

基于时间序列分析的频谱异常 自主检测和稳健估计方法

王 磊, 谢树果, 苏东林, 王国玉

(北京航空航天大学电子信息工程学院, 北京 100191)

摘 要: 复杂电磁环境和无用频先验知识条件下有效检测电磁频谱异常使用信息, 是无线电监测和电磁环境评估等领域的重要难题. 本文基于时间序列分析理论, 通过构建反映有限频谱占用度序列动态依存关系且包含频谱异常值的时序模型, 实现对无线电频谱异常的自主检测和稳健估计. 研究结果表明, 该方法无需用频数据库或无线电监测历史数据支持, 能够有效识别典型频谱异常类型、发生时间以及异常影响强度等信息; 同时通过对频谱占用度时序模型的稳健估计, 能够显著降低模型拟合误差, 提高模型对外部干扰环境的适应性和鲁棒性.

关键词: 频谱异常; 自主检测; 稳健估计; 电磁环境

中图分类号: TN98 **文献标识码:** A **文章编号:** 0372-2112 (2014)06-1055-06

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2014.06.003

An Autonomous Detection and Robust Estimation Method of Spectrum Anomaly Based on Time Series Analysis

WANG Lei, XIE Shu-guo, SU Dong-lin, WANG Guo-yu

(School of Electronics and Information Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: How to precisely detect electromagnetic spectrum anomaly is a major challenge for radio monitoring and electromagnetic environment evaluation, especially in the condition of complex electromagnetic environment and lack of pre-knowledge information about frequency use. Based on time series analysis theory, a timing model which presents the correlation between the previous and the following sequence of spectrum occupancy, is built to help us realize the autonomous identification and robust estimation of typical spectrum anomaly. The analysis results indicate that, without actually requiring pre-knowledge of frequency database and radio monitoring historical data support, this method can effectively identify the types of spectrum anomaly, occurrence time, anomaly effect intension and other relative information. Furthermore, through the robust estimation of spectrum occupancy model, we can significantly improve the model's fitting performance and raise the adaptability and robustness of the model to external interferences.

Key words: spectrum anomaly; autonomous detection; robust estimation; electromagnetic environment

1 引言

通过无线电监测实现对电磁频谱非法占用、辐射超标和电磁干扰等异常行为的有效检测和识别, 对于无线电管理和电磁环境评估等具有重要意义. 现代城市、战场等复杂电磁环境中频谱使用行为往往具有很强的随机性和突发性, 导致无线电监测频谱状态表现出显著的多变性和波动性. 传统上对无线电监测中频谱异常进行检测的方法主要有数据库比对法和历史模型分析法, 前者需要较完备的频谱执照数据库支持^[1], 后者需要长期

监测数据积累以及诸多用频参数先验概率分布知识^[2,3], 两者均很难满足复杂电磁环境下对无线电频谱状态异常进行快速自主检测的需求^[4,5].

时间序列分析是一种根据动态数据揭示系统动态结构和规律的统计方法. 它不需要样本概率分布等先验信息支持, 仅依靠对系统有限长度的观察样本数据, 建立能够精确反映序列中所包含动态依存关系的数学模型, 并能够运用时序模型对系统行为进行预测和异常检测^[6,7]. 频谱占用度是无线电监测中反映频谱使用状态的主要统计量, 利用持续的频谱监测数据可以获取频谱

占用度时间序列,文献[8~10]对频谱占用度时序的建模和波动性问题进行了讨论,提出基于自回归移动平均模型的建模方法,但并未对异常干扰下频谱占用度建模和用频状态检测问题开展进一步研究.实际上,构建能够对频谱异常实现自主检测和稳健估计方法,是应用频谱占用度建模理论解决工程问题的关键环节.本文基于时间序列分析理论,选取频谱占用度衡量频谱状态时域变化信息,在构建频谱占用度时间序列数学模型的基础上,提出了一种频谱异常自主检测和稳健估计方法,该方法无需用频先验知识和长期监测数据积累,能够有效识别典型频谱异常类型,给出异常发生时间以及影响强度等信息,并通过对模型的稳健估计显著提高模型对外部干扰的适应性和鲁棒性.该方法对于实现复杂电磁环境下基于无线电监测的频谱资源智能监管和自主检测具有重要应用价值.

