

遗传门限自回归模型在感潮河段水位预测中的应用

吴玲莉, 张 玮

(河海大学交通学院、海洋学院, 江苏 南京 210098)

摘要:应用门限自回归(TAR)模型建立了同时受潮汐和径流双重影响的长江下游感潮河段高桥水文站月水位 TAR 预测模型,建模过程中运用遗传算法来实现模型参数的优化.计算结果显示,门限自回归模型可以拟合感潮河段的非线性特性,拟合及预测精度均满足水文预报规范要求,遗传算法的引入简化了建模过程,提高了模型的预测精度并保证了其预测性能的稳定性.研究结果表明用遗传门限自回归模型预测感潮河段的水位是可行的,该模型在感潮河段其他水文要素的非线性时序预测中也具有广泛的实用价值.

关键词:感潮河段;河流水位;门限自回归模型;遗传算法

中图分类号:P338

文献标识码:A

文章编号:1006-7647(2005)05-0020-04

Application of genetic threshold auto-regressive model to water stage forecasting for tidal river section//WU Ling-li, ZHANG Wei(College of Traffic and College of Ocean , Hohai University , Nanjing 210098 , China)

Abstract: A threshold auto-regressive (TAR) model was established for forecasting monthly water stage at the Gaoqiao hydrological station in the lower tidal reaches of the Yangtze River, which is under the impacts of both tides and runoff. The genetic algorithm was adopted for parameter optimization in modeling of TAR. The calculated result shows that the TAR model can effectively reflect the nonlinear characteristics of tidal rivers, and the precision of fitting and simulation accords with the standard of hydrological forecasting. The application of the genetic algorithm simplifies the process of modeling, improves the forecasting accuracy, and ensures the high stability of the model. Therefore, the TAR model based on the genetic algorithm for water stage forecasting of tidal rivers is practical and has high adaptability to forecasting of other hydrological nonlinear time series in tidal rivers.

Key words: tidal river; river water stage; threshold auto-regressive model; genetic algorithm

河道的水位预测是水文预报的一个重要组成部分,它的基本依据是大量的水文、气象、地形等实测资料,一般可采用缓变非恒定流的圣维南方程组建模计算.但是,由于受外海潮汐和内陆径流等众多不确定因素的影响,感潮河段的水位系列在时序上常常表现出相依性、突变性和随机性等复杂非线性特征,至今对此进行有效描述的数学模型仍不很成熟.常见的应用缓变非恒定流的圣维南方程组来建立考虑潮水顶托影响的河道水位过程预报模型,虽已成为一个通用的方法,但它所需的资料包括了一般不易具备的河道地形资料,不但求解繁琐,实时校正也较困难,故离实用尚有较大差距.20 世纪 70 年代以来,随着统计理论与时间序列理论的发展,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模拟与预报取得了满意的成果.特别是 Box 和 Jenkins^[1]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^[2].其中, Tong 等^[3]提出的门限自回归 (Threshold Auto-Regressive, 简称 TAR) 模型,能有效地描述具有极限点、极限环(准周期性)、跳跃性、相依性、谐波等复杂现象的非线性动态系统,门限的控制作用保证了 TAR 模型的预测精度和适应性,是解决含有较复杂循环的非线性时间序列的建模与预报的理想模型之一,已成为当今工程界应用最广泛且技术较为成熟的非线性时序模型^[3],并且吸引了不少学者将之应用到水科学领域中^[4-6].

鉴于门限自回归模型在表征非线性特性上具有独到之处,笔者将之引入感潮河段的水位过程模拟并以长江下游感潮河段高桥水文站的月水位资料为

作者简介:吴玲莉(1978—),女,安徽无为,人,博士研究生,从事港口航道及海岸工程研究.

基础,探讨这种模型在感潮河段模拟中的可行性和模拟效果.计算结果表明,遗传门限自回归模型是可行而有效的,可以推广到感潮河段其他水文要素的拟合及预测中.

1 遗传门限自回归模型简介

由汤家豪博士首先提出的门限自回归模型的一般形式为

$$\begin{cases} x(i) = b(j,0) + \sum_{k=1}^{n_j} b(j,k)x(i-k) + e(j,i) \\ x(i-d) \in [r(j-1), r(j)] \\ r(0) = -\infty \quad r(L) = +\infty \\ (j = 1, 2, \dots, L) \end{cases} \quad (1)$$

式中: $r(j)$ 为门限值 ($j = 1, 2, \dots, L-1$); L 为门限区间的个数; d 为延迟步数; $b(j,k)$ 为第 j 个门限区间内的自回归系数; n_j 为第 j 个门限区间自回归模型 (简称 AR 模型) 的阶数; $\{e(j,i)\}$ 为对应每一固定 j 的固定方差的白噪声序列 ($j = 1, 2, \dots, L$), 各 $e(j,i)$ 之间互相独立.

