

李昂, 王洪丰, 赵一, 等. 基于犹豫模糊聚类的 FAST 脉冲星观测数据射频干扰抑制方法研究[J]. 智能计算机与应用, 2025, 15(12): 47-54. DOI: 10.20169/j.issn.2095-2163.25052001

基于犹豫模糊聚类的 FAST 脉冲星观测数据射频干扰抑制方法研究

李昂^{1,2}, 王洪丰², 赵一², 陶振钊²

(1 吉林化工大学 信息与控制工程学院, 吉林 吉林 132022; 2 德州学院 计算机与信息学院, 山东 德州 253023)

摘要: 针对五百米口径 FAST 射电望远镜 (Five-hundred-meter-Aperture Spherical Telescope, FAST) 脉冲星搜寻数据中普遍存在的射频干扰 (Radio Frequency Interference, RFI) 问题, 本文提出了一种基于犹豫模糊聚类 (Hesitant Fuzzy Clustering, HFC) 的干扰识别与抑制方法。通过通道均值、标准差与最大值构建低维统计特征向量, 结合 K-Means 初始化与动态模糊度更新机制实现对非稳态 RFI 的鲁棒识别与高效去除。实验选取 20 组包含已知脉冲信号的 PSRFITS 格式观测数据, 将 PSRFITS 数据处理流程中的射频干扰抑制环节作为研究重点, 分别采用四分位距 (Interquartile Range, IQR)、K-Means 聚类、模糊 C 均值 (Fuzzy C-Means, FCM) 和 HFC 4 种方法进行 RFI 抑制, 并在统一流程与指标体系下, 从信噪比提升和异常通道删除数量两个维度展开量化比较与可视化分析。实验结果表明, HFC 在 3 种聚类方法中表现最优, 平均信噪比提升达 17%, 显著优于传统的 IQR 方法, 验证了其在非稳态干扰场景中的抑制优势。

关键词: FAST 射电望远镜; 脉冲星搜寻; 射频干扰; PSRFITS 数据处理; 犹豫模糊聚类

中图分类号: P161

文献标志码: A

文章编号: 2095-2163(2025)12-0047-08

Radio frequency interference suppression in FAST pulsar observation data based on hesitant fuzzy clustering

LI Ang^{1,2}, WANG Hongfeng², ZHAO Yi², TAO Zhenzhao²

(1 School of Information and Control Engineering, Jilin University of Chemical Technology, Jilin 132022, Jilin, China;

2 School of Computer and Information, Dezhou University, Dezhou 253023, Shandong, China)

Abstract: This paper proposes an interference identification and suppression method based on Hesitant Fuzzy Clustering (HFC) to address the common issue of Radio Frequency Interference (RFI) in the pulsar search data of the Five hundred meter Aperture Spherical Telescope (FAST). This method constructs low dimensional statistical feature vectors through channel mean, standard deviation, and maximum value, and combines K-Means initialization and dynamic ambiguity update mechanism to achieve robust recognition and efficient removal of non-stationary RFI. Twenty sets of PSRFITS format observation data containing known pulse signals were selected for the experiment, and the RF interference suppression step in the PSRFITS data processing flow was taken as the research focus. Four methods, namely Interquartile Range (IQR), K-Means clustering, Fuzzy C-Means (FCM), and HFC, were used for RFI suppression. Under a unified process and indicator system, quantitative comparison and visual analysis were conducted from the dimensions of signal-to-noise ratio improvement and abnormal channel deletion quantity. The experimental results show that HFC performs the best among the three clustering methods, with an average signal-to-noise ratio improvement of 17%, significantly better than the traditional IQR method, verifying its suppression advantage in non-stationary interference scenarios.

Key words: FAST radio telescope; pulsar search; Radio Frequency Interference; PSRFITS data processing; Hesitant Fuzzy Clustering

0 引言

脉冲星是超新星爆发后形成的高速自转中子

星, 在射电波段呈现高度稳定、周期性的脉冲信号, 是研究致密物理、引力波和精密导航的关键天体^[1-2]。

基金项目: 德州学院学科项目(平台)建设(023XKZX017); 德州州市级研发计划(2022dzkj097)。

作者简介: 李昂(2000-), 男, 硕士研究生, 主要研究方向: 脉冲星搜寻。

通信作者: 王洪丰(1980-), 男, 博士, 教授, 主要研究方向: 大数据挖掘, 机器学习。Email: dzuwhf@163.com。

收稿日期: 2025-05-20

实际观测中,射频干扰(RFI)严重制约脉冲星探测效率与准确性。RFI源于通信系统、电力设备等人造发射源,具有非天文特性,通常分为窄带、宽带和脉冲式3类^[3]。

传统基于阈值的剔除方法可识别部分异常值,适用于窄带干扰,但在复杂、非稳态场景下易误删^[4]。为提升去干扰方法的适应性与智能化,近年来研究者提出多种基于机器学习的RFI识别手段。Wolfaardt等^[5]结合K近邻(K-Nearest Neighbors, KNN)与高斯混合模型(Gaussian Mixture Model, GMM)构建分类系统,验证了传统机器学习的有效性;Akeret等^[6]首次将深度卷积神经网络(Convolutional Neural Network, CNN)用于RFI抑制,基于U-Net对频率-时间图进行像素级判别;Fridman与Baan^[7]系统总结了传统RFI抑制技术并提出量化评估标准,为不同望远镜与观测任务提供参考。

