

Post-correlation radio frequency interference classification methods

A.R. Offringa¹, A.G. de Bruyn^{1,2}, M. Biehl³, S. Zaroubi¹, G. Bernardi¹,
V.N. Pandey¹

¹*Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands*
e-mail: offringa@astro.rug.nl

²*ASTRON, PO Box 2, 7990 AA Dwingeloo, The Netherlands*

³*Institute for Mathematics and Computing Science, University of Groningen, P.O. Box 407, 9700 AK Groningen, The Netherlands*

15 September 2009

ABSTRACT

We describe and compare several post-correlation radio frequency interference classification methods for use in radio astronomy. Since data sizes of observations grow with new and improved telescopes, and new telescopes such as LOFAR are being built in populated areas, the urge for completely automated, robust methods for radio frequency interference mitigation is high. We investigated several classification methods and find that on the data sets we used the most accurate among them is the `SumThreshold` method, a new method formed from a combination of existing techniques, including a new way of thresholding. This iterative method estimates the astronomical signal by doing a surface fit in the time-frequency plane. With a theoretical accuracy of 95% recognition and approximately 0.1% false probability rate in simple simulated cases, the method is in practice as good as the human eye in finding RFI. In addition it is fast, robust, does not need a model of the data before it can be executed and works in almost all configurations with its default parameters. By receiver operating characteristics analysis on artificial data, we compare this method to a mitigation technique based on the singular value decomposition of the time-frequency matrix and several others. Even in the best cases, the singular value decomposition shows significant errors in classification.

Key words: instrumentation: interferometers – methods: data analysis – techniques: interferometric – radio continuum: general.

1 INTRODUCTION

Around 1980, when the radio spectrum was getting more and more occupied by the technical advancements of human kind (Pankonin & Price 1981), radio observers started to mitigate the radio frequency interference (RFI) caused by electronic equipment (Thompson, Gergely, & Vanden Bout 1991). Until recently, on-line thresholding and manual flagging of post correlated data used to be sufficient to suppress RFI artefacts in the data. However, as the volume of data increased significantly in size, the required sensitivity of observations and the contamination of RFI by an increased usage of electronical equipment grew, new methods had to be developed to deal with the situation.

RFI mitigation can be applied in two different stages: a pre-correlation stage and a post-correlation stage. Strong RFI can be detected in the first stage of interferometry, before correlation. Any residual RFI has to be removed during the data reduction or imaging stage, which is often per-

formed manually, for example by finding appropriate clipping levels for contaminated baselines until the reduced data is free of artefacts. The pre-correlation mitigation stage is very powerful, as during this step the observational data is available at the highest resolution. For example, methods exist to blank or subtract short periodic radar RFI bursts on-line (Niamsuwan et al. 2005), leaving the astronomical signal intact with only a very slightly increased signal to noise ratio. However, pre-correlation methods have to handle large amounts of data in very short time. Therefore, these methods can often only access limited dimensions of the data, such as the data from a single dish or station or the data from a small time range only, because of hardware constraints.

Several methods from signal processing have been used to perform the first pre-correlation mitigation stage, such as thresholding using a Neyman-Pearson detector (Leshem & van der Veen 2000), spatial filtering with eigenvalue decomposition of a spatial correlation matrix (Leshem & van der

Veen 2000), the CUSUM method (Baan, Fridman, & Millenaar 2004) and adaptive cancellation with a reference antenna (Barnbaum & Bradley 1998). During the post-correlation phase, flagging is often the only option left. Other options for post-correlation mitigation include using an independent RFI reference signal to subtract the RFI (Briggs et al. 2000), fringe fitting (Athreya 2009) or post-correlation spatial filtering, but none of them are applicable or sufficient in all cases or for all types of RFI.

In modern observatories that operate at low frequencies such as the Westerbork Synthesis Radio Telescope (WSRT), the Giant Metrewave Radio Telescope (GMRT), the Low Frequency Array (LOFAR), and the Expanded Very Large Array (EVLA), RFI mitigation is an essential component in the signal processing. In the case of LOFAR, there are high sensitivity requirements, especially from the Epoch of Reionization project (see Jelić et al. (2008); Thomas et al. (2009) for recent progresses), yet in what might be a busier RFI environment, in data sets up to a petabyte in size. RFI mitigation before correlation remains important (Boonstra et al. 2005), yet the amount of data is too large for manual post-correlation flagging, implying the need of automated flagging strategies.

RFI comes in many types (Fridman & Baan 2001). The strong RFI that is problematic is often either local in frequency, such as RFI caused by television stations, airplanes and radar, or is local in time, i.e., broadband RFI, caused by phenomena as lightning, high-voltage power cables and electrical fences. Sometimes, the frequency of RFI drifts over time as shown later in Figure 6(a), which can be caused by Doppler shifting of a satellite signal or by imperfect transmitters. A different class of RFI is caused by weakly transmitting but stationary – and therefore systematic – devices on site. This class of RFI is hard to recognize, as it might cover all channels in a spectral band. In fringe stopping interferometers, the fringe rotation causes this type of RFI to have a sinusoidal response in the time-frequency domain (Thompson, Moran, & Swenson 1986). This RFI can be recognized and subtracted in various ways, as for example described fairly recently by Athreya (2009).

To select an RFI strategy, several considerations should be taken into account:

- The achieved true/false positive ratio of the RFI classification;
- The speed of the algorithm;
- Detection or recovery; whether detection and flagging of contaminated areas is sufficient. In certain situations, it might be necessary to recover contaminated data, i.e., to subtract the RFI from the data;
- The effects of RFI mitigation on the noise. Differences in the observed noise strength caused by RFI is for example fatal for the LOFAR Epoch of Reionization experiment (Jelić et al. 2008).

In this paper we will evaluate the effectiveness of several automatic post-correlation RFI mitigation methods and their combinations. The methods will be compared to each other in order to be able to pick a general optimal post-correlation RFI strategy. We will do this by testing the methods on artificial data and data from WSRT that has been observed at the frequency range of LOFAR. Testing the methods on WSRT data will also provide an indication

of the effects of the RFI environment on future LOFAR observations.

In the next section we describe a new method to flag RFI. We present our results including the comparative study in section 3. In section 4 we discuss the results. In section 5 we discuss some future directions of further efforts in this area.

2 METHODS

Radio astronomers have developed their own ways of dealing with RFI during data reduction within the many astronomical software packages. Still in many cases, this implies flagging by hand – a tedious and time consuming job. Many toolkits, such as AIPS¹, AIPS++², MIRIAD³ and NEWSTAR⁴, provide specific features to perform flagging, such as the FLAGR task in AIPS++. Astronomers have automated the process further by designing scripts in which common signal processing techniques such as thresholding, smoothing, line detection and curve fitting are combined. This has for example been described by Winkel, Kerp, & Stanko (2006). Another common signal processing technique known as Singular Value Decomposition has recently been used for the automatic removal of broadband RFI in GMRT observations (Pen et al. 2008). We will describe some of the techniques available that relate to a new method we will introduce, and finally we will explain the new method itself.

2.1 Post-correlation thresholding

Since RFI increases the measured absolute amplitude, thresholding is an effective method that is often used to remove strong RFI. The threshold level is often globally determined, or sometimes set relative to the variance or mode distribution parameters per baseline, which can be stably estimated with for example the Winsorized variance or mode (Fridman 2008). All values that are outside a certain range around the mean or median are flagged as bad data, and not used in subsequent data reduction. Sometimes, a certain amount of samples around this bad data sample are flagged as well. Most astronomical reduction toolkits provide options to threshold part of a data cube, allowing different thresholds, at the cost of an increased effort for the astronomer. An important consequence of thresholding is that good data is selected with a bias; when many non-contaminated samples are above the threshold, flagged and not used in subsequent data reduction, artefacts such as incorrect flux densities might be caused in the image plane. It is therefore important to have a low false-probability rate of RFI detection.

¹ AIPS: Astronomical Image Processing System, <http://aips.nrao.edu/>.

² AIPS++, sequel of AIPS, <http://aips2.nrao.edu/>.

³ MIRIAD, a data reduction package tailored for the Australia Telescope Compact Array (ATCA), <http://www.atnf.csiro.au/computing/software/miriad/>.

⁴ NEWSTAR, a data reduction package tailored for the 45-m radio telescope of the Nobeyama Radio Observatory (NRO), <http://www.nro.nao.ac.jp/~enewstar/>.

