A NEW APPROACH TO MITIGATION OF RADIO FREQUENCY INTERFERENCE IN INTERFEROMETRIC DATA

RAMANA ATHREYA
National Centre for Radio Astrophysics, P.O. Bag 3, Pune University Campus, Pune 411007, India; rathreya@ncra.tifr.res.in
Received 2008 November 3; accepted 2009 January 16; published 2009 April 20

ABSTRACT
Radio frequency interference (RFI) is the principal factor limiting the sensitivities of radio telescopes, particularly at frequencies below 1 GHz. I present a conceptually new approach to mitigation of RFI in interferometric data. This has been used to develop a software tool (RfiX) to remove RFI from observations using the Giant Metrewave Radio Telescope, India. However, the concept can be used to excise RFI in any interferometer. Briefly, the fringe-stopped correlator output of an interferometer baseline oscillates with the fringe-stop period in the presence of RFI. RfiX works by identifying such a pattern and subtracting it from the data. It is perhaps the only purely software technique which can salvage the true visibility value from RFI-corrupted data. It neither requires high-speed hardware nor real-time processing and works best on normal correlator output integrated for 1–10 s. It complements other mitigation schemes with its different approach and the regime it addresses. Its ability to work with data integrated over many seconds gives it an advantage while excising weak, persistent RFI unlike most other techniques which use high-speed sampling to localize RFI in time–frequency plane. RfiX is also different in that it does not require RFI-free data to identify corrupted sections. Some results from the application of RfiX are presented including an image at 240 MHz with a peak/noise ratio of 43,000, the highest till date at wavelengths greater than 1 m.

Key words: methods: data analysis – techniques: image processing – techniques: interferometric

1. INTRODUCTION
Radio-frequency interference (RFI) signals, which are generated by human activities, are among the main factors limiting the performance of radio telescopes. RFI constriicts the available frequency space, effectively increases system noise and corrupts calibration solutions. The effect is particularly strong at frequencies below 1 GHz. Ironically, weak RFIs cause more problems than strong RFIs; the latter are more easily identified and eliminated and the resulting loss of data can be compensated by longer observations. In radio interferometry (Thomson et al. 1986) each complex visibility is measured by averaging the output of a baseline over typically 1–30 s. RFI fainter than the noise in each visibility will not be detected. During the course of an observation the same baseline could contribute up to tens of thousands of visibilities corrupted by the same RFI. It is usually not possible to integrate out these correlated errors by longer observations.

RFI removal schemes may be broadly classified into two categories: in the first case the corrupted data are identified and flagged, i.e., it is completely lost. In the second case one measures the celestial signal under the RFI, thus salvaging the data. The two categories have not always been labeled differently in literature but increasingly excision and mitigation have been used interchangeably in the literature to denote the latter category.

Most RFI removal techniques use high-speed sampling at nano- to millisecondes followed by a Fourier transform to localize RFI in time and/or frequency (Bhat et al. 2005; Winkel et al. 2007; Fridman 2001; Baan et al. 2004). Statistical quantities of various degrees of sophistication (Fridman 2008), reference signals (Bradley & Barnbaum 1998), and the known properties of RFI signals in a particular context (Briggs et al. 2000; Ellingson et al. 2001) have been used to identify RFI. Generally, these techniques work for the context they were designed while none is universally applicable. Some of the limitations are: (1) high-speed sampling techniques have higher measurement noise and hence are less capable of identifying weak but constantly present RFI, (2) the requirement of clean reference areas to identify RFI, and (3) the necessity of specialized hardware. A different class of techniques has used eigen decompositions of the correlation matrices to isolate and excise RFI in interferometry (Lesham et al. 2000; Kocz 2004; Boonstra & van der Tol 2005; Pen et al. 2008. The reader is referred to the proceedings of the RFI2004 workshop in Radio Science 2005, vol 40, and URSI 2008, Chicago, for a compendium of RFI-related literature).

