

# 多元模糊回归分析在长江流域 旱涝趋势预测中的应用\*

魏凤英 郝跟成\*\*

(中国气象科学研究院)

## 提 要

用长江流域25个站5—9月降水量的第一特征向量的时间系数作为表征长江流域的旱涝指数,分析了它与前期北太平洋海温和500hPa高度场的关系。经过对初选预报因子代表性和稳定性分析后,确定出7个预报因子。利用多元模糊回归分析探讨了这一区域旱涝趋势的预测问题。

## 一、引 言

长江流域旱涝趋势的预测历来是气象工作者关注和重视的研究课题之一。做好这一流域的旱涝趋势预测,可以为政府决策部门提供防洪抗旱信息,对国民经济的发展及合理安排工农业生产有着十分重要的意义。但是,目前旱涝形成的物理机制尚未完全搞清楚,旱涝趋势是非常困难的预测项目。

用多元统计分析处理具有模糊性的问题,即模糊多元统计分析,是近年来发展起来的一种新的数学方法。它体现了多元统计分析与模糊数学相结合的新趋势,具有较高的应用价值。多元模糊回归分析就是其中一种处理多元模糊变量的方法,它近似于普通的多元回归,计算简便,结果清晰直观,有较广的应用范围。这方面的应用在气象领域还很少见。因此,这里用模糊多元统计分析在旱涝趋势预测方面作一尝试。

用长江流域25个站5—9月降水量的第一特征向量的时间系数作为表征这一流域的旱

涝指数。分析了它与前期北太平洋海温和500hPa高度场的关系。对初选预报因子的代表性和稳定性进行了检验,最后挑选出7个预报因子,利用多元模糊回归分析探讨了这一区域的旱涝趋势预测。

## 二、旱涝指数及预报因子的确定

选用长江流域1951—1987年25个站(南京、上海、杭州、合肥、安庆、屯溪、武汉、九江、南昌、宜昌、常德、长沙、恩施、达县、酉阳、重庆、遵义、贵阳、成都、内江、宜宾、毕节、兴仁、雅安、榕江)5—9月降水量。由此构成了一原始数据矩阵。对这一矩阵进行经验正交函数(EOF)分解,分解出的主要因素的特征就是对原场特征的近似。第一特征向量占了总方差的95.5%,权重相当大,它的分布形式代表了这一区域多年平均的降水状况。这一特征向量的各分量均为正值,它所对应的时间系数就表示了这一区域的年际旱涝变化。从图1可以看到,1954、1956、1962、1969、1970、1973、1977、1980和1983年为涝年;1953、1961、1966、1972、1976、1978和1981年为旱年。这些旱

\*国家自然科学基金会资助项目

\*\*北京气象学院86级学生

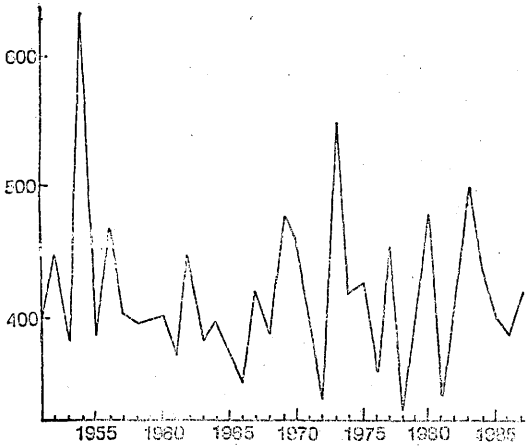


图1 第一特征向量的时间系数

涝年份与中央气象台确定的旱涝年是一致的(1)。因此,取它作为表征这一区域的旱涝指数是可行的。

选用北太平洋(5°S—65°N, 120°E—80°W) 286个网格点逐月平均海温和北半球500hPa 576个网格点逐月高度平均值,与旱涝指数进行了相关计算。结果表明,旱涝指数主要与前期夏、秋和冬季赤道太平洋地区(0°—10°S, 180°—90°W)的海温和冬春季中纬度地区的500hPa高度有较好的关系。尤其与前期赤道东太平洋海温关系最显著。考虑到预报因子的代表性,这里规定初选预报因子的原则是:选取相关系数超过0.05显著水平(0.33),且网格点个数≥8个点的区域为高相关区。从高相关区中选择相关系数绝对值最大者的序列,作为因子序列。按照这样的办法共选出13个序列作为初步选出的因子。