2.2 Surface fitting and smoothing

A surface fit of the correlated visibilities $V(\nu, t)$ as a function of frequency ν and time t can produce a surface $\hat{V}(\nu, t)$ that represents the astronomical information in the signal. Requiring $\hat{V}(\nu, t)$ to be a smooth surface is a good assumption for most astronomical sources, as the observed amplitude of astronomical sources tend not to change rapidly in the time and frequency direction, while specific types of RFI, on the other hand, create sharp edges in the time/frequency domain. The residual between the fit and the data contains the system noise $N_{\text{noise}}(\nu, t)$ and the RFI, $N_{\text{RFI}}(\nu, t)$, which can then be thresholded without the chance of flagging astronomical sources that have visibilities with high amplitude.

Several suitable surface fitting methods exist. As an example, in Winkel et al. (2006) a pipeline is described in which a two-dimensional low order, dimensional independent polynomial is iteratively fitted on time-frequency tiles in the data using a least square fit:

$$\hat{V}_k(\nu, t) = \sum_{i=1}^{N_\nu} a_{k,i} \nu^i + \sum_{i=1}^{N_t} b_{k,i} t^i + c_k, \quad (1)$$

where \hat{V}_k is the fitted surface that represents the astronomical information in the k -th tile, N_ν, N_t are the polynomial order for the frequency and the time respectively and $a_{k,i}, b_{k,i}, c_k$ are the coefficients of the fit for tile k .

The fit is performed iteratively, and values which have been flagged in previous iterations are excluded from the fit. This can be done by introducing a weight function $W_F(\nu, t)$, where $W_F(\nu, t) = 0$ indicates that the value is flagged or outside the boundaries of the measured time or frequency range, and $W_F(\nu, t) = 1$ means the value is accepted. The fit is performed by minimizing an error function E_k for each tile:

$$E_k = \sum_{\nu} \sum_t W_F(\nu, t) f(\hat{V}_k(\nu, t), V(\nu, t)) \quad (2)$$

where $f(a, b) = a^2 - b^2$ for a least squares fit or $f(a, b) = |a - b|$ for a fit with a minimal absolute error.

An example of this approach after a few iterations can be seen in Figure 1. In simple cases, the surfaces that are created with this approach represent the astronomical information reasonably well, and the method is also quite fast. However, as polynomial fits tend to show deviations near the borders, the method is inaccurate near the borders of each tile.

Compared to tile-based approaches, sliding window methods tend to be more accurate. A simple example of a sliding window approach is to calculate the average of a window of size $N \times M$ around each data value:

$$\hat{V}(\nu, t) = \frac{1}{\text{count}} \sum_{i=-\frac{1}{2}N}^{\frac{1}{2}N} \sum_{j=-\frac{1}{2}M}^{\frac{1}{2}M} W_F \cdot V(\nu + i\Delta\nu, t + j\Delta t), \quad (3)$$

with

$$\text{count} = \sum_{i=-\frac{1}{2}N}^{\frac{1}{2}N} \sum_{j=-\frac{1}{2}M}^{\frac{1}{2}M} W_F(\nu + i\Delta\nu, t + j\Delta t) \quad (4)$$

This method is still fast and creates a surface without tile edges. However, the sliding window average represents the astronomical signal less well. For example, peaks in the

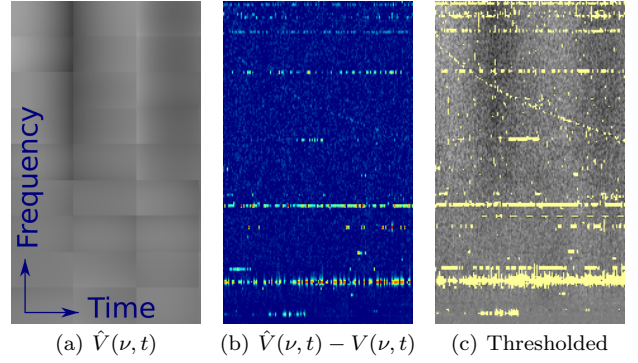


Figure 1. Tile-based polynomial fitting applied to the raw visibilities of a 144m WSRT baseline of a 3C196 observation at 140 MHz (see §3.3). Panel (a) shows the tiled fit of the astronomical signal. Panel (b) shows the difference between the fitted astronomical signal and the observed signal used for thresholding. Panel (c) shows the flags on top of the original signal. The flags established by single pixel thresholding cover the RFI when verified by eye, though many false positives can be seen which are caused by “normal” noise. The tile size used for this image is 30 frequency channels with 10 kHz width \times 50 time scans with 10s integration time.

original function cause square-shaped edges in the fit, which in the end cause classification inaccuracies.

One way to overcome this problem is to calculate the local median instead of the local average. Values that have been flagged in a previous iteration should be ignored by the median calculation. The median is insensitive to peaks and the surface created by the local median remains smooth when the window is slid over the data. The median however is not always a good estimator of the sliding window centre sample specifically, as all samples have equal weight.

Another way to overcome the problem is to calculate a weighted average. Consider a weight function $W_d(i, j)$ that depends on the two components i, j that represent the distance in respectively time and frequency direction to the centre of the window:

$$\hat{V}(\nu, t) = \frac{\sum_{i=-\frac{1}{2}N}^{\frac{1}{2}N} \sum_{j=-\frac{1}{2}M}^{\frac{1}{2}M} W_d(i, j) (W_F \odot V)(\nu_i, t_j)}{\text{weight}} \quad (5)$$

With

$$\text{weight} = \sum_{i=-\frac{1}{2}N}^{\frac{1}{2}N} \sum_{j=-\frac{1}{2}M}^{\frac{1}{2}M} W_d(i, j) W_F(\nu + i\Delta\nu, t + j\Delta t) \quad (6)$$

This can be calculated very fast, since (5) is the convolution operation $W_d * (W_F \odot V)$ and (6) is another convolution $W_d * W_F$, giving:

$$\hat{V} = ((W_F \odot V) * W_d) \oslash (W_F * W_d) \quad (7)$$

Where \odot and \oslash are the elementwise multiplication and division operators. A good choice for W_d is the two-dimensional (dimensional independent) Gaussian function:

$$W_d(i, j) = \exp\left(-\frac{i^2}{2\sigma_\nu^2} - \frac{j^2}{2\sigma_t^2}\right) \quad (8)$$

Together, equation (7) and (8) essentially describe a weighted Gaussian smoothing operation, or more specific, a Gaussian smoothing operation with missing data. The

parameters σ_ν and σ_t can be used to specify the level of smoothing in respectively the frequency and time direction. Since the weight function is dimensionally separable, the convolutions can be dimensionally separated:

$$\hat{V} = \frac{(W_F \odot V) * U_\nu * U_t}{W_F * U_\nu * U_t} \quad (9)$$

With $U_\nu(i) = W_d(i, 0)$ and $U_t(j) = W_d(0, j)$. Each of the convolutions in (9) is a one dimensional convolution, and this is therefore a fast operation.

2.3 The cumulative sum method

The cumulative sum (CUSUM) method is a well known analysis method used to detect changes in the distribution parameters (Page 1954), such as in quality control in productional environments. If the cumulative sum of sequential samples exceeds an adaptive threshold, the system enters an alarmed state and changes can be made to correct quality. In its common form, the likelihood for two values of the distribution parameters is used to compute the threshold.

To turn this method into an RFI mitigation strategy, the total observed power or power received at a certain frequency of a single dish can be used as the sequential input values to the CUSUM method and the likelihoods of either variance or mean of RFI can be estimated using the variance of the signal (Baan et al. 2004). Observing can be stopped as soon as RFI is detected, and continue when reception has turned back to normal. This method can be easily implemented for on-line RFI detection, as it is simple and fast.

The CUSUM method does not estimate the start time of the change, it only detects the change quickly, which nevertheless might cost time and thus some bad data may be leaked until the method detects faint RFI. Hence, the method is more applicable for a first rigorous check on the data then to actually perform flagging. The subsequent sections will describe a method that combines the detection strength of the CUSUM method with the possibility to perform flagging off-line.

