The Impact of Simulator Choice and Parameterization on Fringe-Rate Signal Loss Calculations

Aaron Ewall-Wice

January 2022

Abstract

In this memo, we explore the impact of the choice of different visibility simulators on calculations of signal loss for fringe-rate filtering as a first step towards a more comprehensive analysis along the lines of [Aguirre et al., 2022] and [Kern et al., 2021]. We find very similar estimates for signal loss predictions are achieved between the pyuvsim simulator package and vis_cpu and healvis simulators with their default settings, though ≈1% time domain artifacts are prevalent in the pixelized beam prescription for vis_cpu which may motivate using higher beam pixelization or a different simulation. We also find that simulations with ~10000 sources (Nside=32) either randomly distributed or regularly spaced in a healpix grid give very similar results to ~300000 source simulations (Nside=128). Validation may find it expedient to run random fluctuations simulations at relatively low resolutions. We also observe spike-like artifacts in fringe-rate that are present only in healpix distributed source models but are not present in source models where the sources are distributed randomly.

1 The Simulations

We explore the impact of simulator choice on the signal loss for a single visibility in the HERA array at 155 MHz using four different simulation configurations listed in Table 2. We compare the pyuvsim [Lanman et al., 2019], healvis [Lanman et al., 2020], and vis_cpu simulators. We also experimented with increasing the sampling resolution from every ≈100 seconds to every 10 seconds (4250 time samples instead of 425 time samples) and found this to have a negligible impact on our results. We use the CST Vivaldi beam located at [https://github.com/HERA-Team/HERA-Beams/blob/master/NicolasFagnoniBeams/NF_HERA_Vivaldi_efield_beam.fits] Fagnoni et al., 2021 in all of our simulations. For our vis_cpu and pyuvsim simulations, we use the same source locations using the skyh5 format to pass the same consistent sky models between simulators. In all simulations, our sky model consists of discrete point sources either arranged on a healpix grid or randomly distributed with independently distributed fluxes drawn from an exponential distribution with a mean of two Jy. In our healvis simulation, the sources are drawn from a mean with two Jy / Sr. In all simulations, we use the baseline between antennas 91 and 92 in the HERA layout which is a 12 m EW and 8 m NS baseline.

In Fig. 1 we show the PSD in time at 155 MHz For the same baseline with various simulation choices. Most notably, vis_cpu with its default settings is comprised of significant fringe-rate artifacts at the ≈10^{-3} level in power. For the other simulators, we observe differences between the various simulation prescriptions at the sub 1% level. In the regularly gridded healvis simulations for the NN pol, we see spike artifacts at ±1.5 mHz for the NSide=32 simulation which appear at larger fringe-rates and at a lower level the Nside=128 simulation. These same spikes appear with larger amplitudes in the vis_cpu run. While pyuvsim agrees fairly well both healvis resolutions, the interpolated vis_cpu PSD is skewed at the 10^{-4} level. Another interesting aspect from these simulations is that none are free of artifacts that the 10^{-8} level so advances in simulation fidelity will be required validate time-domain techniques that must suppress systematics by greater than a factor of 10^{4} level from visibilities
Figure 1: Averaged PSD over two-hundred independent simulations in the fringe-rate domain for our various simulator choices. We see that substantial wings exist for our pixelized beam vis_cpu runs. We also see that the healvis nside=128 and healvis nside=32 yield similar fringe-rate profiles (the nside=32 profile is slightly wider) so we may get away with lower resolution simulations with ≈ 10k sources.
Table 1: Common simulation and fringe-rate filtering parameters shared by all simulations. We found that using the full time resolution of the HERA observations of 10s did not have a substantial effect on our results. Hence we stick with simulations with 100s time resolution in this memo.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Resolution</td>
<td>100 s</td>
</tr>
<tr>
<td>Fringe Rate Filter Standoff</td>
<td>0.05 mHz</td>
</tr>
<tr>
<td>Minimum Fringe Rate Filter Half Width</td>
<td>0.15 mHz</td>
</tr>
<tr>
<td>Lower histogram percentile for Fringe-Rate Limits</td>
<td>5%</td>
</tr>
<tr>
<td>Upper histogram percentile for Fringe-Rate Limits</td>
<td>95%</td>
</tr>
<tr>
<td>Number of Draws</td>
<td>200</td>
</tr>
</tbody>
</table>

Table 2: Parameters for our different simulation models. All simulations with 12288 sources and random source locations used the same sky model (indicated by the “matched” column. Our healvis simulations use sources distributed at healpix pixel positions. We draw the flux of each source from an exponential distribution with a mean of 2 Jy.

