

## RFI Mitigation Techniques in Radio Astronomy: Postprint

**Authors:** Hailong Zhang, Zhang Yazhou, Wang Jie, Ye Xinchuan, Wang Wanqiong, Li Jia, Zhang Meng, Du Xu

**Date:** 2022-04-14T16:43:20+00:00

### Abstract

A detailed analysis of RFI mitigation strategies employed by radio observatories both domestically and internationally is presented, specifically addressing the problem of radio frequency interference (RFI) in radio astronomy observations. Based on the RFI issues encountered during actual observations at various astronomical observatories, prevention strategies and mitigation methods for RFI are investigated from the perspectives of proactive prevention stage, pre-correlation stage, post-correlation stage, machine learning, and deep learning. The methods that can be adopted during the proactive prevention stage are analyzed in detail, along with adaptive filtering and spatial filtering methods for the pre-correlation stage; and VarThreshold, SumThreshold, and singular value decomposition methods for the post-correlation stage. The application of related techniques and methods based on machine learning—including principal component analysis, support vector machines, fully convolutional neural networks, convolutional neural networks, and U-Net—in RFI signal processing is also discussed.

### Full Text

#### Application of RFI Mitigation Technology in Radio Astronomy

**Zhang Hailong<sup>1,2,3,4</sup>, Zhang Yazhou<sup>1,2</sup>, Wang Jie<sup>1,4</sup>, Ye Xinchuan<sup>1,4</sup>, Wang Wanqiong<sup>1</sup>, Li Jia<sup>1</sup>, Zhang Meng<sup>1,2</sup>, Du Xu<sup>1,2</sup>**

<sup>1</sup> Xinjiang Astronomical Observatory, Chinese Academy of Sciences, Urumqi, Xinjiang 830011, China

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup> Key Laboratory of Radio Astronomy, Chinese Academy of Sciences, Nanjing,

Jiangsu 210008, China

<sup>4</sup> National Astronomical Data Center, Beijing 100101, China

## Abstract

This paper provides a comprehensive analysis of radio frequency interference (RFI) mitigation strategies employed at radio observatories worldwide, addressing the persistent challenge of RFI in radio astronomical observations. We systematically examine prevention and mitigation approaches across three distinct stages: active prevention, pre-correlation processing, and post-correlation processing, as well as emerging techniques based on machine learning and deep learning. For the active prevention stage, we detail practical shielding methods. In the pre-correlation stage, we analyze adaptive filtering and spatial filtering techniques. For post-correlation processing, we evaluate threshold-based methods including VarThreshold, SumThreshold, and singular value decomposition. Furthermore, we explore the application of machine learning techniques such as principal component analysis, support vector machines, fully convolutional neural networks, convolutional neural networks, and U-Net architectures for RFI signal processing and identification.

**Keywords:** RFI; filtering; thresholding; machine learning

---

## 1. Introduction

In radio astronomy, radio frequency interference (RFI) is defined as any undesirable signal that may affect astronomical observations. Throughout the development of radio astronomy, RFI suppression has remained a focal point of research for astronomers. The rapid advancement of information technology and the expansion of human activities have led to deteriorating electromagnetic environments around observatories, significantly impacting normal telescope operations.

RFI originates from diverse sources, both external and internal. External sources primarily consist of equipment outside the observatory site, such as artificial satellites (including BeiDou and GPS navigation satellites), aircraft, base station signals near the site, and television broadcast signals. Internal sources arise from electronic devices used within the observatory, including computers, video surveillance systems, network switching equipment, and wireless input devices. Astronomical signals are typically broadband and vary smoothly over time, whereas RFI exhibits high amplitude intensity in both time and frequency domains, with most RFI showing clear distinctions from astronomical signals. Unlike thermal noise, RFI generally consists of interference generated by communication systems, artificial radars, or electronic devices, and often possesses complex temporal and frequency structures. Common communication signals have power levels many orders of magnitude higher than astronomical signals

and vary with time, making it impossible to reduce their intensity through long-term integration (accumulation), thereby severely compromising data quality.

Astronomers have proposed numerous RFI mitigation methods, yet these approaches remain constrained by several factors. First, the rapid development of wireless technology and the expansion of human activity have led to year-over-year increases in anthropogenic interference from both ground-based and space-based sources. Second, advances in manufacturing and information technology have driven radio telescopes toward larger apertures and massive arrays, while the deployment of multi-beam or phased array feed (PAF) receiving systems has significantly enhanced observation sensitivity and data collection capabilities, simultaneously complicating RFI mitigation.

