

自激励门限自回归模型在枯水径流预报中的应用

冯国章 王双银[√] 王学斌

(西北农业大学水利与建筑工程学院, 陕西杨陵 712100)

摘 要 简介了门限自回归(TAR)模型的概念和自激励门限自回归(SETAR)模型的一般形式与建模方法,建立了泾河、北洛河和渭河三河流主要灌溉取水断面月平均流量的 SETAR 模型,将模型用于预报枯水期河川径流,精度与合格率均满足水文预报规范要求。

关键词 时间序列分析, 枯水径流预报, 自激励门限自回归模型

中图分类号 P338.3, S273.29

河流枯水径流的预报,是枯水季节水资源科学调配的重要前提。传统的枯水径流的预报方法在枯水径流预报中发挥了重要作用,但也有不足之处。70年代以来,随着时间序列分析理论与实用模型的发展,在包括水文学在内的许多领域内应用时间序列模型进行动态数据的模拟与预报已取得了满意的成果^[1,2]。特别是 Box 和 Jenkins^[3]提出的自回归求和滑动平均(ARIMA)模型,为处理线性时间序列提供了有力的数学工具。但是,实际工作中遇到的数据往往具有复杂的非线性变化规律,不能用线性模型来描述。因此,非线性时间序列模型的研究和应用越来越受到各界学者的关注^[4]。Tong^[5]提出的门限自回归模型(TAR),就是一种非线性时间序列模型,为解决非线性时间序列的建模与预报开拓了新的领域。本文主要研究门限自回归模型簇中自激励门限自回归模型(SETAR)在河流枯水径流预报中的应用。

1 门限自回归模型

门限自回归(Threshold autoregression)模型,是 Tong^[5]于 1978 年提出的,建模的基本思路是用逐段线性化手段来处理非线性系统;此外,门限的控制作用保证了系统的稳定性。所以它可以描述复杂的周期序列。

门限自回归模型根据其特点又可分为自激励门限自回归模型(SETAR)、开环门限自回归模型(TARSO)和闭环门限自回归模型(TARSC)^[4,5],其中 SETAR 模型为单输入模型, TARSO 和 TARSC 模型为多输入模型。

2 自激励门限自回归模型

2.1 基本模型

自激励门限自回归(Self-Excited Threshold Autoregression)模型的一般形式为

$$X_t = a_0^{(i)} + \sum_{j=1}^{p_i} a_j^{(i)} X_{t-j} + \varepsilon_t^{(i)} \quad T_{i-1} < X_{t-1} \leq T_i \quad (1)$$

式中 $\{X_t\}$ 为数据序列; p_j 为第 j 个模型的阶数 ($j=1, 2, \dots, n$); $a_i^{(j)}$ 为自激励门限自回归系数 ($i=0, 1, 2, \dots, p_j$); $\{\epsilon_t^{(j)}\}$ 是方差为 $\sigma_{\epsilon_j}^2$ 的白噪声序列, 当 $j \neq j'$ 时, $\{\epsilon_t^{(j)}\}$ 与 $\{\epsilon_t^{(j')}\}$ 相互独立; T_j 称门限值 ($j=1, 2, \dots, n-1$), 有 $-\infty = T_0 < T_1 < \dots < T_n = \infty$; d 称为延迟参数或延迟步长; n 是模型个数。

模型(1)表明, 门限把实数轴 $(-\infty, \infty)$ 依门限值划分为段, 当 X_{t-d} 落入第 j 段时, X_t 就采用第 j 个阶数为 p_j 的自回归模型描述, 则可将模型(1)简化为 SETAR($d, n, p_1, p_2, \dots, p_n$)。

2.2 建模方法

门限自回归模型虽然是一种非线性模型, 但又是基于对线性系统依状态取值逐段线性化处理。因此门限自回归模型可借助于线性自回归模型的参数估计与定阶准则, 不存在实质上的困难, 这是门限自回归模型优于其他非线性模型的一大特点。

