第4卷 第2期 应用气象学报 Vol.4,№2
1993年5月 QUARTERLY JOURNAL OF APPLIED METEOROLOGY May 1993

多步预测的降水时序模型

曹鸿兴 魏凤英

(中国气象科学研究院,北京 100081)

刘生长

(气象出版社,北京 100081)

提 要

该文设计了一个能作多步预报的时间序列模型,先生成时间序列及其差分的均生函数,再运用双评分准则对所有均生函数延拓序列作粗选和精选,以期建立一个拟合和预报效果均好的模型。就长江中下游6—8月降水总量的序列进行了计算,证实该模型可用在制作逐年气候预报或分月长期预报中。

关键词:气候预测,差分均生函数时序模型,多步降水预报。

1 引 言

在气象、水文、地震等部门,目前都需要在年末对未来一年可能出现的情形、趋势或灾情作出粗略预测,在春天则要对汛期作出预报。因此一个能作长时间的多步预测模型是为实际工作者所迫切需要的。

在文献[1][2]中提出了均生函数时间序列分析。近年来,已在一些气象、水文部门的长期业务预报中应用,证明均生函数时序分析在作多步预测上是有实效的。在文献[3][4]中,又提出了双评分准则(*CSC*),用以对统计模型的变量进行筛选,也收到较好的效果。考虑到作序列的差分是高通滤波中最常用且简便的滤波方法,滤波后的序列的均生函数有可能对建立更稳健的预测模型作出贡献,基于这种想法,在本文中我们设计了一个差分均生函数参加筛选的新建模方案,以期建立预报效果更佳的模型。首先计算原序列的一阶差分和二阶差分序列,连同原序列,派生出三组均生函数,然后在这众多的均生函数中用 *CSC* 进行粗选,通过 χ^2 检验的均生函数作为模型的备选变量。以 *CSC* 达到极大为确定方程变量个数为原则,从备选变量中再进行精选,最后建立模型。

2 计算方案

1991年9月27日收到,12月26日收到修改稿。

2.1 均生函数的计算

对有 N 个样本的观测序列进行标准差规格化，得

$$(\text{Ⅰ})\quad X(t)=\{x(1),x(2),\cdots,x(N)\}$$

作一阶差分运算

$$\Delta x(t)=x(t+1)-x(t)\quad t=1,2,\cdots,N-1$$

得到序列

$$(\text{Ⅱ})\quad X^{(1)}(t)=\{\Delta x(1),\Delta x(2),\cdots,\Delta x(N-1)\}$$

同理，作二阶差分运算

$$\Delta\Delta x(t)=\Delta^2x(t)=\Delta x(t+1)-\Delta x(t)\quad t=1,2,\cdots,N-2$$

得到序列

$$(\text{Ⅲ})\quad X^{(2)}(t)=\{\Delta^2x(1),\Delta^2x(2),\cdots\Delta^2x(N-2)\}$$

对序列（Ⅰ）、（Ⅱ）、（Ⅲ）计算均生函数，即对 $X(t),X^{(1)}(t),X^{(2)}(t)$ 计算

$$\bar{x}_l(i)=\frac{1}{n_l}\sum_{j=0}^{n_l-1}x(i+jl)\quad i=1,2,\cdots,l\quad 1\leqslant l\leqslant M \tag{1}$$

式中 n_l 为满足 $n_l\leqslant\left[\frac{N}{l}\right]$ 的最大整数，$M=\left[\frac{N}{2}\right]$ 为不超过 $\frac{N}{2}$ 的最大整数。当 N 为偶数时，$\left[\frac{N}{2}\right]=\frac{N}{2}$，当 N 为奇数时，$\left[\frac{N}{2}\right]=\frac{N-1}{2}$。若样本较大，$M$ 亦可取 $\left[\frac{N}{3}\right]$。把对 $X(t)$ 计算得到的均生函数称为原序列均生函数，对 $X^{(1)}(t)$ 计算得到的称为一阶差分均生函数，对 $X^{(2)}(t)$ 计算得到的称为二阶差分均生函数。对所有均生函数进行周期性延拓，得到三组均生函数序列，记为 $F(t),F^1(t),F^2(t)$。

$$F(t)=\{f_1(t),f_2(t),\cdots,f_L(t)\}$$

其中

$$f_1(t)\equiv 0$$

$$f_2(t)=(\bar{x}_2(1),\bar{x}_2(2),\cdots,\bar{x}_2(i_2))$$

$$\cdots\cdots$$

$$f_L(t)=(\bar{x}_L(1),\bar{x}_L(2),\cdots,\bar{x}_L(L),\bar{x}_L(1),\cdots\cdots,\bar{x}_L(i_L))$$

