

# 汛期大~暴雨的降水概率预报模式

钟元 吴钟浚 李泓 张珊璞

(浙江省气象科学研究所, 杭州 310021)

朱惠群 徐霜芝

(浙江省气象局)

## 提 要

应用阶梯函数变换预报因子, 经过相关比筛选, 进行 REEP 构造分量级降水概率预报方程, 应用不同时刻资料和对 NWP 产品统计释用方法构造分时刻预报方程, 并对其预报结果进行了综合的集合预报。实用表明, 预报模式对大~暴雨的概率具有一定的预报能力。

**关键词:** 大~暴雨 降水概率预报 阶梯函数 集合预报

## 引 言

由于事件概率天气预报比传统的天气预报更具有经济价值<sup>[1]</sup>, 近年来得到迅速发展。目前国内的降水概率预报 (POP) 大多只给出降水可能性的概率, 未涉及其量的大小。随着人们对天气变化同时具有确定性和随机性认识的深入, 以及社会经济发展对经济决策提出愈来愈高的量化与精细化的需要, 要求 POP 给出量级降水概率预报已成为不可少的内容之一。

大~暴雨属小概率事件, 对其发生的有关先兆信息的了解与收集十分有限, 因而进行大~暴雨的概率预报是 POP 中的一个难点。

近年来数值天气预报 (NWP) 的进展和

监测、通讯技术的不断进步, 给大~暴雨降水概率预报提供了丰富的预报信息。本文应用 NWP 产品的统计释用、阶梯函数变换因子和集合预报等技术, 提出一个汛期大~暴雨概率的预报模式。

## 1 大~暴雨的气候概率

本文讨论浙江省的杭州、宁波、温州、湖州、嘉兴、绍兴、金华、衢州、丽水、台州及舟山 11 个城市 4~6 月的大~暴雨预报。

资料为 1990~1996 年 4~6 月的 637 个样本, 其中小雨以上 (日降水量  $R \geq 0.1\text{mm}$ )、中雨以上 ( $R \geq 10\text{mm}$ )、大雨以上 ( $R \geq 25\text{mm}$ ) 及大~暴雨以上 ( $R \geq 38\text{mm}$ ) 事件发生的气候概率 ( $P_c$ ) 见表 1。

表 1 浙江省各量级降水发生的气候概率

	$R \geq 0.1\text{mm}$	$R \geq 10.0\text{mm}$	$R \geq 25.0\text{mm}$	$R \geq 38.0\text{mm}$
事件发生总次数	304~381	99~157	28~72	15~51
$P_c(\%)$	47.8~60.0	15.6~24.7	4.4~11.2	2.8~8.1
$P_r(1 P_l=1)(\%)$	100	32.6~41.2	9.3~18.8	4.9~13.5
$P_r(1 P_m=1)(\%)$	/	100	28.6~45.5	15.8~32.7

从表 1 可见, 浙江省汛期大雨及大~暴雨事件发生的气候概率只有 4%~11% 及 3%~8%。对如此小概率事件, 用线性相关与

事件概率回归估计 (REEP) 的方法构造预报模式几乎不可能。分析与试验表明, 对于  $P_c < 20\%$  的事件, 应用 REEP 构造的预报方程

甚至通不过信度 0.30 的检验,历史拟合率均在 50% 以下,根本无法应用。

## 2 大~暴雨概率预报框架

针对大~暴雨  $P_c$  过小带来构造预报模式的困难,本文采用分级逐步预报的方法。从表 1 可见  $R \geq 10\text{mm}$  中雨事件的  $P_c$  约为 16% ~ 25%,但在有降水条件 ( $P_l = 1$ ) 下,占有降水事件的概率  $P_m(1|P_l = 1)$  约达 33% ~ 41%。同样在有中雨条件 ( $P_m = 1$ ) 下,大雨与大~暴雨事件占有中雨事件的概率  $P_{he}(1|P_m = 1)$  与  $P_{ha}(1|P_m = 1)$  分别约为 29% ~ 45% 与 16% ~ 33%。统计表明,有降水条件下的中雨事件概率及有中雨条件下的大~暴雨事件概率都不属小概率事件,可用 REEP 构造预报模式。

为此,设定构造大~暴雨概率预报模式的框架:

