Towards Imaging with SWAN

Project Report

By: **Satyapan Munshi**
2nd year BS-MS student
IISER Mohali

Supervised By:
Avinash Deshpande

Raman Research Institute
Bangalore

Synopsis

In this project we tried to verify the tile coordinates of SWAN as obtained from gps measurements by monitoring the Sun. We used data of the Sun at 200 MHz frequency taken using the Sky Watch Array Network at Gauribidanur observatory of Raman Research Institute. After suitable integration and RFI affected channel rejection, the phases that arise as a result of the geometric delay between two tiles were used to monitor the changes in the observed delay. They were compared with the expected delay changes due to the Earth's rotation during the period in which the observations were taken. The offsets in the expected and observed delay plots were utilized to estimate the errors in baseline lengths that had been used to estimate the expected delays between tiles.

Contents

1 Introduction

1.1 SWAN - Sky Watch Array Network

The SWAN system uses an array of Murchison Wide-field Antenna. It consists of 8 tiles each containing 16 aluminum dipoles arranged in a 4X4 matrix. The whole setup is located at Gauribidanur, near Bengaluru. Each tile is attached to a Beamformer which gives appropriate delays (electronically) to the antennae in a tile in order to point to a source. The received data are sampled at 33 MHz and arranged in packets of 1024 samples (512 samples each for the 2 channels of a das) and stored in hexadecimal format along with the relevant information stored as headers. There are two modes of observation: obs mode and sniff mode. In obs mode, all the packets are stored. In sniff mode, for each das, one out of 1000 packets are stored and data from all dases are stored in a single file. In this analysis, sniff mode observations were used since in this mode, significant changes in delay can be observed by analysing a lower amount of data.

1.2 Motivation and Objective

The radio waves from a celestial object do not fall on all the tiles simultaneously (unless the object is at zenith) and hence there is a delay in the arrival times at the any two tiles. This delay leads to a linear phase gradient across the frequency channels. As the earth rotates, the geometric delay between the tiles changes and hence the phase gradient changes. If the RA and Dec of the object, tile coordinates, observation date and time, central frequency of observation are known, then we should be able to simulate how the delay and phase are expected to change with time due to the Earth's rotation. This can be cross-checked with the observed delay changes from the data. If the residual delays show a trend, then it can be interpreted as arising due to lack of accuracy of the tile coordinates that were used to simulate how the delays should change with time.

2 Data Used

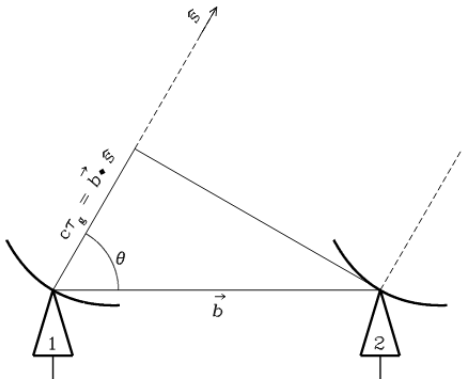
The data used for the analysis were 1 hour observations of the Sun taken using the SWAN on 20/02/2017 at a central frequency of 200 MHz. Four das's (data acquisition systems) were active during the observations. Also, the two channels of a das, instead of being connected to the two polarization components of the same tile, were connected to the same polarization component of two different tiles so that the computations of cross correlations are simpler.

3 Analysis

The .mbr file containing the data was analysed using python 2.7. Several python packages (numpy, scipy, astropy, etc.) were also utilized during the analysis.

3.1 Simulating Expected Delays

The RA and Dec of the Sun on the date of observation were obtained from online records. The tile coordinates were available from gps measurements. The baseline vectors were computed as two dimensional vectors in the xy plane of East-North-Up coordinate system taking one of the tiles as reference. The RA, Dec of the Sun on that date and the observation time were used to get the altitude and azimuth of the Sun at Gauribidanur. A unit source vector was computed for each time point and is given by: (cos(alt)cos(az), cos(alt)sin(az), sin(alt)). If the baseline vector is given by $\vec{b}$ and the source vector is $\vec{s}$ then the geometric delay is given by $\frac{\vec{b}.\vec{s}}{c}$ where c is the speed of light.



