Estimating the burst arrival time offset between SPOTLIGHT's Beam data and Visibility data

Raghav Wani, SPOTLIGHT Team. 21st August 2025

Using the precise localisation of FRBs, host galaxies are studied to weigh the possibilities of progenitor models. Precise sub-arcsecond localisation of FRBs is also useful for searching multi-wavelength counterparts that can differentiate between the progenitor models.

Real-time imaging from SPOTLIGHT requires simultaneous dumping of visibilities as soon as the burst is triggered in SPOTLIGHT's transient detection pipeline. The visibility data is then de-dispersed at the candidate DM, and only the part containing the burst, including 10% overhead on either side, is saved to disk. This allows FRBs to be localised in real time to enable rapid triggers essential for multi-wavelength follow-ups. In order to achieve this, there should be no time lag between the burts' time of arrival (ToA) in beam and visibility data. However, the offset reported earlier between SPOTLIGHT's beam data and visibility data ToA was >200 ms. Dedicated tests were conducted to verify this offset, and this report summarizes the methodology used for offset calculation along with the results obtained from three observations recorded in July–August 2025.

Methodology

The visibility data dumping program records 16 raw visibility files, corresponding to 16 nodes. Each file contains complex visibilities (baseline correlations) in half-precision float format, organized in time slices of records. To make an offline PC beam from these visibility files, Arpan Pal developed a Python script that does the following:

– Reads raw visibilities from the 16 files.
– Applies calibration, which flattens the bandpass and subtracts the baseline.
– Performs beamforming by coherently summing cross-correlations across valid baselines.
– Outputs a time–frequency dynamic spectrum (beamformed data).

This produces a time–frequency beamformed data stream, saved as a `.dat` file with the structure:
Header (44 bytes): 5 doubles (freq start, freq end, channel width, integration time, processing timestamp) + 1 integer (channels).
Data Records: Each record contains an 8-byte timestamp (double, Unix time with microsecond precision) followed by 4096 × 4 bytes of beam power data (32-bit floats, little-endian). Each record represents 1.31072 ms of integrated beam data across all frequency channels, with timestamps in chronological order reflecting the time-multiplexed structure of the 16 input files.

The data from the `.dat` file is then used to create a filterbank file, suitable for PRESTO analysis. The filterbank created contains dummy metadata and visibility data from 16 raw visibility files as a 2D array. The beam data recorded after beam formation is also converted into a filterbank file using the `xtrac2fil` program. Thus, two filterbank files are created—one from visibility data and another from beam data.

Next, both of these filterbank files are de-dispersed at the pulsar DM using PRESTO's `prepdata` module, resulting in de-dispersed time-series data stored in a `dat` file with the start time set to zero.
As visibility data is dumped only after a burst is triggered, the absolute start time of both time series is not the same. To correct for this initial time delay, the timestamps reported in the raw data files were used. For visibility data, the Unix time corresponding to the first sample in the raw visibility file was used, while for beam data, the timestamp of the first sample (in MJD) stored in the header file (`ahdr`) was used. After applying this initial offset, the overlapping chunk of time series common to both was cross-correlated using the `numpy.correlate` function from the NumPy library. The function computes the full discrete linear cross-correlation of the two input sequences and returns an output array that includes all possible overlaps between them, including those where only partial overlap occurs at the beginning and end. The lagged time corresponding to the maximum correlation gives the offset between the two time series.

Data Accumulation

Data used for analysis were recorded for three different pulsars during the dedicated test time slots. The mean flux density of these pulsars spanned 3 orders of magnitude. The table below summarizes the observation details for all three datasets.

Observation Date	19th July 2025	1st August 2025	8th August 2025
Observer	Sanjay	Jyotirmoy, Mekhala	Jyotirmoy, Mekhala
Observing Band	4 (550 - 750 MHz)	3 (300 - 500 MHz)	4 (550 - 750 MHz)
Source Name	J0332+5434	J0528+2200	J0534+2200
Source DM (pc cm^{-3})	26.76410	50.8695	56.7712
Source Period (s)	0.7	3.7	0.0334
Source Flux (mJy)	1500	57	550
Observation length (Beam data)	7.8 min	9.5 min	9.5 min
Observation length (Visibility data)	2.41 min	36.7 sec	36.7 sec

The `.dat` file containing the time-series data was created as outlined in the *Methodology* section. While forming the PC beam from visibility data, antenna masking was taken into account.

For beam data, filterbank files corresponding to the highest SNR beam were used for analysis, as determined from the SNR maps shown in Figures 1 to 3. The expected SNR pattern serves as a verification of the correctness of the data used for analysis.

