

Observational Techniques 2 - Assignment 2

Carys Gilbert (GLBCAR006) and Damien Matthew (MTTDAM001)

October 2024

1 Introduction

We investigate the MeerKAT observation data of source PKS 1934-63 with a focus on understanding and interpreting the radio frequency interference (RFI) present in the data. We begin with some exploratory data analysis (EDA) in section 2 where we present information such as the observation date and duration and the observation bandwidth. In section 3 we explore the UV coverage of this MeerKAT observation and several baseline parameters. Finally, in section 4, we investigate the RFI and determine factors such as RFI frequency regions, ‘clean’ versus RFI region visibilities, assess the statistical parameters of different regions and develop detrending and filtering methods. All code used can be found at <https://github.com/literalcarys/OT2Assignment2>.

2 Exploratory data analysis (EDA)

Q1

Start date and time: 2019-01-31 12:56:12.12. End date and time: 2019-01-31 12:59:48.03. This corresponds to an observation length of 3 minutes and 35.91 seconds.

Q2

The target of the observation is PKS 1934-63 with RA 19h 39m 25.03s and declination -63d 42m 45.62s.

Q3

The bandwidth of the observation covers the frequency range 856.00 MHz – 1711.16 MHz, thus it has a total bandwidth of 855.16 MHz.

Q4

Each channel covers a frequency range of 835.94 KHz – this is the channel width.

Q5

There were 50 MeerKAT antennas used in this observation.

3 Understanding the UVW

Q1

The UV coverage represents the baseline components in the (u, v) plane, in units of wavelengths corresponding to the observation frequency, of a radio interferometry array. Importantly, since the (u, v) plane is a measure of the baseline components, it is not a measure of the antenna locations [1]. We plot the UV coverage for this observation in Fig. 1. Since the UV is a function of the hour angle (H_0) and declination (δ_0) of the observed source, the UV coverage changes as a function of the duration of the observation [2]. Thus the baseline components trace out small paths in the (u, v) plane. The coverage of these paths in the (u, v) plane – the UV coverage – determines the quality of the final image achieved via Fourier transform [1]. From Fig. 1, it is evident that we have significant sampling, due to 50 antennas and thus 1225 baseline pairs, however the tracks traced out in the (u, v) plane are not very extensive due to the limited observing window of ~ 3.5 minutes. So, the extensive baseline pairs evident in Fig. 1 will improve our image but the extent of the (u, v) tracks will not.



Figure 1: The UV coverage for this observation.

Q2

Antenna pair with the shortest baseline: 20.33 m between M000 and M001. Antenna pair with the longest baseline: 7294.42 m between M049 and M059.

Q3

The central frequency is 1283.58 MHz corresponding to a central wavelength of $\lambda = 0.23$ m. The expected short baseline resolution is:

$$\theta = \frac{\lambda}{b} = \frac{0.23}{20.33} = 1.10 \times 10^{-2} \text{ radians}$$

The expected long baseline resolution is:

$$\theta = \frac{\lambda}{b} = \frac{0.23}{7294.42} = 3.20 \times 10^{-5} \text{ radians}$$

Q4

The shortest baseline delay is:

$$t_{\text{delay}} = \frac{b_s}{c} = \frac{20.33}{c} = 6.78 \times 10^{-8} \text{ s}$$

The longest baseline delay is:

$$t_{\text{delay}} = \frac{b_l}{c} = \frac{7294.42}{c} = 2.43 \times 10^{-5} \text{ s}$$

where b_s and b_l are the short and long baseline distances, respectively.

Q5

ENU stands for ‘East North Up’. This thus describes the position of the antennas using the classical cardinal system. These coordinates change as a function of the position of celestial objects as it is in reference to the local horizon [2]. The ‘spatial frequency coordinates’ UVW, in units of wavelength, are baseline and source specific – UV corresponds to baseline coordinates while W “points” in the direction of the source and represents the delay distance between two antennas [1]. The (u, v) plane is thus always perpendicular to the observation direction. It is thus useful to convert from the ENU coordinate system to the UVW coordinate system. This is achieved via the following matrix transformations (adapted from [2]):

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} \sin(H_0) & \cos(H_0) & 0 \\ -\sin(\delta_0) \cos(H_0) & \sin(\delta_0) \sin(H_0) & \cos(\delta_0) \\ \cos(\delta_0) \cos(H_0) & -\cos(\delta_0) \sin(H_0) & \sin(\delta_0) \end{pmatrix} \begin{pmatrix} B_x \\ B_y \\ B_z \end{pmatrix}$$

where B_x, B_y, B_z are the baseline vector distances between a pair of antennas in a coordinate system (x, y, z) that is aligned with Earth's rotational axis (i.e. $B_x = x_2 - x_1$, etc.) [2]. The ENU coordinates are related to the (x, y, z) , and thus UVW, coordinates via:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 & \sin(\alpha) & \cos(\alpha) \\ 1 & 0 & 0 \\ 0 & \cos(\alpha) & \sin(\alpha) \end{pmatrix} \begin{pmatrix} E \\ N \\ U \end{pmatrix}$$

where α is the latitude of the 'U' coordinate.

4 Radio Frequency Interference (RFI)

This section is completed with radio dishes M000 and M001 (the shortest baseline) unless stated otherwise.

Q1

RFI region 1, Fig. 2, occurs at channels 95–105 corresponding to a frequency range of 935.41–943.77 MHz. The major source of RFI at these frequencies is due to GSM (Global System for Mobile communications) down – a telecommunications satellite. Note that the time axis for all visibility profiles in this report is in units of 'number of readouts' which occur every ~ 8 seconds.