Different RFI mitigation strategies must be implemented at various stages of radio astronomical observation and data processing. This paper categorizes RFI mitigation strategies into three phases: active prevention, pre-correlation processing, and post-correlation processing, each employing distinct methods tailored to specific challenges. Effective shielding during the active prevention stage can prevent most RFI from communication base stations and television broadcasts from entering the system. In the pre-correlation stage, methods based on reference antennas and spatial filtering can address specific RFI sources. During post-correlation processing, threshold-based methods can effectively handle RFI with amplitudes far exceeding astronomical signals, while machine learning techniques enable automated RFI flagging in data. The following sections provide detailed analysis of RFI mitigation strategies applicable at each stage.

---

## 2. Active Prevention Stage

### 2.1 Radio Quiet Zone

Radio telescopes observe extremely faint signals from distant celestial objects, necessitating stringent requirements for electromagnetic interference in their vicinity. Selecting geographically favorable sites and establishing radio quiet protection zones represent crucial first steps in RFI mitigation, fundamentally eliminating most interference at its source.

The Five-hundred-meter Aperture Spherical radio Telescope (FAST), the world's largest single-dish radio telescope, is located in Dawodang, Kedu Town, Pingtang County, Qiannan Buyi and Miao Autonomous Prefecture, Guizhou Province. Centered on the FAST site, a 30 km radius radio quiet zone has been established, divided into three regions with different protection requirements. The core protection zone extends 5 km from the site center, the intermediate zone covers a 5-10 km annulus, and the remote zone spans 10-30 km, as illustrated in Fig. 1. This zoning effectively shields the telescope from the vast majority of external interference sources.

The Xinjiang Qitai 110 m Radio Telescope (QTT), situated in Qitai County, Changji Prefecture, Xinjiang, similarly employs a 30 km radio quiet zone tailored to local topography and conditions. This zone comprises three regions: a core zone, a restricted zone, and a coordination zone. The core zone forms a 2.5 km  $\times$  4 km rectangular area, while the restricted zone covers a 10 km  $\times$  15 km rectangular region, as shown in Fig. 2.

## 2.2 Reserved Frequency Bands

The International Telecommunication Union (ITU) coordinates radio spectrum allocation between 8.3 MHz and 300 GHz, dividing it into several bands designated for specific services. Among these allocated bands, approximately 70 to 80 correspond to radio astronomy (the exact number varies by region and local regulations). Radio astronomy bands fall into two categories: dedicated bands (or those shared with passive services) where radio emissions are strictly prohibited, and bands shared with active services where only partial protection can be enforced. Band selection for radio astronomy typically correlates with scientific objectives. For example, the hydrogen line (emitted due to changes in neutral hydrogen atom states) lies near 1420 MHz, prompting the reservation of the 1400–1427 MHz band for radio astronomy.

Even within dedicated radio astronomy bands, interference may persist due to harmonics and power leakage from active components. For such cases, the ITU defines maximum acceptable interference in its ITU-R RA.769 recommendation as interference that introduces no more than 10% error in measured power—a threshold widely accepted by astronomers. This report also provides a recommended threshold list for typical astronomical bands using common telescope and observation parameters.

## 2.3 Limitations of Active Prevention Strategies

The most direct approach in active RFI prevention involves selecting remote observatory locations far from interference sources. Natural terrain features such as mountains can effectively block external interference, though multipath propagation through reflection and diffraction increases mitigation difficulty. Some observatories have found that planting coniferous trees like pines around telescope sites effectively suppresses RFI, as moisture in the needles absorbs signals above 1 GHz. Partial observatories are constructed at high altitudes to reduce atmospheric effects and distance from human activity while minimizing multipath propagation. For internal interference, common practice involves shielding interfering electronic equipment (computers, microwave and RF components) with conductive foil or concentrating equipment in shielded rooms to contain interference within enclosed spaces without affecting device performance.

While active prevention represents the most effective interference mitigation method and serves as the first line of defense, significant limitations remain. First, it can only protect a small portion of the electromagnetic spectrum. Sec-

ond, electromagnetic shielding materials cannot completely block interference for highly sensitive equipment. Third, interference from satellite communications and aircraft cannot be mitigated through active prevention methods due to relative position constraints.

### 3. Pre-correlation Stage

We define the pre-correlation stage as the phase during which raw observation data is processed before being written to disk for scientific analysis. For single-dish observations, adaptive filtering using reference antennas can remove specific interference, while array antennas can employ spatial filtering methods to suppress RFI.