对选出的13个因子序列与其所在高相关区其它序列进行了相关计算。结果相关系数都较高,证明它们是有代表性的。为考查因子的稳定性,分别计算了旱涝指数与13个因子的滑动相关系数 $R_t$ ,滑动长度取10年。滑动相关系数随时间波动得较厉害,有的因子随时间从负相关转变成正相关,如图2A所示;有的前段有较高的相关,到近期下降较

大,甚至达不到显著水平(图2B所示);有些则在近期内相关相对稳定,并达到一定的显著水平,图2C就是一例。根据这种情况,

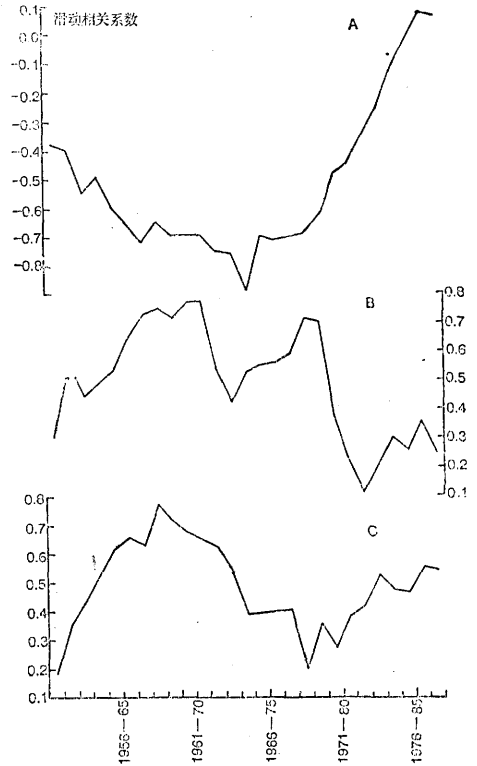


图2 旱涝指数与预报因子间滑动相关系数

分别计算了13个初步因子与旱涝指数滑动相关系数的平均值 $\bar{R}$ ,它表示 $R_t$ 的平均状况,并计算了近期3个滑动相关系数(1976—1985, 1977—1986和1978—1987)的平均值 $\bar{R}_{-3}$ ,以此表示某一因子近期与旱涝指数的相关程度。用既考虑 $\bar{R}$ 要大,也顾及 $\bar{R}_{-3}$ 要高的标准,挑选出7个相对稳定、近期滑动相关系数较高的因子作为建立预报方程的最终因子,即: $x_1$ :前一年7月(0°, 105°W)海温, $x_2$ :前一年9月(0°, 85°W)海温, $x_3$ :前一年11月(0°, 105°W)海温, $x_4$ :前一年12月(0°, 110°W)海温, $x_5$ :当年1月(0°, 110°W)海温, $x_6$ :当年2月(45°N, 20°W)500hPa高度, $x_7$ :当年3月(35°N, 110°W)500hPa高度。

$$\hat{S}_j = \sum_{i=1}^M a_i f(\bar{x}_{ij}) + a_0$$

$$(j=1, 2, \dots, K)$$

### 三、多元模糊回归

多元模糊回归分析<sup>[2]</sup>是处理多元模糊变量的一种方法。它是建立在模糊多变量关系基础之上的。

设样本向量由M个自变量和一个因变量构成：

$$X = (x_1, x_2, \dots, x_M, y)$$

假定自变量与因变量均有N个样本。若X的一个模糊子集 $\tilde{F}$ 可由隶属函数

$$\tilde{F}: (x_1 \times x_2 \times \dots \times x_M) \times y \rightarrow [0, 1]$$

来描述，则称 $\tilde{F}$ 是 $x_1 \times x_2 \times \dots \times x_M$ 与y的一个多元模糊回归关系。相应地， $x_i, i=1, 2, \dots, M$ 和y的模糊子集是 $\tilde{x}_i$ 和 $\tilde{y}$ 。

多元模糊回归分析的具体实施步骤为：

1. 把因变量即预报量y的样本按一定的标准划分为K个等级。根据预报量的等级将预报因子 $x_i, i=1, 2, \dots, M$ 的样本划归在相应的等级里。

2. 计算预报因子的各级平均值 $\bar{x}_{ij}, i=1, 2, \dots, M, j=1, 2, \dots, K$ 。

3. 定义隶属函数 $\tilde{y}_i(X)$ ：

$$\tilde{y}_1(X) = \begin{cases} 1 & \hat{S}_1 \leq \hat{S}_1 \\ \hat{S}_1 / \hat{S}_1 & \hat{S}_1 > \hat{S}_1 \end{cases}$$

$$\tilde{y}_i(X) = \begin{cases} \hat{S}_i / \hat{S}_i & \hat{S}_i \leq \hat{S}_i \\ \hat{S}_i / \hat{S}_i & \hat{S}_i > \hat{S}_i \end{cases}$$

$$\tilde{y}_K(X) = \begin{cases} \hat{S}_i / \hat{S}_K & \hat{S}_i < \hat{S}_K \\ 1 & \hat{S}_i \geq \hat{S}_K \end{cases}$$

$$\begin{matrix} (t=1, 2, \dots, N) \\ (j=2, \dots, K-1) \end{matrix} \quad (1)$$

