水文时间序列的系统自记忆模式研究*

张双虎[1], 黄 强[1], 孙廷容[1,2]
(1 西安理工大学 水利水电学院, 陕西 西安 710048;
2 山西省水利厅, 山西 太原 030002)

[摘 要] 将"系统自记忆模式"引入到水文时间序列预测中, 推导了描述水文系统动力特性的微分方程, 建立了水文时间序列预测的系统自记忆预测模型, 并以黄河上游贵德水文站1919~1997年径流资料对该模型进行了验证。结果表明, 该模型具有预测精度高, 稳定性好, 观测误差对预测值影响小的优点。

[关键词] 系统自记忆; 时间序列; 年径流量预测
[中图分类号] P333.6 [文献标识码] A [文章编号] 1671-9387(2006)12-0221-05

由于水文循环结构的复杂性和影响水文循环因素的不确定性, 使得水文时间序列具有高度复杂的非线性和不确定性。近年来, 随着非线性理论的发展, 水文时间序列分析出现了混沌理论[1-2]、分形理论[3]以及人工神经网络方法[4]等非线性方法。用混沌理论分析水文时间序列, 必须对水文系统进行混沌性识别, 目前虽然有多种方法可用于混沌性识别, 但也只能得出"水文系统可能具有混沌特征"的结论[5]。神经网络方法分析水文时间序列存在着网络结构难以确定, 拟合精度高、预测精度低, 难以预测极值等具体应用问题。

动力系统自记忆性原理是解决非线性系统的一种统计—动力方法。"系统自记忆性"源于对大气运动的研究, 其核心是通过引入记忆函数, 将一个微分方程变换为差分—积分方程, 然后用这个方程来研究系统记忆性, 以对系统未来的演化做出预测[6-8]。对于有微分方程描述的动力系统, 可以直接应用自记忆方程建立预报模型; 对没有微分方程描述的系统, 只要系统具有一定长度的时间序列, 可将其视为描述该动力系统方程的一系列特解, 然后通过数据反演, 导出能近似描写该系统的非线性动力微分方程, 进而对系统未来的演化做出预测(报)。这种自记忆模式的优点是既可以把动力学计算与用历史数据估计参数结合起来, 又可以把统计学中从过去观测资料中提取预报信息的长处吸收进来, 从而提高预报的精度。

由于人们对复杂、非线性的水文系统认识不足, 目前还很难直接建立描述该系统特性的微分方程。本研究将一定长度的水文时间序列视为描述该系统的一系列特解, 建立了能描述水文系统特性的反导微分方程和水文系统自记忆预报模型, 并将其应用于黄河上游贵德站年径流预测, 取得了良好的效果。

1 水文系统反导微分方程

1.1 反导微分方程

设描述水文系统的状态变量(多年平均径流量)为X, 有一组等时间间隔离散观测数据$\{X=[x_t], (t=1, 2, \cdots, N)\}$。一般而言, 趋势、周期是水文时间序列的两种主要成分, 为此, 设变量X随时间变化的方程为:

$$\frac{\mathrm{d}x}{\mathrm{d}t}=f(x_t, \sin(t), \cos(t)) \tag{1}$$

对于具有周期性的水文时间序列, 首先进行分离周期项的预处理, 分离出周期项后的序列记为W $\{W=[w_t], (t=1, 2, \cdots, N)\}$, 则分离周期项后的变量随时间变化的方程为:

$$\frac{\mathrm{d}w}{\mathrm{d}t}=f(w_t) \tag{2}$$

设(2)式为一非线性多项式, 即设$f(w_t)=\sum_{l=1}^{L}\lambda Q_l$, 其中$l$为多项式的项数, Q_l, λ分别为多项式的项及其相应系数。作为一种新的水文时间序列分析建模方法, 这里只考虑线性项和2次幂项(如w_t^2项等),

* [收稿日期] 2006-06-22
[基金项目] 国家科学自然基金项目(50479024)
[作者简介] 张双虎(1976-), 男, 山西阳城人, 在读博士, 主要从事水资源系统工程研究。E-mail: sxslzsh@163.com
[通讯作者] 黄 强(1958-), 男, 四川绵阳人, 教授, 博士生导师, 主要从事水资源系统工程研究。

