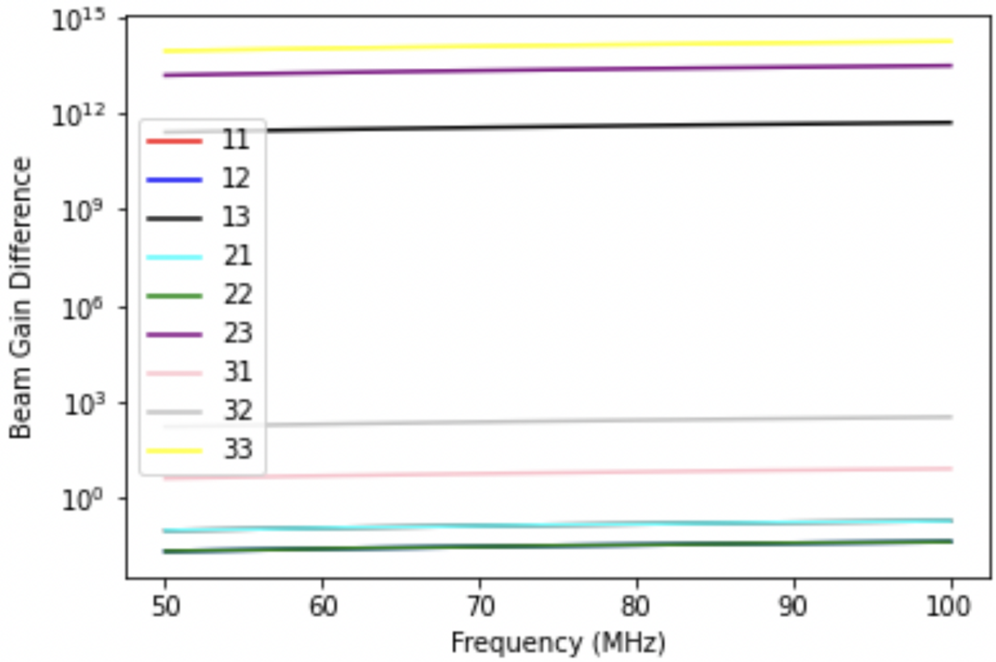

Figure 5: Beam gain difference versus frequency for nine orders using max absolute difference metric.

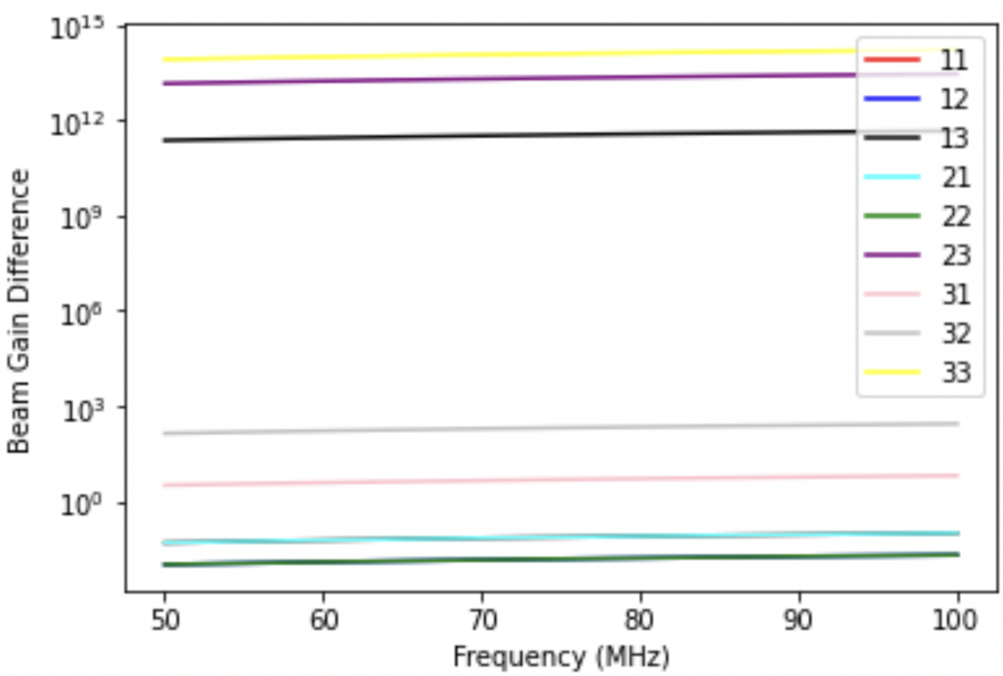

Figure 6: Beam gain difference versus frequency for nine orders using partial median absolute difference metric.