I describe here a new approach to mitigation of RFI from radio interferometric data which can excise contributions from quasi-constant sources of RFI. It was developed for analysing low-frequency data from the Giant Meterwave Radio Telescope (GMRT), Pune, India, and has been implemented as an offline software tool (RfiX). It complements extant techniques in approach and the applicable RFI regime. In particular, a combination of RfiX and real-time high-speed sampled voltage clipping algorithms would go a long way in minimizing RFI in interferometers. Since it works on the usual correlator output it can also be used to process archival data. However, the concept is applicable to any interferometer which uses fringe stopping.

2. FRINGE-STOP PATTERN FOR RFI MITIGATION
The RFI mitigation presented here is applicable to spatially and temporally constant RFI sources. Moving RFI cannot be processed at all and temporal variation will affect the efficacy of the algorithm.

In a standard radio interferometer the fringe pattern due to Earth’s rotation in the correlated output of a baseline is suppressed for the phase center in the sky by multiplying the two antenna data streams by a (co)-sinusoid, i.e, each baseline is fringe-stopped. The fringe-stop frequency is given by

\[ \nu_f(t) = -\omega_E U_3(t) \cos \delta(t), \]  

where \( \nu_f \) is the fringe-stopped frequency, \( \omega_E \) is the Earth’s rotation rate, \( U_3(t) \) is the third-order visibility correlation, and \( \delta(t) \) is the phase difference between the two antennas.
The observed data may be fit by well-defined analytical expressions to estimate the contribution of RFI to the observed visibility. Figure 1. Effect of fringe stopping and stationary RFI on the true visibility. The rectilinear complex-plane plots (not to be confused with the UV plane) show the temporal behavior of the true and observed values of the visibility from a single baseline. Left: schematic illustration—fringe stopping causes the RFI (dashed vector) to circulate around the true visibility (solid vector) at the fringe-stop frequency for the baseline. The resultant amplitude (dotted vectors) and phase (their position angle with respect to Real-axis) vary with time in an analytically calculable manner (see Figure 2). Right: actual data—a 5 minute sequence of GMRT 153 MHz data from C00–C04 baseline exhibiting the visibility circulation. Adjacent points along the curve are separated by 4.18 s.

where $U_j(t)$ is the instantaneous standard spatial frequency component, $\delta$ is the declination of the phase center, and $\omega_E$ is the Earth’s angular speed (Thomson et al. 1986). In general, sources in other locations will not be fringe stopped and will contribute a residual fringe to the baseline which, in fact, is used to construct the sky intensity image. The residual fringe speed increases with the radial distance of the source from the phase center. Thus, a source external to the target field will fringe faster than any interior source along the same radial vector from the phase center. In particular, every source of terrestrial RFI will fringe faster than any astronomical source and, in fact, at exactly the fringe-stopping rate.

The last point may be understood by realizing that a stationary RFI source will contribute a constant amplitude and phase to a baseline which only depends on its location relative to the two antennas. Fringe stopping will cause the RFI signal to pick up the inverse of the sinusoidal pattern. Thus, in a fringe-stopped interferometer the observed modulation of a stationary RFI signal is independent of its location on earth and only depends on the baseline length and orientation relative to the phase center (in the sky). Multiple RFI sources will all be modulated by the same fringe-stop frequency. All such sources, even if they differ in amplitude and phase, will add vectorially to contribute a single sinusoid to the baseline—this key point is worth emphasizing.