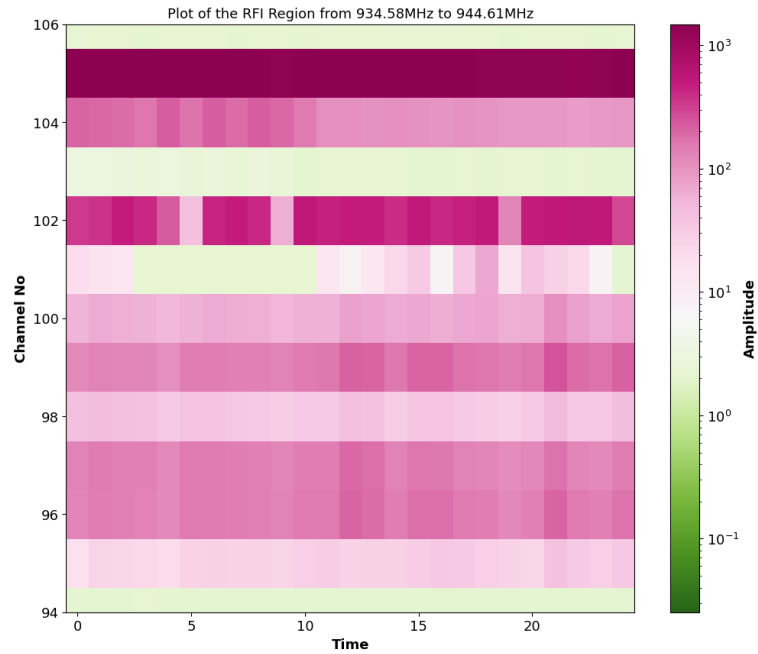


Figure 2: The channel intensity as a function of time for RFI region 1 (channels 95–105 corresponding to a frequency range of 935.41–943.77 MHz).

RFI region 2, Fig. 3, occurs at channels 330–530 corresponding to a frequency range of 1131.86–1299.05 MHz. The major sources of RFI at these frequencies are satellites GPS L2 and GPS L5, Galileo 1 and Galileo 2, GLONASS L2, and possibly aircraft transponders.

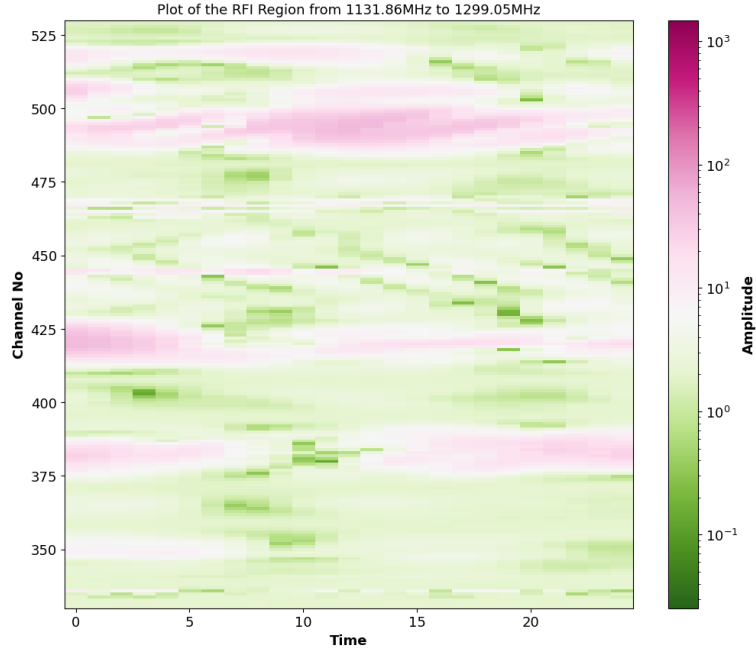


Figure 3: The channel intensity as a function of time for RFI region 2 (channels 330–530 corresponding to a frequency range of 1131.86–1299.05 MHz).

RFI region 3, Fig. 4, occurs at channels 800–900 corresponding to a frequency range of 1524.75–1608.34 MHz. The major sources of RFI at these frequencies are satellites Inmarsat, GPS L1 and GLONASS L1.

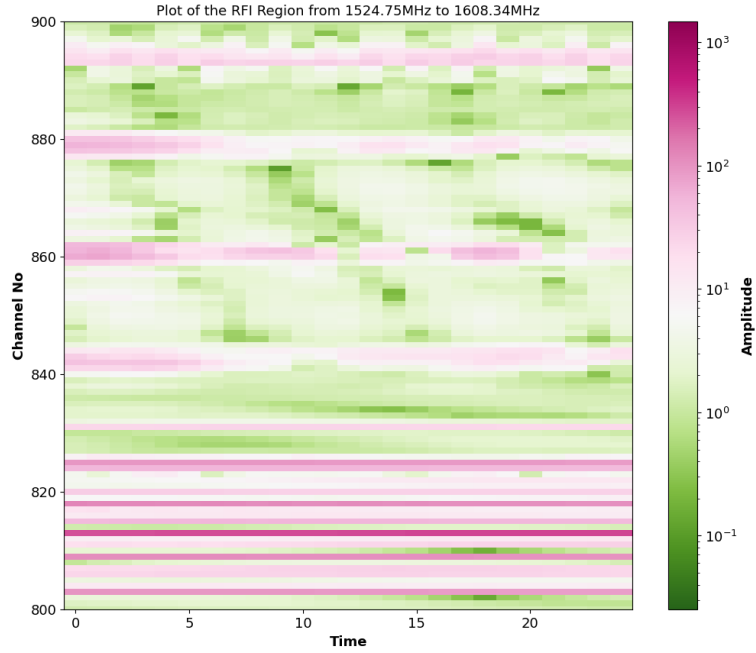


Figure 4: The channel intensity as a function of time for RFI region 3 (channels 800–900 corresponding to a frequency range of 1524.75–1608.34 MHz).

Q2

To analyse the difference between RFI signal and a ‘clean’ region, we plot the amplitude-bins of a single RFI channel and a single ‘clean’ channel as seen in Figures 5 and 6, respectively. In these plots, the amplitude-bins (x-axis) correspond to sources of increasing brightness or intensity while the counts (y-axis) correspond to the occurrence rate at which these sources are detected in the channel. The data from the two channels (RFI and ‘clean’) are independently integrated over the duration of the observation. We combine the data from the 10 shortest baseline antennas comprising dishes m000, m001, m002, m003, m004, m005, m009, m010, m011 and m013.