门限自回归的基本思路为: 在观测时序 $\{x(i)\}$ 的取值范围内引入 $L-1$ 个门限值 $r(j)$ ($j = 1, 2, \dots, L-1$), 将其分成 L 个区间, 并根据延迟步数 d 将 $\{x(i)\}$ 按 $\{x(i-d)\}$ 值的大小分配到不同的门限区间内, 再对不同区间内的 $x(i)$ 采用不同的 AR 模型来描述, 这 L 个 AR 模型的总和构成了对时序 $\{x(i)\}$ 非线性动态系统的完整描述.

门限自回归模型能有效地逼近非线性序列, 能保证系统稳定递推, 反映非线性振荡、极限环、共振跳跃及突变等现象. 因为引入了门限值 $r(1), r(2), \dots$ 以及延迟步数 d , 反映了周期数据的本质和内在规律, 为线性模型所不及.

由于 TAR 模型是分区间的 AR 模型, 因此在建模中只需沿用一般 AR 模型的参数估计方法和模型检验准则, 如最小二乘法与 AIC 准则. 与 AR 建模有所不同的是, 建立 TAR 模型的关键不在于参数估计和模型检验, 而在于确定门限参数 (门限区间数 L , 门限值 $r(1), r(2), \dots, r(L-1)$ 和延迟步数 d). TAR 模型的建模过程, 实质上是一个对模型参数 $d, L, b(j,k), r(1), r(2), \dots, r(L-1)$ 以及 $n_1, n_2, \dots, n_L$ 的高维寻优问题, 常规做法是以 AIC 准则作为模型定阶 (确定各门限区间 AR 模型阶数 $n_1, n_2, \dots, n_L$) 和确定门限参数的目标函数, 可表示为 $f(n_1, n_2, \dots, n_L, d, r(1), r(2), \dots, r(L-1))$, 当该目标函数值最小时对应的模型参数、门限参数即为所求. 可见常规方法的计算量相当大, 在一定程度上限制了其实用性. 鉴于此, 笔者在建模过程中将遗传算法引入

到 TAR 模型的参数寻优中, 建立了遗传门限自回归模型. 建模步骤为: 用自相关分析技术^[2]确定时间序列的延迟特性, 据此用点值图 (dot figure)^[3] 确定 TAR 模型的门限区间个数和门限值的寻优范围, 然后直接在模型拟合误差最小准则下, 用遗传算法同时优化门限值和 TAR 模型的自回归系数.

2 感潮河段月平均水位预报模型的建立

分别用长江下游感潮河段高桥水文站 1968 ~ 1984 年逐月最高水位序列和逐月平均水位序列 $\{x(i) | i = 1, 2, \dots, 204\}$ 建立 TAR 预测模型, 用 1985 ~ 1987 年逐月水位序列进行预报检验. 囿于篇幅, 文中具体分析时只列出逐月平均水位序列的建模过程, 对于逐月最高水位仅给出建模结果. 借鉴文献 [7] 的方法, 建模步骤如下:

a. 用自相关分析技术确定 TAR 模型的延迟步数 d 和门限区间 AR 模型的阶数 n_j , 时序 $\{x(i)\}$ 延迟 k 步的自相关系数为

$$R(k) = \frac{\sum_{i=k+1}^n (x(i) - \bar{x})(x(i-k) - \bar{x})}{\sum_{i=1}^n (x(i) - \bar{x})^2} \quad (2)$$

式中: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x(i)$, n 为实测时序 $\{x(i)\}$ 的容量, $k = 1, 2, \dots, n_k$, $n_k < \frac{n}{10}$ 或 $\frac{n}{4}$. 根据 $R(k)$ 的抽样分析理论, 在置信水平 $1 - \alpha$ 的情况下, 当自相关系数值满足下面的条件:

$$R(k) \notin \left[\frac{-1 - u_{\alpha/2}(n-k-1)^{0.5}}{n-k}, \frac{-1 + u_{\alpha/2}(n-k-1)^{0.5}}{n-k} \right] \quad (3)$$

此时则推断时序 $\{x(i)\}$ 延迟 k 步相依性显著, 其中的 $|R(k)|$ 最大所对应的 k 值就是 TAR 模型的延迟步数 d ; 否则时序 $\{x(i)\}$ 延迟 k 步相依性不显著^[1]. 其中, 分位值 $u_{\alpha/2}$ 可从正态分布表中查得. 门限区间 AR 模型的阶数 n_j 不应大于相依性显著的最大延迟步数, 它的自回归系数项应与这些相依性显著的延迟步数相对应.