#### 3.1 Adaptive Filtering

Adaptive filtering algorithms were first introduced to radio astronomy by Barnbaum and Bradley in 1998 to address RFI problems. In 2005, Kesteven et al. conducted field tests demonstrating that adaptive filtering could substantially improve pulsar observations in RFI-contaminated environments.

The signal received by the reference antenna can be expressed as:

$$r(t) = n_r(t) + rfi_r(t) + s_r(t)$$

where  $r(t)$  represents the signal received by the reference antenna,  $n_r(t)$  denotes the system noise of the reference antenna,  $rfi_r(t)$  indicates RFI received by the reference antenna, and  $s_r(t)$  represents astronomical signals received by the reference antenna. Astronomical signals are received through the reference antenna's sidelobes and are sufficiently weak to be negligible. Similarly, the signal received by the main telescope can be expressed as:

$$m(t) = n_m(t) + rfi_m(t) + s_m(t)$$

The system noise received by the main telescope and reference antenna are uncorrelated, while the received RFI exhibits correlation. To eliminate RFI in the main antenna, the reference antenna is aligned with the RFI source, and continuously optimizing the reference antenna's pointing and polarization direction can improve the RFI signal-to-noise ratio in the reference channel.

The specific implementation steps of adaptive filtering are:

1. In the adaptive correlation loop, determine the complex gain coefficient  $g$  (which is continuously iteratively optimized), multiply it with the signal received by the reference antenna  $r(t)$ , and iteratively adjust  $g$  to approximate the RFI signal in the main antenna, thereby maximizing elimination of RFI from a specific direction.

2. When  $g$  reaches its optimal value,  $g \cdot r(t)$  closely approximates  $r f i_m(t)$ , and the difference between the astronomical signal and the reference signal yields the astronomical signal with specific interference removed.
3. Fig. 3 shows the basic design diagram of an adaptive filter. When the correlation coefficient between the original reference signal and the filter output becomes zero, optimal gain is achieved.

The correlation term is calculated as shown in equation (3):

$$\langle m(t) \cdot r^*(t) \rangle = \langle (n_m(t) + r f i_m(t) + s_m(t)) \cdot (n_r(t) + r f i_r(t) + s_r(t))^* \rangle$$

When  $\langle m(t) \cdot r^*(t) \rangle = 0$ , RFI in the main antenna can be eliminated. Substituting the components received by each antenna:

$$\langle m(t) \cdot r^*(t) \rangle = \langle n_m(t) \cdot n_r^*(t) \rangle + \langle n_m(t) \cdot r f i_r^*(t) \rangle + \langle n_m(t) \cdot s_r^*(t) \rangle + \langle r f i_m(t) \cdot n_r^*(t) \rangle + \langle r f i_m(t) \cdot r f i_r^*(t) \rangle + \langle r f i_m(t) \cdot s_r^*(t) \rangle$$

Since the system noise, astronomical signals, and other signals from the two antennas are all uncorrelated with RFI, and the astronomical signal in the reference antenna is extremely weak and can be neglected ( $s_r(t) \approx 0$ ), the expression simplifies to:

$$\langle m(t) \cdot r^*(t) \rangle = \langle r f i_m(t) \cdot r f i_r^*(t) \rangle - g \cdot \langle r f i_r(t) \cdot r f i_r^*(t) \rangle$$

Letting  $g = \frac{\langle r f i_m(t) \cdot r f i_r^*(t) \rangle}{\langle r f i_r(t) \cdot r f i_r^*(t) \rangle}$ , the residual RFI in the output signal is:

$$\epsilon = \frac{1}{1 + INR}$$

where INR (Interference-to-Noise Ratio) is the interference-to-noise ratio in the reference antenna. It is evident that larger INR values yield smaller  $\epsilon$ , indicating better adaptive filter performance. The residual RFI in the output signal is inversely proportional to INR; thus, improving the RFI signal-to-noise ratio in the reference antenna enhances RFI suppression capability in the main antenna.

Several factors affect adaptive filtering performance:

### (1) Mixed RFI Sources

Adaptive filtering performs well in eliminating single-source RFI, but its performance degrades significantly in complex RFI environments. For mixed RFI, when two interfering signals occupy the same frequency domain, the interference with larger amplitude will be suppressed.