其中

$$\hat{S}_t = \sum_{i=1}^M a_i f(x_{it}) + a_0$$

$$(t=1, 2, \dots, N) \quad (2)$$

$a_0$ 是一常数。 $a_i$ 是权重系数，使得回归样本的等级更适于表示自变量的变化。同时，它们又是模糊回归系数。以y作为因变量， $x_i$ 作为自变量，用最小二乘估计确定出 $a_0$ 及 $a_i$ 。并假定

$$f(x_{it}) \equiv x_{it}, \quad f(\bar{x}_{ij}) \equiv \bar{x}_{ij}$$

将系数 $a_0$ 及 $a_i$ 代入(2)式即可得到 $\hat{S}_j$ 及 $\hat{S}_i$ 。

4. 计算样本的隶属度。对于 $\tilde{y}_i(X)$ ，若

$$\tilde{y}_L(X) = \max_{1 \leq i \leq K} (\tilde{y}_i(X))$$

则预报量的样本归属在L级中。显而易见，隶属度是度量一个样本对每一等级接近的程度。

### 四、旱涝趋势的拟合和预测

对于旱涝的长期预测，人们关心的是未来会不会出现异常旱或涝的情况。报出基本趋势比给出具体数值有时更清晰明了。因此，可以把旱涝预测视作具有模糊性的问题。利用多元模糊回归分析探讨旱涝趋势的预测可能要比多元统计方法计算更简便，结果更明确。

按图1把长江流域旱涝指数划分为3级，即 $<3800$ 为1级， $3800-4400$ 为2级， $>4400$ 为3级。这3个等级分别表示旱、正常和涝。按这样的标准1951—1987年37年旱涝指数有7年为1级，有18年为2级，有12年为3级。

分别计算以上选出的7个预报因子在各个等级中的平均值 $\bar{x}_{ij}, i=1, 2, \dots, 7; j=1, 2, 3$ 。表1给出了这些平均值。利用最小二乘估计确定出 $a_0$ 及 $a_i$ （见表2）。

在具体计算时我们假定：

$$\hat{S}_i = \sum_{i=1}^7 a_i f(x_{it}) + a_0 = a_1 x_{it}$$

表1 预报因子在各个等级中的平均值

因子	等级		
	1 级	2 级	3 级
	<3800	3800—4400	>4400
x <sub>1</sub>	24.16	23.48	23.48
x <sub>2</sub>	22.29	22.46	21.84
x <sub>3</sub>	28.27	28.14	27.74
x <sub>4</sub>	25.60	25.06	24.60
x <sub>5</sub>	26.13	26.09	27.29
x <sub>6</sub>	50.70	51.80	60.50
x <sub>7</sub>	65.70	65.50	62.00

表2 系数值

a <sub>0</sub>	a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>
6303.865	0.907	3.651	-3.187
a <sub>4</sub>	a <sub>5</sub>	a <sub>6</sub>	a <sub>7</sub>
-15.875	11.691	20.881	-38.887

表3 S<sub>t</sub> 值 (1951—1987)

年 代	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
S <sub>t</sub>	4002.88	4745.36	4115.83	4742.68	3932.69	4341.25	3840.19	4571.90	4077.25	3662.19
年 代	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970
S <sub>t</sub>	3960.38	4752.70	4034.95	4237.83	4290.95	3878.83	3953.24	3773.40	4343.50	4836.86
年 代	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
S <sub>t</sub>	4287.07	3429.06	4844.35	4065.75	4376.10	3800.94	4379.80	3674.30	3884.58	4377.50
年 代	1981	1982	1983	1984	1985	1986	1987			
S <sub>t</sub>	4477.02	3620.51	4872.19	4409.04	4090.24	3614.96	4341.66			

