

# HERA data vs pyuvsim simulation comparison

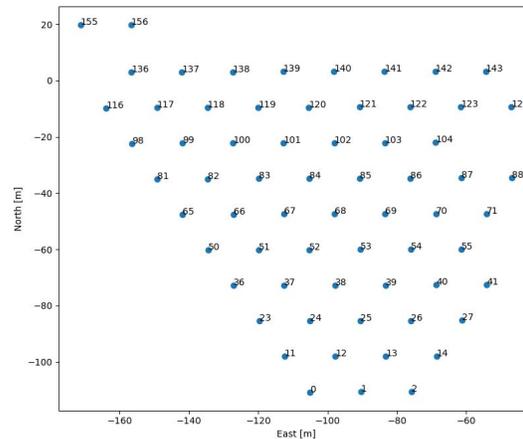
Daniya Seitova

## ABSTRACT

This memo presents a comparison between HERA H1C data and pyuvsim simulation. We compare the waterfall plots of the one of the baselines in one ten-minute data file and the corresponding baseline in pyuvsim. We also compare the waterfall plots in delay space. We find that the data and simulation look similar, but data files contain fringing that pyuvsim does not. It is not clear whether pyuvsim is missing a source or there are systematics in the HERA data complicating the comparison.

## 1. Introduction

We chose 137\_2 baseline for comparison in both HERA data file and pyuvsim, as it is one of the longest baselines. This baseline also allows us to conduct further comparison with shorter baselines that are in the same direction or perpendicular to 137\_2<sup>1</sup>.



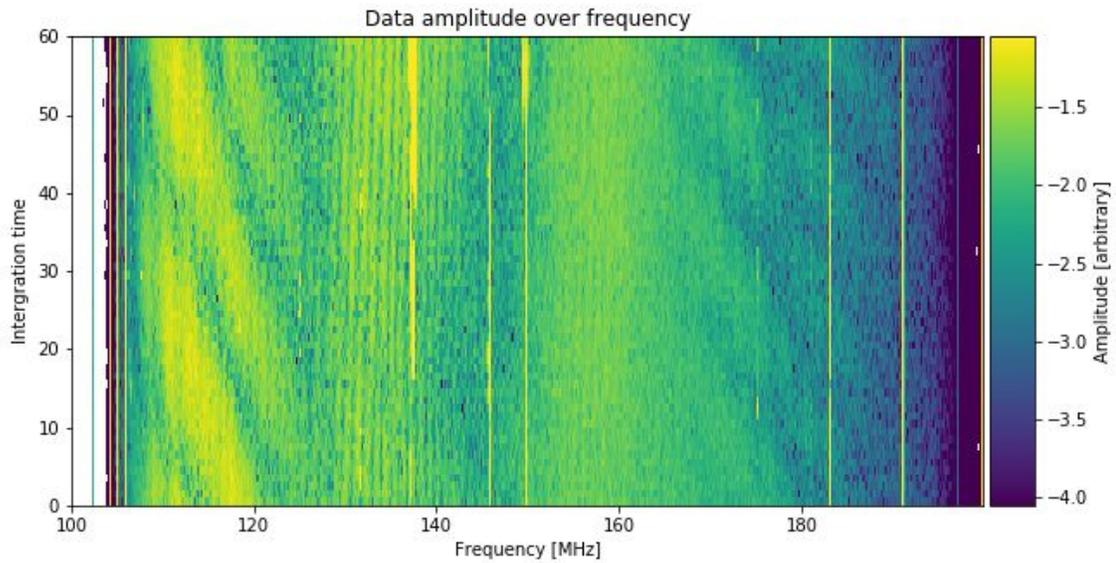
We chose HERA data file zen1.2458110.18517.HH.uv, which is at RA~0 to avoid bright sources off zenith. The HERA data file contains 60 times, 1024 frequencies and 2 polarizations. The corresponding pyuvsim simulation contains the same 60 times, 1024 frequencies and 4 polarizations. Time resolution is 11 seconds and frequency resolution is 97.66 kHz. The GLEAM source catalog and sources brighter than 0.5Jy were used in the simulations with a simple analytic airy beam model.

---

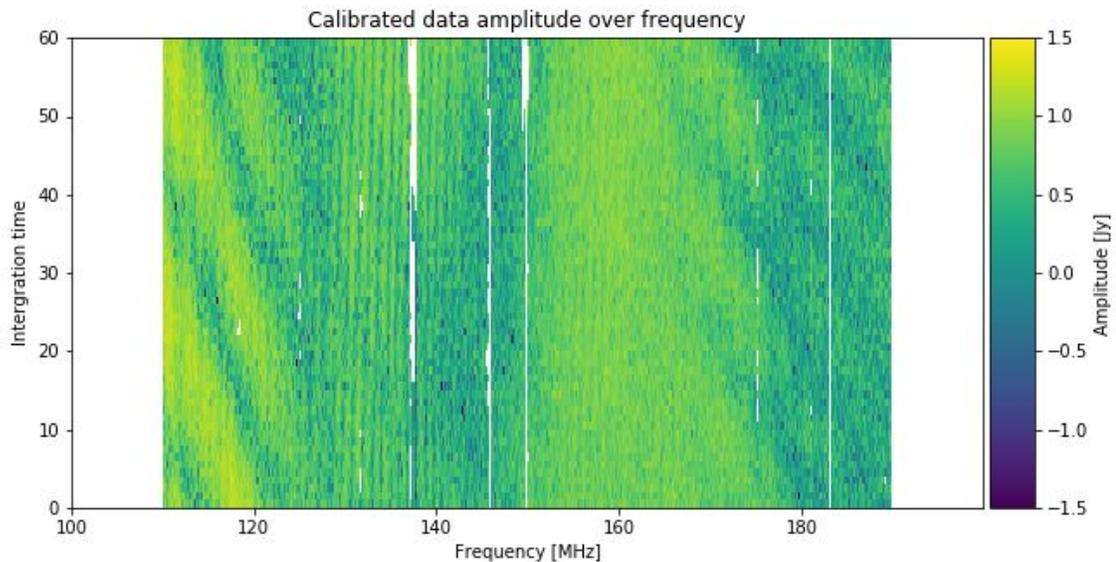
<sup>1</sup> We found out after completing the analysis that antenna 2 might have questionable quality. But when we switched to different baselines, the results were unchanged.