The RFI region contains a heavily left-skewed amplitude-bin distribution. There are a large number of low amplitude detections and an exponentially decreasing number of very high amplitude detections. The ‘clean’ region instead has several (~ 8) Gaussian shaped amplitude-bin distributions all at very low amplitudes likely corresponding to ~ 8 distinct astronomical sources. This is consistent with what we expect from astronomical sources – they are faint (low amplitudes) and Gaussian in their brightness distribution. Note how the amplitude-bins extend beyond 2000 for the RFI channel but only extend to ~ 2.4 in the ‘clean’ channel. This is also an expected outcome as RFI has a significantly larger amplitude than astronomical sources.

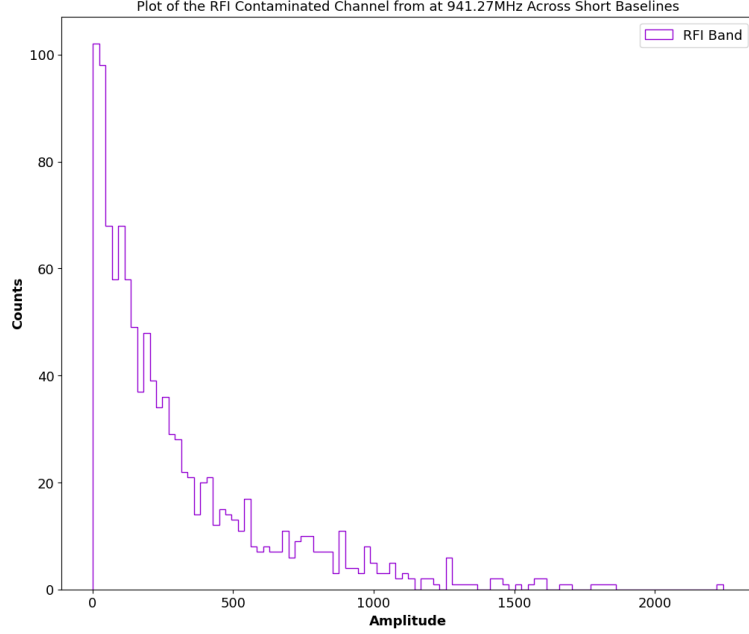


Figure 5: Histogram of the amplitude-bins across 10 of the shortest baselines at a single RFI channel centred on 941.27 MHz. The channel data is integrated across the duration of the observation to produce the histogram.

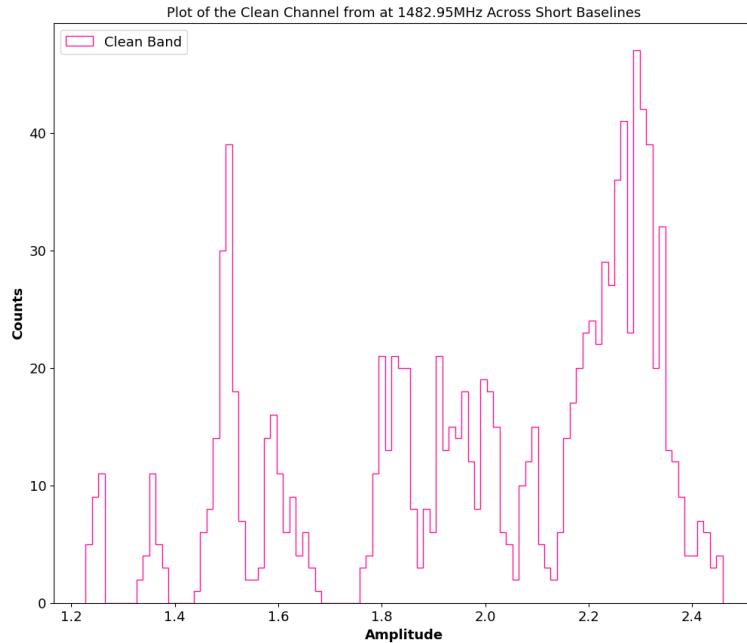


Figure 6: Histogram of the amplitude-bins across 10 of the shortest baselines at a single ‘clean’ channel centred on 1482.95 MHz. The channel data is integrated across the duration of the observation to produce the histogram.

Q3

The measure of spread (standard deviation) in the clean region displayed in Fig. 6 is $\sigma_{clean} = 0.32$ whereas in RFI region 3 (Fig. 5) $\sigma_{RFI} = 332.27$. This is consistent with what is observed in the two plots: the RFI channel has

amplitude bins ranging from 0 to ~ 2300 which should correspond to a large spread and indeed does, whereas the ‘clean’ channel has amplitude bins ranging from 0 to ~ 2.4 which should correspond to a small spread as it does. Physically, this result is also sound as we know that astronomical sources will have low amplitudes whereas RFI will have a vast range of amplitudes and up to very large amplitudes compared with the ‘clean’ data.

Q4

The skewness is a measure of the symmetry, or lack thereof, of a dataset around its centre. For normal distributions or symmetric data, the skewness is zero. Data that is not symmetric may have a longer left or right tail corresponding to left-skewed and right-skewed classifications, respectively. Left-skewed data will have a negative skewness while right-skewed data will have a positive skewness. Multi-modal data, like that in the ‘clean’ region, will have a more complex reason for the sign of the skewness. [3]

The kurtosis is a measure of the peaked or flat nature of the dataset which corresponds to the amount of data falling within the tails of the distribution. A large kurtosis is consistent with heavy tails and outliers and thus a flatter distribution, while a low kurtosis indicates lighter tails and thus a more peaked distribution. The kurtosis for a standard normal distribution is 3, but in our analysis we use the scipy kurtosis which instead considers the ‘excess kurtosis’ measurement. The ‘excess kurtosis’ adjusts the kurtosis value such that a standard normal distribution has a kurtosis of 0. [3]

In the clean region: Kurtosis = -0.80 and skewness = -0.64. Since the kurtosis in the clean region is slightly less than 0, the clean region is slightly more peaked. The slight negative skewness in the clean region indicates that the histogram is slightly skewed to the left. We can see this is the case in Fig. 6.

