研究论文

双线性模型在江淮旱涝序列预测中的应用①

金菊良 (四川大学，成都 610065)
杨晓华 (河海大学)
金保明 (福建省南平市水电局)
丁 晶 (四川大学)

提 要

提出了一套用双线性模型预测旱涝序列的简便方案。实例的计算结果说明了该方案的实用性和通用性，在各种自然灾害序列预测中具有重要的理论意义和广泛的应用价值。

关键词： 旱涝序列 预测 双线性模型 遗传算法

引 言

旱涝是影响中国农业生产的重大自然灾害，对其进行科学预测，可使当地农业、水利等有关部门及时采取防涝抗旱对策，减少灾害损失。旱涝序列受众多不确定性因素影响，在时序上常表现出弱相依性、突变性、稀遇性等复杂非线性特征，对其进行准确的数学描述至今仍需进一步探索[1,2]。作为目前时间序列分析中应用最广的自回归滑动平均模型(auto-regressive and moving average model，简称 ARMA 模型)在非线性情形下的推广，双线性模型(bilinear time series model，简称 BM 模型)是一类具有广泛应用价值的非线性时序模型[3,4]。为解决 BM 建模这一复杂问题，本文给出了一套基于遗传算法(genetic algorithm)的简便方案，并结合江淮旱涝序列预测实例说明了该方案的应用。

1 双线性模型的建模方案

BM 模型的一般形式为[3,4]

$$x(i)=\sum_{k=1}^{p}a(k)x(i-k)-\sum_{j=1}^{q}b(j)e(i-j)-\sum_{k=1}^{m}\sum_{j=1}^{n}c(k,j)x(i-k)e(i-j)+e(i)\quad (i=1,2,\cdots,N) \tag{1}$$

其中，$x(i)$ 为零均值时序，N 为时序样本容量；残差 $e(i)$ 为固定方差的白噪声序列，不同 $e(i)$ 之间相互独立；$a(k)$ 为自回归项的系数，$b(j)$ 为滑动平均项的系数，$c(k,j)$ 为双线性项的系数，p、q、m 和 n 分别为这三项的阶数。该模型简记为 $BM(p,q,m,n)$。显然，BM$(p,q,0,0)$ 就是 ARMA(p,q)。BM 模型的主要特点是，在 ARMA 模型的基础上进一步考虑了残差的过去值 $e(i-j)$ 与系统行为的过去值 $x(i-k)$ 之间的交互作用对系统当前行为 $x(i)$ 的影响，当 $x(i)$ 固定时 BM 模型就成为关于 $e(i)$ 的线性模型，而当 $e(i)$ 固定时 BM

① 国家自然科学基金(49871018)、中国博士后科学基金和四川大学高速水力学国家重点实验室开放基金(9904)资助项目。

模型则成为关于 $x(i)$ 的线性模型，故称为双线性模型。由于BM模型有效利用了预测过程中产生的残差信息进行反馈校正，在实用中常常表现出模型结构简明、参数少、适应性强等显著优点，因而得到了广泛应用[3,4]。

由式(1)知，残差同时出现在滑动平均项、双线性项和当前残差项中，用常规的残差平方和最小来估计BM模型参数显得很困难。目前常采用Subba Rao提出的反复残差法[3,4]，是在模型阶数已定的条件下反复应用最小二乘法来近似估计模型参数，计算量大，操作不便。BM建模的复杂性在一定程度上限制了它的广泛应用。

由于BM模型是ARMA模型的直接推广，因此建立BM模型的方法可以从构造它的自回归项、滑动平均项和双线性项入手，而它们的参数估计问题实质上就是残差最小化问题，可用遗传算法处理。基于这一思路，下面给出一套简便实用的BM建模方案，它包括如下3大步骤：

