第7卷 第4期 空间科学学报 Vol. 7, No. 4
1987 年 10 月 GHINESE JOURNAL OF SPACE SCIENCE Oct., 1987

用开环门限自回归分析方法探讨短波衰退事件的预报

张桂清 王艳琳
（中国科学院北京天文台）

摘 要

本文基于电离层突然骚扰对强太阳活动的依赖关系，用开环门限自回归分析方法，以每日最大 2800MHz 射电爆发峰值流量为预报因子，探讨短波衰退事件（SWF）的预报．通过对 1980.1—6 月和 1980.1—2 月的短波衰退事件作中、短期预报的检验，表明，用此方法作中期预报的效果优于经验方法的预报；用于短期预报（3 天）效果较好．

一、引 言

通过半定量的数字化分析将有助于对太阳活动的预报，过去在实行定性预报的基础上，人们一直在探讨研究数字化预报方法．1974 年北京天文台探讨了用多元判别分析方法作短波衰退事件（SWF）和质子事件（proton）的预报*；1976 年进行过用多维时间序列与经验相结合建立中期短波衰退事件和质子事件的预报方法的研究**．1980 年云南天文台利用综合指数建立时间序列的自回归模型用于短波衰退事件和质子事件的中期预报***．1980 年 Vecchia 等在预报中也采用多元判别分析和逻辑回归技术[1]．1981 年 Neidig 等建立了多元判别分析的耀斑预报模型[2]．1984 年以来，云南天文台、北京天文台均在研究，试验模糊聚类[3]、时间序列的自回归滑动平均模型[4]及自回归求和滑动平均模型[5]应用于太阳活动预报．以上所用各种数理统计手段有一个共同的特点，就是要求所用资料数据序列满足随机平稳条件，正态分布．

太阳活动是一个复杂的过程，观测数据很难达到平稳条件．1978 年，Tong, H. 提出了适合于非平稳随机时间序列的模型，门限自回归模型．1983 年，Tong, H. 使其更完善化[6]．本文试图用这种模型，以每日最大 2800MHz 射电爆发峰值流量为预报因子（可观测输入数据），讨论 SWF 事件的预报．

本文于 1987 年 2 月 19 日收到。

* 杨自强，张建中，范岳华等，北京天文台台刊，No. 3(1974)，p.46.

** 范岳华，王家龙，张建中，北京天文台台刊，No. 7(1976)，p.51.

*** 罗葆荣，张柏荣，刘之煌，云南天文台台刊，No. 3 -4(1981)，p.93.

二、资 料 选 取

短波衰退事件与 2800MHz 射电爆发密切相关，而国内还没有对短波衰退事件的直接观测测量，所以本文选取 2800MHz 射电爆发作为预报因子.

本文所用资料有 SWF 事件的强度和每日最大 2800MHz 射电爆发峰值流量.

SGD, part 1 提供了世界范围内的电离层突然骚扰 (SID) 资料[7] (包括 SWF 事件, SEA 事件, SPA 事件和 SES 事件). 为了数字化的需要, 本文把 SGD, part 1 上给出的 SWF 的 0 —$3^{+}10$ 个级别按表 1 给出数据.

表 1

I	0	1^{-}	1	1^{+}	2^{-}	2	2^{+}	3^{-}	3	3^{+}
I_0	0.0	0.8	1.0	1.2	1.8	2.0	2.2	2.8	3.0	3.2

SGD, part 1 提供了所有每日的射电爆发峰值流量及频率. 本文取 OTTAWA 2800MHz 每日最大射电爆发峰值流量作为当天的射电爆发数据[7], 以北京天文台 2840MHz 每日最大射电爆发资料加以补充[8]. 当北京天文台无观测时, 用紫金山天文台的 3000MHz 的观测资料换算出北京天文台的观测数据. 经对两个天文台 1980 年射电爆发峰值流量的相关统计得出以下关系:

$$\Delta S_B = 7.26 + 0.39\Delta S_p. \tag{1}$$

ΔS_B 为北京天文台 2840MHz 射电爆发峰值流量 (绝对增量), ΔS_p 为紫金山天文台 3000MHz 射电爆发峰值流量.