2.4 Combinatorial thresholding

RFI bursts often affect multiple samples which are connected either in frequency or time. We will now introduce a new threshold mechanism that makes use of this knowledge: we will flag a combination of samples when a property of this combination exceeds some limit. Assume that A and B are neighbouring samples. In normal thresholding, we will look at each of the samples A and B individually and flag one of them if it exceeds some ‘‘single sample’’ threshold χ_1 . For combinatorial thresholding, a new flagging criterium is added: if A and B do not exceed the single sample threshold χ_1 individually, they still can get flagged when A and B are *both* exceeding a somewhat lower threshold χ_2 . If not, they are combined with a third neighbour, C , and thresholded at χ_3 , etc. The more connected samples are combined, the lower the sample threshold.

Given a set of strictly decreasing thresholds, $\{\chi_i\}_{i=1}^N$, a value will be classified as RFI in case it belongs to a combination of i values in either time or frequency direction, of which all absolute values are above threshold χ_i . To determine whether a single sample should be flagged because of

an RFI sequence in the frequency direction, the following rule is applied:

$$\text{flag}_{\nu M}(\nu, t) = \exists i \in \{0 \dots M - 1\} : \forall j \in \{0 \dots M - 1\} : |V(\nu + (i - j) \Delta\nu, t)| > \chi_M$$

Where M is the number of samples in a combination. The flagging rules for the time direction are correspondingly determined. Finally, a sample is flagged if any of the two rules evaluates to true. We will call this method the **VarThreshold** method.

We will show a simple example to demonstrate the method. Consider the following values:

$$V = \begin{pmatrix} 1 & 2 & 1 & 4 \\ 4 & 1 & 1 & 4 \\ 2 & 2 & 1 & 4 \end{pmatrix} \quad (10)$$

Each row represents a frequency channel and each column represents a time scan. Assume the high values in the fourth column were caused by broadband RFI. When using a normal threshold $\chi = 3$, all samples with value 4 would be thresholded, including one false positive. If however we would use combinatorial thresholding, with $\chi_1 = 5$ and $\chi_2 = 3$, we would threshold only the three broadband RFI samples.

The method can be described as thresholding all samples at a certain χ_M and subsequently filtering all samples that have less than M samples in both directions. From this perspective, it is easy to add other morphological constraints. The above method looks for straight lines in the horizontal or vertical direction. An enhanced version might flag connected shapes covering a specific area size, or shapes that form a line or curve in the plane, possible not connected, that are likely to be caused by RFI.

2.4.1 VarThreshold parameters

The following list of parameters need to be optimized to make efficient use of this approach:

- The false positive rate on uncontaminated samples. The lower the value, the more RFI remains. The higher the value, the more uncontaminated samples will be flagged. We will discuss this in §2.4.2.

- A set that defines which samples are combined. For this we define \mathcal{M} , a set containing the number of samples that will be combined for each sample in each of the four directions. Ideally, each sample will be combined with all samples of either the same frequency or the same time, i.e., $\mathcal{M} = \{i \in \mathbb{Z} : 1 \leq i \leq \max(N_\nu, N_t)\}$, with \mathbb{Z} the set of integers. Empirically, a small subset $\mathcal{M} = \{1, 2, 4, 8, 16, 64\}$ works almost as good and saves summing and comparing many samples.

- The set of thresholds $\{\chi_M : M \in \mathcal{M}\}$ for the different number of combinations M . The total set of thresholds is expressed by two parameters, χ_1 (the threshold on a single sample) and ρ , using the following formula:

$$\chi_i = \frac{i \cdot \chi_1}{\rho^{\log_2 i}} \quad (11)$$

A value of $\rho = 1.5$ empirically seems to work well for the **VarThreshold** and the below defined **SumThreshold** method.

To find χ_1 for a desired false probability rate, ρ is kept constant and the χ_1 value is binary searched by performing mitigation on data selected from the distribution of the noise, with the values $\{\chi_i\}_{i \in \mathcal{M}}$ computed as in (11), until the false probability rate is close to the desired rate.

Since the method is combined with a surface fitting strategy, the following parameters are added:

- The number of iterations to be performed. The resulting accuracies are good with about 5 iterations.
- The iteration sensitivity as a function of the iteration number, $\eta(i)$. In each iteration, the threshold sensitivity is increased (more samples are flagged). To accomplish this, all the thresholds $\{\chi_i\}_{i \in \mathcal{M}}$ are decreased by dividing them with a factor of $\eta(i)$. Only during the last iteration a sensitivity of 100% will be used. By slowly increasing the sensitivity a first bad fit of the background won't have much effect, since only the very strongly RFI contaminated samples are removed. Using an exponential function for $\eta(i)$ was found to work well.

2.4.2 The VarThreshold false positive ratio

Assume that $R \sim \mathcal{D}(\sigma_{N_s})$, where R is the residual of the complex correlated visibilities V and the surface fit \hat{V} , and \mathcal{D} is a distribution with parameter σ . The probability that a non-RFI contaminated sample from the residual is larger than χ can be determined with:

$$\forall \nu \forall t : P(|R(\nu, t)| \geq \chi) = \int_{-\infty}^{-\chi} \varphi_{\sigma}(x) dx + \int_{\chi}^{\infty} \varphi_{\sigma}(x) dx, \quad (12)$$

where $\varphi(x)$ is the probability density function of the distribution $\mathcal{D}(\sigma_{N_f})$. Note that the term $\int_{-\infty}^{-\chi} \varphi_{\sigma}(x) dx$ is only relevant when the distribution contains negative values – unlike the Rayleigh distribution – and the values are thresholded above χ as well as below $-\chi$.

The combined threshold false positive rates can best be calculated numerically, since an analytic calculation is rather complex, even for \mathcal{M} with a single combined threshold χ_M . This is because the thresholding of samples is dependent on each other, hence it needs to take into account that the probability of flagging a single sample is not the expected rate of flagged samples. For example, assume that $R(\nu_1, t_1)$ and $R(\nu_2, t_2)$ are independent when they are not RFI contaminated. This is the case if the fit represents the astronomical data and system noise is uncorrelated. With the VarThreshold strategy, the probability P_{false} for a single non-contaminated sample R_1 to be flagged in one of the four combinations with $M = 2$ with one of its neighbours $R_{2\dots 5}$ can be calculated with:

$$\begin{aligned} P_{\text{false}} &= P(|R_1| > \chi \wedge \exists i \in [2 \dots 5] : |R_i| > \chi) \\ &= P(|\mathcal{D}| > \chi) - P(|\mathcal{D}| > \chi) (1 - P(|\mathcal{D}| > \chi))^4. \end{aligned} \quad (13)$$

Larger M will make the formula become more complex. The total false positive ratio (the total percentage of samples that will be flagged being false) will differ from (13), since the probability of each flagged sample depends on the probability of its neighbouring samples.

Figure 2 shows the result of calculating the total false positive ratio numerically, for several values of M .

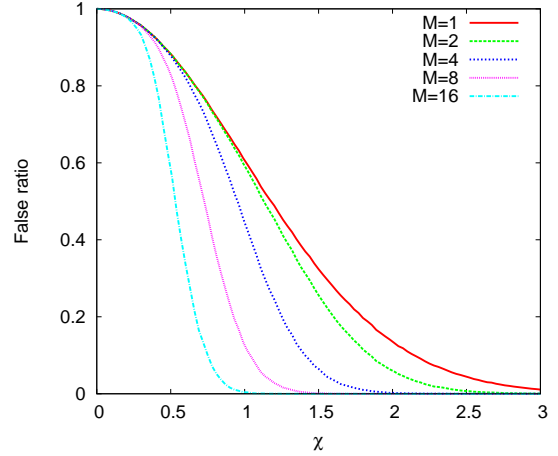


Figure 2. The false positives of the VarThreshold method when flagging with a single combination $\mathcal{M} = \{M\}$ without surface fitting. Samples were selected from a Rayleigh distribution, which is the distribution of the visibility amplitudes. χ is relative to the mode of the distribution.

2.5 The SumThreshold method

Now we will present a variation on the VarThreshold method that improves the classification performance. This method, named the SumThreshold method, is a flagging method that combines samples like the VarThreshold method. In this alternative case, the sum of one or more other samples is used as threshold criterium. As in the VarThreshold method, the threshold χ_M is variable and depends on M , the number of samples that are summed.

