Chapter 4

Hierarchical RFI mitigation system for MRT

"It is too rash, too unadvised, too sudden."

- William Shakespeare in *Romeo and Juliet*

Due to increasing demand for commercial usage of the electromagnetic spectrum, it is becoming difficult to carry out interference free astronomical observations, especially at low frequencies. In this chapter, we present the hierarchical Radio Frequency Interference (RFI) mitigation system developed and implemented for offline processing of the recorded visibilities at MRT. It uses a conjunction of a variety of techniques involving linear and nonlinear methods in the visibility as well as in the image domain. It is important to investigate the nature of interference at an observatory site and develop appropriate techniques based on its statistics for its mitigation. The 20,000 hours of astronomical observations for the MRT survey have been used for such an analysis. The interesting aspects of interference statistics are highlighted in the second part of the chapter.

4.1 Introduction

The measurement of signals from celestial sources in the radio spectrum is affected by undesirable RFI due to a variety of sources. RFI has a complex nature : - rapidly changing or may be long lasting, very strong while it may also be weak. Even though several signal processing methods are used to handle RFI but in practice there is no universal fool-proof technique for RFI mitigation. In view of this RFI mitigation is achieved by using a combination of several engineering practices and techniques which include thresholding, temporal filtering, spatial filtering, canceling algorithms, notching and hole poking etc.. In addition, knowledge of RFI environment and the use of an RFI database at a particular site are valuable ingredients in combating the RFI signals (Sachdev & Udaya Shankar, 2001a; Fridman & Baan, 2001; Bell et. al., 2000; Baan et. al., 2001). RFI mitigation is generally carried out at three principal stages of astronomical data processing, namely : - real-time pre-detection and pre-correlation baseband processing, real-time post-correlation processing and offline processing (Fridman & Baan, 2001). In this chapter our discussion pertains to RFI mitigation during offline processing of the visibilities. We first describe data editing and interference in general and then discuss the RFI mitigation system for MRT data.

4.2 Data editing and interference

Data editing is an important aspect during the process of imaging. In this we identify and discard discrepant and severely corrupted data. Before a visibility file is imaged, it is important to ensure that measurements which are error prone or corrupted do not pass through to further stages of processing. Interference can occur due to a variety of reasons such as terrestrial sources like ignition systems, arcing sources, satellite systems, active Sun, inclement weather, malfunctioning receivers, incorrect observing parameters, data

recording errors, variations in troposphere and ionosphere etc.. The strength of interference can be at times up to 10^{11} times as strong as the signal from the celestial sky, like in case of iridium mobile communication systems etc.. Interference tends to generally affect short baselines more as compared to longer baselines as on longer baselines it gets bandwidth decorrelated. There exist several forms of editing automatic and using various handy software tools for discovering faulty data. While gross errors can be generally handled using automatic methods, data visualization is generally the main form of editing for subtle data.

Data editing is generally a four step process (Fomalont & Perley, 1999). The first step is to find out the antennas which are not working or dead. The second step is to look for inconsistent data and remove the outliers i.e. detection of interference in the data. This is usually achieved by a combination of both automatic and manual data visualization methods. The third step is to look for after applying the calibration, any longer term problems remaining such as unacceptable gains and variations, or bad solutions etc.. Here, more subtle data problems may generally surface. The fourth stage of data editing is carried out after the image is synthesized in terms of unexpectedly high noise, artifacts etc.. In such cases one needs to find out the likely causes and eliminate them if possible.

It is generally preferable to be consistent during the data editing process to be effective. Since the visibility function of a source or celestial sky is generally a slowly varying function of time, relatively simple tests of data consistency can be made. The source visibility is generally limited by noise and in such cases only interference points having very high strength i.e. bursts of interference can be detected. Further, the rms spread in the visibilities should be consistent with the expected noise. In case when the source is strong the instability in amplitude and phase can be ascertained. If the phase scatter is large due to troposphere and ionosphere, then it may be better to look at the visibility amplitude in each *cosine* and *sine* channel separately. In cases when self calibration can be carried out, the visibility phase variations can be corrected in principle. Further, if there are phase jumps during an observation between two calibrators, unless there is some way to ascertain the exact time of phase jump, it is better to flag the data for intervening observations. Usually one wants to retain the data which can be corrected and delete/flag/edit those which are incorrigible.

4.3 RFI mitigation system at MRT

RFI poses a serious problem at MRT due to its low frequency of operation and wide primary beam (EW×NS ≈2°×56°). RFI affects observations due to sources which are in-house

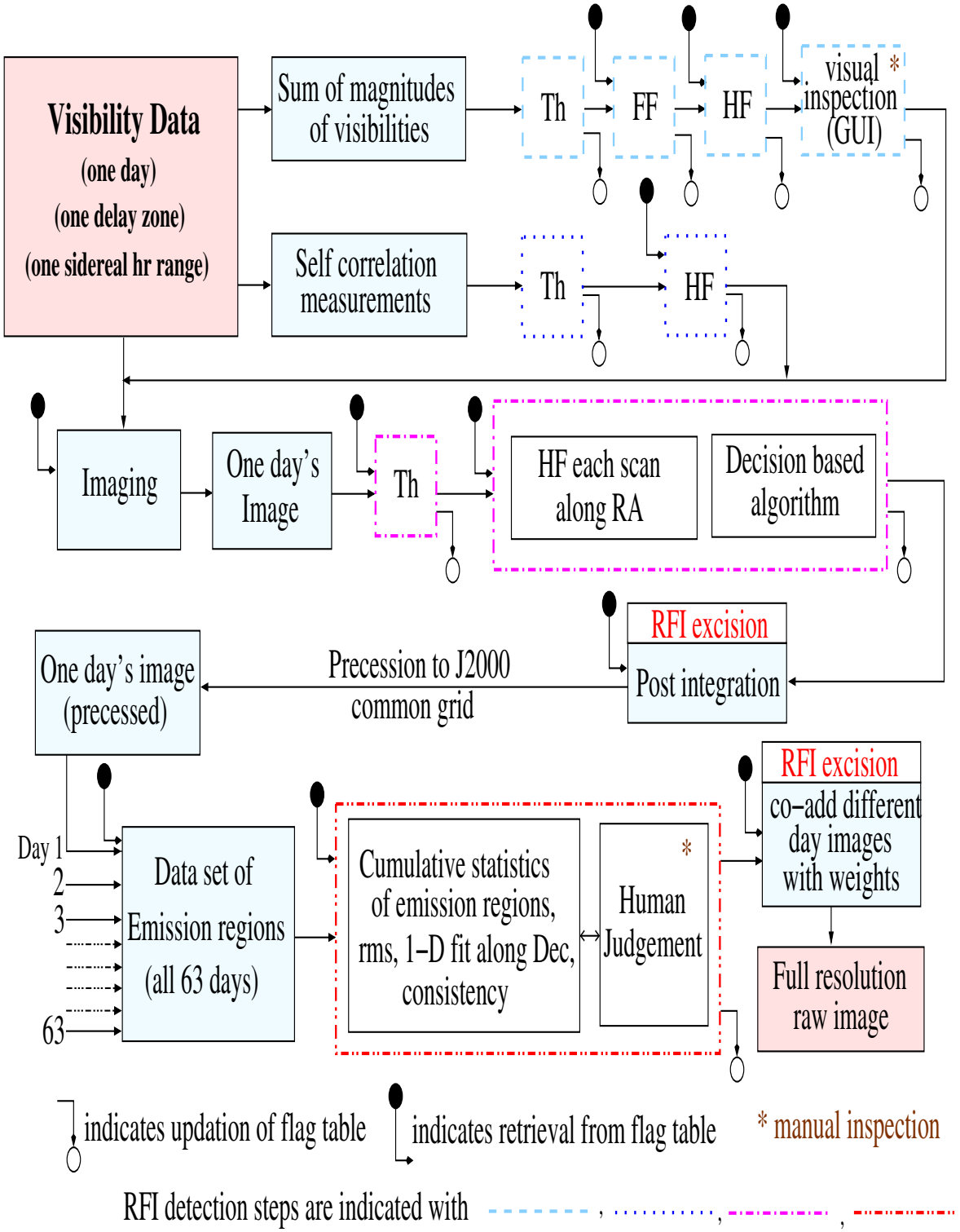


Fig. 4.1: Block diagram of hierarchical RFI mitigation system. The abbreviations used are: Th - thresholding (both upper and lower), FF - Fourier Filtering, HF - Hampel Filtering. The original data is not modified by any of the interference detection procedures. The interference information is stored in a centralized flag table which is referred and updated as and when required.

as well as external. The interference created "in-house" is generally due to harmonics from oscillators used in the instruments in the observatory and oscillations of malfunctioning amplifiers used in the system. In view of heavy interference due to in-house as well as external sources in the observing band of around 150 MHz, the observing frequency was shifted to 151.5 MHz which is relatively interference free. To reduce the effects of remaining in-house interference a variety of measures were employed, the details of which are given in Golap (1998); Sachdev (1999); Sachdev & Udaya Shankar (2001a). After sorting out most of the problems related to "in-house" interference, we still had interference in our data mostly from external sources.

At MRT since we use XF correlators, we cannot reject parts of the observing band in which the interference occurs and use the rest of the band. Interference from man made sources is generally polarized and is generally narrow band. Since we do not measure the polarization of the signal, we cannot use this property of interference in differentiating the interference from the sky signal. In addition, the large amount of data for the survey necessitated the need for an automatic RFI mitigation system. An hierarchical scheme has been developed with a view to achieve an effective, reliable and non-toxic automatic RFI mitigation with minimum possible human intervention. It uses a conjunction of techniques involving linear and nonlinear methods. These include Thresholding, Fourier filtering, Hampel filtering, Model fitting, Visual inspection and the fact that the sky signal is correlated in each day images but the interference is most likely not. The RFI mitigation is carried out in the visibility domain (sum of magnitudes of visibilities on all the baselines, self correlation measurements) as well as in the image domain (each day's image[1] before post-integration, by using all the one day images collectively for interference detection). Fig. 4.1 shows the block diagram of hierarchical RFI mitigation system at MRT. The original data is not modified by any of the interference detection procedures and the complete information about interference is kept in an updated flag table from which information is retrieved as and when required. We now briefly describe the characteristics of linear and nonlinear filters before discussing their application to MRT data.

4.3.1 Linear and Nonlinear filters

Linear system characteristics have profound advantages that can be exploited to develop useful classifications such as low pass filters, high pass filters, band pass filters and design procedures. As an example one fundamental property of the linear systems is that

[1]At MRT on any given day the visibilities are measured in one of the 63 different sets of configurations (allocations) of the trolleys in the NS arm. A day's image simply refers to an image made using the visibilities observed with one of these allocations on a given day. In this chapter an allocation's image and a day's image would be used interchangeably and literally mean the same unless specified otherwise.

they preserve the shape and frequency of the sinusoids and can alter only their magnitude and phase. This fact has been the basis for frequency response characterization which has been extensively used in the Fourier filtering technique for interference detection by separating signals according to their frequency content (See Sec. 4.3.3).

Data cleaning - automatically removing outliers or anomalous data points is a domain where nonlinear filters are very effective. Nonlinear filters are capable of implementing a vastly wider range of qualitative behavior, but they are harder to design and analyse as compared to linear filters. One of the reasons for this difficulty is that the range of possible nonlinear filter types is enormous and extremely heterogeneous. The difficulty is partly inherent in the term nonlinear which characterizes an enormous class of systems by their lack of unique property. So two classes of nonlinear filters may be as dissimilar to each other as they are to the class of linear filters.

The basic requirements of any data cleaning filter are twofold :- it must first reduce or eliminate the effect of outliers and secondly minimize the distortion of the nominal part of the input data sequence. The first requirement is met by filters, whose measure of data sequence center is less sensitive to outliers and thus robust. For the second requirement among the nonlinear digital filters a very useful class is the one which exhibits the property of scale invariance. For example if $F(x) = y$ then $F(px+q) = py+q$ for any positive constant p and real number q, where F is the filter function. The most important property of scale invariant filters is that their behavior is dependent only upon the shape of the signal and not on the units in which the inputs are expressed.

Hampel filter is less sensitive to outliers as its measure of data sequence center is based on median and thus robust. In addition this nonlinear filter is scale invariant due to which it achieves both the requirements of a data cleaning filter nicely. In contrast more general nonlinear filters based on Artificial Neural Networks (ANN), polynomials or piecewise linear functions can respond differently to the same input signal when expressed in different units. At MRT in the first stage, a technique based on Fourier filtering to detect interference in the visibility data was developed by Sachdev & Udaya Shankar (2001a). With the above mentioned attributes of Hampel filter, it was a natural choice for RFI mitigation at other steps of data processing (self correlation measurements, each day's image before post-integration) in addition to detect occasionally residual interference in the visibilities after Fourier filtering. Now we first briefly describe the Hampel filter and subsequently discuss the interference detection at several steps during the MRT data processing.

4.3.2 Hampel filter

One of the popular schemes to detect outliers is to use their distance from the mean which is the most common measure of the data sequence center. A fundamental difficulty with mean is its sensitivity to outliers due to which it fails, as the outliers have generally high values and a few of them are enough to change the value of mean itself by a significant amount. The median is an alternative measure of the data sequence center with much less sensitivity to outliers and provides the basis for the nonlinear scale invariant Hampel filter. An important property both mean and median have is of scale invariance. The mean is scale invariant because it is a linear combination of data values and the coefficients in the linear combinations add on to unity. In contrast, median is scale invariant because the sorting operation to estimate the median is itself scale invariant. Let us consider an input sequence $x(1), x(2), ..., x(N)$, of N samples where the index in the bracket indicates the sequence number of the sample. If we define a symmetric moving window filter function $F\{\}$ as,

$$y(k) = F\{x(k-a), ..., x(k), ..., x(k+a)\} = median\{x(k-a), ..., x(k), ..., x(k+a)\} \tag{4.1}$$

we obtain the median filter. Here $x(k)$ and $y(k)$ represent the current filters input and output signals, k and a are positive integers and F is the filter function which maps the $2a$+1 input samples sequences from $x(k-a)$ to $x(k+a)$ to the real output $y(k)$.