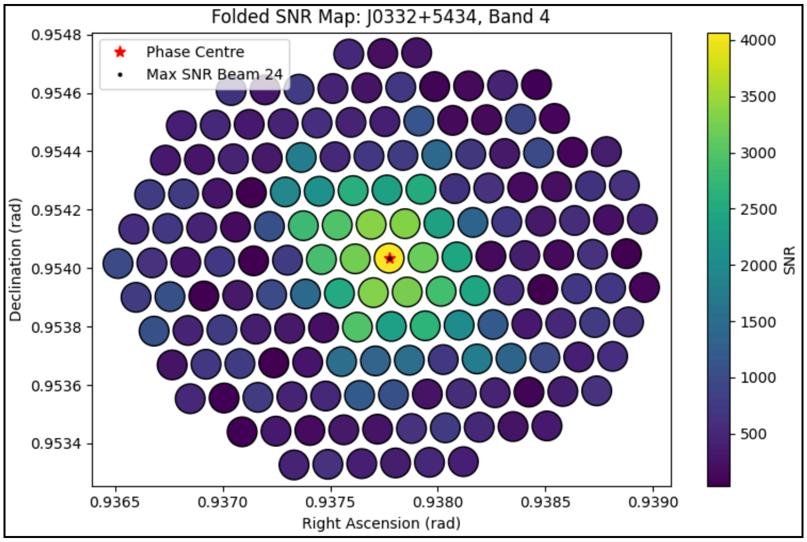


Fig 1: *J0332+5454, band 4 19th July 2025*

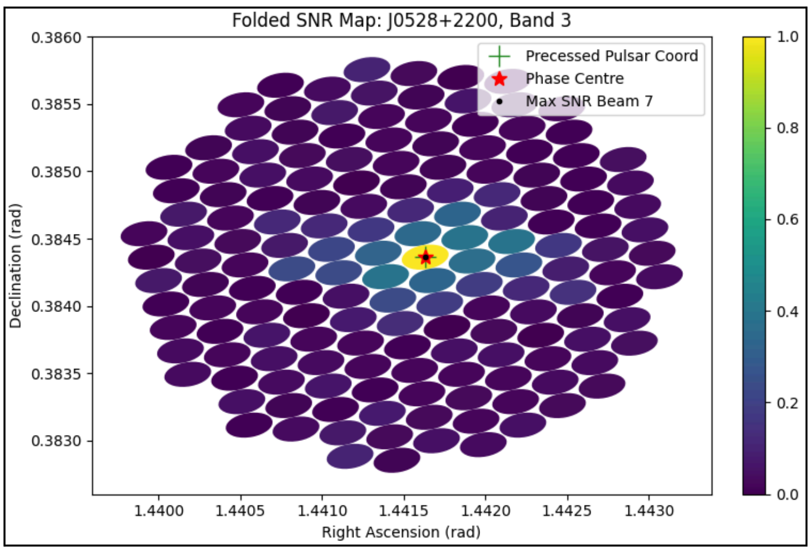


Fig 2: *J0528+2200, band3 1st August 2025*

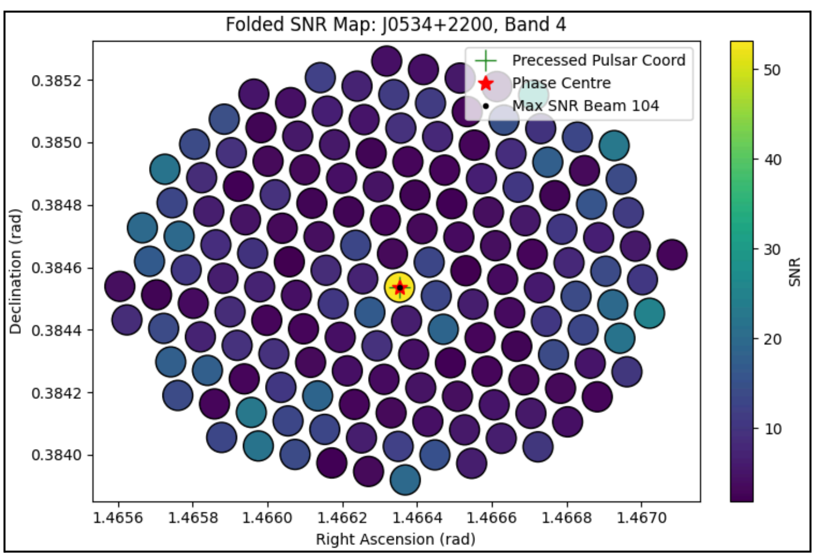


Fig 3: *J0534+2200, band4 8th August 2025*

Test results

As visibility data is stored in half-precision float format (16-bit) and beam data in quarter-precision float format (8-bit), a baseline difference is expected in the two time series, and this is also evident from the plots (difference of order 2) shown in Figure 4. After multiplying the beam data by 100, the baselines match approximately.

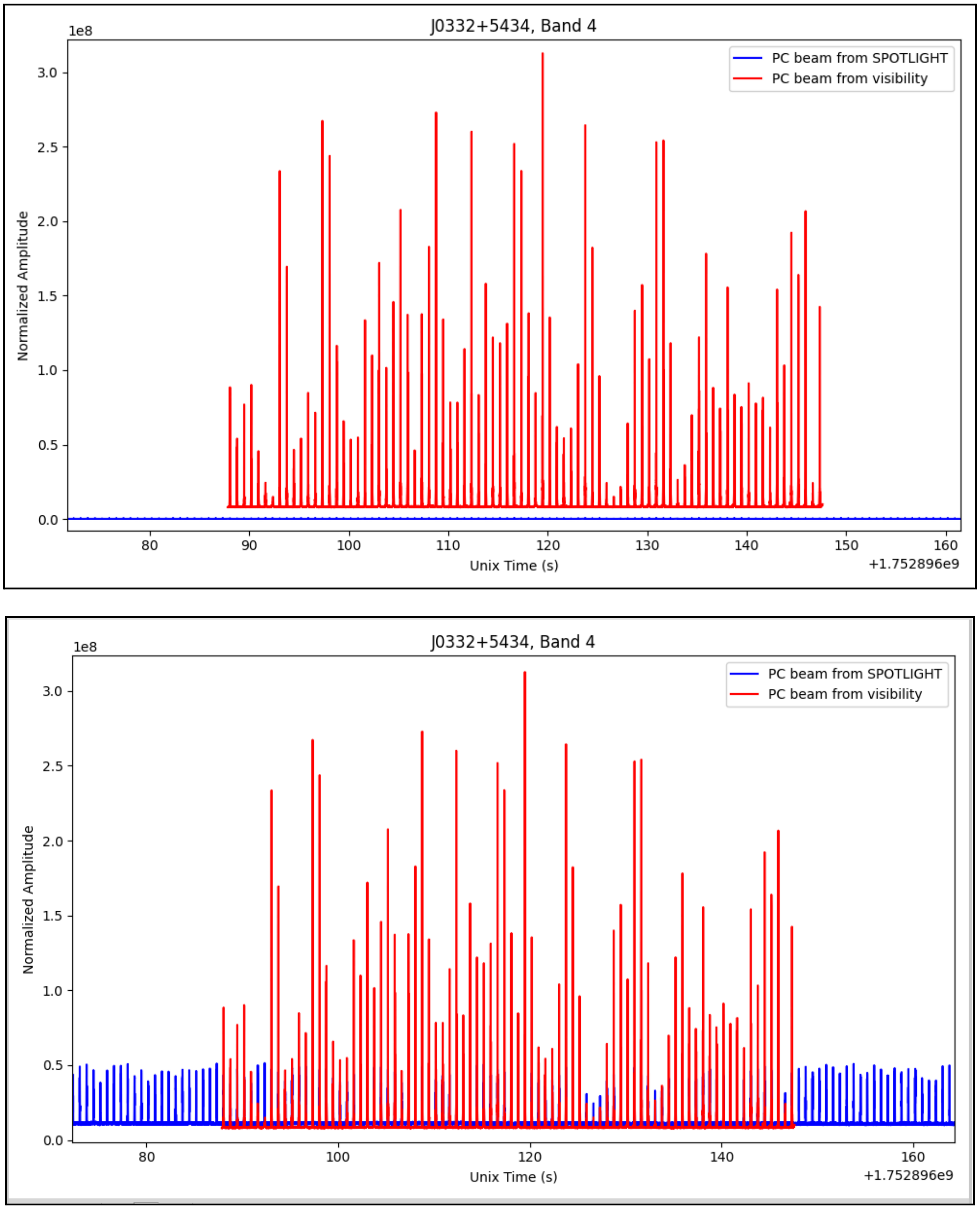


Fig 4: *Baseline RMS variation in beam data and visibility before (top) and after (bottom) multiplying all beam data points by a factor of 100.*

To correct for this, both time series were normalized before cross-correlation to estimate the burst arrival time offset. The normalized time series are overplotted in the figure 5.