The combined effect of RFI and fringe stopping is shown schematically in Figure 1 along with actual data; the observed visibility will then be given by

$$V_{OBS} = V_{TRUE} + A e^{i(2\pi \nu_F t)/(\nu - \Phi)} + \text{Noise},$$

where all terms are complex, $A$ and $\Phi$ are the amplitude and phase of the RFI in the baseline, and $\nu_F$ is the fringe-stop frequency (Equation (1)). Provided the RFI remains constant over a good fraction of the fringe period, the observed data may be fitted for fit-average values of $A$, $\Phi$, and $V_{TRUE}$. The observed visibility amplitude and phase obtained in a closed/analytical form from Equation (2) are plotted in Figure 2. The salvaged/RFI-free visibility will then be

$$V_{RFIX}^j = V_{OBS}^j - A e^{i(2\pi \nu_F t)/(\nu - \Phi)} = V_{TRUE} + \text{Noise}^j,$$

where $j$ is an index identifying a particular data point in the subset over which the RFI was presumed to be constant and a model was fit. Note that $V_{TRUE}$, $A$, and $\Phi$ do not carry the index $j$ and are average fit values for the entire subset. The average $V_{TRUE}$ obtained from the fit is not carried to the next stage. Instead, the model RFI contribution ($A$ and $\Phi$) for the fitted subset is subtracted from the $j$th observed visibility (within the subset) to obtain the $j$th modified visibility, i.e. the modified values continue to carry the original unsmoothed noise component and hence adjacent values remain independent.

This technique does not affect the all-important closure relationships (Thomson et al. 1986). The mean value of $V_{OBS}$ (i.e., the center of the circle) is a good estimator of $V_{TRUE}$ and can be used to quantify deviations from closure, if any. Alternatively, one can consider the case of an inaccurate estimate of the RFI amplitude and phase. Since the modified visibilities are obtained by subtracting the RFI circle, an inaccurate RFI model is equivalent to the original RFI corruption being replaced.

Figure 2. Observed visibility in the presence of RFI for the model shown in Figure 1. The amplitude and phase, derived from Equation (2) are plotted against time. The observed data may be fit by well-defined analytical expressions to estimate the contribution of RFI to the observed visibility.
by a new RFI corruption, whose amplitude should be close to zero under ideal conditions but in any case smaller than the original RFI amplitude. In practice, a simple before-and-after rms calculation and $\langle V_{RFIX} \rangle - \langle V_{OBS} \rangle$ can be used to quantify closure deviations, if any, and provide the choice of either using the model fit or retaining the original data—at worst the processed data are no worse than the raw data.

The relative strengths of the RFI and true visibility is not relevant either since the algorithm is being applied on the rectilinear complex plane.

3. IMPLEMENTATION AT GMRT—RFIX PACKAGE

The RFI environment at GMRT can be quite intimidating, especially at frequencies below 400 MHz. While the strongest source in a field is typically only a few Jy, the RFI can occasionally reach several hundred Jy in some channels in the 150 and 240 MHz bands.

However broadband RFI of up to several tens of Jy causes even more damage: currently available tools, including the Astronomical Image Processing System (AIPS) and FLAGIT (a locally developed tool) would not even recognize the RFI contamination—there is no "clean" reference area for identification by contrast.

This may be seen in Figure 3 which plots the visibilities in grayscale on a two-dimensional time–frequency plot. In the absence of RFI all the visibilities in a small subsection of the time–frequency plane should more or less represent the same Fourier component for a continuum source and hence the intensity pattern should be devoid of features. A very narrow RFI line would show up as a ripple along the time axis in a single frequency channel. A coherent broadband RFI source would introduce a ripple across many channels and the direction of the pattern would depend on the location of the RFI with respect to the baseline. The ripples seen in the left-hand panel show that the visibility data shown has RFI in every single time–frequency pixel. The different slopes seen in the pattern point to the presence of multiple sources of RFI which have emission broad enough to cover many channels (of up to several MHz) but not broad enough to cover the whole band, which complicates data analysis as will be described shortly.

Broadband RFI causes calibration errors which severely limit the dynamic range (peak-to-rms ratio) to less than a few thousand at these frequencies, even for observations of several full synthesis sessions (each of 8–10 hr). Even though the GMRT is currently the most sensitive low-frequency instrument, the sensitivity achieved at 150 and 240 MHz is more than a factor of 10 poorer than expected for long integrations.

