

# 多级逐步判别在台风路径预报中的应用

李麦村 姚棣荣 杨自强

(中国科学院大气物理研究所) (杭州大学地理系) (上海计算技术研究所)

台风是我国东南沿海重要的灾害性天气之一。广大气象台站在作台风路径预报中往往采用逐步回归分析<sup>[1]</sup>, 在应用判别分析中也仅采用两级判别。为了进一步提高台风预报的精确度, 我们应用了多级判别的方法, 并且在因子的筛选上采用了两段筛选法。逐步判别分析是近几年才发展起来的<sup>[2,3]</sup>, 在天气预报中尚未得到应用。我们在应用这一方法作台风路径预报中收到了较好的效果。

## 一、预报因子的初选

我们取 1958—1965 年 7, 8, 9 三个月, 进入预报区域的 100 个样本的台风作为基本资料<sup>[4]</sup>, 选取 43 个预报因子。并把台风分成三类: 在汕头以南登陆、汕头以北登陆和海上转向的, 分别以 I, II, III 表示。每类样本数分别为  $n_1 = 38$ ,  $n_2 = 29$ ,  $n_3 = 33$ , 总样本数  $N = n_1 + n_2 + n_3$ 。

根据多级判别分析的原理, 我们按下式作为初选因子的原则:

$$I_i = \frac{\sum_{g=1}^G ng(\bar{x}_{ig} - \bar{x}_{i...})^2}{\sum_{g=1}^G \sum_{k=1}^{ng} (x_{igk} - \bar{x}_{ig})^2} \quad (i = 1, \dots, p), \quad (1)$$

式中

$$\bar{x}_{ig} = \frac{\sum_{k=1}^{ng} x_{igk}}{ng} \quad (i = 1, \dots, P, g = 1, \dots, G),$$

$$\bar{x}_{i...} = \frac{\sum_{g=1}^G \sum_{k=1}^{ng} x_{igk}}{N}, \quad i = 1, \dots, P.$$

按(1)式计算, 取  $I_i$  最大的作为初选的因子。我们从 43 个预报因子中经初选得到 20 个(见文末附表)。

为了适宜大量的计算, 我们<sup>[5]</sup>又采用如下伪变量来计算  $I_i$ :

$$(x_{igk} - x_{ig}) = \begin{cases} 1, & \text{当 } x_{igk} - \bar{x}_{ig} > 0, \\ 0, & x_{igk} - \bar{x}_{ig} = 0, \\ -1, & x_{igk} - \bar{x}_{ig} < 0. \end{cases} \quad (2)$$

这样, 对  $I_i$  的计算就大大简化了。同样, 从 43 个预报因子中初选出 20 个(见附表)。

## 二、逐步判别精选因子

与逐步回归的思路类似, 我们采用逐步判别函数。详细的计算过程和程序参阅资料<sup>[2,3]</sup>。经过计算, 由上述两个方案初选的 20 个因子进行三级逐步判别精选所得的判别函数是一致的(所选进因子的  $F$  值见附表)。

$$\begin{aligned} d_1 &= -82.766 + 7.630x_2 + 0.013x_6 \\ &\quad + 0.658x_{11} - 0.474x_{13} - 0.227x_{23} \\ &\quad + 0.163x_{25}, \\ d_2 &= -72.584 + 6.986x_2 + 0.004x_6 \\ &\quad + 0.808x_{11} - 0.655x_{13} - 0.623x_{23} \\ &\quad + 0.046x_{25}, \\ d_3 &= -107.021 + 8.785x_2 + 0.004x_6 \\ &\quad + 1.150x_{11} - 0.405x_{13} - 0.700x_{23} \\ &\quad - 0.112x_{25} \\ &\quad (F_1 = F_2 = 3.0). \end{aligned}$$

本文 1976 年 3 月 15 日收到。

\* 具体方法可参见杨自强: 判别分析与逐步判别一文 (1975 年全国概率统计会议资料)。

对样本的分类(等概率情况下)见表 1.

表 1

| 实 况<br>计 算 | I  | II | III | 合 计 |
|------------|----|----|-----|-----|
| I          | 30 | 2  | 0   | 32  |
| II         | 7  | 26 | 1   | 34  |
| III        | 1  | 1  | 32  | 34  |
| 合 计        | 38 | 29 | 33  | 100 |

同时,对局部的判别效果和全部的判别效果采用 Mahalanobis 距离  $D^2$  和 Wilks 量  $A$  进行显著性检验,其结果如表 2.