2 频谱占用度检测模型

根据时间序列分析理论,对于满足平稳性条件的频谱占用度时间序列 X_t ,可以建立其自回归移动平均模型(Auto Regressive Moving Average model, ARMA),即

$$\begin{cases} X_t = \frac{\theta(B)}{\varphi(B)} a_t \\ \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p \\ \varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_q B^q \end{cases} \quad (1)$$

式(1)中 p 和 q 分别为 ARMA(p, q)模型的自回归和移动平均阶数, a_t 是高斯白噪声随机序列.

设 $Y_t (t=1, 2, \dots, n)$ 表示包含异常值的频谱占用度时间序列,则根据异常值类型可用以下两种模型对式(1)进行修正^[11,12].

(1) 加性异常(Additional Outlier, AO)模型

加性异常定义为只影响该干扰发生的那一时刻 T 上的序列值,而不影响该时刻以后的频谱占用度序列值,加性异常值通常通过未知的影响因子 w 而起作用,且 Y_t 可以表示为

$$Y_t = \frac{\theta(B)}{\varphi(B)} a_t + w I_t^{(T)} = X_t + w I_t^{(T)} \quad (2)$$

加性异常值干扰形式如图 1 所示:

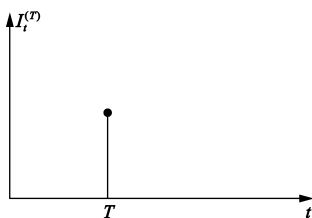


图1 加性异常值干扰示意图

(2) 新息异常(Innovational Outlier, IO)模型

新息异常值定义为影响了 T 时刻及以后的所有频谱占用度序列,且影响的效应与 X_t 的模型形式有关,通过 $\theta(B)/\varphi(B)$ 所描述的系统动态特性而影响后面所有频谱占用度序列. Y_t 可以表示为

$$Y_t = \frac{\theta(B)}{\varphi(B)} (a_t + w I_t^{(T)}) = X_t + w \left(\frac{\theta(B)}{\varphi(B)} \right) I_t^{(T)} \quad (3)$$

其中 w 为影响因子,且

$$I_t^{(T)} = \begin{cases} 1, & t = T \\ 0, & t \neq T \end{cases} \quad (4)$$

新息异常形式如图 2 所示:

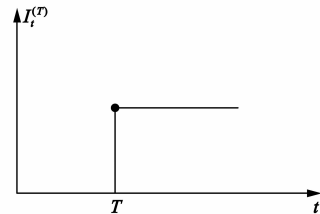


图2 新息异常值干扰示意图

3 基于 ARMA 模型的异常值检测方法

实际无线电监测中,根据是否掌握异常值发生时刻信息,可分别采用如下两种方法进行异常值检测.

3.1 异常值发生时刻已知时的检测方法

假设频谱占用度序列 Y_t 在不同的时间点上受不同类型异常值的影响,由式(1)~(4),可导出包含多个异常值的 Y_t 的 ARMA 模型可以表示为:

$$Y_t = X_t + \sum_{i=1}^l w_i L_i(B) I_t^{(T)} \quad (5)$$

其中

$$L_i(B) = \begin{cases} \frac{\theta(B)}{\varphi(B)}, & \text{IO} \\ 1, & \text{AO} \end{cases} \quad (6)$$

相应的模型残差可以表述为

$$\hat{e}_t = \pi(B) \left\{ Y_t - \sum_{i=1}^l w_i L_i(B) I_t^{(T)} \right\} = \sum_{i=1}^l w_i K_i(B) I_t^{(T)} + a_t \quad (7)$$

其中

$$K_i(B) = \begin{cases} 1, & \text{IO} \\ \pi(B), & \text{AO} \end{cases} \quad (8)$$

$$\pi(B) = \frac{\varphi(B)}{\theta(B)} = 1 - \pi_1 B - \pi_2 B^2 - \dots \quad (9)$$

T 为异常值发生时刻,当 $t = T$ 时,异常值最小估计为

$$\hat{w}_i = \frac{\sum_{t=T}^n \hat{e}_t x_{it}}{\sum_{t=T}^n x_{it}^2} \quad (10)$$

其中:

$$x_{it} = \begin{cases} 0, & t < T, k = 0 \\ 1, & t = T, k = 0 \\ 0, & t > T, k \geq 1, \text{IO} \\ -\pi_k & t > T, k \geq 1, \text{AO} \end{cases} \quad (11)$$

当出现极端的情况 $T = n$ 时, \hat{w} 相同. 它表示当异常值出现在频谱占用度时间序列的最后时, 异常值的类型将无法确定. 异常值检测相关统计量如下:

$$\begin{cases} \hat{T}_{\text{IO}}^{(T)} = \hat{w}_{\text{IO}} / \hat{\sigma}_a \\ \hat{T}_{\text{AO}}^{(T)} = \hat{w}_{\text{AO}} \sqrt{\sum_{i=T}^n x_{2i}^2} / \hat{\sigma}_a \end{cases} \quad (12)$$

式(12)中, $\hat{\sigma}_a$ 为含异常值的残差的标准差, \hat{T} 统计量的极限分布为标准的正态分布, 并服从 t 分布.