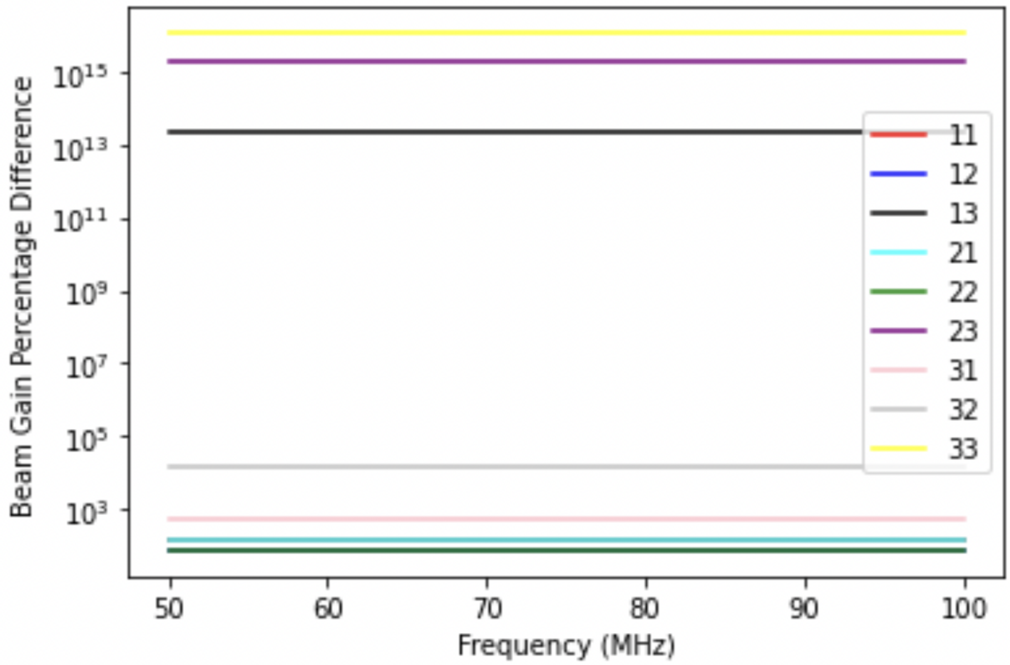

Figure 7: Beam gain versus frequency for nine orders using mean percentage difference metric.

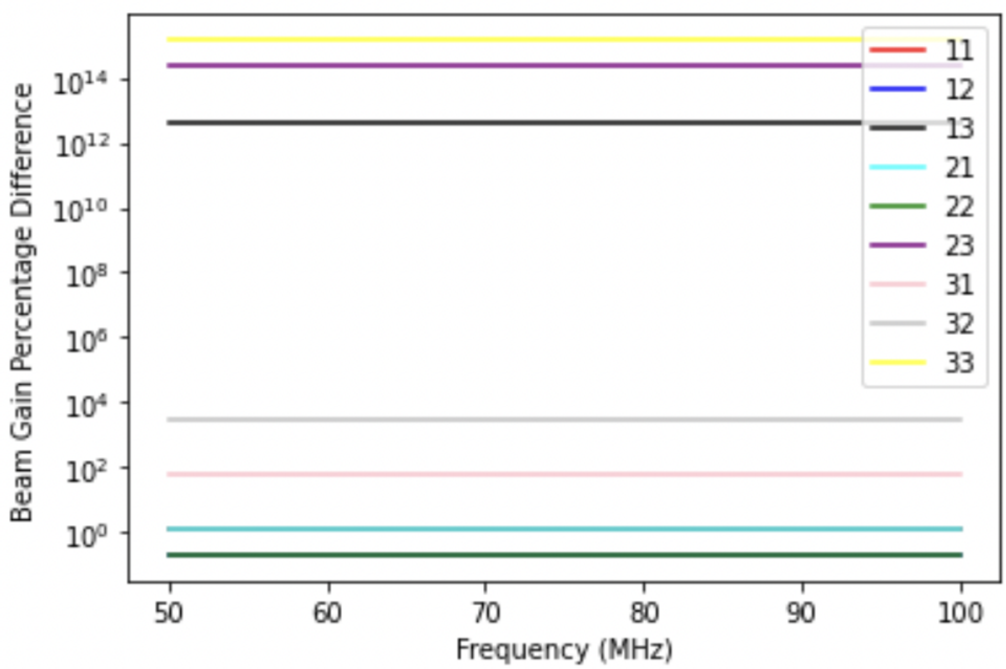

Figure 8: Beam gain versus frequency for nine orders using partial mean percentage difference metric.