## 2. Discussion

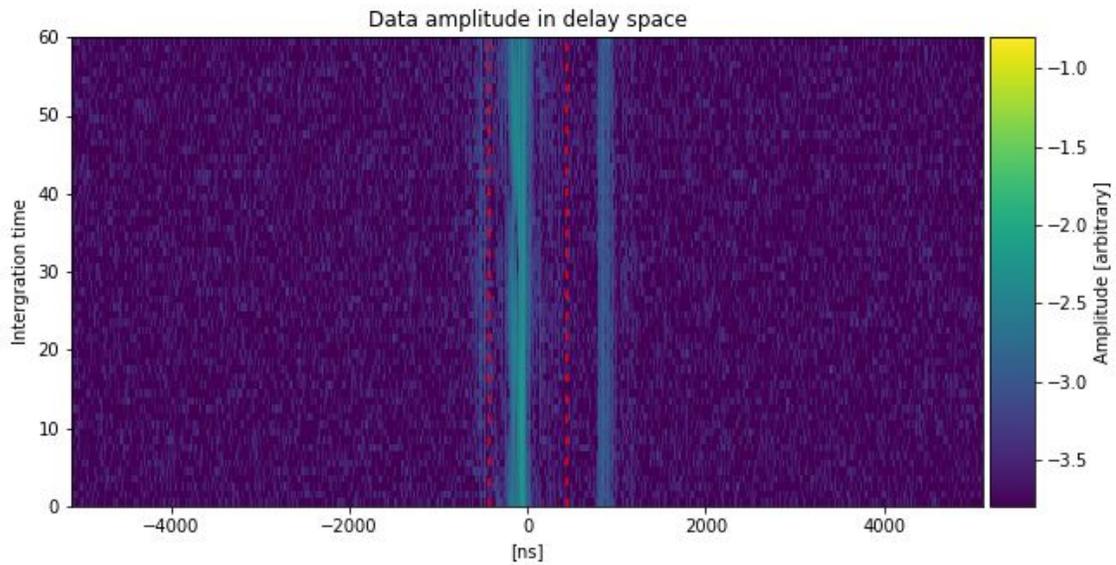
### 2.1 Comparison with calibrated data



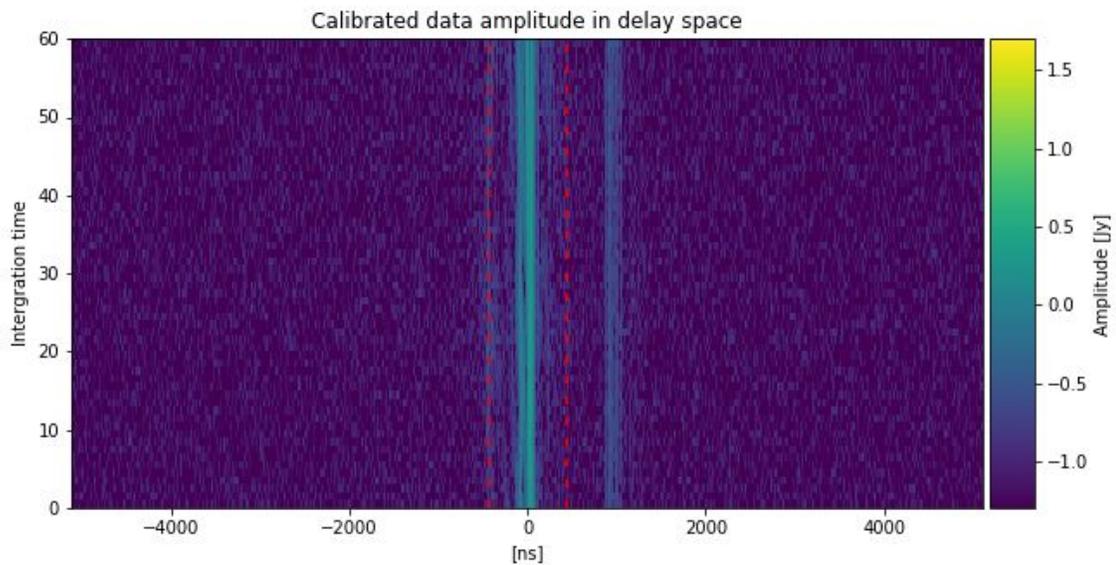
In the plot above, we have a waterfall plot of raw data, and below we have a plot of calibrated data, using Josh Kerrigan's calibration with FHD. We can see that the plot does not change much when we calibrate the data. For the rest of the analysis, we are going to use calibrated data.