In RFI region: Kurtosis = 3.98 and skewness = 1.85 Since the kurtosis in RFI region 1 is significantly larger than 0, this RFI region is more tail heavy. The positive skewness in RFI region 1 indicates that the histogram is skewed to the right. We can see this is the case in Fig. 5.

Q5

In Fig. 7, we plot the full spectrum visibilities for a short (M000–M002) and long (M048–M060) baseline. Notice how the amplitude of the overall short baseline RFI signal (pink) is approximately an order of magnitude greater than the RFI signal in the long baseline, as indicated by the colourbar. We focus on RFI region 1 (channels 95–105) to assess the difference between the two baselines within an RFI contaminated region. Figures 8 and 9 show the baseline visibilities for this contaminated region. Fig. 8 presents these two visibilities with their own amplitude scales represented by the colourbars. There appears to be greater amplitudes of interference in the short baseline (left) due to the greater ‘pink’ coverage. This observation is emphasised in Fig. 9 where the same visibility plots are simply scaled to that of the short baseline. Here, the RFI amplitude is clearly significantly larger in the short baseline than the large baseline.

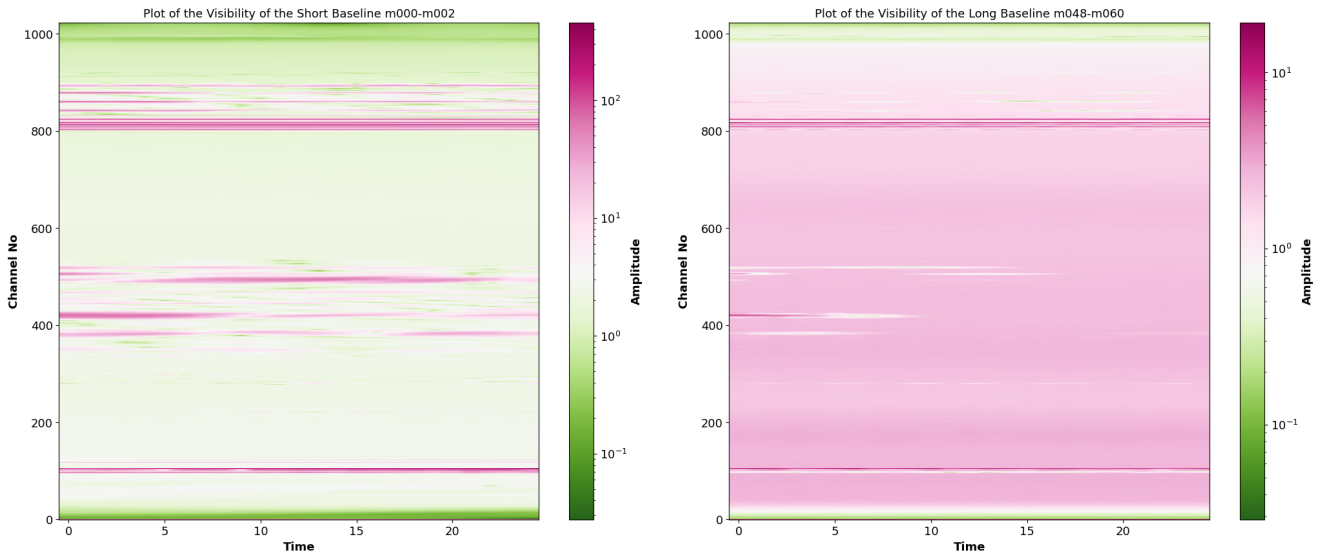


Figure 7: The complete channel intensity as a function of time for short (left) and long (right) baseline pairs.

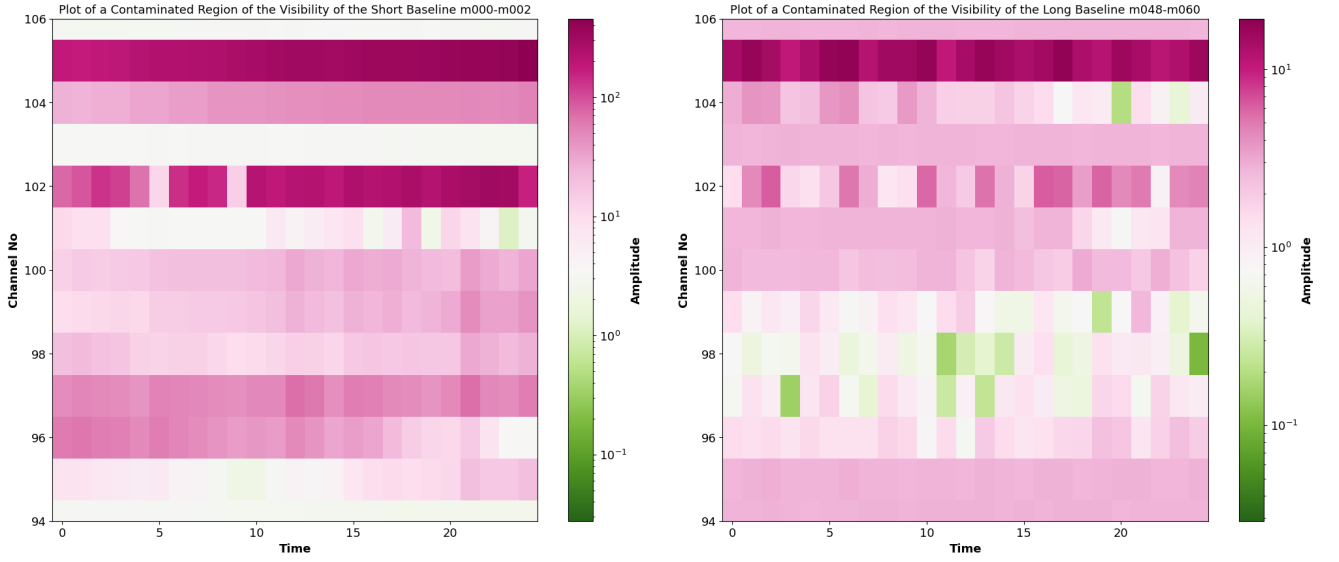


Figure 8: The RFI region 1 channel intensity as a function of time for short (left) and long (right) baseline pairs.

