

# 用非线性逐步回归方法 作台风预测试验

宋文玲 何 敏

(国家气候中心,北京 100081)

## 提 要

利用非线性逐步回归方法作台风活动趋势的预报。在回归方程中考虑了预报因子与预报量之间的非线性关系,回归方程中选入的非线性因子具有较高的相关显著性,对1995~1997年台风预测试验表明:非线性因子在预报中起着及其重要的作用,取得了较好的预报效果。

**关键词:** 非线性逐步回归 台风 预测

## 引 言

在西北太平洋生成的台风和热带风暴(以下简称台风)及登陆我国台风的多少不仅与大尺度环流的变异有关,而且与海温场、地面降水和温度场的变化有关。因此,在环流场和下垫面要素场中存在着各种影响台风活动的超长期预报信号。在过去的台风活动趋势的预测中,大多考虑单因子、线性的相关关系。而台风活动与各预报因素之间并非单一要素的线性关系。黄嘉佑等人过去用非线性逐步回归方法预报其它要素方面也取得了较好的效果<sup>[1,2]</sup>。因此在本文中综合考虑影响台风活动的物理因素,试用非线性逐步回归方法作台风活动趋势的预测。

## 1 资料与方法

### 1.1 预报对象

我们取1951~1997年初次登陆台风时间(初台)、未次登陆台风时间(终台)、编号台风个数(编台)、登陆台风个数(登台)、登陆强台风个数(强台)等序列为预报对象。

### 1.2 预报因子

预报因子分别从前期12个月的4个场面资料中选取,其中(a)反映大气环流作用的环流场:74个北半球环流特征量(副热带高

压、极涡、环流型、西风带环流指数、东亚槽位置和强度、以及西藏高原高度等)资料组成第一预报因子场。(b)海洋场:在赤道太平洋和北太平洋范围内,选取反映海温主要变化特征的9个海域的海温为第二因子场。(c)降水场:选全国15个区的降水指数为第三因子场。(d)温度场:选全国8个区的温度等级为第四因子场。显然,制作预报的初始时间不同,所取预报因子的时段也不一样。1月份作预报时,在上一年1~12月资料中选因子。3月份作预报时则在上一年3~12月和当年1~2月的资料中选因子。其余可类推。

### 1.3 建立预报方程

为了考察各个因子场在台风预报中的情况,我们分别对4个因子场建立台风预报方程。

以第一预报因子为例,在建立预报方程中,对前期12个月74个物理量场共863(888个因子除去夏季月份无资料的项)个因子进行筛选。一般因子的筛选是以因子与预报量的线性相关系数是否显著(显著水平5%临界相关系数为0.35)作为选择的标准。由于气象要素之间关系的复杂性,仅考虑它们之间的线性关系是不够的,还应考虑非线

性关系<sup>[3]</sup>,因此选择因子时,选取了线性和非线性相关系数显著的因子。预报量  $y$  与因子  $x$  的线性和非线性相关关系常用的有线性、指数、自然指数和对数等 4 种类型,函数形式可用最优化方法确定<sup>[3]</sup>。然后把函数作为新变量用常用方法求出其各类型的非线性相关系数<sup>[4]</sup>。其一般关系式为:

$$y_i = f(x_i, \gamma) + e_i$$

式中  $x_i$  为含  $p$  个因子变量  $i$  个样品向量,  $\gamma$  为非线性回归系数向量,  $e_i$  为误差,  $y_i$  为第  $i$  个样品的预报量,  $f(x_i, \gamma)$  为  $x_i$  的任意非线性函数。这里主要有:

$$f(x_i, \gamma_1) = b + ax_i$$

$$f(x_i, \gamma_2) = bx_i^a$$

$$f(x_i, \gamma_3) = be^{ax_i}$$

$$f(x_i, \gamma_4) = b + \lg x_i$$

取非线性因子后做逐步回归,阈值为 0.3~0.5,每一个预报对象,一个非线性方程,选取  $n$  个超过规定置信度的高相关预报因子进方程。

## 2 预报方程中各参数及因子的意义

以第一因子场预报方程预报 1996 年编号台风为例,所选因子及方程中的各项参数见表 1。

表 1 挑选出的预报因子及方程中的参数

序号	M	V	L	a	b	r	Ma	Mi	Xm	S
1	3	41	4	100.000	0.188	0.412	2.165	-1.622	-1.284	1.511
2	4	39	3	-8.719	0.104	0.463	2.391	-1.365	0.215	0.285
3	4	50	4	100.000	0.251	0.269	1.859	-2.261	-1.425	2.648
4	8	17	3	-0.916	0.058	0.391	2.555	-1.209	0.766	0.163
5	8	24	4	100.000	0.078	0.307	1.963	-2.163	-0.974	1.459
6	8	47	4	100.000	0.108	0.301	2.590	-1.760	-1.294	1.480
7	9	31	2	1.093	-0.293	0.482	1.823	-2.528	0.635	0.251
8	9	40	3	-6.597	0.219	0.522	1.851	-1.824	0.140	0.214
9	10	20	4	100.000	0.215	0.335	2.460	-2.147	-1.067	1.425
10	10	37	2	1.562	-0.334	0.414	2.897	-1.230	0.466	0.379
11	10	44	2	0.378	0.014	0.420	2.743	-1.304	0.122	0.093
12	10	45	4	100.000	-0.246	0.248	2.162	-1.497	-1.653	2.286
13	10	72	4	100.000	-0.186	0.304	2.410	-2.158	-1.039	1.404
14	11	50	4	100.000	-0.212	0.248	2.715	-1.876	-1.225	1.415
15	11	54	4	100.000	0.180	0.271	2.431	-1.754	-1.229	1.442
16	11	58	3	1.538	0.355	0.403	1.840	-1.733	2.310	0.989

复相关系数:0.813,历史拟合率:0.87

M:表示月份(1~12月);V:表示物理场序号(1~74项);L:表示因子与预报量关系函数类型(1为线性,2为幂函数,3为指数函数,4为对数函数);a:表示函数中的常数;b:表示方程中标准化变量回归系数;r:表示非线性相关系数;Ma:表示标准化变量最大值;Mi:表示标准化变量最小值;Xm:表示标准化变量平均值;S:表示标准化变量标准差

从表 1 可以看出,选入预报方程的因子有 16 个,其中幂函数形式的 3 个,指数形式的 4 个,对数形式的 9 个,而线性函数的 0 个,反映了物理量场与预报量关系的复杂性,以及非线性因子的重要性。从表中 V 物理场序号可以查出,未来生成台风的多寡与前一年北半球极涡面积和强度指数、北半球极涡中心强度和位置、W 环流型、西太平洋副高强度、南海副高强度等特征量有关,反映了影响台风活动的超长期信号,既与西风带的经向程度有关,也与副热带高压的变化有关。从表中相关系数  $r$  可以看出:非线性相关系数

75% 达 0.05 信度,且 50% 达 0.01 信度。其复相关系数  $R^2 = 0.813$ ,拟合正确率达 0.87。复相关系数实际是衡量所选因子对预报量的非线性解释方差的百分率,即  $R^2$  反映了预报因子对预报量的非线性回归解释的部分,可作为衡量回归方程拟合量的一个指标。按此预报方程得出 1996 年的预报值为负距平。1996 年编号台风为 25 个,比常年偏少,预报与实况相符。

第二因子场对编号台风的预报为:选入 3 个因子,2 个幂函数形式,1 个线性函数形式;第三因子场对编号台风的预报为:选入 1

个因子,为幂函数形式;第四因子场对编号台风的预报为:选入1个因子,也为幂函数形式。表明选入因子以非线性为主。对于台风其余几个预报项目和因子场的情况大致相仿。

### 3 结果与检验

表2给出了1995~1997年初台、终台、强台、编台、登台的试报与检验,Hc表示环流特征量因子场、Sc表示海温因子场、Rc表示降水因子场、Tc表示温度因子场。由于影响台风活动的主要因素是大气环流和海洋状