An important feature of the median filter and other nonlinear filters is the existence of root signals which are sequences that the filter does not modify. Both increasing and decreasing sequences represent the root signal of the median filter. This holds good for all choices of window width parameter $2a$+1. Other root signals also exist for the median filter but they are generally dependent on the window width parameter on which the filter acts. In general, a desirable nonlinear filter for applications such as data cleaning is the one whose root signals include all the nominal data sequences of interest. The median filter often tends to be too aggressive as a data cleaning filter which may also introduce undesirable distortions in the nominal part of the input signal. This observation means that the class of root signals for the median filter is often too small. One way to enlarge this class is by replacing the median filter by Hampel filter which has a rich class of root signals (Pearson et. al., 2002).

The Hampel filter belongs to the class of decision based filters. The filter uses $2a$+1 data values in the moving data window to decide the representativeness of the current data sample $x(k)$. If the filter deems $x(k)$ as representative it makes no change and the filters output is simply $y(k) = x(k)$. Otherwise the filter decides that the current data sample is

unrepresentative and it computes a more representative replacement from the available data, which is actually the median. The most important task (and where it differs from the median filter) is to establish the criterion for deciding whether $x(k)$ is representative or not.

Outlier characterization:

To decide whether a data value is representative or an outlier, any filter must have some quantitative characterization of nominal behavior. One of the most popular working assumptions concerning nominal behavior is the Gaussian, or normal random variable model. There are two important reasons for this. First that this assumption is often reasonable as an initial approximation in describing nominal behavior. Second is that if we adopt this for some data sequence $x(k)$, the mean and the standard deviation provide a complete characterization of the statistical behavior of $x(k)$. Since the median is less sensitive to outliers, the Hampel filter replaces the mean with the median. It replaces the standard deviation with another measure of the spread of the data sequence called Median Absolute Deviation (MAD) scale whose value is equal on an average to the standard deviation in case of a Gaussian data sequence. It was proposed by F. R. Hampel as an alternative to standard deviation. Specifically let x_m denote the median of the current moving window of data sequence $x(k-a), ..., x(k), ..., x(k+a)$, the MAD scale is given by the expression,

$$\begin{aligned} \text{MAD} &= 1.4826\,\text{median}\{d(k)\} \\ &= 1.4826\,\text{median}\{\,|x(k-a)-x_m|, ..., |x(k)-x_m|, ..., |x(k+a)-x_m|\,\} \end{aligned} \tag{4.2}$$

Here $d(k)$ gives a measure of how much far is each data value from this reference value (x_m) and median$\{d(k)\}$ gives measure of how far the input sequence in the moving data window typically lies from the median value x_m. So for each point $x(k)$ if $|x(k)-x_m| \leq t\times$MAD, the filter classifies the point $x(k)$ as nominal otherwise it is classified as an outlier. Here the parameter t is called the aggressiveness parameter of the filter. Selecting a higher value of t actually makes the filter behave very conservative, while lower values tends to make it more aggressive.

Implementation: The Hampel filter uses a moving window implementation of this outlier detection procedure. For each sample $x(k)$ the filter forms a collection of samples from $x(k-a)$ to $x(k+a)$ from which it computes the median and MAD scale estimate and performs the outlier test. For the initial data sequences when $k < a+1$ and for end of the data sequences when $k > N-a$, additional data sequences are padded both before the start and after the end of the data sequence by mirror imaging the part present on the other available side of the first and the last data value in the sequence. Thus, the entire sequence is arranged as $\underline{x(a+1), x(a), ..., x(2)}, x(1), x(2), ..., x(N-1), x(N), \underline{x(N-1), ..., x(N-a)}$ where the

underlined samples represent the padded data sequences before the start and after the end of the data sequence.

Hampel filter has two independent tunable parameters namely the size (width) of the moving data window ($w = 2a + 1$) and the aggressiveness parameter (t), which characterize the filter. They are estimated after experimentation for best performance. Generally a value of t in the range of 3 to 5 is reasonable. Now we discuss interference detection at various steps of data processing starting from the sum of magnitudes of visibilities on all the baselines.

4.3.3 Sum of magnitudes of visibilities

Fourier filtering :

A technique to detect interference in the visibility data was developed by Sachdev & Udaya Shankar (2001a). This technique is based on the understanding that the interference is generally 'spiky' in nature and has Fourier components beyond the maximum frequency which can arise from the radio sky and therefore can be identified. In interferometer arrays, the visibility data is generally sampled in time at rates much faster than the rates required for a given baseline. This oversampling can be made use of in identifying this interference. At MRT, the baseline with the longest EW component ($\sim 512\lambda$) has a maximum fringe rate of about 0.04 Hz. Therefore one needs to sample this data at 0.08 Hz to satisfy the Nyquist sampling criterion. The data has been sampled at about every one second which allows us to identify frequencies up to 0.5 Hz. Therefore interference can be identified by its contribution to frequencies in the range 0.04 Hz to 0.5 Hz.

The maximum frequency[2] to which the sky contributes at the output of a complex correlator is given by $\omega_e x$, where $\omega_e = dH/dt = \frac{2\pi}{(24*60*60)} = 7.27 \times 10^{-5}$ radians per sidereal second and x is the east-west component of the baseline in wavelengths. Since x depends on the baseline used, the ability to distinguish interference from the sky signal decreases with increasing east-west component of the baseline. To overcome this baseline dependence, instead of detecting interference in the *cos* and the *sin* channels separately, interference is detected in the magnitude domain, i.e., in $\sqrt{\cos^2 + \sin^2}$. The magnitude of visibility has a rate of change which is independent of the baseline used. This rate depends on the primary beam of the interferometer which is the product of the voltage patterns of the group of antennas (EW and NS groups) forming the interferometer. The expected maximum fre-

[2]The fringe frequency $f_{osc} = \frac{1}{2\pi}\frac{d\phi}{dt} = \frac{1}{2\pi}\frac{d\phi}{dH} \times \frac{dH}{dt}$ where $\phi = 2\pi(xl + ym + zn)$. Since the fringing is predominantly due to the EW component of the baseline, the fringe frequency may be approximated as $f_{osc} = \frac{1}{2\pi}\frac{d}{dt}(2\pi x.l) = x \cos H \cos\delta \times \frac{dH}{dt}$. In terms of u where $u = x \cos H$, the fringe frequency is $\omega_e u \cos\delta$ where $\omega_e = dH/dt = \frac{2\pi}{(24*60*60)} = 7.27 \times 10^{-5}$. The fringe rate is fastest for a source at $\delta = 0$ and is $\omega_e u$. For small H, i.e., near the meridian $\cos H \approx 1$ and therefore $\omega_e u = \omega_e x$.

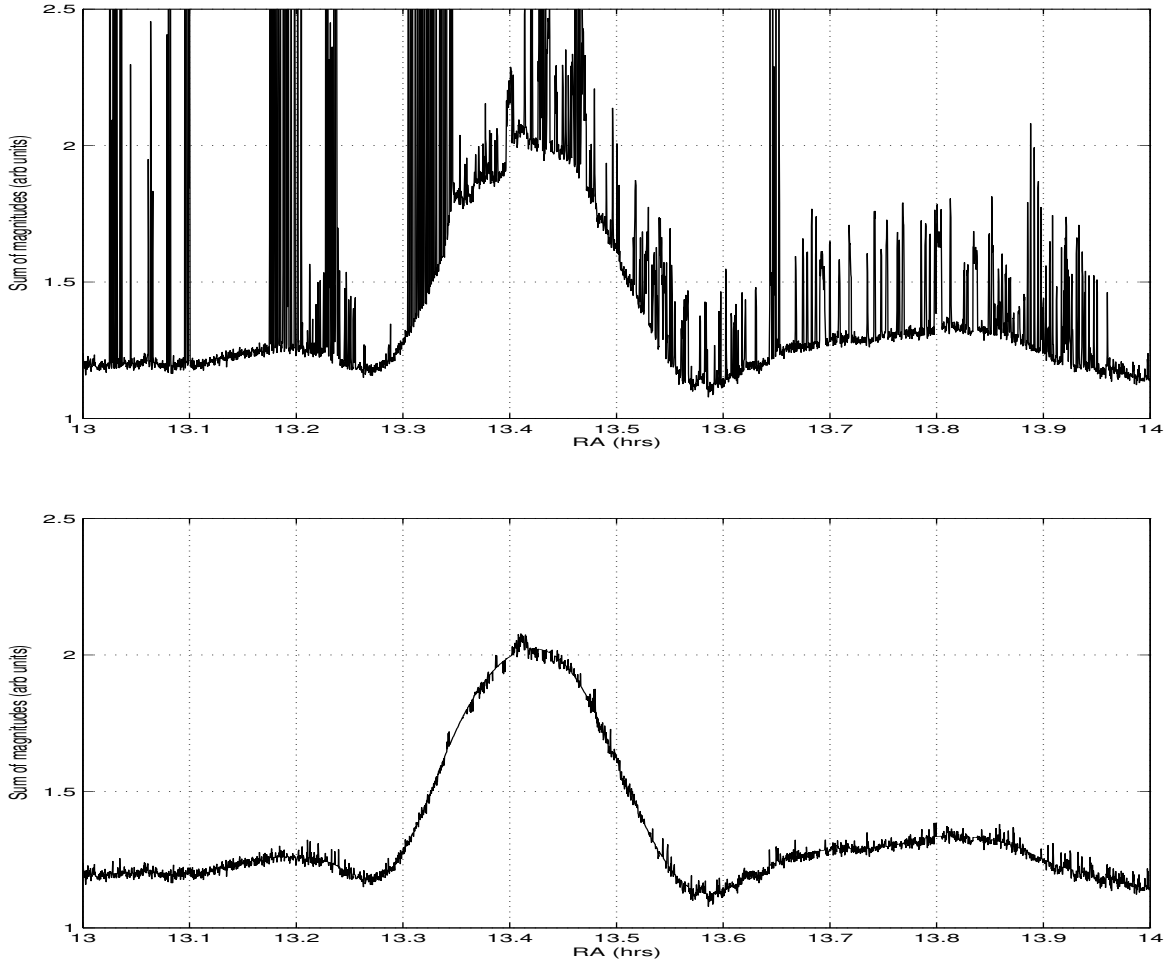


Fig. 4.2: Interference detection using only Fourier filtering in the sum of magnitudes of visibilities on all the baselines. The upper plot is for the original visibility file while the lower plot is obtained after the interference points detected by the Fourier filtering technique have been removed. There are 355 interference points detected in this visibility file having a total of ≈3,300 points.

quency at MRT in the magnitude of the visibility is ≈0.003 Hz.

In practice first an upper and lower thresholding is introduced based a priori information of the highest and lowest level of signals from the region of the sky under process. This is followed by Fourier filtering and performing a CLEAN like operation to detect interference peaks in the sum of magnitudes of visibilities on all the baselines[3]. Fig. 4.2 shows the sum of magnitudes of visibilities on all the baselines before and after successful interference detection carried out using Fourier filtering technique. More details are available in Sachdev & Udaya Shankar (2001a).

[3]We recall that at MRT we measure 512 complex visibilities (32 EW×(15 NS+E16)) each second out of which we use 480 complex visibilities (32 EW×15 NS) for imaging (See Sec. 5.3). In this chapter, all the baseline refer to 480 baselines used for imaging unless specified otherwise.

Hampel filtering :

The analysis of the effectiveness of the Fourier filtering technique for interference detection in the visibilities revealed that, it in fact does successfully detects most of the interference. The analysis also revealed that interference having a complex nature does escape through it occasionally and in a few cases genuine visibility points are also detected as interference. Some of the possible reasons for such occasional failures could be the following. The Fourier filtering technique does not detect slowly varying interference. There are erroneous detections when there is an offset between the beginning and the end of the data being Fourier transformed as DFT causes ripples which is detected as interference at the ends. A strong source which is partially appearing at the beginning or the end of the visibility file under process has a rate of change which is higher than that expected from the sky, which also may cause erroneous detection. When there is a continuous patch of interference the calculation of rms noise itself is affected and hence the reliability of interference detection in such cases is low. All these issues were understood earlier but were not incorporated in the software available to implement the Fourier filtering technique.

Initial attempts to detect RFI using ANN, adaptive filters namely Weiner filter and Widrow-Hoff algorithm, median and Hampel filters indicated that none of these would improve the detection if they simply replace the existing Fourier filtering technique (except in cases where there is very nominal interference like shown in Fig. 4.3 where Hampel filter alone works as good as the Fourier filtering technique). So we decided to detect the residual interference in the data from which the interference detected by the Fourier filtering technique has been removed. Hampel filter was used for this purpose as it was found to be more suitable compared to other mentioned methods.

