

# 宽带频谱序列干扰信号识别与统计方法\*

王利云<sup>1</sup> 刘奇<sup>2†</sup> 陈卯蒸<sup>2</sup> 刘晔<sup>2</sup> 王玥<sup>2</sup> 朱春花<sup>1‡</sup>  
王洋<sup>2</sup>

(1 新疆大学物理科学与技术学院 乌鲁木齐 830046)

(2 中国科学院新疆天文台 乌鲁木齐 830011)

**摘要** 随着科学技术的不断进步,射电天文台站趋于自动化,各类电子设备的广泛使用使得射电天文台站的电磁环境变得尤为复杂,如何有效识别和统计复杂频谱中的干扰信号是当前射电天文台站亟需解决的问题,故提出一种宽带频谱序列干扰信号识别与统计方法.首先,对每组宽带频谱进行信噪分离、识别频谱中的干扰信号;然后,对第1组宽带频谱信号识别结果及信号特征建立模板库,后续每组频谱的信号识别结果与模板库中对应频率的信号进行相似性分析,根据相似性分析结果,统计信号次数,更新模板库;实现宽带频谱序列干扰信号的识别与统计.针对QTT (QiTai Radio Telescope)台站实测频谱,运用该方法进行干扰信号识别与统计,能够有效识别并标记频谱中的干扰信号,并统计干扰信号随时间、方向的变化趋势.

**关键词** 射电天文, 频谱序列, 信号识别, 信号统计

**中图分类号:** P161; **文献标识码:** A

## 1 引言

射电天文望远镜具有极高的灵敏度,可捕捉来自宇宙的微弱信号,但极易受到其他业务(如:通讯业务、广播业务等)的干扰,且随着科学技术的不断进步,台址内各类电子、商用设备使得台站电磁环境变得越来越复杂<sup>[1]</sup>,导致射电天文台站的频谱分析和射频干扰信号识别、标记、统计等工作尤为困难和重要.

射电天文望远镜建设前期会选择合适的观测台址并建立射电宁静区,建设及运行阶段将对台址内各类电子设备进行屏蔽,可以缓解甚至消除台址内电磁干扰的影响,台址外电磁干扰依据射电宁静区管理办法进行协调.平方千米阵列(Square Kilometer Array, SKA)和中国FAST (Five hundred Meters Aperture Spherical Telescope)都在射电天文

2019-02-13收到原稿, 2019-05-05收到修改稿

\*国家自然科学基金项目(11473061, 11763007)、国家重点基础研究发展计划项目(2015CB857100)、中国科学院“西部之光”人才培养引进计划、中国科学院天文台站设备更新及重大仪器设备运行专项经费支持项目、中国科学院天文大科学研究中心前瞻课题资助

<sup>†</sup>liuqi@xao.ac.cn

<sup>‡</sup>chunhuazhu@sina.cn

台站周围建立电磁环境监测站<sup>[2-3]</sup>, 以便对监测频谱中干扰进行识别, 并为确定干扰源及排除干扰源提供依据. 然而, 监测频谱数量随着时间增加, 宽带频谱中信号繁杂, 降低了人工识别信号的准确度, 也增加了信号在时间、方向和极化上的统计难度. 由此可见, 宽带频谱序列干扰信号的识别及统计方法的研究, 对射电天文台站射频干扰缓解意义重大.

现有的信号识别方法主要有统计模式识别算法、似然函数比较分类识别算法、基于卷积神经网络(Convolutional Neural Networks, CNN)和长短期记忆(Long Short Term Memory, LSTM)的瞬态射频干扰分类方法等. 统计模式识别算法是从已知信号样本中提取特征参数并给定误差对信号进行识别, 该方法在低信噪比的情况下难以提取信号特征参数, 识别能力低<sup>[4-5]</sup>. 似然函数比较分类识别算法是通过信号的似然函数进行处理, 得到用于分类的充分统计量, 但该方法需要许多先验知识, 否则将导致计算量大, 难以实时处理<sup>[5]</sup>. 基于卷积神经网络和长短期记忆的瞬态射频干扰分类方法是针对时域瞬态信号分类, 该方法需要事先建立数据样本库<sup>[1]</sup>.

