A Feasibility Study For Real-Time Narrowband RFI Filtering In The GMRT Wideband Backend

Kshitij Aggarwal
Department of Electrical Engineering
Indian Institute of Technology, Ropar

December 2015

Supervised by

Prof. Yashwant Gupta
Dean
GMRT

Mr. Kaushal D Buch
Engineer 'D'
GMRT

NCRA • TIFR

Abstract

The GMRT Wideband Backend aims to provide a better RFI rejection capabilities which calls for an exploration of techniques other than being implemented. One of the motive of this project was to locate an optimal point for spectral RFI filtering in the GWB receiver chain, considering various factors like strength of spectral line and sensitivity of receiver. The long term overall target to be achieved is to develop a deliverable product to GMRT observatory, which should be agnostic to the RFI filtering algorithm and robust enough to be able to enable the user to use RFI filtering and efficiently process the data products. In the current project, various parameters and some ground work for this product has been defined. This project also aims at studying various techniques of narrowband RFI filtering, providing a comparative analysis of them. Techniques like normalisation of the band shape, median based compensation, mean/rms based filtering were considered. These methods have indeed proven better at RFI filtering than the present schemes.

Acknowledgement

I thank my guides Prof. Yashwant Gupta and Mr Kaushal D Buch for giving me a chance to work on this project. I would also like to thank my guides for finding time out of their busy schedule, for guiding me. I would also like to thank Swapnil for helping me with simulated data. I would like to express my gratitude towards the members of GMRT and NCRA for their kind co-operation and encouragement which helped me in completion of this project. I had a wonderful experience of working in a free and stimulating environment over the this month.

Contents

List of Figures

Chapter 1

Introduction

1.1 About GMRT

The Giant Metre wave Radio Telescope (GMRT) consists of thirty 45 m diameter antennas spread over a 25 km region. Half the antennas are in a compact, quasi randomly distributed array with a diameter of about 1 km. The remaining antennas are on 3 arms of length of about 14 km (North West, North East and South) with 5 or 6 antennas on each arm. The longest baseline is about 25 km and the shortest is about 100 m without foreshortening. Astronomical observations are currently possible at 5 different RF bands 150, 235, 325, 610, 1420 MHz with an instantaneous bandwidth of 32 MHz from each of two polarisation channels[1].

1.2 GMRT Wideband Backend(GWB)

The GMRT Backend Group is in the process of developing wide-band digital backends as part of a major upgrade of the GMRT. The main requirements are for processing 400 MHz bandwidth signals from 30 antennas for interferometry and array mode operations, along with additional features like narrow band modes and RFI cancellation schemes. This new GMRT Wideband Backend(GWB) will replace its ancestor GMRT Software Backend(GSB). It will also support seamless frequency coverage from 30 MHz to 1500 MHz[2].

Chapter 2

Radio Frequency Interference

2.1 What is RFI?

Radio Frequency Interference (RFI) refers to the unwanted signals that are detected along with the actual radio astronomical signals. It has the potential to either degrade or inhibit the successful conduct of the observations, by corrupting the data. Therefore it becomes necessary to employ some techniques for successful filtering of these interference signals. There are two main categories of RFI:-

- Narrow Band RFI: Also called frequency domain or spectral domain RFI. In fourier domain, it is visible as a peak in spectrum, therefore it is best to filter it in frequency domain. It emanates from intended transmissions such as radio or TV stations, cell phones.

- Broad Band RFI: Also called as time domain RFI. In fourier domain, it is visible as whole spectrum, while in time domain it would appear as peaks. Therefore its best to filter it in time domain. It emanates from unintentional radiation from sources such as electric power transmission lines, inductive load switching.

2.2 RFI Mitigation

The first step in mitigation is elimination of potential RFI sources. This involves a ban in mobile phones in the GMRT campus, non-usage of petrol vehicles (RFI due to the spark plug can be pretty intense) and tube lights (RFI from starter choke), but there are numerous other factors (like lightning) which cannot be stopped by policies[5]. Hence, a statistical method of RFI mitigation is required to clean the acquired data.

RFI mitigation is based on the fact that signals from astronomical sources follow gaussian probability distribution function [4]. Most of the Radio Frequency Interference (RFI) that the GMRT sees is non-gaussian in nature. The RFI mitigation algorithms used here essentially exploit this property of being non-gaussian of RFI. The probability distribution of gaussian noise is given by:-

$$P(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}}$$

A robust estimator is required for the RFI filtering, because estimators like mean and RMS of the incoming signal would be biased towards the larger values. There are various robust filtering algorithms for RFI filtering[9][10], but the one we are interested is Median Absolute Deviation[11]. A threshold of 3σ would then be applied, as we are aware that a Gaussian probability density curve falls to less than 1% beyond this value. The robustness of MAD estimator will then determine the quality of the filtering.

Chapter 3

Sensitivity and Radiometer Equation

3.1 Radiometer

Natural radio emission from the cosmic microwave background, discrete astronomical sources, the Earth's atmosphere, and the ground is random noise that is nearly indistinguishable from the noise generated by a warm resistor or by receiver electronics. A radio receiver used to measure the average power of the noise coming from a radio telescope in a well-defined frequency range is called a radiometer. Noise voltage has zero mean and varies randomly on the very short time scales comparable with the inverse bandwidth of the radiometer[7].

By averaging a large number N of independent noise samples, an ideal radiometer can determine the average noise power with a fractional uncertainty as small as $\sqrt{\frac{2}{N}} \ll 1$ and detect a faint source that increases the antenna temperature by a tiny fraction of the total noise power. This will be further explained in the following sections.

The simplest radiometer consists of four stages in series, as can be seen from the figure:

1. An ideal(lossless) bandpass filter that passes input noise only in the desired frequency range

2. An ideal square-law detector whose output voltage is proportional to the square of its input voltage which effectively is the input power

3. A signal integrator that smoothes out the rapidly fluctuating detector output

4. A voltmeter or other device to measure and record the smoothed voltage

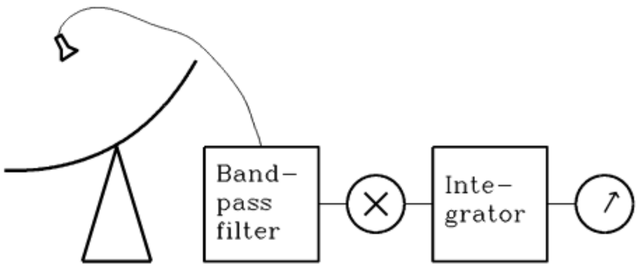


Figure 3.1: A simple Radiometer

3.1.1 Ideal Radiometer Equation

Ideal Radiometer equation* for a total-power receiver is[7]

$$\sigma_{rms} = \frac{T_{sys}}{\sqrt{\Delta\nu_{RF}\tau}}$$

Where: T_{sys} - is the system noise of the receiver
$\Delta\nu_{RF}$ - is the Bandwidth
τ - is the integration time

Please note that in the time interval τ there are $N = 2\Delta\nu_{RF}\tau$ independent samples of the total noise power T_{sys}. Also, as per the Central Limit Theorem, for a heavily smoothed ($\Delta\nu_{RF}\tau \gg 1$) output voltages also have a nearly Gaussian amplitude distribution. Therefore the weakest detectable signal only has to be several times the output rms σ_{rms} and not several times T_{sys}.

3.1.2 What is Signal to Noise Ratio(SNR)?

As is obvious from the name, SNR is the ratio of signal to noise received at the output of a total power detector attached to a radio telescope. From the radiometer equation we know that for a total power detector with instantaneous rms $\mathrm{T_{sys}}$, the rms after integrating a signal of bandwidth $\Delta\nu$ Hz for τ seconds is $\mathrm{T_{sys}}/\sqrt{\Delta\nu\tau}$. The increase in system temperature is just GS, where S is the flux density of the source and G is the gain of the telescope. The signal to noise ratio is hence[4]

$$\mathrm{SNR} = \frac{\mathrm{GS}\sqrt{\Delta\nu\tau}}{\mathrm{T_{sys}}}$$

This is the fundamental equation for the sensitivity of a single dish telescope.

Provided the signal to noise ratio is sufficiently large, one can be confident of having detected the source.

3.2 Integration Time vs SNR

We know[11], that on increasing the integration time, the variance of a gaussian distribution decreases. This is clear from the following figure:-

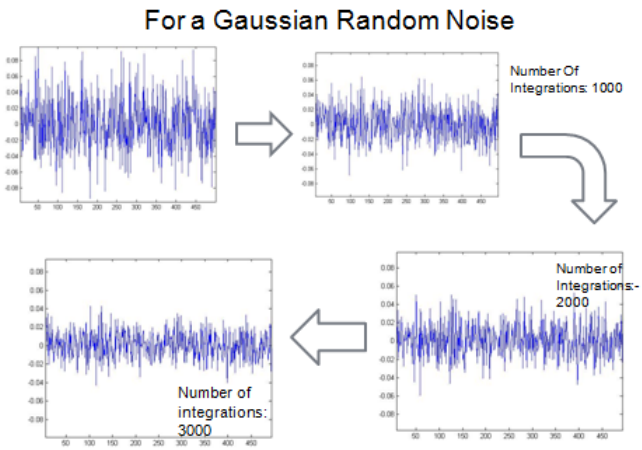


Figure 3.2: Effect of increasing integration time on Gaussian Distribution

Therefore, as could also be seen from the formula given in the previous section, SNR increases by $\sqrt{\tau}$ on increasing the number of integrations, so RFI being non Gaussian in nature stands out as the variance of the gaussian distribution decreases on larger integration times.

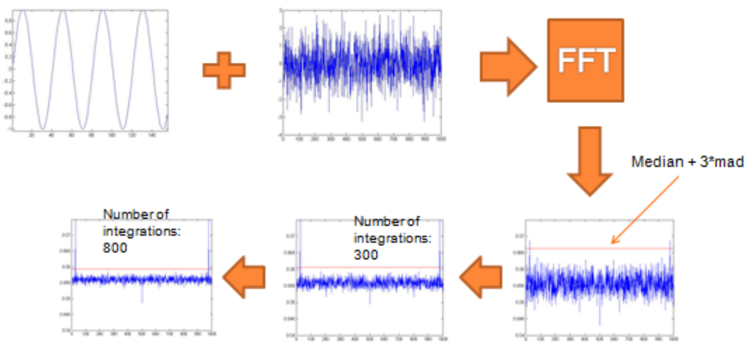


Figure 3.3: Effect of increasing integration time on a signal with RFI

3.2.1 Optimum Integration Time For Given SNR

A MATLAB simulation was done to find the number of integrations required for a given input Signal to Noise ratio. In it a gaussian random noise was generated, and a sine wave (of amplitude depending upon the amplitude of random noise as per a pre-defined SNR) is added to it for all the time samples. An FFT was taken of the sum, showing a peak as expected. A certain number of integrations are then done, followed by MAD filtering (with 3σ threshold). If the peak of sin is removed, then the value of ratio is incremented and the whole process is repeated, if the peak is not removed by the filter then the number of integration is increased, and the signal is integrated till the Peak(due to sine) is removed. This gives us amount of integration required, that given an input SNR, MAD filter would be able to detect it.

