

时间序列的周期性预测模型

姚棣荣 王敬方*

(杭州大学地理系, 310028)

提 要

本文利用新安江流域(10站平均)汛期(4—9月)降水量资料,采用五种试验方案,研究时间序列隐含周期的提取以及对应的预测模型。结果表明:新安江流域的汛期降水量具有明显的周期性变化;采用对时间序列各试验周期的方差比直接进行显著性检验,提取主要周期,然后作周期的叠加外延预报,能取得较好的预报效果。

关键词 时间序列;周期性;方差贡献;回归;预测。

一、引 言

旱涝的变化规律与预测研究已受到国内外的普遍关注,它具有重大的现实意义。但这是一个十分困难的课题,动力模式的预测还处于探索阶段,且尚限在月及季的预测试验中,至于年度或更长时间的预测问题,目前大多采用统计学方法^[1]。

众所周知,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却是反映了曾经发生过的所有因果与结构关系的影响,而且大气运动的多频振动现象,在气候资料的时间序列中也能得到充分的反映,因而利用时间序列自身演变的规律进行气候预测已被广泛采用^[2,3]。而采用时间序列周期性变化规律作预测的方差分析方法更为常用^[4,5]。魏凤英等^[6]曾采用逐步回归周期分析方法进行预测,最近,黄嘉佑^[7]则针对方差分析的不足之处,提出了周期模式的概念。本文以新安江流域汛期(4—9月)降水量资料为例,采用五种方案对时间序列隐含周期的提取以及预测问题,进行分析和研究。

二、试 验 方 案

已知时间序列的观测值为:

$$x(1), x(2), \dots, x(n)$$

对此,欲求出隐含的周期 l_i , $i = 1, 2, \dots, k$, 方差分析周期外推法则是通过 F 检验来提取的^[8], 即对 $1 < l \leq [n/2]$, 记

$$\bar{x}^{(l)}(i) = \frac{1}{n_i} [x(i) + x(i+l) + \dots + x(i+(n_i-1)l)], \quad 1 \leq i \leq l, \quad (1)$$

1993年1月20日收到,4月27日收到修改稿。

* 现在中国科技大学研究生院

其中 n_i 为满足 $i + (n_i - 1)l \leq n$ 的最大整数。并记

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x(j), \quad S = \sum_{j=1}^n [x(j) - \bar{x}]^2, \quad S^{(l)} = \sum_{i=1}^l n_i [\bar{x}^{(l)}(i) - \bar{x}]^2. \quad (2)$$

取统计量

$$F^{(l)} = \frac{S^{(l)} / (l - 1)}{(S - S^{(l)}) / (n - l)} \quad (3)$$

若 $F^{(l)} \geq F_{\alpha}(l - 1, n - l)$, 则存在 l 周期

若 $F^{(l)} < F_{\alpha}(l - 1, n - l)$, 则不存在 l 周期 (4)

本文采用如下五种方案来提取周期。

方案 I 方差分析方法(1)

这是一种在实际工作中较常用的传统做法,按以下的递推方法经多步来实现。开始时置 $\hat{x}(t)$ 为零,然后对原序列进行分析,第 i 步的做法是:

(1) 按(3)式算出 $F^{(l)}$, $2 \leq l \leq [n/2]$ 。

(2) 取 l_i 使 $F^{(l_i)} = \max_l F^{(l)}$, 且在给定显著性水平 α 下, $F^{(l_i)} > F_{\alpha}(l_i - 1, n - l_i)$,

否则运算结束。在许多实际问题中,对各试验周期经 F 检验后, $F^{(l_i)} > F_{\alpha}(l_i - 1, n - l_i)$ 的,往往也是 $F^{(l_i)} = \max_l F^{(l)}$ 。

(3) 按(1)式,形成与原序列长度相同的由周期 l_i 所组成的序列。即令 $u_i(t) = \bar{x}^{(l_i)}(j)$, $t = j \bmod l_i$ 。