Hampel filtering is applied on the relatively interference free data after Fourier filtering (which also helps in ensuring non-toxic behavior) to detect any remaining RFI. In order to ensure a robust estimation of median and MAD, their estimation is carried out in two steps. First all likely outliers detected by Hampel filtering are noted down. In the next step estimation of median and MAD scale is carried out by excluding these likely outliers detected in the first step. The entire data (including all likely outliers) is then passed through the Hampel filter which uses the robust median and MAD scale to decide whether a point is an outlier or not. *In the entire RFI mitigation system the median and MAD scale are always estimated in this manner for Hampel filtering unless specified otherwise.*

For non-toxic behavior the number of data samples affected by interference should be less than half the total number of samples in the moving data window. The width of the moving data window is decided depending on HPBW of the functional form of the data and the frequency and duration of interference present. A smaller width can give more

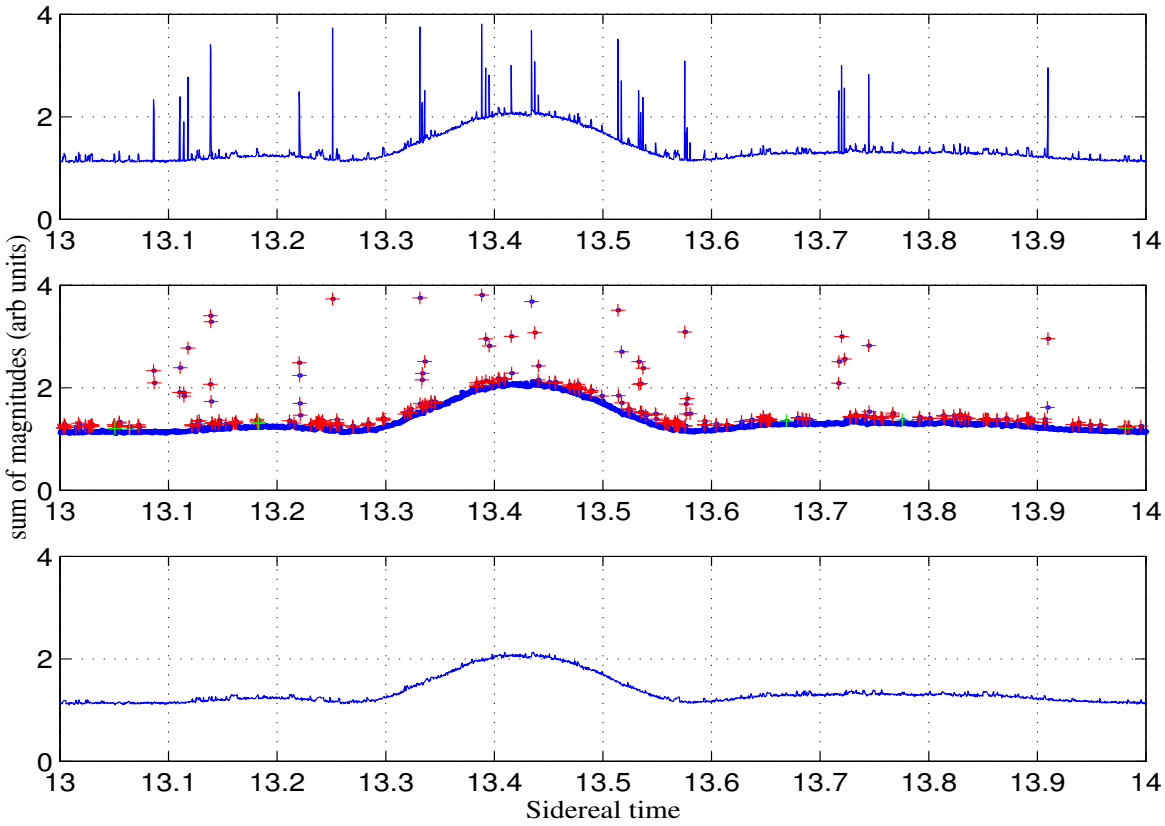


Fig. 4.3: Interference detection using only Hampel filtering in the sum of magnitudes of visibilities on all the baselines. The top plot shows the original visibility file. In the middle plot, points detected as interference are marked as + (red) while others are marked as dots (blue). The bottom plot is obtained after interference points detected by Hampel filtering have been removed.

efficient performance in case there are very few isolated interference data samples but can be toxic if interference is continuous and very frequent while a larger width may decrease the efficiency. After experimentation we found a non-causal moving data window of width 121 s (one minute of data on both sides, HPBW is ≈480 s) to be suitable. The aggressiveness parameter used is 5 (equivalent to 5σ). Fig. 4.3 shows the sum of magnitudes of visibilities on all the baselines before and after successful interference detection using only Hampel filtering.

In actual practice Hampel filter is applied in conjunction with Fourier filtering. Fig. 4.4 shows the application of Fourier filtering and Hampel filtering in conjunction on the sum of magnitudes of visibilities on all the baselines. It is clear from the comparison that the Fourier filtering technique detects most of the interference and a few remaining interference points of slightly lower strength are successfully detected by the Hampel filtering. During the practical implementation and examining the files for interference using visual inspection (See GUI based visual inspection below) we found that the Fourier filter-

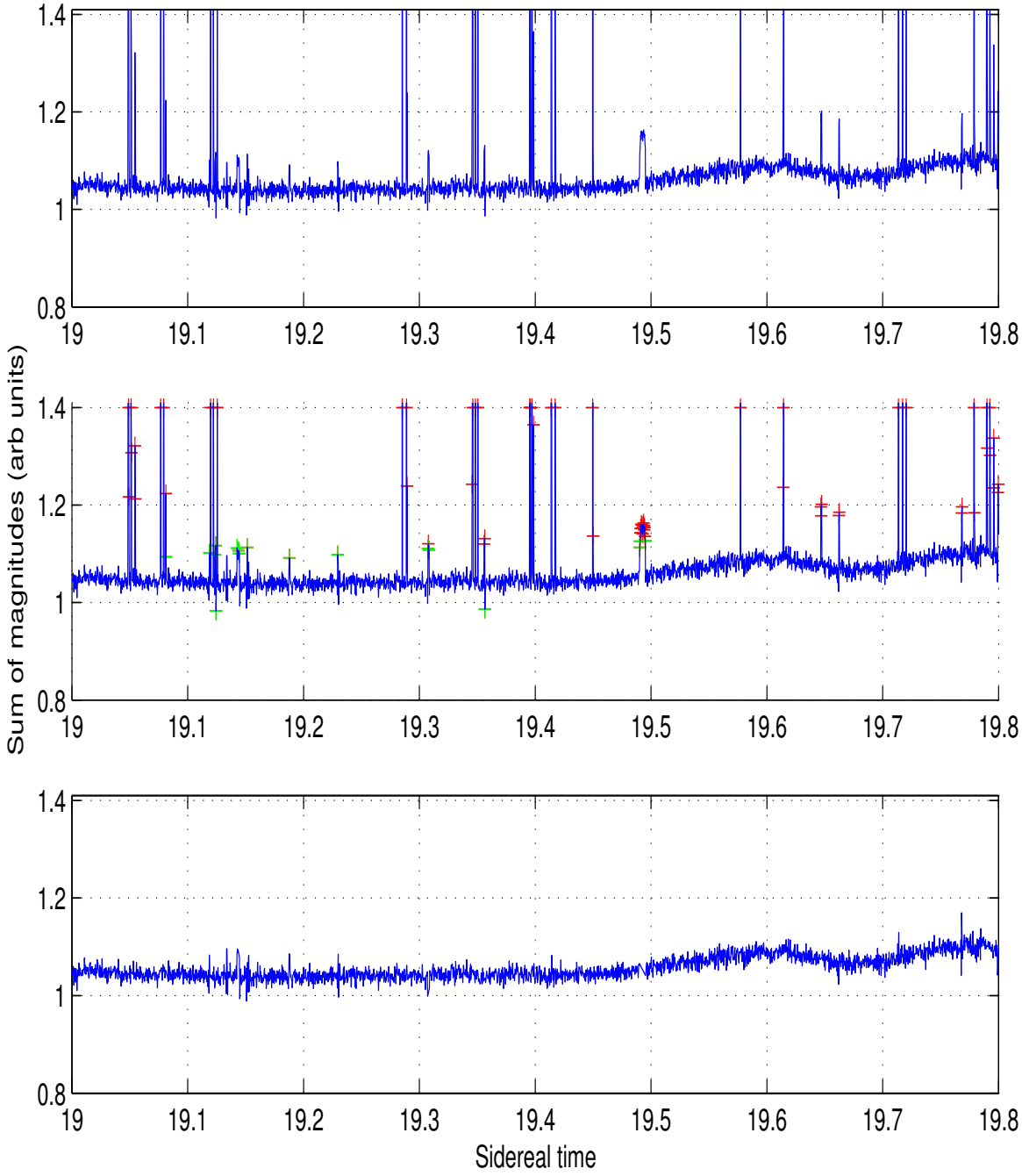


Fig. 4.4: Detection of interference using a conjunction of Fourier filtering and Hampel filtering in the sum of magnitudes of visibilities on all the baselines. The top plot shows the original visibility file. In the middle plot the interference points detected by Fourier filtering are marked as + (red), the residual interference points detected (in the relatively interference free data after Fourier filtering) by Hampel filter are marked as + (green). The bottom plot is obtained after all the interference points detected by Fourier filtering and Hampel filtering have been removed.

ing scheme for interference detection erroneously detects genuine visibility data points as interference in some cases when there is an abrupt level shift in the sum of magnitudes of visibilities and when a strong source partially appears at the beginning or the end of the visibility file under process. For such cases we decided to skip Fourier filtering technique and directly subject the visibilities to Hampel filtering followed by manual inspection. Fig. 4.5 shows one such example of toxic behavior of Fourier filtering technique which detects genuine visibility points at the extreme end as interference (near the transit of a very strong source Cygnus A near Sidereal time 20 hrs). When only Hampel filtering is applied it successfully detects only the interference points. The number of such files is very low ($\approx 0.5\%$). This conjunction of Fourier filtering and Hampel filtering in a symbiotic mode has proved to be very effective and has significantly reduced the number of escaping interference points to a minimum.

GUI based visual inspection :

To evaluate the performance of the automatic RFI excision schemes on the sum of magnitudes of visibilities on all the baselines and to handle remaining discrepant subtle data which can be best handled by human inspection, an interactive Graphical User Interface (GUI) based program was developed. It is used to flag any escaped interference or unflag a wrongly flagged genuine visibility. The program takes a list of data files as input so that batch editing can be carried out and facilitates fast scanning of the visibility data as a graph plot where flagging and unflagging can be carried out quickly in a user friendly way. Any earlier existing interference information is also preserved as a separate flag table so that detailed analysis of the nature of the interference can be carried out later if required. A snapshot of the GUI based interference detection on the sum of magnitudes of visibilities in all the baselines is shown in Fig. 4.6.

The GUI based interference detection is carried out only for the good data selected for imaging. Due to this the number of anomalous points likely to be seen are very few and the inspection can be achieved quickly. The GUI based visual inspection helped during the initial stages of Hampel filter characterization, by enabling quick examination of the results for different parameters for a large number of files while experimentation. We now describe the next step of interference detection in self correlation measurements.

4.3.4 Self correlation measurements

At MRT, to estimate the normalized correlation coefficient from the digital correlation counts, we use self correlation measurements[4] of the interferometer elements involved in

[4]It is to be noted that for self correlation measurements without AGC (unlike the ones discussed here with the AGC) used to estimate the unnormalized correlation coefficient also, Hampel filtering is applied to detect interference. This is discussed later in Sec. 7.2.1.3.

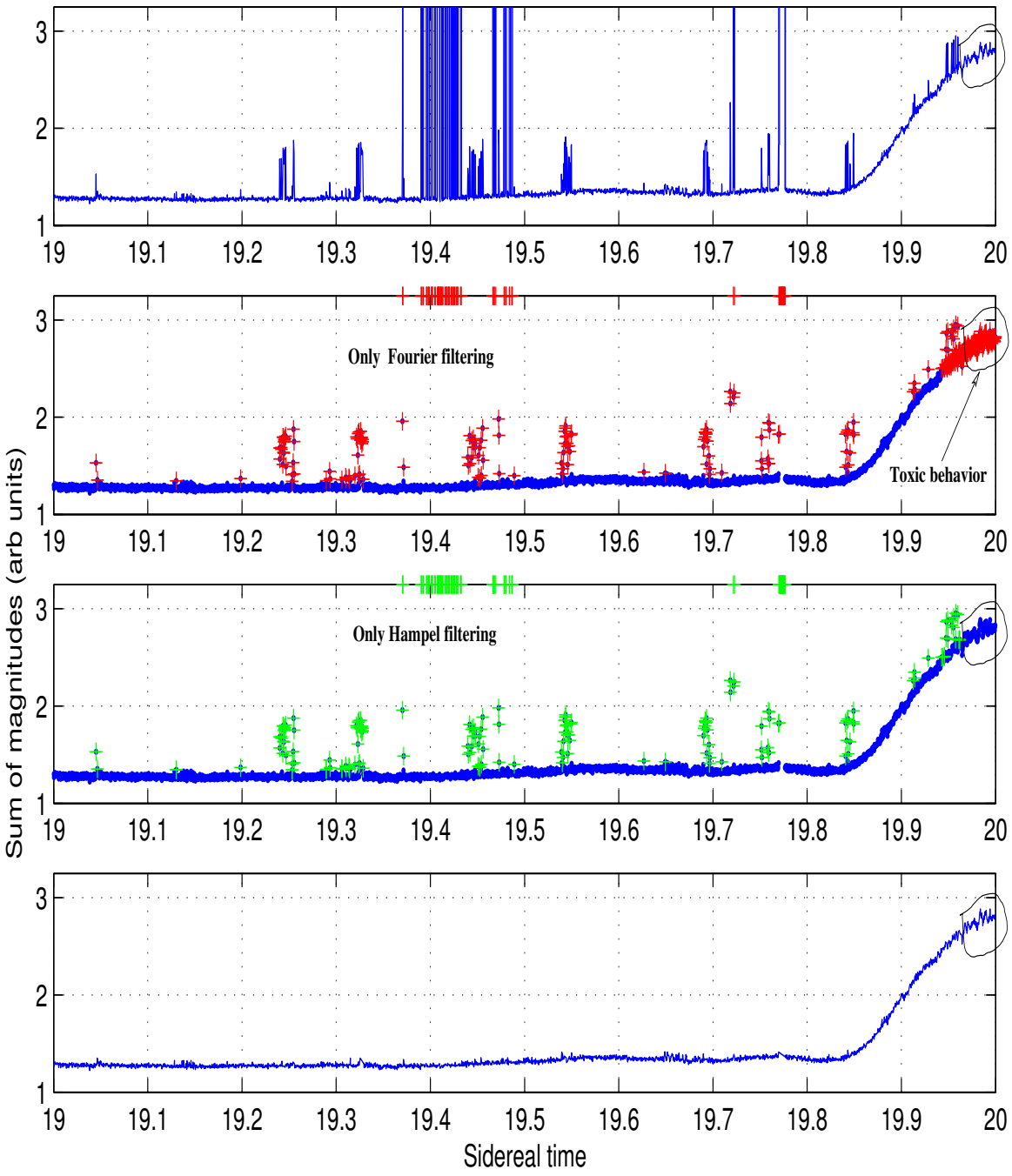


Fig. 4.5: Comparison of detection of interference using Fourier filtering and Hampel filtering separately in the sum of magnitudes of visibilities on all the baselines. The top plot shows the original visibility file. The second plot (from the top) shows the interference points detected by Fourier filtering marked as + (red). The third plot (from the top) shows the interference points detected by Hampel filtering marked as + (green). The lowest plot is obtained after all the interference points detected by Hampel filtering have been removed. The comparison clearly reveals the toxic behavior of Fourier filtering technique. As shown the Fourier filtering erroneously detects genuine visibility points at the extreme end of the visibility file (close to transit of strong source Cygnus A near sidereal time ≈20 hrs), while the Hampel filter successfully detects only the interference points. The number of such files in which Fourier filtering behaves toxic is very low (≈ 0.5%).