1.1 用自相关分析技术确定BM模型的自回归项

时序 $\{x(i)\}$ 延迟 k 步的自相关系数 $R(k)$ 为[5]

$$R(k)=\sum_{i=k+1}^{N}(x(i)-Ex)(x(i-k)-Ex)/\sum_{i=1}^{N}(x(i)-Ex)^2 \quad (2)$$

$$Ex=\sum_{i=1}^{N}x(i)/N \quad (3)$$

其中，N 为实测时序 $\{x(i)\}$ 的样本容量。根据抽样分布理论，当

$$|R(k)|<r_m=t_{\alpha/2}(t_{\alpha/2}^2+n-2)^{0.5} \quad (4)$$

时推断时序 $x(i)$ 与 $x(i-k)$ 之间的相依性不显著；否则推断时序 $x(i)$ 与 $x(i-k)$ 之间的相依性显著[5]。式(4)中，α 为显著水平，n 为计算 $R(k)$ 所用的样本容量，$t_{\alpha/2}$ 为自由度为 $n-2$ 的 t 分布双侧检验的临界值，r_m 为相关显著所需的最低相关系数绝对值。BM模型的自回归项应与这些相依性显著的 $x(i-k)$ 项相对应，其中相依性显著的最大延迟步数即为自回归项的阶数 p。

1.2 根据自回归项和建模经验确定BM模型的滑动平均项和双线性项

基于控制理论中误差反馈校正的思想，BM模型的滑动平均项和双线性项是利用过去预测所产生的残差来影响当前的预测值，因此这两项的具体结构形式与残差对系统行为的作用有关。根据应用BM模型的实际建模经验，实用中时序样本容量 N 一般不大。作者认为，这两项中残差的延迟阶数(即 q 和 n)倾向于取较小值(一般取1，有时取2，很少取3或以上)，以保持较少的模型参数，提高模型预测的稳健性，而双线性项中的 $x(i-k)$ 应与第一步(**1.1**)确定的BM模型的自回归项相一致，故可取 $m=p$。

1.3 用作者研制的加速遗传算法(accelerating genetic algorithm，简称AGA)[6]直接在BM模型的残差最小化下同时优化BM模型各参数

假定式(1)所表示的BM(p,q,m,n)模型具有可逆性，则可用下式递推出残差系列 $\{e(i)\}$[7]

$$e(i)=\begin{cases}0, & i\leqslant \max(p,q,m,n)\\ x(i)-\sum_{k=1}^{p}a(k)x(i-k)+\sum_{j=1}^{q}b(j)e(j-k)+\sum_{k=1}^{m}\sum_{j=1}^{n}c(k,j)x(i-k)e(i-j), & i>\max(p,q,m,n)\end{cases} \quad (5)$$

$$(i=1,2,\cdots,N)$$

其中，max() 为取最大值函数。BM(p,q,m,n)模型各参数的优化估计问题就可转换为求如下极小化问题：

$$\min Q(a(k), b(j), c(r,s); k=1\sim p, j=1, q, r=1\sim m, s=1\sim n)=\sum_{i=1}^{N} e^{6}(i) \tag{6}$$

式(6)中取残差的 6 次方是为了控制灾害序列中大的残差。作为一种通用的优化方法，AGA 显然可同时求解式(6)中的所有参数。AGA 的具体算法可参见文献[6]，这里不再多述。

2　双线性模型在江淮旱涝序列预测中的应用

例 1：根据文献[1]表 3 中基于 Z 指数的江淮地区 1951～1992 年夏季(6～8 月)旱涝序列$\{x(i)|i=1\sim 42\}$来建立 BM 模型。其中 $x(i)=1$、2、3、4、5、6 和 7 分别表示特涝、大涝、偏涝、正常、偏旱、大旱和特旱。计算该序列前 15 阶自相关系数值$\{R(k)|k=1\sim 15\}$，在置信水平取 70% 下，只有 $R(3)=-0.18$、$R(7)=0.18$ 和 $R(13)=-0.30$ 的绝对值不小于对应的相关显著所需的最低相关系数绝对值 r_m(即 0.17、0.18 和 0.20)，故这里只取延迟 3 步、7 步、13 步作为 BM 模型的自回归项，取 $p=13$。考虑到该资料序列较短，模型参数不宜太多，根据经验确定该双线性模型的结构为 BM(13,1,13,1)，它的具体形式为

$$\begin{aligned}x'(i)=&a(3)x(i-3)+a(7)x(i-7)+a(13)x(i-13)-b(1)e(i-1)-c(3,1)\\&x(i-3)e(i-1)-c(7,1)x(i-7)e(i-1)-c(13,1)x(i-13)e(i-1)\end{aligned} \tag{7}$$

