

# Reference Simulation Renovation

Mitchell Burdorf\*

Morgan Lee†

Jonathan Pober‡

September 8, 2025

## Abstract

This memo discusses the impetus and work performed for re-implementing the historical reference simulations due to data loss and output change. We run the reference simulations on historical and modern pyuvsim installs and discover that the output shifted significantly at multiple points in time. We then investigate changes to the reference simulations. We finally discuss alterations to the reference\_simulations directory of the pyuvsim repository, including fixing minor errors, documentation cleanup, and addition and removal of input files.

## 1 Introduction

The reference simulations for pyuvsim were developed from 2018 to 2021 with pyuvsim versions between 0.1.0 and 1.2.1. The simulations were developed to serve as an important tool for regression and integration testing – quoting the current README: “The goal of a reference simulation is to provide a simulated instrument output for a set of precisely defined inputs (sky model, antenna positions, primary beam, etc.) to serve as a point of comparison for other simulators and later versions of pyuvsim.” Unfortunately, the reference simulations were never successfully integrated into the CI/CD pipeline of pyuvsim as they were both computationally intensive and memory intensive, and have now been partially lost to time. The pyuvsim branch used to run the reference simulations no longer exists remotely, and the dependencies used to run the reference simulations were not fully documented. Multiple reference simulation output files used for comparison are now no longer accessible and thus those reference simulations cannot be used for any measure of regression testing. Additionally, the current state of the reference\_simulations repository contains broken functionalities, minor errors in test configuration, and out-of-date documentation. We have also recently identified failures in equality checking for multiple reference simulations. Due to test failures and the inability of the historical reference simulations to satisfy their original testing purposes, we perform a close investigation for changes responsible for modifying the reference simulation output, and justify a renovation of the reference\_simulations repository using well-documented simulation input and output.

## 2 Installation of “historical” pyuvsim versions and testing

We installed “historical” versions of pyuvsim in conda environments using Git and Python Package Index data. The term “historical” is used to indicate that all dependencies for the package existed at or before the time of release of the package. The Radio Astronomy Software Group packages pyuvdata and pyradiosky were pinned and their versions were iterated along with pyuvsim. We subsequently ran the 1.1\_uniform reference simulation across all the installed versions and computed statistical differences in the data between versions. We found that minor output changes had occurred for versions 1.1.2-1.2.0, 1.2.2-1.2.3, 1.2.4-1.2.5, and 1.3.0-1.3.1. These changes were all about a million times smaller than the major changes that have occurred to simulation output, which occur between versions 1.2.0-1.2.1 and 1.2.5-1.3.0. We can see these results via the comparison table 1, with explicit environment configurations in table 2. While we only display the results for 1.1\_uniform, we have successfully performed a similar comparison for all the useable first generation simulations starting from version 1.2.0 – with minor patches to get mpi working – with identical result. We were able to locate the specific alteration responsible for the entirety of the change in output between pyuvsim 1.2.5 and 1.3.0 through first isolating the package responsible – astropy – and creating a minimal set of environments for which the only difference was the difference in astropy version. We then performed a manual bisection of the version of the package for which the change occurred to identify the specific commit responsible.

---

\*mitchell\_burdorf@brown.edu

†morgan\_lee1@brown.edu

‡jonathan\_pober@brown.edu

vers	diff of abs				abs of complex diffs		outcomes	
	abs_mean	abs_var	rel_mean	rel_var	max_abs_dif	max_rel_dif	result	num_err
1.1.3-1.1.2	-8.74e-17	7.04e-29	6.10e-17	3.06e-28	1.14e-13	7.35e-13	True	0/15664
1.2.0-1.1.3	0.e+00	0.e+00	0.e+00	0.e+00	0.e+00	0.e+00	True	0/15664
1.2.1-1.2.0	-9.19e-11	2.99e-16	-9.15e-10	9.05e-16	7.51e-06	1.29e-06	True	0/15664
1.2.2-1.2.1	0.e+00	0.e+00	0.e+00	0.e+00	0.e+00	0.e+00	True	0/15664
1.2.3-1.2.2	7.81e-16	1.81e-27	2.68e-15	1.23e-26	4.74e-13	5.17e-12	True	0/15664
1.2.4-1.2.3	0.e+00	0.e+00	0.e+00	0.e+00	0.e+00	0.e+00	True	0/15664
1.2.5-1.2.4	-7.81e-16	1.81e-27	-2.68e-15	1.23e-26	4.74e-13	5.17e-12	True	0/15664
1.2.6-1.2.5	0.e+00	0.e+00	0.e+00	0.e+00	0.e+00	0.e+00	True	0/15664
1.3.0-1.2.6	-9.14e-10	5.05e-14	-1.18e-08	1.56e-13	7.96e-05	1.72e-05	False	2/15664
1.3.1-1.3.0	-7.23e-16	8.80e-28	-2.73e-16	4.12e-27	4.76e-11	5.24e-12	True	0/15664

Table 1: Comparison statistics of relative installs, with each version compared to the release preceding it. Output is for ref\_1.1\_uniform. The first major column is statistics from taking the absolute of each complex visibility, then taking differences between each set of visibilities in the two output arrays. The second major column is maximum difference from taking the absolute of the subtracted difference between each pair of complex visibilities for the output arrays. The final column is the comparison result, along with the number of visibility pairs that are not `numpy.isclose` with relative error set to  $1e-5$  and absolute error set to  $1e-8$ .