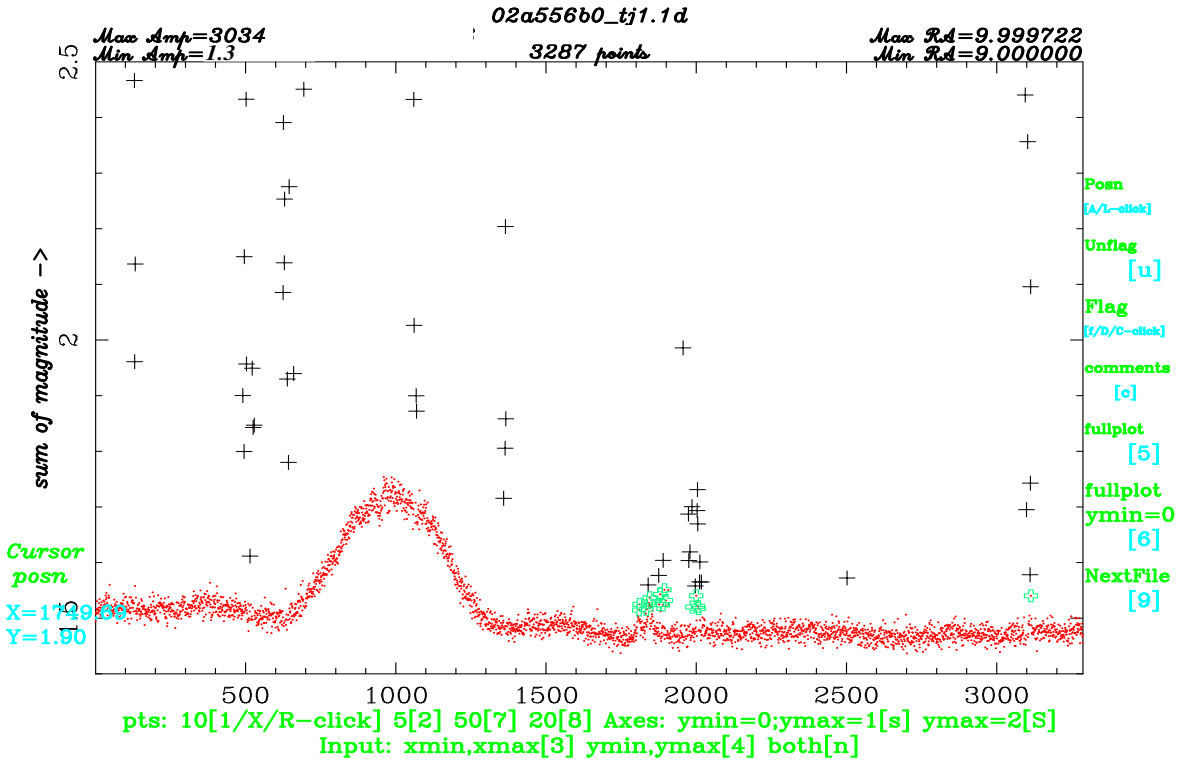


Fig. 4.6: A snap shot of the interference detection in the domain of sum of magnitudes of visibilities using interactive GUI. Most of the options can be carried out using the buttons on the display or equivalent keys. The points marked by + are detected by Fourier filtering followed by Hampel filtering while the points detected on manual inspection are marked as green (These points detected manually are shown only as a example and in practice generally all the points perceivable as interference are successfully detected automatically).

the baseline (Addario et. al., 1984). Sixty four self-correlators (32EW, 16NS *cosine*, 16 *sine*) are measured wherein each channel is correlated with itself with zero delay. Since self correlation measurements are very susceptible to interference it is desirable to carry out its detection and excision. For each of the sixty four self correlation values measured every integration period (≈1.1 s), a first level of thresholding for both upper and lower cut-offs based on a priori information is introduced. Hampel filtering is subsequently applied with a moving data window of width 301 s and an aggressiveness parameter of 5 for both EW and the NS groups (HPBW$_{EW}$≈2°=480 s; HPBW$_{NS}$≈15°=3,600 s).

Due to broad primary beam of both the EW and the NS groups in RA, the rate of change of self correlation measurements is a slow function of time. Taking advantage of this, their integration time was increased from ≈1.1 s to ≈67 s to increase the signal to noise ratio. The time for averaging the self correlation values was decided after carefully examining the actual data. Fig. 4.7 shows the self correlation measurements as recorded in a visibility file, after interference removal and after increasing the integration time. Only those data files which are good and thus selected for imaging are passed through for interfer-

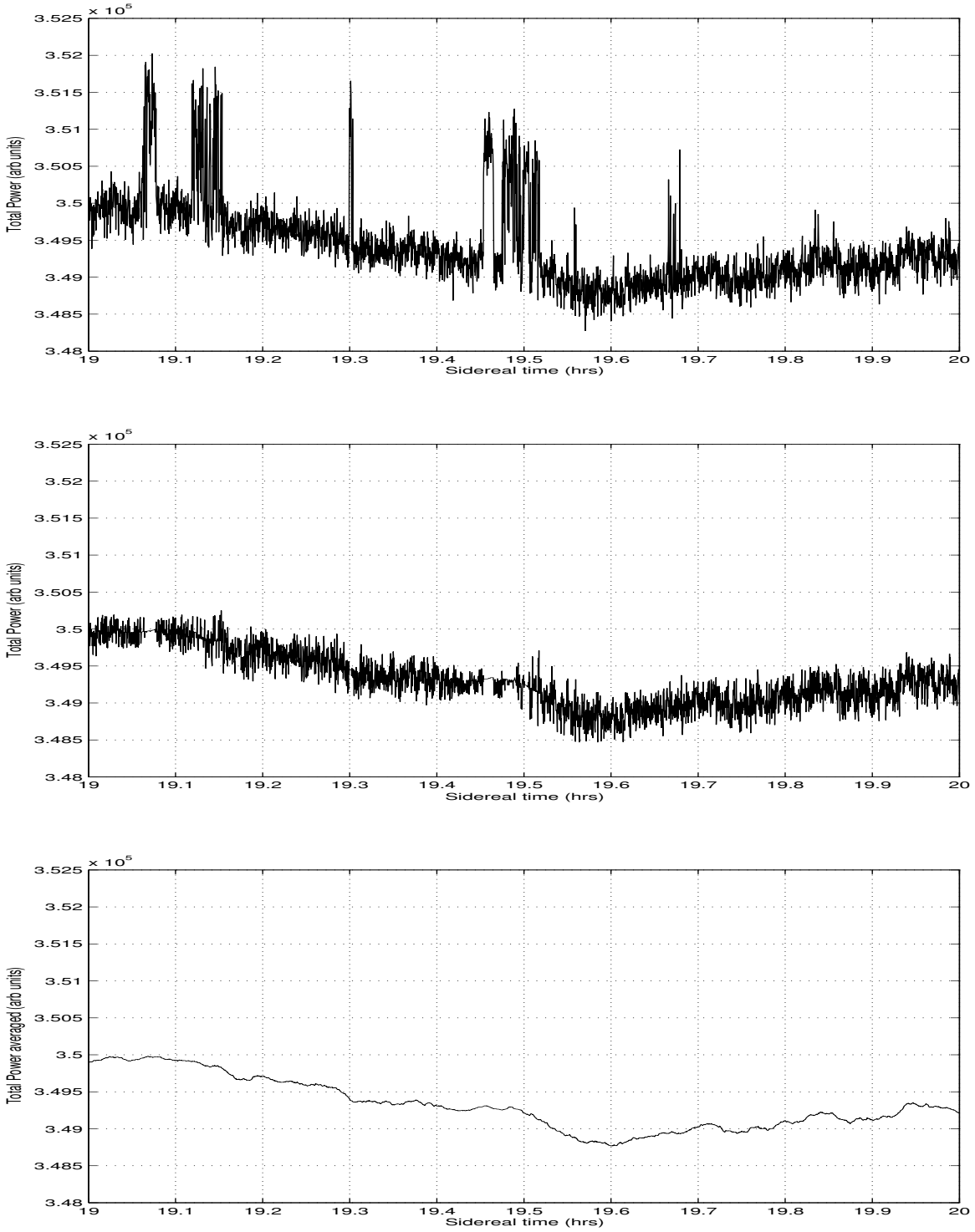


Fig. 4.7: Detection of interference using the Hampel filter and subsequent averaging in self correlation measurement of one EW group. The top plot shows the self correlation measurement as recorded in the visibility file, while in the middle plot all interference points detected by Hampel filter are removed. The plot at the bottom shows the self correlation values after averaging using a non-causal moving window width of 61 measurements (equivalent to 67 s) to improve the signal to noise ratio.

ence detection in self correlation measurements (which ensures non-toxic performance). These interference excised and smoothed self correlation values (with increased signal to noise ratio) are stored in binary format in a separate file along with the corrected sidereal times (see Sec. 5.1) for the visibilities. In all stages of data processing these improved values of self correlation measurements are used.

4.3.5 Each day's image before post-integration

In synthesis imaging a sharp feature in the visibility domain transforms to a sinusoid in the image domain and vice versa. For best interference detection it is desirable to exploit both domains. At MRT, we carry out meridian transit imaging and the 2-D image is formed by stacking the 1-D images on the meridian along the declination at different sidereal times (See Sec. 5.3). The analysis of one day images led to the conclusion that there were interference points which were not eliminated in the visibility domain by the detection schemes used. In view of this we decided to detect interference at the level of each day's image[5] as the next step in our hierarchical scheme. It is important to detect any residual interference at this stage because during the next step in data processing this interference also spreads to adjacent sidereal times during post-integration (Sec. 4.3.6). It is to be noted that since we do a meridian transit imaging, rejecting the entire 1-D scan along the meridian at the sidereal time affected by interference in the image domain is equivalent to flagging the corresponding observed visibilities at that sidereal time.

First in each day's image the RA scans at each declination are Hampel filtered with two aggressiveness parameters (I_{low}=5, I_{high}=10) with a moving data window ≈21 s wide (RA beam HPBW≈16 s). The window width and the aggressiveness parameters are chosen keeping in mind that the filter does not distort a *sinc* function of HPBW 16 s (expected beam in RA in each day's image at δ=0°) which ensures non-toxic performance. The number of times interference has been detected with the two aggressiveness parameters [N(>I_{low}), N(>I_{high})] and the maximum strength of interference (I_{max}) are noted down for each pixel on the RA axis. Once the scanning at all the declination points is completed, all the information related to interference detected at each RA point is stored. Subsequently, at each RA point the intensity along declination (corresponding to the delay zone) is integrated to obtain a one dimensional integrated intensity profile as a function of RA. Hampel filtering is also applied on this integrated intensity profile with a window width of 16 s and aggressiveness parameter of 5. For each RA point information is stored if it has been detected as interference or not (D_{ind}=1 in case of interference else D_{ind}=0) and the

[5]We recall that a day's image simply refers to an image made using the visibilities observed with one allocation on a given day.

strength of detected interference (I_{int} has the actual strength in case of interference else I_{int}=0).

The decision to classify if there is interference at some RA or not, is carried out via a multi-parameter decision based algorithm based on the cumulative statistics including the interference information of the neighbouring pixels. The decision based algorithm applied has a conservative approach. It takes into account N(>I_{low}), N(>I_{high}), (I_{max}), (D_{ind}, I_{int}) and interference details about the adjacent sidereal times. As an example, if at a given RA pixel, the Hampel filter detects an interference of strength more than 10σ even once in the RA scans the pixel will be classified as affected by interference, but if it detects an interference of strength 5.5σ only once in the RA scans, it will be classified as interference only if the neighbouring RA pixels on both the sides are affected by interference. The theme behind the algorithm is to classify a point as affected by interference only if there is more than one consistent indication to confirm the interference. The statistics of interference indicate that ≈88% of the interference bursts last for less than 4 integration periods (each integration period ≈1.1 s). Due to this whenever continuous interference is encountered for four or more integration periods, the algorithm does not classify those points as interference. Such suspected interference cases are very few and are re-investigated after post-integration as described in the next Section (See Sec. 4.3.6).

4.3.6 Data set of all one day images

Each day's images is post-integrated for 4 s in time and precessed to a common grid in the epoch J2000. Since generally the interference is short lived (Sec. 4.4), the RFI mitigation (excision) is carried out while post integrating each one day's image by giving zero weights to the points corrupted by interference (all interference detected up to the level of each day's image before post-integration). The analysis of these post-integrated images (in J2000 epoch on a common grid), revealed that occasionally there is some residual interference still present in the images. This is most likely due to the conservative implementation of the interference detection at the level of each day's image before post-integration which does not classify any interference lasting for four or more integration periods (See Sec. 4.3.5). We have developed a technique by which these interference points can now be detected in each day's image and the corresponding visibilities can be flagged. The scheme works on the assumption that the sky signal is correlated in each day's image at the same sidereal time while the interference is most likely not.

A flow chart of steps to collectively detect any remaining interference in these post-integrated one day images is shown in Fig. 4.8. During the practical implementation the detection of interference in each day's image is carried out in batch mode. In one batch all

one day images (allocation's) for a given sidereal hour for one delay zone are considered. The four batches of images for a sidereal hour are processed separately.

Emission regions : A program goes through each allocation's image. For each image first the maximum value of the intensity is found and its location is noted down. An emission region covering an angular extent corresponding to the first sidelobe level of the full resolution beam ($16\times \sec(\delta)$ s in RA and $4\farcm6 \sec(\delta+20\fdg14)$ in declination) is associated around the located maximum in RA and declination. In a similar way all emission regions are found in the image whose peak is above a threshold (generally 5σ where σ is the rms noise measured in the image) excluding the already detected emission regions. All the images in a batch are processed in a similar manner to produce a database of emission regions.