$F^1(t),F^2(t)$ 与 $F(t)$ 类同。式中 $\bar{x}_2(i_2)$ 表示取 $\bar{x}_2(1)$ 或 $\bar{x}_2(2),\bar{x}_L(i_L)$ 的意义类似。

2.2 粗选均生函数

由于有 $\left[\frac{N}{2}\right]+\left[\frac{N-1}{2}\right]+\left[\frac{N-2}{2}\right]$ 个均生函数，如果一起建模筛选计算量很大，尤其对于样本较大的序列，内存要求和计算时间都十分可观。因此，这里首先用双评分准则对所生成的均生函数进行粗选。双评分准则定义为[4]：

$$CSC=S_1+S_2 \tag{2}$$

其中

$$S_1=(N-k)\left(1-\frac{Q_k}{Q_y}\right) \tag{3}$$

$$S_2=2I=2\left[\sum_{i=1}^{G}\sum_{j=1}^{G}n_{ij}\ln n_{ij}+N\ln N-\left(\sum_{i=1}^{G}n_{i\cdot}\ln n_{i\cdot}+\sum_{j=1}^{G}n_{\cdot j}\ln n_{\cdot j}\right)\right] \tag{4}$$

式中 S_1 为数量评分，即为精评分，S_2 为趋势评分，称为粗评分，N 为样本长度，k 为统计模型中变量个数。由此可见，双评分准则旨在使模型拟合的精度要好，趋势亦准。这一准则对于判别长期预测模型更适用。

Q_k 为模型的残差平方和

$$Q_k = \frac{1}{N}\sum_{t=1}^{N}(x(t) - \hat{x}(t))^2 \tag{5}$$

Q_y 为模型的总离差平方和

$$Q_y = \frac{1}{N}\sum_{t=1}^{N}(x(t) - \bar{x})^2 \tag{6}$$

式中

$$\bar{x} = \frac{1}{N}\sum_{t=1}^{N}x(t) \tag{7}$$

在 S_1 中取 $(N-k)$ 而不是 N，是为了给 S_1 加进一个惩罚因子，即尽可能入选少的变量进入方程。S_2 中 G 为预测趋势类别数，趋势计算公式为

$$\Delta x(t) = x(t+1) - x(t) \qquad t = 1,2,\cdots,N$$

$$u = \frac{1}{N-1}\sum_{t=1}^{N-1}|\Delta x(t)|$$

对降水预报来说，人们最关心的是明年将出现旱、涝还是正常。又如市场预测来说，价格将上升还是下跌还是少变，所以这里将观测和预测趋势均分为三类，即

$$\text{I 类}: \Delta x(t) < -u$$

$$\text{II 类}: |\Delta x(t)| \leqslant u$$

$$\text{III 类}: \Delta x(t) > u$$

计算观测-预测列联表中的样品数 n_{ij} 以及 $n_{i\cdot} = \sum_{j=1}^{G} n_{ij}$，$n_{\cdot j} = \sum_{i=1}^{G} n_{ij}$，然后计算最小判别信息量 $2I$ 的值。

在模型为线性模型的情况下，在 k 固定为一个不大的量时，当 $N \to \infty$，$(N-k) \approx N \to \infty$，$S_1 \approx NR^2$ 的渐近分布为 $\chi^2_{\upsilon_1}$[5]，$\nu = k$，这里 R^2 为复相关系数。$2I$ 也为渐近 $\chi^2_{\upsilon_1}$ 分布[6]，$\upsilon_2 = (G-1)(G-1)$，根据 x^2 分布的可加性，

$$\chi^2_{\upsilon} = \chi^2_{\upsilon_1} + \chi^2_{\upsilon_2} \qquad \upsilon = k + (G-1)(G-1)$$

因此可对 CSC 进行 χ^2 检验。

对每一均生函数作为一个拟合序列计算与原序列 $X(t)$ 的 CSC 值。当 $CSC > \chi^2_{\upsilon,\alpha}$ 时入选。式中 α 视问题情况取 $\alpha = 0.05, 0.01$ 或 0.001.