则(2)式可以写为:

$$\frac{dw}{dt} = (a_1 w_t + a_2 w_{t-1} + \cdots + a_p w_{t-p+1}) + (b_1 w_t^2 + b_2 w_{t-1}^2 + \cdots + b_p w_{t-p+1}^2) \quad (3)$$

式中, $a_1, a_2, \cdots, a_p, b_1, b_2, \cdots, b_p$ 为待定系数; p 为回溯阶, 其表示 $t+1$ 时刻变量 w 的取值 w_{t+1} 与前 p 个时刻 $(t, t-1, \cdots, t-p+1)w$ 的取值有关。

1.2 待定系数求解

设观测数据时间间隔 $\Delta t \equiv 1$, 改写(3)式为差分形式:

$$\Delta w = (a_1 w_t + a_2 w_{t-1} + \cdots + a_p w_{t-p+1}) + (b_1 w_t^2 + b_2 w_{t-1}^2 + \cdots + b_p w_{t-p+1}^2) \quad (4)$$

由数值差分知, 差分 Δw 有向前差分和向后差分2种形式, 分别表示为:

向前差分

$$\Delta_f w_t = w_{t+1} - w_t = a_1 w_t + \cdots + a_p w_{t-p+1} + b_1 w_t^2 + \cdots + b_p w_{t-p+1}^2 + \epsilon_{ft} \quad (5)$$

向后差分

$$\Delta_b w_t = w_t - w_{t-1} = a_1 w_{t-1} + \cdots + a_p w_{t-p} + b_1 w_{t-1}^2 + \cdots + b_p w_{t-p}^2 + \epsilon_{bt} \quad (6)$$

式中, ϵ_{ft}, ϵ_{bt} 分别表示向前差分误差和向后差分误差。

由(5)式、(6)式可得:

$$\epsilon_{ft} = (w_{t+1} - w_t) - (a_1 w_t + \cdots + a_p w_{t-p+1} + b_1 w_t^2 + \cdots + b_p w_{t-p+1}^2) \quad (7)$$

$$\epsilon_{bt} = (w_t - w_{t-1}) - (a_1 w_{t-1} + \cdots + a_p w_{t-p} + b_1 w_{t-1}^2 + \cdots + b_p w_{t-p}^2) \quad (8)$$

双向差分的目的是为了使差分误差达到最小, 即:

$$\epsilon = \sum_{i=1}^{n} (\epsilon_{bt}^2 + \epsilon_{ft}^2) \to \min \quad (9)$$

式中, $n = m - p - 1$, m 为用来求解参数的时间序列的长度。

将(7)式、(8)式代入(9)式, 用最小二乘法求解 $2p$ 个系数 $a_1, a_2, \cdots, a_p, b_1, b_2, \cdots, b_p$。系数求出后, 用相对方差来筛选系数。令

$$c_i = \begin{cases} a_i & i = 1, 2, \cdots, p \\ b_{i-p} & i = p+1, p+2, \cdots, 2p \end{cases} \quad (10)$$

取判别依据

$$\sigma_i = c_i^2 \Big/ \sum_{i=1}^{2p} c_i^2 \quad (11)$$

若 $\sigma_i < \xi$(ξ 为事先给定的一个较小的值, $\zeta = 0.01$), 认为第 i 项在(3)式中起的作用很小, 则在(3)式中剔除此项, 最后可以确定微分方程的项和各项的系数。

2 水文系统自记忆预报模型的建立

2.1 数学理论[9]

将上述确定的微分方程 dw/dt 作为水文系统动力方程, 并令 $dw/dt = F(w, t)$, 运用系统自记忆性原理[5], 建立新的预报模型。

引进记忆函数 $\beta(t)$, 将 $dw/dt = F(w, t)$ 两端同时乘以 $\beta(t)$, 并对其在 $[t-p+1, t+1]$ 上积分:

$$\int_{t-p+1}^{t+1} \beta(\tau) \frac{dw}{d\tau} d\tau = \int_{t-p+1}^{t+1} \beta(\tau) F(w, \tau) d\tau \quad (12)$$