A plot of location of peaks of all the emission regions in the entire database as a function of RA and declination is displayed as shown in Fig. 4.9(a). A strong source appears as cluster of points as it is likely to be detected in more or less each day's image (since the sensitivity of the array is same for all the days). A strong interference appears as a vertical stripe along declination at the sidereal time affected by interference (as we carry out meridian transit imaging). This plot gives a first hand rough idea of the amount of interference still remaining in the images.

Identification of interference affected images : In the next step we identify those images which are likely to be effected by interference. This is accomplished by comparing all images with each other on the basis of rms (σ) noise, number of emission regions and distribution of emission regions along declination in each day's image. The images effected by interference generally show some unusual behavior, in at least one or more of the above parameter space.

A plot of distribution of positions of emission regions peaks along declination in each image as a function of image number is displayed as shown in Fig. 4.10(a). An image affected by strong interference is likely to reveal stripes along declination differently compared to images of nearby allocations. After this a plot of total number of emission regions detected in each day's image as a function of image number is displayed as shown in Fig. 4.11(a). An image affected by strong interference generally shows unusually higher number of emission regions compared to images of nearby allocations. Based on these lines all the images suspected to be affected by interference are noted down. In the next stage only these suspected subset of images are individually investigated.

Interference detection in individual images : Each image suspected to have interference is individually inspected by plotting the location of emission region peaks as a function of RA and declination. Fig. 4.12(a) shows an example of one such image affected by interference. Another plot showing the positions of strong sources (estimated using the MRC

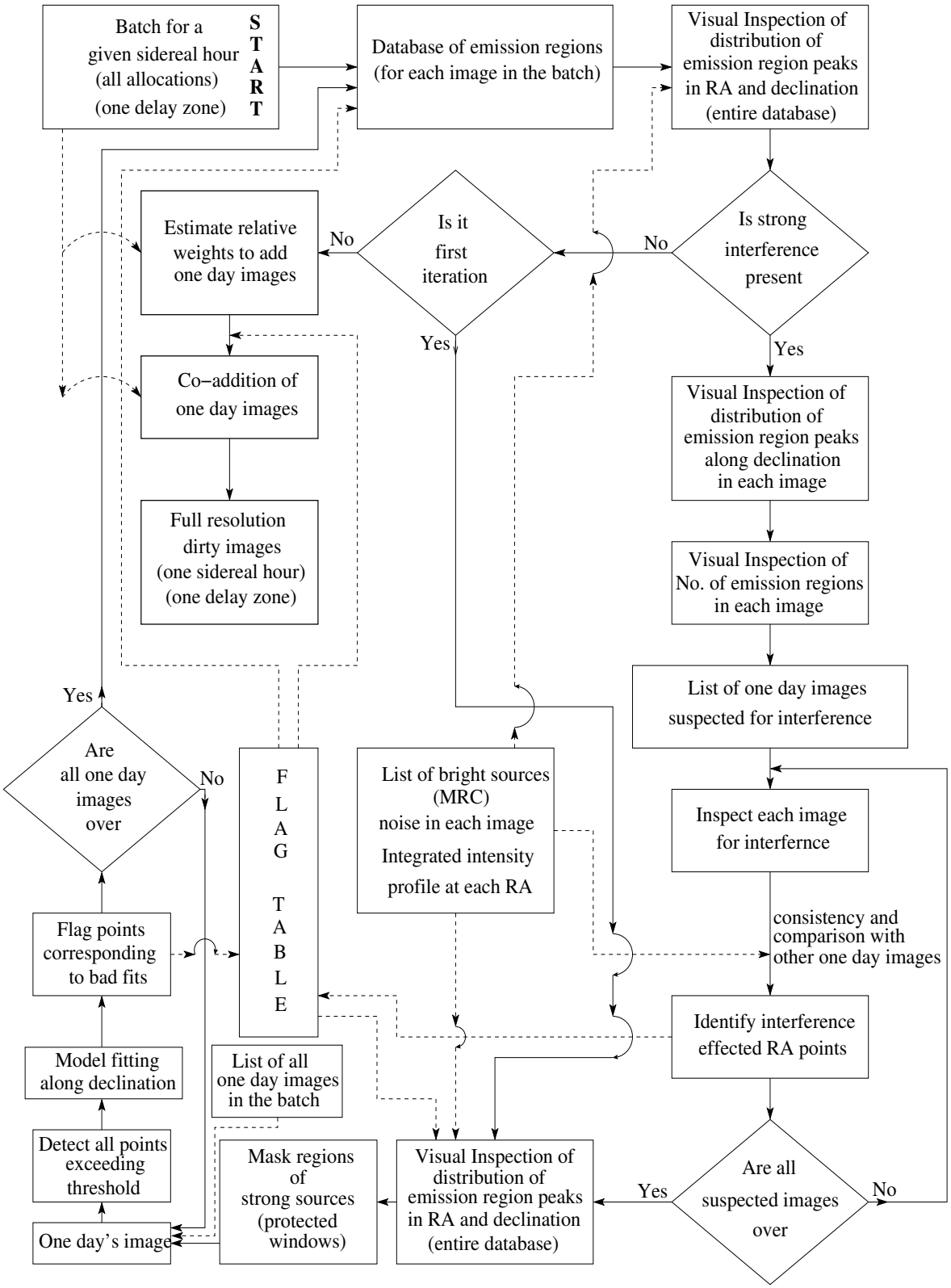


Fig. 4.8: Flow chart showing the steps of collective interference detection using all the one day images corresponding to a given sidereal hour and a delay zone. The solid lines indicate the actual flow of logic of the interference detection procedure while the dashed lines indicate transfer of information between the different processes.

catalogue) expected in the region is used to assist in differentiating interference from genuine signal due to a strong radio source. When the position of a strong source coincides with the stripe along declination, a comparison with the plots of other images (preferably of nearby allocations) of the location of the emission regions at that sidereal time helps in distinguishing interference from a genuine astronomical signal. At each RA point the intensity along declination (corresponding to the delay zone) is integrated to finally obtain a one dimensional integrated intensity profile as a function of RA. The integrated intensity profiles around the sidereal time suspected to be effected by interference is compared with the profiles of other images preferably of nearby allocations. This comparison further helps in demarcating between the interference and genuine astronomical signal. A decision based on these plots and consistency with the other images is made and sidereal times affected by interference are noted down. After this in a similar way all the images suspected to be affected by interference are inspected. This entire process is carried out in *matlab* (using the emission region database) which is user friendly and the user can carry out all his required operations quickly using inbuilt optimized commands.

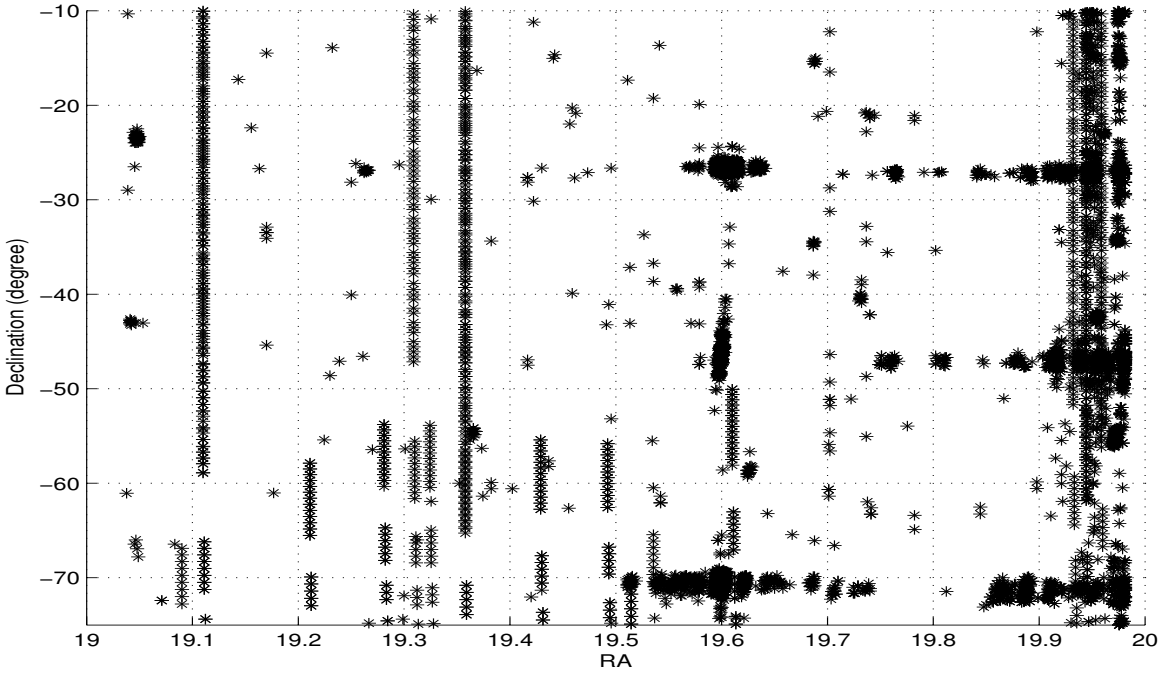
Model fitting :

To remove interference which is not sufficiently strong to appear as a stripe along declination, the following method has been developed. A two dimensional region in RA and declination is masked around emission regions based on visual inspection (See Fig. 4.9) and using well known strong sources from the Molonglo Reference catalogue (MRC) (Large et. al., 1981). These masked regions are called as high emission regions or protected windows. We do not expect strong astronomical sources in each day's image in the remaining regions to have strength more than 5σ. A program goes through each of the individual images and finds out any region of the image in which intensity is more than 10σ (excluding the protected windows).

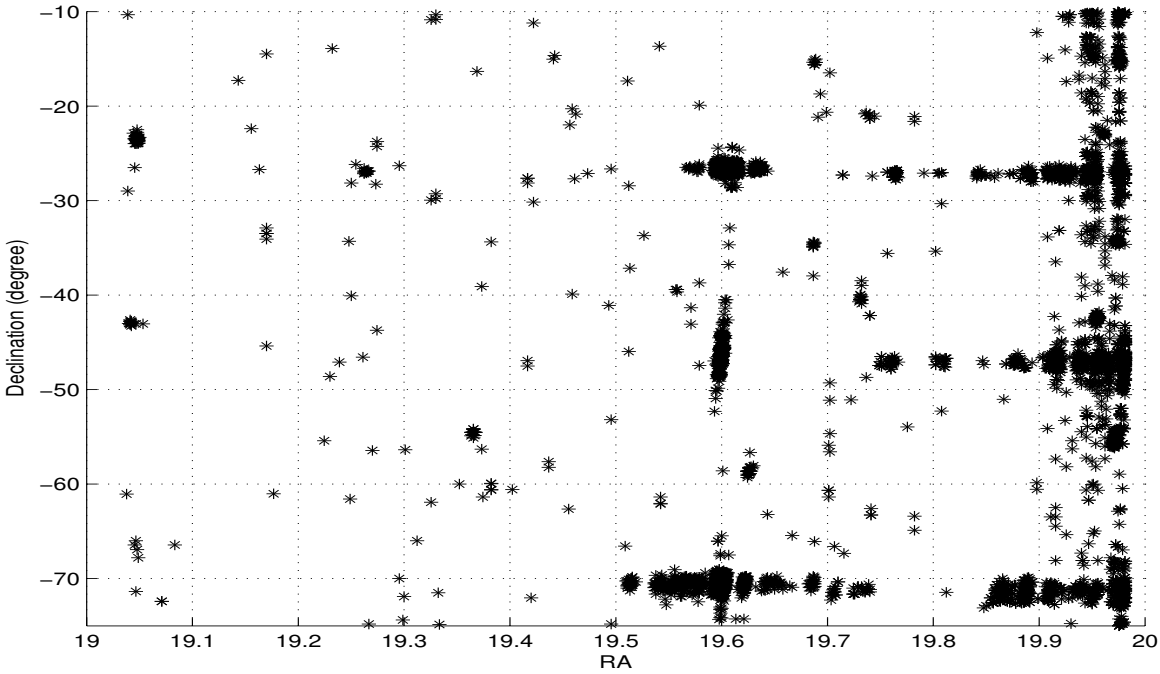
The sidereal times at which pixels have intensity more than 10σ are noted down and a one dimensional fit at the peak is carried out in declination using the expected beam. When the peak amplitude of the fit is more than 10σ or the obtained χ^2 is very high, as in case of a bad fit, the entire 1-D scan of the image along declination at that sidereal time is classified as interference. A one dimensional fit along declination is carried out as the interference at a given sidereal time generally affects a large angular extent in declination while it does not affect the adjacent points in RA. This step is carried out automatically by calling a C program which also creates a flag table for each image containing the interference information.

Interference free images :

The entire cycle (from generating database of emission regions up to model fitting) is

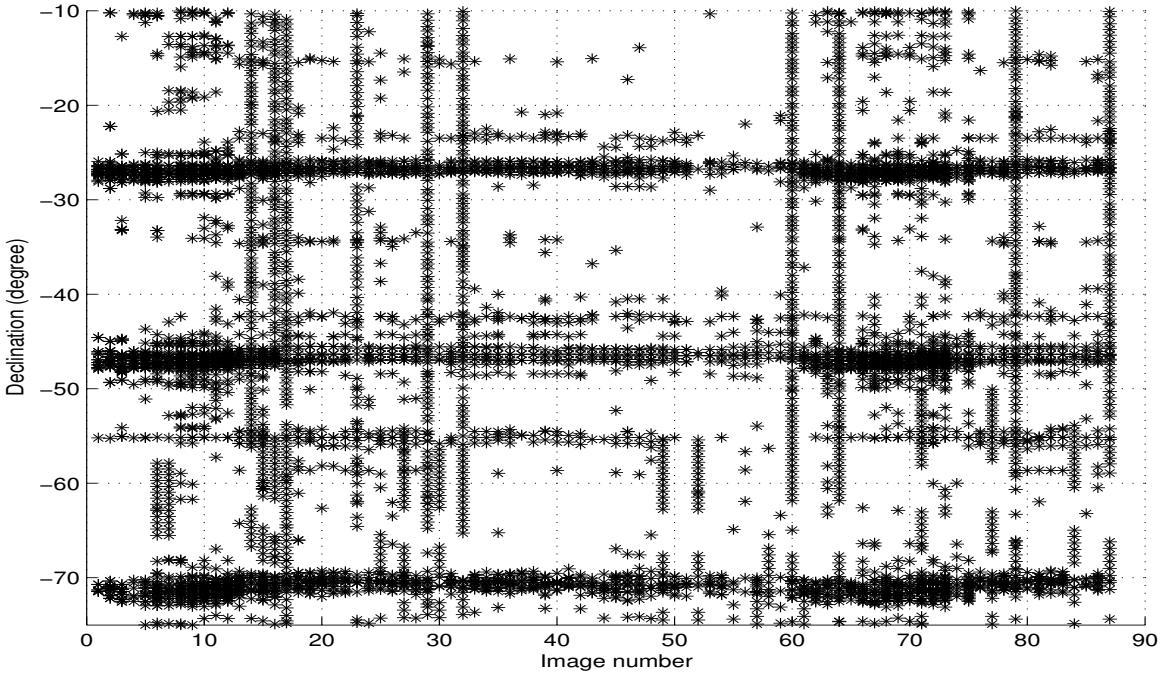


(a) Position of emission region peaks before interference detection in all one day images superposed.