①构造  $R \geq 0.1\text{mm}$  小雨以上降水概率预报模式:

$$\hat{P}_l = (Y = 1|x) = f_l(x)$$

②构造有降水条件下中雨以上降水概率预报模式:

$$\hat{P}_m = (Y = 1|x, P_l = 1) = f_m(x)$$

③构造有中雨条件下大雨以上降水概率预报模式:

$$\hat{P}_{he} = (Y = 1|x, P_m = 1) = f_{he}(x)$$

④构造有中雨条件下大~暴雨以上降水概率预报模式:

$$\hat{P}_{ha} = (Y = 1|x, P_m = 1) = f_{ha}(x)$$

## 3 阶梯函数与因子变换

### 3.1 阶梯函数

预报因子性能的优劣是预报成功与否的关键。在 POP 中,预报量由降水量转化为大~暴雨事件发生的概率,即 (0,1) 化。随后必须对各种因子作相应处理,以使它们与 (0,1) 化的预报量相适应。目前多数 POP 预报的处理方法是把因子序列  $x_j$  作标准化处理:

$$\hat{x}_j = (x_j - \bar{x})/\sigma_x$$

$$\bar{x} = \frac{1}{N} \sum_{j=1}^N x_j \quad (j = 1, 2, \dots, N)$$

$$\sigma_x = \sqrt{\frac{1}{N-1} \sum_{j=1}^N (x_j - \bar{x})^2}$$

式中  $N$  为序列样本总量。

然后统计  $\hat{x}_j$  与预报量  $y_j$  的线性相关,再进行 REEP 分析构造预报模式。

事实上,绝大多数因子  $x$  与事件发生概率  $P(Y = 1|X = x)$  的关系是非线性的。设事件  $Y$  发生的概率随因子  $x$  的分布为一简单的准正态分布函数(图 1)。当通过线性相关得到一直线  $l = ax + b$  来拟合  $P(1|x)$  时,显然仅在区间  $[x_1, x_2]$  内  $l$  可近似拟合  $P(1|x)$ ,而在  $x < x_1$  或  $x > x_2$  的广大区间,  $l$  与  $P(1|x)$  有很大的误差。因此,线性相关的  $l$  难以拟合非线性的  $P(1|x)$  分布。

在此,把因子序列  $\{x_j\}$  在其值域  $[x_{\min}, x_{\max}]$  内分为  $H$  个区间  $h_i$ ,在  $i$  区间内,事件  $Y$  发生的平均概率为:

$$\hat{P}_i(1|x) = \frac{M_i}{N_i} \quad \begin{cases} i = 1, 2, \dots, H \\ x \in h_i \end{cases} \quad (1)$$

式中  $N_i$  为落在区间  $h_i$  的总样本数,  $M_i$  为该区间内事件  $Y$  发生的样本数。

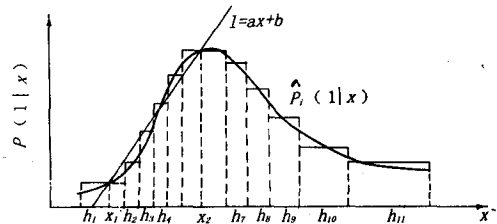


图 1 事件发生概率  $P(Y = 1|X = x)$  随  $x$  的分布及阶梯函数  $\hat{P}_i(1|x)$

因子  $\{x_j\}$  在所有区间的  $\hat{P}_i(1|x)$  构成了一阶梯函数。由此,因子序列  $\{x_j\}$  变换为  $\hat{P}_i(1|x)$  分布的序列,即

$$\{x_j\} \Rightarrow \hat{P}_{ij}(1|x) \quad \begin{cases} i = 1, 2, \dots, H \\ j = 1, 2, \dots, N \\ x \in h_i \end{cases}$$

从图 1 可见阶梯函数  $\hat{P}_i(1|x)$  对  $P(1|x)$  的拟合明显优于  $l$ ,尤其在  $[x_{\min}, x_1]$  与  $[x_2, x_{\max}]$  的区间内。