We are now going to compare the raw data and calibrated data in delay space.

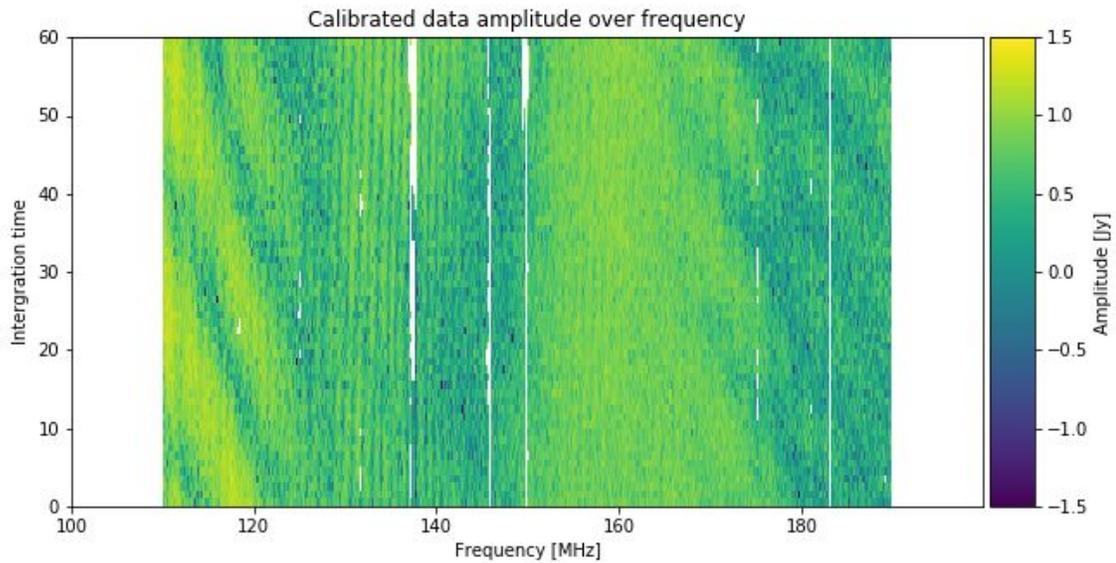
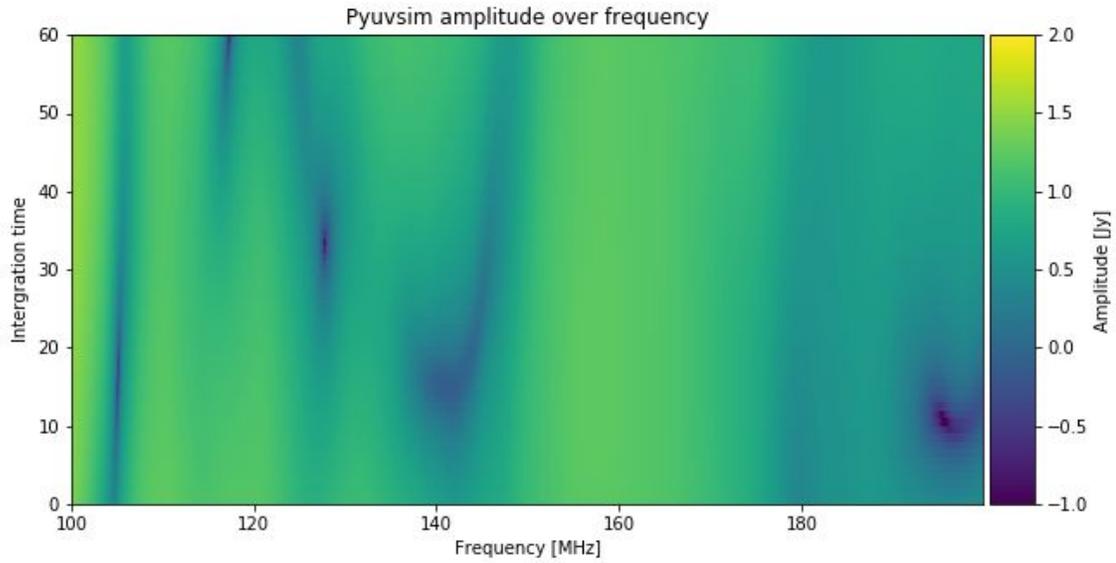


Similarly, we now have a plot of raw data in delay space above, and a plot of calibrated data in delay space below. We can also notice that in delay space, the raw data does not differ much from the calibrated data.

