

多预报量双重筛选逐步回归 在台风路径预报中的应用

金一鸣

(浙江省气象科学研究所)

提 要

本文对通常的逐步回归进行了分析,指出在同时制作多个预报量的预报时,这方法并不十分可靠。因为这样得到的预报方程常常包含着比较大的随机误差,之所以产生上述情况,问题在于对多个预报量建立方程时,没有考虑预报量之间的关系。因此,有必要对通常的逐步回归进行改进,考虑对多个预报量同时进行筛选的逐步回归。

台风路径预报的试验结果表明:由多预报量双重筛选逐步回归建立的方程组,对相近预报时段的相同预报量,其影响的预报因子大体相同,而对于不同预报时段的不同预报量,则影响的预报因子有着显著的差异,这样得到的方程组物理意义比较明确,不仅有利于提高预报方程的预报精度,而且还有利于对不同预报方程的预报因子进行分析和改进。

一、引 言

逐步回归分析是目前统计预报中比较成熟的一个预报方案,对分析和研究气象随机过程是一种重要的工具。但是,在台风路径预报中,同时制作 $\lambda_{24}, \varphi_{24}, \dots, \lambda_{60}, \varphi_{60}$ 等多个预报量(其中 λ 表示经度, φ 表示纬度,下标表示预报时效,以小时计)的预报时,用通常的逐步回归方法所得到的预报方程,对不同的 y_i , 选入的因子可以是完全不同的。而实际情况是对于相近预报时段的相同预报量,其影响的环流因子一般来说应该大体相同,因此,由通常的逐步回归建立的预报方程,就有可能包含一些统计上不显著的预报因子,从而带来比较大的随机误差。同时,也给方程的物理解释带来了困难。另外,对于由通常的逐步回归得到的方程,当给出一定的置信水平 α 时,就可以得到每个 y_i 的区间估计,这些区间就是预报范围。因此,我们就知道每一个 y_i 在该区间成立的概率是 $1-\alpha$,但是却不知道每个 y_i 在这些区间同时成立的概率。因此,用通常的逐步回归所得到的预报方程来做多个预报量的预报时并不十分可靠,有个别的情况,甚至会得出相互矛盾的预报意见,以致使相近预报时段的台风预报路径发生虚假的曲折现象。之所以产生上述情况,问题在于对 $\lambda_{24}, \varphi_{24}, \dots, \lambda_{60}, \varphi_{60}$ 建立方程时,没有考虑预报量 y_i 之间的关系。而对台风路径预报来讲,各时段的预报位置并不互相独立,它们彼此有一定的关联,即实际情况是:因变量的一部分与自变量的一部分有较密切的关系。而在通常的逐步回归中,我们只求自变

1981年10月28日收到,1982年1月8日收到修改稿。

量 x_i 分别与 $y_1, y_2, \dots, y_p$ 的关系,问题的严重性在于我们希望判别 x_i 是否对全部因变量或者部分因变量都重要,因此有必要考虑对多个预报量同时进行筛选的逐步回归^[4,7],使得既能按因变量和自变量的关系进行分组,又能使每个自变量对各组因变量的影响都反映出来。

二、应用和效果

Anderson^[1] 和 Rao^[2] 对双重筛选多元逐步回归的模型和算法,曾先后作过详细的介绍,张尧庭等^[3]将此方法应用于实际的预报。本文又将此方法应用于西北太平洋的台风路径预报,并且与通常的逐步回归进行了比较。

根据 1958—1975 年 7—9 月西北太平洋 293 个台风个例,预报对象为台风未来的中心位置 $\lambda_{24}, \varphi_{24}; \lambda_{36}, \varphi_{36}; \lambda_{48}, \varphi_{48}; \lambda_{60}, \varphi_{60}$, 预报因子与台风路径统计预报 1975 年方案相同^[4](包括持续性因子、天气经验因子、固定网格因子、活动网格因子和福氏级数展开的系数),采用多预报量双重筛选逐步回归的方法^[5],为了控制进入方程组的因子及预报量的个数,我们在筛选自变量与因变量时分别采用不同的 F 值,记 F_x 为筛选自变量的临界值, F_y 为筛选预报量的临界值。当取不同的 F_x 和 F_y 时,根据预报量入选的次序,可分别得到 $\lambda_{24}, \lambda_{36}; \varphi_{24}, \varphi_{36}; \lambda_{48}, \lambda_{60}; \varphi_{48}, \varphi_{60}$ 等四组回归方程,它们分别入选的因子和系数列如表 1。

从表 1 的四组方程可以看出,对于不同预报时段的台风路径,其影响的因子有着显著的差异,对于相近预报时段的不同预报量,影响的因子也有不同,而对于相近预报时段相同的预报量,则影响的因子基本上一致。近期 24—36 小时持续性因子占的比重较大,西风带因子主要是长江中下游的环流组合因子,副高的影响主要反映在台风北侧的高度。另外,引导气流和展开福氏级数的系数对台风短期的移动也有一定的作用。而对于 48—60 小时比较长的预报时段,持续性因子占的比重就小一些,天气经验因子的比重就增大,东亚西风槽的强度和位置反映了西风带因子对台风远期路径的影响,副高的影响除反映在台风与副高的距离外,还反映了副高的位置、形状和强度的影响。另外,我们比较第一组和第二组的方程,注意到除相同的因子外,在短期内影响台风纬向位移的因子主要是长江中下游地区的风场和温度场特征,而影响径向位移的因子主要是台风中心初始移速的变化和前期的路径特征;比较第三组和第四组的方程,同样注意到在 48—60 小时,影响台风纬向位移的因子主要是东亚西风槽的强度、台风中心与副高脊线的距离、台风北侧副高强度和台风东侧副高南落数以及长江下游的风场特征,而影响径向位移的因子,主要是中纬度西风指数、副高强度指数、副高轴线的变化、华北低槽的位置和强度。这些都是和预报员的经验相一致的。因此,这样得到的方程组物理意义比较明确,有利于对不同预报时段预报方程的因子进行分析和改进。