综上所述, 统计模式识别方法、似然函数比较分类识别算法、基于卷积神经网络和长短期记忆的瞬态干扰分类方法都需要先验知识, 很难对宽带频谱中的干扰进行实时识别及统计. 然而, 随着数字技术和无线通讯技术的发展, 信号种类日益增多, 信号特征越来越复杂, 人工识别信号和提取信号特征的难度越来越大, 人工建立数据库也越来越困难, 在后续统计过程中不易实时更新数据库.

本文针对宽带频谱序列, 提出了一种干扰信号识别及统计方法. 该方法处理流程如图1所示: (1)针对获取的第1组频谱进行信噪分离(Signal Noise Separation, SNS), 检测和识别宽带频谱中的各类信号, 将信号特征信息存入信号模板库; (2)对获取的后续宽带频谱 $I$  ( $I = 1, 2, \dots$ ), 对每组宽带频谱进行信噪分离, 检测和识别宽带频谱中的各类信号, 检测识别后的信号与模板库中对应频率的信号进行相似性分析、统计信号次数、更新模板库, 实现频谱序列中干扰信号的识别与统计.

## 2 信号处理方法

### 2.1 需求分析

实现宽带频谱序列中干扰信号的识别与统计需确定如下信号处理问题:

- (1)实现宽带频谱中信号和噪声的有效分离, 以便进一步识别信号;
- (2)研究信号识别方法, 确保宽带频谱内的各类信号有效识别;
- (3)依据识别的干扰信号, 提取信号关键特征信息, 建立信号模板库;
- (4)选取信号相似性分析方法, 分析频谱序列中干扰信号与信号模板库中信号的相似性, 统计信号次数、更新信号模板库, 实现宽带频谱序列中干扰信号的识别和统计.

### 2.2 信噪分离方法

信号分离方法采用频谱噪声统计法<sup>[6]</sup>, 该方法在无线电频谱监测中应用较多, 且算法简单, 文献<sup>[7]</sup>对该方法进行优化, 数据处理流程如下.

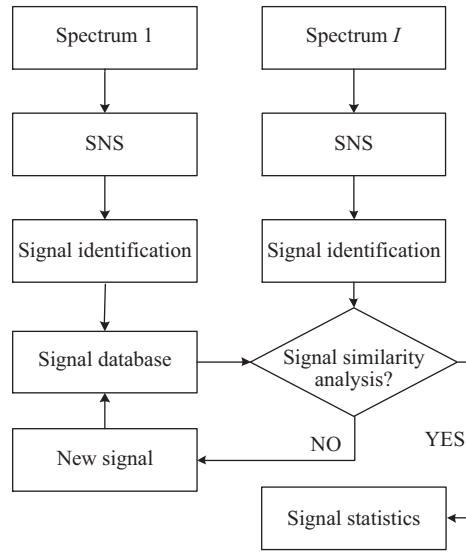


图1 频谱序列干扰识别及统计流程

Fig. 1 Interference identification and statistical process of the spectral sequences

步骤1: 频段划分. 对于给定的2维频谱信号 $\mathbf{P}\{\mathbf{F}, \mathbf{V}\}$ 进行频段划分, 每个频段包含 $m$ 个样本点. 其中 $\mathbf{F}$ 表示频率信息,  $\mathbf{V}$ 表示幅度信息, 频谱信号 $\mathbf{P}$ 包含 $n$ 个样本点. 分频段后的频谱信号 $\mathbf{P}$ 可表示为:

$$\mathbf{P} = \{S_1, S_2, \dots, S_i, \dots, S_{k_1}\}, \quad (1)$$

其中,  $k_1$ 是频段数,  $k_1 = \lceil \frac{n}{m} \rceil$ ,  $\lceil \cdot \rceil$ 表示向左取整,  $i = 1, 2, \dots, k_1$ .

步骤2: 计算信噪分离阈值. 对每个频段 $S_i$ 幅度值从小到大排序, 计算前20%幅度值的平均值 $t_{\text{mean}}^i$ , 则每个频段的阈值可表示为:

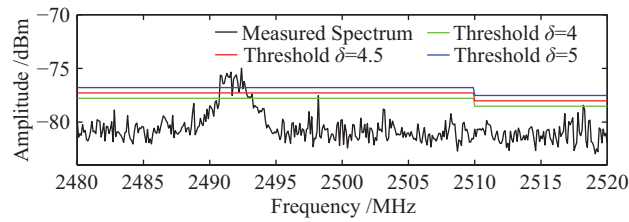
$$t_{\text{threshold}}^i = t_{\text{mean}}^i + \delta, \quad (2)$$

其中,  $\delta$ 为修正系数, 文献[6]中建议取值为 $\delta \in [3, 5]$ . 频谱中每个频点对应的信噪分离阈值 $t_s$ 为:

$$t_s = t_{\text{threshold}}^i, i = \left\lceil \frac{s}{m} \right\rceil, \quad (3)$$

其中,  $s = 1, 2, \dots, n$ .