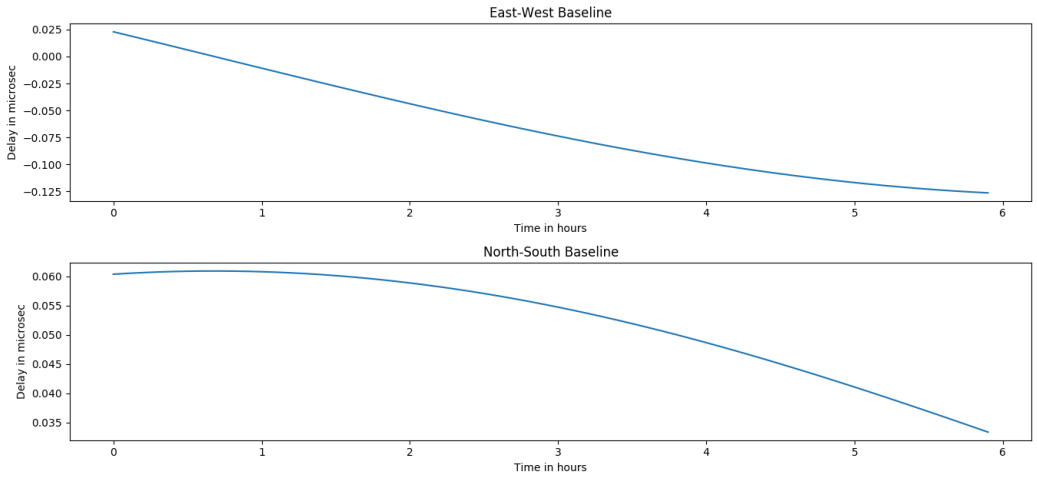


Figure 1: Simulated Expected Delays for two specific baselines

Here we see that the changes in delay over 6 hours for the North-South baseline is much lower than those seen for the East-West baseline. Also, the delay for the East-West baseline becomes 0 at the point when the source crosses the meridian. At that same time, the delay for the North-South baseline passes through an extremum which is also expected.

In this analysis, we assume that the RA, Dec of the Sun remains the same during the observation. This can be justified on the following grounds. We know that RA, Dec of the Sun changes by 5' in a day, i.e., 12.5" in an hour. Hence we simulate the expected delay vs time plots for an hour for two values of RA,Dec differing by 15". The results of a least squares fit to the plots are as follows:

```
Slope:                -0.000551969
Intercept:             0.0333055
Error in slope:        4.49699e-07
Error in intercept:    1.55232e-05
```

(a) Least Squares Fit - 1

```
Slope:                -0.000551676
Intercept:             0.0332888
Error in slope:        4.50725e-07
Error in intercept:    1.55586e-05
```

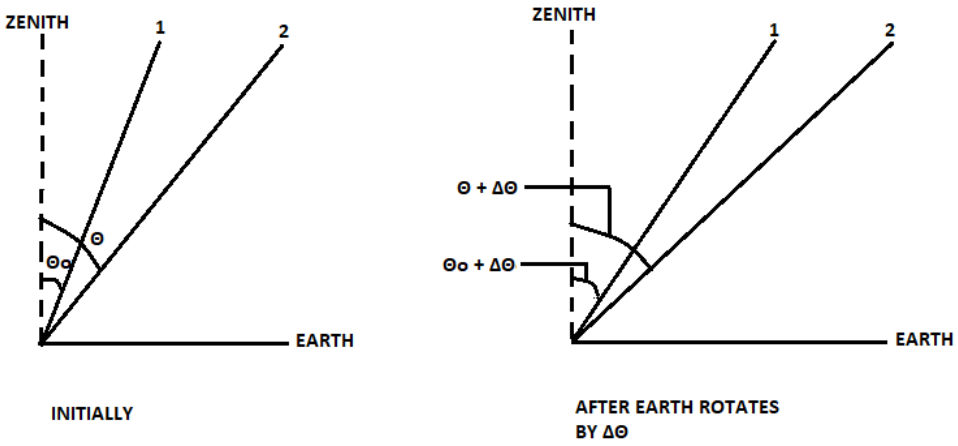
(b) Least Squares Fit - 2

Hence we see that the difference in computed slopes is 2.93×10^{-7} which is less than the error bars in the computed slopes that will be used for refining baseline lengths. Hence the change in RA, Dec of the Sun in an hour will not be a factor in the analysis for 1 hour observation.

3.2 Selecting Integration Time

In order to build up SNR, the data needs to be integrated for some time. But too high an integration time may lead to bandwidth decorrelation due to uncompensated delay. Hence selecting the appropriate integration time is very important in order to get the maximum SNR without losing too much information. So we need to know how fast the correlation is changing and, given that we have compensated for the central direction, what should be the maximum integration time such that sources farthest from the centre of the beam do not cause decorrelation due to uncompensated delay.

3.2.1 Method 1

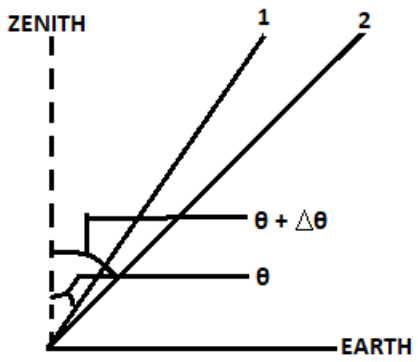


Let us assume that we can tolerate a 10° phase within a channel due to uncompensated delay and that we have compensated for delay at a direction θ_o from zenith. So, phase difference for a source at θ given by:
$\phi_o = \frac{2\pi d}{\lambda}(\sin\theta - \sin\theta_o)$ where λ is the wavelength of radiation.
After the earth rotates by $\triangle\theta$, the phase difference is given by:
$\phi = \frac{2\pi d}{\lambda}[\sin(\theta + \triangle\theta) - \sin(\theta_o + \triangle\theta)]$
How fast this changes is determined by the angular frequency of the Earth. Hence change in phase is given by:
$\triangle\phi = \phi - \phi_o$
$= \frac{2\pi d}{\lambda}[\sin(\theta + \triangle\theta) - \sin(\theta_o + \triangle\theta)] - \frac{2\pi d}{\lambda}(\sin\theta - \sin\theta_o)$
$= \frac{2\pi d}{\lambda}[\triangle\theta\ (\cos\theta - \cos\theta_o)]$ (assuming $\triangle\theta \to 0$)

$< 10^\circ$ (By the problem)
Now considering the worst case where $\theta_o = 0$, $\theta = 30^\circ$, we get:
$\frac{2\pi d}{\lambda}\omega\triangle$t(1 - cos30°) = 10° (where ω is the angular frequency of rotation of Earth)
$\Rightarrow \frac{2\pi d}{\lambda}\frac{360^\circ \triangle t}{86400}$(1 - cos30°) = 10°
$\Rightarrow \triangle t = \frac{1200\lambda}{\pi d(1-cos30^\circ)} \approx$ 9.5 seconds (putting λ=1m, d=300m)
This method gives us a maximum integration time of 9.5 seconds.