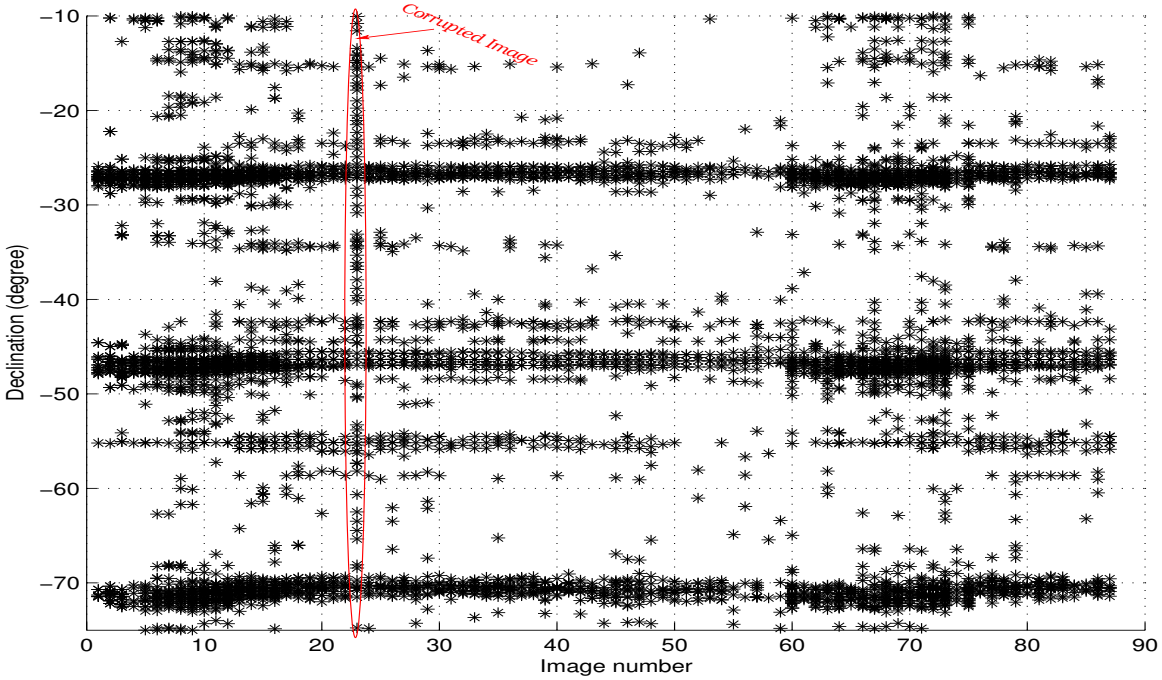


(b) Position of emission region peaks after interference detection and removal in all one day images superposed.

Fig. 4.9: The plot shows the peaks (marked as *) of emission regions in all the one day images in a batch superposed. The lower plot shows the resulting images after interference excision. The vertical strips at constant RA along declination (upper plot) are suspected to be due to interference. Very strong sources appear as a cluster of points with a spread in both directions. Multiple images of very strong sources in declination is seen due to grating response along declination due to discrete sampling of visibilities in NS direction. The vertical strip present near RA 19.6 hrs is due to strong source MRC1932-464 (≈96 Jy at 151.5 MHz) and near RA 20 hrs is due to CygnusA (3C405, ≈9,000 Jy at 151.5 MHz) in grating lobe.

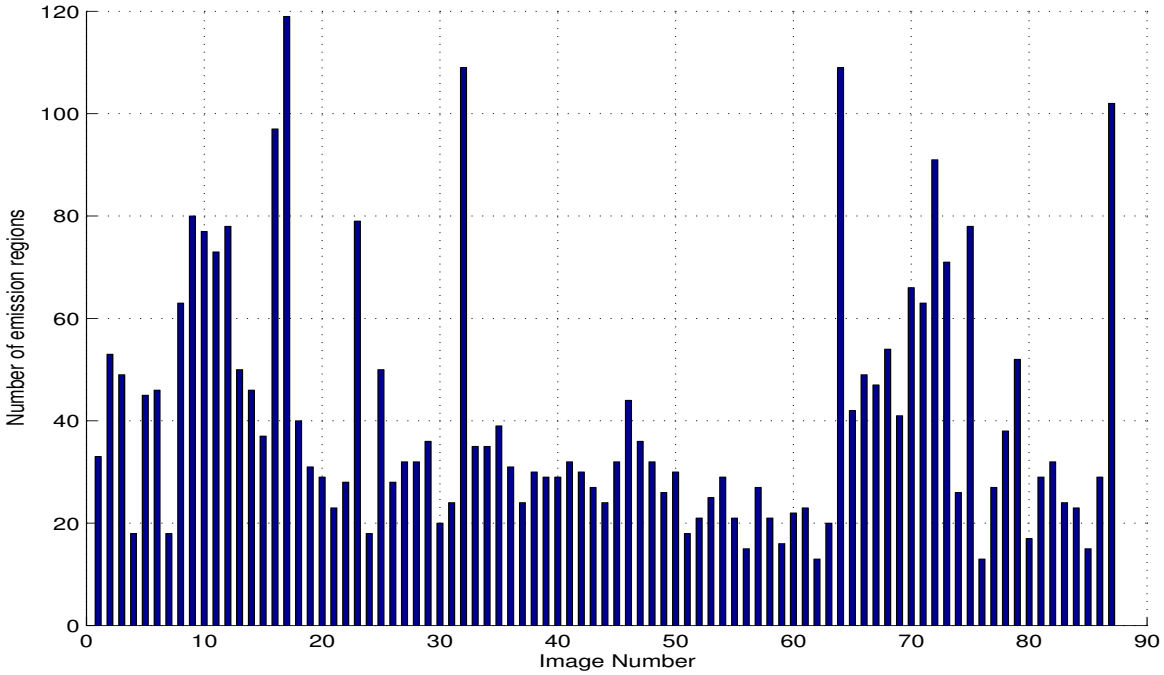


(a) Distribution of positions of emission regions peaks (marked as *) along declination in each day's image before interference detection.

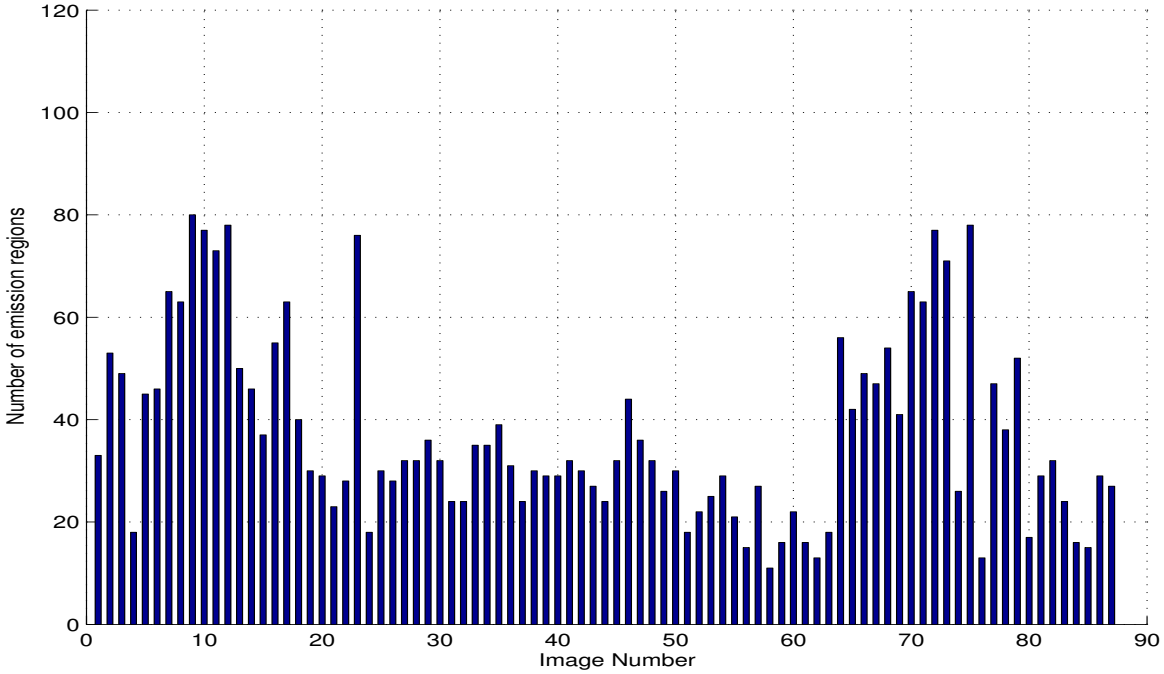


(b) Distribution of positions of emission regions peaks (marked as *) along declination in each day's image after interference detection and its removal.

Fig. 4.10: Distribution of position of emission region peaks (marked as *) along declination in each one day image as a function of image number. The lower plot shows the resulting images after interference excision. The images having continuous vertical strips along declination (upper plot) are suspected to be affected by interference and are individually investigated. The lower plot shows that interference in all the one day images could be detected and removed except incase of image number 23 which is heavily corrupted and is finally rejected.

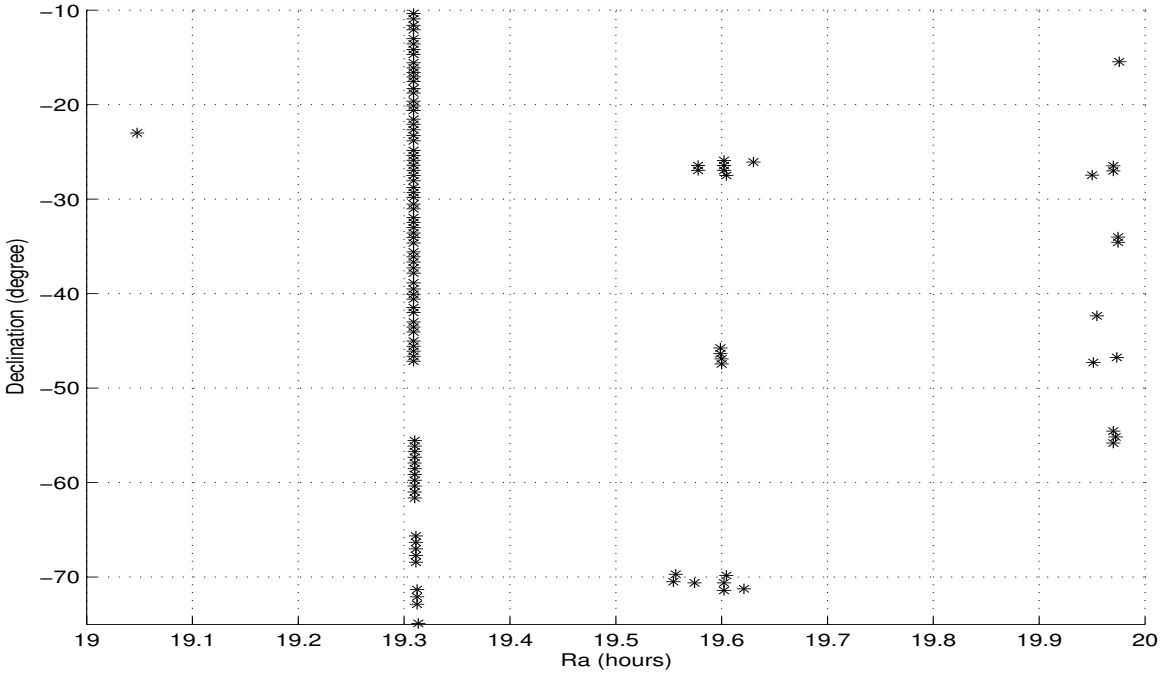


(a) Total number of emission regions in each day's image before interference detection.