对于自激励门限自回归模型, 若要用其拟合动态数据 $\{X_t, 1 \leq t \leq N\}$, 则可从动态数据 $\{X_t, 1 \leq t \leq N\}$ 出发, 通过辨识 SETAR($d, n, p_1, p_2, \dots, p_n$) 模型的参数而达到建模的目的。具体建模步骤如下:

①模型个数 n 和延迟步长 d 的确定 模型个数 n 和延迟步长 d 可通过物理成因分析或经验确定, 在无法做物理成因分析或无经验作根据时, 可采用统计识别方法, 如 AIC 准则^[6], 本文即采用 AIC 准则。

②门限值 T_j 的确定 模型个数 n 确定后, 门限值 T_j 的选择是重要的一步, 因为门限自回归模型的优劣与各段线性化的程度有关。一般通过试算确定。

③模型阶数 p_j 的确定 模型阶数 p_j 用 AIC 准则确定^[6]

按经验给定模型最大阶数 P , 并对 n 个模型各用最小二乘估计算出自回归系数的估计值和残差的方差, 便可用下式计算各模型的 AIC 值:

$$AIC(p_j) = N_j \ln(\sigma_{\epsilon_j}^2) + 2(p_j + 1) \quad (2)$$

式中 N_j 和 $\sigma_{\epsilon_j}^2$ 分别表示第 j 个模型的序列长和残差的方差。

在 n, d 和 T_j 固定情况下, 则 $p_j (1 \leq p_j \leq P)$ 所对应的 $AIC(p_j)$ 中的 $\min AIC(p_j)$ 的 p_j 及其相应的参数, 即为第 j 个模型的阶数和模型参数。

2.3 预报方法

由实测动态数据建立数学模型的主要目的是进行预报, 当对 $\{X_t\}$ 已建立 SETAR($d, n, p_1, p_2, \dots, p_n$) 的具体模型如式(1)后, 则 m 步的最小方差预报为条件期望

$$\hat{X}_{N+m|N} = E(X_{N+m} | X_1, X_2, \dots, X_N) \quad (3)$$

当 $m \leq d$ 时, X_{N+m-d} 的样本值已获得数据, 它属于那一个区间是已知的, X_{N+m} 就用相应的模型来预报。这样用实测数据和逐步递推预报可得到 m 步预报值。这就是常说的一步预报。

当 $m > d$ 时, X_{N+m-d} 的样本值未观测到, 无法判别属于哪一个区间, 但在预报 $\hat{X}_{N+1}$ 时, 因 X_{N-1-d} 已知, 可直接预报。此后, 再用预报的 $\hat{X}_{N+1|N}$ 作为观测值 (实际上未观测到) 来预报 $\hat{X}_{N+2|N}$ 以此类推可得到 m 步预报值。这就是常说的多步预报。一般来说多步预报中由于误差的累积, 预报结果不及一步预报结果好。

3 枯水径流预报

3.1 模型的建立与识别

本文以研究区泾、洛、渭三河流的干流上陕西省关中地区三大自流引水灌区泾惠渠(泾河)、洛惠渠(北洛河)和宝鸡峡引渭灌区(渭河)的引水口所在断面张家山水文站(泾河)、淤头水文站(北洛河)和林家村水文站(渭河)1961~1980年逐月平均流量作为建模数据,1981~1986年逐月平均流量作为预报检验数据。

建模过程中,采用了4种方式:①用全年逐月平均流量数据建模;②用全年逐月平均流量的模比系数建模;③用10~4月平均流量建模;④用10~4月平均流量的模比系数建模,各类模型的门限数均取1~7即建立2~8个模型,模型阶数取1~12,延迟步长根据成因分析确定为1。