Unlike the VarThreshold method, this method allows a sequence of samples to be flagged when there are samples in the sequence that contain values below the sequence threshold value. However, without an additional rule, there are situations in which this method might flag too many samples. For example, consider the sequence $[0, 0, 5, 6, 0, 0]$ that represents a strong RFI contamination in two samples. When the SumThreshold method without a second rule is applied with average threshold values $\chi_1 = 7$, $\chi_2 = 5$, $\chi_3 = 4$, ..., $\chi_6 = 1.8$, all six values would be thresholded, as their average exceeds $6\chi_6$. The following rule is therefore added: the values are thresholded in the order $\chi_1, \chi_2, \dots, \chi_M$. When a lower threshold has already classified samples as RFI contaminated, the samples will be left out of the sum and replaced by the average threshold level. In the example case, the values 5 and 6 will be classified as RFI by the second threshold, and therefore will be replaced by χ_6 when combining all the six samples. The average of the sequence for the sixth threshold is therefore calculated as $(0 + 0 + \chi_6 + \chi_6 + 0 + 0) / 6 = \frac{2}{6}\chi_6$. As a consequence, only the samples with values 5 and 6 are flagged.

2.5.1 The SumThreshold false positive ratio

We calculate the theoretical false positive ratio for $M = 2$ like in the VarThreshold method. The probability $P(T_{\chi,1,2})$ that the sum of two independent random samples exceeds a

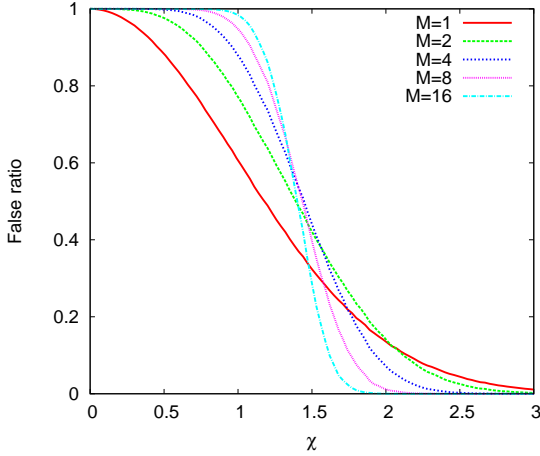


Figure 3. The probability of a false positive when thresholding with a single combination $M = \{M\}$ using the `SumThreshold` method without surface fitting. The Rayleigh distribution was used for the simulation. χ is the average threshold relative to the distribution mode, thus a combination of samples was thresholded when their sum exceeds $\chi \times M \times \sigma$. The false ratio for $M \geq 2$ is different from the `VarThreshold` method (Figure 2). Because of this difference, the parameter ρ used to calculate the set of thresholds as in (11) need to be optimized for the methods individually. Although the false ratio is not decreased when comparing this method with the `VarThreshold` method, the true ratio is often increased (Figure 8).

certain value χ is given by:

$$\begin{aligned} \forall \nu_1 \nu_2 t_1 t_2 : P(T_{\chi,1,2}) &= P[R(\nu_1, t_1) + R(\nu_2, t_2) \geq \chi] \\ &= P(\mathcal{D}(2\sigma N_s) \geq \chi) \\ &= \int_{\chi}^{\infty} \varphi_{2\sigma}(x) dx \end{aligned} \quad (14)$$

When thresholding the average of a combination of two samples, each sample will occur four times in a hypothesis test, once with each of its neighbours. On uncontaminated samples, the probability of a false positive for each of these tests is given by (14). The probability for a false positive with the four tests applied on each sample becomes:

$$P(T_{\chi,1 \times 4}) = P(T_{\chi,1,2} \cup T_{\chi,1,3} \cup T_{\chi,1,4} \cup T_{\chi,1,5})$$

Because the tests $\{T_{\chi,1,i}\}_{i=2}^5$ are dependent on each other, it is much easier to calculate the false positive rates numerically. This can be performed by applying `SumThreshold` on a large amount of data selected from the distribution \mathcal{D} . The result of such a simulation is in Figure 3.

2.6 Singular Value Decomposition

A singular value decomposition (SVD) is a mathematical tool for finding the singular values of a matrix, which can exhibit certain properties of the matrix.

A singular value decomposition consists of finding the complex unitary $M \times M$ and $N \times N$ dimensional matrices U and V containing respectively a left and right singular vector in each row, and the diagonal, $M \times N$ dimensional real matrix Σ containing the singular values, such that:

$$A = U \Sigma V^T \quad (15)$$

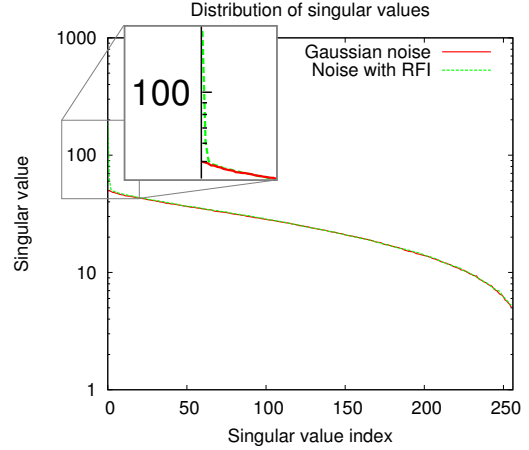


Figure 4. The distribution of the singular values of two artificial measurements: one containing Gaussian noise only, the other containing Gaussian noise polluted by broadband RFI. In this example, the first five singular values are affected by the broadband RFI. In general, the number of singular values that are affected by RFI and the possibility to recognize them varies depending on the orthogonality properties of the RFI.

RFI is mitigated from the data set by performing this decomposition on a matrix A . Each element A_{ij} represents the measured flux, where i is a baseline/frequency index and j a time index. Each given matrix A has a unique solution for the singular values Σ , if the singular values are sorted, but there is no unique solution for U and V (for example, A remains equal when all values in U and V are negated). It is assumed that the highest singular values represent the singular values of the RFI data. To mitigate the RFI, the highest singular values in Σ are set to zero and the new matrix \hat{A} is recomposed from U , Σ and V .

The number of singular values to be removed or set to zero has to be chosen in such a way that only the RFI is removed. The singular values that correspond to RFI are often strong outliers, while the singular values of Gaussian noise form a smooth curve. The position of the abrupt change in the curve of the singular values is used as the number of singular values to be removed, as is shown in Figure 4.

2.6.1 Properties

Let $L = \min(M, N)$, then:

$$A_{ij} = \sum_{k=1}^L U_{ik} \Sigma_{kk} V_{jk} \quad (16)$$

U and V are unitary, $U\bar{U} = I$ with \bar{U} the Hermitian transpose, and the rows and columns of the matrices form by definition a complex orthonormal basis. This implies:

$$\forall i \in [1..M] : \sum_{j=1}^L U_{ij}^2 = 1 \quad (17)$$

Hence there is at least one non-zero value in each row and column of the matrices U and V , and setting a non-zero singular value to zero changes A . If A contains real values only, equation (17) implies that all values in U and V are

between -1 and 1 , and removing a singular value Σ_{ii} can alter each value in A by at most Σ_{ii} . In the complex case, removing a singular value can alter the absolute value of each value in A at most by Σ_{ii} . In general, setting Σ_{ii} to zero subtracts a matrix Γ_i with rank 1 from A , as $(\Gamma_i)_{jk} = U_{ji}\Sigma_{ii}V_{ki}$, and thus all columns are linear dependent on each other.

The orthogonality properties imply that the order of the rows and columns in the original matrix A do not change the singular values: the order of the rows and columns is irrelevant for the SVD method to detect RFI. Intuitively, the SVD method does not “distinguish” between a smoothly increasing amplitude, caused by astronomical sources, and RFI, and might fail to correctly subtract/detect RFI because of the astronomical signal.

If RFI is to be separated from the signal, the RFI and the signal have to adhere to the following properties:

- All columns containing RFI (and consequently all rows) have to be orthogonal to the astronomical signal. In other words, for any column or row \mathbf{a} in the matrix, $\mathbf{a}_{\text{RFI}} \cdot \mathbf{a}_{\text{signal}} = 0$, with \mathbf{a}_{RFI} the RFI component and $\mathbf{a}_{\text{signal}}$ the signal component in the data.
- The singular values of the RFI are substantial higher than the singular values of the astronomical signal. This requires the RFI to be strong.
- The individual RFI columns are either fully linear dependent or fully orthogonal to each other. If the individual RFI components are partially dependent, the largest part of the RFI is removed and the singular value of what is left of the RFI might not have enough ‘gain’ to be removed or flagged.

Iteratively fitting a surface and subtracting the surface, as in §2.2, might improve the compliance to the first requirement, although it increases the execution time of the method. Another way to improve compliance to the requirement is to remove the astronomical signal by subtracting a good model beforehand.