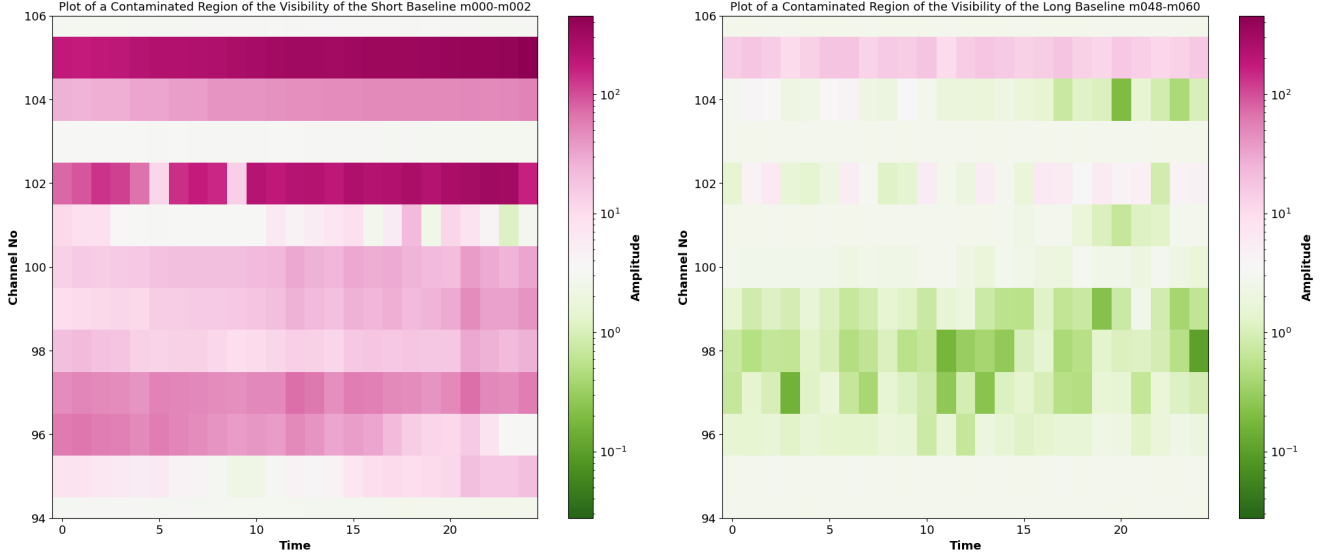


Figure 9: The RFI region 1 channel intensity as a function of time for short (left) and long (right) baseline pairs. For better comparison, the amplitude of the long baseline visibility (right) has been scaled to match the short baseline visibility scale (left).

Q6

Fig. 10 shows the signal amplitude across the entire bandwidth for the M000–M001 baseline. The RFI region that we investigate is located at ~ 940 MHz. A cropped plot of this RFI region, along with a fitted Gaussian, is shown in Fig. 11.

The half power beam width (HPBW) of the RFI at ~ 940 MHz for the shortest baseline (M000–M001) is $\text{HPBW} = 8.67$ MHz. This is determined by fitting a Gaussian to the RFI region, as shown in Fig. 11 and extracting the best-fit standard deviation (σ). The HPBW is related to σ via

$$\text{HPBW} = 2\sqrt{2 \ln 2} \sigma$$

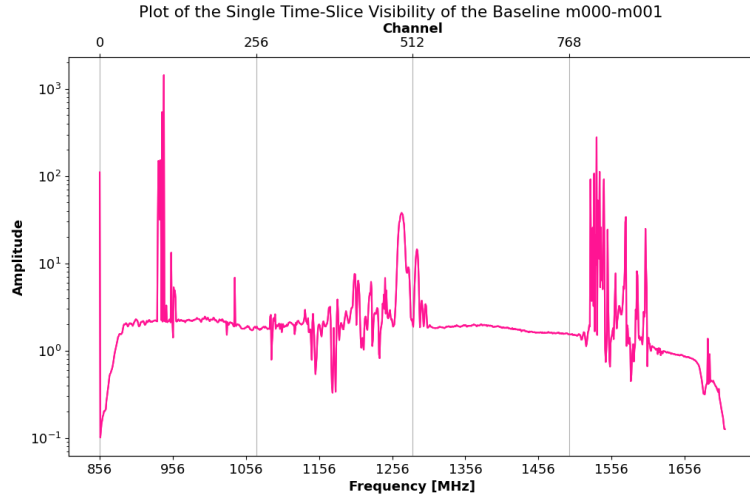


Figure 10: The signal amplitude as a function of the complete bandwidth frequency for the shortest baseline (M000–M001).