其中 $x'(i)$ 为第 i 年旱涝的预测值；序列 $x(i)$ 值都已经去掉资料序列的均值 4.071(中心化)；$e(i-1)$ 为预测第 $i-1$ 年旱涝的残差，通过把式(7)代入式(5)求得。把$\{e(i)\}$代入式(6)，即得此例的目标函数，用 AGA 优化 BM 模型的参数，它们的初始变化区间和优化结果参见表 1。把这些参数优化值代入式(7)，即得预测该旱涝序列的 BM(13,1,13,1)模型。用该模型拟合 1964～1992 年旱涝序列和预测 1993～1995 年旱涝序列，结果见表 2。表 3 列出了该模型拟合误差分析的结果。

表 1　用 AGA 优化江淮夏季旱涝序列 BM(13,1,13,1)模型的参数

加速次数	优秀个体的变化区间							最佳目标函数值
	$a(3)$	$a(7)$	$a(13)$	$b(1)$	$c(3,1)$	$c(7,1)$	$c(13,1)$	
1	−0.500,0.000	0.000,0.500	−0.500,0.000	−1.000,1.000	−1.000,1.000	−1.000,1.000	−1.000,1.000	396.92
4	−0.281,−0.067	0.097,0.442	−0.482,−0.375	−0.040,0.447	−0.579,−0.093	−0.267,0.149	−0.303,0.144	253.81
8	−0.201,−0.142	0.246,0.332	−0.478,−0.452	0.261,0.405	−0.517,−0.337	−0.239,−0.134	0.026,0.097	216.84
优化值	−0.1853	0.2693	−0.4773	0.4047	−0.5016	−0.2130	0.0614	216.84

表 2　江淮夏季旱涝序列实测值、BM(13,1,13,1)模型的拟合结果和预测结果(单位：级)

年份	实测值	BM 计算值	残差	年份	实测值	BM 计算值	残差
1964	4.0	3.7	0.3	1980	2.0	2.1	−0.1
1965	2.0	4.0	−2.0	1981	5.0	3.4	1.6
1966	7.0	7.0	0.0	1982	4.0	4.0	0.0
1967	6.0	5.5	0.5	1983	4.0	4.7	−0.7
1968	4.0	4.5	−0.5	1984	4.0	3.9	0.1
1969	3.0	4.3	−1.3	1985	6.0	4.9	1.1
1970	4.0	3.0	1.0	1986	4.0	3.6	0.4
1971	4.0	3.7	0.3	1987	4.0	3.2	0.8
1972	4.0	2.3	1.7	1988	6.0	4.6	1.4
1973	4.0	5.2	−1.2	1989	4.0	2.9	1.1
1974	4.0	3.9	0.1	1990	4.0	3.6	0.4
1975	4.0	4.0	0.0	1991	1.0	2.4	−1.4
1976	5.0	4.8	0.2	1992	5.0	4.7	0.3
1977	4.0	4.0	0.0	1993	4.0	4.9	−0.9
1978	7.0	5.1	1.9	1994	7.0	6.1	0.9
1979	4.0	2.2	1.8	1995	4.0	4.9	−0.9

例 2：利用文献[8]表 7.8.1 中长江中下游五站 1952～1978 年 6～8 月平均降水量资料序列$\{x(i)|i=1\sim 27\}$来建立 BM 模型。根据类似于例 1 的自相关分析计算，可得该双线性模型的结构为 BM(15,1,15,1)，即：

$$\begin{aligned}x'(i)=&a(4)x(i-4)+a(9)x(i-9)+a(15)x(i-15)-b(1)e(i-1)-c(4,1)x(i-4)e(i-1)-c(9,1)x(i-9)e(i-1)-\\&c(15,1)x(i-15)e(i-1)\end{aligned} \tag{8}$$

其中，$x'(i)$ 为第 i 年 6～8 月平均降水量的预测值；$x(i)$ 值都已经去掉其均值 452.8mm；

$e(i-1)$ 为预测第 $i-1$ 年 6～8 月平均降水量的残差，通过把式(8) 代入式(5) 求得。为控制大的残差和便于计算，目标函数取相对残差四次方和最小形式，用 AGA 优化 BM 模型各参数所得的结果分别为 $a(4)=-0.1242$、$a(9)=-0.1795$、$a(15)=0.3666$、$b(1)=-0.6102$、$c(4,1)=-0.0171$、$c(9,1)=0.4352$ 和 $c(15,1)=-0.8118$，把它们代入式(8)即得预测该降水量序列的 BM(15,1,15,1)模型。用该模型拟合 1967～1978 年降水量序列和预测 1979～1980 年降水量序列，结果见表 4。