$$SNR = \frac{rms\ of\ sin\ wave}{rms\ of\ random\ noise}$$

The relation can be observed from the figure

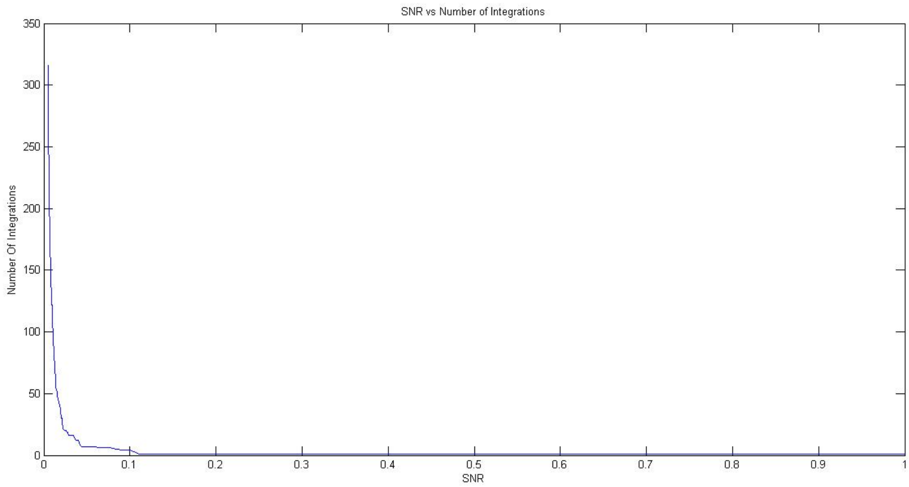


Figure 3.4: Number of integrations Vs SNR

As can be observed, as the SNR increases, the value of integration time required to filter the peak is reduced, which is intuitive. Also, this realtionship is not exactly linear or straightforward. Several plots of $y = \frac{1}{x}$,$y = \frac{1}{x^2}$,$y = \frac{1}{\sqrt{x}}$, etc were overlapped, but none was found consistent throughout the plot. This relationship could be worked out in future.

3.3 Minimum Detectable Signal Estimation

For a given radio telescope, system temperature can be written as

$$T_{sys} = T_{sky} + T_{spill} + T_{loss} + T_{rec}$$

wherein RFI has to be detected. The rms fluctuation of this T_{sys} is given by the radiometer equation:-

$$\sigma_{rms} = \frac{T_{sys}}{\sqrt{\beta\tau}}$$

Therefore the Minimum Detectable Signal(MDS) would be,

$$MDS > median \pm \eta * \sigma_{rms}$$

$$MDS > M \pm \eta * \frac{T_{sys}}{\sqrt{\beta\tau}}$$

Using the above formula for self-correlation, a look-up table could be generated using various values of β, τ for the MDS, to estimate the power level beyond which points will be flagged as RFI before the actual filtering is applied on the data. The values in the following tables are that of SNR(dB) $= 10 * \log \frac{\sqrt{\beta\tau}}{\eta}$.

Channels:- 2048			
.	100MHz	200MHz	400MHz
0.671sec	17.8058	19.3109	20.8161
2.6840sec	20.8161	22.3212	23.8264
16sec	24.6927	26.1979	27.7030

Channels:- 4096			
.	100MHz	200MHz	400MHz
0.671sec	16.3006	17.8058	19.3109
2.6840sec	19.3109	20.8161	22.3212
16sec	23.1876	24.6927	26.1979

Channels:- 16384			
.	100MHz	200MHz	400MHz
0.671sec	13.2903	14.7955	16.3006
2.6840sec	16.3006	17.8058	19.3109
16sec	20.1773	21.6824	23.1876

Chapter 4

Generating and Bookkeeping of Flags

From the real data being used, a 2D matrix containing Time Traces on one axis and Frequency channels on the other axis was generated. After filtering, another 2D matrix of same dimension was generated which contains the flags indicating whether the point was marked as RFI or not by the MAD filter. This 2D flag matrix will have 1 for the point which was left untouched after filtering, and 0 for the point which was removed after filtering. This information could be used to visually check if an RFI point shown in the original time-freq plot was marked in the Flag plot. Also this flag matrix is written out for the user.

4.1 Bookkeeping of Flags

The writing out of 2D flag matrix could be done in two ways:-

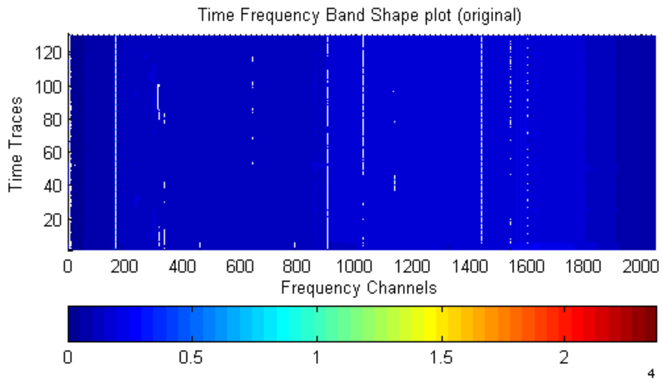
1. Convert these 2D flags to a 1D array by aligning all the channels of a time trace behind the channels of the previous time trace and so on. Here each point will represent a bit, therefore this output file will be of $\frac{1}{8}^{th}$ size of the input integrated file. This will occupy more space but it would provide the user an accurate location information of RFI points in time-freq space.

2. Another simpler way is to define a parameter(X) for the maximum allowable percentage of RFI in a channel across all the time traces. If a channel is found with more than X% of RFI then it must be flagged as "BAD". The output file will have as many bits as number of channels(1k,2k...) each indicating whether the channel was bad or good as per the above criterion. This would take very less space, but would have limited functionality, as

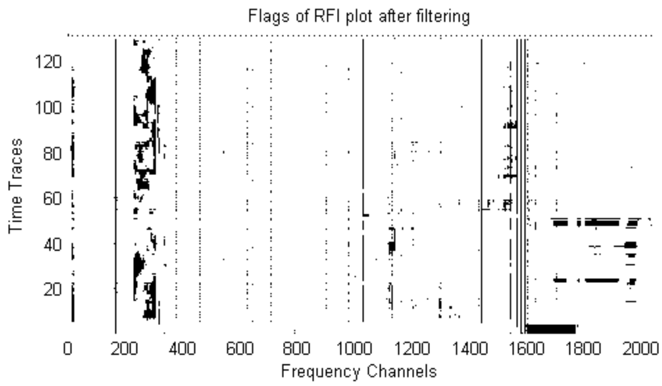
whole information of the channel was being discarded if it exceeded the threshold.

Figure 4.1: A typical 2D flag plot

(a) Time - Frequency plot of original data



(b) Flags generated after filtering



4.2 Data used for the project

The plots given in this report are from the following real data sets:-

1. $b1508 + 55\_10jun15.raw.1.001$ - a 500mb binary file obtained from splitting the original data set $b1508 + 55\_10jun15.raw.1$

2. $B1133 + 16\_gwb\_03jun2015\_pa.raw.001$ - a 400mb binary file obtained from splitting the original data set $B1133 + 16\_gwb\_03jun2015\_pa.raw$

The data has an integration time of 0.00131072 sec initially, then more number of integrations are done in the program to reach an integration time of 1sec.
Number of channels - 2048
Bandwidth - 200MHz

Chapter 5

MAD Based Filtering

MAD [1] (Median Absolute deviation) is a measure of statistical dispersion of the variability of a quantitative data set. Moreover, the MAD is a robust statistic, being more resilient to outliers in a data set than the standard deviation. In the standard deviation, the distances from the mean are squared, so large deviations are weighted more heavily, and thus outliers can heavily influence it. In the MAD, the deviations of a small number of outliers are irrelevant.
For a univariate data set $X_1, X_2, ...X_n$, the MAD is median of absolute value of deviations from the datas median.

$$MAD = median_i|X_i - median_j(X_i)|$$

The Robust Standard Deviation is given by (for Gaussian Distribution)[11]

$$\sigma = 1.4826 * MAD$$

5.1 Spectral Domain RFI Removal using MAD at GMRT

Two methods of MAD based spectral RFI filtering is proposed to be carried out in the GMRT wideband backend(GWB)[6].

- MAD over time:- In this case, individual threshold calculation is done by using median and MAD calculated for each spectral channel as shown by the equation:-

$$T(i) = median(i) \pm \eta * \sigma_{mad}(i)$$

[1] in MATLAB, use the command `mad(array,1)` for Median Absolute Deviation.
Note:- `mad(array)` gives Mean Absolute Deviation(and not Median Absolute Deviation) in MATLAB

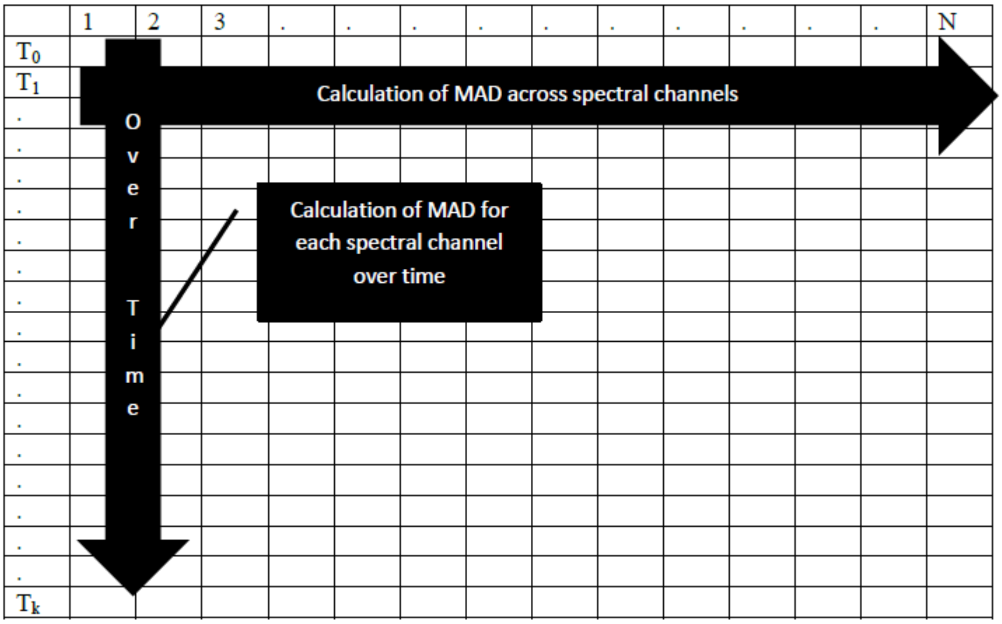


Figure 5.1: Two proposed schemes for spectral domain filtering

This technique is computationally cumbersome, particularly in real-time as there is a requirement to calculate MAD for each spectral channel.

- MAD across channels:- In this scheme, threshold calculation is done by using median and MAD calculated across the channels as shown by equation

$$T = median \pm \eta * \sigma MAD$$

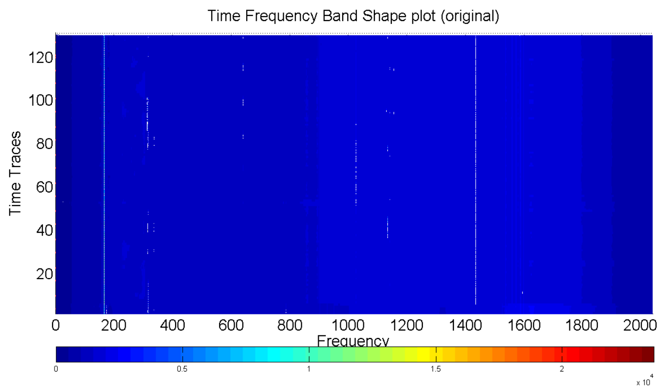
Threshold computation and filtering are carried out independently on the real and imaginary parts of the signal at the FFT output. This is computationally light as compared to the MAD-over-time since only one MAD computation for real part and one for imaginary part is required in real-time. Also, in case of persistent RFI, this method performs better than the MAD across channels[6].