表 2

| Mahalanobis 距离   |       |        | Wilks 量            |                     |
|------------------|-------|--------|--------------------|---------------------|
| $D^2$            |       |        | $A$                |                     |
| II-I             | III-I | III-II |                    |                     |
| 3.94             | 8.83  | 10.62  | 0.20               |                     |
| F(6.92)          |       |        | F(12.184)          | $\chi^2(12)$        |
| 10.24            | 24.64 | 25.91  | 18.73              | 152.46              |
| $F_{0.01}(6.92)$ |       |        | $F_{0.01}(12.184)$ | $\chi^2_{0.01}(12)$ |
| 3.03             |       |        | 2.27               | 26.22               |

结果表明,这一判别效果是相当显著的.另外,我们采用同样的初选得到的因子,进行两级逐步判别,即台风登陆或海上转向,以 I, II 表示,此时两类的样本数分别为  $n_1 = 67$ ,  $n_2 = 33$ . 由此所得的判别函数 ( $F_1 = F_2 = 3.0$ ) 和对样本的分类分别(不等概率情况下)如表 3, 4.

(1) 原始初选方案

$$d_1 = -1725.92 + 28.00x_1 - 7.18x_4 + 0.23x_{11} + 0.72x_{18} - 2.66x_{23} - 1.69x_{25},$$

表 3

| 实 况<br>计 算 | I  | II | 合 计 |
|------------|----|----|-----|
| I          | 64 | 1  | 65  |
| II         | 3  | 32 | 35  |
| 合 计        | 67 | 33 | 100 |

$$d_2 = -1805.97 + 28.36x_1 - 6.11x_4 + 0.62x_{11} + 0.81x_{18} - 3.04x_{23} - 1.94x_{25}.$$

(2) 简化初选方案

$$d_1 = -1999.69 + 32.26x_1 - 7.15x_4 + 0.29x_{11} + 0.88x_{18} - 3.28x_{23} - 3.11x_{25} - 0.32x_{29} + 2.04x_{35},$$

$$d_2 = -2109.15 + 32.84x_1 - 6.07x_4 + 0.69x_{11} + 0.98x_{18} - 3.68x_{23} - 3.44x_{25} - 0.34x_{29} + 2.13x_{35}.$$

表 4

| 实 况<br>计 算 | I  | II | 合 计 |
|------------|----|----|-----|
| I          | 66 | 1  | 67  |
| II         | 1  | 32 | 33  |
| 合 计        | 67 | 33 | 100 |

对二级情况也经过判别效果显著性检验,表明它们的判别也是相当有效的.

### 三、试 报 结 果

我们应用上述方案,对 1975 年进入预报区域(即图 1 梯形框区域)的 3 号、4 号、7 号和 11 号台风共作了 12 次预报. 预报结果表明,不论是原始方案,还是简化方案;也不论是二

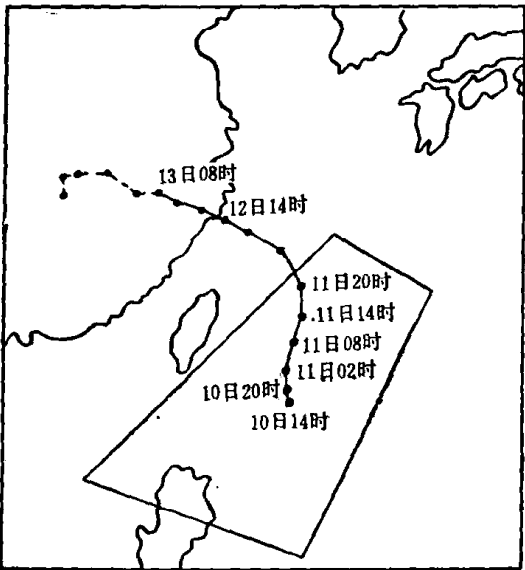


图 1 1975 年 4 号台风路径图  
(圆点为不同时刻台风位置,梯形框为本模式起报区)

级判别,还是三级判别,结果是一致的,其预报准确率都为  $\frac{11}{12} = 92\%$ 。试报具体结果如表 5 和表 6。

表 5 两级判别分类表

| 计 算 \ 实 况 | 登 陆 | 转 向 | 合 计 |
|-----------|-----|-----|-----|
| 登 陆       | 8   | 0   | 8   |
| 转 向       | 1   | 3   | 4   |
| 合 计       | 9   | 3   | 12  |