对实测数据统计分析, 频谱中信号最大带宽为30 MHz, 结合文献[7]设置频谱带宽为30 MHz. 如(2)式中 $\delta$ 设置过大, 影响微弱信号的检测能力,  $\delta$ 设置过小可能将噪声误认为信号; 故针对实测的频谱样本进行分析, 分析结果如图2所示, 结合文献[8]中对 $\delta$ 的推荐值, 取 $\delta = 4.5$ .

图 2 修正系数 $\delta$ 与信噪分离阈值的关系Fig. 2 Relationship between correction factor  $\delta$  and SNS threshold

### 2.3 信噪识别方法

传统的信号识别方法中, 大于信噪分离阈值的认为是信号, 小于阈值的认为是噪声; 这种方法忽略了信噪分离阈值对单个干扰信号的影响, 存在将一个信号误认为是多个信号的可能. 如图3给出了实测频谱中的信号样本, 可以看出传统的信号识别方法可能将一个信号误判为多个信号, 故本文引入边界修正系数 $\alpha$ , 提高信号识别的准确性. 信号识别流程方法如下.

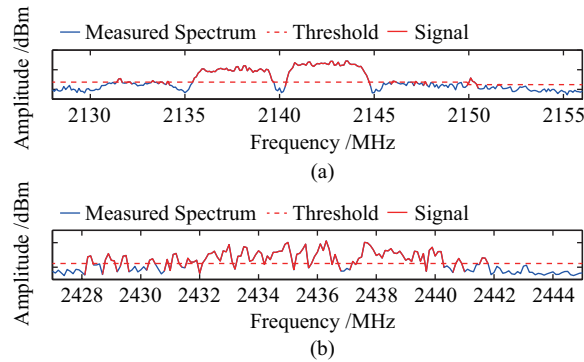


图 3 传统的信号识别方法

Fig. 3 Traditional signal recognition

步骤1: 寻找小于SNS阈值的波谷 $\mathbf{N}$ :

$$\mathbf{N} = \{N_1, N_2, \dots, N_r, \dots, N_t\}, \quad (4)$$

(4)式中,  $r = 1, 2, \dots, t$ .  $\mathbf{N}$ 为一维矩阵, 表示小于SNS阈值的波谷的采样点序号集合,  $N_r$ 表示小于SNS阈值的波谷的采样点序号,  $t$ 为该频谱中小于SNS阈值波谷的数量.

步骤2: 寻找频谱与SNS阈值的交点. 用频谱 $\mathbf{P}$ 幅度值序列与对应的SNS阈值比较, 如果第 $s$ 个幅度值小于等于对应的SNS阈值, 且第 $s + 1$ 个幅度值大于对应的SNS阈值, 则第 $s + 1$ 个采样点作为信号的左突变点; 如果第 $s$ 个幅度点大于对应的SNS阈值, 且第 $s + 1$ 个幅度点小于等于对应的SNS阈值, 则将第 $s$ 个采样点作为信号的右突变点.

$$\begin{cases} L_j = s + 1 & t_s \geq v_s \text{ 且 } v_{s+1} > t_{s+1}, \\ R_j = s & v_s > t_s \text{ 且 } t_{s+1} \geq v_{s+1}, \end{cases} \quad (5)$$

$$\mathbf{L} = \{L_1, L_2, \dots, L_j, \dots, L_l\}, \quad (6)$$

$$\mathbf{R} = \{R_1, R_2, \dots, R_j, \dots, R_l\}, \quad (7)$$

(5)式中,  $j = 1, 2, \dots, l$ .  $l$ 为频谱中干扰信号左(右)突变点数,  $v_s$ 表示第 $s$ 个采样点对应的幅度值,  $v_{s+1}$ 表示第 $s + 1$ 个采样点对应的幅度值;  $t_s$ 表示第 $s$ 个采样点对应的SNS阈值,  $t_{s+1}$ 表示第 $s + 1$ 个采样点对应的SNS阈值, (6)–(7)式中,  $\mathbf{L}$ 为一维矩阵, 表示信号左突变点,  $L_j$ 表示左突变点的样本点序号;  $\mathbf{R}$ 为一维矩阵, 表示信号的右突变点,  $R_j$ 表示右突变点的样本点序号.