### 3.2 相关比

在统计非线性因子  $\hat{P}_{ij}(1|x)$  与事件  $Y$  发生的相关时,宜用相关比  $\eta_{yx}$ <sup>[2]</sup>:

$$\eta_{yx} = \sqrt{1 - \frac{1}{\sigma_y^2} E[(Y - E[Y|x])^2]} \quad (2)$$

式中  $Y$  的标准差  $\sigma_y = \sqrt{E[(Y - E[Y])^2]}$ ,  $E[(Y - E[Y|x])^2]$  代表  $Y$  的自回归曲线  $y_x = E[Y|x]$  的最小均方离差。

当  $Y$  只取 0,1 时,  $\hat{P}_c$  为  $Y = 1$  出现的气候概率估计值,  $\hat{P}_i(1|x)$  为阶梯函数, 则式(2)可化为:

$$\eta_{yx} = \sqrt{\frac{\sum_{i=1}^H \hat{w}_i (\hat{P}_i(1|x) - \hat{P}_c)^2}{\hat{P}_c(1 - \hat{P}_c)}} \quad (3)$$

$\left\{ \begin{array}{l} i = 1, 2, \dots, H \\ x \in h_i \end{array} \right.$

式中  $\hat{w}_i = N_i/N$ ,  $\hat{P}_c = M/N$  ( $M = \sum_{i=1}^H M_i$ ,  $M$  为  $N$  个样本中  $Y = 1$  出现的总次数)。

通过式(1)的变换和式(3)的估算,便可评估因子  $\{x_j\}$  与事件  $Y$  发生的非线性相关。

### 3.3 变换对因子预报性能的影响

对 500hPa 高度场  $H$  (20 ~ 90°N, 90 ~ 180°E, 5° × 5° 网格)、地面气压场  $P$  (20 ~ 90°N, 90 ~ 180°E, 5° × 5° 网格)、200hPa 与

表 2 因子变换前后与杭州市大~暴雨事件相关

	$H$	$P$	$U_{200}$	$V_{200}$	$U_{850}$	$U_{850}$	$\zeta_{200}$	$D_{200}$	$\zeta_{850}$	$D_{850}$
$r_{\max}(\hat{x}_j)$	0.32	-0.35	0.29	0.24	0.24	-0.20	0.27	-0.26	0.27	0.23
$\eta_{\max}(\hat{P}_{ij}(1 x))$	0.48	0.50	0.44	0.41	0.45	0.39	0.45	0.45	0.42	0.39
$n(\hat{x}_j)$	2	24	0	0	0	0	0	0	0	0
$n\hat{P}_{ij}(1 x)$	187	123	108	49	63	41	38	35	41	38

### 3.4 因子变换的若干问题

在进行因子变换时,区间  $h_i$  的划分至关重要。一方面  $h_i$  应充分小,这样可避免  $\hat{P}_i(1|x)$  在区间内过多平滑,以确保其精度;另一方面  $h_i$  又不宜过细,以使每个区间内有足够的样本数,  $\hat{P}_i(1|x)$  才具有代表性和稳定性,避免其值受个别样本的异常影响而产生大的波动。

由于区间  $h_i$  的大小要适应  $P(1|x)$  的非

850hPa  $u$  和  $v$  风场 (35°N ~ 35°S, 90 ~ 180°E, 5° × 5° 网格)、涡度场  $\zeta$  和散度场  $D$  (30°N ~ 30°S, 95 ~ 175°E, 5° × 5° 网格) 等因子的标准化序列  $\hat{x}_j = (x - \bar{x})/\sigma_x$  处理及经过阶梯函数变换后的  $\hat{P}_{ij}(1|x)$  序列计算其与杭州市汛期大 ~ 暴雨事件概率序列  $Y$  的最高线性相关系数  $r_{\max}$  和最高非线性相关比  $\eta_{\max}$  以及  $|r| \geq 0.30$  和  $\eta \geq 0.30$  的高相关因子数  $n$ , 结果如表 2 所示。

从表 2 可见,因子经阶梯函数变换后,最高相关与高相关因子数均有大幅度提高。

原始(标准化)因子与大~暴雨事件的线性相关不好,是由于因子往往只在某些区间与事件发生有关,而在另一些区间则完全无关。线性相关只统计因子在所有区间的平均相关,完全不考虑在不同区间具有不同的相关性。经过阶梯函数变换后,相关比分别统计每个区间与事件的相关,在某些区间因子与事件发生高相关,而另一些区间因子与事件不发生高相关,这样便提高了因子与事件的总体相关比。这表明因子序列用阶梯函数变换后,能反映其与大~暴雨事件非线性概率分布的内在关系,其预报性能明显优于原因子序列。