3.2.2 Method 2

In this method, we see from the perspective of the source. As the earth rotates, the projected baseline as seen by the source changes in length. Now the ambiguity in position of a tile depends on the length of a tile. Let us assume that the ambiguity in baseline length is $\frac{1}{10}$th of a tile length. Hence, if the baseline appears significantly different, we should treat it as a different measurement. So we should integrate for a time interval until the change in the projected baseline equals the uncertainty in position.



Let θ be the angle between the source and zenith at the beginning and $\triangle\theta$ be the change in that angle during the integration. Hence considering this logic, we conclude:
bcos($\theta + \triangle\theta$) - bcos$\theta = \frac{l}{10}$ (where l = tile length and b = baseline length)
$\Rightarrow$ b$\triangle\theta$sin$\theta = \frac{l}{10}$
$\Rightarrow \omega\triangle$t $= \frac{l}{10bsin\theta}$ (where $\omega = \frac{2\pi}{86400}$ and $\triangle$t = integration time)
$\Rightarrow \triangle$t $\approx$ 23 seconds (taking b = 300m, l = 5m and $\theta = 90^\circ$ for worst case)
This method gives us an integration time of 23 seconds. Hence for the analysis, we use the lower of the two integration times i.e. 9.5 seconds.

3.3 Efficiency of Receiver and RFI Affected Channel Rejection

The signal received by the antennae may contain RFI (Radio Frequency Interference) produced by local sources apart from the signal from the sky. The RFI are usually restricted to few frequency channels and hence it can be removed by going to the frequency domain and removing the frequency channels affected by RFI. But in order to remove them, we first need to identify the RFI affected channels.

Here we make the assumption that the signal from the sky is is purely Gaussian and hence the intensities follow an exponential distribution. We know that in exponential distribution, $\frac{mean}{standarddeviation} = 1$.
Now we take the observed mean as a proxy for the expected standard deviation and define efficiency as:
Efficiency = $\frac{Mean}{(StandardDeviation)\sqrt{TimeBandwidthproduct}}$

This definition can be justified in the following manner:
Suppose we average for 'N' packets, then the mean remains the same, but the standard deviation reduces by $\sqrt{N}$. The time bandwidth product is N. If the signal is from the sky, then mean should equal standard deviation before averaging. After averaging, mean = $\sqrt{N}$ (standard deviation). Hence the calculated efficiency should equal 1. For RFI, mean < $\sqrt{N}$ (standard deviation) and the calculated efficiency should be less than 1. During the analysis, the efficiency of each frequency channel in the frequency domain was calculated for individual power spectra.

For selecting the threshold efficiency below which a channel should be rejected, the following logic was adopted. Suppose the ensemble size for which mean and rms are computed is M. Then the uncertainty in $\frac{\mu}{\sigma}$ is $1/\sqrt{M}$. A Gaussian is constructed for each channel whose standard deviation equals the rms of the channel across time. Now if we have k frequency channels, we find the minimum value of 'a' such that P($|x| > a\sigma$) < $\frac{1}{k}$. Now this 'a' should reflect the confidence level in our efficiency calculation. Hence, a channel should be rejected if its efficiency is less than: (1 - $\frac{a}{\sqrt{M}}$)

This analysis was carried out on the individual power spectra and was necessary to get sharp isolated peaks in the cross correlation plots.