步骤3: 边界整合. 通过判断(6)–(7)式中左右突变点之间的频率值间的差值与修正系数 $\alpha$ 的关系, 确定信号提取时的信号边界, 以此来降低信噪分离阈值对信号提取的影响. 当 $j > 1$ 时, 如果第 $j$ 个左突变点的频率值减去信号边界阈值 $\alpha$ 小于等于第 $j - 1$ 个右突变点的频率值时, 则从 $\mathbf{L}$ 中删除第 $j$ 个左突变点, 从 $\mathbf{R}$ 删除第 $j - 1$ 个右突变点.

$$\begin{cases} L_j = 0 \\ R_{j-1} = 0 \end{cases} \quad f_{L_j} - \alpha \leq f_{R_{j-1}}, \quad (8)$$

(8)式中,  $L_j = 0$ 与 $R_{j-1} = 0$ 表示删除并非赋值为0.  $f_{L_j}$ 表示第 $L_j$ 个左突变点对应的频率值,  $f_{R_{j-1}}$ 表示第 $R_{j-1}$ 个右突变点对应的频率值.

步骤4: 确定信号边界. 将离左(右)突变点最近的波谷频率对应的信号作为信号的左(右)边界. 如果第 $r$ 个波谷采样点序号小于等于第 $j$ 个左(右)突变点, 且第 $r + 1$ 个波谷大于等于(大于)第 $j$ 个左(右)突变点, 则将第 $(r + 1)$ 个波谷频率作为信号的左(右)边界.

$$\begin{cases} L_k = N_r & N_r < N_j \leq N_{r+1}, \\ R_k = N_{r+1} & N_r \leq R_j < N_{r+1}, \end{cases} \quad (9)$$

(9)式中,  $k = 1, 2, \dots, e$ ,  $e$ 为该频谱中包含的信号个数.

用 $\mathbf{H}$ 向量表示对应频谱 $\mathbf{P}$ 中包含的信号, 识别出的信号用 $\mathbf{H}_o$ 表示. 则有

$$\mathbf{H} = \{\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_o, \dots, \mathbf{H}_e\}, \quad (10)$$

$$\mathbf{H}_o = \{(f_{L_o}, v_{L_o}), \dots, (f_{R_o}, v_{R_o})\}, \quad (11)$$

(11)式中,  $L_o$ 表示第 $o$ 个信号的起始频率在频谱 $\mathbf{P}$ 中对应的序号,  $R_o$ 表示第 $o$ 个信号的终止频率在频谱 $\mathbf{P}$ 中对应的序号.

(8)式中引入边界修正系数 $\alpha$ ,  $\alpha$ 设置过小, 导致1个信号被误认为多个信号;  $\alpha$ 设置过大将相邻的多个信号及噪声被误判为1个. 对实测样本数据进行分析, 取 $\alpha = 3$  MHz时, 信号误判率较低, 表1给出了随机3组频谱在不同 $\alpha$ 下信号识别的个数, 可以看出, 本文的信号识别方法与传统信号识别方法相比, 减少了将多个信号误判为1个信号的概率.

表 1 不同 $\alpha$ 值下信号提取个数  
Table 1 Number of signal extractions under different  $\alpha$  values

	MI <sup>a</sup>	$\alpha$			
		0 MHz	2 MHz	3 MHz	4 MHz
Spectrum 1	9	44	12	9	9
Spectrum 2	10	44	12	11	11
Spectrum 3	3	7	5	4	4

<sup>a</sup> MI is manual identification

## 2.4 信号模板库建立考虑

2.3节给出了宽带频谱的干扰信号识别方法,对于信号识别后获得的各类信号,需提取信号完整的特征信息,存入信号模板库,为进一步的信号相似性分析提供支撑.信号特征信息主要包括信号的起始频率(Starting Frequency, SF)、终止频率(Termination Frequency, TF)、带宽(Bandwidth, B)、幅度(Amplitude, A)、信号方向(Direction, D)、时间(Time, T)、极化(Polarization, P)、信号来源(Source, S)等,其中,信号来源信息获取依据《中华人民共和国无线电频率划分规定》[8].将完整的信号信息存入信号模板库.信号模板库包含数值表和特征信息表,数值表包含信号的幅度向量,特征信息表包含信号的起始频率、终止频率、带宽、方向、极化、时间和来源等信息.信号模板库结构如图4所示.