(12)式左端按时间间隔段分段积分, 得:

$$\int_{t-p+1}^{t+1} \beta(\tau) \frac{dw}{d\tau} d\tau = \int_{t-p+1}^{t-p+2} \beta(\tau) \frac{dw}{d\tau} d\tau + \int_{t-p+2}^{t-p+3} \beta(\tau) \frac{dw}{dt} d\tau + \cdots\cdots + \int_{t}^{t+1} \beta(\tau) \frac{dw}{d\tau} d\tau \quad (13)$$

为简便计, 令 $\beta_t = \beta(t)$, $w_t = w(t)$。对(13)式右端第1项进行分部积分计算, 则有:

$$\int_{t-p+1}^{t-p+2} \beta(\tau) \frac{dw}{d\tau} d\tau = \beta_{t-p+2} w_{t-p+2} - \beta_{t-p+1} w_{t-p+1} - \int_{t-p+1}^{t-p+2} w(\tau) \beta'(\tau) d\tau \quad (14)$$

对(14)式中第3项运用积分中值定理得:

$$\int_{t-p+1}^{t-p+2} \beta(\tau) \frac{dw}{d\tau} d\tau = \beta_{t-p+2} w_{t-p+2} - \beta_{t-p+1} w_{t-p+1} - w_{t-p+1}^m (\beta_{t-p+2} - \beta_{t-p+1}) \quad (15)$$

式中, w_{t-p+1}^m 为中值, $\min(w_{t-p+1}, w_{t-p+2}) \le w_{t-p+1}^m \le \max(w_{t-p+1}, w_{t-p+2})$。

同理, 对(13)式右端 2~p 项分别进行分部积分、并对分部积分的第3项运用积分中值定理, 则(13)式可写为:

$$\int_{t-p+1}^{t+1} \beta(\tau) \frac{dw}{d\tau} d\tau = \beta_{t+1} w_{t+1} - \beta_{t-p+1} w_{t-p+1} - \sum_{i=t-p+1}^{t} w_i^m (\beta_{i+1} - \beta_i) \quad (16)$$

将(16)式代入(11)式得:

$$w_{t+1} = \frac{1}{\beta_{t+1}} \Big[(\beta_{t-p+1} w_{t-p+1}) + \sum_{i=t-p+1}^{t} w_i^m (\beta_{i+1} - \beta_i) + \int_{t-p+1}^{t+1} \beta(\tau) F(w, \tau) d\tau \Big] \quad (17)$$

(17)式表明, 如果已知 $w_{t-p+1}, w_{t-p+2}, \cdots, w_t$, 则

预报w_{t+1}依赖于两部分, (17)式右端前两项表示历史值对预报值的影响; 最后一项表示源项的影响。若令$w^m_{t-p}=w_{t-p+1}$, $\beta_{t-p}=0$则(17)式可进一步写为:

$$w_{t+1}=\frac{1}{\beta_{t+1}}\left[\sum_{i=t-p}^{t}w^m_i(\beta_{i+1}-\beta_i)+\int_{t-p+1}^{t+1}\beta(\tau)F(w,\tau)\mathrm{d}\tau\right] \tag{18}$$

但是(18)式中的w^m_i是未知的, 为此作近似:

$$y_i\equiv w^m_i\approx\begin{cases}\frac{1}{2}(w_i+w_{i+1}) & t-p\le i\le t-1\\ w_i & i=t\end{cases} \tag{19}$$

并且用左积分近似计算(17)式的第3项

$$\int_{t-p+1}^{t+1}\beta(\tau)F(w,\tau)\mathrm{d}\tau\approx\sum_{i=t-p+1}^{t}\beta_iF(w,i)\Delta t=\sum_{i=t-p+1}^{t}\beta_iF(w,i) \tag{20}$$