5.2 MAD Filtering Across Channel(MFAC)

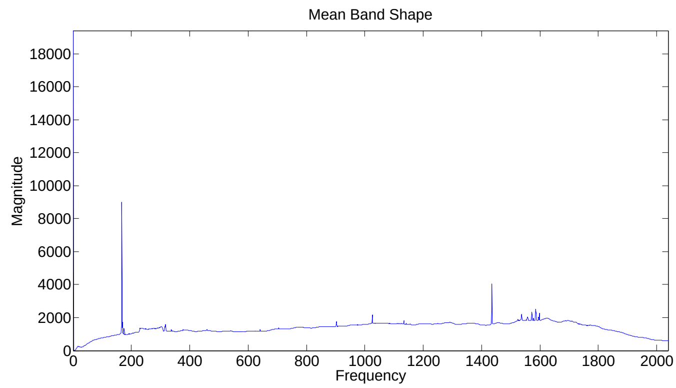
As explained in the previous section, MFAC is a better technique over MAD across time. Some results of filtering using MFAC on real data are shown in the following pages.

Figure 5.2: MFAC on $b1508 + 55\_10jun15\_raw\_1$

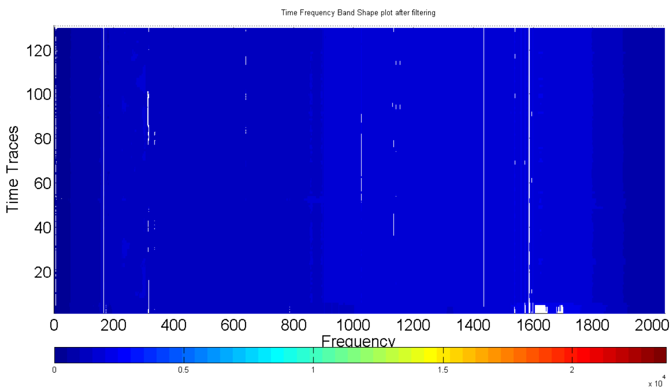
(a) Time Frequency Plot



(b) Mean Band Shape Original



(c) Time Frequency plot after filtering



(d) Mean Band Shape after filtering

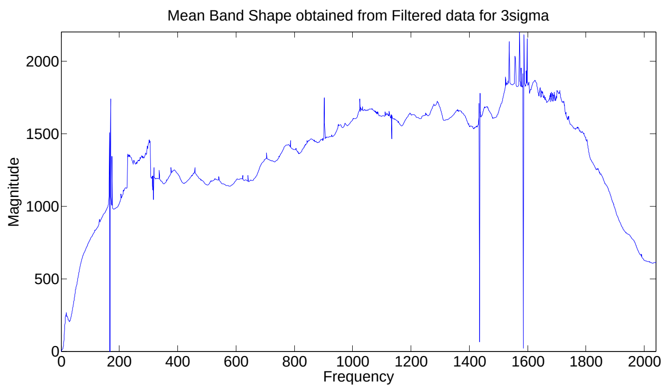
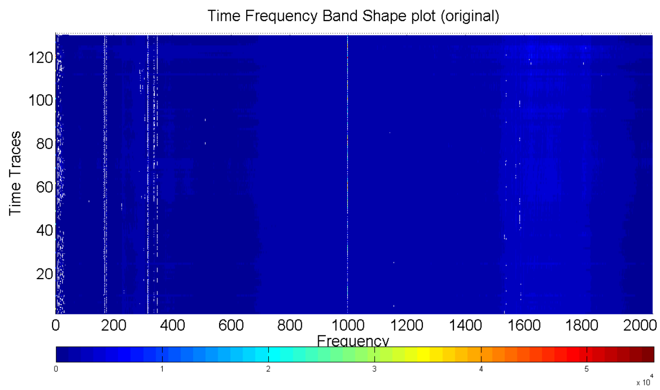
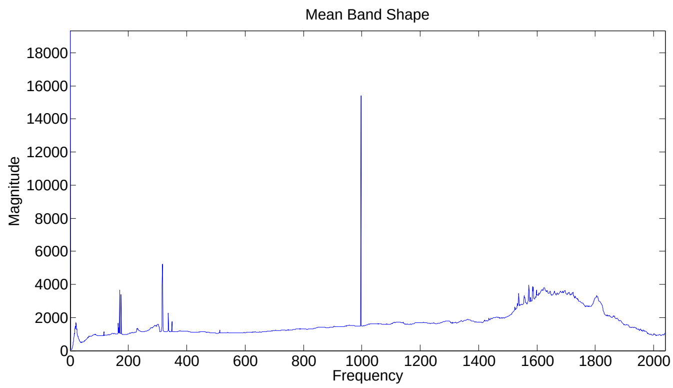


Figure 5.3: MFAC on $B1133+16\_gwb\_03jun2015\_pa\_raw\_001$

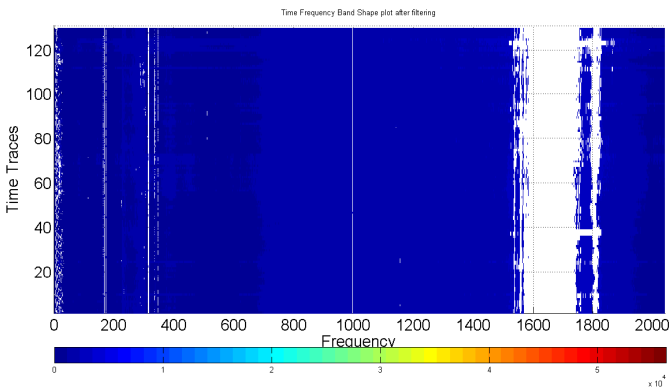
(a) Time Frequency Plot



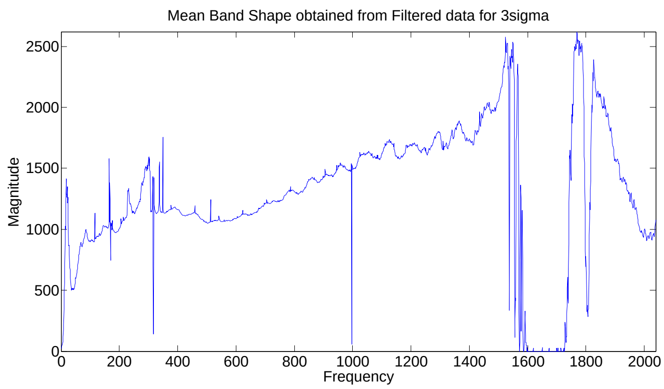
(b) Mean Band Shape Original



(c) Time Frequency plot after filtering



(d) Mean Band Shape after filtering



5.3 Observations

1. The MFAC is able to detect and remove very strong RFI lines, while many medium and small RFI lines could still be observed after filtering.
2. As can be seem from fig 5.3(c), this filtering removes the high power points which are a part of band shape and are not RFI.
3. It is not able to account for irregularities in the band shape while filtering, which leads to loss of useful data.

Therefore some other technique like normalsation needs to be explored, which can account for the irregularities in band shape, and also is more sensitive to low level RFI.

Chapter 6

Normalisation of the Band Shape

Normalisation of the Band Shape is a new approach for RFI filtering, which was tested and analysed on various real and simulated data sets, and the results obtained will be presented in this chapter. The procedure behind this method involves calculating a temporal band shape(by mean or median of all the time traces). Applying some method, to filter it from RFI. Using this RFI free, smooth bandshape to normalise the band shapes at all time traces by diving them by this template Band Shape. This would be advantageous as now ideally the mean and median of the data set will scale down to 1, and RFI peaks will stick out of the data. Therefore Median Absolute Filtering approach across channel should be highly effective in filtering of this normalised data set.
Primarily there can be two approaches for Normalising the band shape:-

1. Using Mean Band Shape

2. Using Median Band Shape

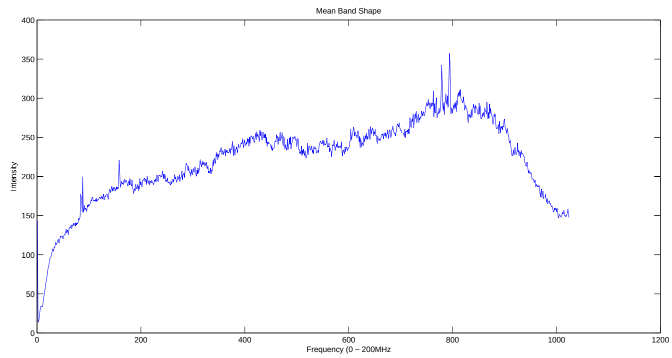
A small 'Cheat' was applied before starting the analysis of normalisation using mean and median band shapes:- From a simulated data, a mean band shape was generated and the RFI points were manually removed from this band shape. Normalisation and filtering was then done using it. This exercise was done, to check prematurely the effects of normalisation before other rigorous analysis was done.

6.0.1 Why use RFI Free Band Shape for Normalisation?

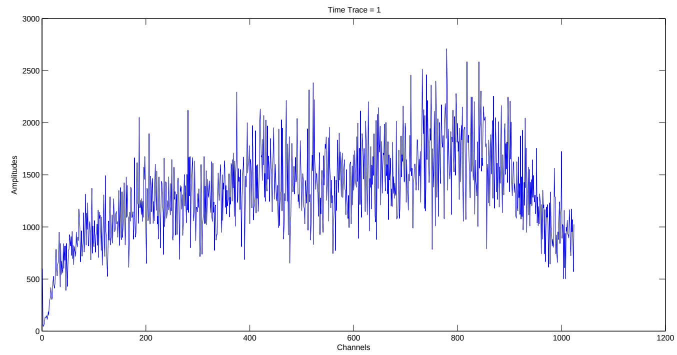
It is necessary to use a RFI free band shape for normalisation as otherwise, in the normalised band shape the channels with RFI will be masked and a new mean bandshape constructed from the normalised data will NOT show the RFI in an easily detectable manner. This is demonstrated by the following plots:-

Figure 6.1: Effect of normalising with a RFI corrupt band shape

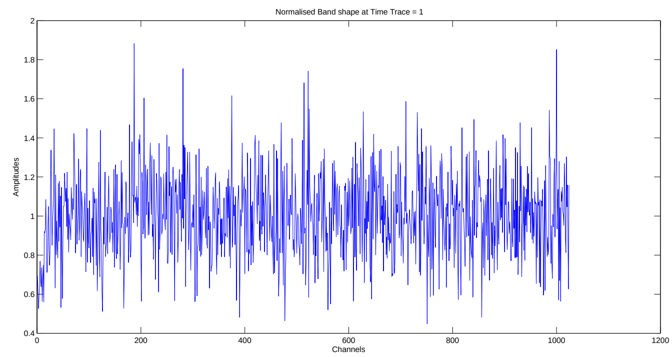
(a) Mean Band Shape (Note the RFI)