It is useful to note that unitary transformations do not change the singular values of a matrix, although they might change the singular vectors. Since the Fourier transform is a unitary transformation according to Parseval’s theorem, the following equation holds:

$$A = USV \Leftrightarrow \mathcal{F}(A) = U'SV' \quad (18)$$

The consequence of this is that it does not matter whether the SVD method is executed in the time-frequency domain, the time-lag domain, or another Fourier domain, since setting singular values to zero in the Fourier domain would set the singular value to zero in the original domain.

2.7 Input data types

The combined thresholding methods described in this paper can be applied on several types of data: auto-correlated or cross-correlated, on specific polarizations or Stokes parameters, on amplitude or on phase, etc.

We have compared flagging on cross-correlations and auto-correlations. The cross-correlations of each baseline can be processed with one of the flagging methods, resulting in $N(N-1)/2$ correlations to be processed. Alternatively, every

antenna can be individually flagged by processing the auto-correlations, and samples in a baseline might be flagged if either of the corresponding samples in the individual antenna auto-correlations have been flagged. Only N correlations need to be searched for RFI in this case. On top of the benefit of speed, the RFI is stronger in auto-correlations and the data contain no fringes from astronomical sources, as auto correlations do not have interference patterns, offering an improved accuracy in RFI detection. On the down side, some RFI might be present in auto-correlations that would have been mitigated by cross-correlation, and detecting RFI in auto-correlations potentially throws away some usable data in the cross-correlations.

In cases where the polarization of the observed electromagnetic waves is measured, the polarization might contain valuable information for RFI classification. For now, we have processed each polarization individually, without exploiting relations between polarizations.

3 RESULTS

3.1 Surface fitting results

In §2.2 we described several surface fitting methods to estimate the astronomical signal in the frequency-time domain. We found that the surface fitting methods when combined with one of the classification methods do not differ much in accuracy. A sliding window approach was found to be more stable compared to a tile based approach. The Gaussian weighted average, a polynomial fit and the window median for the subtracted surface were found to be approximately equal in their accuracies after optimizing their parameters such as the window size, the Gaussian kernel size and the order of the polynomial, though their parameters do influence the accuracy.

Finding global parameters that always work well (or automatic procedures to find the parameters) is not trivial. The algorithm can handle data with very different characteristics: it can be applied on XX, XY, YX or YY polarizations, auto- or cross-correlations from either long or short baselines, for LOFAR or for WSRT data, before or after calibration, et cetera. To use the same surface fitting parameters in all these different situations, the window size and if applicable the Gaussian kernel size needs to be rather small. The expected amplitude changes of celestial signals are usually much less in the frequency direction, and setting the window size larger in the frequency direction improves stability. We used a typical size of the sliding window of 40 frequency channels \times 20 time scans and Gaussian kernel parameters of $\sigma_\nu = 15$ and $\sigma_t = 7.5$. The numbers are based on trials on different observed and artificial data sets. The parameters are relative to the number of channels and number of time steps. For WSRT data, a channel is 10 kHz wide and a time scan is 10 seconds long. LOFAR will have a 1 kHz \times 1 second correlation output resolution. For best results, the length and width of the window should be about three times the Gaussian kernel size or larger.

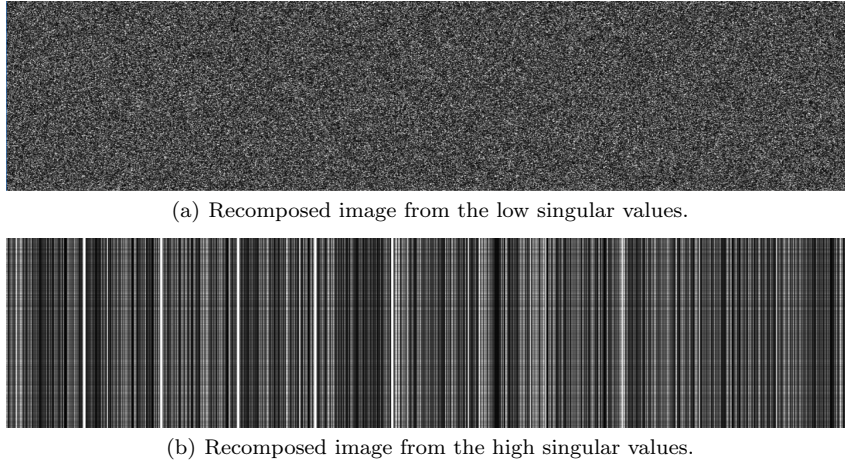


Figure 5. *SVD decomposition of test set A (Figure 7(a)): noise with broad band RFI covering all channels homogeneously. The recomposed image from the low singular values (top panel) looks very promising: none of the RFI is left and the noise seems to be untouched. However, a recomposition of the matrix with only high singular values (bottom panel), i.e., the part that has been subtracted from the image, shows that the noise is affected in an unpredictable way by the decomposition. This is the best case for the SVD method; in more realistic scenario's, the data should include residual astronomical signal and broadband RFI that might not be linearly dependent.*

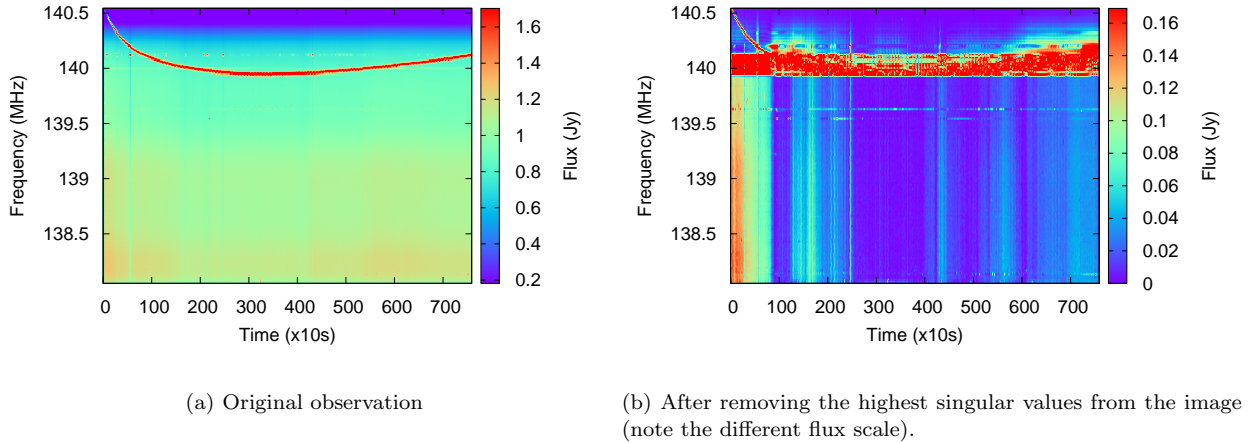


Figure 6. *The auto-correlated data in this image demonstrates the inability of the SVD method to remove sources that slowly change frequency over time (e.g., because the source has a changing velocity in the direction of the antenna). This type of RFI seems to be relatively common in low frequency WSRT data. The RFI in this particular example is so strong that it can be easily removed by thresholding, but this plot is to illustrate the effects of such RFI. When the frequency-changing signal is faint and cannot be removed by thresholding, applying SVD will, like in this example, change the astronomical information in the data in an unpredictable way.*

3.2 RFI classification on artificial data

Both the SVD and threshold methods show accurate results on removing line RFI and broadband RFI. The SVD method is not suitable for removing frequency-varying RFI, and thus has to be complemented with other techniques to remove all RFI. However, the SVD method can be used to subtract and remove the RFI from the image, leaving the astronomical signal intact. For this to be successful, considerable assumptions about the mathematical properties of RFI and the astronomical signal have to be true: the time-frequency matrix with the RFI components has to be orthogonal to the time-frequency matrix of the astronomical signal, and

the different RFI components have to be either orthogonal to each other or linear dependent on each other.

As it is hard to quantitatively compare RFI mitigation methods based on data sets of which the characteristics of the RFI can not be known for sure, several artificial test sets were created. These sets are shown in Figure 7 and contain broadband RFI only. Since the RFI was added artificially, the location of the RFI in the time-frequency domain is known, and the accuracy of the methods can be tested quantitatively. The results are drawn as receiver operating characteristic (ROC) curves in Figure 8. ROC curves show the true probability rate against the false probability rate.