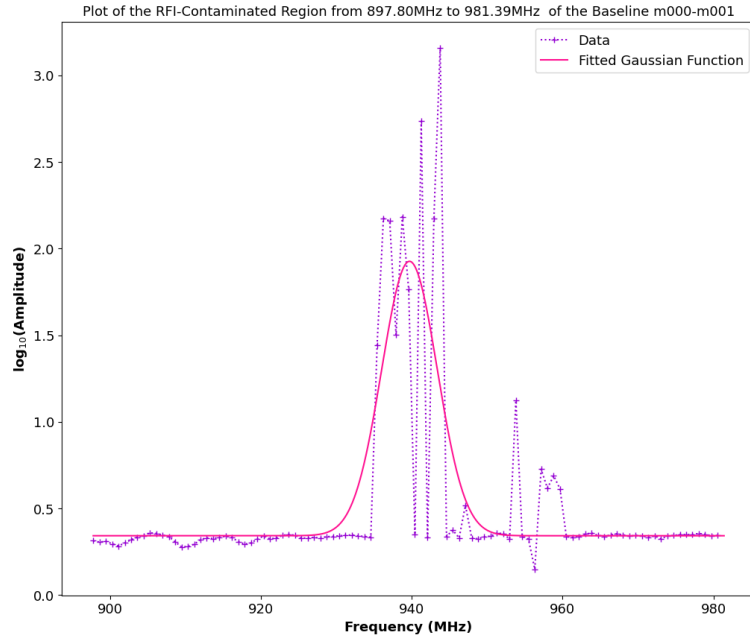


Figure 11: The signal amplitude as a function of frequency for the RFI region at ~ 940 MHz for the shortest baseline (M000–M001). A Gaussian function is fitted to the RFI region from which the HPBW is determined.

Q7

The HPBW of 8.67 MHz corresponds to the following number of affected channels:

$$\text{Affected Channels} = \frac{8.67 \times 10^3 \text{ KHz}}{835.94 \text{ KHz/Channel}} \approx 10 \text{ Channels}$$

Number of channels affected: 10

Q8

Assuming that the objective of the observation is to study HI emission at 1420 MHz, we can determine the redshift associated with frequency shift that would result in the HI emission detected at 940 MHz. The redshift of a galaxy with HI at 940 MHz is: $z = 0.51$. This corresponds to a galaxy ~ 8.52 Gyr after the big bang or ~ 5.18 Gyr in the past, relative to Earth. Since the channels centred around 940 MHz are contaminated with RFI, we will not be able to detect HI emission from galaxies at a redshift of $z = 0.51$.

We investigate the harmonic effects of the fundamental RFI frequency at 153 MHz. Only the 6^{th} – 11^{th} harmonics fall within the bandwidth range of the observation. The harmonics are integer multiples of the fundamental frequency and the strength of these harmonics decreases for larger integer multiples. Since only the last 6 harmonics fall within the observation bandwidth, it is likely that the data will suffer no significant, additional RFI. Fig. 12 shows the location of these harmonics within the full spectrum of the shortest baseline pair (M000–M001). Indeed, from this plot, it is noticeable that these harmonics either fall within other regions of RFI and thus do not affect our data, or fall within ‘clean’ regions but do not cause a noticeable difference to our ‘clean’ signal. To emphasise the negligible effect of those harmonics that fall within the ‘clean’ signal, we display a zoomed in amplitude-frequency plot of the 7^{th} harmonic in Fig. 13 in which there appears to be no discernable difference in signal with respect to the continuum.

| Frequency Type | Frequency (MHz) | Location Feature |
|----------------|-----------------|------------------|
| Fundamental | 153 | Out of bandwidth |
| 2nd harmonic | 306 | Out of bandwidth |
| 3rd harmonic | 459 | Out of bandwidth |
| 4th harmonic | 612 | Out of bandwidth |
| 5th harmonic | 765 | Out of bandwidth |
| 6th harmonic | 918 | Clean |
| 7th harmonic | 1071 | Clean |
| 8th harmonic | 1224 | RFI |
| 9th harmonic | 1377 | Clean |
| 10th harmonic | 1530 | RFI |
| 11th harmonic | 1683 | RFI |

Table 1: Assuming a fundamental frequency of 153 MHz, the consequential harmonics are as listed above.

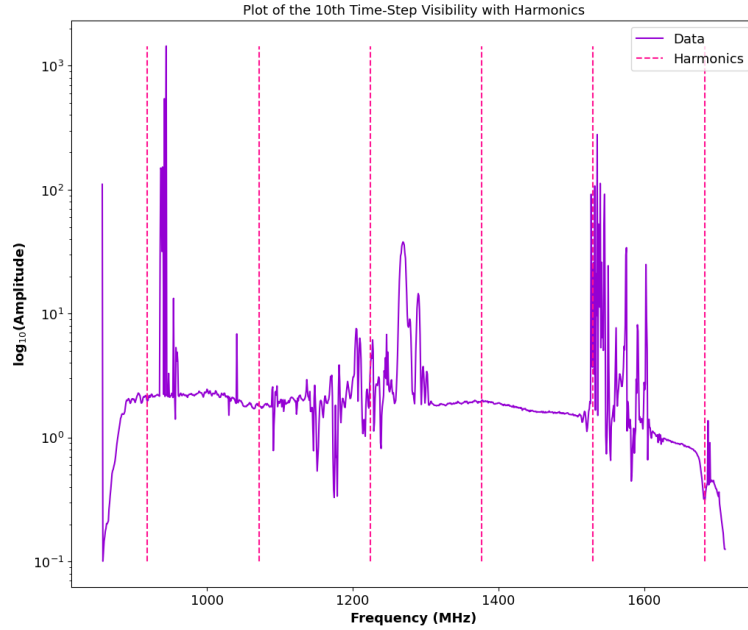


Figure 12: The signal amplitude as a function of the complete bandwidth frequency for the shortest baseline (M000–M001). The 6^{th} – 10^{th} harmonics, due to a fundamental RFI frequency of 153 MHz, are shown.