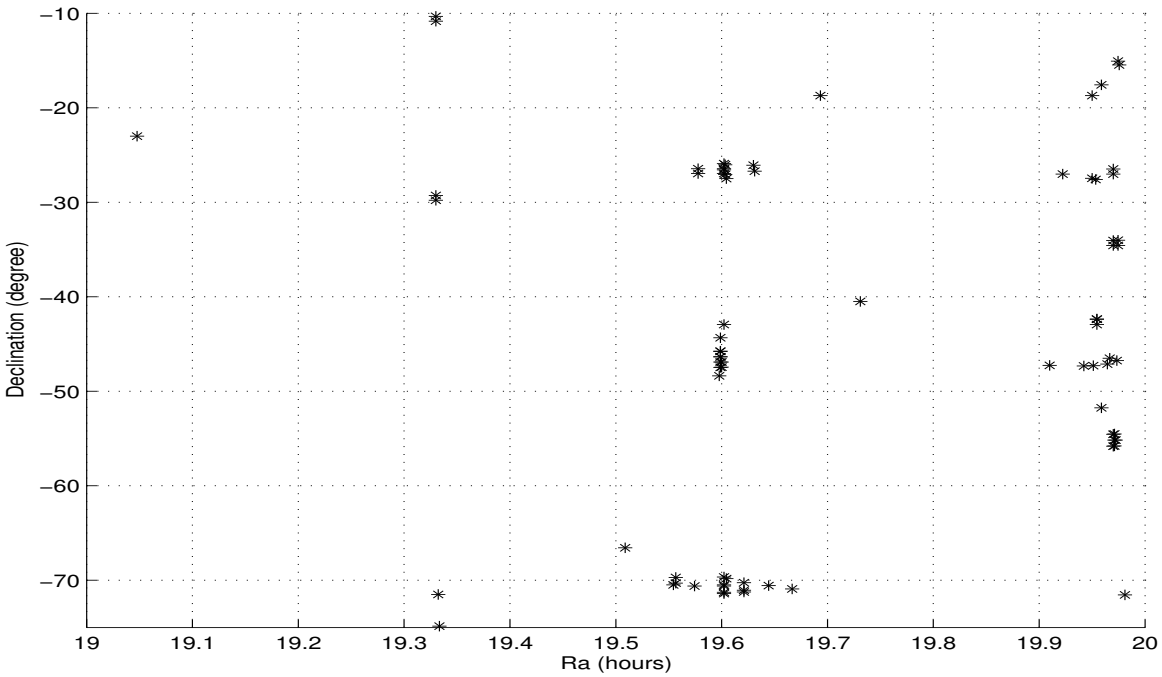


(b) Total number of emission regions in each day's image after interference detection and its removal.

Fig. 4.11: Histogram of the total number of emission regions detected in each day's image before (upper plot) and after interference detection (lower plot). Images with unusually high number of emission regions are suspected to be effected by interference and individually investigated. Images affected by interference after its removal show normally expected number of emission regions except in case of heavily corrupted images which need to be discarded.



(a) Position of emission region peaks before interference detection in one day's image suspected to be affected by interference.



(b) Position of emission region peaks after interference detection and its removal in one day's image.

Fig. 4.12: The plot shows the peaks (marked as *) of emission regions in one day image suspected to be affected by interference. The lower plot shows the resulting image after interference excision. The vertical strips at constant RA along declination (upper plot) is due to interference which disappear once interference has been detected and removed (lower plot).

repeated iteratively untill we obtain images free from any perceivable interference. Images with high interference are summarily rejected. Generally, two iterations have been found to be sufficient. Comparison of various plots before and after interference detection is complete are shown in Fig. 4.9, Fig. 4.10, Fig. 4.11 and Fig. 4.12. It clearly demonstrates that interference detection and excision is successful and there is no perceivable interference in the images.

To obtain the full resolution dirty image covering the given sidereal hour and declination range corresponding to a delay zone, images made for all the 63 allocations (days) are co-added with appropriate weights. During co-addition the 1-D image scans along declination at the sidereal times affected by interference (detected using the combined data set of all one day images after post-integration) in the individual one day images are estimated by linearly interpolating the nearby 1-D scans unaffected by interference (See Sec. 5.4.1). The hierarchical RFI mitigation system as described ensures that the full resolution dirty image obtained after combining images of different allocations is free from interference. This is also a crucial requirement before deconvolution else it can lead to instabilities. In the next section we discuss interesting aspects of statistics of interference in our data.

4.4 RFI statistics

We present interesting results of the statistical analysis of interference in the database of 20,000 hours of astronomical observations for the survey. The statistics revealed that ≈10% of the data is affected by interference. The total data affected (≈15%) during the day time (6 a.m.-6 p.m.) is three times compared with (≈5%) in the night (6 p.m.-6 a.m.). This confirms the general belief that night time observations are relatively interference free compared to day time observations[6].

Statistics of strength :

The number of interference points falls monotonically with the strength of interference as shown in Fig. 4.13 except for an increase around $\approx 260\sigma$, which indicates the presence of some nearby strong source of interference. A histogram of the number of interference points as a function of the strength of interference (up to $10{,}000\sigma$) is shown in Fig. 4.14 for the day time, night time observations and for the total data. The number of interference points is highest in the strength bin around 10σ, during both day and night time observations. About 50% interference points have strengths less than 28σ in the total data, less than 31σ during the day and less than 23σ during the night. There are 90% interference points having strengths less than 580σ in the total data, less than 700σ during the day and

[6]Comparison of interference during day and night time is always carried out by considering equal amount of data.

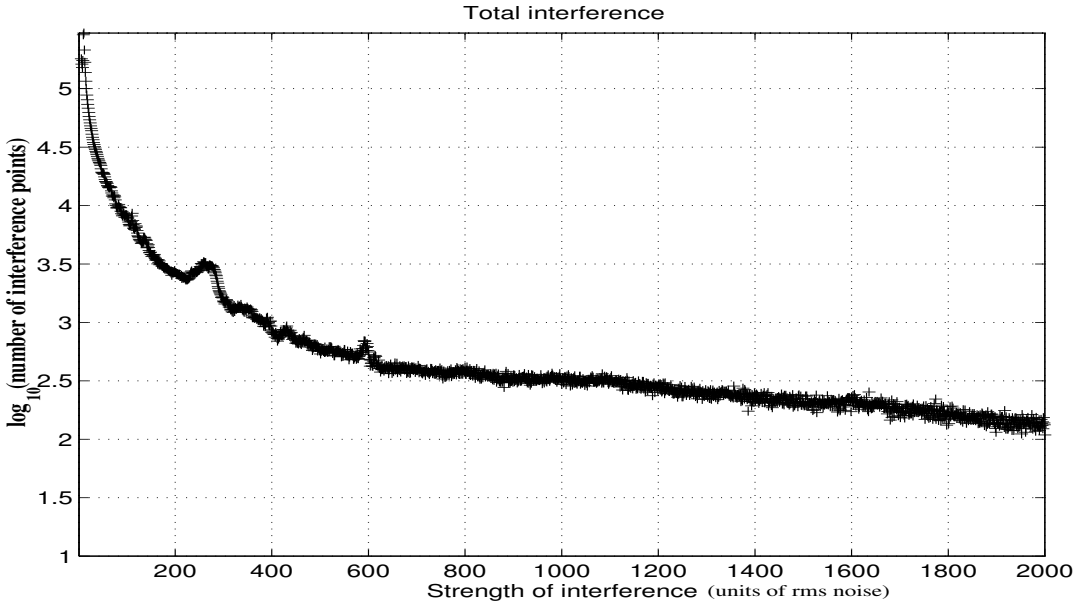


Fig. 4.13: The number of interference points as a function of its strength fall monotonically with the strength of interference except for a significant increase *i.e.* bump at around 260σ and a small bump around 600σ. These sudden increase indicates presence of some nearby interference source.

less than 380σ during the night. This reveals that the interference generally has higher strength in the day time as compared to during the night.

The number of interference points above a strength N(>S), can be approximated as a sum of two exponentials, $N(>\ S) = A + B\ \exp(-S/C) + D\ \exp(-S/E)$, where S is strength of interference in units of σ. The values of the parameters and their uncertainties (indicated in brackets) obtained after fitting are A=415192 (0.55%), B=1.73639e+06 (0.67%), C=198.764 (0.93%), D=3.75137e+06 (1.11%) and E=18.6835 (1.47%). Fig. 4.15 shows the number of interference points above a given strength, the fitted sum of two exponentials, along with the individual exponentials, as a function of strength of interference. As seen in the plot the obtained fit is in good agreement with the measured value. At this stage we are unable to associate physical processes likely to be responsible for such a behavior.

Statistics of duration :

A histogram of the frequency of interference bursts as a function of their duration is shown in Fig. 4.16 for the observations during the day, night and for the total data. In the total data about 45% of the interference bursts last for less than one integration period (≈1.1 s) while it is 41% and 56% for the observations during the day and the night respectively. **About 88% of the interference bursts last for less than 4 integration periods (≈4.4 s) in the total data whereas it is 87% and 92% for the observations during the day and the night respectively**. It reveals that the interference is generally short lived and justifies our approach that most of the interference excision can be carried out at the level

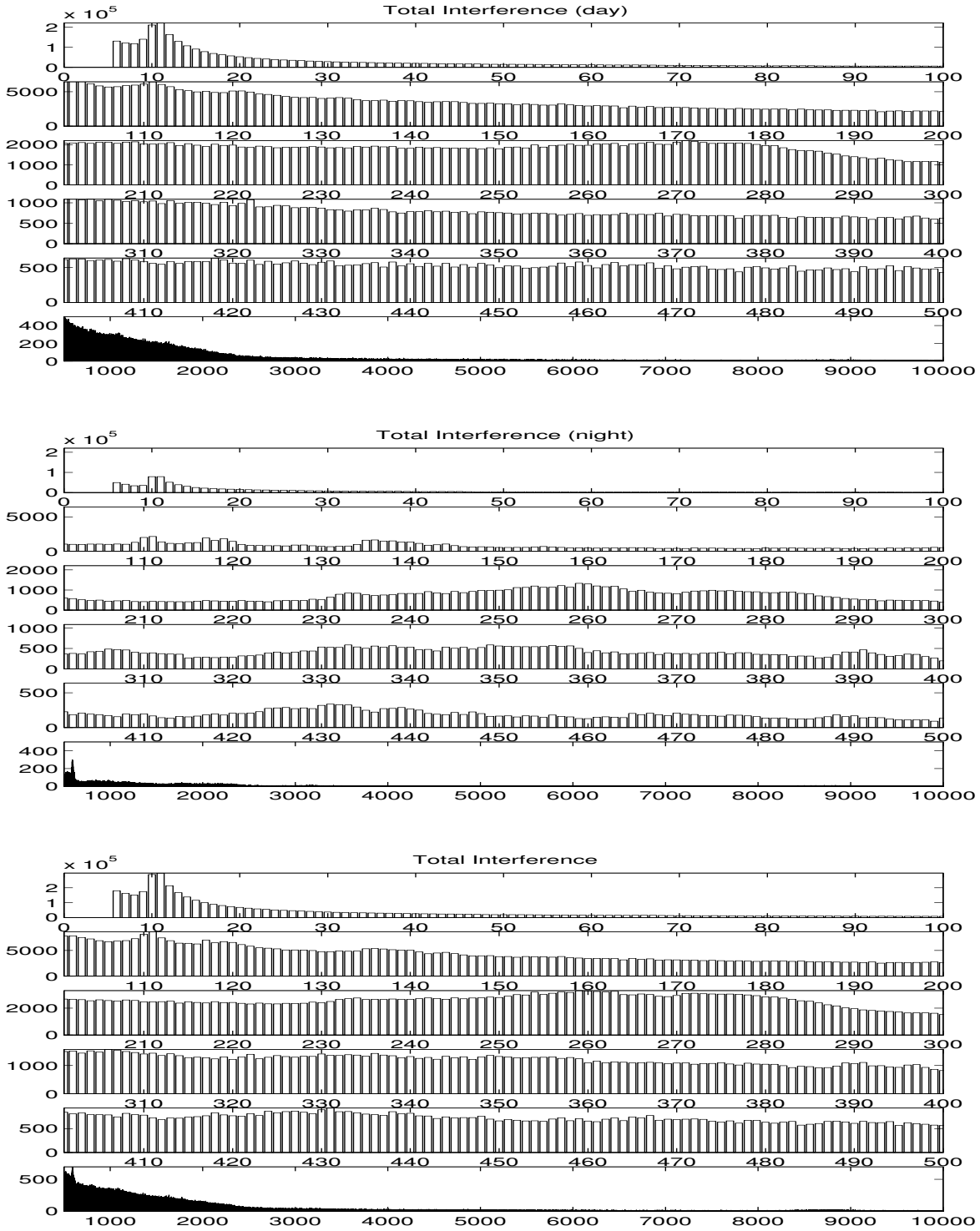


Fig. 4.14: Histogram of number of interference points detected as a function of their strength for the observations during the day (top-most plot), night (middle plot) and total data (lower plot). The number of interference points, peak in the strength bin $\approx 10\sigma$ and fall monotonically with the strength of interference. The comparison also shows that night time observations are less affected by interference as compared to day time observations.

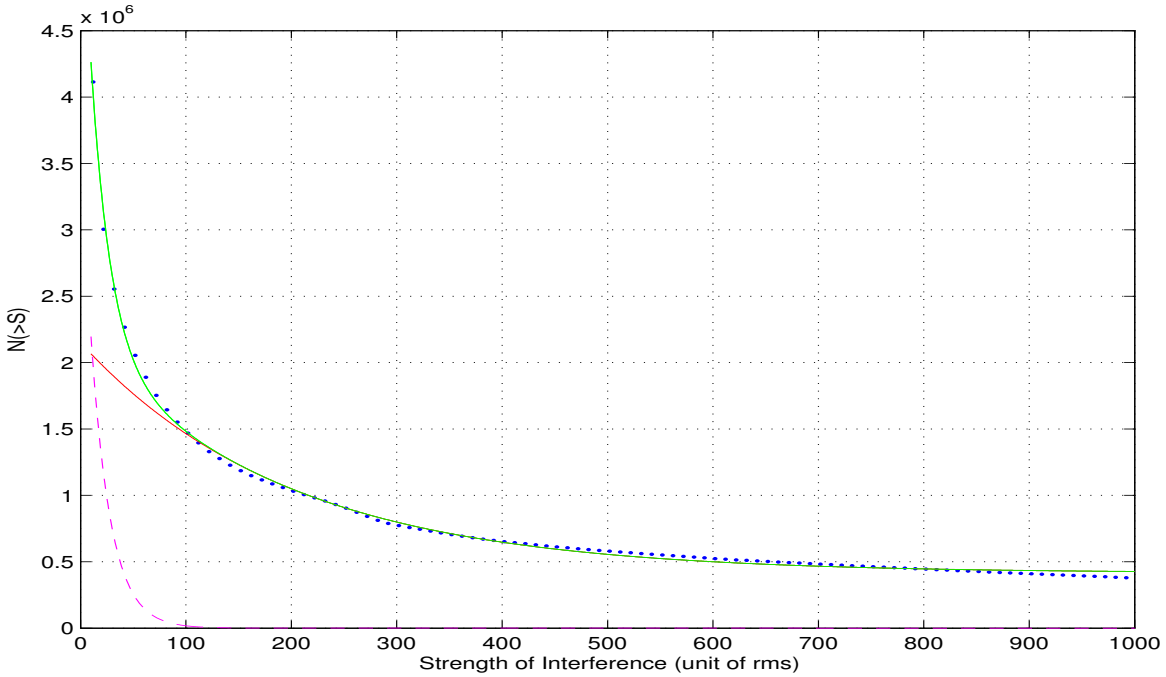


Fig. 4.15: The number of interference points above a given strength (N>S blue dots) as a function of interference strength. The fitted curve obtained by modeling it as a sum of two exponentials (shown as green) is in good agreement with the measured data (blue dots). The individual two exponentials are also shown (red (solid) and magenta (dashed)).

of each day's image by giving zero weights to the points effected by interference during post-integration. It also reveals that the duration of interference is lesser during night as compared to day time.