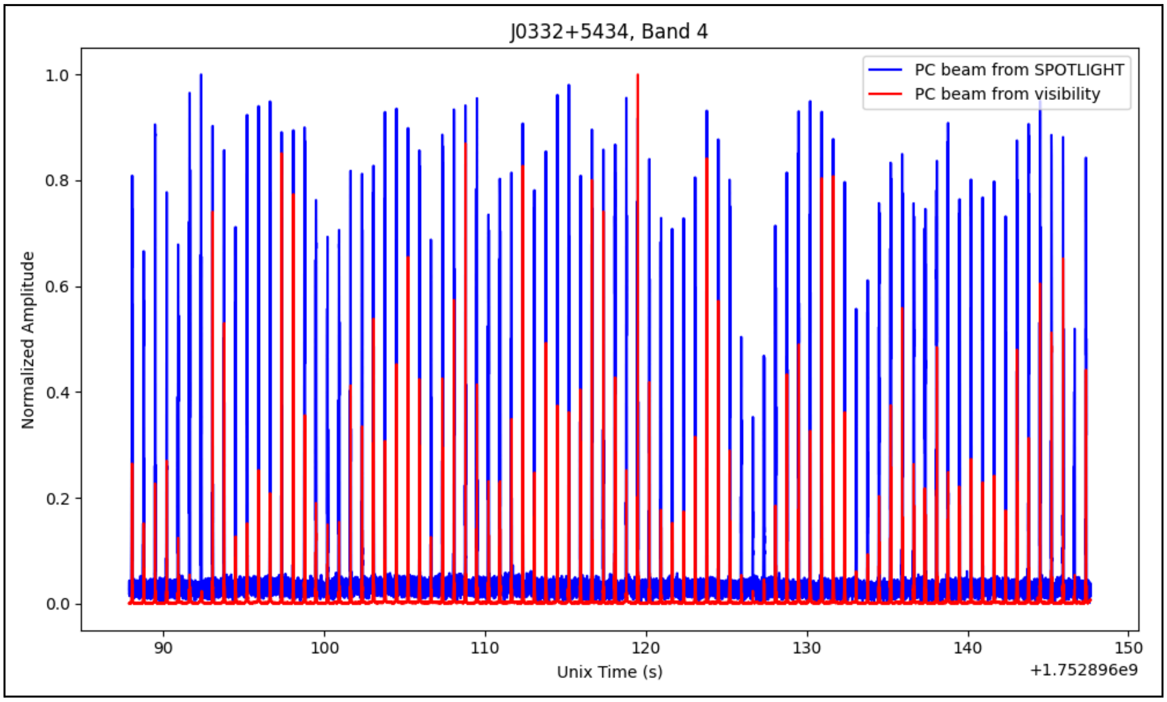
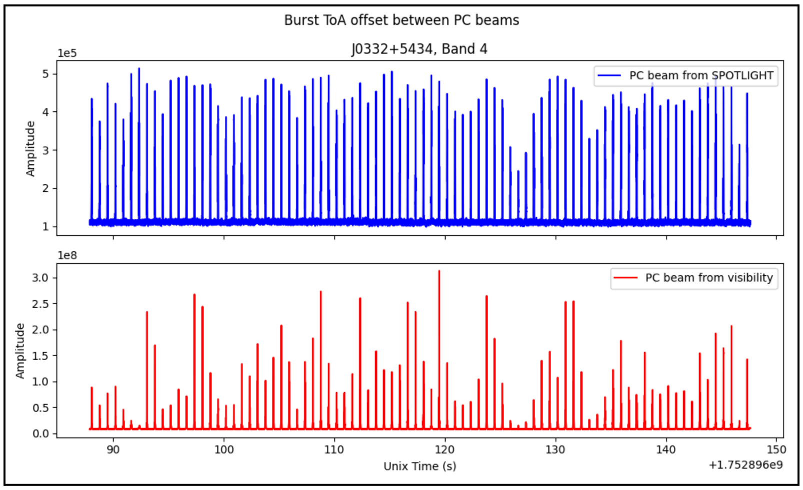


Fig 5: *Normalized timeseries from beam data and visibility data*

The cross-correlation was performed on normalized time series in two ways—first using the entire time series, and second using a cropped time series containing a single or a few bursts. The resulting burst arrival offsets for all datasets are summarized in the table below. The unnormalized time series and cross-correlation versus lag plots are shown in Figures 6 to 14 for all three datasets, covering different scenarios.

Observation Date	19th July 2025	1st August 2025	8th August 2025
Source Name	J0332+5434	J0528+2200	J0534+2200
Start time in beam data (in MJD)	60875.148679717953	60888.343137086100	60895.207071288321
Start time in visibility data (in MJD)	60875.1491651698	60888.34829501202	60895.21176803499
Start time offset (in sec)	41.9430396556	445.6447997093	405.7989122868
ToA offset considering the entire timeseries (in msec)	1.310720	7.864320 *6 times 1.310720*	14854.389760 *11333 times 1.310720*
ToA offset considering a few bursts (in msec)	**0.00**	**1.310720**	**2.621440** *2 times 1.310720*
ToA offset considering only a single burst (in msec)	**0.00**	**1.310720**	**1.310720**