目前,世界主要射电望远镜包括澳大利亚的64米口径射电望远镜Parkes、美国的305米口径射电望远镜Arecibo、德国的100米口径射电望远镜Effelsberg和中国的500米口径球面射电望远镜(FAST)等,在灵敏度、频段覆盖与后端体系上各具特点,对脉冲星搜寻与时域天文学做出重要贡献,值得注意的是,这些大型设备因其高灵敏度特性,在捕获微弱脉冲星信号的同时,也更容易受到各类RFI的影响,这使得发展高效的射频干扰抑制技术成为数据处理流程中的关键环节^[8-11]。美国的305米口径射电望远镜Arecibo于2020年因设备坠落导致望远镜反射盘表面被砸毁,从而退出历史^[12]。FAST作为目前世界上口径最大、灵敏度最高的单口径射电望远镜,为脉冲星搜寻提供了强大支撑,其生成的数据以PSRFITS格式存储,结构丰富^[13-14]。本文所用数据来自一次基于FAST的观测任务,中心频率1250 MHz,总带宽500 MHz,划分为4096个频率通道和4个极化通道,包含128个子积分,每个子积分1024个采样点,时间分辨率为98.304 μs。格式化处理后,形成维度为4096×131072的二维矩阵,充分反映信号结构特征,为后续干扰识别与算法研究奠定基础。

聚类作为一种无监督学习算法,具有良好的模型独立性与数据适应性^[15]。本文提出一种改进的犹豫模糊聚类(HFC)方法,基于通道均值、标准差与最大值等低维统计特征构建描述向量,结合K-Means初始化聚类中心和动态模糊度更新机制,实

现对非稳态RFI的高效识别与抑制。

1 数据准备

本文选取20组来自FAST射电望远镜的PSRFITS观测数据,数据均包含已知脉冲信号及其色散量(Dispersion Measure, DM)和到达时间信息,基于Song等^[16]提出的脉冲相位与窗口模型,结合消色散处理与信噪比(Signal-to-Noise Ratio, SNR)计算,构建了干扰识别与评估流程。

预处理阶段,从PSRFITS文件中提取频率和时间轴信息:频率由DAT_FREQ字段给出;时间起始点由STT_IMJD、STT_SMJD与STT_OFFS确定,结合采样间隔(Time Bin, TBIN)与总采样点数,构建标准(Coordinated Universal Time, UTC)时间轴^[17]。

观测数据经维度重排、极化通道合并及归一化后,形成形状为(频率,时间)的二维频率-时间矩阵。消色散后对各采样点的频率通道求平均可生成脉冲轮廓图。

为提升方法可比性,本文提出犹豫模糊聚类方法及对比方法均基于上述标准化数据结构进行干扰检测与信号保留处理。

2 方法

本文比较4种异常通道干扰去除方法,分别为基于四分位距(IQR)的异常检测法、K-Means聚类法、模糊C均值聚类(FCM)法,以及改进的犹豫模糊聚类(HFC)法。

数据处理过程如图1所示。射电望远镜信号经初步处理生成PSRFITS文件;数据转换为二维频率-时间矩阵后,复制两份:一份保留原始数据,另一份分别应用4种干扰抑制方法;所有数据经消色散处理,计算信噪比并绘制脉冲轮廓图,得到每种方法的结果用于后续分析。

2.1 基于IQR的异常通道检测方法

IQR方法通过分析通道功率分布的分位数识别异常。计算每个通道的时间平均功率,在全体通道均值集合中求第一四分位数(Q1)和第三四分位数(Q3),得出四分位距IQR:

$$IQR = Q3 - Q1 \quad (1)$$

以 $Q3 + n \cdot IQR$ 作为异常判定阈值,一般情况下,当 $n = 1.5$ 时,满足条件的观测值为异常值,超过该阈值的通道视为异常并置零。箱线图的判定公式如下:

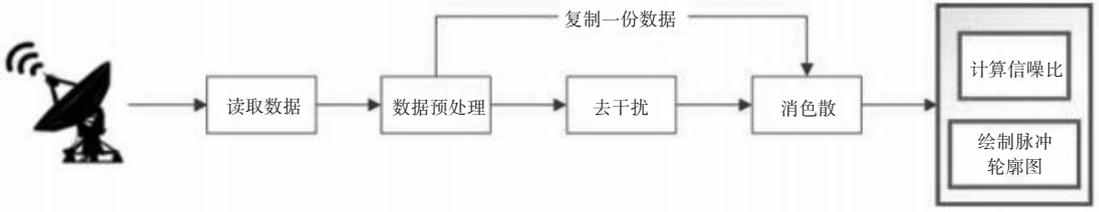


图 1 数据处理过程

Fig. 1 Data processing process

$$\begin{cases} \text{Out} > Q3 + n\text{IQR} \\ \text{Out} < Q1 - n\text{IQR} \end{cases} \quad (2)$$