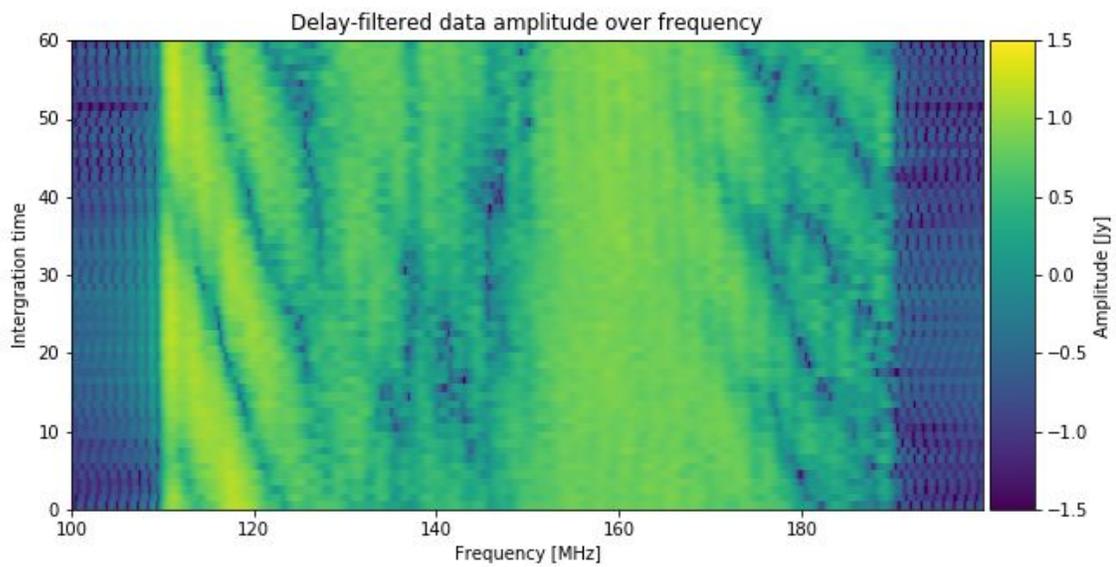


To analyse the data better, we are going to apply a delay filter. We take a delay transform of the visibility with respect to frequency and then apply a top hat filter, where everything outside the horizon lines is set to zero. Then we take an inverse delay transform to make the waterfall plot of the visibilities. However, the delay filter can only be applied to calibrated data, which is why we are comparing raw data with calibrated data to make sure we don't lose any information by looking at calibrated data, instead of raw data.