线性分布,因而不应是均匀的,以保证  $\hat{P}_i(1|x)$  的精度和可靠性。

阶梯函数  $\hat{P}_i(1|x)$  表达了区间  $h_i$  内事件发生的平均概率,其值在区间内恒定。这有悖事实,并导致变换误差。 $\hat{P}_i(1|x)$  在相邻区间变化时存在间断点,在间断点附近  $x$  值的微小变化可能导致  $\hat{P}_i(1|x)$  剧烈振荡,产生变换误差。为解决这个问题 Parzen 提出一种概率密度的无偏估计模型<sup>[3]</sup>,它在  $x$  的区间  $h_i$

内用双曲函数代替应用直方图法<sup>[4]</sup>的阶梯函数,使概率拟合曲线在相邻区间连续而无间断点,从而避免了拟合误差。然而在实用中,面对大量因子的变换与筛选,该模型的计算量剧增,操作不便;且实际效果与阶梯函数变换相差不大。相比之下,阶梯函数变换更简洁实用。

#### 4 分级预报模式构造

##### 4.1 预报因子

考虑到大~暴雨降水概率预报的难度,必须尽量应用各种预报信息场资料,包括各

$$\hat{P}_r(Y = 1 | X_1, X_2, \dots, X_M) = b_0 + \sum_{k=1}^M b_k X_k$$

$$X_K = \hat{P}_{ik}(1 | x)$$

$$\left. \begin{array}{l} r = l, m, he, ha \\ k = 1, 2, \dots, M \\ i = 1, 2, \dots, H \\ x \in h_i \end{array} \right\}$$

##### 5 大~暴雨降水概率的集合预报

由于 NWP 初始场存在不可避免的误差,在积分过程中可能迅速增长,通过非线性相互作用影响大气运动,从而在有限的时间内整个大气运动变得不确定。这时的大气状态可认为是随机性的,预报就是对它的概率分布进行估计。为了能更好地描述初始场状况,需要尽可能多的样本与集合。将大气环流模式作用于该集合,所得到的全部解便组成大气未来发展的概率分布。国外在 90 年代发展了一种集合预报 Ensemble Prediction<sup>[5~7]</sup>:在实际观测和分析所得的初始场基础上叠加适当小扰动,形成稍有差别的初始场,作出多个动力延伸预报,然后对其结果进行综合预报。由此,可以对不同时刻、不同种类的 NWP 产品,用不同预报方法分别作统计释用预报后,进行综合得到集合预报。

由于实际预报的起始时刻比 NWP 初始时刻落后 24 小时(图 2),而降水事件发生在距 NWP 初始时刻后 24~48 的时段内,因而 NWP 产品的 24 与 48 小时预报产品  $\hat{x}_{24}$  及  $\hat{x}_{48}$  分别与降水的开始与结束时刻对应,而均不与降水的全过程完全对应。可见,  $\hat{x}_{24}$  与  $\hat{x}_{48}$  均含有降水的信息,只有同时引入,才能较好地反映降水事件的发生。

层次高度场、温度场、流场、物理量场的客观分析,测站观测实时地面与探空及 NWP 预报产品资料。

##### 4.2 预报因子预处理

应用式(1)将预报因子序列变换成阶梯函数分布的新序列,再应用式(3)计算每个因子的相关比  $\eta_{yx}$ ,从中选取通过信度  $\alpha = 0.05 \sim 0.10$  检验的预选因子。

##### 4.3 分级预报模式构造

分别对不同量级降水事件的预选因子应用 REEP 构造预报方程:

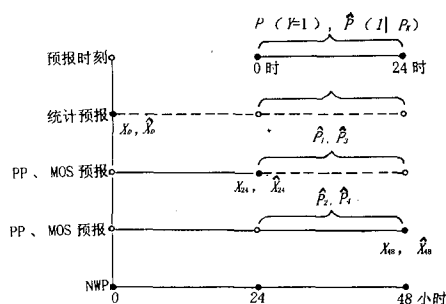


图 2 24 小时降水概率预报时段及多时刻预报因子的预报与集合预报

由此,用客观分析或实测资料  $x_0, x_{24}$  和  $x_{48}$  分别构造大~暴雨概率的统计天气学预报方程和完全预报方程:

$$\hat{P}_0(Y = 1 | X) = f_0(x_0)$$

$$\hat{P}_1(Y = 1 | X) = f_{PP}(x_{24})$$

$$\hat{P}_2(Y = 1 | X) = f_{PP}(x_{48})$$

用 NWP 输出的 24、48 小时的预报产品  $\hat{x}_{24}, \hat{x}_{48}$  分别构造大~暴雨概率预报方程:

$$\hat{P}_3(Y = 1 | X) = f_{mos}(\hat{x}_{24})$$

$$\hat{P}_4(Y = 1 | X) = f_{mos}(\hat{x}_{48})$$

最后,对以上 5 个分时刻预报方程的预报结果应用全回归分析构造集合预报方程:

$$\hat{P}(Y = 1 | \hat{P}_k) = B_0 + \sum_{k=0}^L B_k \hat{P}_k$$

$$(k = 0, 1, 2, 3, 4)$$

表 3 为杭州市大~暴雨概率的分时刻预报方程与集合预报的回归参数。表中  $F$  为  $F$  检验值,  $L$  为因子个数,  $R$  为复相关系数,  $S$  为剩余标准差,  $\hat{P}'$  为应用标准化因子构造的预报方程。

表 3 杭州市大~暴雨概率预报方程的回归参数

	$\hat{P}'$	$\hat{P}_0$	$\hat{P}_1$	$\hat{P}_2$	$\hat{P}_3$	$\hat{P}_4$	$\hat{P}$
$F$	1.7	4.0	5.0	9.0	8.0	4.0	21.0
$L$	14	13	13	12	11	12	5
$R$	0.54	0.83	0.86	0.93	0.92	0.88	0.96
$S$	0.46	0.28	0.22	0.17	0.18	0.20	0.12

由表 3 可见,用标准化因子构造的预报方程  $\hat{P}'$  的效果最差,应用阶梯函数变换因子构造的预报方程  $\hat{P}_0 \sim \hat{P}_4$  效果甚优于  $\hat{P}'$ 。应用 NWP 产品的  $\hat{P}_1 \sim \hat{P}_4$  的效果优于应用初

始时刻资料的  $\hat{P}_0$ ,在 PP 方法中,应用  $x_{48}$  的  $\hat{P}_2$  效果略优于应用  $x_{24}$  的  $\hat{P}_1$ 。而在 MOS 方法中,应用  $\hat{x}_{24}$  的  $\hat{P}_3$  效果略优于应用  $\hat{x}_{48}$  的  $\hat{P}_4$ 。集合预报  $\hat{P}$  的拟合效果明显优于所有分时刻的预报。这是由于集合预报综合了较多的信息,并使分时刻预报的误差得到了平滑的结果。

### 6 预报结果的评估

应用预报模式对浙江省 1997~1998 年 4~6 月 11 个城市进行了降水概率预报(见表 4),预报结果分别应用  $B$ 、 $B_s$  和  $Bias$  评分<sup>[8,9]</sup>(见表 4),其中  $\hat{P}_i$  有 24~120 小时 5 种预报时效,其余 3 种降水概率预报仅为 24 小时一种时效,同时列出标准化因子与阶梯函数变换因子两种结果对比。

表 4 浙江省汛期降水概率预报评估

预报时效/h	$\hat{P}_i(R \geq 0.1mm)$					$\hat{P}_m(R \geq 10.0mm)$		$\hat{P}_{he}(R \geq 25.0mm)$		$\hat{P}_{ha}(R \geq 38.0mm)$	
	24	48	72	96	120	标准化	阶梯变换	标准化	阶梯变换	标准化	阶梯变换
$B$	0.13	0.16	0.20	0.23	0.28	0.38	0.11	0.31	0.10	0.26	0.07
$B_s(\%)$	22	19	13	10	6	4	18	-15	11	-26	5
$Bias(\%)$	8	-3	-1	-8	-5	-55	35	-104	52	-168	78

对于  $\hat{P}_i$ ,  $B$  与  $B_s$  评分均随时效增加而变差,这是由于较长预报时效的 NWP 产品精度下降所致。但在所有预报时效内,评价预报技巧的  $B_s$  评分均为正的预报技巧。从  $Bias$  看,除 24 小时有过多预报倾向外,48~120 小时均为过少预报倾向。