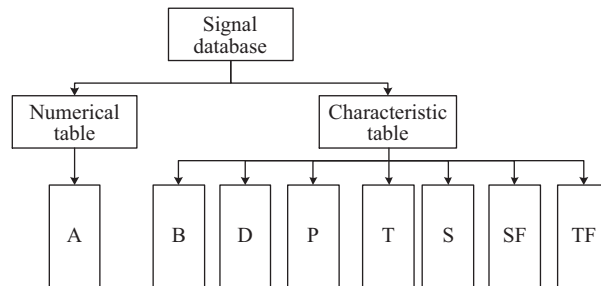


图 4 模板库结构框图

Fig. 4 Structure diagram for template database

## 2.5 信号相似性分析方法

通过信号相似性判断两个信号是否来自于同一信号,采用计算两个信号的皮尔逊系数的方式来分析模板信号 $x$ 与样本信号 $y$ 的相似程度.皮尔逊系数广泛用于度量两个变量之间相关程度,其优点在于:(1)能够区分正相关与负相关;(2)对相关程度的变化敏感;(3)在线性和非线性情况下都只有较小的误差<sup>[9-10]</sup>.皮尔逊系数可表示为:

$$\rho_{x,y} = \frac{\sum x_a y_a - \frac{\sum x_a \sum y_a}{b}}{\sqrt{\left(\sum x_a^2 - \frac{(\sum x_a)^2}{b}\right) \left(\sum y_a^2 - \frac{(\sum y_a)^2}{b}\right)}}, \quad (12)$$

(12)式中,  $b$ 为信号 $x$ 或 $y$ 包含的样本点个数,  $a = 1, 2, \dots, b$ .

皮尔逊系数的绝对值 $|\rho_{x,y}| \in [0, 1]$ ,  $|\rho_{x,y}|$ 的值越趋近于1, 表明信号 $x$ 与信号 $y$ 相似程度越高,  $|\rho_{x,y}|$ 的值越趋近于0, 表明信号 $x$ 与信号 $y$ 相似程度越低. 通过 $\rho_{x,y}$ 分析2个信号的相似性, 确定相似阈值 $\gamma$ , 若 $\rho_{x,y}$ 大于 $\gamma$ , 认为两个信号相似, 反之则不相似.

相似阈值 $\gamma$ 的选取影响信号相似分析的精度, 文献[9-10]建议 $\gamma$ 取0.4; 另外, 我们针对实测样本进行分析, 不同 $\gamma$ 取值信号相似分析的误判数量如图5所示. 综上分析, 相似阈值 $\gamma$ 设置为0.4.

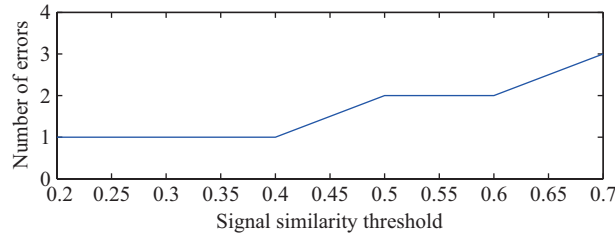


图5  $\gamma$ 取值对信号相似性分析的影响

Fig. 5 Influence of  $\gamma$  on signal similarity analysis

### 3 频谱序列信号统计方法

依据第2节给出的信号处理方法, 频谱序列干扰信号统计方法如下:

步骤1: 建立信号模板库, 即针对获取的第1组宽带频谱, 依照2.2节给出的信号分离方法实现宽带频谱的信噪分离; 按照2.3节提出的信号识别方法实现宽带频谱中各类信号识别; 按2.4节建立模板库的要求, 提取识别出的信号特征信息, 包括起始频率、终止频率、带宽、幅度、信号方向、时间、信号来源存入信号模板库, 完成信号模板库的建立.

步骤2: 信号统计, 统计流程如图6所示. (1)针对后续获取的宽带频谱, 每组宽带频谱经过信号识别, 提取信号的特征信息; (2)依据提取的信号特征信息, 检索信号模板库中是否存在相关信号, 依照2.5节提出的信号相似性分析方法, 计算两个信号的相似性; (3)若识别的信号为模板库中现有信号, 依据信号来源, 统计信号次数; (4)若识别的信号与模板库中现有信号不相关, 提取该信号特征信息, 存入信号模板库.