对每日最大的 2800MHz (北台为 2840MHz) 爆发峰值流量 ΔS 作如下变换:

$$R_t = \lg(\Delta S + 1). \tag{2}$$

由以上资料得到两列数据时间间隔为一天的时间序列. SWF 事件的时间序列以 $\{I_{0t}\}$ 表示; 每日最大射电爆发峰值流量的时间序列以 $\{R_t\}$ 表示.

三、数 学 模 型

(一) 建立开环门限自回归模型

设 $\{x_t\}$ 为系统的可观测输出动态数据, $\{y_t\}$ 为系统的可观测输入动态数据, $t = 1, 2, \cdots\cdots, N$, 则称

$$x_t = \begin{cases} a_0^{(1)} + \sum\limits_{i=1}^{p_1} a_i^{(1)} x_{t-i} + \sum\limits_{i=1}^{q_1} b_i^{(1)} y_{t-i} + \varepsilon_t^{(1)}, & \text{当 } y_{t-d} \leqslant r \\ a_0^{(2)} + \sum\limits_{i=1}^{p_2} a_i^{(2)} x_{t-i} + \sum\limits_{i=1}^{q_2} b_i^{(2)} y_{t-i} + \varepsilon_t^{(2)}, & \text{当 } y_{t-d} > r \end{cases} \tag{3}$$

为 $\{x_t, y_t\}$ 的开环门限自回归模型, 简记为 TARSO $(2, (p_1, q_1), (p_2, q_2))$. 其中, p_1、

q_1、p_2、q_2 为模型阶数；$a_i^{(j)}(i=1,2,\cdots\cdots,p_j)$，$b_i^{(j)}(i=1,2,\cdots\cdots,q_j)$，$(j=1,2)$ 为模型系数；$\{\varepsilon_t^{(j)}\}(j=1,2)$ 是均值为零，方差为 $\sigma^2_{\varepsilon_t(j)}$ 的白噪声序列，$\varepsilon_t^{(1)}$，$\varepsilon_t^{(2)}$ 互相独立；r 为门限值；d 为延迟参数.

预先给定模型中两分段的最大阶数 L 和最大延迟参数 D. 令 $n_0=\max\{d,L\}(d=1,2,\cdots\cdots,D)$. 对于每一个固定的 r 和 d，将数据 x_{n_0+1}，x_{n_0+2}，$\cdots\cdots$，x_N 分成两类：$x_{n_0+i_{11}}$，$x_{n_0+i_{12}}$，$\cdots\cdots$，$x_{n_0+i1N_1}$（当 $y_{n_0-d+i}\leqslant r$ 时）；$x_{n_0+i_{21}}$，$x_{n_0+i_{22}}$，$\cdots\cdots$，$x_{n_0+i2N_2}$（当 $y_{n_0-d+i}>r$ 时），N_1 和 N_2 分别为两类数据的个数. 将两类数据代入 (3) 式得

$$\underline{X}_t^{(j)}=\underline{A}^{(j)}\underline{\beta}^{(j)}+\underline{\varepsilon}_t^{(j)},\quad(j=1,2) \tag{4}$$

其中，

$$\underline{X}_t^{(j)}=\begin{pmatrix}x_{n_0+i_{j1}}\\ x_{n_0+i_{j2}}\\ \vdots\\ x_{n_0+i_jN_j}\end{pmatrix},$$

$$\underline{A}^{(j)}=\begin{pmatrix}1 & x_{n_0+i_{j1}-1}\cdots\cdots x_{n_0+i_{j1}-p_j} & y_{n_0+i_{j1}-1} & y_{n_0+i_{j1}-2}\cdots\cdots y_{n_0+i_{j1}-q_j}\\ 1 & x_{n_0+i_{j2}-1}\cdots\cdots x_{n_0+i_{j2}-p_j} & y_{n_0+i_{j2}-1} & y_{n_0+i_{j2}-2}\cdots\cdots y_{n_0+i_{j2}-q_j}\\ \vdots & \vdots & \vdots & \vdots\\ 1 & x_{n_0+i_jN_j-1}\cdots\cdots x_{n_0+i_jN_j-p_j} & y_{n_0+i_jN_j-1} & y_{n_0+i_jN_j-2}\cdots\cdots y_{n_0+i_jN_j-q_j}\end{pmatrix},$$

$$\underline{\beta}^{(j)}=(a_0^{(j)}\ a_1^{(j)}\ \cdots\cdots\ a_{p_j}^{(j)}\ b_1^{(j)}\ b_2^{(j)}\ \cdots\cdots\ b_{q_j}^{(j)})^\tau,$$

$$\underline{\varepsilon}_t^{(j)}=(\varepsilon_{n_0+i_{j1}}\varepsilon_{n_0+i_{j2}}\cdots\cdots\varepsilon_{n_0+i_jN_j})^\tau.$$