There are two ways of removing RFI from the data. A Fourier transform of the time–frequency visibility data would efficiently localize the RFI signal in frequency-lag space (time lag ≡ path difference from the RFI source to the two antennas). A Fourier transform requires many dozens or even hundreds of data points within the period of constancy of the RFI source for efficient inversion. This requires higher (than usual) speed sampling and storage for consequent larger data volume. Furthermore, abrupt changes along either the time- or the frequency-axis, like those seen in Figure 3, will make the frequency-lag plane noisier, reducing the ability to detect RFI.

The alternative chosen for RfiX was motivated by the realization that while RFI ripples along the frequency axis could...
Figure 5. Efficacy of RfiX in cleaning the bandpass. The panels show the average bandpass for three 5 minute scans spanning a 4 hr period of the strong calibrator 3C446 at 153 MHz using the GMRT. The solid histograms are the pre-RfiX (raw) data and the dots are the post-RfiX visibilities. The source flux density is 21.5 Jy. The strongest RFI (middle scan) is almost 200 Jy. The bottom scan has at least 25 Jy of RFI in every channel and 4 times as much in some. In spite of the very strong RFI the post-RfiX bandpass is more or less flat in amplitude and linear in phase and the amplitude level and phase gradient is similar in all three scans.

Figure 4 illustrates the application of the algorithm to actual data. The plot shows a 5 minute data sequence from a single channel (same as in Figure 1). The outer annulus of points is the raw data. The compact cluster at the center is the processed/RFI excised data. The variability of RFI within the 5 minute scan was surmounted by separate fits to subsets of the data of length 54 s (=1 fringe period). Nevertheless, this variability resulted in imperfect RFI mitigation, which is reflected in a larger scatter in the modified visibilities. However, RfiX succeeded in reducing the scatter by a factor of 5, from the radius of the outer annulus (RFI amplitude) to the post-processing scatter represented by the small inner circle. This much smaller scatter facilitates further cleaning-up through sigma-clipping.

The fully automated implementation (1) checks if the baseline fringe-stop period is within the appropriate range (see under discussion) and if not, either skips it or flags it as specified, (2) compares the scan rms to a fiducial value to determine if the particular channel has significant quantum of RFI and therefore worth processing, (3) identifies subsections with constant RFI amplitude, and (4) robustly fits and subtracts the RFI fringe.

Figure 5 shows the results of RfiX on the averaged bandpass of amplitude and phase of three different 5 minute scans (separated by 4 hr). The data was obtained with the GMRT at 153 MHz on the calibrator source 3C446 (21.5 Jy). The upper panel shows RFI nudging 200 Jy in some channels. Traditional methods of offline RFI clipping (in AIPS, for instance) would completely fail in
Figure 6. GMRT image of 3C286 at 240 MHz using RfiX processed data. The left panels show the celestial sources (top) and the rms noise (bottom), respectively. They were made using just 2 hours of data with the GMRT. The field covers the half-power extent of the primary beam. The peak flux density is 28 Jy (of 3C286) and the noise is 2 mJy beam$^{-1}$ near 3C286 and 0.65 mJy beam$^{-1}$ at the periphery (peak/noise $\equiv$ centre: 14,000, periphery: 43,000). This is perhaps the highest dynamic range image made at wavelength greater than 1 m (frequency $<$ 300 MHz). The plot illustrates, rather starkly, the problem of finding isolated point sources for calibration at low frequencies! The top-row middle panel shows a more detailed plot of the immediate area around 3C286. As may be expected artifacts are most numerous here, but even so the image is quite clean with only few artifacts in excess of 5$\sigma$, with the highest at 8$\sigma$. In the peripheral regions (top-right) there are no artifacts in excess of 4$\sigma$. Histograms of the noise pixels (bottom row) near the contour locations show well-behaved residuals. This is perhaps the highest dynamic-range image achieved till date at frequencies below 300 MHz (wavelength greater than 1 m).