同时,就上述门限数及模型阶数用4种方式建模,并经大量计算,发现在4种建模方式中均以SETAR(1,1,1,1)模型的预报精度和合格率最高,其中门限值用试算法求得,如在SETAR(1,1,1,1)模型中的②、④建模方式中,取模比系数 $K_{t-1}=0.2\sim1.8$,步长0.05和0.01试预报,结果以前者为0.50,后者为0.45左右时预报效果最佳,为方便计,分别取0.50和0.45。

3.2 预报与检验

用上述4种方式分别对各水文站建模并识别出最佳模型后,进行了全年(①、②建模方式)和枯水期(③、④建模方式)的一步和多步预报。结果显示,以预报值与实测值间的相对误差 $\epsilon \leq 20\%$ 为合格,全年预报的丰水期预报误差很大,多步预报的合格率低于一步预报的合格率,模型只适于枯水期预报。枯水期按6个月(11~4月)和按5个月(11~3月)计时,后者合格率略高于前者。

建模结果显示,直接用平均流量建立模型,在各种条件下其精度和合格率均小于用模比系数建立的模型的精度与合格率。

限于篇幅,本文仅给出用模比系数建立的SETAR(1,1,1,1)模型(表1,表2)。

表1 全年逐月模比系数建立的SETAR(1,1,1,1)模型

站 名	模型序号	模 型	适用条件
张家山	1	$K_t = 0.1532 + 0.8020K_{t-1} + \epsilon_t$	$K_{t-1} \geq 0.50$
	2	$K_t = -0.3363 + 1.7830K_{t-1} + \epsilon_t$	$K_{t-1} < 0.50$
淤 头	1	$K_t = 0.3539 + 0.6487K_{t-1} + \epsilon_t$	$K_{t-1} \geq 0.50$
	2	$K_t = 0.4483 - 0.0806K_{t-1} + \epsilon_t$	$K_{t-1} \geq 0.50$
林家村	1	$K_t = 0.1454 + 0.8188K_{t-1} + \epsilon_t$	$K_{t-1} \geq 0.50$
	2	$K_t = -0.1512 + 1.3363K_{t-1} + \epsilon_t$	$K_{t-1} \geq 0.50$

注:模型中, K_t 和 K_{t-1} 分别表示时段 t 和时段 $t-1$ 的流量的模比系数($K_t = Q_t/Q_1$, $K_{t-1} = Q_{t-1}/Q_1$, t 以月为单位), ϵ_t 是随机项。

表1和表2所列两类模型的预报精度(用平均相对误差 ϵ_m 表示)和合格率分别统计于表3和表4,从表3、4可以看出,用全年逐月模比系数和用10~4月模比系数所建模型的预报值与实测值间的相对误差的各站平均值分别为12.6%(预报6个月)和11.7%(预报5个月),说明用10~4月数据所建模型总体精度略高于全年数据所建模型的精度。两