3.2 异常值发生时刻未知时的检测方法

该方法分为以下几个步骤:

(1) 以 ARMA 模型建立序列 Y_t 不含异常值的模型, 计算估计出模型的残差

$$\begin{cases} \hat{e}_t = \hat{\pi}(B) Y_t = \frac{\hat{\varphi}(B)}{\hat{\theta}(B)} Y_t \\ \hat{\theta}(B) = 1 - \hat{\theta}_1 B - \hat{\theta}_2 B^2 - \dots - \hat{\theta}_p B^p \\ \hat{\varphi}(B) = 1 - \hat{\varphi}_1 B - \hat{\varphi}_2 B^2 - \dots - \hat{\varphi}_q B^q \end{cases} \quad (13)$$

令 $\hat{\sigma}_a^2 = \frac{1}{n} \sum_{i=1}^n \hat{e}_i^2$ 为 σ_a^2 的初估计;

(2) 用估计出的模型残差 \hat{e}_t 和 $\hat{\sigma}_a^2$, 计算 $T = 1, 2, \dots, n$ 时刻的异常值检测统计量, 再令

$$\eta_j = \max_T |\hat{T}_j^{(T)}|, T = 1, 2, \dots, n; j = \text{IO}, \text{AO} \quad (14)$$

即 η_j 表示 $\hat{T}_{\text{IO}}^{(T)}$ 和 $\hat{T}_{\text{AO}}^{(T)}$ 中的最大值, 如果 $\eta_j > C_j$ (C_j 是针对 AO 和 IO 异常预先设置的常数), 则异常值发生时刻在 T_j , 异常值类型由 j 决定, 因为每个 j 都对应一个相应异常值类型. 然后利用残差表达式对模型进行修正, 得到相应的残差 \hat{e}_t^1 , 若没有发现异常值, 即 $\eta_j < C_j$, 则直接进入第(4)步, 否则进入第(3)步;

(3) 对修正后的 \hat{e}_t^1 和 σ_a^2 , 再计算 $\hat{T}_{\text{IO}}^{(T)}$ 和 $\hat{T}_{\text{AO}}^{(T)}$ 中的最大值, 重复第(2)步骤, 直到最终没有新的异常值被发现.

(4) 通过第(2)、(3)步骤处理后, 假定已找出当前模型下干扰发生的时刻分别为 T_1, T_2, \dots, T_m 的 m 个异常值, 剔除这些异常值影响, 可以重新估计一个新的含异常值的模型, 针对新的模型重复前面的步骤, 再找出是否有新的异常值出现, 直到所有 k 个异常值都被找出, 最后进行联合估计, 得出含异常值的拟合模型.

4 异常检测算法性能分析

检测算法性能分析是根据典型数据集和应用场景, 对上述基于 ARMA 模型的异常检测算法的有效性和性能进行仿真实验分析, 定量评估其对典型异常类

型的检测效果, 为算法实际应用提供先决条件和基础支撑.

4.1 数据来源和仿真参数设置

实验数据选择作者前期监测获取的 CDMA 下行业务(870 ~ 885MHz)某两天的监测数据及相应的频段占用度序列(SOB), 异常类型包括加性异常(AO)、新息异常(IO)以及两者混合三种模式. 三种模式仿真中 AO 和 IO 的幅值均设为 0.09, 加性异常模式中 AO 发生时刻从 20 到 400, 发生时刻间隔以 2 个采样点步进; 新息异常模式中 IO 的发生时刻从 10 到 400, 持续时长均为 3 个采样点, 发生时刻均以 5 个采样点为单位步进; 混合模式中 AO 和 IO 发生时刻分别从 20 到 400 和从 10 到 400, 发生时刻均以 5 个采样点为单位步进. 如图 3 和图 4 所示. 为衡量算法对异常检测的准确性, 用正确检测率表示算法正确检测到异常发生时刻及类型的概率; 用误检率表示算法正确检测到异常发生的时刻但未正确判别异常类型的概率; 用漏检率表示算法既未正确获取异常发生的时刻又未正确判别异常类型的概率.