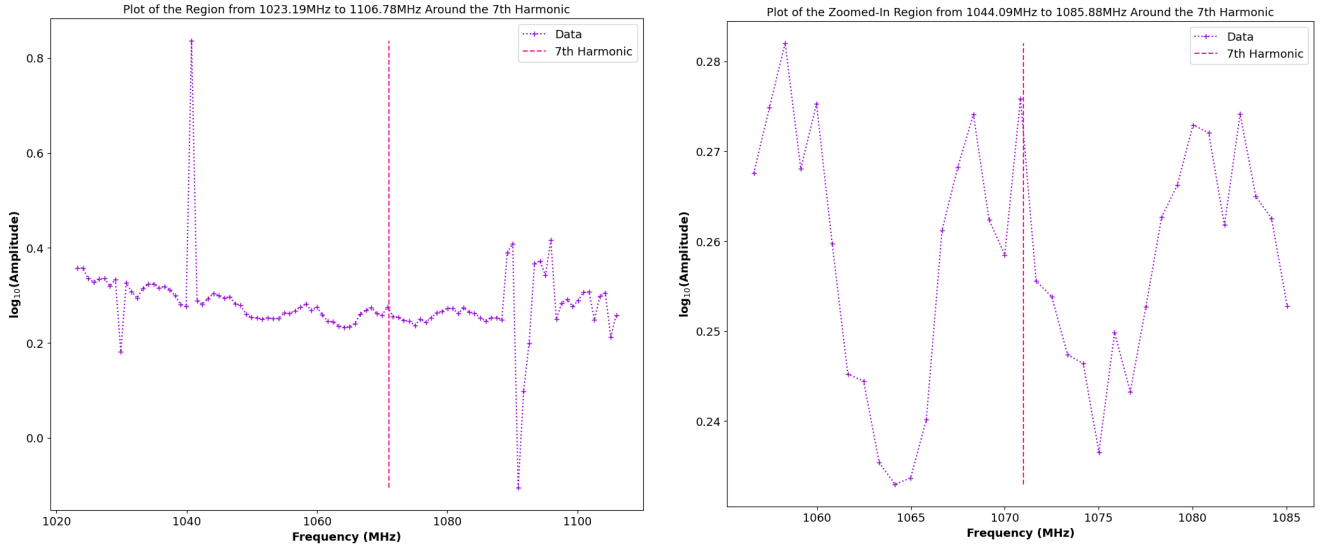


Figure 13: Two zoomed in plots of the signal amplitude as a function of the frequency is shown for the 7th harmonic. The left plot is less zoomed in while the right plot is more zoomed in.

Q10

Signal detrending involves the subtraction of some function from ones dataset with the goal of removing trends within the data. By removing trends in the data, other characteristics of the dataset may become isolated and amplified allowing for improved recognition. We explore several methods of detrending. We investigate the detrending capabilities of scipy and online detrending code to generate our own adapted detrending method.

We utilise both the linear and constant line detrending methods in scipy with the results shown on the left of Fig. 14. The linear detrending method fits a linear function to the data and subtracts the fitted linear function from the data. Due to the constant nature of the dataset, the linear function subtraction does not effectively remove the trend in the data, as seen on the left of Fig. 14 – there is still a positive incline to the linear detrended data. The constant detrending method subtracts a constant function from the dataset. This achieves a more reliable result with our dataset due to the continuum nature of much of the ‘clean’ regions, as evident on the left of Fig. 14.

Using a spline function – a constant function which tends to zero at the beginning and end of a defined window – we compute a moving window which determines the mean value of the dataset as the window progresses along the domain. We effectively compute the mean value of a spline window convolved with our dataset. We then subtract this moving mean value from the dataset in order to detrend our data. We present the results on the right of Fig. 14. Indeed, this method appears to be the most reliable of the three methods explored at detrending our data, evident by a near zero amplitude within the continuum regions and clear amplitude peaks in RFI regions.

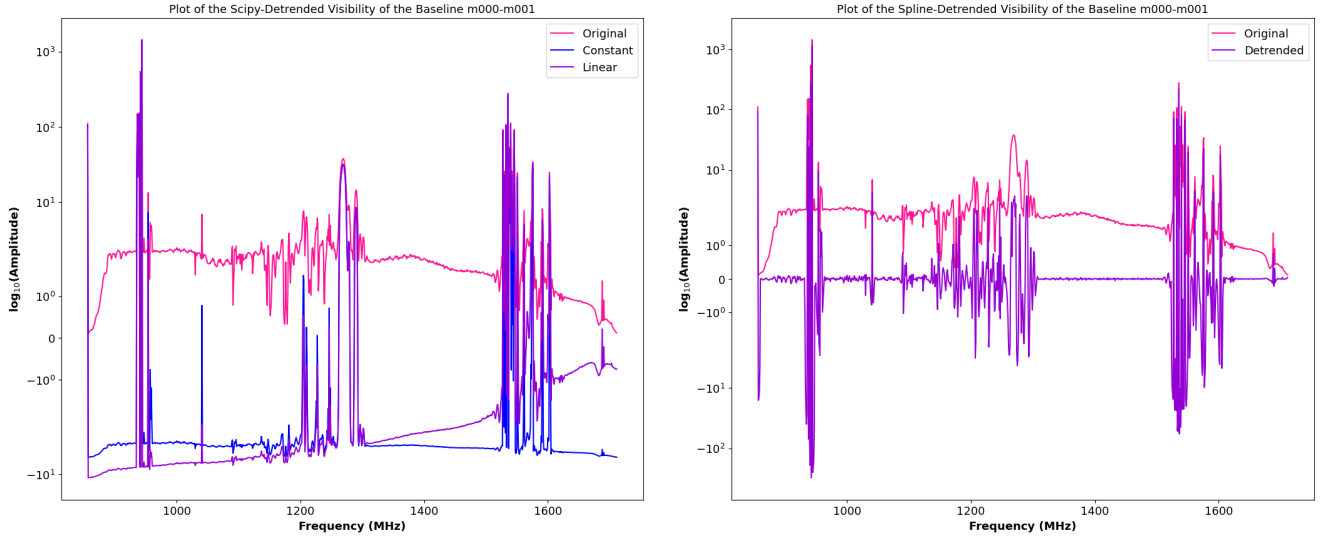


Figure 14: Three detrending methods are explored with scipy’s linear and constant results displayed on the left and our own adapted convolution results on the right. Both plots show the amplitude as a function of the frequency for the shortest baseline (M000–M001).