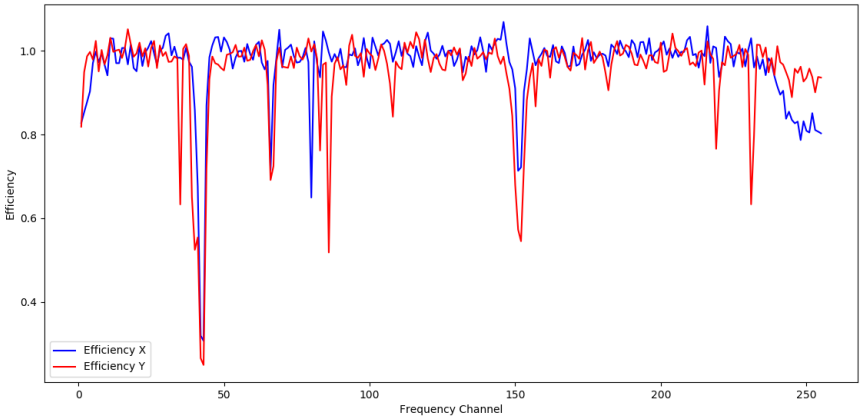


Figure 3: A sample efficiency plot before channel rejection

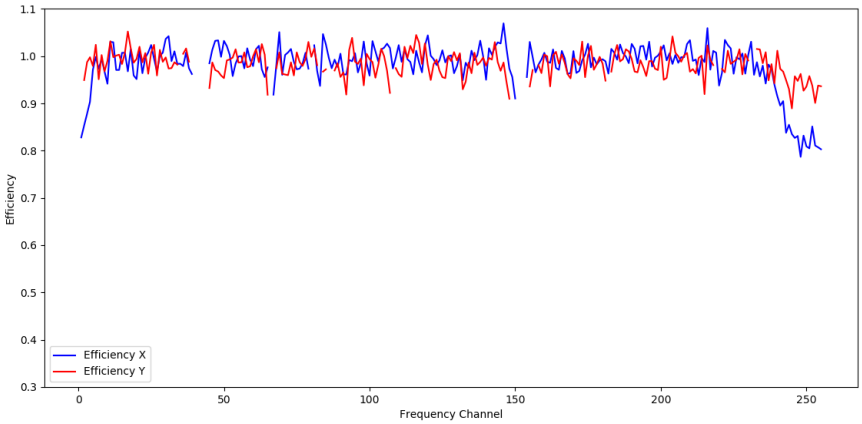
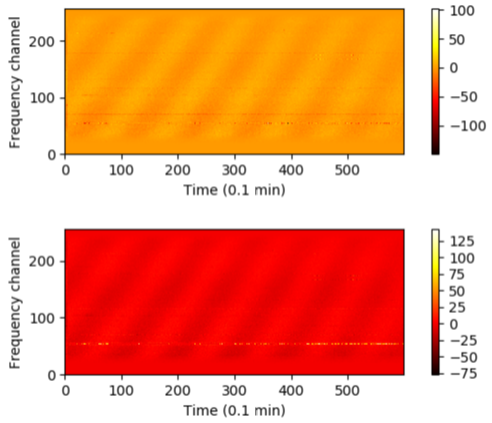


Figure 4: A sample efficiency plot after channel rejection

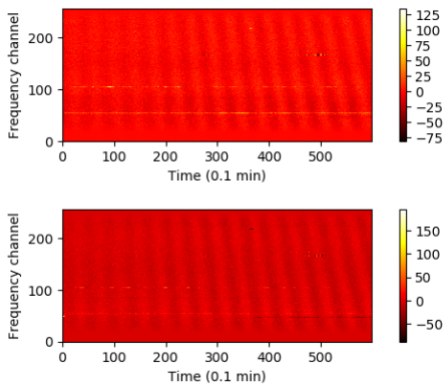
3.4 Auto and Cross Correlation Dynamic Spectra

The auto-correlation dynamic spectrum is the plot of Power (XX* or YY*) as a function of time and frequency. The cross-correlation dynamic spectra are plots of real and imaginary parts of the complex cross correlation (XY*) as a function of time and frequency. If the X and Y voltages are from two

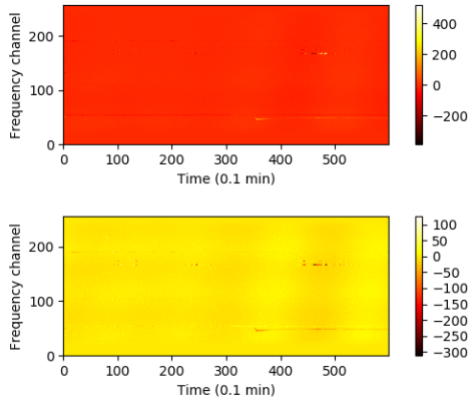
different tiles, then we expect to see fringe patterns in the dynamic and cross spectra which are produced since the geometric delay results in a phase gradient across frequency. Hence these plots tell us if there indeed is some signal from the source. During the analysis, the dynamic spectra were produced after integration and RFI affected channel rejection.



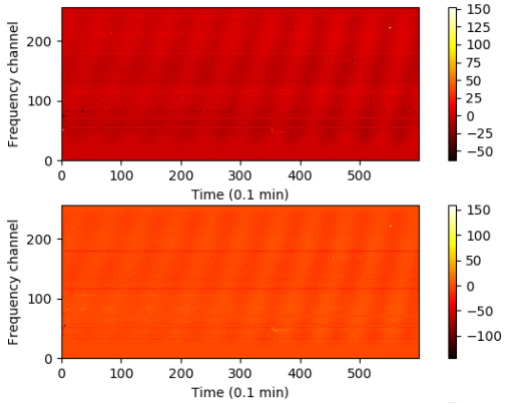
(a) Real and Imaginary parts of Cross spectrum of tiles connected to das 3



(b) Real and Imaginary parts of Cross spectrum of tiles connected to das 7



(c) Real and Imaginary parts of Cross spectrum of tiles connected to das 1



(d) Real and Imaginary parts of Cross spectrum of tiles connected to das 6

3.5 Estimating the Observed Delay using Hilbert's Transform

Here we make use of the cross correlation theorem which states that: a $\star$ b = $\mathcal{F}(\overline{\mathcal{F}(a)}\mathcal{F}(b))$ (where $\star$ denotes cross correlation, $\mathcal{F}$ denotes Fourier Transform and $\overline{A}$ = complex conjugate of A.)