4. DISCUSSION

The most important feature of RfiX is its ability to salvage RFI-corrupted data—it not only identifies RFI but also excises it to obtain uncorrupted data. This is perhaps the only RFI mitigation tool which is entirely software based. It can be applied offline on the normal correlator data typically output every 0.1–30 s. This much longer integration compared to high-speed sampling techniques makes RfiX sensitive to fainter RFI. It complements high-speed techniques which are best for clipping short but intense bursts of RFI but cannot target weak but constantly present RFI. Also, RfiX does not require a clean reference area to identify RFI.

The technique is applicable to any interferometer which uses fringe stopping. It can also be applied at any frequency provided
only that the correlator output matches the Nyquist rate for the fringe-stop frequency of the particular baseline. For instance, the GMRT correlator outputs data every 2–16 s in the standard mode. A 2 s sampling rate limits the application of RfiX to baselines with $U_b < 3–12 \, \text{km}$ (declination range $0^\circ–75^\circ$). At the GMRT the RFI is most severe at $v_{\text{obs}} < 400 \, \text{MHz}$ and at these frequencies the upper limit corresponds to baselines of several kilometers. We do not usually find correlated RFI on such long baselines and so the sampling rate does not pose a constraint in practical terms. In any case, the present correlator at GMRT can be configured to provide data every 0.128 s and, if necessary, the longer baselines and/or higher frequency observations can also be processed for RFI using RfiX. In general, a visibility sampling interval of 0.1 s, which can be very easily achieved using present-day hardware, will be sufficient for excising RFI (using RfiX) from baselines up to a few km in observations up to a few GHz. Of course, shorter visibility integrations in the correlator can relax both these limits.

RfiX will not work effectively with interferometric observations near the Celestial Poles where the fringe rate is very slow. Even in regions away from the Poles, baselines with small $g$ $U$-component (see Equation (1)) have very long fringe periods and so cannot be satisfactorily processed. RfiX works best for RFI which manifests and remains constant over the fringe period of a baseline. Fitting over a full fringe period is very safe for conserving closure. Fitting intervals shorter than a third of the fringe period could lead to incorrect estimation of RFI amplitude and hence closure-violating baseline errors. At the GMRT the RFI amplitude can be quasi-constant for up to 150–200 s which limits the application of RfiX to baselines with fringe periods $\lesssim 500 \, \text{s}$, corresponding to baseline component $U_b \sim 25–100\, \text{km}$ (declination range $0^\circ–75^\circ$). Of course, we do see instances of RFI varying on much shorter timescales and RfiX will fail in those cases. Depending on the goal one can either flag all shorter baselines or retain them without RfiX processing. Therefore, this algorithm will not improve the quality of image structures larger than $0.5^\prime–1.5^\prime$. This is not a serious limitation since the GMRT primary beam FWHM is not much larger. Note that the constraint comes from the timescale of RFI variability and not RfiX. Pen et al. (2008) also have a similar limitation for excising RFI in baselines with long fringe periods.

Briefly, the limiting fringe stopping frequency and the corresponding baseline for which RfiX is effective are related through the rather simple Equation (1). It must be noted that only the baseline component $U_b$ is involved in the equation and not $V_b$. The lower baseline limit is defined by the timescale of RFI variability while the upper limit is set by the visibility integration period of the correlator. It is often claimed that interferometric fringe stopping in itself washes out RFI, but this is not entirely appropriate for low-frequency arrays. If the system temperature ($T_{\text{sys}}$) dominates over the RFI amplitude ($A_{\text{RFI}}$), the noise in a single visibility of integration period $\tau$ will be $T_{\text{sys}}/\sqrt{\tau}$. When RFI dominates, the effective noise will be $A_{\text{RFI}}(v, \tau)$, where $v$ is the fringe-stop frequency for the particular baseline. Thus, RFI suppression due to fringe stopping is most effective for short baselines for which $v \tau \gg 1$. In the context of GMRT observations at 153 MHz (longest baseline 27 km) the shortest fringe period is of the order 1 s (see Equation (1)) compared to the recommended integration time of 2–4 s. Therefore RFI suppression due to fringe stopping would only have a marginal impact. Even at 1 GHz RFI would be suppressed only in baselines formed by the outer antennas; but correlated RFI is anyway mostly limited to baselines shorter than a few km. RFI suppression due to fringe stopping would be most effective in arrays operating at $v \gg 1 \, \text{GHz}$ as even the shorter (RFI-prone) baselines will have sufficiently short fringe period to achieve RFI suppression.