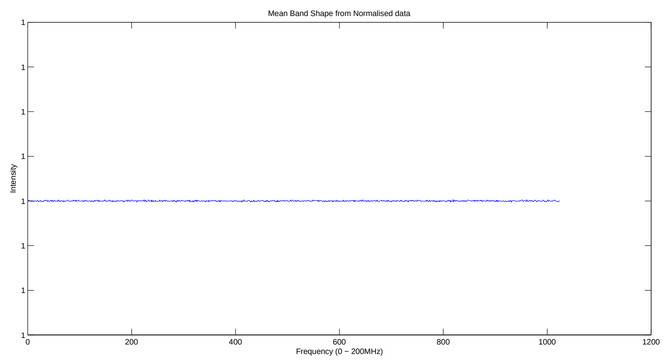
(b) Band Shape of a single time trace



(c) Band Shape of a single time trace after normalisation



(d) Mean Band Shape after normalisation



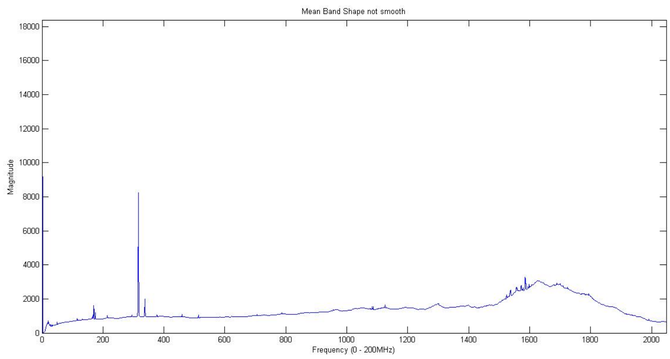
6.1 Mean Band Shape

In this case, the template band shape is obtained by taking mean across time of all the band shapes, thus Mean Band Shape. But the Mean Band Shape obtained by this procedure is corrupted by RFI, and has a lot of fluctuations, as can be seen from the figure.

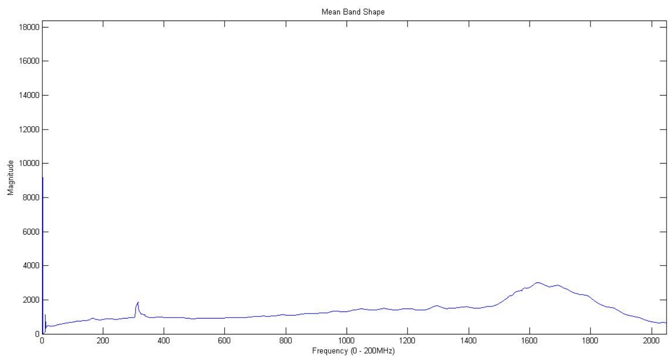
A method to filter this Mean Band shape from RFI and also to reduce the variations, is to take a running mean on the whole width. Running mean implies that, each point on the band is replaced by the average of a window of points around that point. For example for channel number 80, we take average of channels from 70 to 90 and replace the value at the channel by this average. This procedure is repeated for all the channels of the band, resulting in a smooth and RFI free band shape as given in the figure.

Figure 6.2: Mean Band Shapes before and after smoothing

(a) Mean Band Shape Corrupted by RFI



(b) Mean Band Shape smoothed



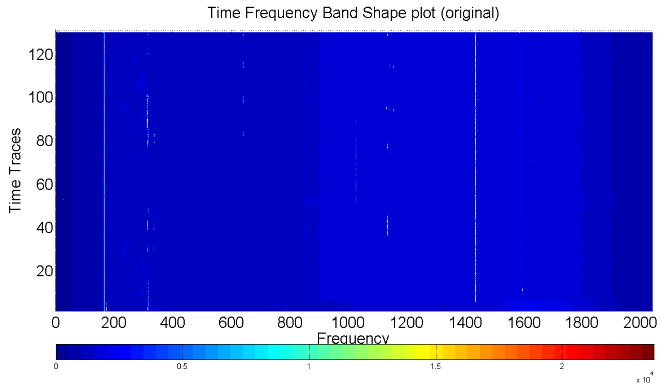
As can be observed from the two plots RFI is removed by using running mean on the whole band. But this smooth band shape would still be biased by high RFI points because mean is being used which is not a robust estimator.

6.1.1 Normalisation using Mean Band Shape

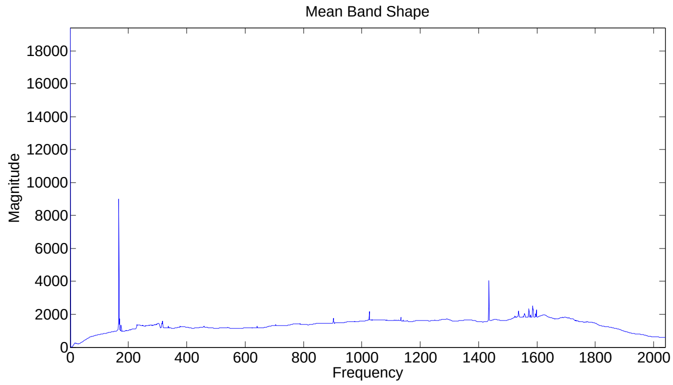
After the mean band shape is obtained, band shape for each time trace is normalised by dividing it with the mean band shape. This normalised matrix is passed through a MAD filter, which filters the points above 3σ threshold, and generates a 2D Flag matrix as discussed in Chapter 4. Filtering is also done on the unnormalised data, to compare the effects of normalisation. The following pages shows the plots for this scheme applied to 3 data sets.

Figure 6.3: Mean Band Shape Filtering on $b1508 + 55\_10jun15\_raw\_1$

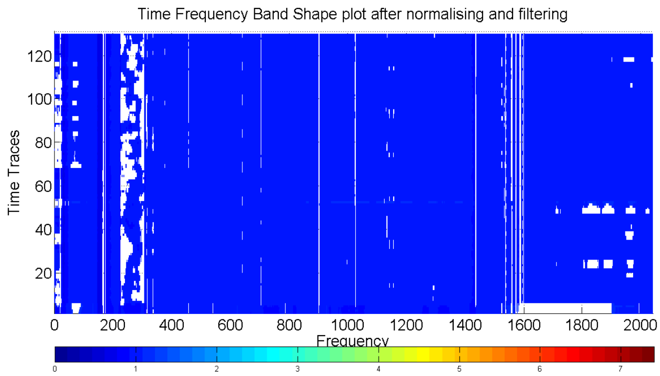
(a) Time Frequency Plot



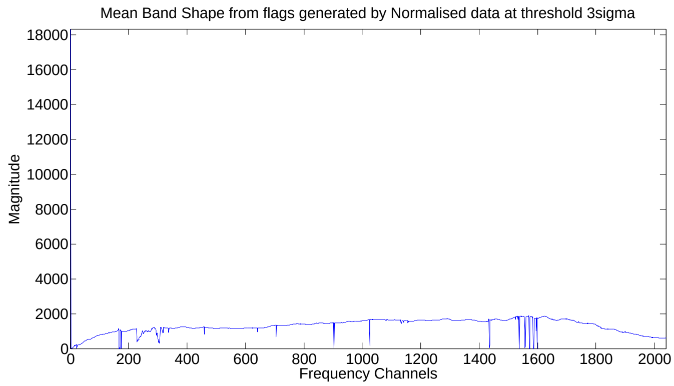
(b) Mean Band Shape Original



(c) Time Frequency plot after Normalisation and Filtering



(d) Mean Band Shape from Flags obtained from Filtering using Normalisation



(e) Normalised Mean Band Shape

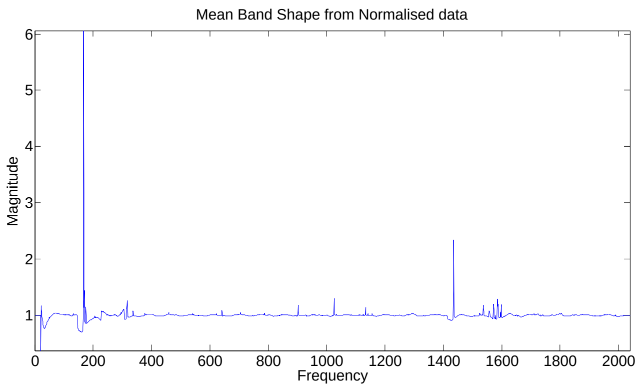
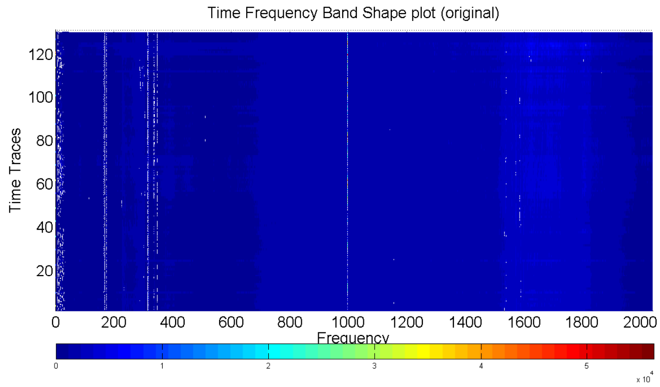
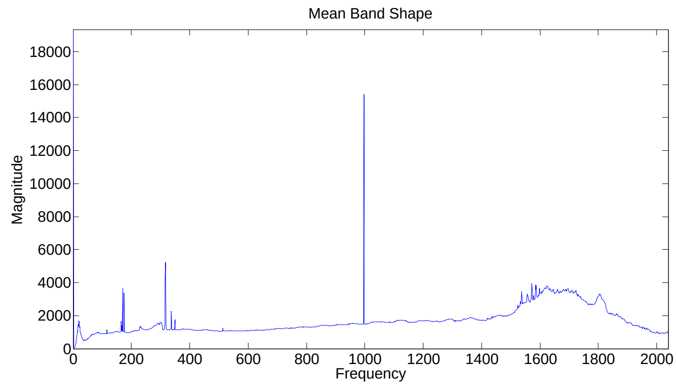


Figure 6.4: Mean Band Shape Filtering on $B1133 + 16\_gwb\_03jun2015\_pa\_raw\_001$