During the analysis, we first compute the FFT (Fast Fourier Transform) of the signals from two tiles and find their complex cross correlations (AB*) in the frequency domain. Then we retain only the positive frequencies so that we get the analytic signal corresponding to the cross correlation function upon Inverse Fourier Transform. This whole procedure is equivalent to adding the signal x(t) with j$\mathcal{H}$[x(t)] (where $\mathcal{H}$[x(t)] is the Hilbert's Transform of x(t)) which has the same effect of filtering out the negative frequencies, giving rise to a complex valued function after IFFT. Before IFFT, zero padding might be done in order to improve the accuracy of finding the position of the peak in the cross correlation function. Now the delay can be found just by noting the position of the peak in the cross correlation function. Alternatively, the phase at the peak can be found by computing the value of arctan($\frac{Imaginarypart}{RealPart}$) at the peak.

This value of phase is monitored for every averaged spectrum and results in a Saw-tooth like function of Phase vs Time. The 360° wraps are removed and each value is divided by the central frequency to get the plot of Delay against Time.

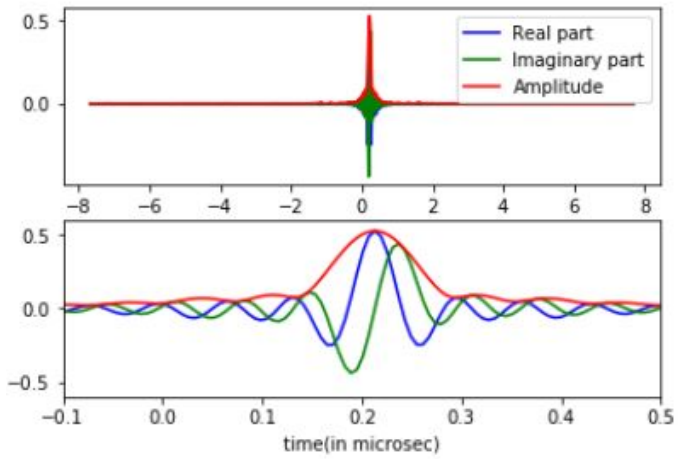
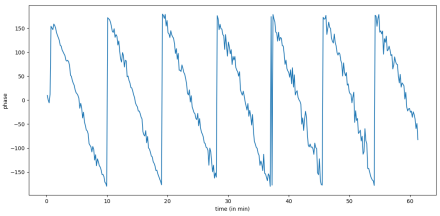
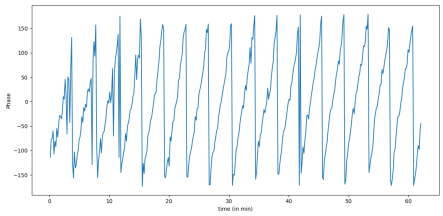


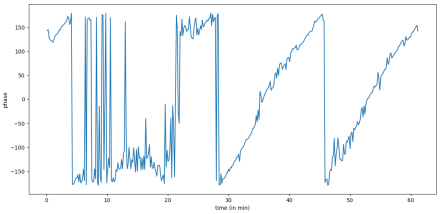
Figure 6: An example of complex valued cross-correlation function



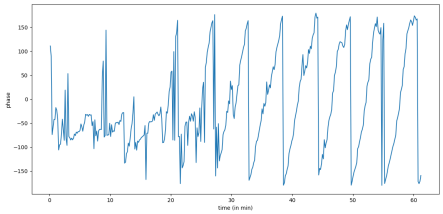
(a) Phase vs time plot das 3



(b) Phase vs time plot das 7



(c) Phase vs time plot das 1



(d) Phase vs time plot das 6

3.6 Dynamic Delay Plot

This is a plot of the amplitude of the cross-correlation function described in 3.5 as a function of time and describes how the geometric delay is changing with time. Also, since the geometric delay depends on the RA, Dec of the source, each source should have a unique peak in the cross correlation function and these plots are useful to monitor what sources were detected by the telescope.