A detailed description of the implementation of RfiX for GMRT data is presented in R. M. Athreya (2009, in preparation) and may be of use to astronomers intending to implement the tool at other facilities.

RfiX is somewhat related to the RFI-to-North-Pole scheme tried by Perley et al. (2005). Briefly, a stationary RFI and a source at the North Pole are both stationary with respect to the interferometer. Thus, one can attempt to map RFI to pseudo sources at the North Pole and subtract them from visibilities. This will work when the RFI contributes the same antenna temperature to all the antennas, just like a celestial source. However, in general the intensity of an RFI source varies from antenna to antenna due to $1/R^2$ fall-off and shielding by man-made or geographical structures. RfiX surmounts this problem by processing each baseline separately allowing for variable intensity in each.

In summary, RfiX is a powerful tool based on a simple concept for mitigating RFI in interferometric data. Indeed, its strength lies in its simplicity which provides a clear understanding of its domain of applicability. It not only identifies RFI but also salvages RFI-affected visibilities. It is purely a software approach that requires neither additional hardware nor real-time analysis nor complicated mathematical operations and can even be applied to archival data. It complements other extant excision techniques in approach and domain of applicability. It will work for any interferometer which uses fringe-stopping.

I thank Pramesh Rao and Dipanjan Mitra for useful discussions during the protracted incubation of this algorithm. K. S. Jeeva contributed data display and FITS IO tools. Rahul Basu’s work helped to efficiently flag residual RFI from 3C286 data. I thank the GMRT staff for helping with observations presented in this work. The comments of an anonymous referee considerably improved the manuscript.

Facilities: GMRT

REFERENCES

Baan, W. A., Fridman, P. A., & Millenaar, R. P. 2004, AJ, 128, 933
Bhat, N. D. R., Cordes, J. M., Chatterjee, S., & Lazio, T. J. W. 2005, Radio Science, 40 (arXiv:astro-ph/0502149v1)
Boonstra, A. J., & van der Tol, S. 2005, in Proc. RFI2004 Workshop, Radio Science, 40
Bradley, R., & Barnbaum, C. 1998, AJ, 116, 2598
Briggs, F. H., Bell, J. F., & Kesteven, M. J. 2000, AJ, 120, 3351
Ellingson, S. W., Bunton, J. D., & Bell, J. F. 2001, ApJS, 135, 8793
Fridman, P. A. 2001, A&A, 369, 369
Fridman, P. A. 2008, AJ, 135, 1810
Kocz, J. 2004, Honours thesis, Australian National Univ.
Kotch, J. 2008, in AIP Conf. Ser. 1035, The Evolution of Galaxies Through the Neutral Hydrogen Window, ed. R. Minchin & E. Momjian (Melville, NY: AIP), 362
Perley, R. A., Cornwell, T. J., Bhatnagar, S., & Golap, K. 2005, in Proc. RFI2004 Workshop, Radio Science, 40
Thomson, A. R., Moran, J. M., & Swenson, G. W. 1986, Interferometry and Synthesis in Radio Astronomy (New York: Wiley InterScience)
Winkel, B., Kerp, J., & Stanko, S. 2007, Astron. Nachr., 328, 68