因此, 以上两次近似中误差为$0(\Delta t)$。记$k_i=\frac{\beta_{i+1}-\beta_i}{\beta_{t+1}}$, $\theta_i=\frac{\beta_i}{\beta_{t+1}}$, 从而(18)可近似为:

$$w_{t+1}=\sum_{i=t-p}^{t}k_iy_i+\sum_{i=t-p+1}^{t}\theta_iF(w,i) \tag{21}$$

(21)式即系统自记忆预测模型, 式中k_i, θ_i为记忆系数, y_i由(19)式确定; $F(w,i)$由(3)式确定。

2.2　记忆系数的估计

正如只有少量的微分方程有解析解, 自记忆方程也一样, 因此需借助计算机求解其数值解。若将(21)式右端y_i, $F(w,i)$视为系统输入, 左端w_{t+1}视为系统输出, 则可以用最小二乘法等方法求解由(21)式构成的方程组。

3　应用实例

将上述模型用于黄河上游贵德水文站年径流量预测。现有贵德站1919～1997年79年长系列年径流量资料(资料略), 其中用1919～1992年径流量资料建立模型、率定参数, 用1993～1997年径流量资料进行预测检验。

3.1　周期项

本文采用修正的周期图法确定时间序列的周期[10]。经计算分析得, 黄河上游贵德水文站年径流量序列的显著周期为18年和5.14年。为此, 该时间序列的周期项可表示为:

$$P_t=\sum_{i=1}^{2}\left(a_i\cos\frac{2\pi}{T_i}t+b_i\sin\frac{2\pi}{T_i}t\right) \tag{22}$$

用最小二乘法对(22)式进行拟合, 结果见表1。

表1　贵德水文站年径流量序列的周期及其系数

Table 1　Period and its coefficient of annual runoff in Guide station

周期编号 Period number	周期 Period	系数 Coefficient	系数 Coefficient
1	18	1.941 8	-2.195 4
2	5.14	2.848 0	1.203 5

3.2　回溯阶的确定

关于回溯阶的确定, 本研究采用"试错"的方法, 即分别取$p=5, 6, \cdots\cdots$进行建模、预测, 以最好的拟合结果来筛选回溯阶p的取值。

3.3　年径流量过程的模拟及预测

根据前述方法, 首先令$p=t$, $\xi=0.01$, 推导得分离周期项后的贵德水文站年径流量过程反导微分方程为:

$$\frac{\mathrm{d}w}{\mathrm{d}t}=-0.389\,9w_t+0.954\,1w_{t-1}+1.568\,0w_{t-2}-1.501\,1w_{t-4} \tag{23}$$

式中无w_{t-3}, w^2_t, w^2_{t-1}, w^2_{t-2}, w^2_{t-3}和w^2_{t-4}项, 这是因为进行筛选时, 上述6项已被剔除。进一步记(23)式右端为F。根据已经建立的贵德水文站年径流量自记忆预报模型, 用最小二乘法求解记忆系数k_i, θ_i, 得回溯阶$p=5$时分离周期项后的贵德水文站年径流量预报方程为:

$$w_{t+1}=\sum_{i=t-5}^{t}k_iy_i+\sum_{i=t-5+1}^{t}\theta_iF(w,i) \tag{24}$$

方程的系数如表2所示。

表2　回溯阶$P=5$时贵德水文站年径流量预报方程系数

Table 2　Prediction equation coefficient of annual runoff when backward moment is 5

系数 Coefficient	系数值 Coefficient value	系数 Coefficient	系数值 Coefficient value
k_{t-5}	0.490 0	θ_{t-4}	-0.014 4
k_{t-4}	-0.076 5	θ_{t-3}	-0.186 6
k_{t-3}	0.372 8	θ_{t-2}	0.027 5
k_{t-2}	0.031 3	θ_{t-1}	-0.084 8
k_{t-1}	-0.293 1	θ_t	0.071 1
k_t	0.596 7		

加上周期项后, 回溯阶$p=5$时贵德水文站年径流量预测模型为:

$$X_{t+1}=\sum_{i=t-5}^{t}k_iy_i+\sum_{i=t-t+1}^{t}\theta_iF(w,i)+$$

$$\sum_{i=1}^{2}\left(a_i\cos\frac{2\pi}{T_i}t+b_i\sin\frac{2\pi}{T_i}t\right) \quad (25)$$