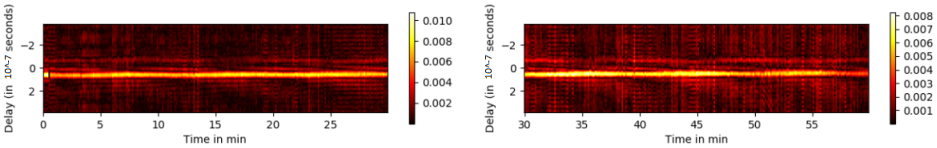


Figure 8: Dynamic delay plot of tiles connected to das 3

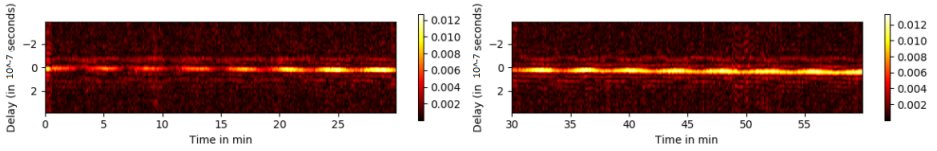


Figure 9: Dynamic delay plot of tiles connected to das 7

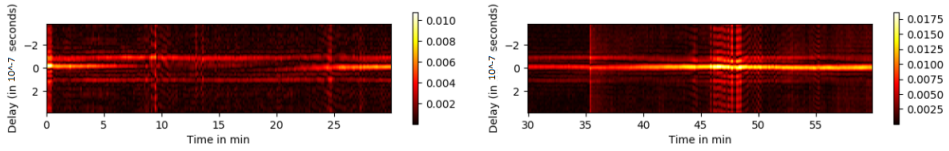


Figure 10: Dynamic delay plot of tiles connected to das 1

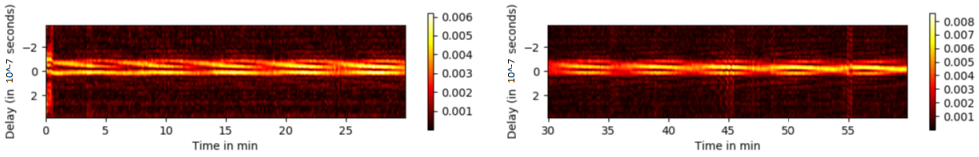


Figure 11: Dynamic delay plot of tiles connected to das 6

3.7 Residual Delays

The residual delays were calculated from the expected and observed delay plots. The trends in the residual delay plots can be attributed to lack of accuracy of the gps coordinate measurements.

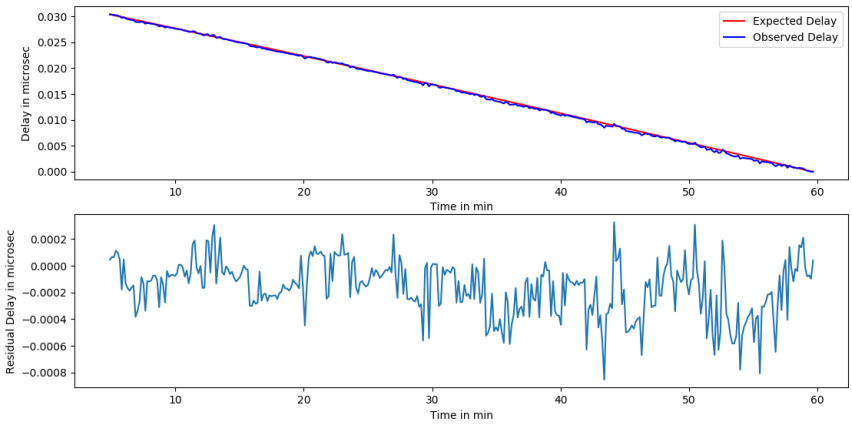


Figure 12: Expected, Observed and Residual delays between tiles connected to das 3

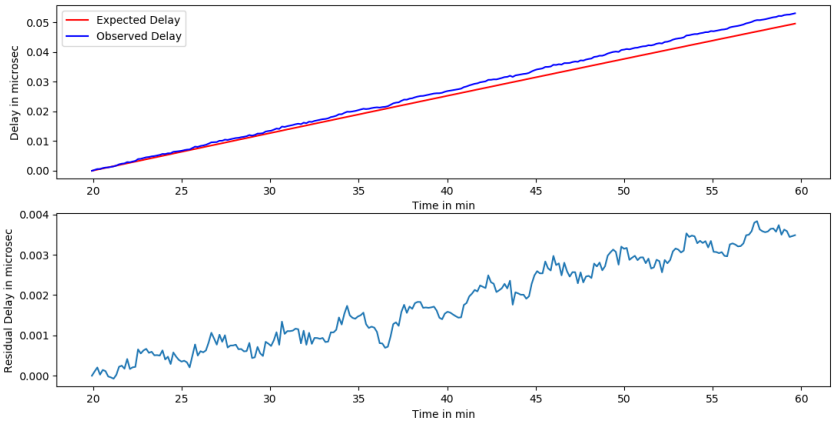


Figure 13: Expected, Observed and Residual delays between tiles connected to das 7

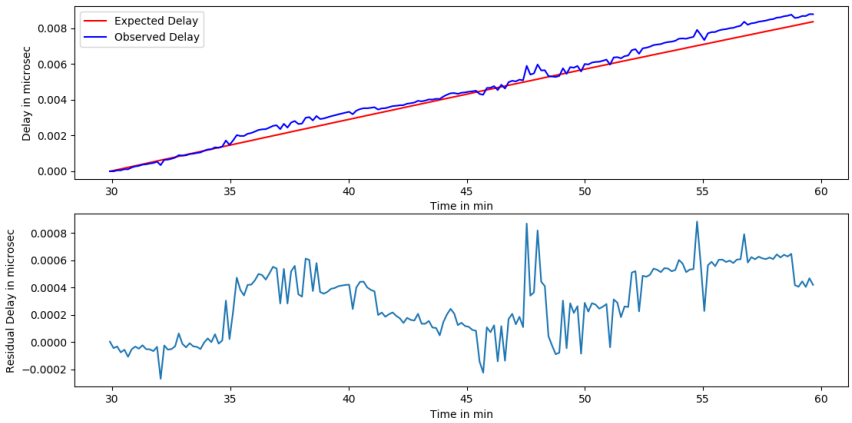


Figure 14: Expected, Observed and Residual delays between tiles connected to das 1