该方法实现简单高效,适合快速筛除规则干扰,但对非对称或多模态干扰适应性有限。

2.2 基于 K-Means 的通道聚类方法

K-Means 适用于通道信号强度差异显著的情况^[18]。提取每个通道的均值、标准差和最大值构成三维特征向量。设置聚类数 $k=2$ 进行聚类,比较两簇的平均功率,将高功率簇识别为异常簇,属于该簇的通道即被视为干扰通道并置零处理。K-Means 聚类公式如下:

$$\min \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (3)$$

其中, K 是预设聚类簇数; x_i 是第 i 个通道的特征向量; C_k 是第 k 个簇; μ_k 是第 k 个簇的质心; $\|x_i - \mu_k\|^2$ 是通道特征与其簇中心的欧式距离平方。

2.3 基于 FCM 的通道聚类方法

FCM 通过隶属度提高聚类柔性,适用于干扰幅度不均、边界模糊场景^[19]。同样使用通道均值、标准差、最大值三维特征。初始化随机隶属度矩阵后迭代更新簇中心和隶属度至收敛。每个通道划入隶属度最高的簇,若该簇平均功率显著偏高,则其通道置零。

聚类目标函数:

$$J = \sum_{i=1}^N \sum_{k=1}^C u_{ik}^m \|x_i - c_k\|^2 \quad (4)$$

隶属度更新公式:

$$u_{ik} = \left(\sum_{j=1}^C \left(\frac{\|x_i - c_k\|}{\|x_i - c_j\|} \right)^{\frac{2}{m-1}} \right)^{-1} \quad (5)$$

聚类中心更新公式:

$$c_k = \frac{\sum_{i=1}^N u_{ik}^m x_i}{\sum_{i=1}^N u_{ik}^m} \quad (6)$$

其中, N 是通道总数; C 是簇数; x_i 是第 i 个通道的特征向量; c_k 是第 k 个簇的中心; u_{ik} 是第 i 个通道

属于第 k 个簇的隶属度; m 是模糊因子,控制隶属度的模糊程度。

2.4 基于 HFC 的通道聚类方法

为提升非稳态 RFI 识别能力,本文提出改进的犹豫模糊聚类(HFC)方法,沿用三维统计特征并归一化。初始聚类采用 K-Means 估计初始隶属度。在迭代过程中动态更新模糊因子 m ,使其随隶属度分布不确定性自适应调整,优化聚类边界与置信度^[20]。最终根据簇中心功率确定异常簇,筛选隶属度大于设定阈值的通道置零。HFC 方法融合软聚类的灵活性与模糊度调节自适应性,在干扰抑制与信号完整性间取得更好平衡,提升复杂非线性背景下的识别精度与鲁棒性。

模糊度更新公式:

$$H(U) = - \sum_{i=1}^N \sum_{k=1}^C u_{ik} \log u_{ik} \quad (7)$$

$$m = m_{\text{base}} + \alpha \cdot H(U) \quad (8)$$

其中, $H(U)$ 为模糊熵,表示隶属度分布的模糊程度; m_{base} 为基础模糊因子; α 是调节系数,用于控制 m 的变化幅度,其他符号与 FCM 中一致。

3 方法评估

本文系统评估 4 种 RFI 去除方法在信噪比提升(ΔSNR)、异常通道删除数量、信号完整性及运行效率等方面的性能。

3.1 评估指标与实验设置

评估采用两个核心指标:信噪比提升和异常通道删除数量,分别衡量不同方法的干扰抑制能力与信号的保留程度。

信噪比提升用于量化干扰去除前后脉冲信号质量的变化。具体而言,以 FITS 文件中记录的已知脉冲到达位置为参考,在脉冲中心的一定时间窗口内积分信号强度,减去背景区域的均值并归一化噪声标准差,计算除去干扰前后的 SNR 值,二者之差即为 ΔSNR 。该指标不仅能直接反映去干扰的效果,还间接表征了信号完整性的保持程度,具体公式如

下式:

$$\text{SNR} = \frac{u_{\text{pulse}} - u_{\text{bkg}}}{\sigma_{\text{bkg}}} \quad (9)$$

$$\Delta\text{SNR} = \text{SNR}_{\text{denoised}} - \text{SNR}_{\text{raw}} \quad (10)$$