表 6 三级判别分类表

| 计 算 \ 实 况 | 北 登 | 南 登 | 转 向 | 合 计 |
|-----------|-----|-----|-----|-----|
| 北 登       | 8   | 0   | 0   | 8   |
| 南 登       | 0   | 0   | 0   | 0   |
| 转 向       | 1   | 0   | 3   | 4   |
| 合 计       | 9   | 0   | 3   | 12  |

这里需要特别提出的是对 4 号台风的预报。4 号台风的路径比较特殊,在 8 月 10 日 14 时中央气象台开始编号,此时便开始向西北方向移动,到 11 日 02 时以后改为向东北方向移动,到 11 日 20 时以后忽然又折向西北方向前进,最后在浙江温岭登陆(图 1)。由于在一段不长的时间中,台风移动方向反复改变,所以一些台站预报结果较差。但用此法在 10 日 08 时、10 日 20 时和 11 日 08 时,分别对这次台风移动作了三次预报。无论是二级判别还是三级判别,三次计算结果都是预报在汕头以北登陆,与实况符合。特别是在 11 日 08 时那次计算,虽然此时台风已经转向东北方向,似乎要在海上转向北上,一些台站错报也就在于一时受了这种暂时现象所迷惑。而我们的计算结果,仍预报登陆。这说明本文所提供的方法在预报台风路径中是相当有效的。

#### 四、结 论

通过以上讨论可以得到以下几点结论:

1. 本文针对大量资料对台风路径进行多级逐步判别分析,结果令人满意。一方面进入预报方程的因子数目不多,另一方面预报方程的历史拟合率三级判别达 90% 左右,而二级判别达 98%。在 1975 年预报实践中,共作 12 次预报,仅错报一次,准确率都达 92%。这说明近年来,按数理统计发展起来的逐步判别方法,在台风路径这样一类天气预报中是一种行之有效的方法;也说明两段筛选这种方法能选出主要的、本质的因子。事实上,本文提出的预报方程中的那些因子,都反映了台风本身位置、副热带高压结构及台风周围引导气流的特征,都有一定的天气物理意义。

2. 三种初选因子方案中,以人为初选效果最差\*。这表明客观、定量而简便的初选方法是很必要的。简易判别初选效果也很好。这种方法简便,适于大量计算和广大气象台站使用。关于简易判别,将另文讨论。

3. 与资料[4]相比,本文提供方法在下列两方面有所改进:首先不需要利用因子作点聚图和计算条件概率,而是单因子直接进入方程。其次,我们在预报方程中,略去了本身需要事先估计的台风中心后 12 小时位置作为预报因子。这对预报方程的稳定性和使用上都有一定的好处。

#### 参 考 资 料

- [1] Jarrell, J. D., Mauck, C. J. & Renard, Mon. Wea. Rev., D 103(1975), 674—684.
- [2] Afifi, A. A., Azen, S. P., Statistical Analysis: A Computer Oriented Approach, Chap. 5, Academic press, New York and London, 1972.
- [3] Dixon, W. J., BMD. Biomedical Computer programs, University of California Press, London, 1973.
- [4] 刘韵清、姚稼荣、祝启恒、金一鸣,数值预报和数理统计预报会议文集,科学出版社,1974, 160—170.
- [5] 李麦村,近代气象学若干问题的进展,科学出版社, 1975, 46—74.

\* 人为初选因子,三级判别历史拟合率为 85%,二级判别为 91%。

附表 预报因子初选、精选表 (表中有 F 值的表示已入选的因子)