类模型的总体合格率分别为 80%、81% 和 83%、87%, 具有与相对误差同样的规律, 但合格率的提高并不一定减小相对误差。

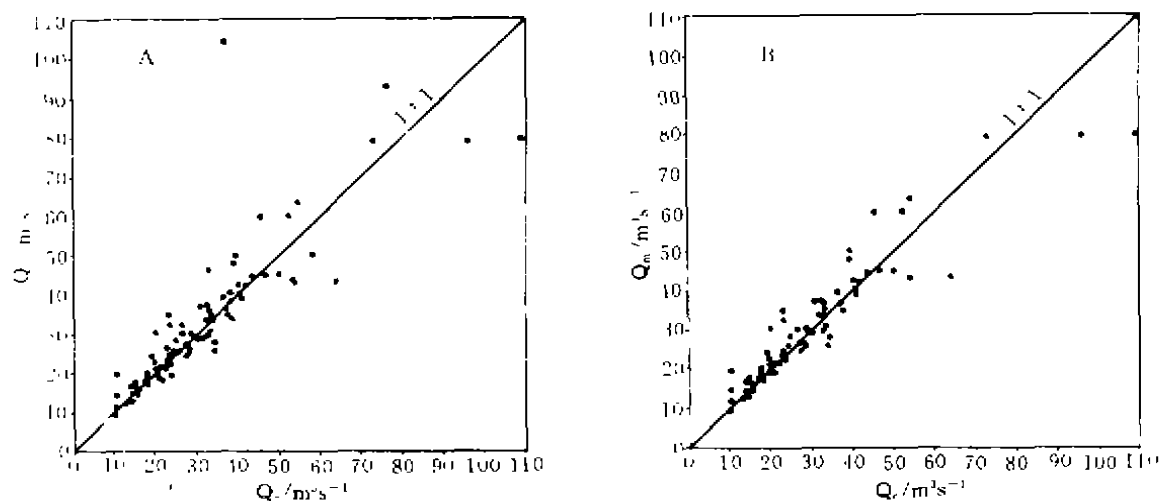
表 2 用 10~4 月模比系数建立的 SETAR(1,1,1,1) 模型

站名	模型序号	模型	适用条件
张家山	1	$K_t = 0.1265 + 0.8462K_{t-1} + e_t$	$K_{t-1} \geq 0.45$
	2	$K_t = -1.0597 + 3.7267K_{t-1} + e_t$	$K_{t-1} < 0.45$
状头	1	$K_t = 0.3735 + 0.6214K_{t-1} + e_t$	$K_{t-1} \geq 0.45$
	2	$K_t = 4.7357 - 11.1819K_{t-1} + e_t$	$K_{t-1} < 0.45$
林家村	1	$K_t = 0.1065 + 0.8821K_{t-1} + e_t$	$K_{t-1} \geq 0.45$
	2	$K_t = -0.3035 + 1.8122K_{t-1} + e_t$	$K_{t-1} < 0.45$

表 3 全年模型枯水期一步预报精度与合格率统计

站名	枯水期 (月)	各枯水期合格率(%)					总合格率 (%)	e_m (%)
		81~82	82~83	83~84	84~85	85~86		
张家山	11~4	83	83	67	83	83	80	11.1
	11~3	80	80	80	80	80	80	11.5
状头	11~4	83	83	100	100	83	90	10.7
	11~3	80	100	100	100	80	92	8.3
林家村	11~4	83	83	50	67	67	70	16.1
	11~3	80	80	60	80	60	72	14.0
平均	11~4	83	83	72	83	78	80	12.6
	11~3	80	87	80	87	73	81	11.7

从附图中可看出, 绝大部分点据分布在 1:1 斜直线附近, 表示预报值十分接近实测值。用 10~4 月模比系数建立的模型, 其结果比用全年数据据所建模型稍好一些, 限于篇幅, 不再赘述。

附图 全年逐月模比系数所建模型预报的枯水期流量(Q_p)与实测流量(Q_m)相关图

A. 枯水期为 11~4 ($r=0.855$) B. 枯水期为 11~3 月 ($r=0.934$)

表 4 10~4 月模型枯水期一步预报精度与合格率统计

站名	枯水期 (月)	各枯水期合格率(%)					总合格率 (%)	ϵ_m (%)
		81~82	82~83	83~84	84~85	85~86		
张家山	10~4	100	83	83	83	83	87	9.8
	10~3	100	80	100	80	80	92	11.0
状头	10~4	100	83	100	83	83	90	9.4
	10~3	100	100	100	100	80	92	8.1
林家村	10~4	83	83	50	67	83	73	16.3
	10~3	80	80	60	80	80	76	14.3
平均	10~4	94	83	78	78	83	83	11.8
	10~3	93	87	87	87	80	87	10.6

4 结论与讨论

本文通过建立泾、洛、渭三河干流上主要灌溉取水断面的枯水期月径流预报的自激励门限自回归模型,可得以下结论和认识。

1)用月平均流量的模比系数建立的模型,其精度与合格率均高于直接由月平均流量所建模型的精度与合格率。

2)模型阶数与门限数的确定,以成因分析与统计分析相结合为好。模型阶数高、门限数多不一定能取得好的预报结果。相反,本文显示出低阶模型和较少门限可能得到较满意的结果。