其中, u_{pulse} 是脉冲窗口内的平均强度; u_{bkg} 是背景窗口的平均强度; σ_{bkg} 是背景窗口的标准差; SNR_{raw} 为原始数据的信噪比; $\text{SNR}_{\text{denoised}}$ 为去干扰后的信噪比; ΔSNR 则是信噪比提升量。

异常通道删除数量指在处理过程中被识别并置零的通道数目,用于衡量方法的敏感性与保守性,删除通道过少可能导致 RFI 残留,过多则可能误伤有效信号,因而需与 ΔSNR 结合分析以全面评价算法的精确性与鲁棒性。

实验流程统一,包括数据加载、干扰去除、GPU 消色散、脉冲轮廓提取及 ΔSNR 计算,所有方法均在一致的硬件平台下执行,确保结果可比性。

表 1 3 种聚类方法的超参数范围设置

Table 1 Hyperparameter range settings for the three clustering methods

方法	聚类数 c	迭代次数	模糊因子	收敛阈值	比例阈值
K-Means	2	10,20			
FCM	2	30,50	1.5,1.7,1.9	1×10^{-2} , 5×10^{-3} , 1×10^{-3} , 5×10^{-4}	
HFC	2	30,50	1.5,1.7,1.9	1×10^{-2} , 5×10^{-3} , 1×10^{-3} , 5×10^{-4}	0.5,0.55,0.6

经过搜索后,在相同数据集和统一评价指标下,3 种聚类经过超参数调整后均有不同程度的性能提升。K-Means 聚类数 $c = 2$ 、迭代次数为 10;FCM 的最佳配置为 $c = 2$ 、迭代次数为 50、模糊因子 $m = 1.7$ 、收敛阈值为 1.0×10^{-2} ;HFC 在多个样本中取得了最高评分,其最优参数为 $c = 2$ 、迭代次数 50、模糊因子 $m = 1.7$ 、收敛阈值为 5.0×10^{-3} 、比例阈值为 0.5,实现较大 ΔSNR 提升并显著降低误删风险。

该实验结果进一步验证了不同方法在参数配置下的干扰通道去除效果具有明显差异。这种差异未必反映方法本身的优劣,更可能源于参数选择,建议在实际应用中根据数据特征进行微调或自动化参数搜索。

3.3 性能对比分析

20 组 FITS 文件的去干扰结果见表 2,其中加粗部分标示了当前 FITS 文件中信噪比提升最显著的方法及对应的通道删除数量,可以看出 4 种方法均提升 SNR。HFC 方法在多数样本上表现最优或次优。例如,在 FRB121102_tracking-M01_1300 中,HFC 将原始 SNR 从 101.76 提升至 133.03,提升幅度达 31.27,略优于 K-Means 和 FCM。

为了更直观的看出不同方法的效果,将 4 种方法在最优配置下的平均信噪比进行了计算,4 种不

3.2 聚类超参数

使用 5 个独立样本对 K-Means、FCM、HFC 进行超参数搜索,以 ΔSNR 和通道保留率为基础计算综合评分,综合评分函数:

$$\text{Score} = 0.8\Delta\text{SNR} + 0.2 \left(\frac{4096 - N_{\text{removed}}}{4096} \right) \quad (11)$$

其中, ΔSNR 表示去干扰前后脉冲轮廓的局部信噪比增益, N_{removed} 为被识别并置零的通道数量。

3 种聚类方法的超参数范围设置见表 1。在超参数搜索中,聚类数统一设定为 2,将通道分为干扰与非干扰,设定迭代次数防止算法陷入死循环,模糊因子 m 用于控制聚类的模糊程度。在改进的犹豫模糊聚类(HFC)方法中,模糊因子还会根据当前聚类的不确定性动态调整,以提高对非稳态 RFI 的适应能力。本文使用收敛阈值来判断隶属度的更新是否已达到稳定,使用比例阈值判断一个通道是否异常。

同方法的信噪比提升对比见表 3,可以看出 HFC 的平均 ΔSNR 最高,达到了 26.17,3 种聚类方法均优于 IQR 方法,验证了聚类方法在处理复杂非稳态干扰方面的优势。

但评价方法优劣不仅需关注 ΔSNR ,还需考察其识别精准性。若某方法通过大量删除通道获得高 SNR,可能以牺牲信号完整性为代价。为此,将异常通道删除数量纳入考量,作为去干扰精度的衡量依据。理想方法应在提升 SNR 的同时,减少对正常通道的误删,实现干扰识别与信号保留的双重优化。

结合信噪比提升和异常通道删除数量两个指标来看,HFC 方法在多个样本中实现了最高的 SNR 提升效果。且对单个文件进行分析时可以发现,若 HFC 删除的通道数量较多,其 SNR 提升也最为显著;而在 HFC 的 SNR 并非最高、甚至处于最低的情况下,HFC 与其他方法的信噪比提升差距也较小,但是保留了更多的信号。以样本 FRB121102_tracking-M01_0407 为例,HFC 通过删除 352 个通道实现了 121.54 的 SNR 值,而 K-Means 与 FCM 分别删除了 363 和 360 个通道,且 SNR 略低于 HFC,表明尽管三者同为聚类算法,HFC 并非简单地滥删通道,而是更具策略的精准干扰识别。