(4) 作周期叠加,并形成残差序列,即

$$\hat{x}(t) \simeq \hat{x}(t) + u_i(t), \quad x(t) \simeq x(t) - u_i(t)$$

上式表示:第 i 步结束时的周期叠加结果为

$$\hat{x}(t) = \sum_{j=1}^i u_j(t),$$

往后就是对新序列 $x(t) - \sum_{j=1}^i u_j(t)$ 进行周期分析。

方案 II 方差分析方法(2)

这一方案与方案 I 基本相同,只是在确定主要周期上有所不同。

(1)、(3)、(4)同方案 I

(2) 从小到大改变 α 值,按(3)式检验,取 l_i 使 $F^{(l_i)} > F_{\alpha}(l_i - 1, n - l_i)$, 若取不出 l_i , 则运算结束。

采用这一方案是出于这样的考虑:由(3)式可见,在一般情况下,当分子自由度 $(l - 1)$ 增大时, $F^{(l)}$ 减小;当分母自由度 $(n - l)$ 减小时, $F^{(l)}$ 减小。即当 l 增大时, $F^{(l)}$ 就减小,于是就会出现这样的情况,若 $F^{(l_i)} = \max_l F^{(l)}$, 但 $F^{(l_i)} < F_{\alpha}(l_i - 1, n - l_i)$, 而虽然 $F^{(l_i)} < F^{(l_j)}$, 但却有 $F^{(l_j)} > F_{\alpha}$ 。例如,在本文的分析中,发现新安江流域汛期降水量的 4 年周期 $F^{(4)} = 1.770$, 36 年周期 $F^{(36)} = 1.613$, 显然有 $F^{(4)} > F^{(36)}$, 但 $F^{(4)} < F_{0.0525}$, 不能通过 $\alpha = 0.0525$ 的显著性检验;而 $F^{(36)} > F_{0.0525}$, 却通过了 $\alpha = 0.0525$ 的显

着性检验。为了进行比较,本文设计了第二试验方案。

方案 III 周期模式逐步回归模型(1)

因为由方差分析所得到的主要周期,其波型是对应周期真实波型的平均状态,它无法反映不同类型的真实波型,于是出现了周期模式的概念^[7]。

(1) 由方案 I, 选出周期 $l_1, l_2, \dots, l_k$ 。

(2) 对每个周期 $l_i, i = 1, 2, \dots, k$, 将分析该周期时所用的对应的序列, 按周期 $l_i (i = 1, 2, \dots, k)$ 进行分组排列, 构成一个二维数组矩阵, 对该数组矩阵采用经验正交函数方法^[8]进行分解, 得到的特征向量即为周期 l_i 的主要模式。对应的特征值与所有特征值之和的比值可作为对应周期模式代表真实波型个例的百分率(详参文献[8])。

(3) 对于选出的周期 $l_i (i = 1, 2, \dots, k)$, 共得到 p 个主要周期模式, 并按对应的周期长度排列, 形成 p 个与原序列长度 n 相同的序列, 记作 $V_i(t), i = 1, 2, \dots, p$ 。

(4) 取 $\hat{x}(t) = \sum_{i=1}^p b_i V_i(t)$, 其中 $b_i (i = 1, 2, \dots, p)$, 使残差

$$Q = \sum_{t=1}^n \left[x(t) - \sum_{i=1}^p b_i V_i(t) \right]^2$$

达到极小。

方案 IV 周期模式逐步回归模型(2)

(1) 由方案 II, 选出周期 $l_1, l_2, \dots, l_k$ 。

(2)(3)(4)向方案 III。

方案 V 周期分量逐步回归模型

(1) 对 $x(1), x(2), \dots, x(n)$ 按(1)式算出 $\bar{x}^{(l)}(i), l = 2, 3, \dots, [n/2], 1 \leq i \leq l$ 。