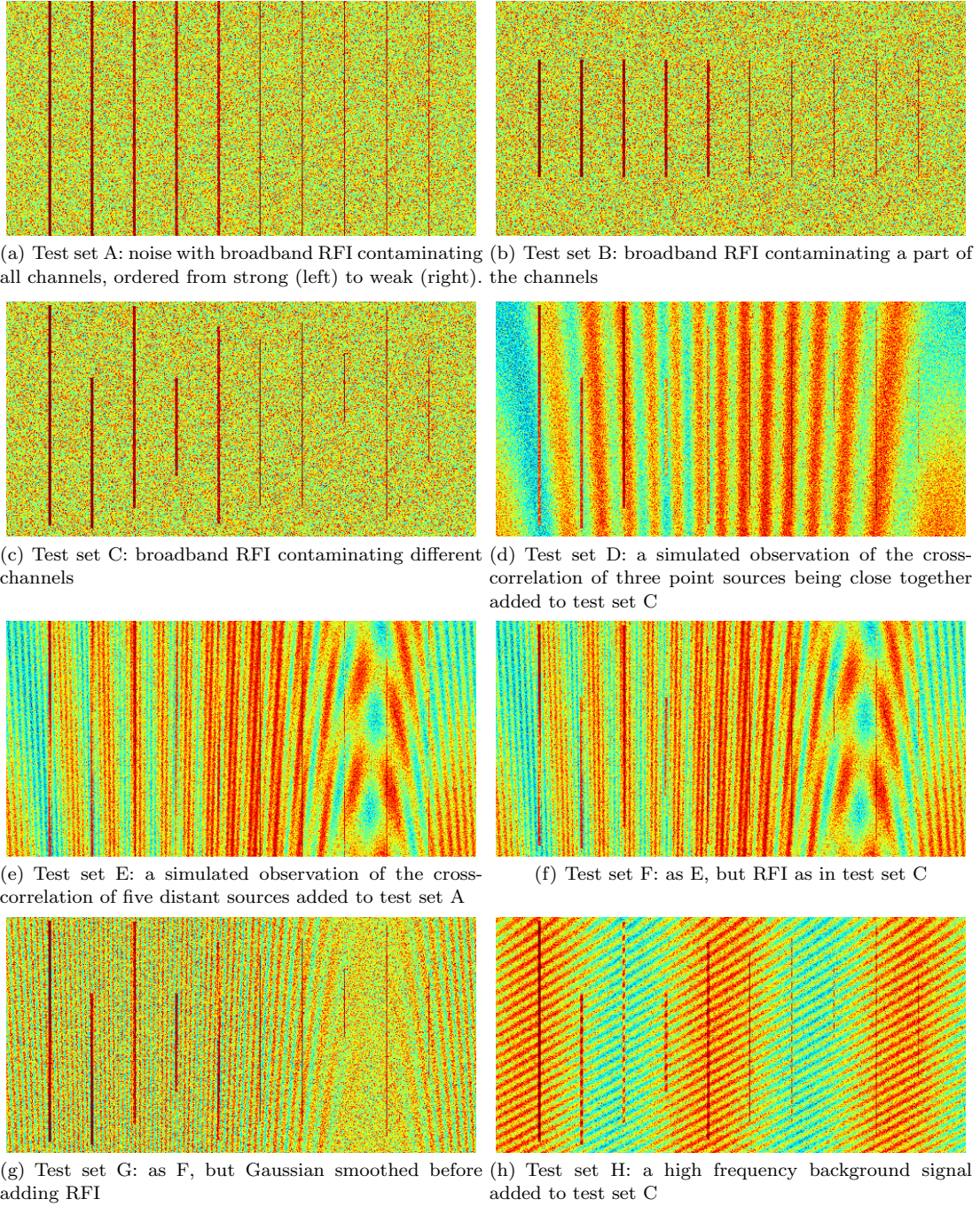


Figure 7. The artificial test sets containing broad band RFI, used for testing and parameter optimization. In all images, time is along the horizontal and frequency along the vertical axis. All test sets simulate a similar baseline.

Different accuracies and characteristics of the methods can easily be compared in ROC graphs.

The `SumThreshold` shows a considerably better accuracy in all the test sets. Test sets A and B contain RFI that is completely linear dependent, and the SVD method works also very well in these sets. The SVD method could actually be used to subtract the RFI, instead of flagging the data and not using the data. However, to mitigate the RFI in test set C, the methods have to deal with RFI that is neither or-

thogonal nor completely dependent on each other, and thus the SVD methods starts showing a decreased accuracy.

A normal thresholding strategy was also tested to compare the results. When performing normal thresholding with a surface fit as in the `SumThreshold` method, the accuracy for thresholding actually decreases in the test cases without astronomical signal (see the curves with label “Fit + simple threshold” in Figure 8). This is partially due to the fact that we optimized the surface fit for the `SumThreshold` method. Further more, since the accuracy of the threshold-

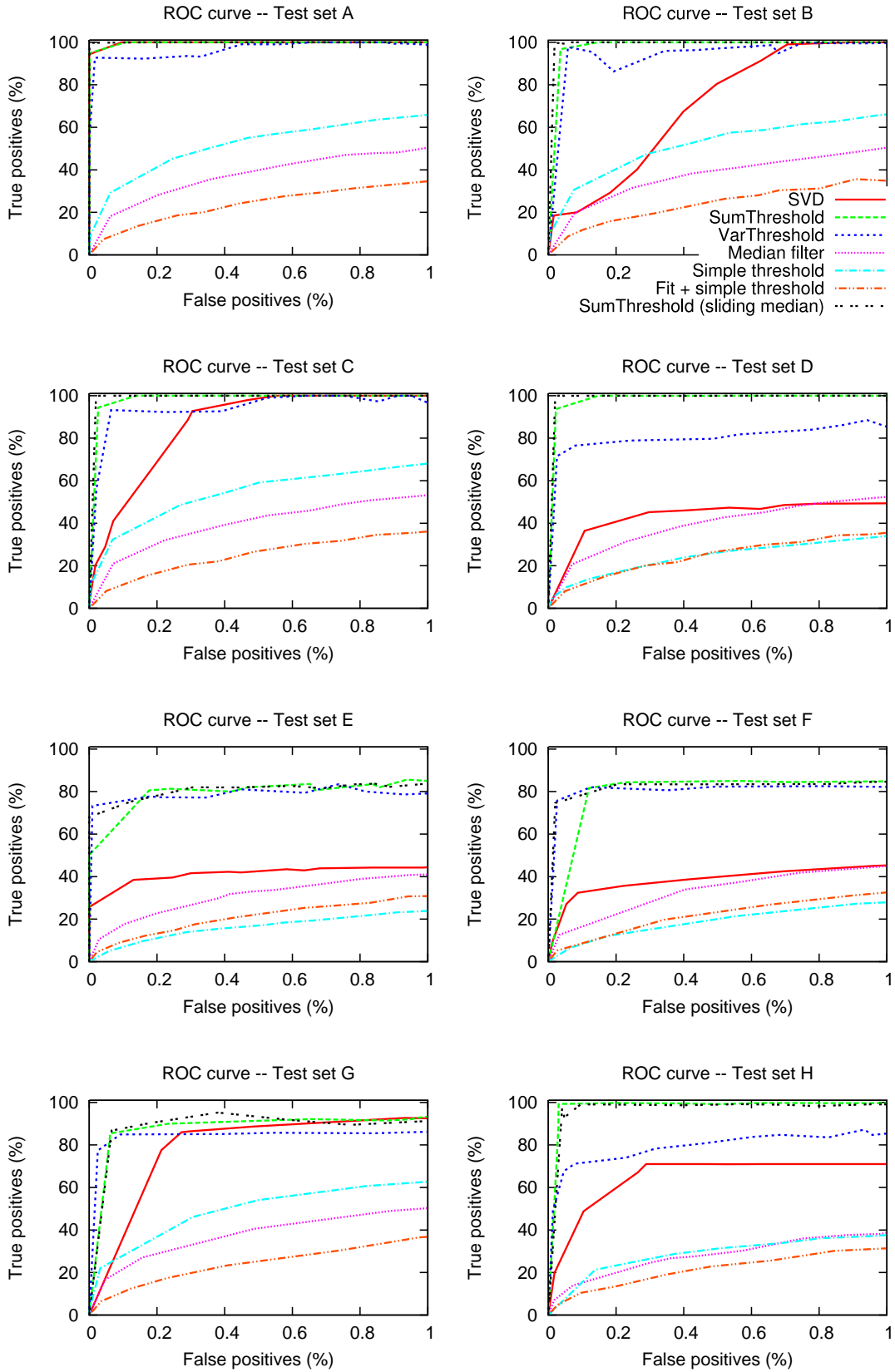


Figure 8. The ROC curves produced by applying various RFI detection methods to the test sets. The closer an ROC curve passes the left-top of the graph at 100% true positives with 0% false positives, the more accurate the method is.