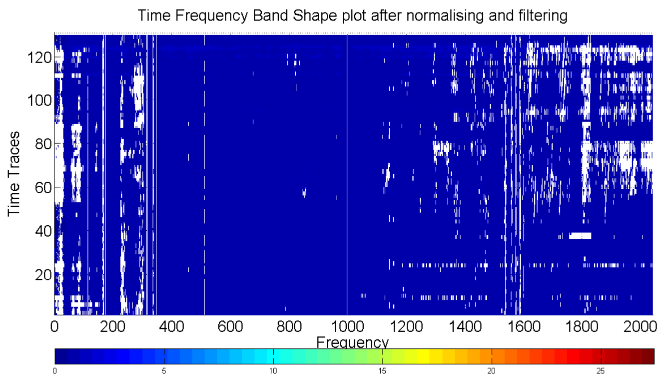
(a) Time Frequency Plot



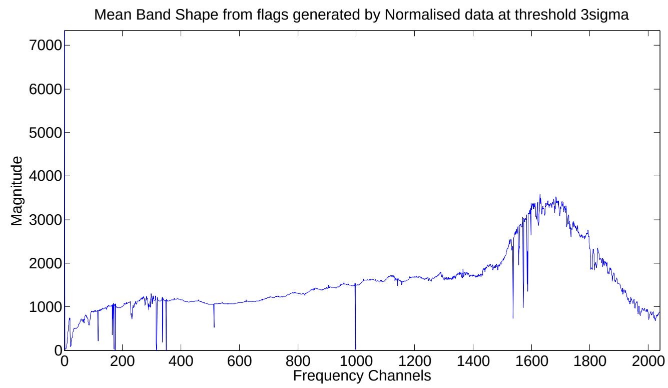
(b) Mean Band Shape Original



(c) Time Frequency plot after Normalisation and Filtering



(d) Mean Band Shape using Flags from Filtering using Normalisation



6.1.2 Observations

1. RFI peaks are distinctly visible in the normalised band shape(fig 6.3(e)), and are thus filtered out.

2. This filtering scheme is able to remove high as well as low power RFI successfully.

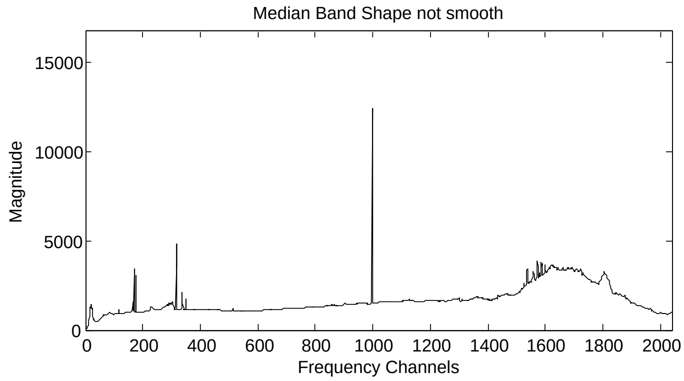
3. Even in the case of band shape irregularity, the filtering gives much better results than MFAC.

6.2 Median Band Shape

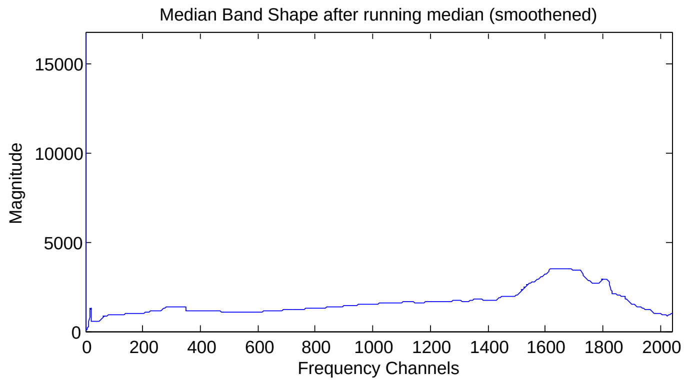
In this case, similar to the mean band shape the template band shape is obtained by taking median(instead of mean) across all the time traces. Though the median band shape obtained is much "cleaner" than the mean band shape, still it is corrupted by RFI. Again, similar to the case with mean band shape,here we will use running median to smoothen the band. The resulting band shape can be seen in the corresponding figure.

Figure 6.5: Median Band Shapes before and after smoothing

(a) Median Band Shape Corrupted by RFI



(b) Mean Band Shape smoothed

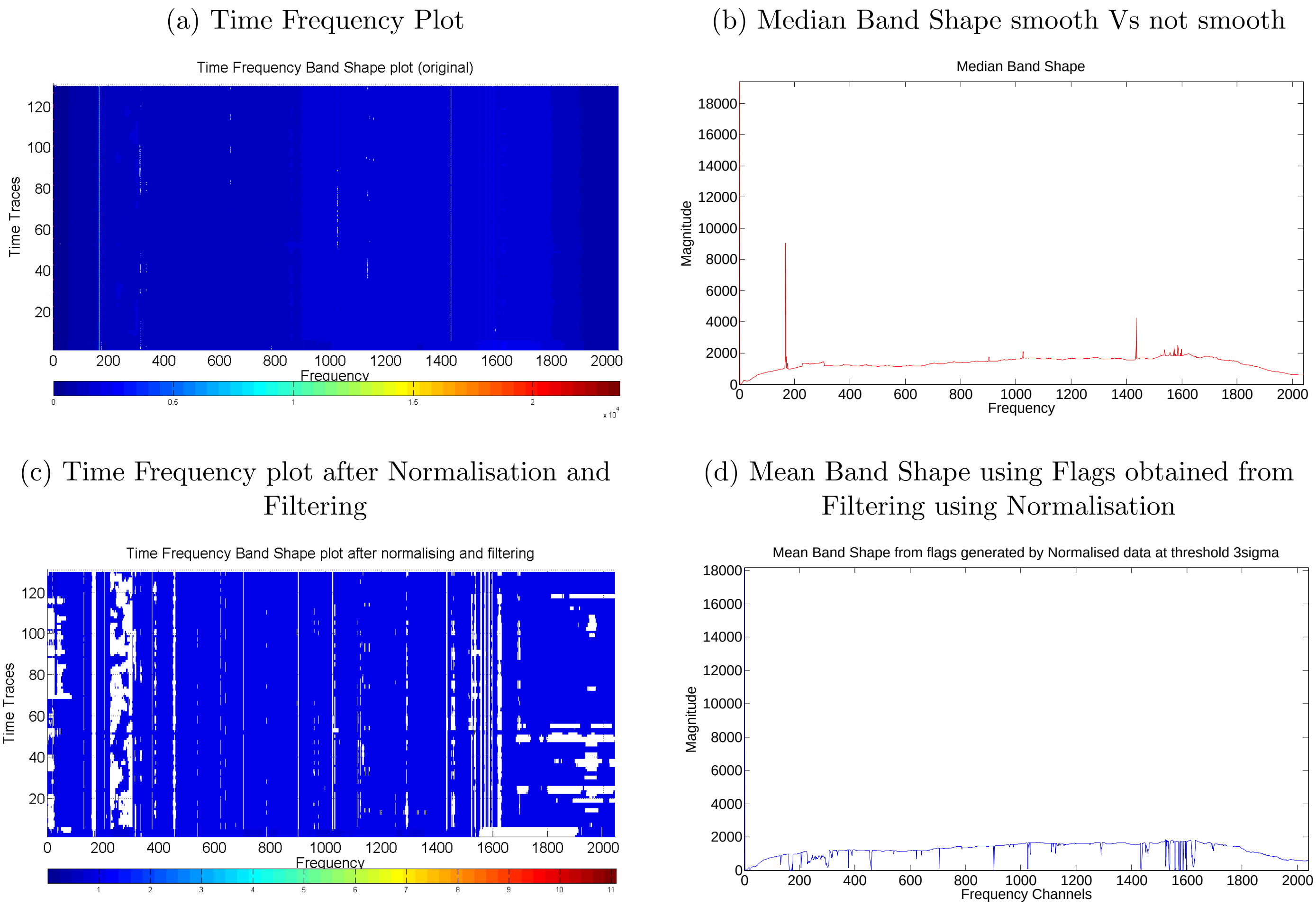


Though we are using median, which is a robust estimator still our smoothened median band shape would be biased by RFI, but the bias in this case would definately be less than in the previous case.

6.2.1 Normalisation using Median Band Shape

After the median band shape is obtained, band shape for each time trace is normalised by dividing it with the median band shape. This normalised matrix is passed through a MAD filter, which filters the points above 3σ threshold, and generates a 2D Flag matrix as discussed in Chapter 4. Filtering is also done on the unnormalised data, to compare the effects of normalisation. The following pages shows the plots for this scheme applied to a few data sets.

Figure 6.6: Median Band Shape Filtering on $b1508 + 55\_10jun15\_raw\_1$



(a) Time Frequency Plot

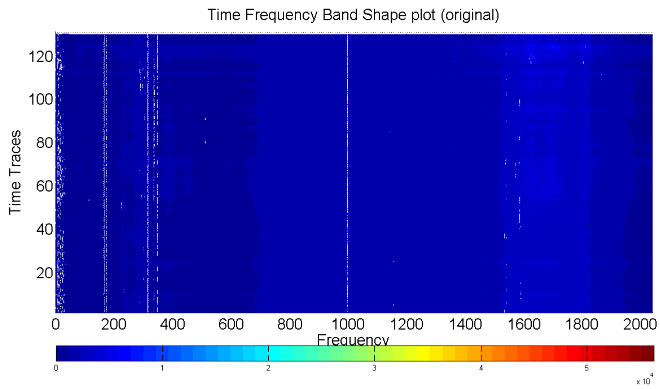
(b) Median Band Shape smooth Vs not smooth

(c) Time Frequency plot after Normalisation and Filtering

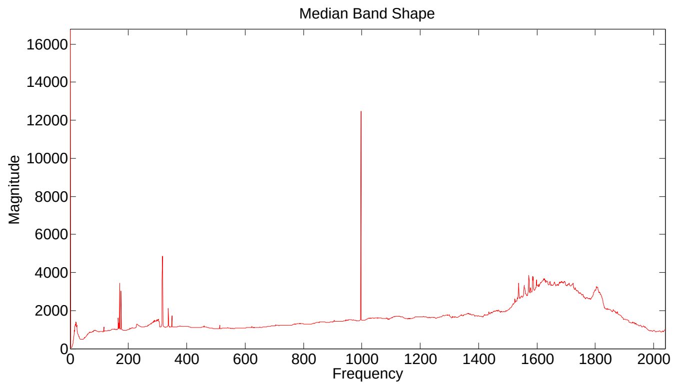
(d) Mean Band Shape using Flags obtained from Filtering using Normalisation

Figure 6.7: Median Band Shape Filtering on $B1133+16\_gwb\_03jun2015\_pa\_raw\_001$

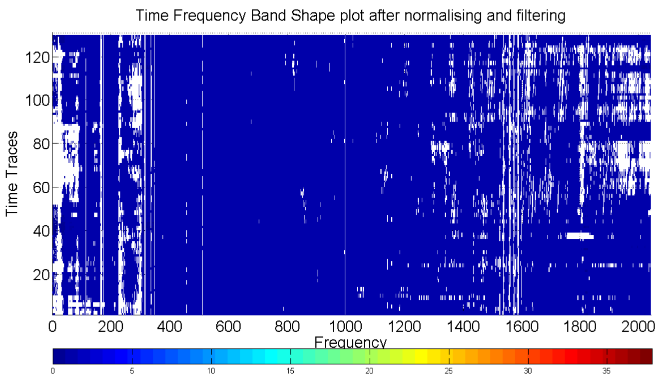
(a) Time Frequency Plot



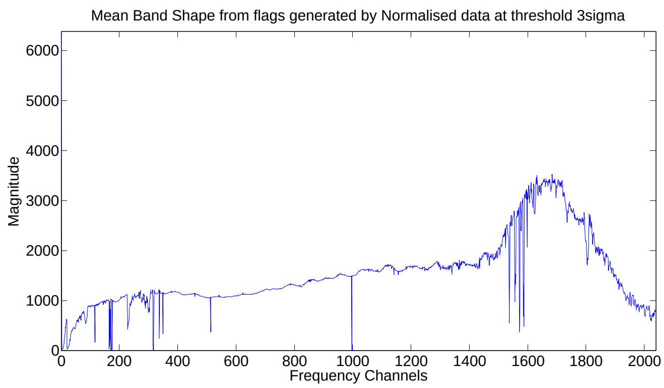
(b) Median Band Shape smooth Vs not smooth



(c) Time Frequency plot after Normalisation and Filtering



(d) Mean Band Shape using Flags obtained from Filtering using Normalisation



6.2.2 Observations

1. Although median is a more robust estimator than mean, the filtering results in this case are similar in the previous filtering scheme using Normalisation by mean band shape.