表 3　江淮夏季旱涝序列 BM(13,1,13,1)模型的拟合误差分析结果

残差绝对值(级)落在下列区间的百分比(%)				残差标准差(级)	平均残差绝对值(级)	平均相对误差(%)
[0.0,0.5]	[0.0,1.0]	[0.0,1.5]	[0.0,2.0]			
48.3	62.1	82.8	100.0	1.02	0.773	23.313

表 4　长江中下游 5 站 6～8 月平均降水量序列实测值、BM 模型的拟合和预测结果

年份	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
实测值/mm	247.1	257.6	885.3	497.2	368.4	390.9	333.1	510.8	609.7	407.7	470.5	182.3	454.2	766.0
BM 计算值/mm	442.4	211.1	858.9	569.4	446.2	403.6	365.4	416.9	569.6	475.2	370.6	245.4	404.8	519.3
相对残差/%	−79.0	18.1	3.0	−14.5	−21.1	−3.2	−9.7	18.4	6.6	−16.6	21.2	−34.6	10.9	32.2

表 2～表 4 说明：BM 模型虽然仅利用旱涝序列几阶自相关项信息，但是由于在预测过程中进一步利用了上次预测的残差信息进行反馈校正，模型的拟合精度和预测性能令人满意，该模型基本上可描述江淮夏季旱涝序列这一复杂的非线性动态系统，可作为预测地区旱涝的一种新的有效途径。

3　结语

由于灾害现象的复杂性及其预测的重要性，旱涝序列的预测至今仍是一片广阔的值得进一步探索的领域。作为自回归滑动平均模型在非线性情形下的推广，双线性模型(BM）是一类具有广泛应用前景的非线性时序模型。为处理 BM 建模这一复杂问题，文中提出了一套基于遗传算法的简便方案，并在江淮旱涝预测中得到成功应用。实例的计算结果表明，这套方案是可行而有效的，由于 BM 模型利用在实际预测过程中产生的残差信息进行反馈校正，保证了模型高的拟合精度和稳健的预测性能，对复杂非线性动态系统具有很强的适应性。文中提出的 BM 建模方案具有一般性，为 BM 模型在各种自然灾害序列预测中的应用提供了新的有力工具。

参考文献

1　张清．1994 年江淮伏旱及其影响研究．灾害学．1998，13 (2)：58～62.

2　胡小晖，延军平，欧维新．1950 年以来广西洪涝灾害及趋势预测．灾害学．1990，14 (4)：27～31.

3　杨位钦，顾岚．时间序列分析与动态数据建模．北京：北京工业学院出版社，1986：411～414.

4　杨叔子，吴雅等．时间序列分析的工程应用（下册）．武汉：华中理工大学出版社，1992：356～386.

5　金光炎．水文水资源随机分析．北京：中国科技出版社，1993：114～154，320～355，396～399.

6　金菊良，丁晶，魏一鸣．加速遗传算法在地下水位动态分析中的应用．水文地质工程地质，1999，26 (5)：4～7.

7　安鸿志，陈敏．非线性时间序列分析．上海：上海科学技术出版社，1998：186～189.

8　施能．气象统计预报中的多元分析方法．北京：气象出版社，1992：253～261.

Application of Bilinear Time Series Model to Drought and Flood Prediction Series in Jianghuai Area

Jin Juliang
(Sichuan University, Chengdu 610065)

Yang Xiaohua
(Hehai University)

Jin Baoming
(Water & Electricity Bureau of Nanping, Fujian Province)

Ding Jing
(Sichuan University)

Abstract

A simple scheme for establishing bilinear time series model (BM) is presented to predict drought and flood series. The examples show that the scheme applied to the drought and flood prediction series in Jianghuai area is practical and universal. The scheme has major theoretic value and wide-ranging application for predicting the time series of various natural disasters.

Key Words: drought and flood series prediction bilinear time series model genetic algorithm