19th July 2025



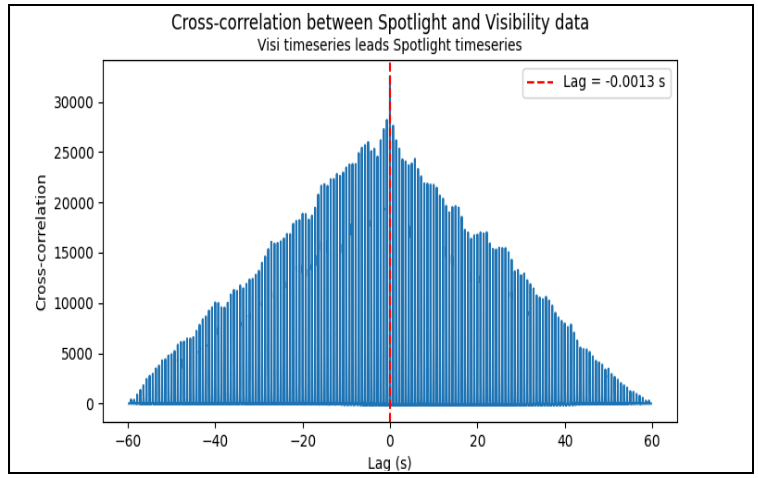
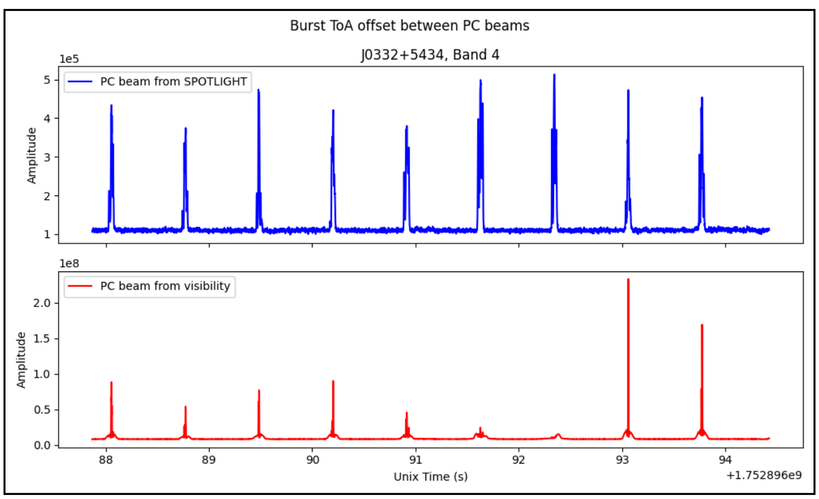


Fig 6: *The entire time series and corresponding cross-correlation. Offset = 0.001310720 sec*



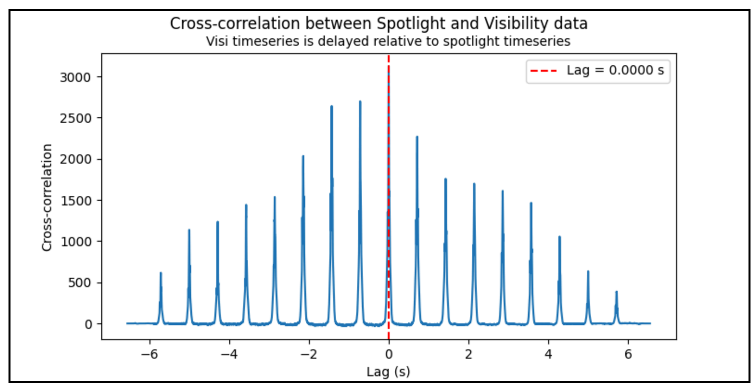
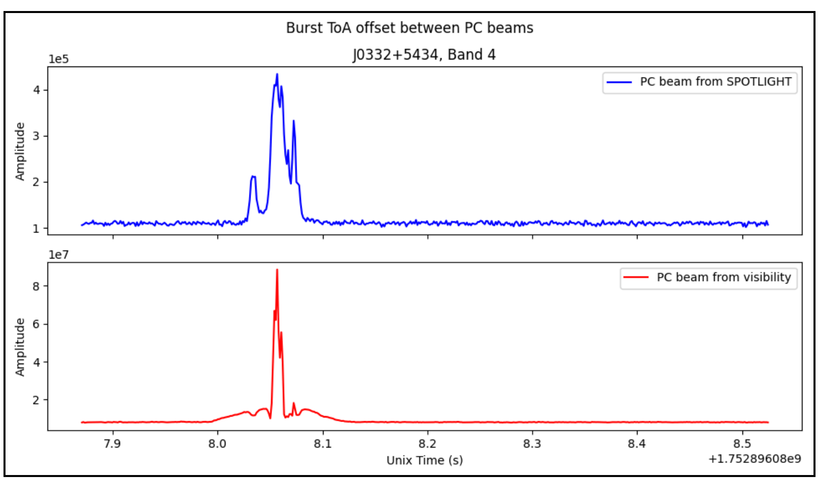


Fig 7: *Cropped time-series and cross-correlation for a few bursts. Offset = 0sec*



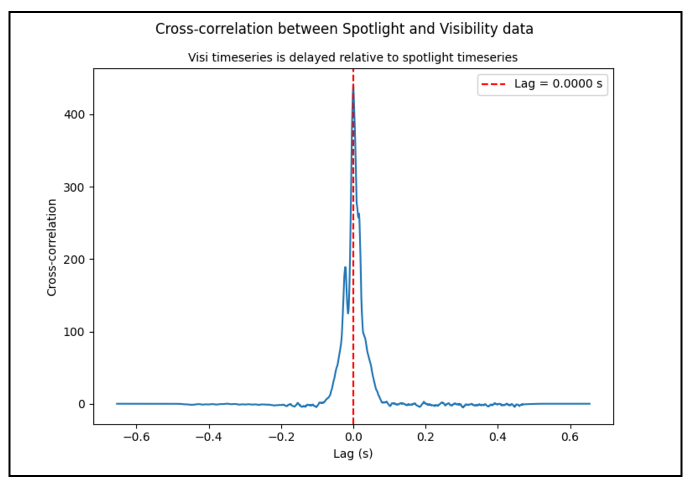
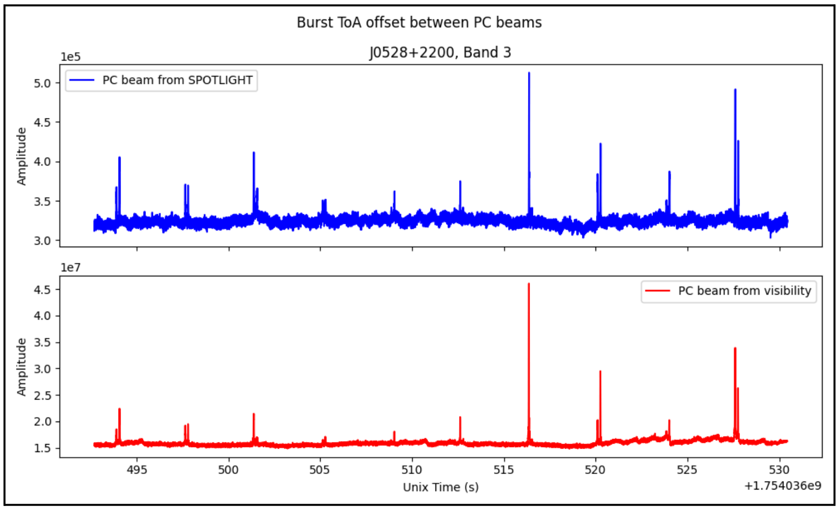