2.3 精选均生函数

设已粗选得 q 个均生函数，从中要选出 p 个 $(p < q)$，再用最小二乘建立回归模型。均生函数进入的次序有三种方案：

（Ⅰ）按 CSC 的值由大到小排列，采用前向筛选逐个进入方程。（Ⅱ）先建立 q 个变量的回归模型，按回归系数值由大到小，前向筛选逐个进入方程。（Ⅲ）按最优子集回归原理，穷尽所有的可能组合。当模型的 CSC 值出现极大值时停止筛选。

当然第 Ⅱ 种方案最好，但计算量很大。其它两种方案各有千秋。在下节的计算实例中采用第 Ⅰ 种方案。

2.4 建模

假设方程引进 p 个均生函数后，CSC 出现极大值，即确定由 p 个均生函数建立预报模型。对构成的均生函数矩阵 $\underset{N\times q}{F}$，进行 Gram-Schmidt 正交化，得到 $\underset{N\times q}{\tilde{F}}$，它与 $x(t)$ 的线性模型的矩阵形式为：

$$X = \underset{N\times qq\times 1}{\tilde{F}}\ \underset{}{\tilde{\Phi}} \tag{8}$$

求最小二乘解

$$\tilde{\Phi} = (\tilde{F}^T\tilde{F})^{-1}\tilde{F}^T X$$

由于均生函数向量经正交化处理，故精选过程中系数不必重复计算。再求出原均生函数 F 的系数 Φ 即可建立最终预报方程：

$$\hat{x}(t) = \varphi_0 + \sum_{i=1}^{q}\varphi_i f_i(t) \tag{9}$$

对相应的均生函数作 S 步周期外延，利用式(9)就可得到 S 步预报。

3 长江中下游降水模型

用上述建模方案预测长江中下游地区汛期降水量。取南京、合肥、上海、杭州；安庆、屯溪、九江、汉口、钟祥、岳阳、宜昌、常德、宁波、衢县、贵溪、南昌、长沙等 17 站平均的 6—8 月降水总量作为预报对象，资料从 1951—1984 年，1985—1990 年作为试报。其中 $N = 34$，$M = 17$。

(1) 将观测数据进行标准差规格化，并用公式(1)计算原序列、一阶差分和二阶差分三个序列的均生函数，作周期延拓，由此生成了 48 个均生函数序列。

(2) 对均生函数进行粗选。将上述均生函数序列依次与降水量建立方程。分别建立以降水为预报量，以单个均生函数为预报因子的一元回归方程。由方程得到预报量的拟合值。这时就可以利用公式(3)和(4)分别计算 CSC 的精评分和粗评分。以 CSC 值最大者($f_{15}(t)$)为例说明计算过程。表 1 给出长度为 15 年的均生函数 $f_{15}(t)$ 的列联表。

表 1　$f_{15}(t)$ 的列联表

n_{ij}		预报 Ⅰ类	预报 Ⅱ类	预报 Ⅲ类	行总计 $n_{i\cdot}$
实况	Ⅰ类	5	1	1	7
	Ⅱ类	3	5	4	12
	Ⅲ类	0	1	13	14
列总计 $n_{\cdot j}$		8	7	18	$n = 33$

列联表中数字表示各类别出现的频数。例如，第一行表示实况为 Ⅰ 类而预报为 Ⅰ、Ⅱ 和 Ⅲ 类的频数。表中最后一行和最后一列的 $n_{i\cdot}$ 和 $n_{\cdot j}$ 反映了预报与实况频数的差异。

将列联表中的数字代入(4)即可算出粗评分 $S_2 = 2I = 21.99$。根据一元回归的残差平方和 Q_t 和总离差平方和 Q_y，代入(3)可算出精评分 $S_1 = 19.37$。$CSC = S_1$

$+S_2=41.36$。这时 $\nu_1=k=1,\nu_2=(3-1)(3-1)=4$，故，$\nu=5$.