### (2) Multipath Propagation Effects

Multipath propagation refers to the phenomenon where multiple copies of the

same signal arrive at the receiving antenna via different paths, typically caused by large terrestrial objects (mountains, buildings), large water bodies (lakes, lagoons), and ionospheric reflections. Fig. 4 illustrates a multipath interference propagation scenario. Because multiple copies arriving at the telescope and reference antenna have different time delays, if the delay is sufficiently large, the added signals may be considered completely uncorrelated by the telescope. In this case, multipath propagation is equivalent to multiple RFI sources, making it difficult to obtain a general expression for multipath propagation in filters. The best measure to prevent multipath propagation is to place the telescope and reference antenna as close as possible to minimize propagation distance differences, and to increase FFT points to improve frequency resolution or reduce system bandwidth while keeping FFT points constant.

### (3) Violation of Linear Time-Invariant (LTI) Conditions

The interference propagation path is assumed to satisfy LTI conditions at least during filter convergence time. If LTI conditions are not met, the RFI path exhibits nonlinear or time-varying propagation characteristics somewhere. However, no natural medium exhibits sufficiently strong nonlinearity or rapidly changing propagation characteristics to affect filter performance, particularly considering that final algorithm convergence occurs on millisecond timescales. When the telescope tracks celestial objects, the sidelobe capturing RFI moves with the telescope, and the measured RFI path characteristics change gradually over time. Adaptive filtering methods must possess sufficient computational capability to avoid being significantly affected by the telescope's slow movement.

Some analog components in the receiver front-end, such as amplifiers and frequency mixers, exhibit signal saturation. When a component reaches saturation, the amplitude of the input signal is clipped, generating harmonic distortion that affects the system's LTI conditions. This distortion is treated as an additional corrupted signal with frequencies several times the original input signal's center frequency. Moreover, harmonic signals are often aliased into the baseband after digitization, corrupting astronomical data. Harmonic distortion from RFI in the main channel can severely degrade filter performance, as the additional harmonic signals do not appear in the reference channel, making such RFI more difficult to suppress.

## 3.2 Spatial Filtering

Spatial filtering methods primarily address RFI mitigation in multi-channel radio astronomy observations and are applicable to interferometric radio telescope arrays such as the Westerbork Synthesis Radio Telescope (WSRT) in the Netherlands, the Very Large Array (VLA) in the United States, or single-dish telescopes equipped with phased array receiving systems. Fig. 5 illustrates the application of spatial filtering techniques in antenna array systems at different processing stages.

The signals received by radio telescopes consist of three components: radio

source signals (the intended observation targets), system noise, and RFI. System noise comprises cosmic background noise, atmospheric noise, receiver noise, and other noise components. According to the central limit theorem, system noise is temporally independent and Gaussian-distributed, short-term stationary over brief intervals, with all radio sources assumed to be statistically independent of one another.

RFI characteristics in radio astronomy are highly diverse. We select three distinct signal properties for RFI modeling: narrowband Gaussian stationary, second-order aperiodic, and cyclostationary signals.

Consider an antenna array with  $N$  antennas, where  $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_N(t)]^T$  represents the array outputs, with  $x_i(t)$  denoting the output of the  $i$ -th antenna at time  $t$ . Then  $\mathbf{x}(t)$  can be expressed as:

$$\mathbf{x}(t) = \mathbf{A}_s \mathbf{s}(t) + \mathbf{A}_i \mathbf{i}(t) + \mathbf{n}(t)$$

where: 1.  $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_K(t)]^T$  represents the signal characteristics of  $K$  radio sources at time  $t$ , with dimension  $K \times 1$ . 2.  $\mathbf{A}_s = [\mathbf{a}_s(\theta_1), \mathbf{a}_s(\theta_2), \dots, \mathbf{a}_s(\theta_K)]$  is the spatial signature matrix of the radio sources, with dimension  $N \times K$ , where  $\mathbf{a}_s(\theta_k) = [a_{s,1}(\theta_k), a_{s,2}(\theta_k), \dots, a_{s,N}(\theta_k)]^T$  is the spatial signature vector of the  $k$ -th radio source. 3.  $\mathbf{A}_i$  and  $\mathbf{i}(t)$  represent the spatial signature matrix of RFI and the RFI signal characteristics at time  $t$ , respectively. 4.  $\mathbf{n}(t) = [n_1(t), n_2(t), \dots, n_N(t)]^T$  is the system noise vector at time  $t$ , with dimension  $N \times 1$ .