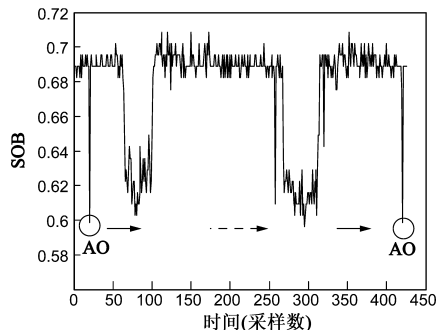


图3 AO模式仿真参数设置示意图

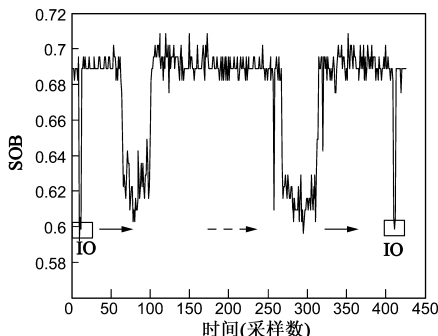


图4 IO模式仿真参数设置示意图

4.2 仿真结果及分析

假设异常发生时刻未知, 根据上述数据来源和仿真参数设置, 根据 3.2 中提出的基于 ARMA 模型的频谱占用度异常检测算法, 表 1 给出了所选取的 SOB 序列分别在 AO 异常、IO 异常以及两者混合模式下异常检测的结果. 为便于比较分析本算法性能, 表 1 中同时给出了基于小波系数分析方法对相同序列异常检测结果,

10月18日凌晨2:04)开始无发射信号,图6(c)中879~883MHz在1330时刻(对应监测时间为2011年10月18日晚上21:57)信号电平明显提高.可见,使用本文所提出的方法能够准确识别实际无线电频谱状态变化的时间和类型,并定量给出频谱异常行为的影响强度信息.

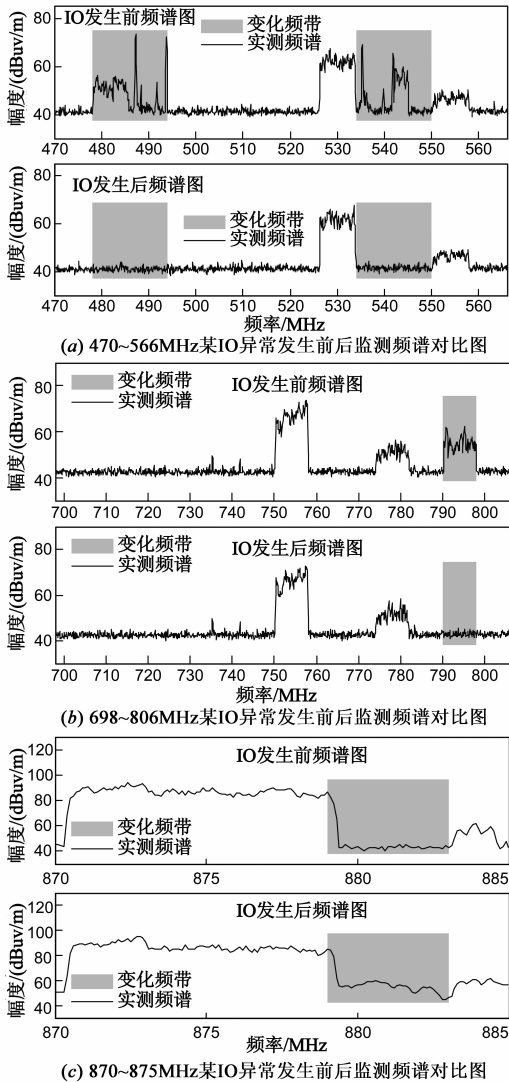


图6 无线电业务频段监测频谱异常变化示意图

5 频谱占用度模型的稳健估计

对含有异常值的时间序列进行建模和分析研究,有时也称为稳健估计(robust estimation)方法,该方法由Box和Anderson首先提出^[11,12],后来Masarotto提出了稳健样本偏自相关和自相关函数^[13].对频谱占用度的稳健估计是频谱占用度异常检测研究的进一步深化,这里着重分析两个内容,一是异常值存在条件下频谱占用度时间序列模型的精确构建,二是异常值剔除后频谱占用度模型误差的改进量.