用最小二乘方法对参数 $\underline{\beta}$ 进行估计：

$$S(\underline{\beta})=\varepsilon^\tau\varepsilon=\|\underline{X}_t-\underline{A}\underline{\beta}\|^2, \tag{5}$$

$$\frac{\partial}{\partial\underline{\beta}}S(\underline{\beta})=-2\underline{A}^\tau(\underline{X}_t-\underline{A}\underline{\beta})=0,$$

得

$$\underline{\beta}=(\underline{A}^\tau\underline{A})^{-1}\underline{A}^\tau\underline{X}_t. \tag{6}$$

用最小信息准则 (AIC)[9] 确定模型阶数 p_j 和 q_j，门限值 r 以及延迟参数 d：

$$\mathrm{AIC}(\hat{p}_j,\hat{q}_j)=\min[N_j\ln\hat{\sigma}^2_{\varepsilon_t(j)}+2(p_j+q_j+2)], \tag{7}$$

其中

$$\hat{\sigma}^2_{\varepsilon_t(j)}=\frac{1}{N_j}S(\underline{\beta})^{(j)}.$$

$$\mathrm{AIC}(\hat{r},d)=\min[\mathrm{AIC}(\hat{p}_1,\hat{q}_1)+\mathrm{AIC}(\hat{p}_2,\hat{q}_2)]. \tag{8}$$

$$\mathrm{AIC}(\hat{d})=\min[\mathrm{AIC}(\hat{r},d)/(N-n_0)]. \tag{9}$$

$\hat{p}_j$，$\hat{q}_j$，$\hat{r}$ 和 $\hat{d}$ 分别为 p_j，q_j，r 和 d 的估计值.

(二) 预报

由历史数据 $\{x_t\}$ 和 $\{y_t\}$ 确定了 TARSO$(2,(p_1,q_1),(p_2,q_2))$ 模型以后，对于时间序列 $\{x_t\}$ 未来第 T 时段的预测值 $\hat{x}_{N+T}$ 的预报公式为

$$\begin{cases} x_{N+T} = a_0^{(1)} + a_1^{(1)}\hat{x}_{N+T-1} + a_2^{(1)}\hat{x}_{N+T-2} + \cdots\cdots + \\ \qquad + \sum_{i=T}^{p_1} a_i^{(1)} x_{N+T-i} + \sum_{i=1}^{q_1} b_i^{(1)} y_{N+T-i}, \quad y_{N+T-d} \leqslant r \\ x_{N+T} = a_0^{(2)} + a_1^{(2)}\hat{x}_{N+T-1} + a_2^{(2)}\hat{x}_{N+T-2} + \cdots\cdots + \\ \qquad + \sum_{i=T}^{q_1} a_i^{(2)} x_{N+T-i} + \sum_{i=1}^{q_2} b_i^{(2)} y_{N+T-i}, \quad y_{N+T-d} > r \end{cases} \tag{10}$$

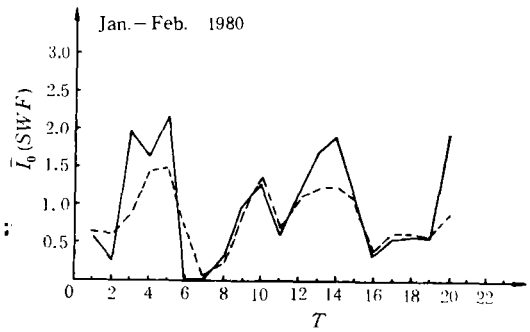
四、结果及讨论

用前面得到的随机时间序列 $\{I_{0t}\}$ 和 $\{R_t\}$ 经预处理建立模型，并作短期和中期预报检验．两列随机时间数据序列的起始时间取作 1978 年 1 月 1 日．