### 4 实例分析

QTT台址电波环境测量方法依据文献[11]提出的准实时电波环境测量方法, 测量及数据信息如下: 测试天线的3 dB波束宽度为 $60^\circ$ ,  $360^\circ$ 覆盖6个测试方向, 单次测量时间为12 min, 1轮共计6个测量方向, 1 d 24 h共计19轮测量数据; 运用干扰信号统计方法对4 d共456组宽带频谱数据进行统计, 其中一半数据为采用水平极化(Horizontal Polarization, H)方式测量, 另一半为采用垂直极化(Vertical Polarization, V)方式测量. 表2统计了不同方向、采用不同极化方式测量频谱序列中干扰信号的次数; 同时, 针对1 d 24 h水平极化和垂直极化测量数据进行统计, 结果如图7所示. 依据统计结果, 分析如下: (1) 2135.4-2144.6 MHz为中国联通信号(China Union, CU), 全方

向、全时段、采用水平和垂直极化方式均能测量到. (2) 2187.5–2196.1 MHz的瞬态信号(Transient Signal, TS)来自120°和180°方向, 该信号由3–5个带宽为0.3 MHz左右的窄带信号组成, 采用水平极化方式和垂直极化方式均能测量到该信号, 水平极化测量结果强度更为明显, 该信号来源可能为数字通信信号, 需进一步排查信号来源. (3) 2429.1–2475.1 MHz为WLAN信号, 来自60°、120°和180°方向, 该信号工作时间与人们的作息时间相关. (4) 2489.7–2492.7 MHz的点对点无线通信业务(Point-to-point wireless communication, PTPWC)来自120°和180°方向, 该信号为固定信号, 强度相对稳定.

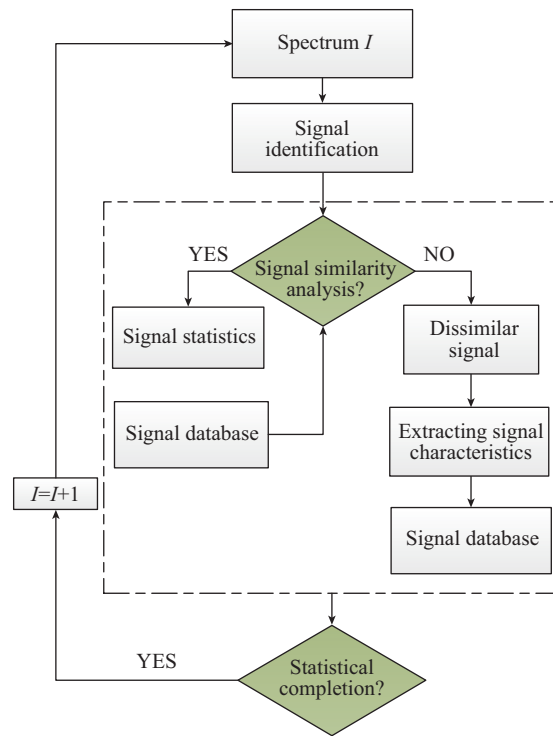


图 6 信号统计流程图

Fig. 6 The flow chart of signal statistics

表 2 频谱序列干扰信号统计

Table 2 Interference signal statistics of spectral sequences

	SF/MHz	TF/MHz	H	V	0°	60°	120°	180°	240°	300°	B/MHz
CU	2135.4	2144.6	228	228	76	76	76	76	76	76	9.2
TS	2187.5	2196.1	61	33	0	0	54	40	0	0	8.6
WLAN	2429.1	2475.1	33	22	4	14	22	13	1	1	46
PTPWC	2489.7	2492.7	96	70	0	22	73	63	8	0	3