## 2.2 Amplitude waterfall plots comparison

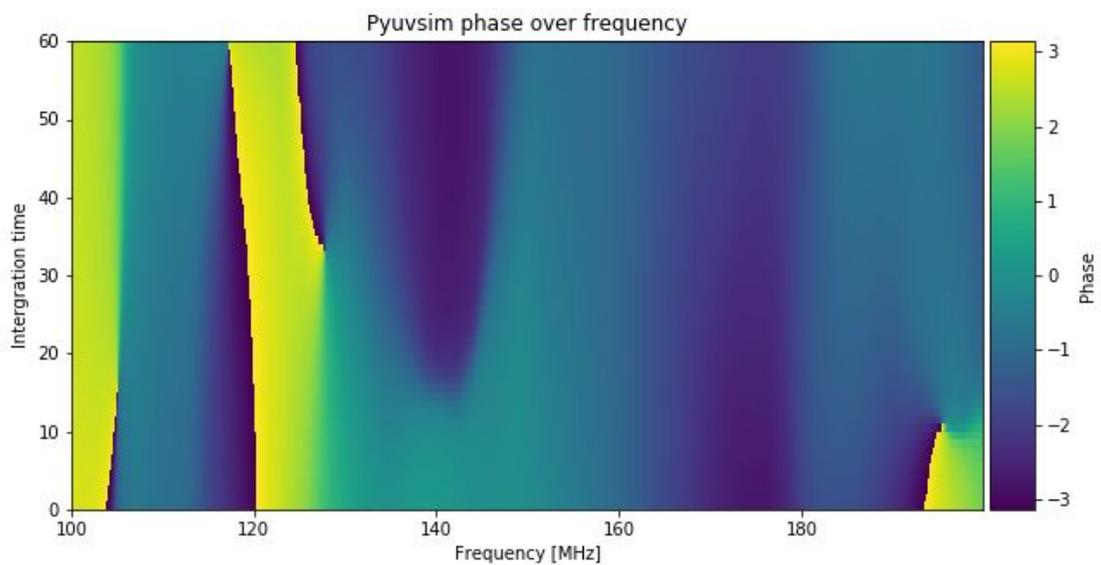


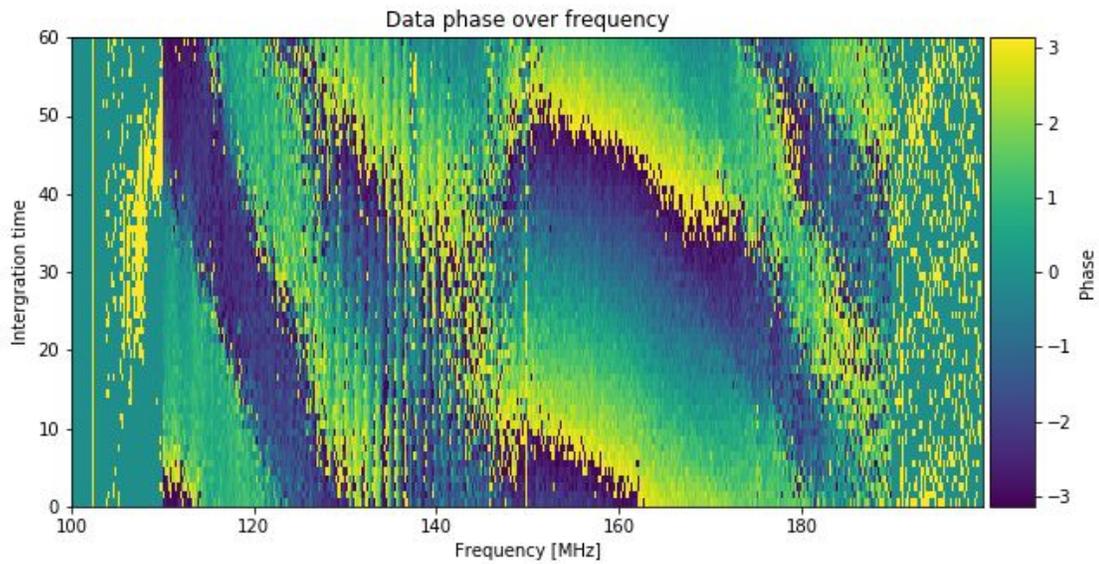
To remove systematic errors at high delays outside the horizon, we applied a delay filter to the data. There are ripples in the data that we do not see in pyuvsim, so we wanted to check if it would still be there if a delay filter was applied. The fast ripple seems to disappear, however the slow ripple is still there as we can see in the following waterfall plot:



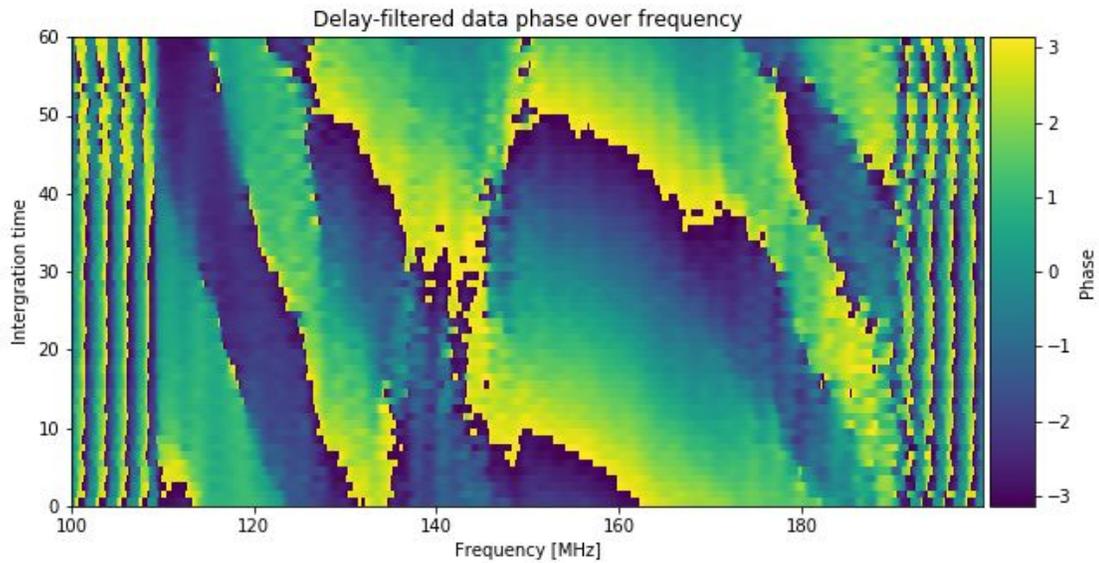
To further investigate the cause of this fringe, we conduct a comparison of the phases, instead of amplitudes.