(一) 短期预报

对 SWF 事件的短期预报系对未来三天 SWF 事件平均活动水平的预报．数据时间间隔 ΔT 取作三天，即对原序列 $\{I_{0t}\}$，$\{R_t\}$ 分别作每三天取一平均值的处理，得新的随机序列 $\{x_t\}$，$\{y_t\}(t = 1, 2, \cdots\cdots, N)$．建模样本长度 N 取作 243，预报步长取作 1．模型状态及对 1980 年 1 月和 2 月最佳预报结果分别由表 2 和图 1 给出．

显然，模型阶数，门限值，延迟参数等都不是稳定不变的．由图 1 可见，用 2800MHz 每日最大射电爆发峰值流量对未来三天短波衰退事件平均活动水平作预报效果比较好．当 SWF 事件的强度小于 1.5 时，预报与实测拟合的非常好．但当 SWF 事件强度大于 1.1 时预报值偏低．本模型及预报方法基本可作为 SWF 事件的短期预报方法．



——实测 (reality)；　------预报 (prediction)

图 1　SWF 的最佳预报结果与实测的比较

Fig. 1 The best prediction of SWF versus realities.

(二) 中期预报(一个月)

首先对原时间序列 $\{I_{0t}\}$，$\{R_t\}$ 分别作 3 天滑动，7 天平均预处理，得相应的时间序列 $\{x_t\}$，$\{y_t\}(t = 1, 2, \cdots\cdots, N)$．建模样本长度仍取 $N = 243$，预报步长取 10．各时段模型各种参数等见表 3．对 1980 年 1 月到 6 月的最佳预报结果见图 2．很明显，以本方法作中期预报时其模型阶数，门限值等也是时变数．图 2 说明中期预报效果不如短期预报，只能作到预报与实测趋势基本一致．但是，比过去的经验方法的预报结果明显好些．用原来经验方法对同一时段的预报与实测的相关系数仅为 0.23**** (自由度为 60)，而本方法的预报结果与实测的相关系数达 0.69 (自由度仍为 60)．

值得提出的是，本预报是在假设对未来一个月的 2800MHz 的每日最大射电爆发流量的预报相当准确的条件下得出的结果．以公式 (10) 可知，预报 x_{N+T} 需要知道未来

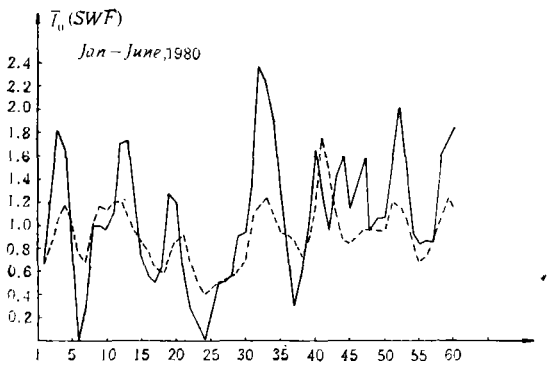
**** 董士岑，汤柏盛，云南天文台台刊，No. 3—4(1981)，p.100.

表 2

预报时段	d	r	p_1	q_1	p_2	q_2	$\sigma^2_{\varepsilon_t(1)}$	$\sigma^2_{\varepsilon_t(2)}$
80.1.1—3	5	0.6623	2	1	5	1	0.2884	0.3385
1.4—6	4	1.0477	15	10	1	1	0.2460	0.3208
1.7—9	4	1.0718	12	10	10	15	0.2603	0.2152
1.10—12	2	1.2425	11	6	1	1	0.2263	0.3938
1.13—15	3	0.8731	5	1	5	5	0.2909	0.3251
1.16—18	3	1.2543	5	5	5	14	0.2892	0.2715
1.19—21	1	1.4484	1	2	1	2	0.2885	0.4560
1.22—24	3	1.4484	5	7	1	1	0.3110	0.3670
1.25—27	2	0.6824	11	12	10	4	0.1305	0.3256
1.28—30	5	0.6829	4	4	5	1	0.2881	0.3455
1.31—2.2	3	1.4683	5	5	1	10	0.3167	0.2803
2.3—5	3	1.3016	5	1	1	10	0.3138	0.3202
2.6—8	5	1.3245	5	5	2	2	0.2655	0.3961
2.9—11	5	1.3251	5	6	2	2	0.2665	0.3987
2.12—14	1	1.4765	1	2	1	3	0.2862	0.4453
2.15—17	3	1.1397	5	1	5	11	0.3064	0.2940
2.18—20	3	0.6917	1	1	5	5	0.2996	0.3495
2.21—23	1	1.4738	1	3	1	2	0.2787	0.4570
2.24—26	5	1.1397	5	5	2	2	0.2758	0.3748
2.27—29	1	1.3016	11	6	1	2	0.2118	0.4509