Fig 8: *Cropped time-series and cross-correlation for a single burst. Offset = 0.00 sec*

1st August 2025



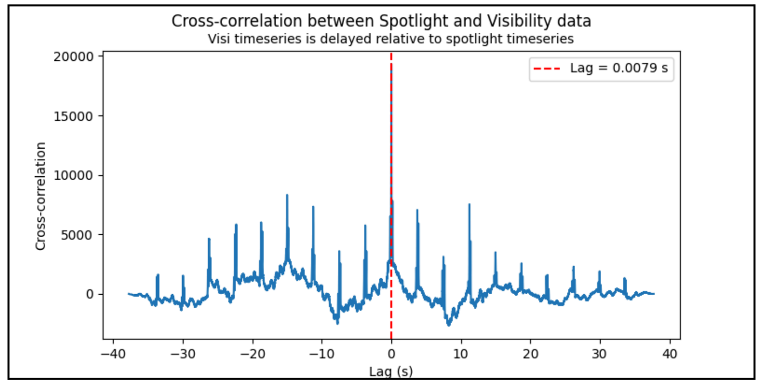
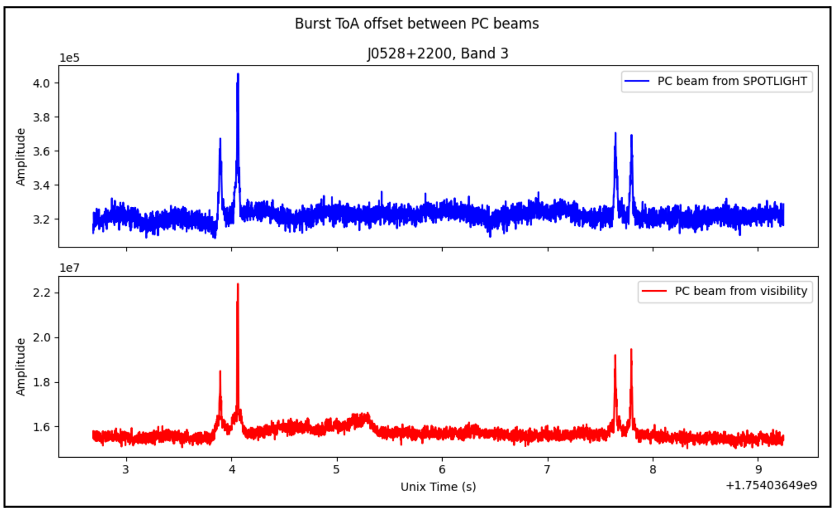


Fig 9: *The entire time series and corresponding cross-correlation. Offset = 0.007864320 sec*



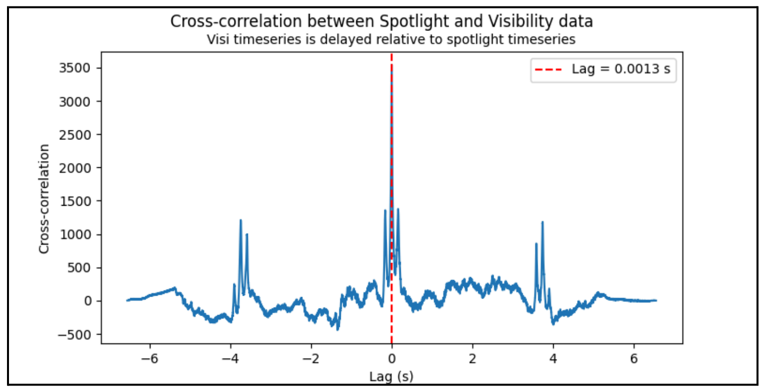
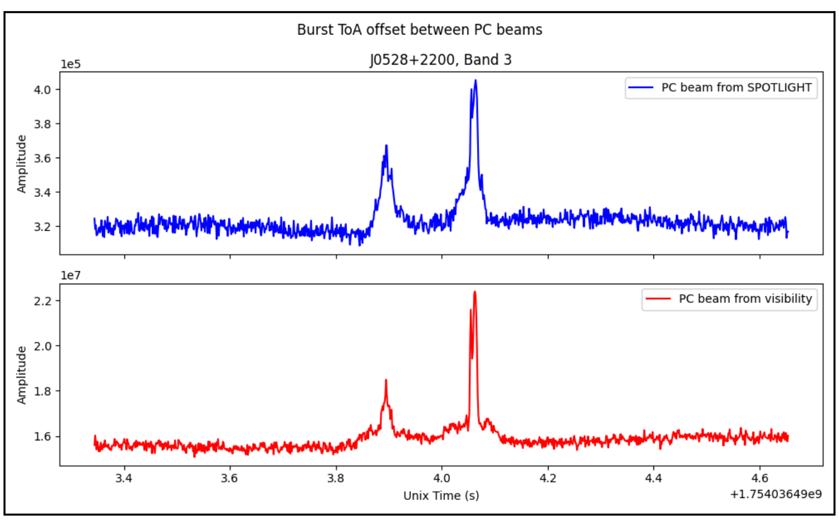


Fig 10: *Cropped time-series and cross-correlation for a few bursts. Offset = 0.001310720 sec*



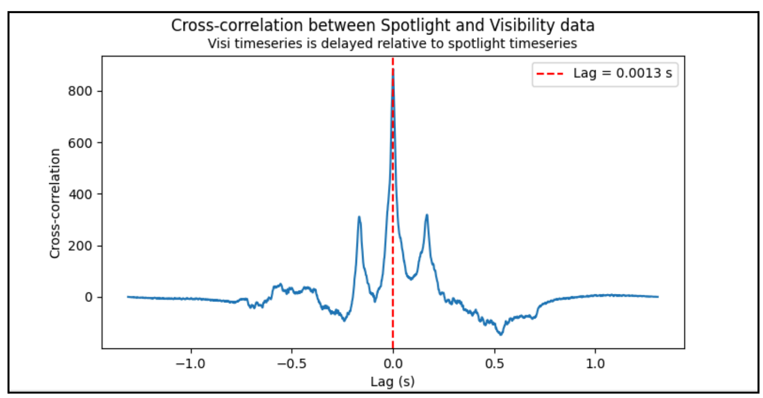


Fig 11: *Cropped time-series and cross-correlation for a single burst. Offset = 0.001310720 sec*