### 2.3 Phase waterfall plots comparison



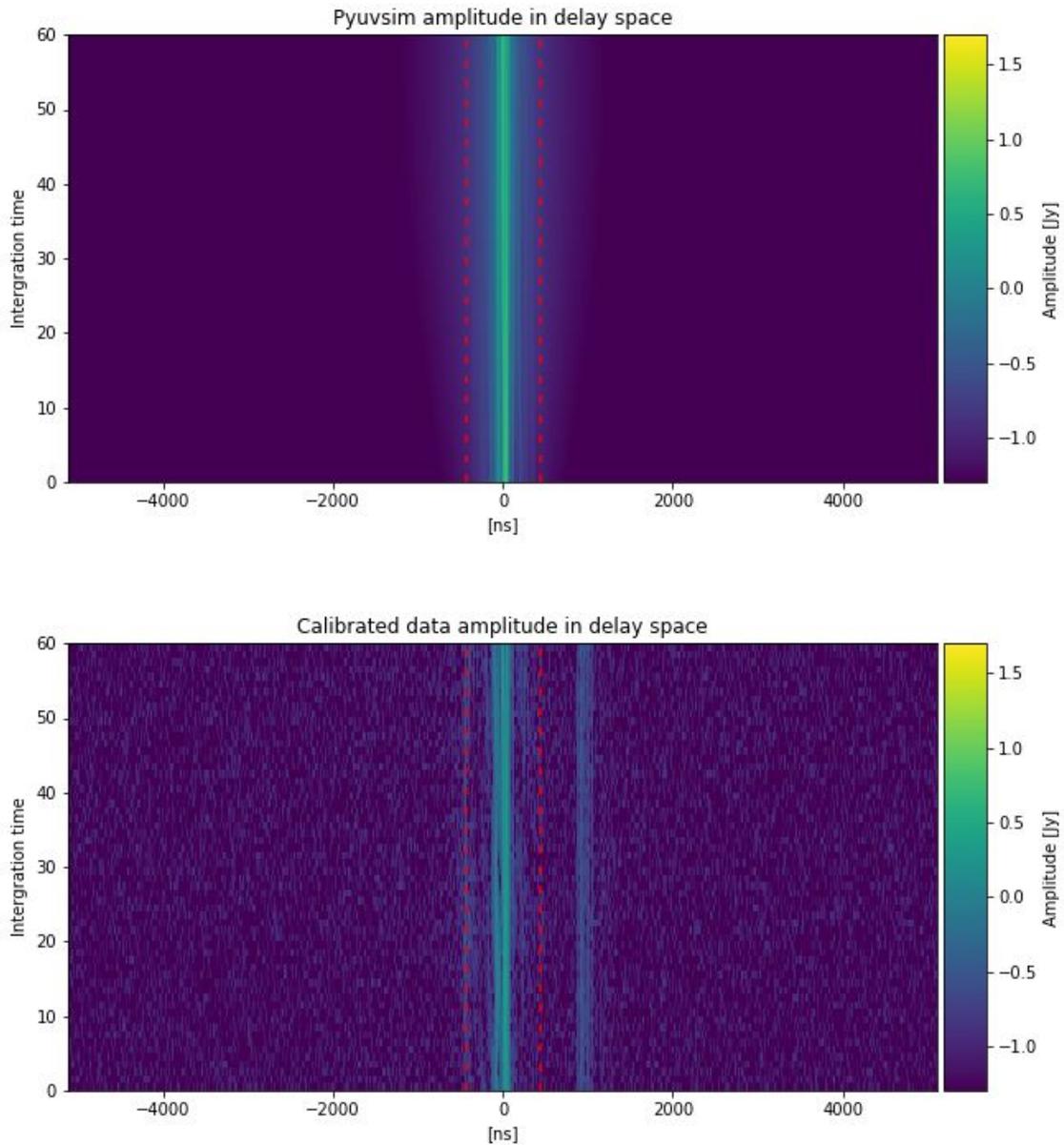


As we can see the phase plots conspicuously show that the fringe we see in data is absent in pyuvsim. Even when we apply a delay filter (described in Section 2.1), we can still see the fringe in the following plot.

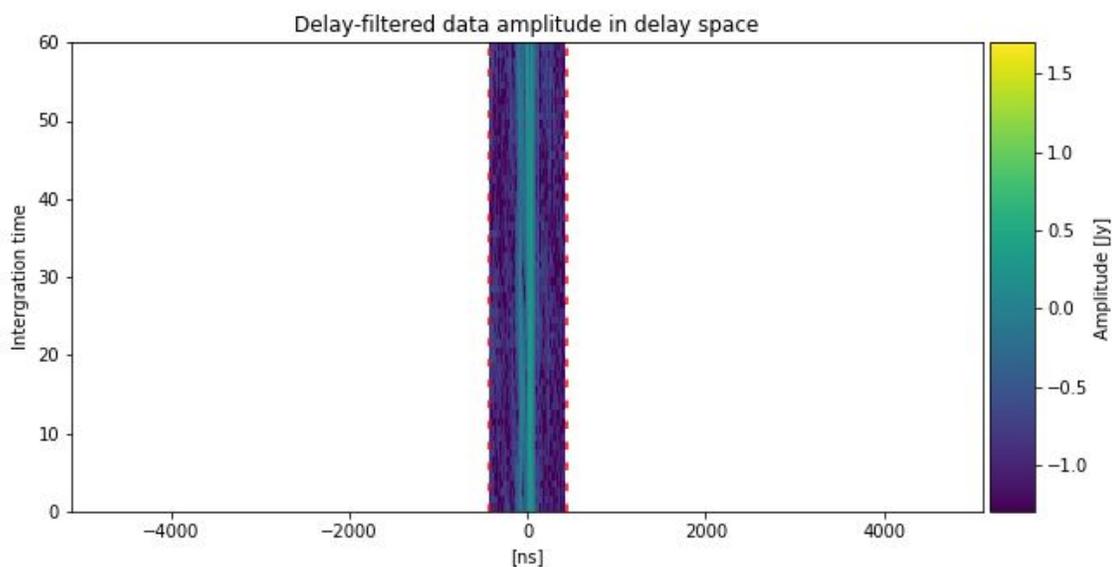


## 2.4 Amplitude waterfall plots in delay space comparison

We plotted the amplitudes of pyuvsim and data in delay space. Dashed lines indicate the horizon limits.