计算出前 20 阶自相关系数值 $R(k)$ 及对应的上、下限 $R_1(k)$ 和 $R_2(k)$ 值见表 1, 其中置信水平取 90%. 由表 1 可看出, $R(1), R(2), R(5) \sim R(7)$ 以及 $R(11) \sim R(13)$ 的相依性在置信水平 90% 的条件下都比较显著, 最为显著的是 $R(1)$ 和 $R(12)$. 其中 $R(2)$ 近似为 $R(1)^2$, 这正是 AR(n) 模型自相关系数的拖尾性特征, 因此仅用 $R(1)$ 就可大致反映 $R(2)$ 的相依特性. 同时 $R(12)$ 的显著性表明该序列具有

准年周期性,至于其他较为次要的相依性显著的延迟步数,考虑到稳健的预测模型的参数不宜太多,故这里只取延迟 1 步、延迟 12 步作为门限区间 AR 模型的自回归系数项,TAR 模型的延迟步数 d 取为 1。

表 1 高桥水文站月平均水位自相关系数及其上、下限值

k	$R_1(k)$	$R(k)$	$R_2(k)$	k	$R_1(k)$	$R(k)$	$R_2(k)$
1	-0.120	0.795	0.111	11	-0.124	0.717	0.113
2	-0.121	0.452	0.111	12	-0.124	0.829	0.114
3	-0.121	0.031	0.111	13	-0.124	0.710	0.114
4	-0.121	-0.356	0.111	14	-0.125	0.391	0.114
5	-0.122	-0.638	0.112	15	-0.125	0.001	0.114
6	-0.122	-0.752	0.112	16	-0.125	-0.369	0.115
7	-0.122	-0.647	0.112	17	-0.126	-0.626	0.115
8	-0.123	-0.376	0.112	18	-0.126	-0.707	0.115
9	-0.123	0.019	0.113	19	-0.126	-0.597	0.116
10	-0.123	0.420	0.113	20	-0.127	-0.341	0.116

b. 运用文献[3]的点值图法确定 TAR 模型的门限区间个数 L 和门限值 $r(1) \sim r(L-1)$ 的寻优范围. 考虑数据对集合 $\{(x(i), x(i-p)) | i=1, 2, \dots, M; p=1, 2, \dots, M_k\}$, 把 $x(i-p)$ 作为坐标, 将其分为均匀的 s 段 (s 远大于 L), 设共有 M_j 个 $x(i-p)$ 落在 j 段, 这 M_j 个 $x(i-p)$ 所对应的 $x(i)$ 值记为 $x(i, j)$, 则第 j 段内 $x(i)$ 对 $x(i-p)$ 的条件期望估计值为

$$E\left(\frac{x(i)}{x(i-p)}\right)_j = \frac{1}{M_j} \sum_{i=1}^{M_j} x(i, j) \quad (4)$$

以 $x(i-p)$ 作为横坐标, $E\left(\frac{x(i)}{x(i-p)}\right)$ 作为纵坐标, 在各段的中心位置上标出点 $E\left(\frac{x(i)}{x(i-p)}\right)_j$, 即得点值图. 根据点值图的分段线性的段数可确定门限区间个数 L , 在相邻两线段的转折点附近确定各门限值 $r(1), r(2), \dots, r(L-1)$ 的搜索范围, 从而缩短寻优过程.

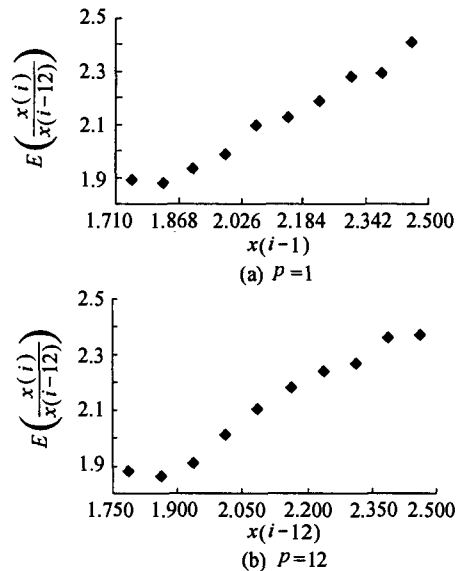


图 1 高桥水文站月平均水位数据点值

月平均水位序列点值如图 1 所示,点值图大致可分为两段直线,据此可确定门限区间个数 $L=2$, 同时可以初步确定门限值 $r(1)$ 的搜索范围大约为 $(1.8, 1.9)$ 或 $(2.34, 2.5)$ 。