况,因此对4个因子场的预报结果进行综合时,首先重点考虑环流场和海温场的预报符号;如果两者为异号时,则以4个场的优势符号为最终的距平符号预报;当因子场为两正两负时,主要考虑选历史拟合率高的三个因子场的优势符号为距平符号预报。结果表明:1995年5个预报项目中有两项报对,1996年5项全报对,1997年5项中有四项报对,因此,总的预报效果是比较好的,正确率达73%以上。

表2 利用非线性回归方法计算的结果

	1995					1996					1997						
	初	终	强	编	登	初	终	强	编	登	初	终	强	编	登		
Hc	-	+	-	+	+	Hc	+	+	+	-	+	Hc	+	+	-	-	+
Sc	+	-	-	+	-	Sc	-	-	-	+	-	Sc	-	-	+	-	-
Rc	+	-	-	+	-	Rc	+	-	-	-	+	Rc	-	-	-	+	-
Tc	-	-	-	-	+	Tc	+	-	+	-	+	Tc	-	-	+	-	+
综合	晚	早	少	多	少	综合	晚	早	少	少	多	综合	早	早	少	少	少
实况	晚	晚	少	少	多	实况	晚	早	少	少	多	实况	晚	早	少	少	少
检验	✓	×	✓	×	×	检验	✓	✓	✓	✓	✓	检验	×	✓	✓	✓	✓

我们在1998年汛期预报时,给出1998年台风活动趋势预报(见表3),1998年初台在8月4日,比常年明显偏晚,终台在9月19日,比常年明显偏早,编号台风、登陆台风以及强台风个数均比常年偏少,预报与实况均相符。1998年台风趋势预报正确。

表3 1998年利用非线性回归方法计算的结果

	初台(1)	终台(5)	强台(6)	编号(7)	登陆(8)
Hc	+	-	+	+	+
Sc	-	-	-	-	-
Rc	+	+	-	-	-
Tc	-	-	-	-	+
综合	晚	早	少	少	少

### 4 结论

非线性逐步回归方法的特点是:

(1)考虑了预报因子和预报量之间的非线性关系,对1995~1997年5个预报量的预报试验表明,非线性因子在预报中起着及其重要的作用,所选择的因子具有最好的效果。

(2)可以选择对台风活动有影响的因子场,使方程中预报因子有较明确的物理意义。

(3)较高的复相关系数表明入选预报因子具有较好的代表性及较强的解释能力。

(4)可以对预报量进行滚动预报,按照年度预报、汛期预报以及台风活动盛期来临之前的补充预报的业务需要,挑选不同的物理因子,利用非线性回归方法,在不同时期作出台风活动趋势的预测。

试验表明,非线性逐步回归方程,由于考虑非线性因子在预报中的重要作用,在三年的试报及一年台风活动趋势预报中取得了较好的效果,利用这种方法是有一定的预报意义的。

### 参考文献

- 1 黄嘉佑等. 黄河三花地区汛期逐日降水 MOS 预报的因子选择试验. 气象学报, 1993, 51(2): 232~235.
- 2 黄嘉佑等. 北京降水缺水资料恢复的研究. 气象, 1996, 22(7): 6~11.
- 3 冯耀煌, 杨旭. 论最优预报因子与最优预报方程. 气象学报, 1989, 47(1): 52~60.
- 4 黄嘉佑. 气象统计分析与预报方法. 北京: 气象出版社, 1990: 387.

# Typhoon Forecast Test with Nonlinear Stepwise Regression Method

Song Wenling He Min

(National Climate Center, Beijing 100081)

## Abstract

The nonlinear stepwise regression method is used to make typhoon activity tendency forecast. Predictors chosen in the nonlinear regression equations are associated with very high correlation significance level. By considering nonlinear relationships between predictors and predictands in regression equations, better forecast results were gained in the typhoon activity forecast tests in 1995—1997.

**Key Words:** nonlinear stepwise regression typhoon forecast tropical storm