对48个均生函数逐一计算 CSC，用水平 $\chi^2_{5,0.05}=11.07$ 进行检验，满足 $CSC>\chi^2_{5,0.05}$ 的有16个(见表2)。用筛选出的这16个均生函数作为备选因子。

(3) 对上述过程构成的均生函数矩阵 $F_{16\times34}$ 进行Gram-Schmidt正交化。用正交化的序列作为自变量，建立与 $x(t)$ 的线性模型。用最小二乘求解，将系数 $\varphi_i,i=1,2,\cdots,16$ 按绝对值由大到小排序。

表2 均生函数序列的粗评分 $2I$ 和精评分 S_1 及 CSC

序号	阶次	周期长度	$2I$	s_1	CSC
1	0	5	8.15	6.36	14.52
2	0	7	7.58	8.54	16.12
3	0	10	9.82	9.13	18.94
4	0	11	7.92	9.49	17.42
5	0	13	20.40	14.68	35.08
6	0	14	18.29	14.17	32.46
7	0	15	21.99	19.37	41.36
8	0	16	4.96	12.48	17.44
9	0	17	17.58	11.85	29.43
10	1	11	8.37	5.52	13.90
11	1	13	11.35	9.02	20.38
12	1	14	11.46	5.55	17.01
13	1	15	13.04	9.78	22.82
14	1	17	10.71	6.95	17.65
15	2	13	9.45	4.18	13.64
16	2	15	13.04	2.51	15.56

注：表中0表示原序列，1表示1阶差分序列，2表示2阶差分序列。

表3 精评分 s_1、粗评分 $2I$ 和 CSC 值

步数	1	2	3	4	5	6	7	8	9	10	……	16
S_1	6.36	13.94	15.49	20.98	23.17	23.77	23.32	22.87	22.31	21.64		
$2I$	8.15	10.52	10.00	21.52	37.43	42.27	36.25	46.71	43.32	43.32		
CSC	14.52	24.46	25.49	42.50	60.60	66.04	59.57	69.58	65.63	64.96		

表4 方程引入8个均生函数时的列联表

n_{ij}		预报 Ⅰ类	预报 Ⅱ类	预报 Ⅲ类	行总计 $n_{i\cdot}$
实况	Ⅰ类	7	0	0	7
	Ⅱ类	1	6	1	8
	Ⅲ类	0	1	17	18
列总计 $n_{\cdot j}$		8	7	18	$n=33$

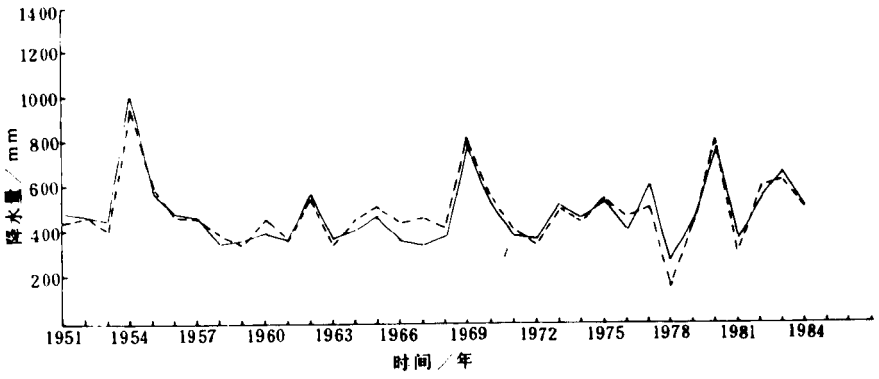

图1 长江中下游降水量变化曲线
(实线为实际值，虚线为拟合值)

(4)根据列联表和回归模型的复相关系数分别计算出方程引进1个，2个，…，16个均生函数的 CSC 值(见表3)。从表3看到，当引进8个均生函数时，趋势评分 $2I$ 由36.25骤升至46.71，CSC 值由59.57变成69.58。但当引进9个均生函数时 $2I$ 明显下降，因此模型的因子数应为8个。