模型的具体形式可写为

$$x'(i) = \begin{cases} b(1,1)x(i-1) + b(1,6)x(i-6) + \\ \quad b(1,12)x(i-12), x(i-1) \leq r(1) \\ b(2,1)x(i-1) + b(2,6)x(i-6) + \\ \quad b(2,12)x(i-12), x(i-1) > r(1) \end{cases} \quad (5)$$

式中: i 为时间序列的序号, $i=1$ 对应于 1964 年 1 月, $i=14$ 对应于 1965 年 2 月, 依此类推; $x'(i)$ 为第 i 个月平均水位的预测值. 序列 $x(i)$ 值都已经进行了中心化处理.

c. 各门限值 $r(1), r(2), \dots, r(L-1)$ 和各门限区间内的自回归系数 $b(j, k)$ 的优化. 本文采用遗传算法直接在模型拟合误差最小准则下同时优化门限值和自回归系数. 考虑到实际估计模型参数时用最小一乘法比用最小二乘法稳健^[7], 这里的优化准则函数为模型的最小拟合绝对偏差:

$$\min H(r(1), r(2), \dots, r(L-1); b(j, k)) = \sum_i |x'(i) - x(i)| \quad (6)$$

式中: $x'(i)$ 为式(1)中除白噪声项以外的所有项(即 TAR 模型估计值), 它是各门限值 $r(1) \sim r(L-1)$ 和各门限区间内 AR 模型的自回归系数 $b(j, k)$ 的函数; $x(i)$ 为观测值.

将式(5)代入式(6), 即得到优化的目标函数, 用 GA 优化 TAR 模型的参数, 采用实数编码, 群体规模为 300, 交叉概率 90%, 变异概率 1%, 替换概率 25%, 进化代数 1000 代, 得模型参数优化结果 $b(1, 1)=0.16, b(1, 6)=-0.20, b(1, 12)=0.63, b(2, 1)=-0.89, b(2, 6)=4.55, b(2, 12)=4.35, r(1)=2.49$, 最佳目标函数值 $H=12.13$ m.

3 模型预测及检验

将优化参数值代入式(5), 即得月平均水位的 TAR 预测模型, 同样可建立月最高水位序列的 TAR 预测模型. 运用 TAR 模型即可对月水位进行拟合和预测. 表 2 分别列出了 TAR 模型拟合高桥水文站 1968~1984 年的逐月水位和预测该站 1985~1987 年逐月水位序列的误差分析结果. 作为比较, 自回归模型的结果也列于表中.

从表 2 可以看出, TAR 模型的预测精度明显高于自回归模型, 其拟合及预测的平均误差均在许可误差以内, 并且各预测检验指标值十分接近于拟合

表2 用 TAR 模型拟合及预测水位时序的误差分析

项 目	模型类型	平均绝对 误差/m	平均相对 误差/%	拟合或预测误差绝对值落在各区间的百分比/%					
				[0,0.1]	[0,0.2]	[0,0.3]	[0,0.4]	[0,0.5]	[0,0.6]
月平均 水位	TAR 模型拟合	0.063	3.01	76.56	100.00	100.00	100.00	100.00	100.00
	TAR 模型预测	0.058	2.88	83.33	100.00	100.00	100.00	100.00	100.00
	自回归模型拟合	0.067	3.18	78.65	98.96	100.00	100.00	100.00	100.00
	自回归模型预测	0.080	3.98	72.22	94.44	100.00	100.00	100.00	100.00
月最高 水位	TAR 模型拟合	0.219	5.07	30.21	57.29	72.40	85.94	90.63	93.75
	TAR 模型预测	0.186	4.36	22.22	52.78	83.33	97.22	100.00	100.00
	自回归模型拟合	0.185	4.24	36.98	65.63	80.21	90.10	95.83	96.88
	自回归模型预测	0.217	5.14	30.56	44.44	69.44	91.67	97.22	100.00