<table>
<thead>
<tr>
<th>label</th>
<th>simulator</th>
<th>locations</th>
<th>matched?</th>
<th>Nsources</th>
<th>beam values</th>
</tr>
</thead>
<tbody>
<tr>
<td>healvis_128</td>
<td>healvis</td>
<td>healpix</td>
<td>No</td>
<td>196608</td>
<td>healpix</td>
</tr>
<tr>
<td>healvis_32</td>
<td>healvis</td>
<td>healpix</td>
<td>No</td>
<td>12288</td>
<td>healpix</td>
</tr>
<tr>
<td>viscpu_bmpix_100</td>
<td>vis_cpu</td>
<td>uniform random</td>
<td>yes</td>
<td>12288</td>
<td>100 × 100 ortho</td>
</tr>
<tr>
<td>viscpu_interp</td>
<td>vis_cpu</td>
<td>uniform random</td>
<td>yes</td>
<td>12288</td>
<td>uvbeam interp</td>
</tr>
<tr>
<td>viscpu_hpx</td>
<td>vis_cpx</td>
<td>healpix</td>
<td>no</td>
<td>12288</td>
<td>u100 × 100 ortho</td>
</tr>
<tr>
<td>pyuvsim</td>
<td>pyuvsim</td>
<td>uniform random</td>
<td>yes</td>
<td>12288</td>
<td>uvbeam interp</td>
</tr>
</tbody>
</table>

2 Fringe Rate Filtering

We perform fringe-rate filtering by fitting time-domain slepian sequences to the data in time [Ewall-Wice et al., 2021, Slepian, 1978] using the select_main_lobe branch of hera_cal which can be found at https://github.com/HERA-Team/hera_cal/pull/736. This allows us to extract a rectangular region in fringe-rate space with maximal power concentration. To set this rectangular region, we histogram the amplitude of the primary beam squared on an equal-area healpix grid against instantaneous fringe-rates on the sky [Parsons and Backer, 2009, Parsons et al., 2016] and set upper and lower limits set by the 5% and 95% limits of the CDF of these beam squared histograms. We also add a small 0.05 mHz offset to these limits and set the minimum half width of our filter to be 0.15 mHz to account for time structures arising from the spatial variations in the primary beam. To avoid polarization confusion, we include both the NN and EE polarized beams in the histogram step so that the same filter which includes both pols, is applied to both polarizations.

We show the residuals of our visibility after applying the fringe-rate filter in FR-space (Fig. 2) and time-space (Fig. 3). With the exception of vis_cpx the residuals all lie on a similar level in fringe-rate space.

3 Signal Loss

To evaluate the level of signal loss in our measurements, we compute the retained power ratio

\[ R_{ret} = \frac{\sum_t |V_{frf}(t)|^2}{\sum_t |V(t)|^2} \] (1)

where \( V(t) \) is the visibility without fringe-rate filtering and \( V_{frf}(t) \) is the fringe-rate filtered visibility (DPSS components fitted to the data in time).

In Fig. 4 we show histograms of \( R_{ret} \) for our various simulators. We find that all simulation options recover the median and 68% confidence intervals within several percent. The fiducial filter setting achieves a median signal loss of 5% with 84% of instances yielding signal loss below 10% or so. We note that the distributions for Nside=128 and Nside=32 are quiet close to each other with relatively small differences in
Figure 2: The same as Fig. 2 except after applying a fringe-rate filter designed to removed.
Figure 3: Mean of absolute value squared residuals over time, normalized to the mean square of the unfiltered data. **Top:** NN pol, **Bottom:** EE pol.
the modes and medians (a few percent) so it may be computationally advantageous to use simulations with only \( \approx 10k \) sources rather than 300k.

## 4 A longer baseline

So far, we have only considered a relatively short baseline (twenty-two meters) to demonstrate convergence between different simulators and source count/resolution choices. However, we expect pixelization effects arising from pixelization to be more prominent in long baselines. To consider this, we also explore the baseline between antennas 0 and 128 which has a 13 meter EW component and a 97 meter NS component.

In Fig. 5 we compare averages of the PSD for baseline 0-128. We do not repeat this experiment for pyuvsim due to run-time constraints. We find that the Nside=32 and Nside=128 Healvis runs are in very good agreement.

### References


Figure 4: Distributions of signal retention for the fringe-rate filter on a 21 m EW 8 m NS baseline at 155 MHz. Vertical solid line indicates the median of the distributions and dashed vertical lines indicate the 68% confidence intervals for the retained power, $R_{rt}$. 
Figure 5: Same as Fig. 1 except now using the visibility formed from antennas 0 and 128; a 123 meter EW, 97 meter NS baseline. Nside=32 and Nside=128 pixel counts are still relatively close together for healvis.
Figure 6: Same as Fig. 4 but now for a 123 meter EW and 97 meter NS baseline.