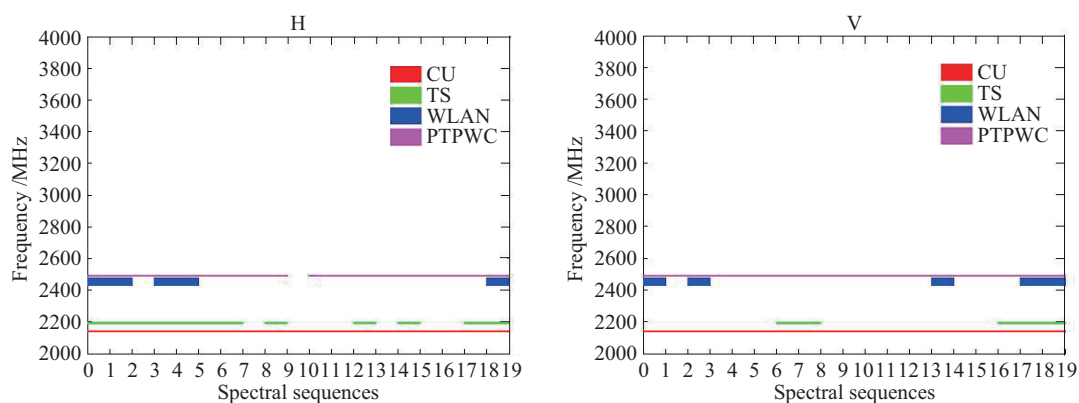


图7 1 d 24 h 频谱序列信号统计

Fig. 7 Signal statistics of 24-hour spectrum sequence in 1 day

## 5 结论

本文基于射电天文台站频谱分析需求, 提出了一种宽带频谱干扰信号识别及统计方法, 该方法可有效统计射电天文台站监测频谱中的干扰信号, 为进一步的干扰信号特征分析和干扰消解提供依据. 但是, 该方法依旧存在以下不足, 需要在未来的工作中改进: (1)信号识别方法需要进一步优化, 确保同一信号识别的一致性; (2)针对信号模板库中信号检索算法需要进一步优化, 提高数据处理效率.

## 参考文献

- [1] Czech D, Mishra A, Inngs M. A&C, 2018, 25: 52
- [2] Schollar C. RFI monitoring for the MeerKAT radio telescope. Cape Town: University of Cape Town, 2015
- [3] 黄仕杰. 500米口径球面射电望远镜(FAST)电磁干扰监测与分析. 贵州: 贵州大学, 2017
- [4] 陈卫东. 数字通信信号调制识别算法研究. 西安: 西安电子科技大学, 2001
- [5] 罗吉. 数字调制信号识别算法研究. 成都: 电子科技大学, 2008
- [6] ITU. Spectrum occupancy measurements and evaluation: ITU-R SM.2256. Geneva: ITU, 2012
- [7] 袁力, 刘奇, 孙正文, 等. 电波科学学报, 2017, 32: 650
- [8] 工业和信息化部无线电管理局. 中华人民共和国无线电频率划分规定. 北京: 人民邮电出版社, 2018
- [9] 彭海. 电子世界, 2017, 7: 163
- [10] 李宏彬, 赫光中, 果秋婷. 化学分析计量, 2015, 24: 33
- [11] 刘奇, 王娜, 王玥, 等. 电波科学学报, 2017, 32: 718

## Identification and Statistics of Interference Signals in Wideband Spectral Sequences

WANG Li-yun<sup>1</sup>   LIU Qi<sup>2</sup>   CHEN Mao-zheng<sup>2</sup>   LIU Ye<sup>2</sup>   WANG Yue<sup>2</sup>  
ZHU Chun-hua<sup>1</sup>   WANG Yang<sup>2</sup>

(1 College of Physical Science and Technology, Xinjiang University, Urumqi 830046)

(2 Xinjiang Astronomical Observatory, Chinese Academy of Sciences, Urumqi 830011)

**ABSTRACT** With the development of science and technology, radio observatories gradually tend to be automated, and the electromagnetic environment of radio observatories becomes more complicated with the introduction of various electronic equipments. How to effectively identify and count the interference signal in a complicated spectrum is an urgent problem to be solved in the radio observatories at present. Therefore, a method of identification and statistics of the interference signals of broadband spectrum sequences is proposed in this paper. Firstly, signal-to-noise separation was performed for each set of wideband spectrum, and signals were identified in the spectrum. Then, the template database was established for the first group of wideband spectrum signal identification results and signal features. The correlation analysis between the signal identification results of each set of spectrum and the corresponding frequency signals in the template database was carried out. According to the correlation analysis results, the number of signals was counted, and the template database was updated. Thus, the signal identification and signal statistics of wideband spectrum interference were realized. According to the measured spectrum of QTT (QiTai Radio Telescope), the interference signal can be identified and counted by this method, and the trend of interference signal varied with time and direction can be calculated.

**Key words** radio astronomy, spectral sequences, signal identification, signal statistics