

南海热带气旋风力的逐步回归分析及预报

杨平章 郭英琼 张慧婵 郭平

(中山大学大气科学系, 广州 510275)

摘要 选取具有天气学意义和物理意义的 9 个预报因子, 采用逐步回归分析, 建立了 84 个统计预报方程, 用于东南沿海 14 个台站预报热带气旋影响的风速. 根据统计样本外的 1988 年的资料作预报检验, 结果表明所建立的一套统计预报方程是可以用于实际预报的.

关键词 南海热带气旋, 逐步回归分析, 统计预报方程

分类号 P444

热带气旋风场的分析和预报, 由于缺乏直接观测资料, 难度较大. 早期用半经验的轴对称径向风速分布律来估计热带气旋风力分布和预报. 李叔廷, 房文鸾等^[1,2]采用天气气候学与统计相结合的方法分析出台风中的不对称的气压场, 通过梯度风关系算出不对称的台风风场分布. 关芬呈等^[3]采用移动极坐标对特征相似的南海热带气旋风场作合成分析, 然后又用实时资料与合成风场相结合, 采用 EOF 分析技术, 分析不同类型热带气旋风场的时空分布规律, 并结合热带气旋路径和强度预报作出热带气旋风场预报. 杨美川等^[4]利用上海台风所业务数值模式提供的预报场, 引入部分其他因子, 组合成多张概率图作组合因子, 建立单站热带气旋大风、暴雨预报方程. 数值预报的发展为提高气象要素预报精度, 包括难度较大的热带气旋大风预报展现了新的前景, 数值预报产品释用是今后业务预报发展的必然趋势. 作者选取 9 个预报因子, 应用计算机软件^[1], 对东南沿海 14 个站分别用同时资料(即完全预报法或 PP 法所用资料^[5])和实时资料(预报量的前 24 小时的数据), 又分别按不分季节和分 2 种季节建立各站的 2 种预报方程, 每站点 6 个方程, 14 个站点共 84 个热带气旋风力统计预报方程.

1 资料的选取和计算方案

1.1 资料范围

本文根据《台风年鉴》和《地面气象要素月报表》(1983~1987)选取 37 个热带气旋, 每个热带气旋从其对站点有影响(指热带气旋中心与站点距离在 11 个纬距之内)开始至影

1) 杨美川, 客观预报方法研制系统(PC 机版本). 上海台风研究所, 1993

收稿日期: 1994-09-28 杨平章, 男, 56 岁, 教授

响结束,累计为此热带气旋对站点影响的天数,每天取 02、08、14、20 时 4 个时次的资料,这样共获得 128 d,512 个样品;选取东南沿海分布较均匀的福州、厦门、广州、汕头、海丰、湛江、阳江、钦州、北海、海口、东方、琼海、崖县、西沙等共 14 个站点作分析和预报。

1.2 预报因子

选取预报因子的指导原则是进入统计的候选因子具有清楚的物理意义^[4,6],符合天气学原理,并能反映预报对象的天气特点.对数值预报产品的应用首先要考虑数值模式本身的预报精度.因此,据天气学、动力学等方面物理上的考虑,结合预报经验,作者选择了与热带气旋风速有密切关系的 9 个因子作为进行统计筛选的候选因子: x_1, x_2 分别是南海热带气旋影响下的本站气压和温度; x_3, x_4, x_5, x_6, x_7 依次为热带气旋中心的纬度位置($^{\circ}\text{N}$)、经度位置($^{\circ}\text{E}$)、最低气压、最大风速、与站点距离; x_8 是台站与热带气旋中心的气压差; x_9 是台站与参考站(武汉或韶关)的气压差,其中仅福州、厦门采用与武汉的气压差^[5]; y 是预报量,即南海热带气旋影响下的站点风速.

1.3 计算方案

采用双重检验的逐步回归方案^[6],其基本原理是将方差贡献显著的因子一个个引入,同时,每引入一个新因子,要对老因子逐个检验,将方差贡献变为不显著的因子剔除.这种方案是应用杨美川设计的软件在微机上的完成的.

在给定信度 α 结合具体自由度查 t -分布表得 t_α ,按下式计算出临界值.

$$(\gamma_{xy})_\alpha = [t_\alpha^2 / (n - 2 + t_\alpha^2)]^{1/2}$$

n 为样本容量,据此初选出相对相关好的 p 个因子, $p \leq 9$;然后用这些因子进行逐步回归.逐步回归时,各站均按下列 6 种情况分别建立回归预报方程:用全部样品资料(不分季节)建立的实时预报方程和 PP 预报方程;用 7~9 月样品资料(盛夏季节)建立的实时预报方程和 PP 预报方程;用全部资料中除 7~9 月后的其余资料(过渡季节)建立的实时预报方程和 PP 预报方程.

2 计算结果及分析

2.1 计算结果

计算结果包括了上述 6 种情况下各站在逐步回归中挑选到的因子所建立的最优回归方程,共 84 个方程.以湛江站为例,计算结果包括各站的单相关系数临界值、复相关系数、残差估计值(s)、正偏差频率、平均绝对误差、误差、单相关系数、方程的回归系数(即回归方程).本文仅列出其中的 s 及回归方程(表 1),其它 13 个站计算输出的项目与湛江站的相同.

