

# 舟山市汛期降水量预报方法<sup>①</sup>

王 雷 曹美兰

(浙江舟山市气象台, 316004)

## 提 要

利用基于预测误差平方和(PRESS)准则的逐步回归分析和基于残差平方和(RSS)准则的逐步回归分析建立了舟山市3个县市区汛期(5~9月)降水量的预报模型,并对两种方法的预测结果进行预报集成,经试报和预报检验表明,该模型的预测效果较好。

关键词: PRESS 准则 RSS 准则 逐步回归 汛期 降水量预报

## 引 言

舟山地处海岛,年平均降水量在936~1620mm之间,其中5~9月降水量占了全年总降水量的60%以上。从月分布情况来看,舟山地区降水量的月际分布曲线呈双峰型,6月和9月出现两个降水高峰,这就是梅汛期和台汛期。由于舟山特定的地理位置,旱涝灾害对舟山的影响很大。如果针对旱涝等短期气候异常及趋势提前作出较为准确的预测,并及时采取必要的防御措施,则可以大大减少灾害造成的损失。因此本文对汛期5~9月降水量的预测问题进行研究,期望能得到较好的预测模型。

## 1 资料与方法

### 1.1 资料

使用了1953~1999年北半球500hPa高度场576点网格资料和西北太平洋海水表面温度286点网格资料以及舟山市3个县市区站(分别是定海国家基准站、普陀、嵊泗气象站)5~9月降水资料和定海基准站2月和4月月平均气温资料,其时间年限定海为1954~1999年,普陀、嵊泗则是1961~1999年。将3个县市区站5~9月降水量作为预报对象,并以1998年和1999年降水资料作为独立样本,用于方程的试报检验。

### 1.2 预报方法

目前关于旱涝预测的方法有经验方法、统计学方法和动力学方法。动力模式的预测还处于探索阶段,且尚限于月及季的预测试验中,统计学方法仍是目前旱涝预测的主要方法<sup>[1]</sup>。本文拟采用基于PRESS准则的逐步回归分析方法来建立舟山市汛期5~9月降水量的预测模型,该方法的PRESS准则以预测为目的,比着眼于拟合为目的的传统的RSS准则的方法预测意义更大,且回归方程中因子的取舍不需通过 $F$ 检验来决定,而 $F$ 值的控制存在着主观性,因而这一方法也更为客观、合理<sup>[2]</sup>。为了便于比较,同时也采用了传统的基于残差平方和(RSS)准则的逐步回归分析的预测模型。基于RSS准则的逐步回归已为大家所熟知,这里不再详述,下面就基于PRESS准则的逐步回归分析作简要介绍:

设有 $P$ 个自变量(预报因子) $X = (x_1, x_2, \dots, x_p)'$ 和因变量(预报对象) $Y$ 和 $n$ 次观测值,考虑如下的线性回归模型

$$Y = X\beta + e$$

式中 $\beta, e$ 分别表示回归系数和残差的向量。若把去掉第 $j$ 个试验点(样品)后的模型记为

$$Y(j) = X(j)\beta + e(j) \quad (2)$$

式中 $Y(j), X(j)$ 和 $e(j)$ 分别是 $Y, X$ 和 $e$ 中删去第 $j$ 行后得到的。由最小二乘法估计,可

① 浙江省自然科学基金资助项目(498022)资助

得到  $\beta$ , 则第  $j$  个试验点  $X_j = (X_{j1}, X_{j2}, \dots, X_{jp})'$  处的预测值为  $X_j' \beta(j)$ , 而预测偏差为  $f_j = y_j - X_j' \beta(j)$ , 对每一个试验点重复上述做法, 可得到  $f_1, f_2, \dots, f_n$ , 其平方和

$$PRESS = \sum_{j=1}^n f_j^2 = \sum_{j=1}^n [y_j - X_j' \beta(j)]^2 \quad (3)$$

称为预测平方和。采用引进和剔除回归自变量的双重筛选逐步算法, 在回归自变量可能产生的一切子集中, 以  $PRESS$  值达到最小的子集, 作为最优回归子集, 这就是选取回归自变量的  $PRESS$  准则。其具体的计算过程详见文献[2]。