2. Most of the low and high power RFI has been successfully filtered.

3. The scheme works fine even in the presence of band shape irregularities.

6.3 Overall Observations

1. Temporal band shape obtained after smoothing is free from RFI, and is being used for normalisation.

2. Filtering after normalisation clearly is more efficient than simple MFAC(MAD filtering across channel).

3. Mean band shape for normalised and filtered data has dips because flagged points were considered as zeros and the band was averaged over all time samples.

4. Filtering using normalisation by Mean Band Shape and Median Band Shape, both show similar results of filtering.

Chapter 7

Using Median Compensation Filtering

In this method, a Median Band Shape is generated by taking the median across time at each channel. This median band shape would be corrupted by RFI and to reduce its effects, a running median window is applied(this technique is discussed in depth in next chapter). Instead of normalising(to be discussed in next chapter), this median band shape would be used during filtering. As discussed in Ch 5.2, MAD across channel is given by:-

$$MAD = median_i|X_i - median_j(X_i)|$$

Now, instead of using the median across channel to calculate the absolute deviations, the median obtained from the median band shape (corresponding to that channel) would be used to calculate the deviation at that channel. Similarly, the absolute deviations could be calculated for all the channels. The MAD obtained by taking median of this absolute deviation matrix, would now be used for filtering.

While calculating the Minimum Detectable signal from the formula-

$$MDS = median_{across\ channel} + 1.4826 * \eta * \sigma_{mad}$$

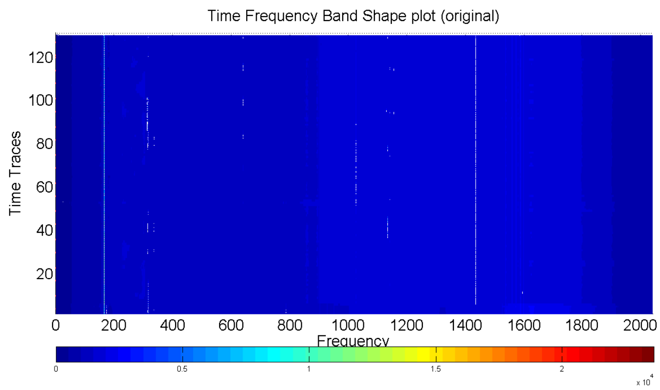
again instead of median(across channel) the value of median of that particular channel would be used from the median band shape. So for i^{th} channel:-

$$MDS_i = median_i + 1.4826 * \eta * \sigma_{mad}$$

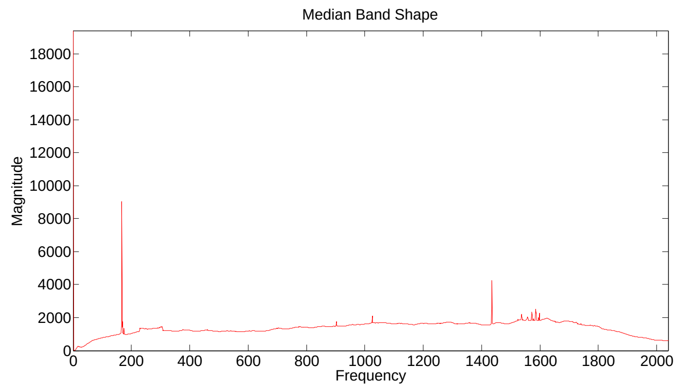
This method is more robust than applying only the MAD filtering. Also it exploits the features of median across time, and filtering across channel both. As we are calculating medians initially across time, but then the filtering is being done across channels(with a change that the median value to be used will be from the median band shape). Some of the results of using this method are given in subsequent pages.

Figure 7.1: Median Compensation Filtering on $b1508 + 55\_10jun15\_raw\_1$

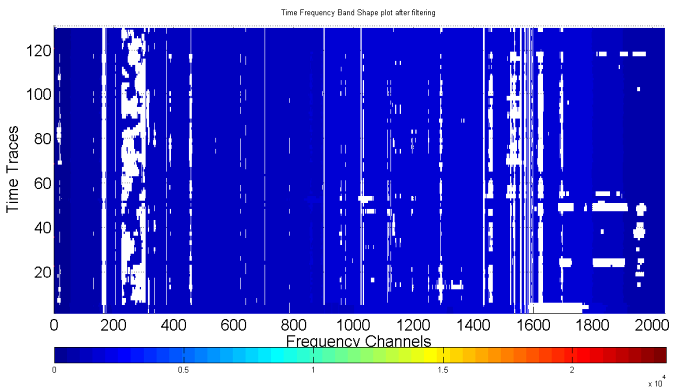
(a) Time Frequency Plot



(b) Median Band Shape



(c) Time Frequency plot after filtering



(d) Mean Band Shape after filtering

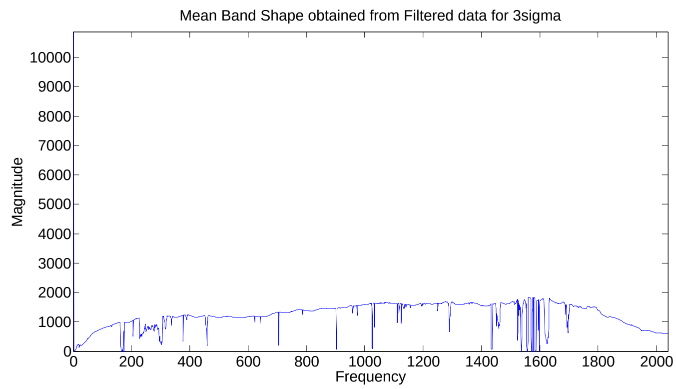
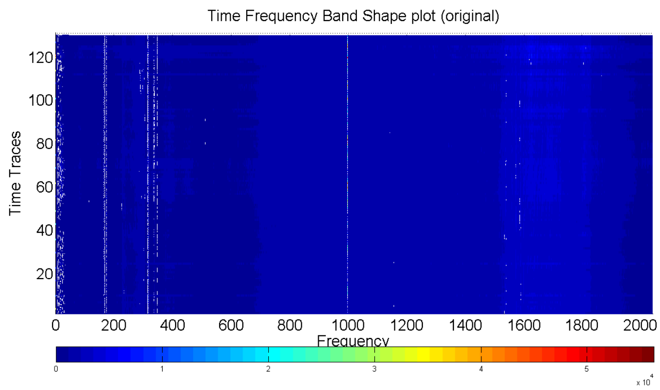
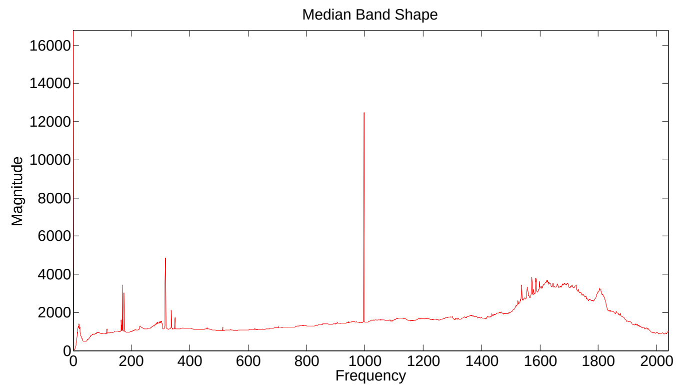


Figure 7.2: Median Compensation Filtering on $B1133 + 16\_gwb\_03jun2015\_pa\_raw\_001$

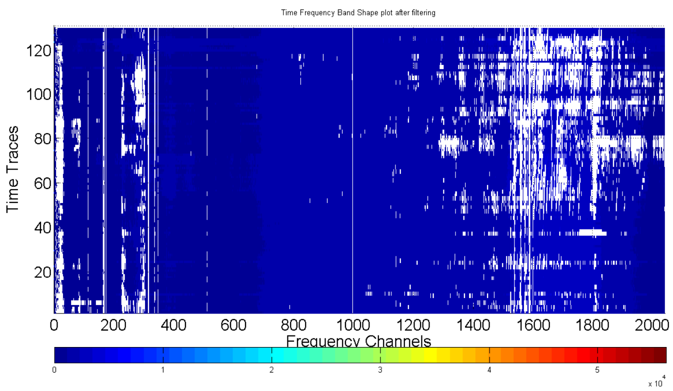
(a) Time Frequency Plot



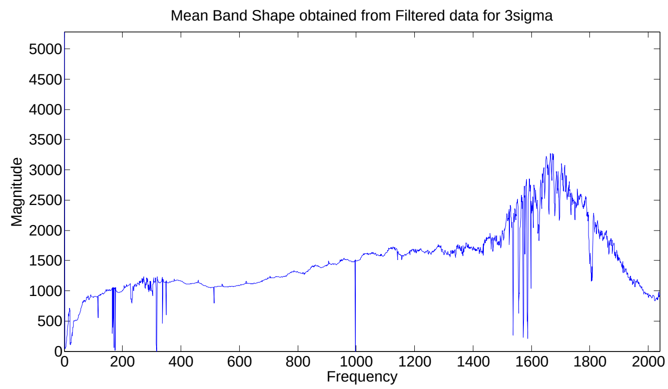
(b) Median Band Shape



(c) Time Frequency plot after filtering



(d) Mean Band Shape after filtering



7.1 Observations

1. The filtering results of this method are similar to the ones obtained using Normalisation Technique.

2. Most of the low and high power RFI has been detected and removed. Also irregularities in the band shape have been taken into account while filtering.

Chapter 8

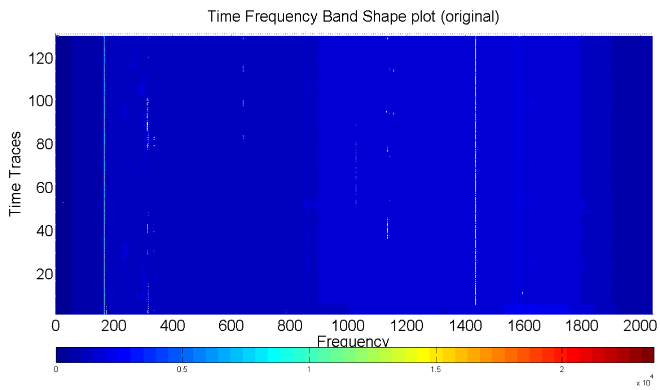
Mean/Rms Ratio Filtering

Another technique to filter the RFI is to evaluate the Mean by RMS ratio across time and filter the plot obtained. In this method, a ratio i.e Mean/RMS [1] value is calculated for each channel across all time traces. A band shape is obtained which will show "Dips" at the locations of RFI, this is because the presence of RFI will increase the RMS (i.e standard deviation) while the mean would not be much increased. So Mean/RFI will drop in the presence of RFI lines at some time traces. This band shape obtained is then subjected to MAD filtering across channel, and the channels marked as RFI are completely flagged as BAD. So although it may lead to some data loss because of whole channel being flagged, it will provide the channels to be used and the flags in a much simpler way using much lesser memory.