检验的相应指标值,显示出 TAR 模型稳健的预测性能.常用的统计模型为了拟合实测水位时序复杂的动态特征,往往采用较多的模型参数,从而导致模型的不确定性较大,表现为模型拟合精度虽然较高而预测精度较低,从而限制了预测模型的实用价值. TAR 模型虽然仅利用月平均水位序列自身延迟 1 步、延迟 12 步和延迟 6 步的相依特征信息,但由于有了门限的控制作用,仅用 7 个模型参数已可以有效地描述长江感潮河段河口处高桥水文站的分段相依特性.同时由于高桥水文站地处长江河口处,濒临外海,月最高水位受风暴潮增水的外界影响较大,所以月最高水位模型的拟合精度和预测精度均低于月平均水位模型,但两者都达到了文献[8]的规范要求.综上所述, TAR 模型用于模拟预测感潮河段月最高、月平均水位过程是可行的,成果是令人满意的.模型的拟合及预测效果见图 2、图 3,其中序号 1~192 对应于 1968~1984 年逐月水位的拟合过程,序号 193~228 对应于 1985~1987 年逐月水位的预测过程.

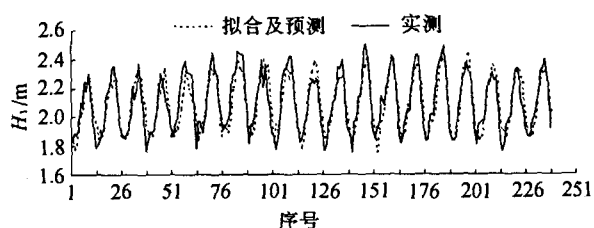


图2 TAR 模型拟合及预测高桥站月平均水位

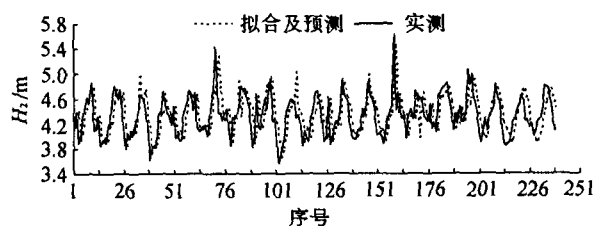


图3 TAR 模型拟合及预测高桥站月最高水位

4 结 语

a. 笔者首次尝试采用时间序列分析结合最优控制理论,将遗传门限自回归模型用于感潮河段的

月水位序列的随机模拟.该方法既可用于受潮汐影响明显的感潮河段的水位预报,也可用于潮汐影响不明显的河段水位预报,为感潮河段水文因素的预报提供了一种不同于水文学途径的随机建模途径.

b. 建立门限自回归模型的关键在于确定门限参数(门限区间个数、门限值 and 门限延迟步数),传统的做法是人为假设一系列门限值,在预测误差最小化准则下通过大量试算来选取,计算过程相当复杂,而本文则结合最优控制理论,将遗传算法应用于 TAR 模型的参数优化问题,简化了建模过程.

c. 遗传门限自回归模型是一种结构简单、概念清楚的非线性时序模型,可用此模型反映感潮河段水位时序上的非线性特性.计算结果表明,用 TAR 模型预测感潮河段月平均水位及月最高水位是可行的,此模型计算简便、精度较高、适用性强,可以推广应用于感潮河段或水情复杂的河段中其他水文要素的预测.

参考文献:

- [1] 丁晶,邓育仁.随机水文学[M].成都:成都科技大学出版社,1988.33—226.
- [2] George E P B, Gwilym M J, Gregory C R.时间序列分析:预测与控制[M].顾岚译.北京:中国统计出版社,1997.23—49.
- [3] 杨叔子,吴雅,王治藩,等.时间序列分析的工程应用(下册)[M].武汉:华中理工大学出版社,1992.309—348.
- [4] 金菊良,杨晓华,金保明,等.具有周期变化和下降趋势的地下水位的预测[J].水利水运科学研究,2000(9):53—57.
- [5] 金菊良,杨晓华,丁晶.年径流预测的遗传门限自回归模型[J].四川水力发电,2001(3):22—24.
- [6] 王文圣,袁鹏,丁晶,等.门限自回归模型及其在水文随机模拟中的应用[J].四川水力发电,2001(7):47—50.
- [7] 金菊良,丁晶.遗传算法及其在水科学中的应用[M].成都:四川大学出版社,2000.91—115.
- [8] SD138-85,水文情报预报规范[S].

(收稿日期:2004-11-22 编辑:高建群)