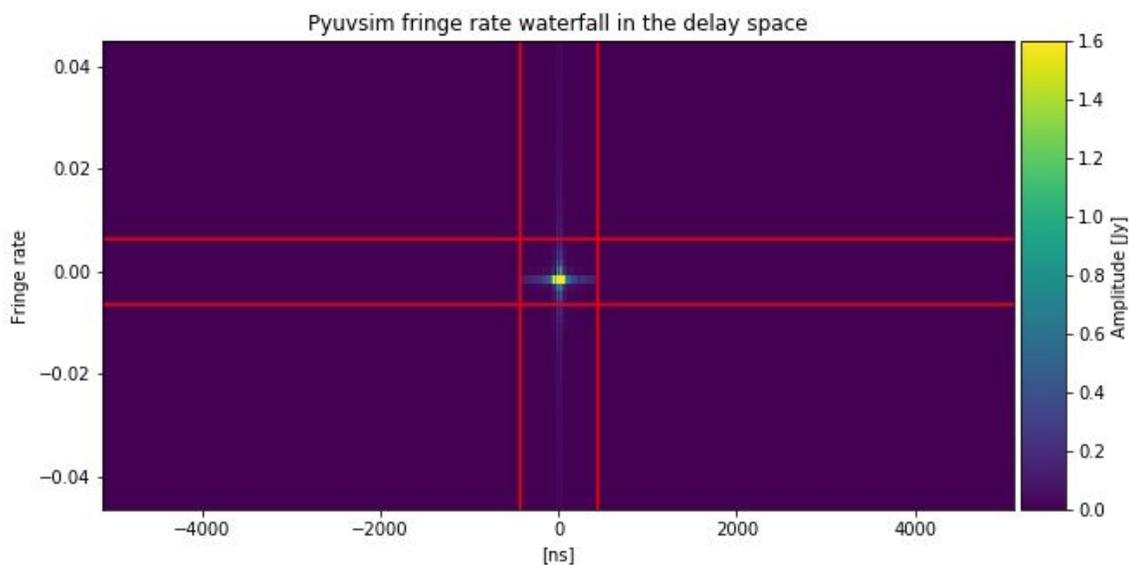
We can see in the plot above that there are additional line, that are not in pyuvsim simulation plot. There's a line on the right, outside the horizon limits, and another fainter line on the left, which is on or just outside the horizon. The brightest line in the middle is also split into two, whereas it's a single line in pyuvsim simulation plot. However, these lines fall either on or outside the horizon lines so we can remove them by applying the delay filter described in Section 2.1.

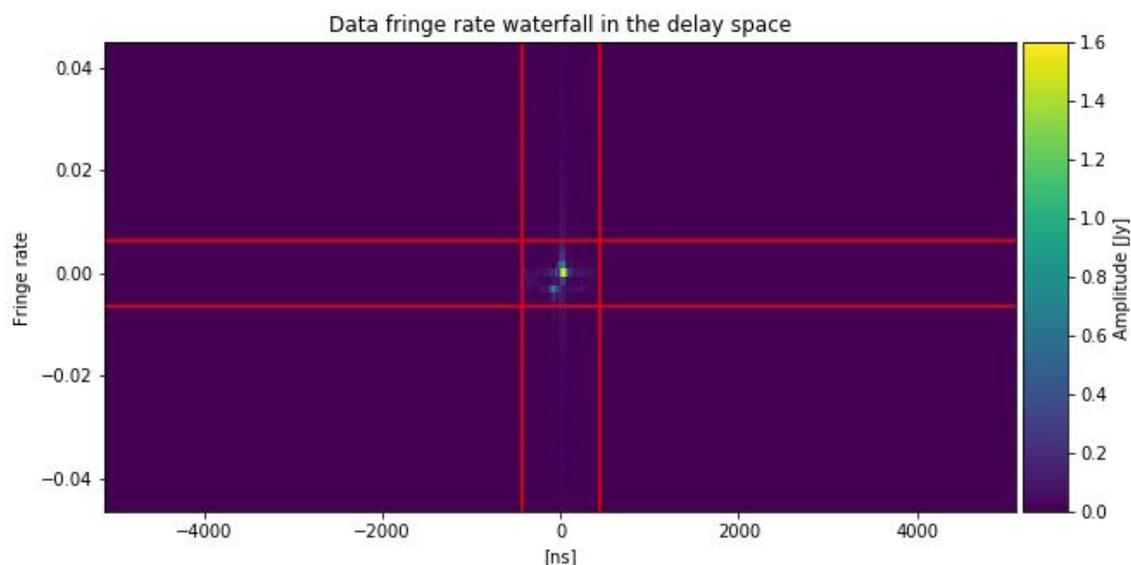


As we can see there are no longer extra lines or “sources” on the plot as everything outside the horizon lines has been taken to zero. However, we can see that the emission in the middle is not as symmetrical as the one in pyuvsim plot, particularly at negative delays. This emission appears responsible for the fringing seen in HERA data but not pyuvsim, so we want to further explore its nature.

### 2.5 Delay-fringe rate comparison

We also took a delay transform of the visibilities and then did a fringe rate transform to analyze the data in delay-delay-rate (DDR) space. (Parsons & Backer 2009)