表 2 20 组 FITS 文件的去干扰结果

Table 2 RFI mitigation results for 20 sets of FITS files

文件名	初始 SNR	IQR SNR	IQR 删除通道数	HFC SNR	HFC 删除通道数	K-Means SNR	K-Means 删除通道数	FCM SNR	FCM 删除通道数
FRB121102_tracking-M01_0053	121.07	131.94	317	132.02	161	131.31	155	131.39	152
FRB121102_tracking-M01_1300	101.76	132.84	566	133.03	476	132.15	580	131.52	564
FRB121102_tracking-M01_0407	109.51	120.44	479	121.54	352	121.26	363	121.20	360
FRB121102_tracking-M01_0937	93.76	166.92	863	170.58	957	171.60	1 085	171.52	1 082
FRB121102_tracking-M01_0405	56.86	59.91	480	60.00	450	59.31	356	59.31	353
FRB121102_tracking-M01_0129	49.21	51.37	260	51.09	223	50.97	218	50.97	218
FRB121102_tracking-M01_0774	58.87	98.91	578	102.14	507	103.20	623	102.61	607
FRB121102_tracking-M01_0634	40.28	62.77	564	62.24	354	62.52	481	62.49	476
FRB121102_tracking-M01_0279	62.43	63.37	489	68.77	369	68.74	378	68.70	374
FRB121102_tracking-M01_1195	40.93	42.45	275	42.60	292	42.26	199	42.23	196
FRB121102_tracking-M01_0069	112.55	123.48	431	121.31	346	121.04	407	121.13	409
FRB121102_tracking-M01_0197	221.93	246.50	367	245.61	310	243.77	287	243.59	285
FRB121102_tracking-M01_0106	136.10	170.10	470	167.26	384	166.93	354	166.44	347
FRB121102_tracking-M01_0013	59.19	80.36	551	79.57	362	79.89	417	79.89	414
FRB121102_tracking-M01_0257	131.01	135.35	341	135.17	341	135.79	335	135.81	346
FRB121102_tracking-M01_0152	55.16	61.85	240	68.39	210	62.69	220	62.77	222
FRB20201124A_tracking-M01_0890	19.20	20.83	611	19.72	301	19.87	401	19.82	397
FRB20201124A_tracking-M01_0471	14.43	16.48	533	16.39	351	16.05	332	16.05	332
FRB20201124A_tracking-M01_0299	15.02	15.35	363	15.37	164	15.23	147	15.23	145
FRB20201124A_tracking-M01_0452	49.72	52.21	508	51.49	307	51.25	366	51.24	360

表 3 4 种不同方法的信噪比提升对比

Table 3 Comparison of SNR improvement among four different methods

方法	Δ SNR
HFC	26.17
K-Means	22.88
FCM	22.82
IQR	21.74

综上所述, HFC 方法在信噪比提升与有效通道保留之间实现了较为理想的平衡, 表现最为稳定。K-Means 和 FCM 可作为资源受限场景下的次优选择, 而 IQR 适合作为轻量化的基准对比方案。

在这 20 个文件中, 前 10 个文件主要受到较强的窄带干扰影响, 11 至 16 则以弱窄带干扰为主, 而 17 至 20 则包含少量弱窄带干扰和部分宽带干扰。可以看出, 在强窄带和弱窄带干扰的场景下, 3 种聚类方法整体优于传统的阈值法。尽管阈值法在部分样本中实现了较大的信噪比提升, 但其代价往往是误删更多的有效通道信号。相较之下, 聚类方法在提升信噪比方面仅略低于阈值法, 但能够更有效地

保留正常通道数据, 表现最优的是改进的 HFC 算法, 这对于后续的脉冲星信号检索尤为关键, 尤其是对于微弱信号的检测。

由于宽带干扰通常作用于多个频率通道, 但其持续时间较短, 若仅依据通道维度进行整体删除, 可能会将包含脉冲信号的通道一并移除, 导致虽然干扰被消除, 但有效信号也随之丢失, 对于存在宽带干扰的样本, 无论是阈值法还是聚类方法, 在提升整体信噪比方面均未取得显著成效。抑制宽带干扰仅依赖通道级的去除是不足的, 有必要引入基于点级的干扰去除机制。