记1919年 $t=1$, 1920年 $t=2$, 依次类推。用(25)式对1928～1992年贵德水文站年径流量进行拟合(由于建模原因, $t=1$ 到 $t=2p-1$ 时刻没有拟合值), 拟合均方根误差为4 952亿m³。同时, 用这一方程对1993～1997年贵德站年径流量进行预测, 结果发现预测值与实际观测值偏差较大。于是, 增大回溯阶 p 再进行拟合、预测。结果表明, 拟合均分根误差、预测值的相对误差随着回溯阶的增加而逐渐减小, 当回溯阶 $p>12$ 时, 拟合均方根误差基本稳定(图1), 为此确定回溯阶为12。

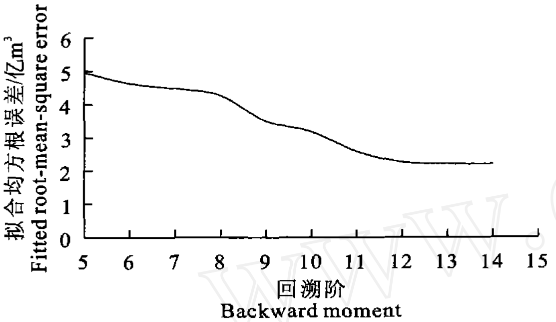

图1 不同回溯阶拟合的均方根误差

Fig. 1 Fitted root-mean-square of different backward moments

当回溯阶 $p=12$ 时, 分离周期项后的贵德水文站年径流量过程反导微分方程为:

$$\frac{dw}{dt}=-0.980\ 5w_t+1.306\ 1w_{t-1}+2.276\ 7w_{t-2}-0.995\ 6w_{t-4}-0.510\ 2w_{t-5}-0.748\ 6w_{t-6}-1.802\ 2w_{t-7}+1.047\ 7w_{t-8}+1.546\ 1w_{t-9}+0.002\ 1w_t^2 \quad (26)$$

将上式作为自记忆函数, 运用已建立自记忆预测模型对1942～1992年径流量过程进行拟合, 并对1993～1997年贵德水文站年径流量过程进行预测。拟合结果见图2, 预测结果见表3。

表3 贵德水文站1993～1997年径流量的预测结果

Table 3 Results of predicted annual runoff in Guide Station from 1993 to 1997

年份 Year	实测值/亿m³ Measured value	模型计算值/亿m³ Calculated value	相对误差/% Relative error
1993	230.0	260.61	13.30
1994	150.4	163.72	8.80
1995	171.2	183.00	6.89
1996	152.3	133.67	-12.23
1997	135.0	150.70	11.63

从图2, 表3可以看出, 本文建立的动力系统自记忆预报模型能很好的用于径流拟合及预测, 尤其极值年份都能很好的拟合出来, 如1975, 1981年的极大值和1959, 1977年的极小值。根据相对误差<20%作为预报合格的判定标准, 拟合、预报精度均达到《水文情报预报规范》中甲等预报的要求。