## 2 预报因子的选取

### 2.1 外部因子

影响旱涝的因子很多, 目前广泛应用的 500hPa 环流和海温因子是其中的二个重要因子, 根据目前气候可预报性研究指出的预报上限为 6~12 个月的特点, 我们确定预报因子时段为上年 10 月~当年 4 月。通过计算舟山市 3 个县、区、站 5~9 月降水量与逐月 500hPa 高度场、西太平洋海温的相关系数, 选取相关系数通过显著性水平  $\alpha = 0.01$  且至少 4 点连续成片的区域平均值作为预报因子, 其中定海 7 个, 普陀 8 个, 嵊泗 7 个。另外我们还计算了本站要素(日照、蒸发量、月平均温度)与 5~9 月降水量的相关系数, 其中定海的 2 月和 4 月月平均温度的平均值与其降水量相关系数通过了  $\alpha = 0.01$  的显著性水平, 因此也作为一个预报因子, 这样定海共有 8 个预报因子。

### 2.2 降水本身周期分量因子

引起气候变化的影响因素错综复杂, 而气候资料的时间序列反映了曾经发生过的所有因果关系和结构关系的影响, 而且大气运动的多频振动的现象在气候资料的时间序列中也能得到充分的反映。因此我们按照方差分析方法提取降水资料序列本身的隐含周期, 即对已知的时间序列的观测值  $x(1), x(2), \dots, x(n)$ , 采用方差分析方法确定其

隐含的周期  $l_j, j = 1, 2, \dots, k$  作为试验周期, 具体的对  $1 < l \leq [n/2]$ , 记

$$\bar{X}^{(l)}(j) = \frac{1}{n_j} [X(j) + X(j+l) + \dots + X(j+(n-1)l)], \quad 1 < j \leq l \quad (4)$$

式中  $n_j$  为满足  $j + (n_j - 1)l \leq n$  的最大整数。

根据资料样本长度的不同, 定海选取了 21 个周期因子 ( $2 \leq l \leq 22$ ), 普陀、嵊泗各选取了 17 个周期因子 ( $2 \leq l \leq 18$ )。其中通过显著性水平  $\alpha = 0.01$  的定海有 11 个、普陀 12 个、嵊泗 11 个周期因子。

## 3 预报方程的建立

这里我们对基于  $RSS$  准则的逐步回归分析(取  $F_1 = F_2 = 4.0, \alpha = 0.05$ )和基于  $PRESS$  准则的逐步回归分析分别对影响降水的外部因子、降水量序列的隐含周期因子和两者综合的因子进行计算, 共计 6 个试验方案, 由此建立了 3 个县、区、站各 6 个预报方程。经过分析比较, 我们发现: 当因子相同时, 应用  $PRESS$  准则的逐步回归方法所得的回归子集优于应用  $RSS$  准则的逐步回归方法 ( $\alpha = 0.05$ ) 所得的回归子集, 而当因子不同, 预报方法(分别是  $PRESS$  方法和  $RSS$  方法)相同时, 应用综合因子所得的回归子集优于应用周期因子所得的回归子集, 而应用周期因子所得的回归子集又优于应用外部因子所得的回归子集。

表 1 给出了  $RSS$  和  $PRESS$  准则的逐步回归方法对综合因子计算的效果。

表 1 基于  $RSS$  和  $PRESS$  的逐步回归对综合因子的计算效果

地名	入选因子数		复相关系数		剩余均方差		平均绝对误差	
	RSS	PRESS	RSS	PRESS	RSS	PRESS	RSS	PRESS
定海	6	9	0.9129	0.9338	94.6649	86.5485	68.3	63.4
普陀	7	11	0.9561	0.9767	61.2525	48.3428	43.2	28.7
嵊泗	6	7	0.9330	0.9383	66.4873	64.9898	46.1	45.3

我们在用上述两种方案对舟山市 3 个县、区、站 1998 年和 1999 年的试报以及 2000 年的预报中发现: 在某些年份, 应用  $RSS$  准则的逐步回归方法所得到的结果优于应用

PRESS 准则的逐步回归方法所得的结果。而 RSS 准则的回归方程的建立是随  $F$  临界值的改变而变化的,它与 PRESS 准则相比具有不客观性,这一试验中我们是选定  $\alpha = 0.05$  时的  $F$  临界值进行运算的,因此上述的分析结果对于本文所研究的问题在实际使用中仍有一定价值,于是我们选取了上述应用综合因子的 PRESS 准则的逐步回归方法和 RSS 准则的逐步回归方法所得的回归方程作为舟山市汛期 5~9 月降水量的预报方程,在实际的预报中,则是把两种方法所得的预报值取平均值。表 2 给出了舟山市定海区 1998 年和 1999 年试报以及 2000 年实际预报的结果。