(2) 令 $u_l(t) = \bar{x}^{(l)}(i), t = j \bmod l, l = 2, 3, \dots, [n/2]$ 。

(3) 取 $\hat{x}(t) = \sum_l b_l u_l(t)$, 其中 $b_l (l = 2, 3, \dots, [n/2])$, 使残差

$$Q = \sum_{t=1}^n \left[x(t) - \sum_l b_l u_l(t) \right]^2$$

达到极小。

三、试验结果

1. 试验资料

本文采用新安江流域(10站平均)1886—1988年汛期(4—9月)的逐年降水量资料。

为了检验预报效果,我们把1886—1981年的96年降水量资料作为依赖样本,而把1982—1988年的7年资料作为试报样本。

2. 计算结果

(1) 方案 I、方案 III 计算结果综合在表 1 中。

由表 1 可见,采用方案 I 我们可选得 7 个主要周期, F 检验的显著性水平 α , 除第一周期(4 年周期)外,其它周期的 α 值均小于 0.05, 对 7 个主要周期采用经验正交函数分解, 累积方差贡献达到 85% 以上的周期模式共有 25 种。其中以第一周期(4 年周期)为例,共得到 3 种周期模式,分别记为波型 1、波型 2 和波型 3, 其累积方差贡献为 90%, 即这 3 种波型能反映出序列实际波型 90% 的信息。而把由方差分析方法(1)得到的 4 年周期记为波型 0, 图 1 中给出了上述四种波型的曲线。

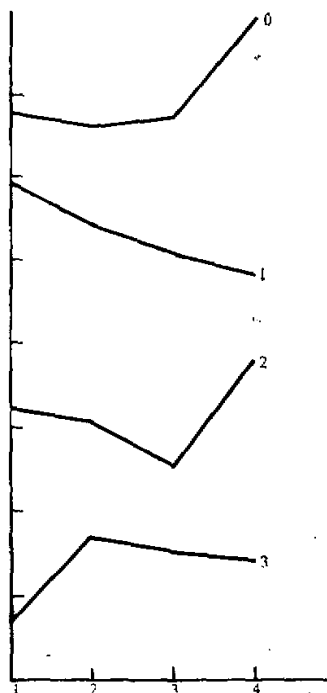


图 1 4 年周期波型曲线

$$+ 106.9251x_{11,2} + 100.0962x_{11,3} - 310.2488x_{23,1} - 126.6578x_{16,1}$$

式中因子 x 下标的数字,前者代表周期,后者表示该周期的波型代号(下同)。

(2) 方案 II、方案 IV 的计算结果综合在表 2 中。

由表 2 可见,方差分析方法(2)共选取 5 个主要周期,显著性水平 α 值均小于 0.06, 与方差分析方法(1)的结果有较大的差异,尤其是在第一周期的选取上,前者 $\alpha = 0.0525$, 而

由图 1 可见,波型 0 与波型 1、波型 2、波型 3 有较大的差异。而在方差分析中只能得出单一的波型 0。但考察新安江流域汛期降水量 4 年周期序列的各波型发现,几乎没有一个个例的波型是属于波型 0, 这是因为波型 0 是由真实波型平均得到的,它只能反映其平均状态,而无法反映不同类型的真实波型。这样用 3 个波型描写原序列 4 年周期的特征,似乎更为合理。

此外,由表 1 还可知,周期模式回归模型(1)所选取的 8 个波型中,以对真实波型贡献最大的波型 1 为主,其回归方程为

$$\hat{y}_s = 1136.6847 + 325.2567x_{23,1} + 853.3039x_{45,1} + 238.3890x_{45,2} + 176.2821x_{17,3}$$

表 1 方案 I、III 的综合结果

序号	l	$F^{(l)}$	α	周期模式个数*	回归模型中所选波型号
1	4	1.770	0.16	3	/
2	23	1.792	0.035	4	1
3	45	1.764	0.03	2	1,2
4	17	2.289	0.01	4	3
5	11	3.164	0.01	5	1,3
6	28	2.046	0.01	3	1
7	16	2.080	0.02	4	1