用筛选均生函数的第Ⅰ种方案，按系数 φ_i 的序号，均生函数依次引入方程。表4给出方程引入8个均生函数时的列联表。其相应方程为：

$$\hat{x}(t) = -0.000763 + 0.0472f_5 + 0.1997f_7 + 0.4759f_{10} + 0.6222f_{13} + 0.5521f_{11} + 0.3006f_{14} + 0.3554f_{17} - 0.0295f_{15}^2 \quad (10)$$

图1给出了1951—1984年降水量的拟合曲线。由图可以看出拟合效果十分显著。值得注意的是，方程对历史罕见的长江流域1954年和1969年的洪水有非常好的拟合。对该区1978年的干旱和1980年的涝情反映得也十分清楚。均方根误差 $RMSE = 52.31$mm，远小于序列标准差 $S_x = 144.65$mm。

表5列出了1985—1990年的预报结果。可以看出，1985，1986，1987和1988这4年的预报值与实际值都相差甚微。1989年的绝对预报误差为106mm，但其距平符号正确。1990年误差为99mm，相对误差 $e = \dfrac{99}{S_x} = 68\%$，也不算大。

为了比较差分均生函数对时序模型的贡献，我们就只有原序列产生的均生函数参加筛选、用文献[2]中的方案进行了计算，从17个原序列产生的均生函数中选出8个，组成时序模型，这时模型的 $RMSE = 81.65$mm，远大于式(10)(称新方案)拟合的 $RMSE$。并对1985—1990年的长江中下游汛期降水作了试报，逐年预报值也列在表5中。由表可见，除1989年以外，其他年份新方案的结果均明显地优于老方案。

表5　1985—1990年预报值与实况值

年份	1985	1986	1987	1988	1989	1990
老方案预报值	385.4	372.3	440.7	151.3	641.9	779.9
新方案预报值	360.1	422.7	513.9	515.3	494.0	533.0
实况值	370.0	481.0	558.0	480.0	600.0	434.0

4　结　语

将时间序列一阶和二阶差分的均生函数连同原序列均生函数作为供筛选的预报因子，并用双评分准则经 χ^2 检验进行粗选，再进行前向筛选，最终建立具有多步预报功能的方程。这一方案旨在由入选适量差分序列的均生函数以改进对原始方程的拟合。长江中下游降水实例的拟合和预报结果表明，这一建模方案确能建立作多步预报且较稳健的时序模型。它提示我们：用兼顾数量和趋势效果的双评分准则对时间序列及其差分序列的均生函数进行筛选的方法，用在需作多步或较长时间预报而又不要求十分精细的预测项目上是适合的。这样的模型对制作国家经济规划中的气候预测以及年度分月预报是十分有用的。

参 考 文 献

1 曹鸿兴,魏凤英.基于均值生成函数的时间序列分析,数值计算与计算机应用,1991,12(2):82—89.

2 Cao HongXing, Wei FengYing, Wang Yongzhong. Time series model of long—range prediction and its application. *Acta Meteorologica Sinica*, 1990,4:120—127.

3 曹鸿兴,牛保山.统计模型选择的双评分准则及其在气象、水文预报中的应用.数理统计与应用概率,1989,4(1):5—10.

4 魏凤英,曹鸿兴.长期预测的数学模型及其应用.北京:气象出版社,1990,29—36.

5 Kendall S. M., A. Stuart. The Advanced Theory of Statistics. New York: Charles Greffin Company Ltd. 1979. 358—366.

6 Kullback S. Information Theory and Statistics. New York: Dorer Publications, Inc. 1968. 113—119.

MULTI—STEP PREDICTION MODEL OF TIME SERIES FOR PRECIPITATION

Cao Hongxing Wei Fengying

(*Chinese Academy of Meteorological Sciences, Beijing* 100081)

Liu Shengchang

(*China Meteorological Press, Beijing* 100081)

Abstract

A time series model for the multi—step prediction is proposed in the paper. Having calculated mean generating functions of the time series and its difference and then screening all extending series of the mean generating functions with the couple score criterion, the model which is well available for both the fitting and the forecast is built. As an example, total precipitation during the period from June to August over the lower and middle reaches of the Yangtse river is modeled. It is shown that the model proposed here is useful for year—to—year climate prediction or month—to—month long—range weather forecast.

Key words: Climatic prediction; Time series model of difference mean generating function; Multi—step precipitation forecasting.