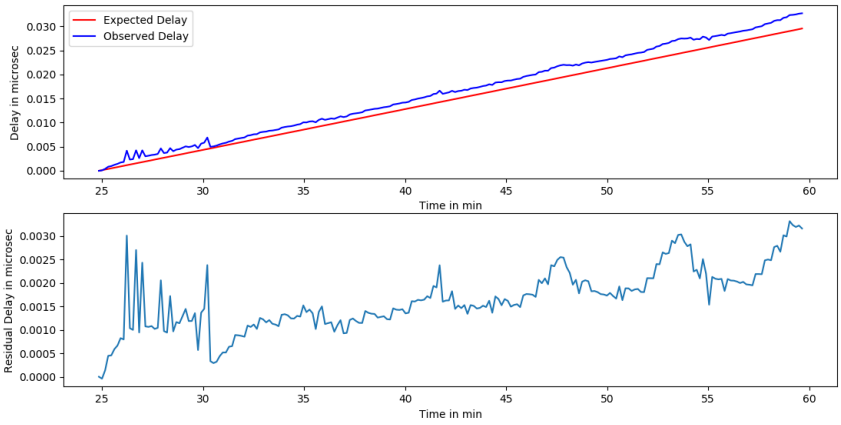


Figure 15: Expected, Observed and Residual delays between tiles connected to das 6

3.8 Updating Baseline Lengths

The geometric delay in arrival, at two tiles, of a signal due to a source at an angle θ from zenith is bsinθ where b is the baseline length. After a time t, the source moves to an angle $(\theta + \omega t)$ (where $\omega = \frac{2\pi}{86400}$). Now the geometric delay becomes bsin(θ+ωt). Hence the change in geometric delay in time t is given by:
b[sin(θ+ωt) - sinθ]
$\approx$ b ωt cosθ (if t $\rightarrow$ 0)
Hence, the slope of Delay vs Time plot is given by bωcosθ.

During the analysis, a straight line is fitted to the Delay vs Time Plots for both expected and observed.
Let the slope of expected delay vs time plot be m_{exp} and that of the observed delay vs time plot be m_{obs}. Let b_{gps} be the baseline length calculated from gps coordinates and $b_{updated}$ be the updated baseline length. Then,
$b_{updated} = b_{gps}(\frac{m_{obs}}{m_{exp}})$.

The uncertainty in the measured baseline length is obtained from the error in slope due to fitting which is given by:
$\sigma_m = \sqrt{[\frac{\sum_{i=1}^{N}(y_i - mx_i - c)}{N-2}][\frac{N}{N\sum_{i=1}^{N}x_i^2 - (\sum_{i=1}^{N}x_i)^2}]}$ (where $y_i = i^{th}$ delay, $x_i = i^{th}$ time, m = slope, c = intercept and N = population size)

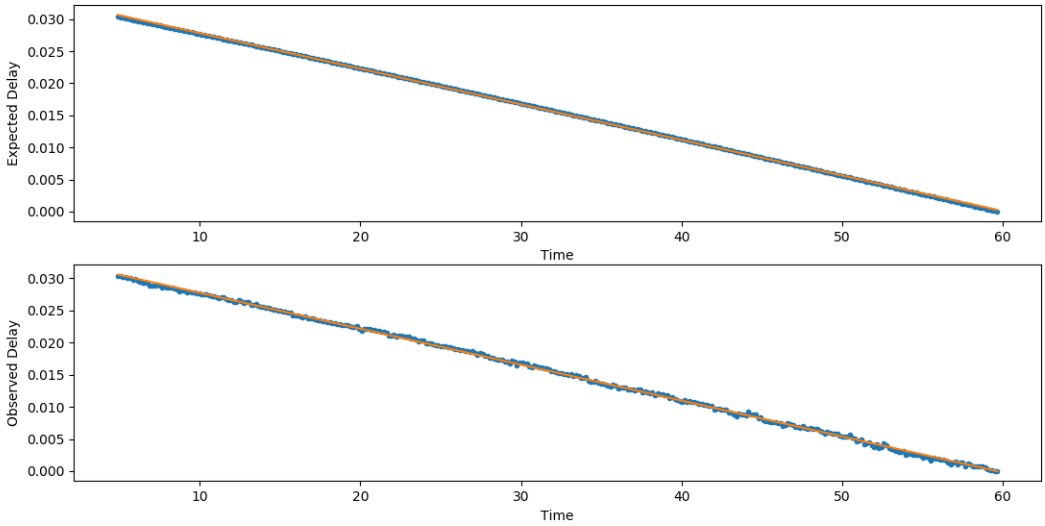


Figure 16: Least Squares fits to the delay time plots (das 3)

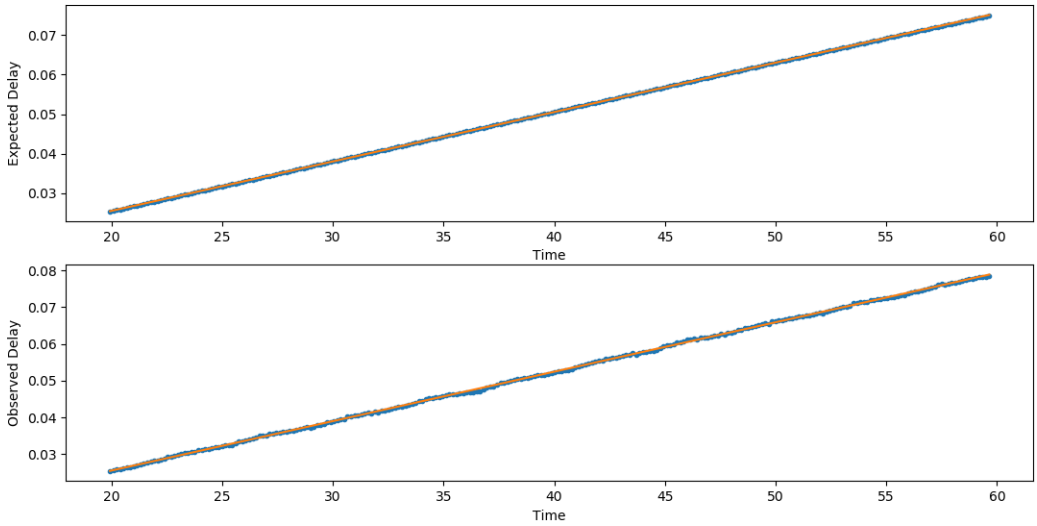


Figure 17: Least Squares fits to the delay time plots (das 7)