表 2 舟山市定海区 5~9 月降水量试报及预报结果(历年平均值 763.3mm)R/mm

年份	实况	RSS 方法		PRESS 方法		平均值	
		预报	绝对误差	预报	绝对误差	预报	绝对误差
1998	738.2	858.9	120.7	972.9	234.7	915.9	177.7
1999	903.7	970.2	66.5	964.1	60.4	967.1	63.4
2000	825.1	754.1	71.0	752.7	72.4	753.4	71.7

#### 4 预报方程中因子的物理意义

在应用回归分析方法选取预报因子时,选择有物理意义的因子是做好预报的关键。从舟山市三个县区站各自的二个预报方程中的预报因子来看,与定海 5~9 月降水量有关的因子是当年 4 月份俄罗斯东部鄂霍茨克海附近的高空 500hPa 环流( $X_1$ ),这与文献[4]所指出的初夏鄂霍茨克海高压的维持往往使我国长江流域梅雨雨带稳定,出现大范围洪涝,以及王克非的研究所指出的中国的梅雨与鄂霍茨克海阻塞系统有很好的关系相一致<sup>[4]</sup>。这说明选入的该因子物理意义明显。同时还与上一年 11 月份的西太平洋副热带高空 500hPa 环流( $X_2$ )以及本站当年的 2 月份和 4 月份的月平均温度的平均值( $X_3$ )有关。从周期分量因子来看,所选入的因子表现为 12 年、13 年、15 年、20 年、21 年以及 22 年周期(分别对应  $X_{12}$ 、 $X_{13}$ 、 $X_{15}$ 、 $X_{20}$ 、 $X_{21}$  和  $X_{22}$ , 以下同)。这可能与太阳活动的 11 年、22 年周期有关。至于在回归方程中出现与太阳活动周期相近的周期,这正是反映了回归拟合是对若干基本周期进行调频调幅的过程,在基本

周期上叠加了一些微小的振动。

与普陀 5~9 月降水量有关的因子有海温因子、500hPa 高度场因子以及降水本身的周期分量因子。其中海温因子有当年 1 月份黑潮区海温( $X_1$ )、当年 3 月份北赤道暖流区海温( $X_2$ )以及当年 4 月份北太平洋暖流区附近的海温( $X_3$ )。这与文献[5]所指出的冬季黑潮区海温与我国东部 6~8 月降水相关明显相一致。从入选的海温因子的时间来看,与文献[5]指出的在洋流区的海气相互作用中,海洋起主导作用,其中暖洋流的主导作用在冬季 1~3 月最明显相一致。这说明选入的海温因子物理意义明显。从高度场因子来看,则有上一年 11 月份阿留申群岛北部白令海附近的 500hPa 环流( $X_4$ )以及当年 2 月和 4 月美国阿拉斯加东部的高纬 500hPa 环流( $X_5$  和  $X_6$ )。从周期分量因子来看,普陀 5~9 月降水量有 8 年、9 年、10 年、15 年、16 年、17 年和 18 年的周期。

与嵊泗 5~9 月降水量有关的外部因子则分别是当年 1 月份欧洲东部的高空 500hPa 环流( $X_1$ )和当年 4 月份北太平洋暖流区的海温因子( $X_2$ )。而周期分量则主要表现为 7 年、9 年、12 年、13 年和 17 年这 5 种周期。

从上述入选的因子分析,回归方程中未能选进与长江流域梅汛期雨量有明显关系的西太平洋副高区环流因子以及对全球气候有重大影响的反映厄尔尼诺现象的赤道东太平洋海温因子,而且回归方程中周期因子占了多数。这一结果一方面反映了舟山市汛期 5~9 月降水量有明显的周期变化,另一方面我们认为入选的周期分量因子本身包含了各种因素(包括海温、大气环流等)对其作用的结果。

#### 5 回归方程中因子数的讨论

用回归方法选取重要因子时,被选中的因子数  $l$  的多少是值得注意的,当样本数  $n$  较小时, $l$  不应选得太大,否则会造成回归方程中预报因子随意性的增大,导致方程预测能力的不稳定。我们分析了上述 6 个回归方程

中的因子数,发现应用 PRESS 方法选取的因子数多于应用 RSS 方法所选取的因子数,特别是定海、普陀两站的因子数较多,分别是 9 个和 11 个。一般认为因子数可控制在样本数  $n$  的  $1/10 \sim 1/3$  范围内,为此,我们通过控制 PRESS 值的大小来控制选入回归方程中因子的数目。在具体的实施时,我们先确定平均预测误差  $e$  为某一合适的数值,以此来求得与它相适应的预测误差平方和(即 PRESS)控制值  $p$ 。当 PRESS 值小于  $p$  时,因子不再引进;当 PRESS 值大于  $p$  时,因子不再剔除。当平均预测误差分别在 105mm 和 65mm 时得到的定海和普陀的回归方程,我们称之为改进方案。表 3 给出了 1998 年、1999 年试报以及 2000 年的实际预报结果,改进方案的预测能力优于原方案。