表 3

预报时间	d	r	p_1	q_1	p_2	q_2	$\sigma^2_{\varepsilon_t(1)}$	$\sigma^2_{\varepsilon_t(2)}$
80年1月	2	0.7519	11	15	5	5	0.0247	0.0806
2月	3	1.4447	5	4	3	7	0.0673	0.0941
3月	2	1.2706	5	6	4	4	0.0599	0.0916
4月	5	1.2812	4	1	2	15	0.0710	0.0650
5月	5	1.4664	4	4	2	1	0.0788	0.0828
6月	4	1.4736	3	1	13	4	0.0850	0.0590



——实测 (reality)；　　----预报 (prediction)

图 2　SWF 事件的最佳预报结果与实测的比较

Fig. 2 The best predictions of SWF versus realities

y_{N+T-1}, ……, y_{N+1} 的精确值，而要预测其精确值是比较困难的，所以，这种方法对于中期预报只能作为一种参考.

（三）讨论

两个时段(短期和中期)的预报检验结果的共同点是对强 SWF 事件的预报偏低. 究其原因，粗略认为有以下几点：

(1) 当时 SGD part 1 上没有完整的完成全天覆盖的 2800MHz 射电爆发资料，所以，我们选取的每日最大爆发峰值流量不一定是最大流量的真值，漏掉的可能性相当大.

(2) 观测数据来源不一，数据不一定精确. 例如，有时爆发峰值流量只给出≥1000的，我们只能取 1000，而其实际值可能是 1500，也可能是 8000. 另外，北台的观测数据与 OTTAWO 的观测数据也有差别. 所以，这样的误差也会影响建模及预报效果.

(3) 时间序列的数据跳动大时，可能模型本身也难以作出响应. 如果对数据作更多的平滑处理，预报效果肯定会好，但无实际使用价值.

另外，模型参数，阶数，门限值以及延迟参数均为时变数，故难以用固定的模型描述.

参　考　文　献

[1] Vecchia, D., P. Tryon, G. Caldwell and R. Jone. Rept. AFGL-TR-0-0336, Air Force Geophysics Lab., Hanscom AFB, 1980.

[2] Neidig, D., P. Wiborg, P. Seagraves, J. Hirman and W. Floers, Rept. AFGL-TR-8-0026, Air Force Geo-Physics Lab., Hanscom AFB, 1981.

[3] 张勤和洪琴芳，天文学报，第 **27** 卷，第 164 页，1986.

[4] 张勤，罗葆荣和柴晓军，天文学报，第 **27** 卷，第 247 页，1986.

[5] 张桂清，天文学报，第 **27** 卷，第327 页，1986.

[6] Tong, H., Threshold Models in Nonlinear Time Series Analysis, Lecture Notes in Statistics, Springer Verlag, No. 21, 1983.

[7] Solar-Geophysical Data Prompt Reports, part 1, No. 403—437, 1978—1980.

[8] 太阳地球物理资料，中国科学院，1978 年 1 期—1980 年 12 期.

[9] Akaike, H., 2nd International Symposium on Information Theory (Ed. by B. N. Petrov and F. Csaki) Budapest, Akademial Kiado, p. 261, 1973.

The Inquiry of the Prediction of the Short Wave Fadeout by the Open-loop Threshold Autoregressive Model

Zhang Gui-qing Wang Yan-lin

(Beijing Observatory, Academia Sinica)

Abstract

The open-loop threshold autoregressive model is used as a method to study prediction of short wave fadeout in this paper, based on a relation that short wave fadeout strongly relies on intense solar activity.

The peak of daily flux of the strongest 2800 MHz radio burnt is chosen as predictor. The medium-term, short-term forecasting verifications for Jan.-June, 1980 and Jan.-Feb., 1980 show: for medium-term forecasting (one month ahead) this method is better than the empirical method, but it is quite difficult to forecast the predictor which is required first in the method; the short-term forecasting (three days ahead) using this method has better forecasting effect.