8th August 2025

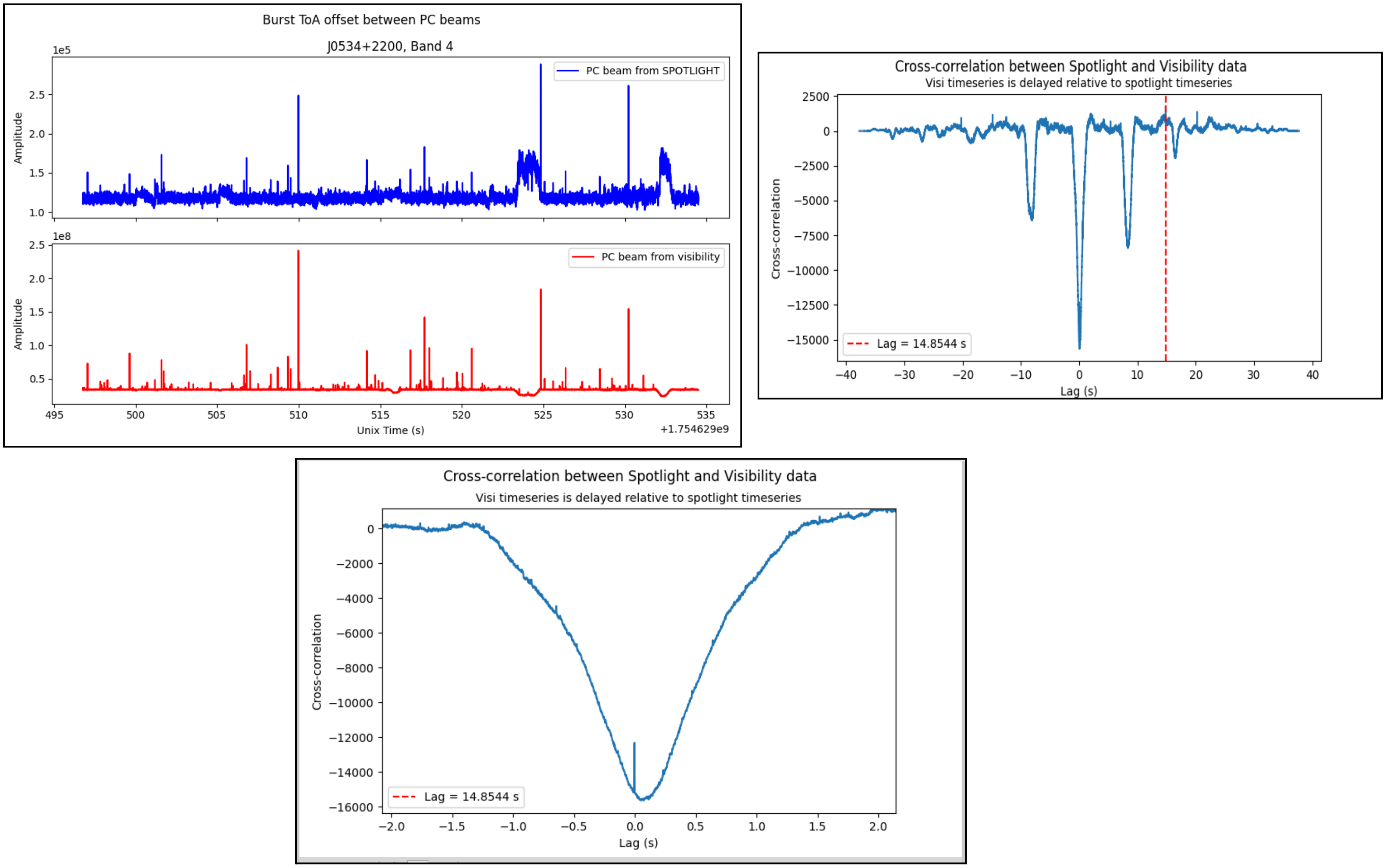
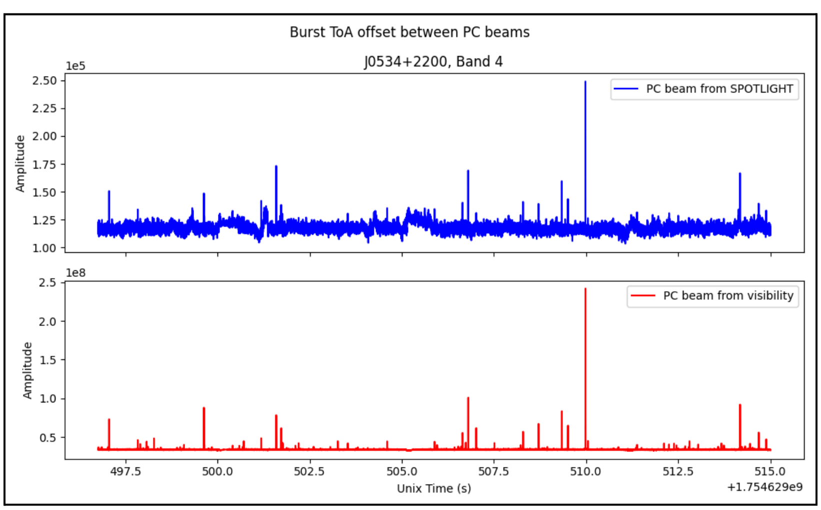


Fig 12: *The entire time series and corresponding cross-correlation. Offset = 14.854389760 sec*



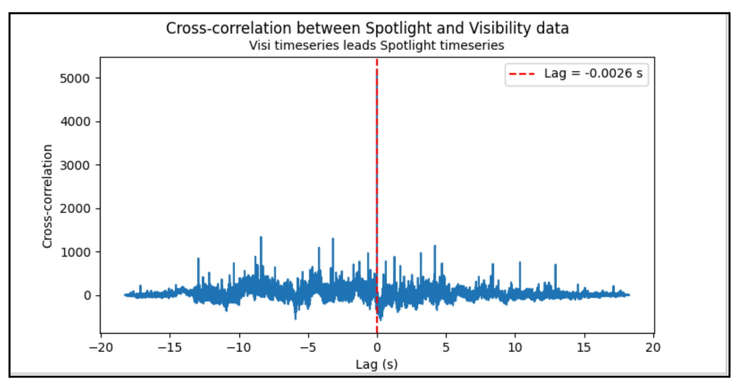


Fig 13: *Cropped time-series and cross-correlation for a few bursts. Offset = 0.002621440 sec*