* 对各主要周期施行经验正交函数分解时,累积方差贡献达到 85% 以上的周期主要模式(波型)数。

后者却是 $\alpha = 0.16$ 。累积方差贡献达到 85% 以上的周期模式共 17 种。图 2 是 36 年周期的波型曲线。

从表 2 同样可见,在周期模式回归模型(2)中选取的 8 个波型仍然以波型 1 为主。

$$\begin{aligned} \hat{y}_t = & 1186.5964 + 836.4825x_{36,1} - 100.9686x_{7,1} \\ & + 157.0981x_{7,3} + 284.8953x_{22,1} \\ & - 231.5942x_{22,2} - 123.1748x_{23,1} \\ & - 113.9935x_{29,1} + 260.0251x_{29,1} \end{aligned}$$

(3) 方案 V 的计算结果

为了比较,与方案 III、方案 IV 一样,在程序中自动控制进入回归方程的周期分量因子为 8 个,我们得到如下的回归方程:

$$\begin{aligned} \hat{y}_t = & -1134.3722 + 0.2366x_{28} + 0.2475x_{35} \\ & + 0.2683x_{36} + 0.2130x_{38} + 0.2254x_{44} \\ & + 0.2665x_{45} + 0.2283x_{46} + 0.2894x_{47} \end{aligned}$$

由上述回归方程可见,入选的周期分量因子主要表现为 28 年左右、36 年左右和 45 年左右这三种周期,这一结果与方案 I、II 所选取的主要周期有相似之处。但它们之间各有差异。至于在回归方程中出现与上述周期相近的周期,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,这与周期模式有类似之处。

(4) 五种方案的拟合和预测效果比较

这里我们着重考虑预测的效果,而对于拟合情况,我们以原序列和拟合序列之间的相关系数和平均绝对误差(用于方案 I、II),以及回归方程的复相关系数和剩余均方差(用于方案 III、IV、V)来衡量。

关于预测效果,我们用上述五种方案,对新安江流域汛期降水量 1982—1988 年进行试报,并算出它们的预报准确率和剩余均方差。在统计预报准确率时,我们按文献[9]的标准来作旱涝级别划分,然后比较预测的级别与实况级别的差异,如果两者的级别相同,

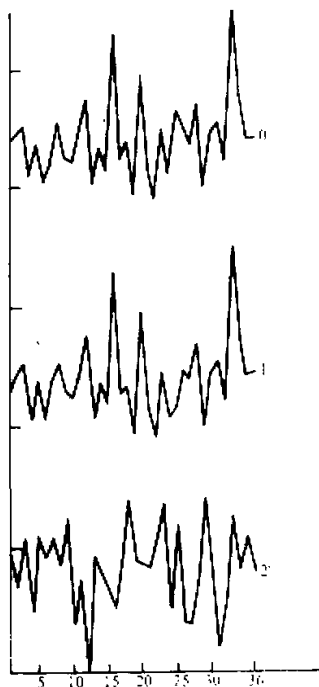


图 2 36 年周期波型曲线

表 2 方案 II、IV 的综合结果

序号	l	$F^{(1)}$	α	周期模式个数	回归模型中所选波型号
1	36	1.613	0.0525	2	1
2	7	3.554	0.005	4	1, 3
3	22	2.117	0.010	4	1, 2
4	20	1.760	0.045	4	1, 4
5	29	1.887	0.0178	3	1

或者相差 1 级但趋势是对的,就认为预报准确。

具体的级别划分为(单位: mm):

1 级: <878.53 ,

2 级: $878.53-1083.12$,

3 级: $1083.12-1243.87$,

4 级: $1243.87-1448.45$,

5 级: >1448.45 。

五种方案拟合和预测效果见表 3。

表 3 五种方案的拟合和试报效果

试验方案	1886—1981 拟合		1982—1988 试报	
	R	S_r	准确率	S_r
I	0.9759	44.0374	1/7=0.14	336.4078
II	0.9495	64.0283	5/7=0.71	307.7546
III	0.8892	117.0357	2/7=0.29	285.2258
IV	0.8499	134.8321	3/7=0.43	309.6411
V	0.9558	75.2344	3/7=0.43	288.7801