The temporal correlation of radio sources is described through the covariance matrix:

$$\mathbf{R}_{ss} = E\{\mathbf{s}(t)\mathbf{s}^H(t)\}$$

where  $\mathbf{R}_{ss}$  describes the correlation of radio source signals over time range  $\tau$  at time  $t$ , and  $^H$  denotes conjugate transpose. Assuming RFI, radio sources, and system noise are mutually independent and uncorrelated:

$$E\{\mathbf{s}(t)\mathbf{i}^H(t)\} = E\{\mathbf{s}(t)\mathbf{n}^H(t)\} = E\{\mathbf{i}(t)\mathbf{n}^H(t)\} = \mathbf{0}$$

From a linear algebra perspective, the covariance matrix of phased array radio telescope data can be viewed as a linear transformation matrix that generates a data vector space composed of RFI subspace, radio source subspace, and system noise subspace. Fig. 5 shows an example of a two-dimensional data vector space in a noise-free scenario, where red and black vectors represent signals associated with radio sources and RFI sources, respectively.

Using orthogonal projection techniques to reduce interference, the data vector space is projected onto a subspace orthogonal to the RFI subspace. The projected RFI subspace becomes completely zero, yielding data containing only

radio sources. As clearly shown in Fig. 6, the recovered radio source energy after projection depends on the angle between the two vectors:

$$P_{\text{recovered}} = P_{\text{source}} \cdot \sin^2(\theta)$$

where  $P_{\text{source}}$  represents the power of the radio source and  $\theta$  is the angle between the RFI vector and the radio source vector. Larger  $\theta$  values result in greater recovered radio source power. However, the angle between the radio source subspace and interference subspace is difficult to measure directly and requires indirect calculation via projection matrices.

Let the projection matrix be  $\mathbf{P}$ , whose eigenvalues are either 0 or 1. Eigenvectors with eigenvalue 0 generate the projection null space, while eigenvectors with eigenvalue 1 generate the range subspace. The rank of the projection matrix equals the dimension of its range subspace. If  $\mathbf{U}$  is the basis of the projection range subspace and  $\mathbf{V}$  is the basis of the projection null space, then:

The projection matrix generated by basis matrix  $\mathbf{U}$  is defined as:

$$\mathbf{P}_U = \mathbf{U}(\mathbf{U}^H \mathbf{U})^{-1} \mathbf{U}^H$$

The orthogonal projection whose range is orthogonal to the subspace of  $\mathbf{V}$  is:

$$\mathbf{P}_V^\perp = \mathbf{I} - \mathbf{V}(\mathbf{V}^H \mathbf{V})^{-1} \mathbf{V}^H$$

In multi-interference scenarios, the interference subspace is multidimensional, generated by the set of independent RFI vectors stored in matrix  $\mathbf{A}_i$ . Therefore, the projection matrix can be expressed as:

$$\mathbf{P}_{RFI}^\perp = \mathbf{I} - \mathbf{A}_i(\mathbf{A}_i^H \mathbf{A}_i)^{-1} \mathbf{A}_i^H$$

In the pre-correlation stage, the available data is the antenna array output vector  $\mathbf{x}(t)$ . Projecting this onto the orthogonal space of the RFI subspace yields the corrected data vector  $\mathbf{x}_{\text{cleaned}}(t)$ :

$$\mathbf{x}_{\text{cleaned}}(t) = \mathbf{P}_{RFI}^\perp \mathbf{x}(t)$$

As evident from the above, spatial filtering requires prior knowledge of interference sources to determine the projection matrix. In complex interference environments, accurately determining the specific locations of interference sources is extremely difficult.

## 4. Post-correlation Stage

Since RFI signals are typically several orders of magnitude stronger than astronomical signals, thresholding methods provide effective RFI suppression. Threshold levels are usually determined based on statistical measures such as the mean or root-mean-square of a data segment, with values exceeding these ranges flagged as RFI. The primary limitation of thresholding methods is the potential misidentification of astronomical signals as RFI, making the reduction of false positive rates a critical research direction in post-correlation RFI mitigation.

### 4.1 VarThreshold Algorithm

In both frequency and time domains, RFI data often affects neighboring data points. Due to their relatively low intensity, these affected points may not be flagged as RFI, increasing detection error rates. VarThreshold is a combined threshold algorithm based on the principle that a combination of samples is flagged as RFI when a certain statistical property of the combined samples exceeds a specified limit. Assuming  $x_i$  and  $x_{i+1}$  are adjacent sample points, conventional thresholding examines whether each sample individually exceeds the set threshold level. In contrast, combined thresholding only flags the combination when both samples exceed the threshold. If samples  $x_i$  and  $x_{i+1}$  are not flagged, they are combined with the next adjacent sampling point  $x_{i+2}$  for continued threshold comparison, iterating multiple times to complete RFI flagging, as shown in equation (18):

$$\text{VarThreshold}(x_i, \dots, x_{i+L-1}) = \begin{cases} \text{RFI} & \text{if } |x_{i+j}| > T_L \text{ for all } j \in [0, L-1] \\ \text{Not RFI} & \text{otherwise} \end{cases}$$

This expression indicates that if all sampling points within a range of  $L$  points starting from time  $i$  have absolute values greater than  $T_L$ , then all  $L$  points are flagged as RFI.