3.4 可视化对比与两组数据分析

从实验数据中选取两个 FAST PSRFITS 文件, 分别对原始数据和 4 种方法的处理结果进行可视化对比。

例 1: FRB121102_tracking-M01_0405

该样本存在集中于 1 200~1 300 MHz 的窄带干扰, 原始 SNR 为 56.86, 例 1 未去干扰情况下的频率时间图和脉冲轮廓图如图 2 所示。

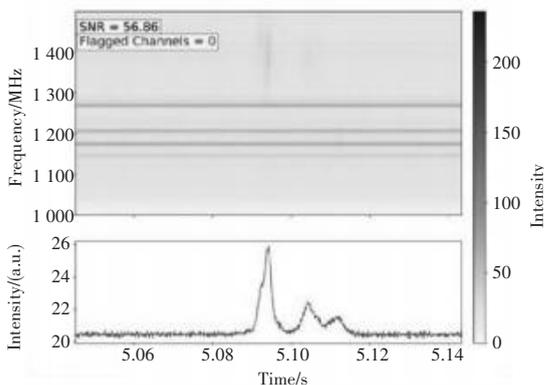


图2 例1未去干扰情况下的频率时间图和脉冲轮廓图

Fig. 2 Frequency-time diagram and pulse profile of Example 1 without interference suppression

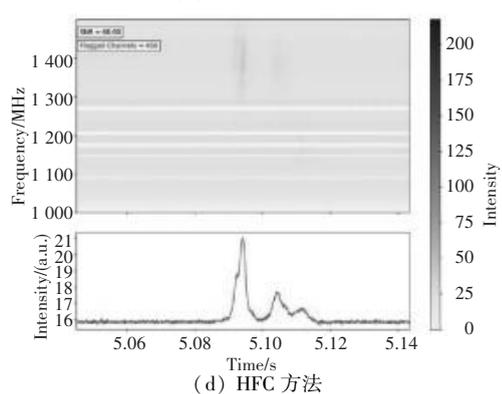
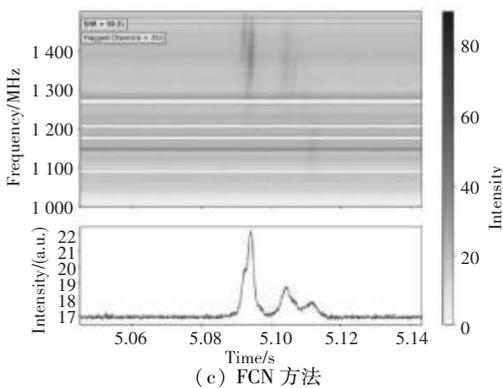
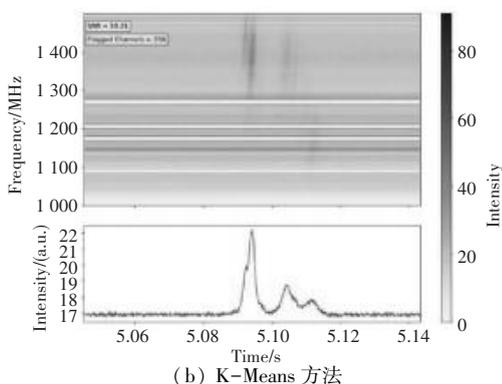
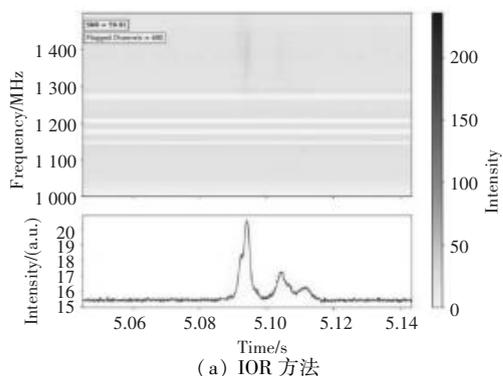


图3 对例1使用4种干扰抑制方法处理后的频率时间图和脉冲轮廓图

Fig. 3 Frequency-time diagrams and pulse profiles of Example 1 after applying four different interference suppression methods

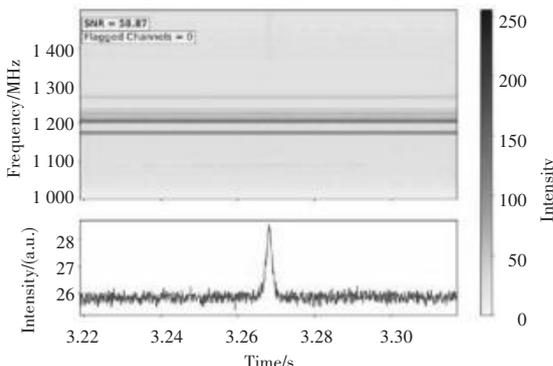


图4 例2未去干扰情况下的频率时间图和脉冲轮廓图

Fig. 4 Frequency-time diagram and pulse profile of Example 2 without interference suppression

对例1使用4种干扰抑制方法处理后的频率时间图和脉冲轮廓图如图3所示。IQR方法SNR提升至59.91,删除了480条通道,出现了整段频率屏蔽现象;K-Means方法SNR提升至59.31,删除了356条通道,明显有干扰残留;FCM方法SNR提升至59.31,删除了353条通道,副脉冲存在轻微拖尾,结果与K-Means接近;HFC方法SNR提升至60,删除了450条通道,干扰定位精准,信号保留完整,综合信噪比提升和信号保留效果达到最佳。