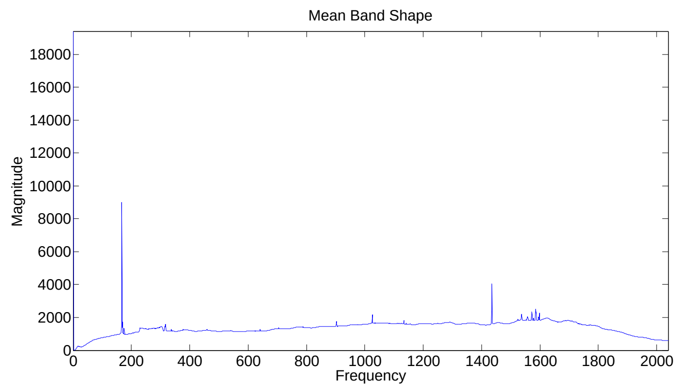
[1] here RMS refers to Standard Deviation

Figure 8.1: Mean/RMS Ratio Filtering on $b1508 + 55\_10jun15\_raw\_1$

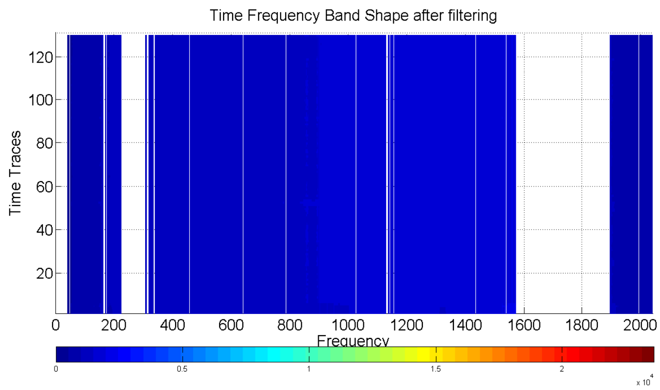
(a) Time Frequency Plot



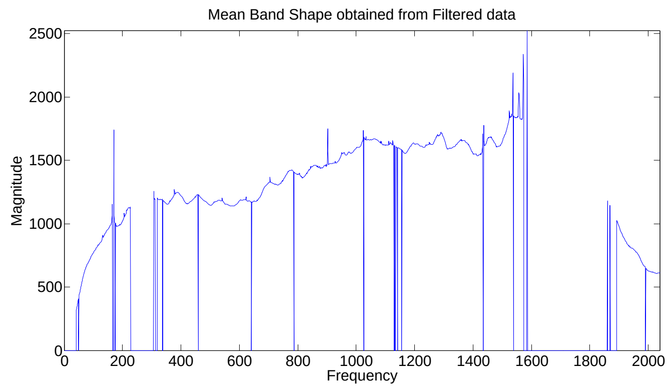
(b) Mean Band Shape



(c) Time Frequency plot after filtering



(d) Mean Band Shape after filtering



(e) Mean/RMS band shape

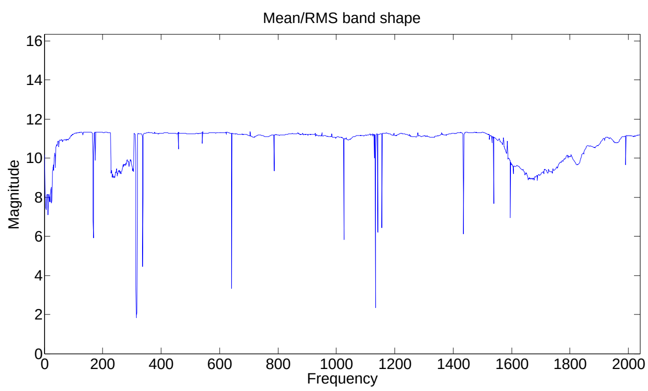
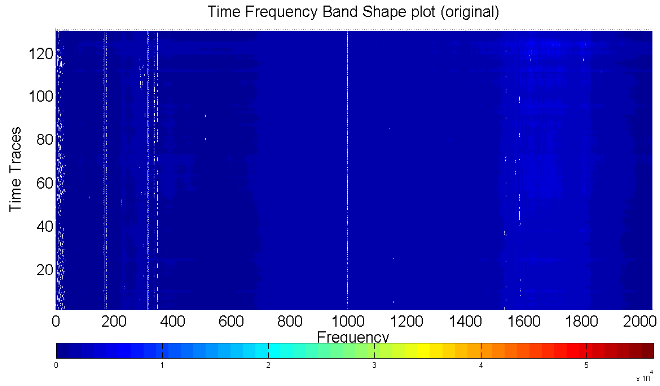
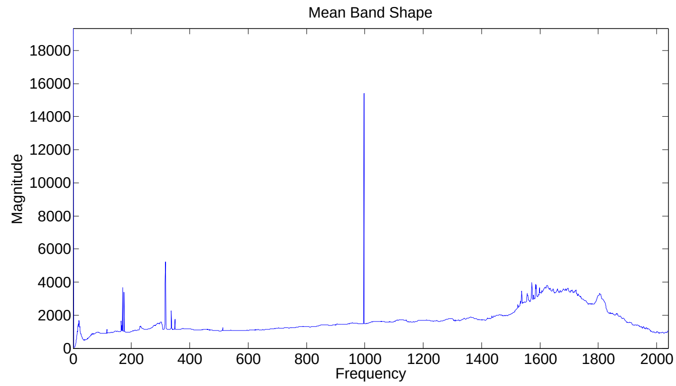


Figure 8.2: Mean/RMS Ratio Filtering on $B1133+16\_gwb\_03jun2015\_pa\_raw\_001$

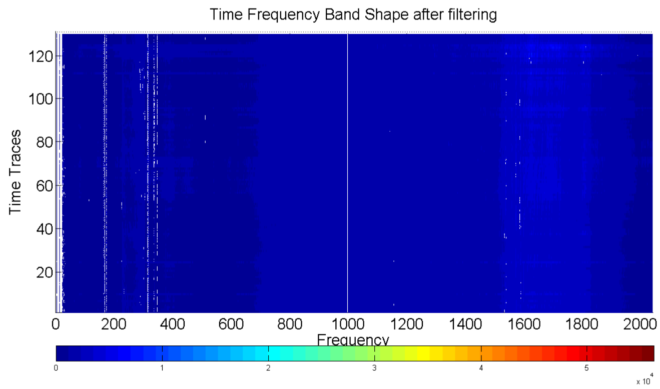
(a) Time Frequency Plot



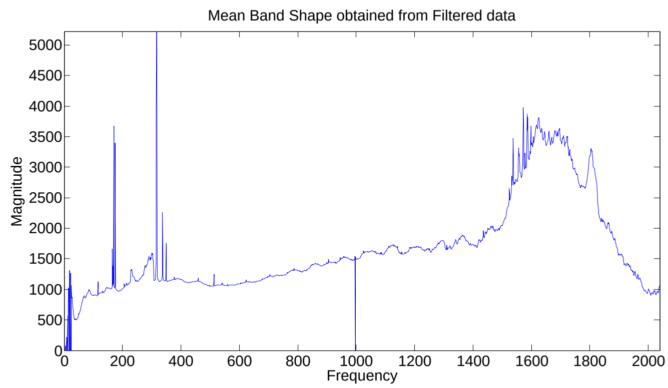
(b) Mean Band Shape



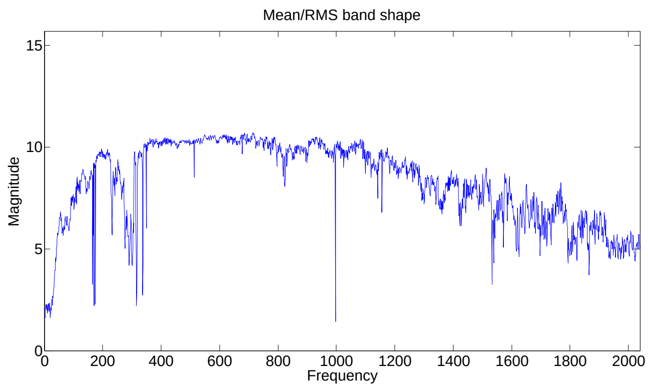
(c) Time Frequency plot after filtering



(d) Mean Band Shape after filtering



(e) Mean/RMS band shape



Chapter 9

Miscellaneous

9.1 Replacement options

This section answers the question of, Once a point is identified as RFI, then what? It is understood that, that point has to be replaced by some other "number", which will not act as RFI. There are a 3 possible replacements:-

1. Replace with Zero
2. Clipping i.e replace with the threshold value at that point which is $Median + 1.4826\ \eta\ \sigma_{mad}$
3. Replace with a random number generated which lies between 0 and threshold

The usefulness of the replacement option could be explored only when some standard post processing is done on the filtered data. Therefore the data written out of the Matlab programs made, is with one of these replacements(User-defined).

9.2 START-STOP feature

A START-STOP feature is added to the Matlab programs made. It gives user liberty to reject channels from starting and end. This facilitates the filtering, as the roll-off errors decrease if only the central part of band shape is chose which is nearly flat. The results of one such simulation is shown in the following figure.

Chapter 10

Conclusion

1. Based on the strength of spectral line it is now proposed to find the location of filter in the GWB receiver chain, using the look-up table generated.

2. MAD filtering across channels(MFAC) was subjected to real data, and the results were reported.

3. Several compensation schemes to account for gain fluctuations in band shape in MFAC were analysed and compared [1], like:-

   (a) MAD filtering across channel

   (b) Normalisation using Mean Band Shape

   (c) Normalisation using Median Band Shape

   (d) Median Compensation Filtering

   (e) Mean/RMS ratio based Filtering

4. It was observed that Normalisation using Median/Mean band shape, and Median compensation Filtering works efficiently as compared to other techniques.

5. Also, it is difficult to characterize the best among these three, though it is understood that Median Band Shape based Normalisation and Median Compensation though appear to differ, are same on statistical level.

[1] Please refer to appendix for filtering results of all the techniques

Chapter 11

Future Plans

1. Analyzing the effects of current procedure if applied on real and imaginary data, and necessary changes required in the current approach.
2. Optimizing the window size for obtaining the Temporal band shape of channels to be used for normalisation.
3. Methodology to obtain the temporal band shape(free from RFI) which has to be used for normalisation. In report median and mean are shown, but still the band shape could be biased towards RFI. So more work is required here.
4. Porting the Matlab code in C and enabling it to be a part of the GWB processing pipeline.
5. Optimization of code for making it amenable to real-time.
6. Work towards a general purpose utility for GMRT observatory. This utility should be agnostic to the RFI filtering algorithm, it should be robust enough to be able to enable the user to use RFI filtering and efficiently process the data products (after RFI filtering) at the output of the GWB.
7. Effects of recursively filtering the data (more than once) could also be explored.
8. Methods to improve the Mean/RMS based filtering technique could also be worked out.