The VarThreshold algorithm uses a strictly decreasing threshold sequence  $(T_1, T_2, \dots, T_L)$  to determine whether a sampling point should be flagged as RFI, as shown in equation (19):

$$T_L = T_1 \cdot \alpha^{L-1}$$

where  $T_1$  is the single-sample threshold, and optimal performance is achieved when  $\alpha = 0.5$ . The value of  $T_1$  is determined based on the statistical level of the data, specifically selecting the threshold that minimizes the false positive rate for RFI detection.

## 4.2 SumThreshold Algorithm

The SumThreshold algorithm is another combined thresholding method similar to VarThreshold in its approach to determining threshold sequences, but with different  $L$  values. Optimal performance occurs when  $L = 8$ . A threshold set  $(T_1, T_2, \dots, T_L)$  is calculated and compared iteratively with samples:  $T_1$  compares with single samples,  $T_2$  compares with the mean of two samples, and  $T_L$  compares with the mean of  $L$  samples. When the sample mean exceeds the threshold, the samples are flagged as RFI and excluded from subsequent iterations. For a one-dimensional sequence  $[x_1, x_2, \dots, x_N]$  with thresholds  $[T_1, T_2, \dots, T_L]$ .

Figs. 7, 8, and 9 illustrate the SumThreshold algorithm process. From an algorithmic perspective, the time complexity is  $O(N \cdot L)$ . To reduce computational cost, when threshold  $L$  takes values of 1, 2, 4, 8, 16, 32, 64, 128, 256, etc., the time complexity reduces to  $O(N \log L)$ .

## 4.3 Singular Value Decomposition

Singular Value Decomposition (SVD) is an important matrix factorization technique in linear algebra capable of extracting information, reducing data dimensionality, and removing noise by mapping data to a lower-dimensional space. SVD decomposes a matrix as:

$$\mathbf{X} = \mathbf{U} \mathbf{V}^T$$

where  $\mathbf{U}$  and  $\mathbf{V}$  are orthogonal matrices, and  $\mathbf{\Lambda}$  is a diagonal matrix with singular values on its diagonal. The vast majority of RFI signals exhibit large amplitudes in either the time or frequency domain.

Let  $\mathbf{X}$  represent the received astronomical signal. After performing SVD,  $\mathbf{X} = \mathbf{U} \mathbf{V}^T$ , the largest singular values are set to zero. Upon inverse reconstruction, a new matrix  $\tilde{\mathbf{X}}$  is obtained, where signals with large amplitudes (RFI) in the original matrix  $\mathbf{X}$  are suppressed.

This algorithm achieves optimal performance when RFI signals are sufficiently strong and  $\text{SNR} \gg 1$ . However, the method is only applicable to broadband RFI signals and is not suitable for RFI signals that follow a Gaussian distribution.

---

## 5. Machine Learning-Based RFI Mitigation Methods

The rapid development of machine learning and deep learning technologies, combined with the exponential growth of astronomical data, has made their integration inevitable. Through learning and training on large datasets, these methods can effectively flag RFI in astronomical signals.

### 5.1 Principal Component Analysis

Principal Component Analysis (PCA) is a method for dimensionality reduction in data features, commonly used to reduce dataset dimensionality while preserving features that contribute most to variance. Lower-order components often retain important information, facilitating feature extraction and classification.

Reference [19] conducted classification tests on nine common transient RFI sources from daily life, evaluating both standard PCA and kernel PCA in the time domain. Using a clustering separation method to compare the clustering accuracy of both approaches (Figs. 10 and 11), the study found that kernel PCA outperformed standard PCA in distinguishing source types, effectively differentiating transient RFI origins.

### 5.2 Support Vector Machine

In machine learning, Support Vector Machine (SVM) is a supervised learning model and associated algorithm used for classification and regression analysis.

Reference [21] selected training data with time length  $T$ , including both non-RFI and RFI signals, extracting features such as root-mean-square, mean, variance, mean-to-variance ratio, skewness, kurtosis, maximum amplitude, minimum amplitude, and peak count. These features were imported into an SVM training model, which then classified test data into RFI and non-RFI categories, as illustrated in Fig. 12. Results demonstrated that this method accurately detected RFI even at very low INR and small RFI duty cycles, showing excellent performance.