| 精 选<br>人为初选 |                       | 二级判别  | 三级判别  | 精 选<br>原始初选 |                               | 二级判别  | 三级判别  | 精 选<br>简化初选 |                                  | 二级判别  | 三级判别  |
|-------------|-----------------------|-------|-------|-------------|-------------------------------|-------|-------|-------------|----------------------------------|-------|-------|
| 序号          | 代 号                   | 引入的 F |       | 序号          | 代 号                           | 引入的 F |       | 序号          | 代 号                              | 引入的 F |       |
| $x_1$       | $\lambda_0$           |       | 4.32  | $x_1$       | $\lambda_0$                   | 3.80  |       | $x_1$       | $\lambda_0$                      | 7.14  |       |
| $x_2$       | $\varphi_0$           |       | 47.92 | $x_2$       | $\varphi_0$                   |       | 47.92 | $x_2$       | $\varphi_0$                      |       | 47.92 |
| $x_4$       | $\varphi_{12}$        | 72.33 | 4.90  | $x_3$       | $\lambda_{12}$                |       |       | $x_4$       | $\varphi_{12}$                   | 72.33 |       |
| $x_7$       | "上海                   |       |       | $x_4$       | $\varphi_{12}$                | 72.33 |       | $x_5$       | $\varphi_0 + \Delta\varphi_{24}$ |       |       |
| $x_8$       | "上海                   | 4.53  | 5.04  | $x_5$       | $\varphi_0 + \Delta\varphi_2$ |       |       | $x_6$       | $\xi_G$                          |       | 5.80  |
| $x_9$       | "大陈                   | 11.51 | 7.88  | $x_6$       | $\xi_G$                       |       | 5.80  | $x_7$       | "上海                              |       |       |
| $x_{10}$    | "大陈                   |       |       | $x_7$       | "上海                           |       |       | $x_{11}$    | "福州                              | 32.74 | 17.50 |
| $x_{12}$    | "汕头                   |       |       | $x_{10}$    | "大陈                           |       |       | $x_{13}$    | " $g_1$                          |       | 4.70  |
| $x_{13}$    | " $g_1$               |       | 5.47  | $x_{11}$    | "福州                           | 32.74 | 17.50 | $x_{15}$    | $G_{02}$                         |       |       |
| $x_{15}$    | $G_{02}$              |       |       | $x_{13}$    | " $g_1$                       |       | 4.70  | $x_{17}$    | $\Delta\varphi_{588}$            |       |       |
| $x_{16}$    | $G_{0\varphi}$        |       | 6.37  | $x_{14}$    | " $g_1$                       |       |       | $x_{18}$    | $\lambda_{588}^w$                | 6.62  |       |
| $x_{17}$    | $\Delta\varphi_{588}$ |       |       | $x_{16}$    | $G_{0\varphi}$                |       |       | $x_{19}$    | " $g_2$                          |       |       |
| $x_{18}$    | $\lambda_{588}^w$     |       |       | $x_{17}$    | $\Delta\varphi_{588}$         |       |       | $x_{23}$    | $\Delta(OT)_2$                   | 7.77  | 10.48 |
| $x_{19}$    | " $g_2$               | 9.69  |       | $x_{18}$    | $\lambda_{588}^w$             | 6.62  |       | $x_{24}$    | $\xi_{25}$                       |       |       |
| $x_{20}$    | " $T$                 |       | 3.65  | $x_{19}$    | " $g_2$                       |       |       | $x_{25}$    | $\xi_{44}$                       | 8.65  | 9.52  |
| $x_{21}$    | " $T$                 |       |       | $x_{23}$    | $\Delta(OT)_2$                | 7.77  | 10.48 | $x_{26}$    | $\xi_{\text{貝}}$                 |       |       |
| $x_{23}$    | $\Delta(OT)_2$        | 6.47  | 7.57  | $x_{25}$    | $\xi_{44}$                    | 8.65  | 9.52  | $x_{29}$    | $\pm  \Delta H_{44} ^2$          | 3.12  |       |
| $x_{24}$    | $\xi_{25}$            |       |       | $x_{37}$    | $H_{81}$                      |       |       | $x_{30}$    | $\pm  \Delta H_{50} ^2$          |       |       |
| $x_{36}$    | $H_{36}$              |       |       | $x_{39}$    | $H_{102}$                     |       |       | $x_{33}$    | $\pm  \Delta H_{72} ^2$          |       |       |
| $x_{38}$    | $H_{96}$              |       |       | $x_{40}$    | $H_{105}$                     |       |       | $x_{35}$    | $\pm  \Delta H_{104} ^2$         | 3.99  |       |

表中因子的意义:  $\xi_{44}$  为日本海附近的涡度,  $H_{105}$  为第 105 点 500 mb 高度,  $\xi_{\text{貝}}$  为贝加尔湖附近的涡度,  $\pm |\Delta H_{72}|^2$  为第 72 点 24 小时变高的绝对值平方, 其它因子的意义参看资料 [4] 中说明。