例2:FRB121102_tracking-M01_0774

该样本存在集中于1180-1230 MHz的窄带干扰,原始SNR为58.87,例2未去干扰情况下的频率时间图和脉冲轮廓图如图4所示。

对例2使用4种干扰抑制方法处理后的频率时间图和脉冲轮廓图如图5所示。IQR方法SNR提升至98.91,删除了578条通道,同样出现了整段频率的屏蔽现象;K-Means方法SNR提升至103.20,删除了623条通道,虽然增强幅度最大,但存在误删低强度正常通道的问题;FCM方法SNR提升至102.61,删除了607条通道,频谱抑制能力稍逊于K-Means,信号略模糊;HFC方法SNR提升至102.14,删除了507条通道,虽然在数值上略低于K-Means和FCM方法,但保留了更多通道。

从两个例子看出,HFC方法能够在提升信噪比

的同时有效避免误删; IQR 方法虽信噪比提升明显, 但干扰剔除策略简单粗暴, 存在信息丢失风险; K-Means 与 FCM 方法可提供较大幅度的 SNR 提升, 但

容易误删有效通道; HFC 方法在实际应用中更具鲁棒性和工程推广价值, 是当前脉冲星搜寻中值得优先考虑的 RFI 去除技术。

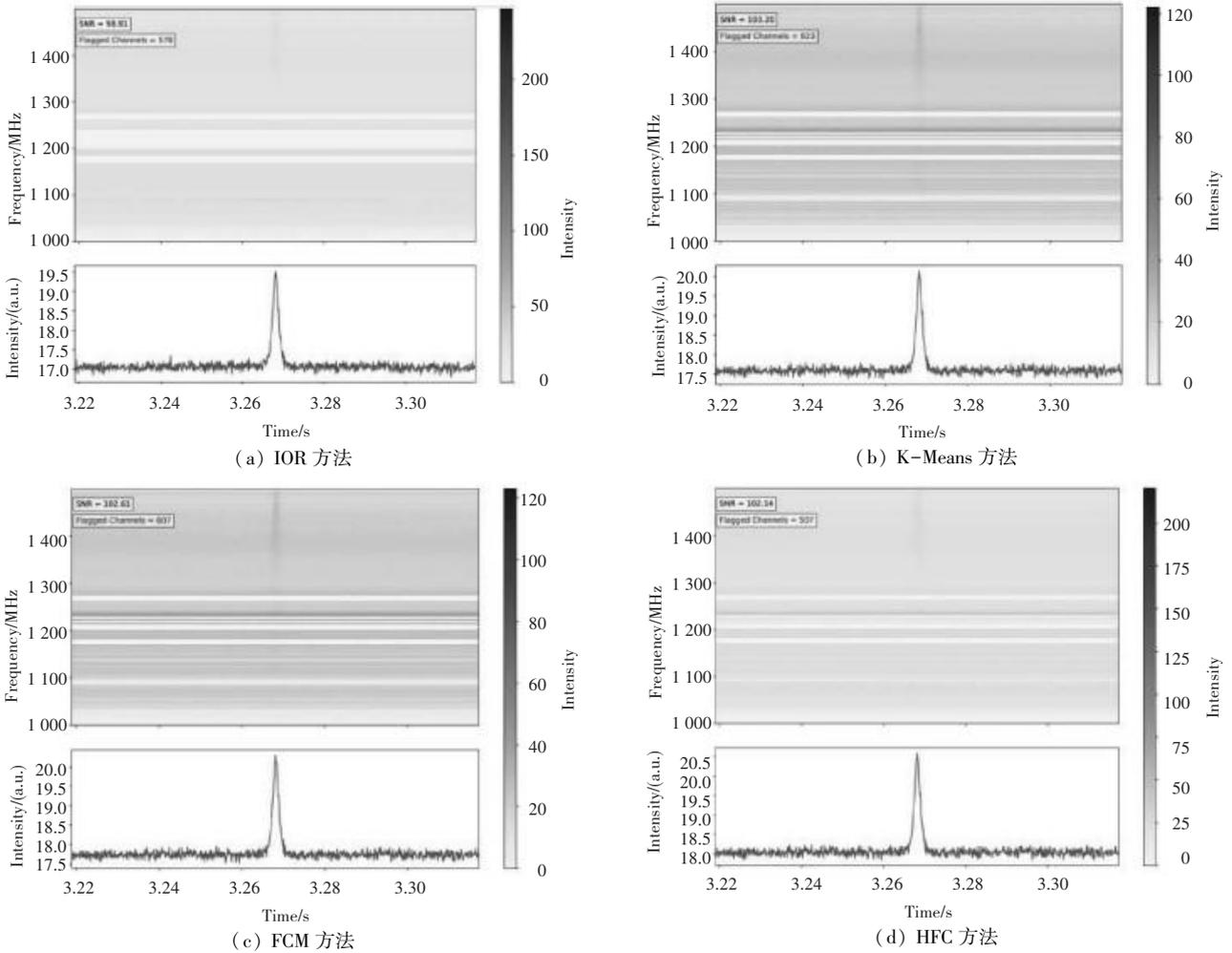


图 5 对例 2 使用 4 种干扰抑制方法处理后的频率时间图和脉冲轮廓图

Fig. 5 Frequency-time diagrams and pulse profiles of Example 2 after applying four different interference suppression methods