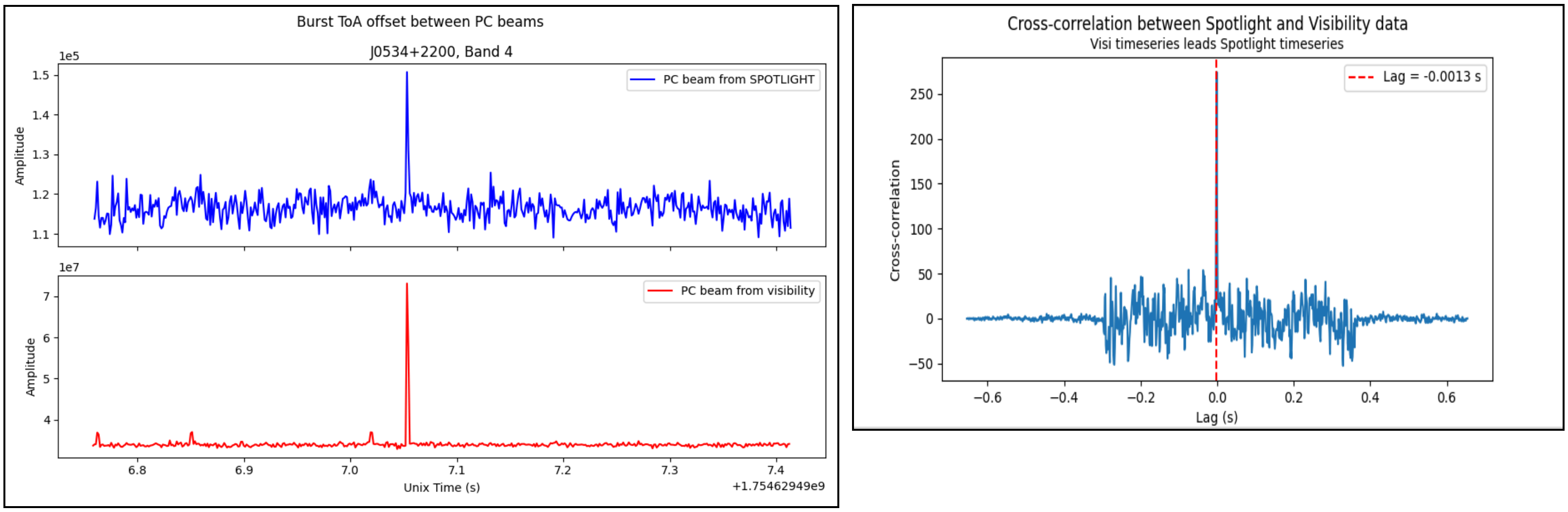


Fig 14: *Cropped time-series and cross-correlation for a single burst. Offset = 0.001310720 sec*

To verify that we are applying the correct initial offset and thus essentially calculating the arrival offset between the same burst, we also compared the burst profiles from SPOTLIGHT's beam data and visibility data with GWB's PC beam data, as shown in Figure 15.