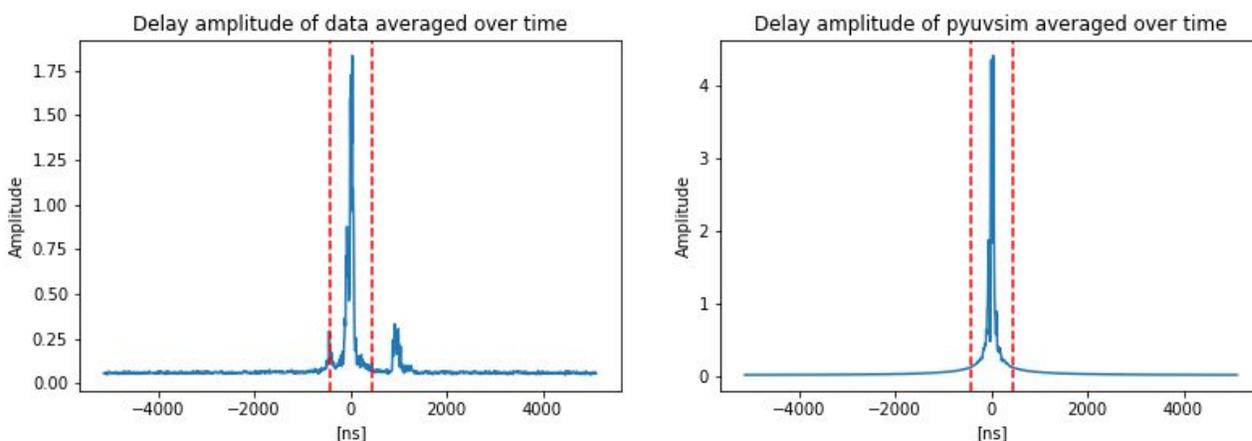




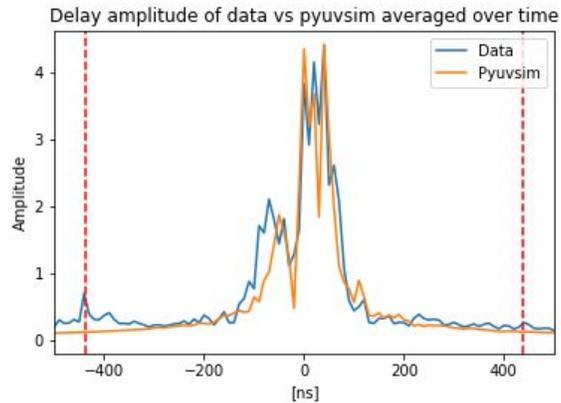
The horizon and maximum fringe rate lines are in red, and as we can see there are two bright points in the data plot, but only one point in pyuvsim. However, as both points are inside the red lines, the signal is potentially consistent with something from the sky, meaning there might be another source in the data that is not included in pyuvsim.

### 2.6 Amplitude plots in the delay space comparison

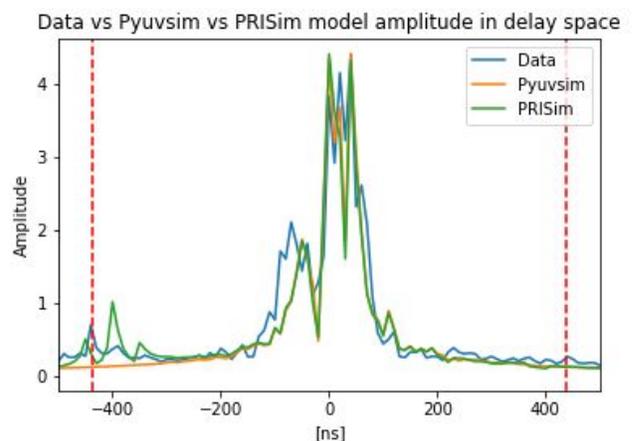
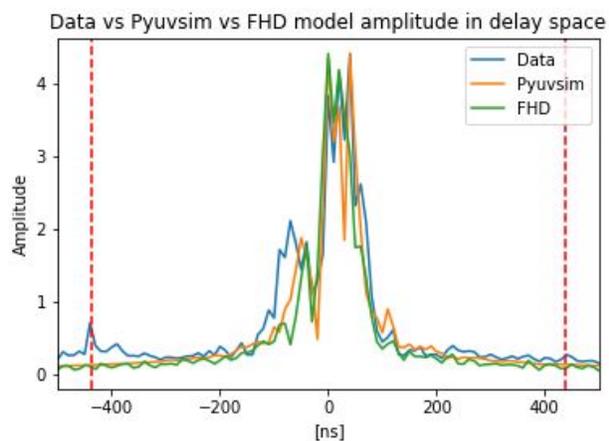
We then made 1D plots of amplitude in delay space averaged over time to see if the additional source we see in the data is still there.



When we zoom in on the plots since we are only interested in what is inside the horizon lines, then we can see that there's a feature in the data to the left of the main source.



As we can see the feature, on the left, is bigger in the data plot than in pyuvsim. We conducted similar simulations with both PRISim and FHD and we found those simulations to be consistent with pyuvsim. This led us to believe that it could be a contaminant in the HERA data but we cannot be certain that it is.



### 3. Conclusion

In conclusion, HERA H1c data and pyuvsim simulation waterfall plots look similar. However, the data files contain fringing that pyuvsim does not. It seems that pyuvsim is missing a source when comparing the plots in delay space. But when comparing the data with other simulations it becomes unclear whether pyuvsim is missing a source. It is possible this source is missing from the GLEAM catalog itself, hence it doesn't appear in any of the simulations based on GLEAM. However, this field is the well-studied MWA EoR0 field and any obviously missing sources from GLEAM would likely have been already identified in MWA analysis. One potential explanation is that the extra “source” is in fact a

contaminant in the HERA data; comparing with H2C data will prove informative in the future.