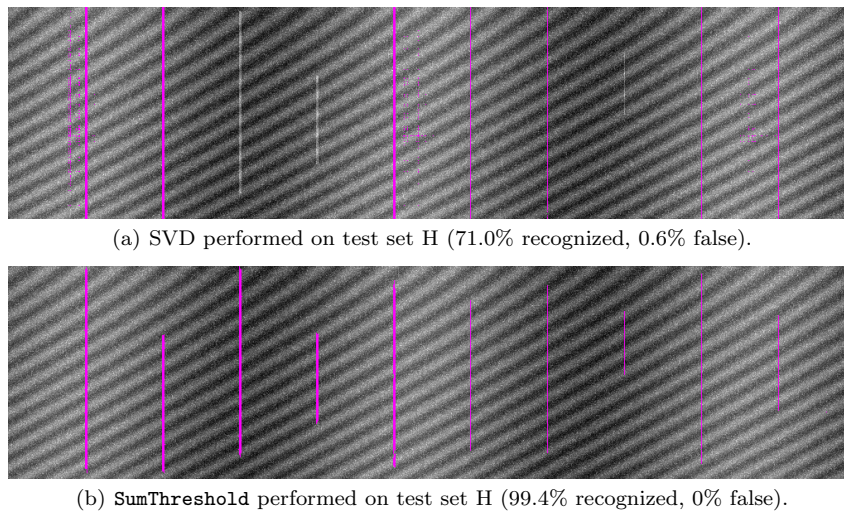


Figure 9. The results of two mitigation methods applied on test set H.

ing is not very good, the fit is influenced by the undetected RFI, causing more errors.

When astronomical information is added as in test set E and a more complex background is added as in test set F, the SVD method shows an increased inaccuracy in mitigating the RFI, as can also be seen in Figure 9. However, in test set G, the background of test set F is Gaussian smoothed and subtracted, as is done before thresholding, and the SVD method shows an improved accuracy, though still not as good as the `SumThreshold` method. Test set H shows that the linear dependency of the RFI is not the only requirement for successful mitigation with the SVD method: the added RFI is completely linear dependent in this test set, but the background is still causing low accuracies in the SVD method.

It should be noted that some of these test sets are measuring the theoretical accuracy on unorthogonal but not completely independent RFI contamination, and this was shown in §2.6 to be the hardest case for the SVD method. When in practise the RFI does behave orthogonal or dependent, the results might be quite different. Nevertheless, it is unlikely that all RFI contaminations that are measured by different antennae at different times are always either linear dependent or orthogonal.

The presented test sets simulate a single baseline, while in a real measurement, the SVD method will exploit the correlation of RFI between different antennae. This will however also decrease the probability that all RFI is either orthogonal or linear dependent.

3.3 Automatic flagging of WSRT data

To test the various RFI flagging algorithms we have used WSRT data in the LFFE band from 138-157 MHz obtained in November/December 2007. These observations have been described and analysed by Bernardi et al. (2009a,b) to which we refer for details about the astrophysical motivation and calibration. For our analysis, however, we used the raw uncalibrated visibilities. The correlator integration time for these data was 10s. A total of 8 bands of 2.5 MHz width

were available. The central frequencies of these bands are located at 139.3, 141.5, 143.7, 145.9, 148.1, 150.3, 152.5 and 154.7 MHz. Each band was covered with 512 spectral channels. The data were Hanning tapered, yielding an effective spectral resolution of 9.8 kHz. Therefore, adjacent spectral channels are highly correlated. A total of 13 telescopes participated in the observations providing a total of 78 interferometers with baselines from 36 till 2736 meters. All four cross-correlations between the orthogonal linearly polarized feeds were used in the analysis.

We have tested the various methods on several data sets. Especially the `SumThreshold` method in combination with Gaussian smoothing shows excellent results. Figure 10 shows a typical time-frequency diagram of WSRT data around 140 MHz and the application of the `SumThreshold` method. Although the smoothed surface is slightly affected by the RFI after five iterations, as faint artefacts are visible in the smoothed surface around places where RFI used to be, the effect is so small that it does not pose a problem for the `SumThresholding` method. However, it makes the calculated false probability rate inaccurate, as the false probability calculations assume independence between the residual samples. When validating the results by visual inspection, we see far less false detections than the calculated false probability rate.

We were able to use the same parameters for any situation in the WSRT data, and therefore completely automate the flagging process. Even at baselines and frequencies with dramatic RFI contamination of up to 50%, the `SumThreshold` flagging method remains stable and accurate. Figure 11 shows for example a badly contaminated band of WSRT data, that is almost perfectly RFI flagged.

4 CONCLUSION AND DISCUSSION

In this article we have shown several approaches to deal with RFI that is left after correlation. The results show that automated flagging with the `SumThreshold` method works well for broadband and peak RFI. In all cases, the default parameters for the method work well, although parameter

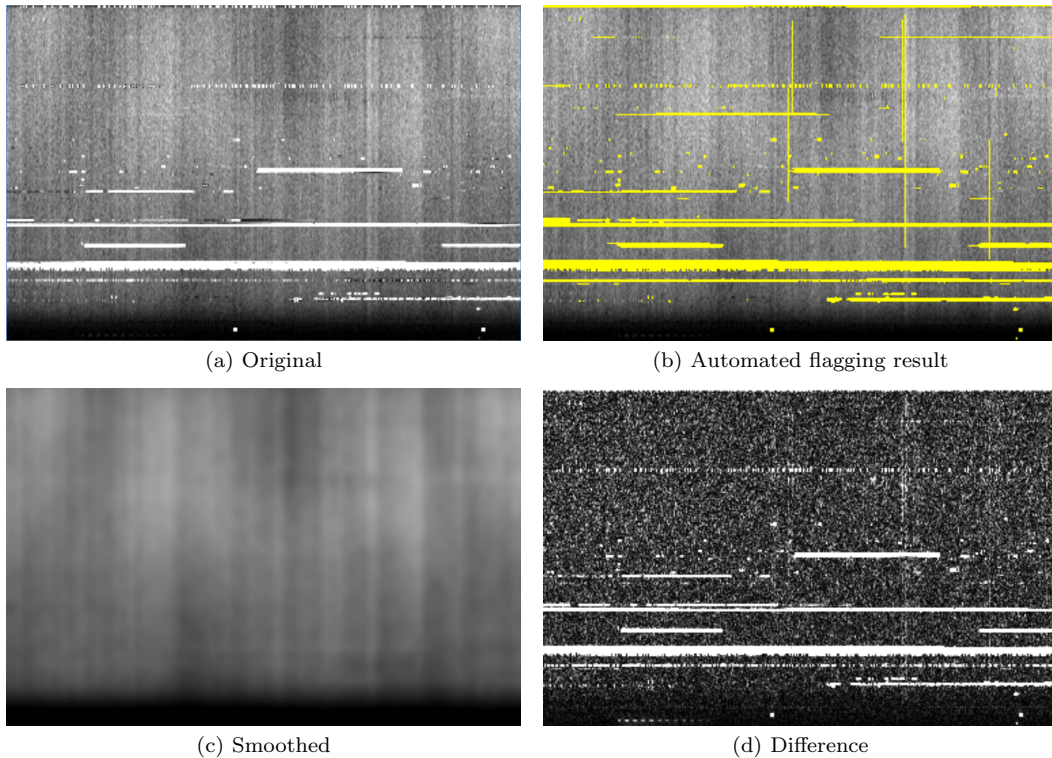


Figure 10. Time (horizontal) vs. frequency (vertical) plots of uncalibrated WSRT data, cross-correlations of antenna C vs. D. and the application of the `SumThreshold` automated flagging procedure. Panel (a) shows one hour of the amplitude of a 3C196 observation, panel (b) shows the result of the flagger, panel (c) shows the fitted surface after 5 iterations, and panel (d) shows the difference between panel (a) and panel (c).

tweaking might in some cases improve the classification. In the artificial broadband RFI situations, it detects 80% of the artificially inserted RFI with less than 0.1% error, and often approaches a 99% recognition almost without error. With this, the accuracy of this method is as accurate as can be expected from manual flagging. In the case of WSRT, the new method does not improve the dynamic range of the data compared to manual flagging, but the method saves a considerable amount of work.