Q11

We develop a filtering algorithm to detect RFI regions within the observation bandwidth. At the heart of the algorithm is a comparison between a rolling standard deviation and a slight overestimate of the ‘clean’ data’s standard deviation. We first generate two ‘windows’ of 5 and 10 channels wide which we then iteratively ‘roll’ across the observing channels, calculating the standard deviation of the 5 or 10 channels at every iterative interval. Those windows that have a standard deviation of $\sigma > 0.35$ are masked as RFI regions. This value is selected as a slight overestimate of the standard deviation determined for the clean region of $\sigma_{clean} = 0.33$ so that there is a slight buffer in preventing clean regions from being wrongly identified as RFI. We have two rolling windows to ensure that both wide RFI regions and narrow RFI regions are flagged. In Fig. 15, we present the results of the filtering algorithm applied to both the standard and detrended data. The filtering method appears to be reliable in detecting RFI regions and produces near-identical results on both the standard and detrended data. There are several low amplitude RFI regions which do not appear to be flagged. This method may be improved by considering a less crude, qualitative threshold of standard deviation comparison.

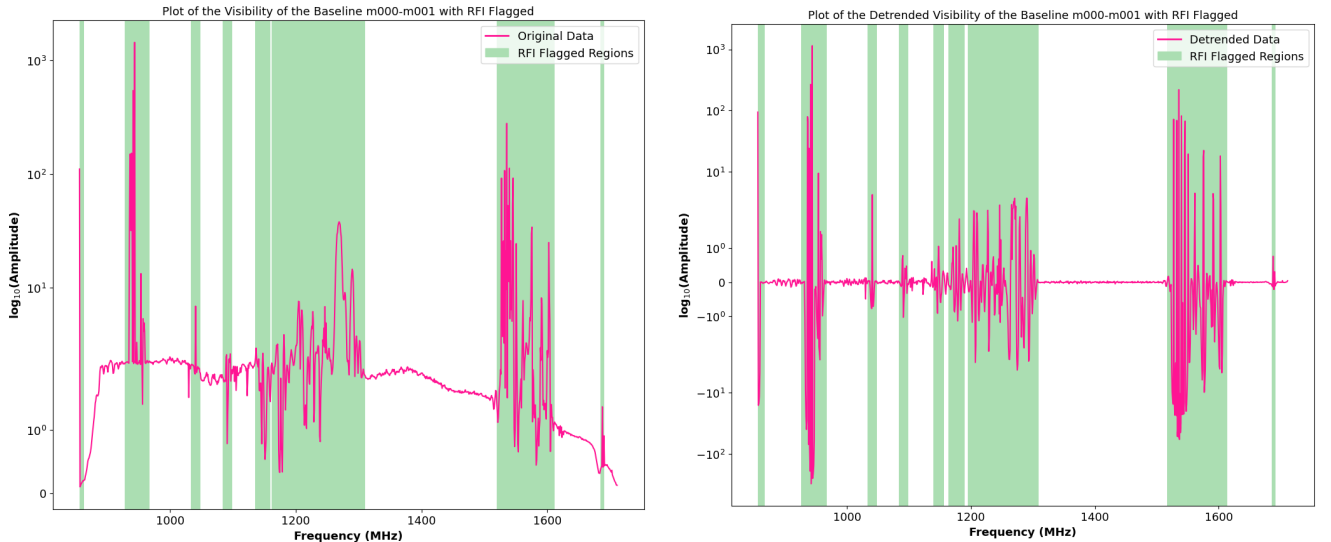


Figure 15: The filtering algorithm developed in this report and applied to the normal dataset (left) and to the detrended dataset (right) to determine the RFI regions.

Q12

The two-line element (TLE) data encodes the orbital position of satellites over time. We obtain the TLE data of 288 satellites via the Celestrak Special Data Request function. Special thanks to Kirsten Elliott for carrying out

that request on behalf of the class. The RA and declination positions of the satellites at the start of the observation (2019-01-31 12:56:12.12) are determined. We then include only those satellites that fall within a cone centred on PKS 1934-63 of radius 45° . This angular threshold is taken due to the source being $\sim 45^\circ$ above the horizon. After filtering for those satellites within the cone, we calculate the angular separation distance using `astropy` and present the results in Table 2. The larger the angular separation between the source and the satellite, the smaller the RFI effect of the satellite. We therefore present the list in ascending angular separation and thus in descending likely RFI.

| Satellite | Angular Separation ($^\circ$) |
|-------------------------|---------------------------------|
| BEIDOU-3 M16 | 7.852 |
| BEIDOU-3 M16 | 7.852 |
| BEIDOU-3 M16 | 7.852 |
| BEIDOU-3 M16 | 7.856 |
| BEIDOU-3 M16 | 7.861 |
| BEIDOU-3S M1S | 30.826 |
| BEIDOU-3S M1S | 30.826 |
| BEIDOU-3S M1S | 30.827 |
| BEIDOU-3S M1S | 30.828 |
| BEIDOU-3S M1S | 30.828 |
| GSAT0205 (GALILEO 9) | 41.184 |
| COSMOS 2529 [GLONASS-M] | 43.981 |
| COSMOS 2529 [GLONASS-M] | 43.981 |
| COSMOS 2529 [GLONASS-M] | 43.981 |
| COSMOS 2529 [GLONASS-M] | 43.981 |

Table 2: The angular separation in degrees between satellites of possible RFI and the pointing direction of MeerKAT when observing PKS 1943-63. The BEIDOU-3S M1S satellites are 5 Chinese telecommunication satellites and were sent as precursors to the BEIDOU-3 generation of satellites. The BEIDOU-3 M16 are 5/36 satellites part of the newer BEIDOU-3 generation. Galileo is a European telecommunications satellite while the 4 GLONASS satellites are Russian telecommunication satellites.

References

- [1] Rick Perley. *Basic Radio Interferometry – Geometry*. Tech. rep. NRAO & ICRAR.
- [2] Dale E. Gary. *Fourier Synthesis Imaging*. Tech. rep. New Jersey Institute of Technology.
- [3] NIST. *Measures of Skewness and Kurtosis*. Tech. rep. National Institute for Standards and Technology.