由表 3 可见,从拟合来看,以方案 I、II、V 较好,其中尤以方案 I 为最好;从预测效果来看,则以方案 II 为最好,方案 IV、V 次之,方案 I 最差;从数值来看,预测的精度还不够高,个别年份预测的误差还比较大。

按上述的评分标准,用方案 II 对 1989—1991 年三年作预报,预报准确率为 0.67。

四、趋势预报

根据上述分析,我们取方案 II 对新安江流域汛期(4—9 月)降水量作了 1992—2000 年的趋势预报(见表 4)。

表 4 新安江流域 1992—2000 年汛期降水量趋势预测¹⁾(单位: mm)

年份	1992	1993	1994	1995	1996	1997	1998	1999	2000
降水量	1274.18	802.98	1050.88	1109.24	987.00	1454.80	1359.66	1357.77	1023.98
级 别	4	1	2	3	2	5	4	4	2

1) 本文完成时,1992 年实况尚未得到,故未检验。

五、结论和讨论

经过上述分析,我们可以得到如下几点结果:

(1) 方案 I、II、IV 对原序列有较好的拟合效果,这说明新安江流域具有较明显的周

期性变化。

(2) 本文分析得到的主要周期,如 22 年、36 年、45 年等周期,可能与海气相互作用、太阳活动等有关,因而新安江流域汛期降水量的周期性变化是有物理背景的,需要继续研究。

(3) 传统的方差分析周期叠加外延预报,在主要周期的选取上以直接对各试验周期进行显著性检验的方案,其预测效果在五种方案中是最好的。

(4) 从预测的数值来看,本文所提出的预测方案II的预测精度还不够高,有待于改进。

(5) 对 1992—2000 年的趋势预测表明:新安江流域汛期降水量在 1992, 1997—1999 年为偏多,在 1997 年为特多;而在其他年份则以偏少为主。

参 考 文 献

- [1] 周家斌等, 1990, 旱涝预测方法的研究, 旱涝气候研究进展, 科学出版社, 134—142.
- [2] 周家斌, 1983, 一种新的时间序列预报方法, 科学通报, 28, 1250—1252.
- [3] 姚稼荣, 1989, 用于气候预测的灰色动态模型, 气候学研究, 气象出版社, 147—155.
- [4] 田承骏, 1982, 关于方差分析周期外推法, 应用数学学报, 5(1), 102—105.
- [5] 武汉中心气象台, 1974, 方差分析周期外推法在长期预报中的应用, 数学学报, 17(3), 156—163.
- [6] 魏凤英等, 1983, 逐步回归周期分析, 气象, No. 2, 2—4.
- [7] 黄嘉佑, 1990, 北京地区旱涝变化规律及其预测试验, 地理学报, 45(1), 102—109.
- [8] Kutzbach, J. E., 1967, Empirical eigenvectors of sea-level pressure surface temperature and precipitation complexes over North America, *J. Appl. Meteor.*, 16, 791—802.
- [9] 中央气象局等, 1980, 中国近五百年旱涝分布图集, 地图出版社, 332.

The Periodicity of Time Series and Forecasting Model

Yao Dirong Wang Jingfang

(Department of Geography, Hangzhou University, Hangzhou 310028)

Abstract

This article studies the extraction of implicit periodicity of time series and relevant forecasting model. The rainfall data of Xinanjiang River drainage basin (averaged over 10 stations) during the flood season (April to September), and five testing schemes are used. The results indicate that there is an obvious periodic change in flood season rainfall in Xinanjiang River drainage basin. Better forecasting result can be achieved by the following way: directly make a significance test and extract the dominant period according to each variance over the testing periodicity of time series; and then make the superimposing extension forecast.

Key words: Time series; Periodicity; Variance contribution; Regression; Foreshadow.