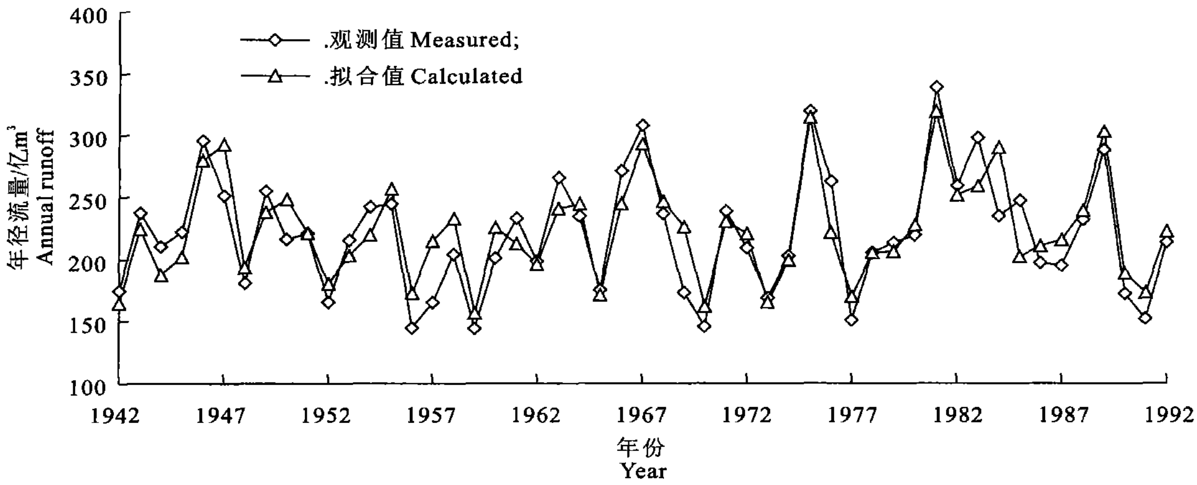

图2 贵德水文站1942～1992年径流量实测值与拟合值的比较

Fig. 2 Comparison between measured and calculated annual runoff in Guide station from 1942 to 1992

4 结 论

复杂动力系统的模拟、预测, 是当今世界系统科学界研究的方向之一。本文将"动力系统自记忆"这一概念引入到水文时间序列的预测中, 建立了水文时间序列预测的动力系统自记忆模型, 并通过实例对该模型进行了验证。研究表明, (1) 动力系统自记忆预报模型既能克服有些模型(如AR(1), ARMA(1,1))只用一个初值进行预报带来的"对初值"极其敏感的局限性, 又能克服传统方法用历史资料进行

统计而与机理性方程无关的局限性; (2) 该模型用历史资料估计记忆系数, 蕴含了历史资料中对预测有用的信息, 能很好地预测出系统极值, 克服了以统计为基础的预报模型在做多步预测时, 预测值偏向平均值的缺点。

[参考文献]

[1] 赵永龙, 丁　晶. 混沌分析方法在水文预测中的应用和展望[J]. 水科学进展, 1998, 9(2): 181-186.

[2] Jayawardena A W. Analysis and prediction of chaos in rainfall and stream flow time series[J]. Journal of Hydrology, 1994, 753: 23-52.

[3] 张少文, 王文圣, 丁　晶. 分形理论在水文水资源中的应用[J]. 水科学进展, 2005: 16(1): 141-147.

[4] 张双虎, 孙廷容, 黄　强, 等. 基于相空间遗传BP 神经网络的径流预测研究[J]. 西北农林科技大学学报: 自然科学版, 2005, 33(4): 122-126.

[5] 张利平, 王德智, 夏　军. 相空间神经网络模型及其在水文预测中的应用[J]. 水电能源科学, 2004, 22(1): 5-7.

[6] 曹鸿兴. 动力系统自忆性原理——预报和计算应用[M]. 北京: 地质出版社, 2002.

[7] 曹鸿兴. 大气运动的自己异性方程[J]. 中国科学(B 辑), 1993, 23(1): 102-112.

[8] Feng Guolin, Cao Hongxing, Gao Xinquan, et al. Prediction of precipitation during Summer Monsoon with self-memorial model[M]. Advances in Atmospheric Sciences, 2001, 18(5): 701-709.

[9] 陆君安, 夏　军, 陈士华, 等. 动力系统的自记忆数值预报[J]. 数学杂志, 1998, 18(增刊): 11-14.

[10] 丁　晶, 刘权授. 随机水文学[M]. 北京: 中国水利水电出版社, 1997.

The study on system self-memory model of hydrologic time series

ZHANG Shuang-hu[1], HUANG Qiang[1], SUN Ting-rong[1,2]

(1 *Institute of Water Resources and hydroelectric, Xi'an University of Technology, Xi'an, Shaanxi* 710048, *China*;
2 *Water Resources Bureau of Shanxi Province, Taiyuan, Shanxi* 030002, *China*)

Abstract: In this paper, the system self-memory model is introduced in the prediction of hydrologic time series, the differential equation describing dynamic characteristic of hydrologic system is deduced, and the prediction model of system self-memory of hydrologic time series is established. Also, with Guide hydrological station on the upstream of Yellow River as the study subject, the annual runoff materials from 1919 to 1997 are used to test the model. Research result indicates the model has the merits of high accuracy, good stability and low effect of observing error on prediction data.

Key words: system self-memory; time series; annual runoff prediction