4 结束语

本文针对 FAST 望远镜脉冲星观测数据中的 RFI 问题, 比较了 4 种去干扰方法即 IQR、K-Means、FCM 和 HFC 方法。在统一实验框架下, 以信噪比提升和通道删除数量为评估指标开展了系统测试。实验结果表明, HFC 方法在信号保留与干扰抑制之间实现了更优平衡, 表现出较强的鲁棒性和实用性。

下一步研究将面向点级干扰识别, 结合 HFC 的模糊建模能力, 实现对频率-时间图中局部异常点的精准识别与抑制。

参考文献

[1] 王双强, 王娜, 张承民, 等. 特殊类型脉冲星的研究进展[J].

天文学进展, 2018, 36(1): 1-16.

- [2] LORIMER D R, KRAMER M. Handbook of Pulsar Astronomy [M]. Cambridge, UK: Cambridge University Press, 2012: 301.
- [3] OFFRINGA A R, VAN DE GRONDE J J, ROERDINK J, et al. A morphological algorithm for improving radio-frequency interference detection[J]. Astronomy & Astrophysics, 2012, 539: A95.
- [4] 穆宝胜, 刘欣, 朱文艳. 基于 n 个标准差法和箱线图法识别变形监测中异常值的应用探究[J]. 南通职业大学学报, 2023, 37(2): 100-104.
- [5] WOLFAARDT C J. Machine learning approach to radio frequency interference (RFI) classification in radio astronomy [D]. Stellenbosch: Stellenbosch University, 2016.
- [6] AKERET J, CHANG C, LUCCHI A, et al. Radio frequency interference mitigation using deep convolutional neural networks [J]. Astronomy and Computing, 2017, 18: 35-39.
- [7] FRIDMAN P A, BAAN W A. RFI mitigation methods in radio astronomy [J]. Astronomy & Astrophysics, 2001, 378(1): 327-

- 344.
- [8] FRATER R H, BROOKS J W, WHITEOAK J B, et al. The Australia telescope - overview [J]. *Journal of Electrical and Electronics Engineering Australia*, 1992, 12(2): 103-112.
- [9] HESSELS J W T, RANSOM S M, STAIRS I H, et al. A 1.4 GHz Arecibo survey for pulsars in globular clusters [J]. *The Astrophysical Journal*, 2007, 670(1): 363.
- [10] LORIMER D R, KRAMER M, MÜLLER P, et al. A 1400-MHz pilot search for young pulsars [J]. *Astronomy & Astrophysics*, 1999, 570(1): 1-18. DOI:10.1051/0004-6361/201424495
- [11] 吴盛殷, 南仁东, 彭勃, 等. FAST 计划的现状和期望[C]//中国电子学会第七届学术年会论文集. 2001: 57-62.
- [12] KRUESI L. Astronomers mourn Arecibo collapse [J]. *Physics World*, 2021, 34(1): 8.
- [13] 姜鹏, 刘彬, 于东俊, 等. 中国天眼工程升级计划——FAST 核心阵设计和关键技术研究[J]. *科技导报*, 2025, 43(5): 55-63.
- [14] HOTAN A W, VAN STRATEN W, MANCHESTER R N. PSRCHIVE and PSRFITS: An open approach to radio pulsar data storage and analysis[J]. *Publications of the Astronomical Society of Australia*, 2004, 21(3): 302-309.
- [15] 孙吉贵, 刘杰, 赵连宇. 聚类算法研究[J]. *软件学报*, 2008(1): 48-61.
- [16] SONG Y, LIU Z, WANG N, et al. An algorithm for mitigating transient RFI in pulsar observation[J]. *The Astrophysical Journal*, 2021, 922(2): 94.
- [17] WANG N, MANCHESTER R N, ZHANG J, et al. Pulsar timing at Urumqi astronomical observatory: Observing system and results [J]. *Monthly Notices of the Royal Astronomical Society*, 2001, 328(3): 855-866.
- [18] 杨俊闯, 赵超. K-means 聚类算法研究综述[J]. *计算机工程与应用*, 2019, 55(23): 7-14.
- [19] 宫改云, 高新波, 伍忠东. FCM 聚类算法中模糊加权指数 m 的优选方法[J]. *模糊系统与数学*, 2005(1): 143-148.
- [20] 陈娜. 犹豫模糊环境下的决策方法及聚类算法研究[D]. 南京: 东南大学, 2015.