为了与通常的逐步回归得到的预报方程进行比较,我们用上述四组方程和统计预报 1975 年方案中由通常的逐步回归得到的综合方程^[4]分别计算了它们对 293 个台风历史样本的拟合误差(表 2)。

从表 2 可见,近期 24—36 小时二个方案的拟合误差相仿,而远期 48—60 小时,双重

表 1 各时段台风路径预报的回归方程

预报时段(小时)		24		36		48		60	
预 报 量		λ	φ	λ	φ	λ	φ	λ	φ
预 报 因 子	b_0	102.5150	6.1381	108.3465	11.0447	112.4322	16.1744	115.3105	18.1867
	$\lambda_0 - 100$	0.7567		0.9445		0.9256		1.0523	
	φ_0		0.5456		0.4565		0.9827		0.7589
	$\lambda_1 - 100$	0.6346		0.6737					
	φ_1		0.3474		0.5027				
	$(\lambda_0 - \lambda_{-4})\cos\varphi_0$	-0.9875	-0.2283	-0.7821	-0.1890	-1.8500		-1.9340	
	$\varphi_0 - \varphi_{-4}$	1.2425	1.0545	1.2523	1.3435		2.0030		1.7975
	$V_{上海}$	0.0345		0.0199		0.0405		0.0373	
	V_{T1}	-0.0336		-0.0438					
	$P_E + P_W - 2P_0$		0.0050		0.0038				
	Z_3	0.0504	-0.0136	-0.0800	-0.0304				
	ΔZ_1^1						0.0857		0.0731
	ΔZ_1^2					0.0565		-0.0424	
	u_2	-0.0208		-0.0436					
	V_2		0.0569		0.0296				
	a_2		0.0047		-0.0032				
	b_2		-0.0073		-0.0088				
	$\Delta 588_{20}$					0.1246		0.1125	
	I_G						0.0684		0.0696
	$a + b - 2c$						0.0532		-0.0671
	$Z_2 + \Delta Z_2$					-0.1813	-0.1104	0.2531	-0.1340
	$\Sigma \Delta H$					-0.0973		-0.1309	
	$\lambda_0 - \lambda_{0.12W}$					-0.0868	0.0213	-0.1133	0.0647
	$C - \varphi_0$					-0.2543		0.1345	
	I_W						-0.0216		-0.0455
	$Z_1 + Z_2 - 2Z_{11}$						-0.0569		-0.0980
备 注		预报因子的含义和表达式与 [4] 相同							

筛选逐步回归方程的误差却小于通常的逐步回归方程。这说明了双重筛选逐步回归方程的方差减少百分率是随着预报时效的延长而增大的。所以从拟合情况可知,随着预报时效的延伸,双重筛选逐步回归比通常的逐步回归具有更好的预报性能。

为了进一步检验双重筛选逐步回归预报方程的预报效果,我们用二个方案分别对

表 2 二个预报方案绝对平均误差比较表(单位: 纬距)

预报时段 (小时)	24		36		48		60	
	$\Delta\lambda$	Δp	$\Delta\lambda$	Δp	$\Delta\lambda$	Δp	$\Delta\lambda$	Δp
双重筛选逐步回归方程	1.4	1.3	2.0	1.9	2.7	2.4	3.2	2.9
通常的逐步回归方程	1.5	1.3	2.1	1.9	2.9	2.7	3.6	3.2

表 3 53 次独立样本预报检验平均位置误差

预报时段(小时)	平均位置误差(单位: 海里)	
	双重筛选逐步回归方程	通常的逐步回归方程
24	136	145
36	189	212
48	270	298
60	335	368

1977—1980 年 7—9 月西北太平洋的 17 个台风进行了 53 次同台风同时次的预报, 它们的平均预报误差列如表 3。

表 3 显示了从 24 小时一直到 60 小时, 用双重筛选逐步回归得到的预报方程的误差明显地小于通常的逐步回归方程的误差。在实际预报时, 相对于通常的逐步回归方程, 随机误差的影响就会减小。由此可见, 在制作多个预报量的预报时, 当提供的可预报因子相同时, 多预报量双重筛选逐步回归方程能提高预报方程的预报精度。

参 考 文 献

- [1] Anderson, T. W., AN Introduction to Multivariate Statistical Analysis, Wiley, New York, 1958.
- [2] Rao, C. R., Linear Statistical Inference and Its Application, Wiley, New York, 1973.
- [3] 张尧庭、赵谦, 双重筛选多元逐步回归, 应用数学学报, 第 3 卷第 2 期, 1980.
- [4] 上海、浙江台风协作组, 台风路径统计预报 1975 年方案, 1976 年台风会议文集, 92—100.