2.2 统计分析

对所建立的 84 个统计预报方程都作了类似于表 2 所列各统计量的统计分析,表 2 仅给出了厦门、湛江、北海、海口站的计算结果,从表 2 可以明显看出:

(1) PP 法预报效果明显比实时预报好,表现为复相关系数前者比后者高,而残差估计则是前者比后者低.

(2) 预报的正偏差频率均大于 50%,表明预报的估计值多数稍高于实况值.

表 1 湛江站的最优回归方程
Tab. 1 Optimal regression equation for Zhanjiang station

6 种情况		s	回 归 方 程
实时 预报 方程	不分季节	1.79	$\hat{y} = 37.8157 + 0.1443x_2 - 0.0474x_4 - 0.0012x_7 - 0.0750x_9$
	盛夏季节	1.98	$\hat{y} = 11.2634 - 0.3074x_3 - 0.023x_7$
	过渡季节	1.53	$\hat{y} = 37.594 + 0.0533x_1 + 0.1564x_2 - 0.0906x_5 - 0.073x_6 - 0.014x_7$
PP 预报 方程	不分季节	1.71	$\hat{y} = 54.2593 + 0.1049x_3 - 0.1733x_4 - 0.0327x_5 - 0.002x_7 - 0.1259x_9$
	盛夏季节	1.75	$\hat{y} = 20.158 - 0.1308x_4 - 0.0029x_7 - 0.1657x_9$
	过渡季节	1.67	$\hat{y} = 59.7317 - 0.1681x_4 - 0.0375x_5 - 0.0010x_7$

(3) 实时预报与 PP 预报所选入的因子不完全相同.

(4) x_1, x_5, x_7 几乎均与预报量 y 呈负相关(表 2), 表明各站气压(x_1)越低, 风速越大; 热带气旋中心与测站的距离(x_7)越近, 风速越大; 热带气旋中心气压(x_5)越低, 即热带气旋强度愈强, 受其影响的台站风速越大.

另外, PP 法计算结果表明, 单纯考虑 7, 8, 9 月热带气旋资料, 因子与预报量的复相关程度较高.

表 2 4 个代表站不分季节的回归方程的统计分析表
Tab. 2 Statistical analysis table of 4 representative stations

类型	站名	复相关系数	残差估计	正偏差频率/(%)
实时 预报 方程	厦门	0.4956	2.25	50.4
	湛江	0.4489	1.79	53.3
	北海	0.4036	2.48	55.3
	海口	0.3889	2.18	56.4
PP 预报 方程	厦门	0.5276	2.22	52.0
	湛江	0.6040	1.71	54.3
	北海	0.5149	2.38	54.1
	海口	0.5757	2.11	51.8

2.3 预报的成功概率

据统计的推断原理^[6,7], 在得出了各站点的各种情况的回归方程后, 用 1988 年的非统计样本实例资料进行预报验证. 将单站各因子实时或同时的相应资料代入相应的方程, 即可算出热带气旋风速的估计值 \hat{y}_i . 因 \hat{y}_i 与 y_i 有差别, 在概率统计理论中必须要算出预报的置信区间, 其估计值 \hat{y}_i 才有预报意义. 在信度为 α 时, 置信区间为

$$((\hat{y}_i - N_{\alpha} s_y) \leq y_i \leq (\hat{y}_i + N_{\alpha} s_y))$$

$$\text{其中, } s_y = \left[\sum_{t=1}^n (y_t - \hat{y}_t)^2 / (n - p - 1) \right]^{1/2}$$

正态分布的双侧分位数 N_α , 在 $\alpha = 0.01, 0.05, 0.10$ 时分别为 $N_{0.01} = 2.58, N_{0.05} = 1.96, N_{0.10} = 1.65$. 概率的保证程度为

$$P(\hat{y}_t - 2.58s_y, \hat{y}_t + 2.58s_y) = 0.99$$

$$P(\hat{y}_t - 1.96s_y, \hat{y}_t + 1.96s_y) = 0.95$$

$$P(\hat{y}_t - 1.65s_y, \hat{y}_t + 1.65s_y) = 0.90$$

也就是说, 实况 y_t 落在上述区间时, 则认为预报成功, 其可信程度分别为 99%、95% 和 90%. 对上述 14 个站点均作了统计, 统计出预报的成功概率, $\alpha = 0.1$ 的统计结果见表 3. 其中 m 为实况落在置信区间的个数, cf 为预报的成功概率 ($cf = m/n \times 100\%$). 预报的样本分三种情况: 不分季节的样本容量为 64; 盛夏季节的样本容量为 26; 过渡季节的样本容量为 38.