### 5.3 Neural Networks

The basic structure of neural networks comprises an input layer, hidden layers, and an output layer, consisting of numerous interconnected nodes. As a nonlinear statistical data modeling tool, neural networks are commonly used to model complex relationships between inputs and outputs or to explore data patterns.

Reference [23] proposed Deep Fully Convolutional Neural Networks (DFCN) for image segmentation, addressing pixel-level image classification. Reference [24] utilized DFCN (Fig. 13) with  $T$  and  $F$  corresponding to input time and frequency visibility dimensions,  $C$  representing the number of filter layers, and  $L$  representing the total number of layers between input and fully convolutional layers. Using amplitude and phase as features, waterfall plots were processed for feature extraction and RFI flagging.

Reference [25] introduced the U-Net neural network, which adds upsampling convolutional expansion paths to convolutional neural networks, enabling fast and effective image segmentation for small datasets. Originally applied to medical cell image segmentation, reference [3] successfully applied the U-Net deep neural network model to identify RFI in single-dish radio telescope data, with its structure shown in Fig. 14.

Reference [26] found that Convolutional Neural Networks (CNN) often produced mislabeled RFI in FAST data, requiring significant manual review and creating substantial additional workload. To overcome this limitation, they proposed a neural network model called RFI-Net (Fig. 15), with results demonstrating that RFI-Net outperformed U-Net, K-Nearest Neighbor (KNN), and SumThreshold algorithms.

Recurrent Neural Networks (RNN) are a class of neural networks designed for processing sequential data, demonstrating excellent performance in handling context-dependent speech data. Astronomical signals are also time-series data, making them highly suitable for RNN processing. Burd et al. utilized RNN for RFI detection in radio telescope interferometric array data, distinguishing RFI from non-RFI data based on 610 MHz RFI amplitude information from the Giant Metrewave Radio Telescope (GMRT).

---

## 6. Summary

This paper systematically elaborates on algorithms and techniques for suppressing and flagging RFI in radio astronomy, analyzing the advantages and limitations of RFI mitigation methods across active prevention, pre-correlation, and post-correlation stages. Effective shielding measures during the active prevention stage can prevent most RFI from entering the system. Pre-correlation methods exploit the correlated nature of astronomical signals across antennas to suppress RFI. Post-correlation algorithms primarily rely on threshold-based RFI flagging. Machine learning techniques for RFI flagging and identification represent current research hotspots, as training on massive astronomical datasets can substantially improve RFI flagging accuracy. However, a notable drawback is the time-consuming manual labeling of RFI and astronomical signals in training samples. Currently available training samples remain limited.

RFI mitigation requires collaborative solutions employing multiple methods across different stages, with RFI considerations necessary from radio environment protection around telescope sites to final astronomical data processing. Each observatory must select appropriate RFI mitigation methods based on actual requirements and the electromagnetic interference conditions of its environment.

---

## Acknowledgments

This work is supported by the National Key R&D Program of China (2021YFC2203502); the National Natural Science Foundation of China (12173077, 11873082, 11803080, 12003062); the Tianshan Innovation Team Program of Xinjiang Uygur Autonomous Region (2022D14020); the Youth

Innovation Promotion Association of the Chinese Academy of Sciences; the National Key R&D Program (2018YFA0404704); the National Astronomical Data Center; and the Special Fund for Equipment Renewal and Major Instrument Operation of Astronomical Observatories of the Chinese Academy of Sciences. This paper benefits from data resources and technical support provided by the China Virtual Observatory, the National Astronomical Data Center, and the Scientific Data Center System of the Chinese Academy of Sciences.