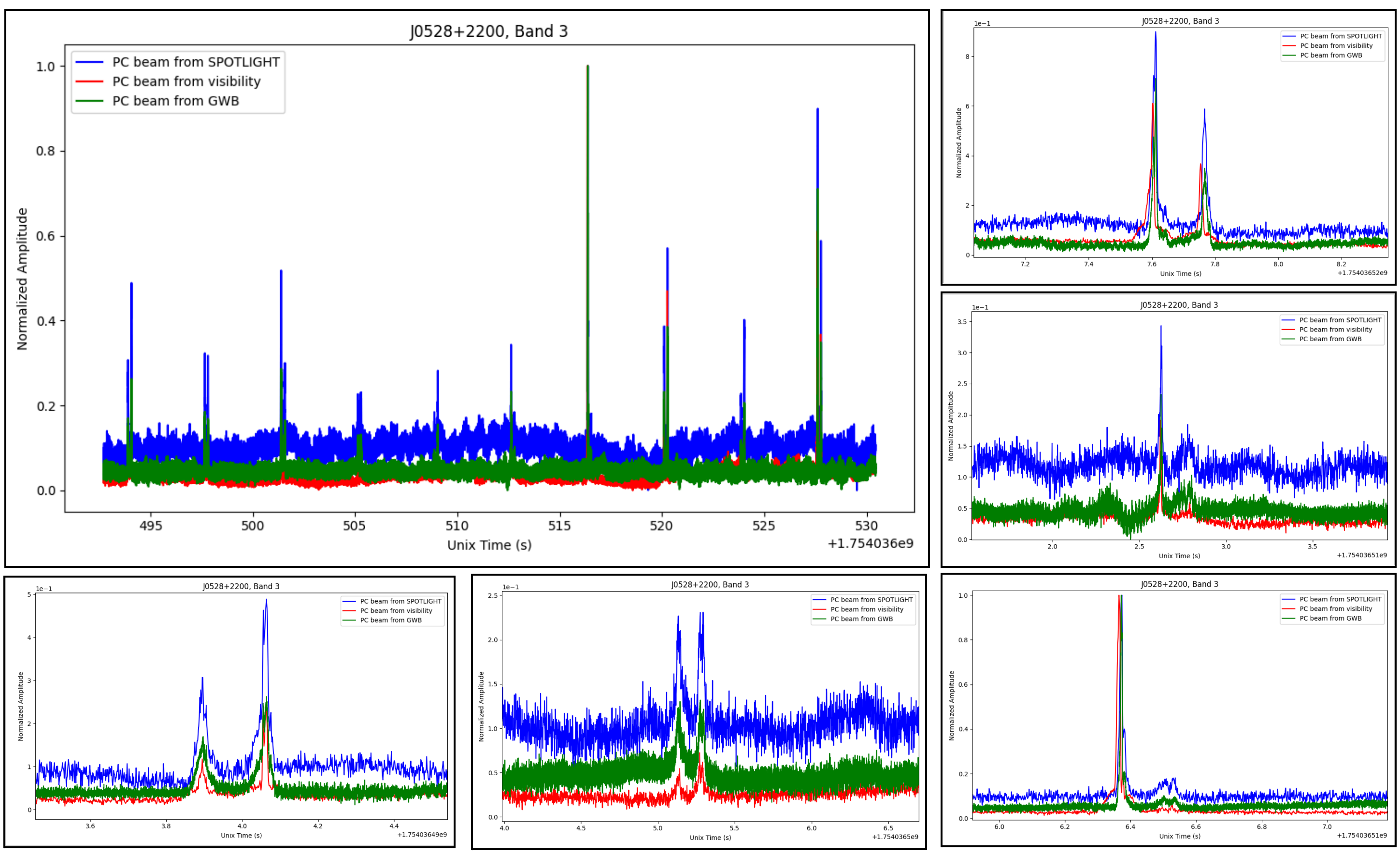


Fig 15: *Burst profiles for PSR J0534+2200 from SPOTLIGHT and GWB data*

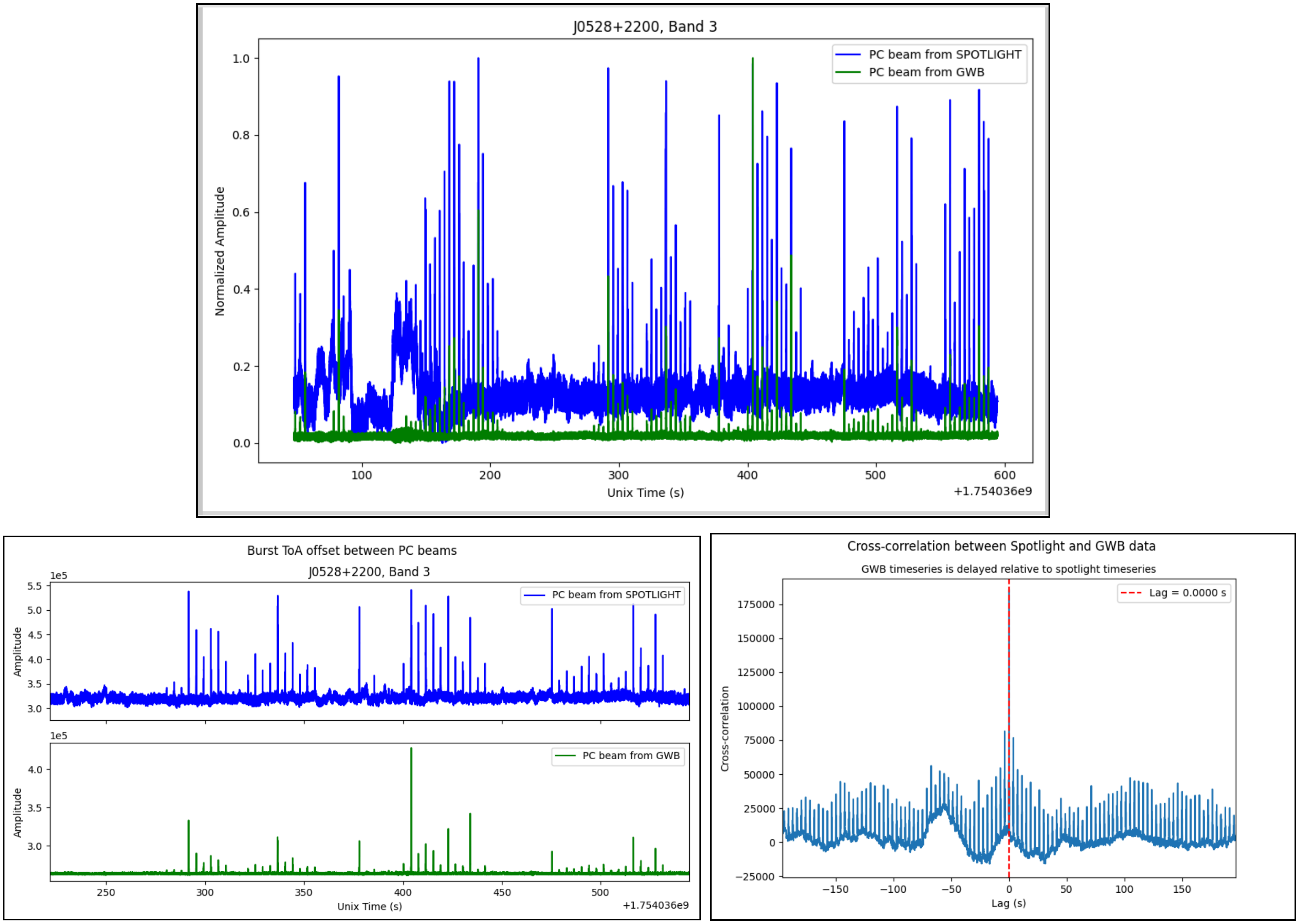


Fig 16: *Overplot of normalized timeseries of SPOTLIGHT's and GWB's PC beams (Top); Cropped time-series (Bottom left) and cross-correlation (Bottom right) for a few bursts. Offset = 0.00 sec*

The cross-correlation between SPOTLIGHT's and GWB's PC beams resulted in no offset (as shown in Figure 16), thus confirming our claim of cross-correlating the same bursts.

Conclusion

1. **No intrinsic timing offset is observed between SPOTLIGHT's beam data and visibility data.**
2. The offsets measured (1.31072 ms or multiples) correspond exactly to the sampling time and arise solely from cross-correlation binning and baseline/statistical variations.
3. Cross-correlation of cropped burst segments provides more accurate offset estimates than using the full time series (since baseline noise dominates otherwise)
4. Time corresponding to the start of the visibility block should be considered for the offset calculation instead of the burst triggering time.
5. After normalization and cropping around bursts, both time series align within a single sample, confirming that the previously reported >200 ms lag was not real.
6. Comparing SPOTLIGHT (beam and visibilities) with the simultaneously recorded GWB PC beam shows good agreement with no offset in burst ToA, validating both the offset estimation methodology and the correctness of the data streams.

Further problems to discuss

1. Reason for the difference in RMS variation in visibility and beam data baselines.
2. Robustness of cross-correlation-based time offset estimation.

References

1. Path to the directory containing all the codes and the test data (check *README* for more info): `/lustre_archive/spotlight/raghav/visibility_vs_beam`
2. A real-time imaging localisation pipeline for the SPOTLIGHT, URSI-RCRS 2024, ARIES/GEHU, Bhimtal, India, 22 - 25 October 2024