表 3 14 测站 1988 年的预报验证统计

Tab. 3 Statistics of the forecast verification of 14 stations in 1988

站名	项目	实时预报			PP 预报		
		不分季节	盛夏季节	过渡季节	不分季节	盛夏季节	过渡季节
厦门	m	62	25	38	62	25	38
	cf	96.85	96.15	100.00	96.88	96.15	100.00
湛江	m	63	26	34	63	26	34
	cf	98.44	100.00	89.47	98.44	100.00	89.47
海口	m	62	26	29	62	26	34
	cf	96.88	100.00	76.32	96.88	100.00	89.47
北海	m	62	26	33	64	26	35
	cf	96.88	100.00	86.84	100.00	100.00	92.11

2.4 主要预报因子

热带气旋造成的东南沿海大风与各因子相关关系的统计结果见表 4. 各站点均有 6 种情况即 6 个方程, 因此共有 84 个方程. 统计每个因子在这些方程中被引进的次数, 分正、负相关两种. 表 4 的数量结果意义分述如下:

表 4 84 个方程中各个因子引进的频率统计

Tab. 4 Frequency statistics of the introduced factors in 84 equations

因 子	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
正相关引进次数	1	18	8	12	1	24	0	15	9
负相关引进次数	32	11	13	25	26	0	51	1	36
总 计	33	29	21	37	27	24	51	16	45
引进频率/(%)	39.29	34.52	25	42.86	32.14	28.57	60.71	19.05	53.57

x_1 : 在热带气旋影响范围内, 站点气压越低越靠近台风中心, 风速也越大.

x_2 : 负相关只出现在钦州, 北海, 厦门和西沙, 其它均为正相关. 这说明在一般情况下, 当热带气旋在南海时广东南部沿海地区处于热带气旋外围下沉气流, 是使人感到闷热的

天气,而钦州等 4 站呈负相关的原因,可能与当地特殊地理位置和气候背景有关。

x_3 :正相关出现在福州、厦门、钦州、北海、广州,其他的为负相关,这主要是由于热带气旋路径的影响所致。正相关多半与在广东登陆的热带气旋路径有关,负相关则与西行热带气旋路径有关。

x_4 :正相关出现在厦门、海丰、汕头、西沙,其余站为负相关,这可能与热带气旋气候路径有关,在南海转向之后在粤东登陆的热带气旋与上面的 4 站呈现正相关关系。西行或在粤西登陆的热带气旋中心所在的经度越小越接近站点,反映为负相关。

x_5, x_6 :均为表示热带气旋强度的量,热带气旋越强风速越大。

x_7 :无论热带气旋强度、路径如何,总是站点距热带气旋中心越近,风力越大,故它是引进预报方程频率最高的因子。

x_8 :台站与热带气旋中心气压差越大反映热带气旋与台站间的气压梯度越大,风力就越大,所以呈正相关关系。

x_9 :福州、厦门、钦州、广州为正相关,其他为负相关,这可能与冷空气南下造成热带气旋靠陆地一侧的气压梯度增大,相应的风力也大有关系。而盛夏无冷空气南下活动的情况下,内陆为热低压,故呈负相关。

3 结 论

(1)14 个站点中,除广州、海丰两站外,其余 12 个站点均表现出 PP 预报的效果比实时预报的好。

(2)无论是 PP 或实时的预报,用 7~9 月时段资料统计所得的预报效果均较佳。

(3)西沙站,过渡季节的预报效果相当显著,其复相关系数达 65%,预报的成功概率达 97.3%。

(4)从因子被引进方程的次数(表 4)可以看出,引进次数最多的是热带气旋中心与站点距离因子(x_7)。

(5)本文所取的因子与预报量的线性单相关系数均较低,但经逐步回归分析选出的若干因子组成的多元回归方程,经检验分析,并用历史资料试报效果是好的,这说明影响热带气旋风力分布的因子是复杂的,简单的回归分析难于建立可靠的统计预报方程。

参 考 文 献

- 1 李叔廷.台风外围风速的计算和预报.1978 年台风会议文集,上海:上海科技出版社,1981
- 2 房文鸾,徐启明,章家琳,等.东海台风域内的风速分布.海洋预报,1987,4(3):1~14
- 3 关芬呈,杨平章.南海台风风场分析及预报.海洋预报,1990,7(2):20~27
- 4 杨美川,朱精福,邓之瀛.台风大风暴雨模式输出概率预报.气象,1981(7):2~5
- 5 国家气象中心编.数值预报产品应用指南.北京:气象出版社,1991.196~200
- 6 黄嘉佑.气象统计分析与预报方法.北京:气象出版社,1990.89~120
- 7 《社会经济统计学原理教科书》编写组编.社会经济统计学原理教科书.北京:中国统计出版社,1990.364~366

The Stepwise Regression Analysis and Prediction of the Wind Speed of Tropical Cyclone over South China Sea

Yang Pingzhang * *Guo Yingqiong* *Zhang Huichan* *Guo Ping*

Abstract Nine synoptic and physics factors are selected to establish 84 statistical forecasting equations by means of stepwise regression analysis. These equations have been used in the prediction of wind speed of the tropical cyclone by 14 stations located in coastal areas of south eastern China. Forecasting verification by using of the data in 1988 that are not included in the statistical samples shows that the equations in the paper are applicable to practical prediction.

Keywords tropical cyclone over South China Sea, stepwise regression analysis, statistical forecasting equation

* Department of Atmospheric Sciences, Zhongshan University, Guangzhou 510275