5.1 频谱占用度模型的稳健估计表达式

以基于ARMA模型的频谱占用度异常检测分析为例.对于一个频谱占用度序列,假定已找出当前模型下发生的时点分别为 T_1, T_2, \dots, T_m 的 m 个异常值,剔除这些异常值的影响,我们可以重新估计一个新的含异常值的模型,针对新的模型重复前面的步骤,再找出是否有新的异常值出现,直到所有 k 个异常值都被找出,最后,进行联合估计,得出拟合的异常值模型

$$Y_t = \sum_{j=1}^k \hat{w}_j L_j(B) I_t^{(T_j)} + \frac{\hat{\theta}(B)}{\hat{\varphi}(B)} a_t \quad (15)$$

其中 Y_t 为含有异常值的频谱占用度序列, \hat{w}_j 为估计出的第 j 个异常值(AO或IO)的权重, $L_j(B)$ 为反映第 j 个异常值影响的时间效应的系数, $\hat{\theta}(B) a_t / \hat{\varphi}(B)$ 为最后一次迭代中为 Y_t 构建的ARMA模型结构.

如对于电视13~24频道(470~566MHz)SOB序列,由4.3中对频谱占用度序列的异常检测结果,可以求出模型各参数估计量,在此基础上,可以得到其稳健估计模型为

$$Y_t = \hat{w}_1 I_t^{(1171)} + \hat{w}_2 I_t^{(1087)} + \hat{w}_3 I_t^{(665)} + \hat{w}_4 I_t^{(1144)} + \hat{w}_5 I_t^{(875)} + \hat{w}_6 I_t^{(1151)} + \hat{w}_7 \frac{1}{\hat{\varphi}(B)} I_t^{(1153)} + \hat{w}_8 \frac{1}{\hat{\varphi}(B)} I_t^{(1298)} + \hat{w}_9 \frac{1}{\hat{\varphi}(B)} I_t^{(876)} + \frac{1}{\hat{\varphi}(B)} a_t \quad (16)$$

将各参数的估计量应用到式(16)中后得到:

$$\begin{cases} Y_t = -0.14 I_t^{(1171)} + 0.13 I_t^{(1087)} + 0.13 I_t^{(665)} - 0.12 I_t^{(1144)} \\ - 0.12 I_t^{(875)} + 0.11 I_t^{(1151)} - 0.14 \frac{1}{\hat{\varphi}(B)} I_t^{(1153)} \\ + 0.13 \frac{1}{\hat{\varphi}(B)} I_t^{(1298)} + 0.13 \frac{1}{\hat{\varphi}(B)} I_t^{(876)} + \frac{1}{\hat{\varphi}(B)} a_t \\ \hat{\varphi}(B) = 1 - 0.8266 B - 0.1394 B^2 \\ a_t \sim \text{NID}(0, \sigma_a^2) \end{cases} \quad (17)$$

同理,根据其它序列模型参数的估计值和异常值,可以得到相应的频谱占用度稳健估计模型表达式.

5.2 基于稳健估计的模型误差减小量

频谱占用度稳健估计模型另一个优点是,经过异常值检测后,通过将异常值带来的序列值偏差剔除,从而得到相应的稳健估计模型,使得模型的误差值得到一定程度的减小,对于频谱占用度的ARMA模型,该误差表现为ARMA模型的残差的平均平方和.通过4.3中对电视13~24频道(470~566MHz)、电视36~49频道(698~806MHz)以及CDMA下行业务(870~885MHz)三个无线电业务频段SOB建模分析,可以得到其异常值检测(剔除)前后模型估计误差的对比量如表3所示.由该表可以得出,经过异常检测和异常值剔除后,频谱占用度模型的误差减小非常明显,证明经过稳健估计修正之后模型的拟合精度得到显著提高,对外部干扰的

适应性进一步增强。

表 3 三个频段频谱占用度稳健估计模型误差减小量

| 误 差 无 线 电 业 务 | 异常检测 前误差 | 异常检测 后误差 | 模型误差 减少率 |
|---------------------------------|-----------------------|-----------------------|-------------|
| 电视 13 ~ 24 频道 (470 ~ 566MHz) | 5.10×10^{-4} | 4.45×10^{-4} | 12.8% |
| 电视 36 ~ 49 频道 (698 ~ 806MHz) | 6.77×10^{-5} | 4.20×10^{-5} | 38.01% |
| CDMA 下行业务 (870 ~ 885MHz) | 1.95×10^{-4} | 1.03×10^{-4} | 46.96% |