New telescopes such as LOFAR and SKA mandate robust automatic procedures, as these telescopes will produce data sets that exceed current measurements in volume by orders of magnitude, losing the ability to flag or check baselines or subbands individually.

The ROC analysis shows that the `SumThreshold` method is to be preferred above the `VarThreshold` and SVD methods. The SVD method can be used in some respect to detect RFI, but is less accurate. If the SVD method were to be used to subtract the RFI in order to use RFI contaminated samples nevertheless, instead of only detecting and flagging the RFI, it can be expected that the method leaves artefacts in the data with yet unknown characteristics. For WSRT data, these artefacts look as bad as the broadband RFI itself.

All methods have been tested without assuming a data model. Subtracting the model before RFI detection might improve the classification further. Nevertheless, the detection accuracy with and without a model do not differ much.

As such, going back and forth between flagging data and creating a model is not necessary in most cases.

5 FURTHER WORK

RFI with a moderate strength that can be detected by eye was found to be of no concern for automatic flagging methods in sensitive telescopes such as WSRT. However, different kind of RFI might still pose problems. Certain weak RFI might be present in many channels for a substantial duration of the observation, such as radiation that leaks from cabins in situ. This might pose a problem for observations that require long integration to achieve its required signal-to-noise ratio, such as the LOFAR-EoR project, if the RFI is persistent in time, since these might cause systematic errors. There are some interesting ways to remove these, and one of them is the fringe-fitting RFI mitigation method described by Athreya (2009). Although this technique works at the GMRT, preliminary tests with the fringe-fitting RFI mitigation method on WSRT and LOFAR data do not show a strong presence of this type of RFI, and removing very weak RFI with a similar method requires more work. Therefore, to determine whether this type of RFI is really present, and whether it might be removable is yet to be seen.

An important next step is to consider practical issues in RFI mitigation techniques. For example, the influence of many RFI mitigation methods, post as well as pre-correlation methods, need to be simulated, since we never

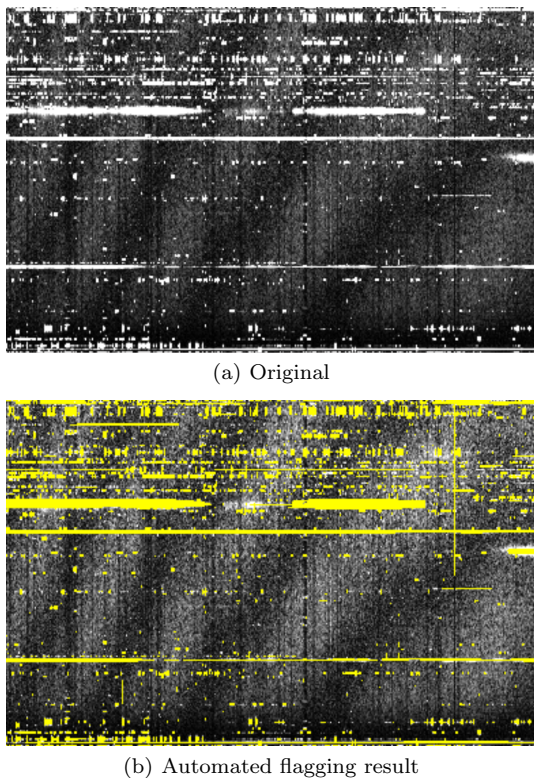


Figure 11. Time (horizontal) vs. frequency (vertical) plots of WSRT data, cross-correlations of antenna 1 vs. 2: a particular bad band at 121.3 MHz - 123.7 MHz of an observation of 3C147, showing that the method remains robust in one of the worst cases at the WSRT.

know what the image plane ought to look like. Also, which post and pre-correlation methods can be combined? Under which practical circumstances do RFI mitigation methods fail? How can we be sure that astronomical detections are not caused by RFI, or by the methods that try to reduce RFI? Answering these questions is important for establishing the reliability of new RFI mitigation methods and their regular use by astronomers.

Although, at this point, it seems to be of little concern to improve the `SumThreshold` automatic flagging method any further, it might be interesting to improve it by combining more information for detection and using fuzzy logic to decide the sample classification. An interesting example is to include phase information in the recognition, as only the amplitude information has been used so far by the threshold methods. For example, Figure 12 shows that the phase contains valuable information about a sample: in uncontaminated samples, the phase is likely to be near zero rotation, while many contaminated samples do have a phase deviating from zero. Other distinguishing information could be contained in the polarization information per sample and in the combination of different baselines.

Based on the low frequency observations with the WSRT, it can be expected that the radio environment of LOFAR is sufficiently clean for sensitive astronomical experiments. In a future paper we will fully analyse and describe the LOFAR environment and the effectiveness of the RFI strategies.

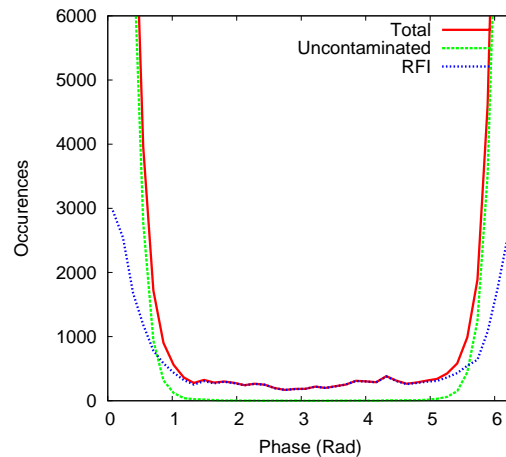


Figure 12. Typical histogram of the phase in a short baseline of a WSRT observation. The RFI was detected by using the `SumThreshold` method. The plot implies that RFI-contaminated samples have a much higher probability to have a phase deviating from zero, and the phase thus contains valuable information for RFI detection.

Finally, we like to express that the methodology of RFI flagging, or any kind of error detection, needs to change by the introduction of telescopes such as LOFAR, that generate very large amounts of data, disallowing the astronomer to browse through the data for “the baseline that was producing this artefact” or “the timestep that corresponds with these stripes in my image”. Therefore, another important next step is to be able to automatically detect errors that are caused by RFI, calibration issues, broken hardware, faulty software or any step in the complicated pipeline of a radio observatory.

ACKNOWLEDGMENTS

A.G. de Bruyn thanks Peter Fridman for many discussions on RFI mitigation.

REFERENCES

- Athreya R., 2009, AJ
- Baan W. A., Fridman P. A., Millenaar R. P., 2004, AJ, 128, 933
- Barnbaum C., Bradley R. F., 1998, AJ, 115, 2598
- Bernardi G., de Bruyn A. G., Brentjens M. A., et al., 2009a, A&A
- , 2009b, in preparation
- Boonstra A. J., Wijnholds S. J., van der Tol S., Jeffs B., 2005, Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing, 5, 869
- Briggs F., Bell J. F., Kesteven M. J., 2000, AJ
- Fridman, 2008, AJ, 35, 1810
- Fridman P. A., Baan W. A., 2001, A&A, 378, 327
- Jelić V., Zaroubi S., Labropoulos P., et al., 2008, MNRAS, 389, 1319
- Leshem A., van der Veen A.-J., 2000, ApJS, 131, 355

- Niamsuwan N., Johnson J. T., Ellingson S. W., 2005, *Radio Science*, 40
- Page E. S., 1954, *Biometrika*, 41, 100
- Pankonin V., Price R. M., 1981, *IEEE Transactions on Electromagnetic Compatibility*, EMC-23, 308
- Pen U.-L., Chang T.-C., Peterson J. B., Roy J., Gupta Y., Hirata C. M., Odegova J., Sigurdson K., 2008
- Thomas R. M., Zaroubi S., Ciardi B., et al., 2009, *MNRAS*, 58
- Thompson A. R., Gergely T. E., Vanden Bout P. A., 1991, *Physics today*, 41
- Thompson A. R., Moran J. M., Swenson G. W., 1986, *Interferometry and Synthesis in Radio Astronomy*. Wiley-Interscience
- Winkel B., Kerp J., Stanko S., 2006, *Astronomische Nachrichten*, 88, 789