3)门限值的确定十分关键,大量分析表明,其他条件不变时,所建模型精度与预报合格率的高低在很大程度上取决于门限值的选取。

4)自激励门限自回归模型是最简单的门限自回归模型。对于河川径流的预报,一阶自激励门限自回归模型则是以前一时段的流量作为当前时段的流量的预报因子,也是水文预报中较简单的情况,但它符合水文预报以最简单的模型和最少的信息获得可靠预报成果的基本原则。

5)考虑水文时间序列的随机性是当今水文学的理论研究和应用研究的热点之一,自激励门限自回归模型与传统回归分析方法的最大区别在于它把水文时间序列作为非线性随机动态过程来研究,是其本质所在。

参 考 文 献

- 1 杜金观,项静怡,戴俊华. 时间序列分析与预报. 合肥:安徽教育出版社,1991
- 2 Bras R L, Rodriguez-Iturbe I. Random functions and hydrology. USA: Addison-Wesley Publishing Company, 1985
- 3 Box G E P, Jenkins G M. Time series analysis, forecasting and control. San Francisco: Holden-Day, 1970
- 4 项静怡,黄文杰. 门限自回归模型在天气预报中的应用. 第二次全国概率统计天气预报会议论文集,北京:科学出版社,1986
- 5 Tong H, Lim K S. Threshold autoregression, limit cycles and cyclical data. The Journal of the Royal Statistical Society Series B(Methodological), 1980 42(3): 245~292
- 6 Akaike H. Information theory and an extension of the maximum likelihood principle. in: Petrov B N and Csaki F eds. 2nd International Symposium on Information Theory. Budapest Akademiai Kiado, 1973, 267~281

Application of Self-Excited Threshold Autoregressive Model to Low Flow Forecast

Feng Guozhang Wang Shuangyin Wang Xuebin

(The College of Hydraulic and Architectural Engineering, Northwestern
Agricultural University, Yangling, Shaanxi, 712100)

Abstract The concept of threshold autoregressive (TAR) model and the general form and modeling method of self-excited autoregressive (SETAR) model are briefly introduced. The SETAR models of monthly flows for three major sites of irrigation water supply on the three rivers, Jing, North Lou and Wei, are formulated respectively. The models are used to forecast stream flows in low flow period of the three sites, and show relatively high accuracy and qualified ratio which satisfy the nation's criterion of hydrologic forecast.

Key words time series analysis, low flow forecast, self-excited threshold autoregressive model

1996年《中国生物防治》(季刊)

征 订 简 介

中国生物防治(原生物防治通报)是由农业部主管,中国农科院生物防治研究所主办的国家级学术性期刊。

主要内容:以虫治虫,以菌治虫,以菌治菌,以菌或虫治草,天敌资源的保护利用,农用抗生素,微生物农药,昆虫信息素,生物技术工程应用等无公害新技术,用来防治农、林、牧、贮粮、卫生方面的害虫、病菌、杂草、鼠害,以减少化防造成的污染和残毒,维持生态平衡。

主要栏目:研究报告、专题综述、基础知识与实验技术、研究简报、国外生防等。

读者对象:农、林、牧、贮粮、卫生各级管理干部、科技人员、院校师生、基层技术推广骨干等。

本刊季中月8日出版,16开本,每期48页。国内外公开发售,国内邮发代号2—507,国外代号Q812。每册定价2.50元,全年4期10元。全国各地邮局均有订售。

本编辑部备有1985~1994年以来各年《生物防治通报》精装合订本,每卷定价:20元,每期零售本(可以成套出售,也可以选期补售),每册定价2.00元。欢迎单位和个人来函订购(平寄免费,挂号另加15%邮费)。

编辑部地址:北京西郊白石桥路30号 邮政编码:100081 电话:8314433—2442