References

1. Baburaj, N (April 2013). *Development of MAD based spectral domain RFI filtering techniques*

2. Ajithkumar, B., Choudhari, S., Buch, K., Gupta, Y. (2013). Next generation digital Backend for gmrt. *Materials Science and Engineering*,44,doi:10.1088/1757-899X/44/1/012024

3. Prasanna, R (2009). *Study of RFI Filtering Algorithms for the GSB*

4. Chengalur, J., Gupta, Y., Dwarkanath, K.S. (2007). *Low Frequency Radio Astronomy*

5. Balaji, S (2015). *Wpmon-An Interface for real-time monitoring, RFI detection and mitigation for GMRT Pulsar Observations*

6. Buch, K (2015). *Spectral Domain RFI filtering for GWB using MAD extimator*

7. *Radiometer*. Retrieved from http://www.cv.nrao.edu/course/astr534/Radiometers.html

8. Thompson, A., Moran, J., Swenson, G. (2004). *Interferometry and Synthesis in Radio Astronomy*

9. Baan, W. (2010). *RFI Mitigation in Radio Astronomy*, RFI mitigation workshop - RFI2010, Groningen, the Netherlands, March 29-31, 2010, DOI: 10.1109/URSIGASS.2011.6051248

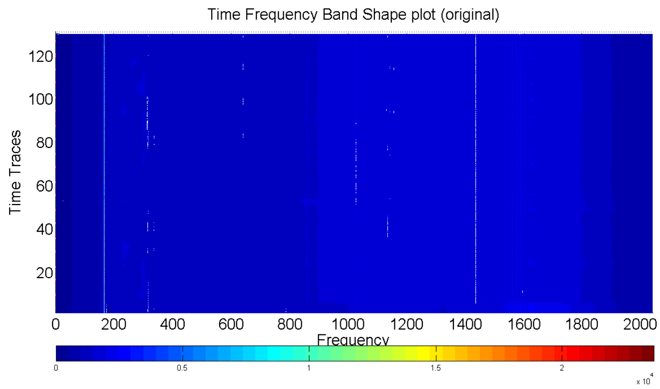
10. Ford, J., Buch, K. (2014), RFI Mitigation Techniques in Radio Astronomy, IEEE Geoscience and Remote Sensing Society, Quebec, Canada, July 13-18 2014

11. Fridman, P. (2008), Statistically Stable Estimates Of Variance In Radio-Astronomy Observations As Tools For Radio-Frequency Interference Mitigation, *The Astronomical Journal*, 135:18101824.,doi:10.1088/0004-6256/135/5/1810

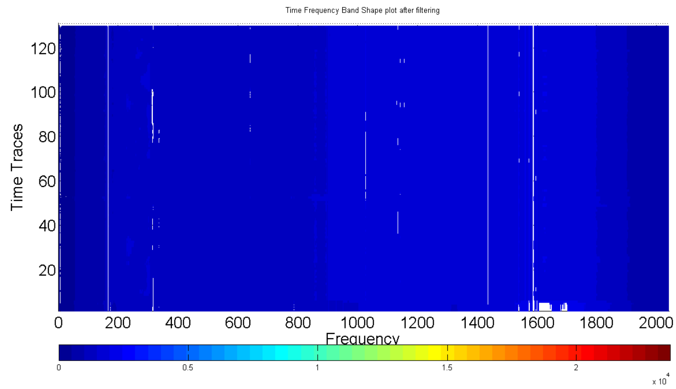
Appendix

Figure 11.1: Filtering Results on $b1508 + 55\_10jun15\_raw\_1\_001$

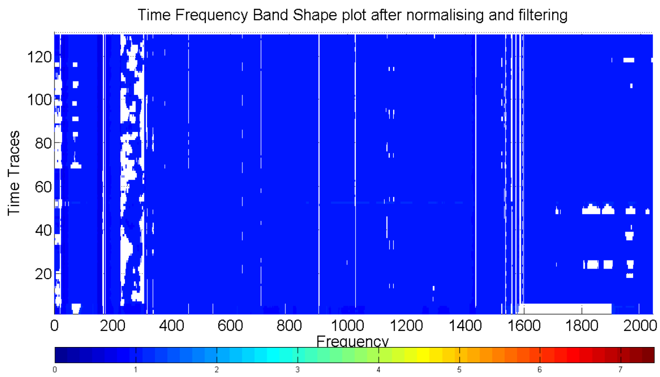
(a) Time Frequency Plot Original



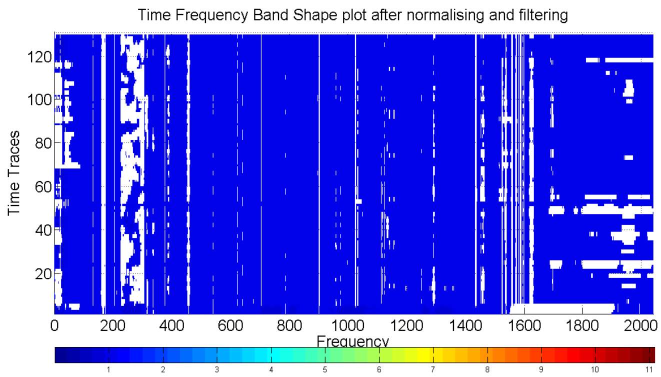
(b) Filtered using MAD Across Channel



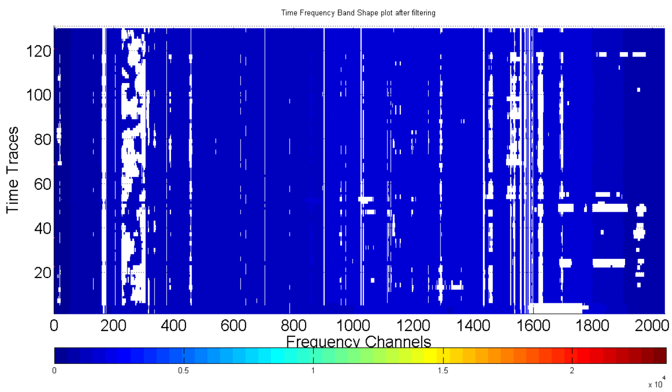
(c) Filtered after Normalisation using Mean Band Shape



(d) Filtered after Normalisation using Median Band Shape



(e) Filtered using Median Compensation Filtering



(f) Filtered using Mean/RMS Ratio Filtering

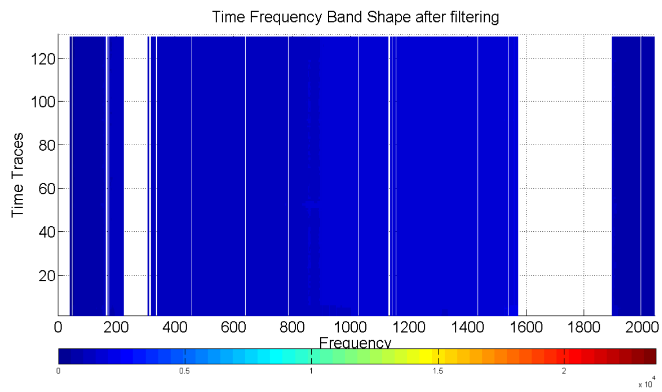
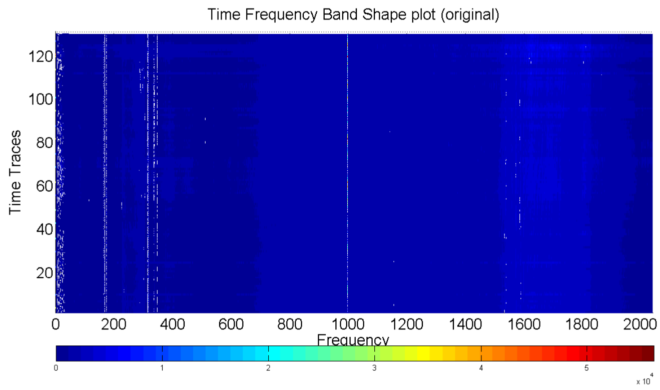
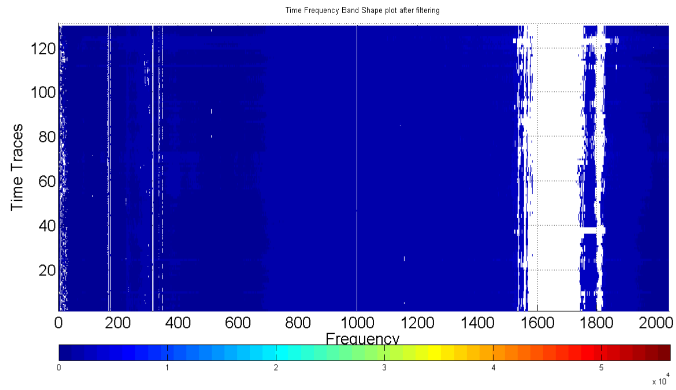


Figure 11.2: Filtering Results on $B1133 + 16\_gwb\_03jun2015\_pa\_raw\_001$

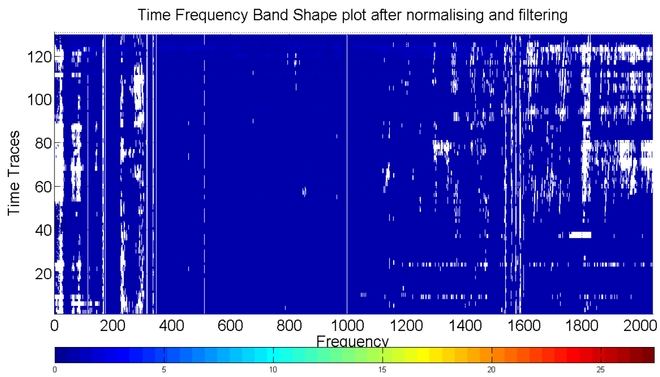
(a) Time Frequency Plot Original



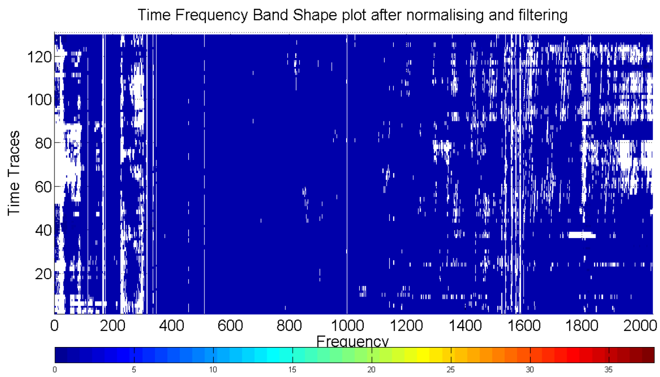
(b) Filtered using MAD Across Channel



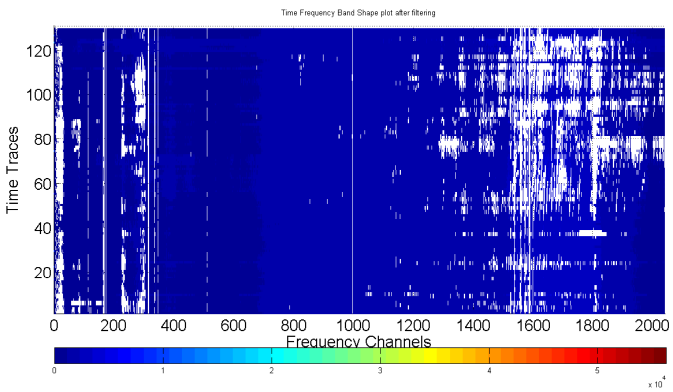
(c) Filtered after Normalisation using Mean Band Shape



(d) Filtered after Normalisation using Median Band Shape



(e) Filtered using Median Compensation Filtering



(f) Filtered using Mean/RMS Ratio Filtering