## References

- [1] Wang Yu, Zhang Hai-Yan, Hu Hao et al. Satellite RFI mitigation on FAST[J]. *Research in Astronomy and Astrophysics*, 2021(21):18.
- [2] Porko, Jukka-Pekka Göran. Radio frequency interference in radio astronomy[D], Aalto University, 2011:77.
- [3] J. Akeret, C. Chang, A. Lucchi, A. Refregier. Radio frequency interference mitigation using deep convolutional neural networks[J]. *Astronomy and Computing*, 2017(18):35-39.
- [4] 胡浩, 张海燕, 黄仕杰. FAST 电波环境保护措施 [J]. *深空探测学报*, 2020, 7(02):152-157. HU Hao, ZHANG Haiyan, HUANG Shijie. Protection Measures of FAST Radio Environment[J]. *Journal of Deep Space Exploration*, 2020, 7(02):152-157.
- [5] 王娜. 新疆奇台 110 米射电望远镜 [J]. *中国科学: 物理学力学天文学*, 2014, 44(08): 783-794. WANG Na. Xinjiang Qitai 110 m radio telescope[J]. *SCIENTIA SINICA Physica, Mechanica & Astronomica*, 2014, 44(08): 783-794.
- [6] 刘晔. QTT 台站限制区域电磁干扰影响分析 [C]. 中国天文学会 2018 年学术年会. LIU Ye. Analysis of electromagnetic interference impact in the restricted area of QTT stations[C]. *Chinese Astronomical Society 2018 Annual Academic Conference*.
- [7] 刘晔, 刘奇. 大口径射电望远镜台址电磁干扰预测方法 [J]. *中国科学: 物理学力学天文学*, 2019, 49(09):121-129. LIU Ye, LIU Qi. A prediction method for electromagnetic interference of large aperture radio telescope[J]. *SCIENTIA SINICA Physica, Mechanica & Astronomica*, 2019, 49(09):121-129.
- [8] Cecilia Barnbaum and Richard F. Bradley. A New Approach to Interference Excision in Radio Astronomy: Real-Time Adaptive Cancellation[J]. *The Astronomical Journal*, 1998(116):2598-2614.
- [9] Kesteven, M. and Hobbs, G. and Clement, R. and Dawson, B. and Manchester, R. and Uppal, T. Adaptive filters revisited: Radio frequency interference mitigation in pulsar observations[J]. *Radio Science*, 2005(40).
- [10] Curotto Molina, Franco Andreas. Design, implementation and characterization of a radio frequency interference digital adaptive filter using a field-programmable gate array[D]. *Universidad de Chile*, 2019.
- [11] Raza J, Boonstra A J, Van der Veen A J. Spatial filtering of RF interference in radio astronomy[J]. *IEEE Signal Processing Letters*, 2002, 9(2): 64-67.

- [12] Hellbourg G. Radio Frequency Interference spatial processing for modern radio telescopes[D]. Université d' Orléans, 2014.
- [13] Kocz J, Briggs F, Reynolds J. Spatial filtering using a multibeam receiver[C]//Proceedings of RFI Mitigation Workshop PoS (RFI2010) proceedings. 2010, 1: 32.
- [14] Offringa, A. R., de Bruyn, A. G., Biehl, M. et al. Post-correlation radio frequency interference classification methods[J]. Monthly Notices of the Royal Astronomical Society, 2010(405):155-167.
- [15] Offringa A R. Algorithms for radio interference detection and removal[D]. University of Groningen, 2012.
- [16] Guo Q, Zhang C, Zhang Y, et al. An efficient SVD-based method for image denoising[J]. IEEE transactions on Circuits and Systems for Video Technology, 2015, 26(5): 868-880.
- [17] Abdi H, Williams L J. Principal component analysis[J]. Wiley interdisciplinary reviews: computational statistics, 2010, 2(4): 433-459.
- [18] Kherif F, Latypova A. Principal component analysis[M]//Machine Learning. Academic Press, 2020: 209-225.
- [19] Czech Daniel, Mishra Amit Kumar, Inggs Michael. Time domain classification of transient radio frequency interference[C]. 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2016:302-305.
- [20] Pisner D A, Schnyer D M. Support vector machine[M]//Machine Learning. Academic Press, 2020: 101-121.
- [21] Nazar I M, Aksoy M. Radio Frequency Interference Detection in Microwave Radiometry Using Support Vector Machines[J]. URSI GASS, 2020.
- [22] Aggarwal C C. Neural networks and deep learning[J]. Springer, 2018, 10: 978-3.
- [23] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.
- [24] Kerrigan J, Plante P L, Kohn S, et al. Optimizing sparse RFI prediction using deep learning[J]. Monthly Notices of the Royal Astronomical Society, 2019, 488(2): 2605-2615.
- [25] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.
- [26] Yang Z, Yu C, Xiao J, et al. Deep residual detection of radio frequency interference for FAST[J]. Monthly Notices of the Royal Astronomical Society, 2020, 492(1): 1421-1431.

- [27] Zaremba W, Sutskever I, Vinyals O. Recurrent neural network regularization[J]. arXiv preprint arXiv:1409.2329, 2014.
- [28] Burd P R, Mannheim K, März T, et al. Detecting radio frequency interference in radio-antenna arrays with the recurrent neural network algorithm[J]. Astronomische Nachrichten, 2018, 339(5): 358-362.

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv –Machine translation. Verify with original.*