6 结论

传统无线电频谱异常检测方法需要基于用频执照数据库和诸多先验信息支持,应用范围受到很大限制.本文面向复杂电磁环境下无线电监测系统自主快速频谱异常检测需求,提出一种基于时间序列分析理论的无线电频谱异常自主检测和稳健估计方法.研究表明,本方法无需用频先验知识和长期监测数据积累,能够依靠频谱监测序列之间的动态依存关系自主识别典型频谱异常类型,给出异常行为信息,并基于对模型的稳健估计实现外部干扰条件下频谱占用度时序模型的精确构建.该方法对提高无线电频谱异常检测中的智能性、建立鲁棒的频谱正常使用模型库具有重要应用价值.

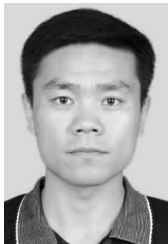
参考文献

- [1] 周鸿顺. 频谱监测手册[M]. 北京: 人民邮电出版社, 2006. 9-10.
- [2] Tao Li, Wai Ho Mow, Vincent K N Lau. Robust joint interference detection and decoding for OFDM-based cognitive radio systems with unknown interference[J]. IEEE Journal on Selected Areas in Communications, 2007, 25(3): 566-575.
- [3] Yin Sixing, Li Shufang. Temporal-spectral data mining in anomaly detection for spectrum monitoring[A]. Proceedings of International Conference on Wireless Communications, Networking and Mobile Computing[C]. Beijing, China: IEEE Press, 2009. 1-5.
- [4] Androulidakis G, Papavassiliou S. Improving network anomaly detection via selective flow-based sampling[J]. IET Communications, 2008, 2(3): 399-409.
- [5] Androulidakis G, Chatzigiannakis V, Papavassiliou S. Using selective sampling for the support of scalable and efficient network anomaly detection[A]. Proceedings of International Conference on GLOBECOM[C]. Washington, DC, USA: IEEE Press, 2007. 1-5.
- [6] 王振龙, 胡永宏. 应用时间序列分析[M]. 北京: 科学出版

社, 2007. 1-10.

- [7] 王黎明, 王连, 杨楠. 应用时间序列分析[M]. 北京: 复旦大学出版社, 2008. 1-10.
- [8] Sixing Yin, Dawei Chen, Qian Zhang, et al. Mining spectrum usage data: a large-scale spectrum measurement study[J]. IEEE Transactions on Mobile Computing, 2012, 11(6): 1033-1046.
- [9] Ali Gorcin, Hasari Celebi, Khalid A. Qaraq, Huseyin Arslan. An autoregressive approach for spectrum occupancy modeling and prediction based on synchronous measurements[A]. Proceedings of 22th International Symposium on Personal, Indoor and Mobile Radio Commission[C]. Toronto, Canada: IEEE Press, 2011. 705-709.
- [10] 王磊, 苏东林, 谢树果, 王国玉. 基于 EGARCH 过程的电磁频谱占用状态波动特性分析[J]. 电子与信息学报, 2012, 34(11): 2767-2773.
Wang Lei, Su Dong-lin, Xie Shu-guo, Wang Guo-yu. Electromagnetic spectrum occupancy state volatility analysis based on EGARCH process[J]. Journal of Electronics & Information Technology, 2012, 34(11): 2767-2773. (in Chinese)
- [11] George E P Box, Gwilym M Jenkins, Gregory C Reinsel. Time Series Analysis: Forecasting and Control[M]. New York: John Wiley & Sons, 2008. 313-323.
- [12] William W S Wei. Time Series Analysis: Univariate and Multivariate Methods[M]. New Jersey: Pearson-Addison Wesley, 2006. 204-226.
- [13] Masarotto G. Robust identification of autoregressive moving average models[J]. Applied Statistics, 1987, 36(2): 214-220.

作者简介



王磊 男, 1981 年 12 月出生于陕西省渭南市. 现为北京航空航天大学电子信息工程学院博士生. 主要研究方向为电磁频谱管理、电磁兼容和电磁环境等.

E-mail: wlei@ee.buaa.edu.cn



谢树果 男, 1968 年 7 月出生于安徽省合肥市. 2001 年毕业于武汉大学无线电物理专业. 现为北京航空航天大学电子信息工程学院教授、博士生导师. 主要研究方向为电波传播、天线及雷达信号处理、复杂系统电磁兼容设计与评估、电磁兼容测试等.

E-mail: xieshuguo@buaa.edu.cn