按上面关系式，计算出历年各等级的隶属度，其最大者就是这一年所在的等级。例如1956年， $S_t = 4341.25$ ， $3930.21 < S_t$ ，所以  $\tilde{y}_1(X) = 3930.21/4341.25 = 0.905$ ； $4046.81 < S_t$ ， $\tilde{y}_2(X) = 4046.81/4341.25 = 0.932$ ； $4567.19 > S_t$ ， $\tilde{y}_3(X) = 4341.25/4567.19$

$$+ a_2 x_{2t} + a_3 x_{3t} + a_4 x_{4t} + a_5 x_{5t} + a_6 x_{6t} + a_7 x_{7t} + a_0 \quad (3)$$

$$\hat{S}_j = \sum_{i=1}^7 a_i f(\bar{x}_{ij}) + a_0 = a_1 \bar{x}_{1j} + a_2 \bar{x}_{2j} + a_3 \bar{x}_{3j} + a_4 \bar{x}_{4j} + a_5 \bar{x}_{5j} + a_6 \bar{x}_{6j} + a_7 \bar{x}_{7j} + a_0 \quad (4)$$

根据(3)式算得  $\hat{S}_t$ ， $t = 1, 2, \dots, 37$ ，列在表3。据(4)式可得：

$$\hat{S}_1 = 3930.21 \quad \hat{S}_2 = 4046.81$$

$$\hat{S}_3 = 4567.19$$

这样我们得到：

$$\tilde{y}_1(X) = \begin{cases} 1 & 3930.21 \geq \hat{S}_t \\ 3930.21/\hat{S}_t & 3930.21 < \hat{S}_t \end{cases}$$

$$\tilde{y}_2(X) = \begin{cases} \hat{S}_t/4046.81 & 4046.81 \geq \hat{S}_t \\ 4046.81/\hat{S}_t & 4046.81 < \hat{S}_t \end{cases}$$

$$\tilde{y}_3(X) = \begin{cases} \hat{S}_t/4567.19 & 4567.19 \geq \hat{S}_t \\ 1 & 4567.19 < \hat{S}_t \end{cases}$$

$$= 0.951。$$

$\tilde{y}_3(X) = \max(\tilde{y}_j(X)) = 0.951$ ，由此得出1956年为3级，即属涝年。依次类推，可以得到历年的等级值。根据1988和1989年7个预报因子资料，可以得到这两年的预报。历年用模糊回归计算的等级值及实况等级列在表4。从表中看出，37年中计算结果与实况完

表4

计算及实况等级

年 份	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964
计 算	2	3	2	3	1	3	1	3	2	1	1	3	2	2
实 况	3	3	2	3	2	3	2	2	2	2	1	3	2	2
年 份	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978
计 算	2	1	1	1	3	3	2	1	3	2	3	1	3	1
实 况	1	1	2	2	3	3	3	1	3	2	2	1	3	1
年 份	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989			
计 算	1	3	3	1	3	3	2	1	3	3	2			
实 况	2	3	1	2	3	2	2	2	2					

全一致的有21年,占57%。但是在不一致的16年中,15年均是错1级,只有1年报错2级,即这一年旱涝趋势完全颠倒。如果按旱涝趋势不反为正确统计,拟合率为97%。1988年预报3级,实况为正常偏涝,1989年预报为2级,实际属正常年景。

### 五、结 语

旱涝趋势的长期预测要建立在坚实的物理基础之上的。目前造成长江流域大范围异常旱涝的物理机制还未完全搞清,因此,不可能期望某种数学分析方法会有很理想的预

报效果。这里只是想用模糊数学与多元统计分析相结合的处理方法在气象应用方面作一尝试。当然,多元模糊回归分析的具体实施步骤,如隶属函数的确定,模糊回归系数的选择等并非最佳方案。我们认为,这一方法有许多待进一步改进之处。

### 参 考 文 献

- (1) 冯佩芝等,中国主要气象灾害分析,气象出版社, P7-8, 1988。
- (2) Chen Shi-quan, Analysis for Multiple Fuzzy Regression, Fuzzy Sets and system, 59-65, North-Holland, 25, 1988

## Application of multiple fuzzy regression analysis of the drought/flood trend prediction over the Changjiang River valley

Wei Fengying Hao Gencheng

(Academy of Meteorological Science)

### Abstract

The time coefficient series of the first characteristic vector of May-September precipitation at twenty-five stations is regarded as the drought/flood index over the Changjiang River valley. The correlation between the index and SST of the North Pacific and 500hPa height in the Northern Hemisphere is analysed and seven factors have been examined and the stability of the factors are selected. The drought/flood trend of the area is predicted by use of the multiple fuzzy regression analysis.