pyuvsim	numpy	pyradiosky	python	pyuvdata	h5py	pyyaml	scipy	astropy	astropy-healpix
1.1.2	1.17.3	0.0	3.7.16	1.5.0	2.10.0	5.1.2	1.3.2	3.2.3	0.5
1.1.3	1.18.1	0.0.2	3.7.16	2.0.2	2.10.0	5.3.1	1.4.1	4.0.1.post1	0.5
1.2.0	1.18.5	0.1.0	3.7.16	2.1.0	2.10.0	5.3.1	1.5.0	4.0.1.post1	0.5
1.2.1	1.21.6	0.1.2	3.9.19	2.1.3	3.3.0	5.4.1	1.7.1	4.3.1	0.6
1.2.2	1.21.6	0.1.2	3.9.19	2.1.5	3.6.0	6.0	1.7.3	5.0	0.6
1.2.3	1.22.3	0.1.2	3.9.19	2.1.5	3.7.0	6.0	1.8.0	5.0.4	0.6
1.2.4	1.22.4	0.1.2	3.10.4	2.2.8	3.7.0	6.0	1.8.1	5.0.4	0.6
1.2.5	1.20.3	0.1.2	3.8.19	2.2.10	3.1.0	5.1.2	1.3.3	5.0.4	0.6
1.2.6	1.20.3	0.2.0	3.8.19	2.2.10	3.1.0	5.1.2	1.3.3	5.2	0.6
1.3.0	1.26.4	0.2.0	3.10.14	2.4.3	3.11.0	5.4.1	1.14.0	6.0.0	1.0.2
1.3.1	2.0.0	1.0.1	3.12.4	3.0.0	3.11.0	6.0.1	1.14.0	6.1.1	1.0.3

Table 2: Documented dependency versions for the specific install environments (python-casacore and lunarsky left out).

What we found is that the change between pyuvsim version 1.2.5 and 1.2.6 occurred for astropy commit `0cac669b85` due to their International Earth Rotation and Reference Systems Service Bulletin B file changing. The monthly updates to this file were enough to cause our reference simulation equality check to fail. This is not incredibly surprising as we require a lot of consistency in our output comparison, and the comparison prescription we use is the default for `np.allclose`. This change is desirable because it is a correction of our output.

We similarly localized the entirety of the change between pyuvsim versions 1.2.0 and 1.2.1 to be due to astropy. Unfortunately, the time between version 1.2.0's release and version 1.2.1's release is quite long, and there were multiple occurrences of significant change in the simulation output all localized to astropy. We have not tracked down the specific commits responsible, but have decent evidence that same Bulletin B file is again responsible – via following the file through git. As we can show that all significant deviations in the simulations output for the first generation reference simulations have occurred external to packages within the Radio Astronomy Software Group, and we already have strong impetus to update the reference simulation output due to the IERS corrections, we consider our current results sufficient evidence of consistency over time.

Lastly, we have verified that the output of the reference simulation `2.2_uvbeam` has also remained consistent over time. The only changes identified were again astropy and an additional deliberate change in how we handle the `"min_flux"` config parameter – we swapped from using `>` to `>=` when identifying sources to keep.

### 3 Specific repository issues

In addition to the issue of the reference simulation output changing over time, we also identified a number of problems with the `reference_simulations` repository of pyuvsim. In short, these are:

- The various scripts to download and run simulations are outdated or broken
- We appear to be using the wrong version of the `"two_distant_points"` text catalog file
- When the simulations were first ran, we mistakenly used ICRS in place of J2000 for coordinates
- Out of date / incomplete documentation
- We no longer have access to the historical output for many reference simulations
- Some of the historical output from reference simulations has been insufficiently documented to be usable

### 4 Conclusion

We were able to accurately reproduce all of the currently existing historical reference simulation output for eight of the first generation reference simulations and one of the second generation reference simulations, and demonstrate that the major output changes which occurred from 1.2.0-1.2.1 and 1.2.5-1.3.0 could be entirely accounted for by astropy. We were able to identify the exact commit responsible for changing the output and causing our tests to fail in the 1.2.5-1.2.6 case, and learned that the change was due to the International Earth Rotation and Reference Systems Service updating their earth orientation parameters – Bulletin B – for increased accuracy. Evidence points to this being true for the 1.2.0-1.2.1 case as well, though we didn't extract the specific astropy commits responsible for the changes. As the major output changes are

- (a) only due to astropy,
- (b) an updating of the earth reference frame to have more accurate data (as far as we can tell with strong evidence to the positive),

we conclude that development of pyuvsim, pyuvdata, and pyradiosky since 2020 has been physically consistent, and that the change in the output is desired. We recognize that there could be another change in astropy responsible for significant drift in the output, though that is unlikely and much less serious than a change in the output of running pyuvsim due to code changes in a repository maintained by the Radio Astronomy Software Group. As we have also verified the remaining historical second generation reference simulation data, we will now move to fix and refactor the `reference_simulations` repository of pyuvsim.