对于  $\hat{P}_m$ 、 $\hat{P}_{he}$  和  $\hat{P}_{ha}$ ,在  $B$  评分中,经阶梯函数变换因子的预报结果明显优于只经标准化因子的预报。这里可注意到,中雨、大雨、大~暴雨的  $B$  评分甚至优于小雨。这是由于小雨、大~暴雨属小概率事件,在  $B$  评分中它们被“淹没”于事件不发生的大概率预报的评分中。可见  $B$  评分是对概率预报总体的评价,不能对小概率事件预报作出评估。而  $B_s$  评分能评价小概率事件的预报技巧。我们看到,经过阶梯函数变换因子的预报结果均为正技巧,虽然对大~暴雨预报的评分不高,但仍为正技巧。而只经标准化因子的预报模式对大雨和大~暴雨的预报均为负技巧,即实际上无预报能力。在  $Bias$  评分上,经变换

因子的预报有较小的正偏差,而标准化因子的预报有很大的负偏差,即它们的预报大大偏离了气候概率。

### 7 结论

①大~暴雨属小概率事件,对其预报宜采用分量级逐步预报方法,使较大量级降水概率预报建立在较小量级降水概率预报的基础之上。

②建立在区间概率分布基础上的阶梯函数较好地反映条件概率的非线性分布,经过阶梯函数变换后的非线性因子预报性能大大优于线性相关因子。阶梯函数区间的适当划分是构造较优变换的关键,用阶梯函数变换因子与相关比相结合有利于进行大~暴雨的 PEEP。

③采用不同时刻资料和 NWP 产品构造分时刻大~暴雨概率预报方程,再对分时刻的预报结果进行综合得到集合预报,由于包含了较多的预报信息和对分时刻预报误差的平滑而提高了预报效果。

④对两年实用预报结果的评估表明,本文构造的大~暴雨概率预报模式具有一定的预报技巧。

参考文献

1 A. H. Murphy 等. 大气科学中的概率、统计和决策. 北京:气象出版社,1991:167~183.  
 2 么枕生. 气候统计学基础. 北京:科学出版社,1984:235~240.  
 3 Parzen E. An estimation of a probability density function and mode, Ann. Math Stat. ,1962,33.  
 4 K. Fukunaya. Introduction to statistical pattern recognition, Academic Press, 1972.  
 5 Traction M. S. and E. Kalnay. Operational ensemble

prediction at the National Meteorological Center: Practical aspects. Wea. and Forecasting, 1993, 8(3): 378~398.  
 6 Molteni E. et al. The ECMWF ensemble prediction system: Methodology and Validation. Q. J. R. Metc. Soc. 1996, 122: 73~119.  
 7 Petraliagis T. et al. Extremen rainfall prediction using the ECMWF ensemble prediction system, 1996.  
 8 Murphy A. H. Comparative evaluation of categorical and probabilistic forecasts. Mon. Wea. Rev. 1986, 114(1): 245~249.  
 9 Murphy A. H, Q. Ye. Comparison of objective and subjective precipitation probability forecasts: The sufficiency relation. Mon. Wea. Rev. 1990, 118(4): 874~882.

## A Probability Forecast Model of Heavy Rainfall in Rainy Season

Zhong Yuan Wu Zhongjun Li Hong Zhang Shanying

(Zhejiang Research Institute of Meteorological Sciences, Hangzhou 310021)

Zhu Huiqun Xu Shuangzhi

(Zhejiang Meteorological Bureau, Hangzhou 310002)

### Abstract

A probability forecast model of heavy rainfall was constructed by developing probability of precipitation amount (PoPA) and based on probability of moderate precipitation. Step function based on estimating POP in suitable intervals more accurately describes nonlinear probability distribution and closely relates to probability of heavy rainfall. By using predictors transformed into step function and estimating correlation ratio between predictors and heavy rainfall event, the regression estimation of event probabilities (REEP) of PoPA was completed. By using multi-time data and NWP products, multi-time predictive equations were constructed, and an ensemble prediction that synthesized multi-time predictions improved forecast effect. Forecasting tests indicate that the probability forecast model of heavy rainfall is efficacious.

**Key Words:** heavy rainfall probability of precipitation step function ensemble prediction