表 3 经改进后的 PRESS 方法对  
5~9 月降水量  $R/\text{mm}$  预测结果

年份	定海			普陀		
	实况	预报	绝对误差	实况	预报	绝对误差
1998	738.2	920.7	182.5	753.1	687.6	65.5
1999	903.7	1005.6	101.9	779.6	553.6	226.0
2000	825.1	762.0	63.1	653.1	626.6	26.5

## 6 试报和预报结果分析

2 年的试报表明:回归方程对定海、普陀的预测效果较好,PRESS 准则的逐步回归的预测结果与 RSS 准则的逐步回归相比,有些年份后者优于前者,但在有的年份,前者优于后者。考虑到 RSS 准则在选取预报因子中的主观因素,如本文中的运算是在选定  $\alpha = 0.05$  时进行的,而 PRESS 准则的逐步回归方法却比较客观、合理的,于是我们选取这两种方案预报值的平均值作为预报结果,以起到取长补短、提高预报效果的作用,经检验,它在实际应用中是可取的。从嵊泗的试报结果来看,1998 年比较好,而 1999 年则很不理想,从其雨量的实况分析,1998 年为早年,其雨量 445.0mm 比历年平均值 555.5mm 偏少了 110mm,而 1999 年则为大涝年,其雨量 904.4mm 比历年平均值多了将近 350mm。从嵊泗这 2 年的试报来看,似乎方程对早年

的预报能力较好,而对大涝年特别是从早年立即转向大涝年的预报能力则较差。对舟山市 3 个县区站 2000 年 5~9 月雨量的实际应用表明,其预测结果相当令人满意。

## 7 结果和讨论

经过上述分析、研究,我们可以得到如下几点结果:

(1)舟山市汛期(5~9 月)降水量具有明显周期变化,且与太阳活动周期有关。

(2)舟山市汛期(5~9 月)降水量的变化是降水量时间序列隐含的周期因子与大气环流、海温因子共同作用的结果,得到了具有物理意义明确的影响因子。

(3)经试验,采用 PRESS 准则的逐步回归分析比采用 RSS 准则的逐步回归分析更为客观、合理,而且前者得出的回归子集优于后者。

(4)针对基于 PRESS 准则的逐步回归分析得出的回归方程中因子数较多的问题,可以通过设置一个适当的平均预测误差值来控制 PRESS 值来解决。试验表明,采用此方法所得方程的预测效果优于原方案。

(5)在试报中发现,PRESS 准则的逐步回归和 RSS 准则的逐步回归的预报效果,对于不同的年份存在差异,所以在这一研究中,我们选取两种方案预报结果的平均值作为预报值,这在实际应用中是可取的和有价值的。

致谢:本文得到了浙江大学姚棣荣教授的指导,在此深表感谢。

## 参考文献

- 1 周家斌,黄嘉佑. 旱涝预测方法的现状. 旱涝气候研究进展,北京:气象出版社,1990:134~142.
- 2 姚棣荣,俞善贤. 基于 PRESS 准则选取预报因子的逐步算法. 大气科学,1992,16(2):129~135.
- 3 姚棣荣. 基于 PRESS 准则的逐步回归周期分析. 气候学研究——气候与中国气候问题,北京:气象出版社,1993:169~173.
- 4 叶笃正,黄荣辉等. 长江黄河流域旱涝规律和成因研究. 济南:山东科学技术出版社,1996:77.
- 5 王绍武,赵宗慈. 长期天气预报基础. 上海:上海科学技术出版社,1987:65~75.

# On Forecasting of Rainfall over Zhoushan in Rainy Season

Wang Lei Cao Meilan

(Zhoushan Meteorological Observatory, Zhejiang Province 316004)

## Abstract

With stepwise regression analysis on the basis of the criterion from Prediction Error Sum of Squares (PRESS) and on the basis of the criterion from Residual Sum of Squares (RSS), the forecasting model of precipitation over Zhoushan in Rainy season (May to September) were studied. Moreover, the scheme of forecast ensemble by forecast results of two predictive method is proposed. Results show that effects of the prediction of this scheme are superior.

**Key Words:** PRESS criterion RSS criterion stepwise regression rainfall