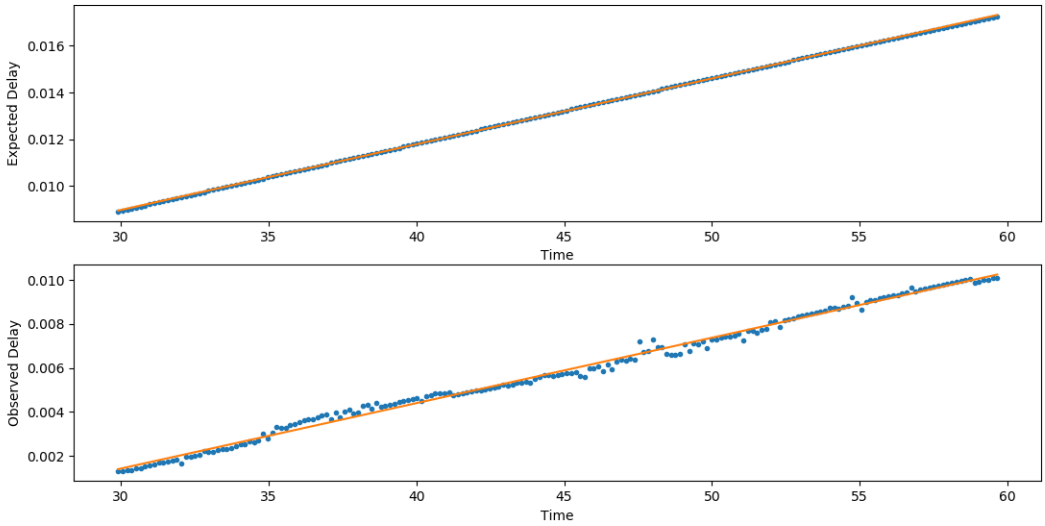


Figure 18: Least Squares fits to the delay time plots (das 1)

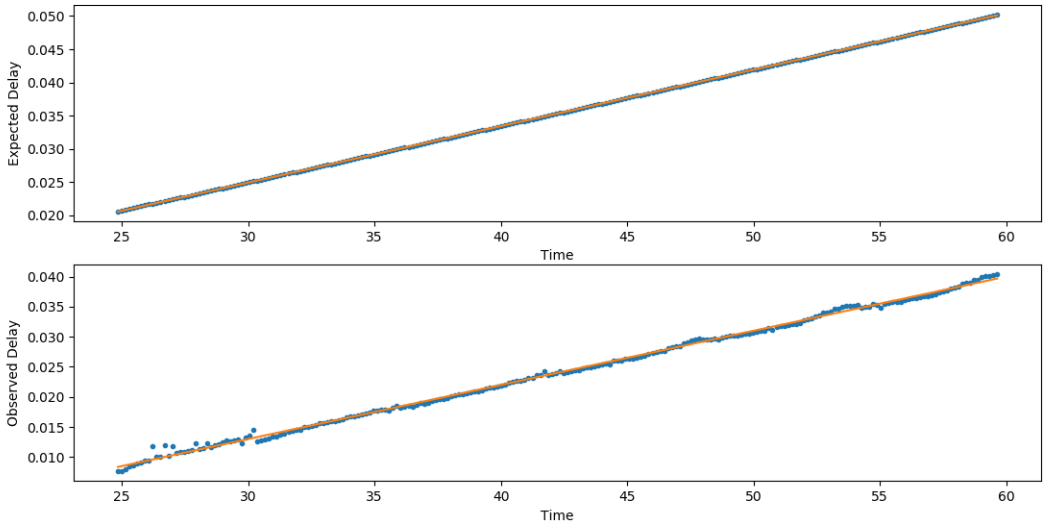


Figure 19: Least Squares fits to the delay time plots (das 6)

4 Results

The results of this analysis can be summarized in the following table:

DAS	TILES (H-V)	ORIGINAL LENGTH	UPDATED LENGTH	DIFFERENCE (UPDATED-ORIGINAL)
3	1 – 3	79.978	80.6 +- 0.1	0.6
7	7 – 4	113.019	121.4 +- 0.1	8.4
1	7 – 3	45.524	48.2 +- 0.3	2.7
6	7 – 1	65.275	69.2 +- 0.2	3.9

Figure 20: Updated baseline lengths

It is seen that the largest difference in original and updated baseline lengths is seen in the tiles connected to das 7 which implies that it is most likely that the gps coordinates of tile 4 might have an error.

5 Acknowledgements

I would like to thank Desh for the guidance provided by him throughout the project. Also, the online SWAN teleconferences and discussions with fellow students were always very enriching and enjoyable.