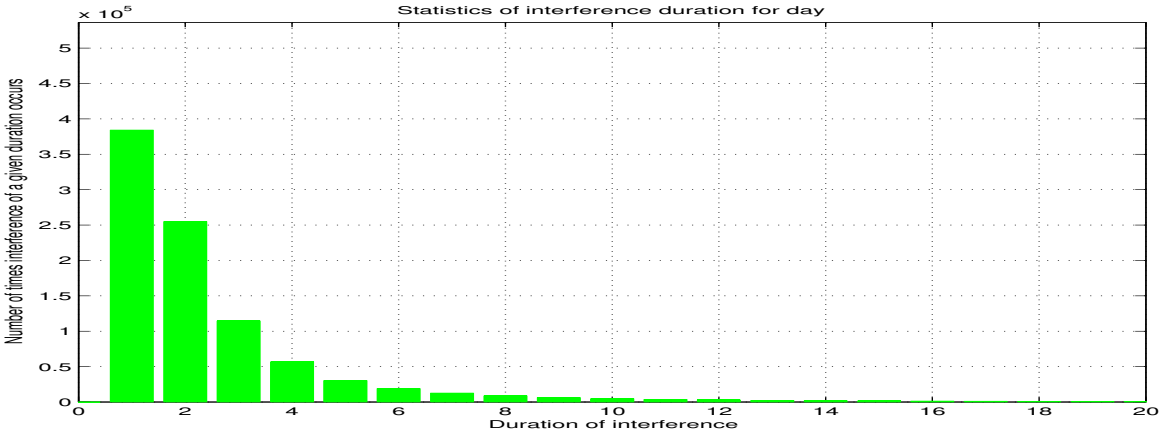
Statistics of interference with MST :

We have complete uninterrupted observations available for 222 Julian Days (JD). Fig. 4.17 shows the total number of interference points in each MST [7] hour range. There is a marked increase in interference during working hours of the local industry (MST 8-16 hrs) which accounts for $\approx$ 62% of the total interference in the entire sidereal day (24 sidereal hours). Thus the interference during the working hours of the industry is nearly four times compared to other times of the day. This increase indicates that the increased interference during the day is perhaps linked to the local industry if effect of solar interference is not the sole cause behind it. We need to compare the interference on public holidays in order to distinguish between the two.

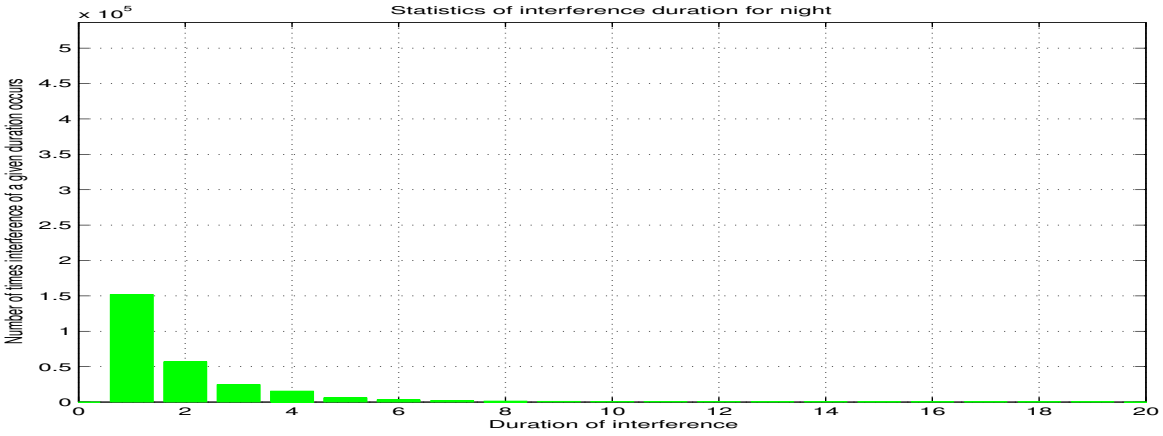
Statistics of interference with JD :

Fig. 4.18 shows the total number of interference points as a function of Julian Day for 222 days. We do not see any systematic trend in the number of interference points with

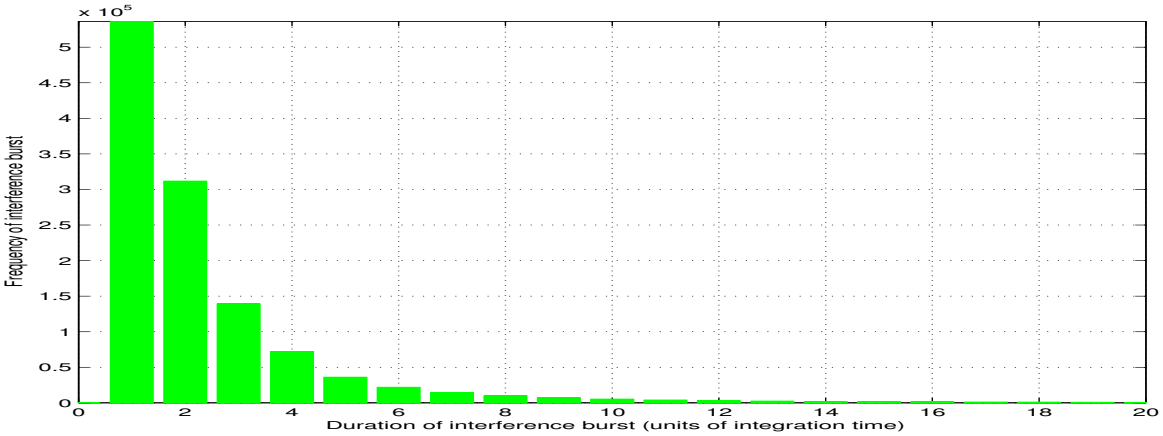
[7] Mauritian Standard Time



(a) Histogram of duration of interference for day time observations.



(b) Histogram of duration of interference for night time observations.



(c) Histogram of duration of interference including both day and night time observations.

Fig. 4.16: Histogram of number of interference bursts as a function of their duration. About 88% of the interference bursts last for less than 4 integration periods (≈4.4 s) in the total data where as it is 87% and 92% for the observations during the day and the night respectively. It reveals that the interference is generally short lived. It also reveals that the duration of interference is lesser during nights as compared to day time.

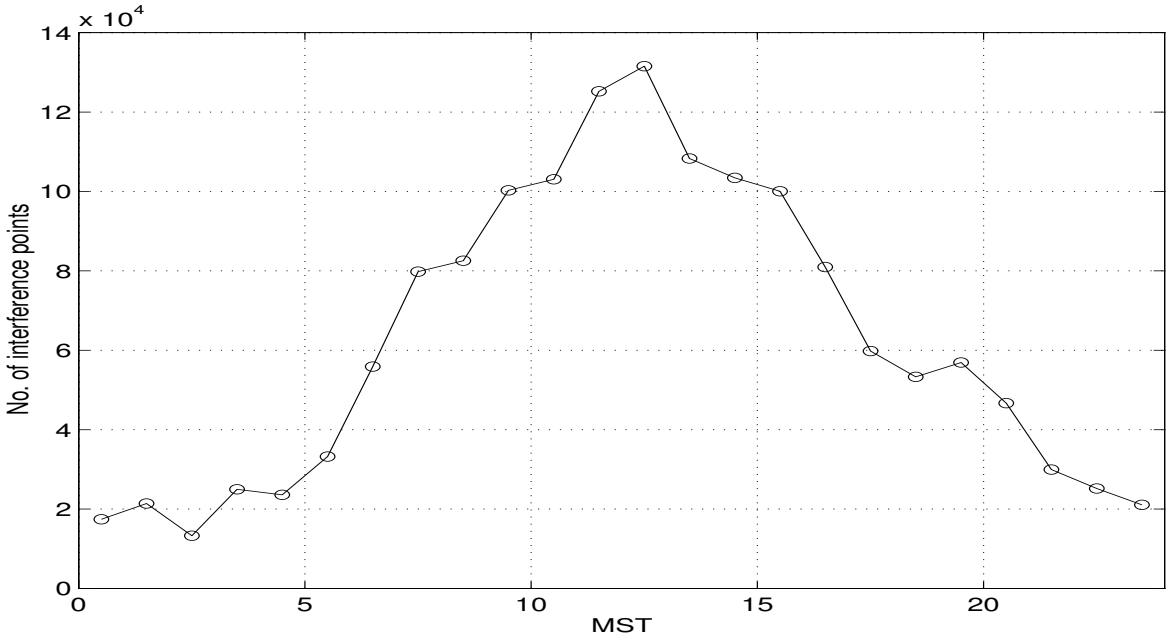


Fig. 4.17: The number of interference points as a function of MST (Mauritius Standard Time) hour ranges. There is a marked increase in interference during working hours of the local industry (MST 8-16 hrs) which accounts for about 62% of the total interference in the entire sidereal day (24 sidereal hours). This increase indicates that the increased interference during the day is perhaps linked to the local industry if effect of solar interference is not the sole cause behind it. We need to compare the interference on public holidays in order to distinguish between the two.

Julian Day. The general belief that interference has significantly increased in the last few years does not seem to be true for MRT.

Interference detected at various stages :

We also analysed the interference detected at different stages during the data processing[8]. The number of interference points detected in the visibilities is ≈9.7% as compared to ≈0.3 % in the images. In the total 9.7% interference detected in the visibilities nearly 9.6% is detected by Fourier filtering and the remaining 0.1% is detected by Hampel filtering. Among the interference points detected in the images ≈0.28% is detected in each day's image before post-integration and remaining 0.02% is detected using the combined data set of all one day post-integrated images. The interference points detected in the images are all isolated in time and there are hardly any interference points which are continuous for more than 3 integration periods (≈3.3 s). It shows that there are very few interference points which escaped the automatic interference excision in the visibilities. The analysis also reveals that more than 99.7% of the interference is detected automatically, the remaining is detected by semi-automatic methods. We have no information about the frequency of interference, but our preliminary analysis suggests that generally the strength of inter-

[8]In this para the statistics is based on about 1,500 hours of good data used for imaging.

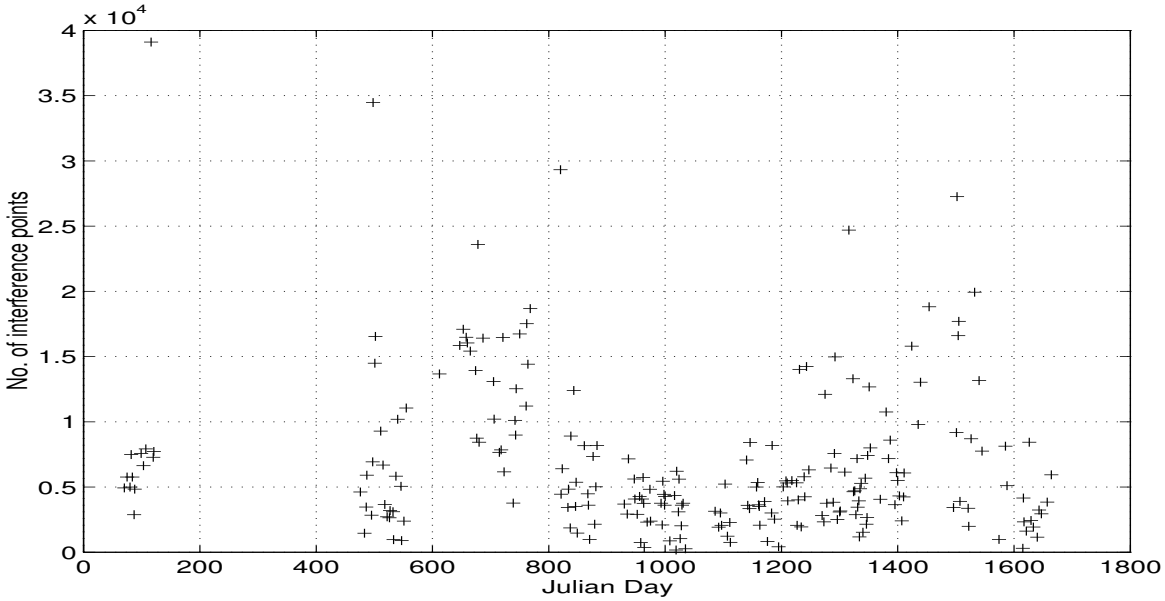


Fig. 4.18: The total number of interference points detected as a function of Julian Day (using uninterrupted observations available for 222 Julian Days). The Julian Day range on the x axis starts from JD 2449500 and extends up to 2451300. We do not see any systematic trend in the number of interference points with JD. The general belief that interference has significantly increased in the last few years does not seem to be true for MRT.

ference is correlated in the different delay zones and interference is narrow band (10 KHz).

RFI mitigation in a low frequency survey comprising of ≈20,000 hours of astronomical observations is a daunting task. Using the developed hierarchical scheme for interference mitigation the time required for manual editing has been reduced to a minimum. The images obtained after applying the RFI mitigation system are free from any perceivable interference and demonstrate its effectiveness. Such an approach exploiting the natural strengths of a variety of signal processing techniques and judiciously applying them at various stages of data processing is an important step towards achieving the ultimate goal of automatic data flagging. In the next chapter we discuss wide field imaging to obtain full resolution images.