

文章编号:1671-6833(2004)04-0067-07

水文时间序列周期叠加预测模型及 3 种改进模型

左其亭¹, 高峰²

(1. 郑州大学环境与水利学院, 河南 郑州 450002; 2. 河南省睢县水利局, 河南 睢县 476000)

摘 要: 自然界中水文现象常暗含周期性, 基于这一特性, 发展了水文时间序列周期叠加预测模型. 介绍了水文时间序列周期叠加预测模型(POL), 分析了水文时间序列周期叠加预测模型(POL)在某些特定条件下应用上的不足, 并提出 3 种改进模型: 趋势一周期叠加组合预测模型(TPOLC), 压缩振幅的周期叠加预测模型(AC-POL), 逆推优势周期的周期叠加预测模型(DPI-POL). 这些模型可根据需要组合建模, 并可广泛应用于相应类型水文时间序列建模.

关键词: 水文时间序列; 预测; 周期叠加预测模型(POL); 趋势一周期叠加组合预测模型(TPOLC); 压缩振幅的周期叠加预测模型(AC-POL); 逆推优势周期的周期叠加预测模型(DPI-POL)

中图分类号: TV 214 **文献标识码:** A

0 引言

水文时间序列呈现周期性是常有的现象. 比如, 降水量总是在每年的雨季比较大, 在其它月份较小, 多年观测值曲线呈现周期变化; 再比如, 季节性河流、以高山冰雪融水为主要补给水源的河流流量; 即使是非季节性大江大河, 其流量、水位序列也暗含周期性特性^[1~3]. 因此, 在这种情况下, 建立水文时间序列预测模型, 需要考虑其序列周期性特点.

时间序列的建模方法很多, 不同方法对应有不同的适用条件^[4~5], 一般无法确定某一方法是绝对的适用或具有绝对的应用优势.

本文作者将讨论的是, 针对时间序列暗含周期性特点, 按照一定公式计算, 提取优势周期, 由若干优势周期叠加, 得到序列周期叠加函数, 从而建立周期叠加预测模型. 实践证明, 把该模型运用到水文时间序列上是可行的.

然而, 有时由于系统的复杂性, 使水文时间序列周期性表现得不是十分明显, 这会常常影响建模效果. 当然, 这不完全是方法本身的问题, 有时是由于观测资料规律性有限, 有时是该方法不太适合. 为此, 作者将针对不同情况提出 3 种改进模型.

(1) 针对水文时间序列有总体下降或上升趋势的情况, 提出“趋势一周期叠加相结合的预测模型(TPOLC)”, 该模型对有明显下降或上升趋势的水文时间序列预测效果很好.

(2) 针对水文时间序列周期性比较明显, 但年际变化较大(或者说振幅跳动较大)的一类情况, 提出“压缩振幅的周期叠加预测模型(AC-POL)”, 该模型对振幅跳动较大的水文时间序列预测效果很好.

(3) 如果水文时间序列中最后几个数值跳动特别大, 常规 POL 建模从第一项提取优势周期的计算, 对后面几个数值模拟影响较大. 针对这一问题, 作者采取从最后一项开始逆向提取优势周期, 从而建立逆推优势周期的周期叠加预测模型(DPI-POL).

1 周期叠加预测模型(POL)

设原始水文时间序列为: $x(k)$, $k = 1, 2, \dots, N$. POL 建模步骤如下.

第一步: 计算均值生成函数. 均值生成函数计算式

$$\bar{x}_l(i) = \frac{1}{n_l} \sum_{j=0}^{n_l-1} x(i+jl),$$
$$i = 1, 2, \dots, l; 1 \leq l \leq M,$$

收稿日期: 2004-08-10; 修订日期: 2004-09-20

基金资助: 水资源与水电工程科学国家重点实验室开放研究基金资助项目(2003B007); 2004 年河南省杰出青年科学基金资助项目; 河南省自然科学基金资助项目(311021600)

作者简介: 左其亭(1967-), 男, 河南省固始县人, 郑州大学副教授, 博士, 主要从事水文学及水资源方面的研究.

式中: $n_l = [N/l]$, $M = [N/2]$, 方括号表示取整.

组成的一个均值生成函数矩阵为

$$\begin{bmatrix} \bar{x}(1) & \bar{x}(1) & \bar{x}(1) & \cdots & \bar{x}(M) \\ & \bar{x}(2) & \bar{x}(2) & \cdots & \bar{x}(M) \\ & & \bar{x}(3) & \cdots & \vdots \\ & & & & \bar{x}(M) \end{bmatrix},$$

对均值生成函数 $\bar{x}_l(i)$ 作周期延拓,即令

$$f_l(k) = \bar{x}_l(i), k \equiv i[\text{mod}(l)],$$

式中: $k = 1, 2, \dots, N$; mod 表示同余; $f_l(k)$ 称作延拓均生函数.

第二步:提取优势周期.目前有如下两种方法:

① 依方差分析基本原理,可用下式来检验 $x(k)$ 是否隐含长度为 l^* 的周期:

$$F^{(l)} = \frac{S^{(l)} / (l - 1)}{S / (N - l)},$$

式中: $F^{(l)}$ 为服从自由度为 $(l - 1, N - l)$ 的 F 分布.

其中,

$$S^{(l)} = \sum_{i=1}^l n_i (\bar{x}_l(i) - \bar{x})^2;$$

$$n_i = [\frac{N}{i}], \bar{x} = \frac{1}{N} \sum_{i=1}^N x(i);$$

$$S = \sum_{i=1}^l \sum_{j=1}^{n_i} \{x[i + (j - 1)l] - \bar{x}_l(i)\}^2.$$

对于事先给定的置信水平 α ,若 $F^{(l^*)} > F(l^* - 1, N - l^*)$,则认为 $x(k)$ 隐含长度为 l^* 的优势周期.

② 欲确定长度为 l^* 的优势周期,只需取

$$\frac{S(l^*)}{l^*} = \max_l \frac{S(l)}{l}, 2 \leq l \leq M.$$

第三步:由原始序列 $x(k)$ 减去周期所对应的延拓均生函数构成一新序列.即

$$x'(k) = x(k) - f_{l^*}(k),$$

再对此新序列 $x'(k)$ 重复第二、第三步,可以进一步提取其它优势周期.

第四步:叠加.假设寻找到 t 个周期,则令不同周期的同一时刻取值的叠加

$$\tilde{x}(k) = f(k) = \sum_{i=1}^t f_{l_i}(k)$$

作为 $x(k)$ 的一个近似.这就是所建的周期叠加预测模型(POL).

选择新疆大山口水文站 1978 ~ 1997 年月平均流量值序列(见表 1),作为应用实例,建立 POL 模型,计算值与实际值对比如图 1 所示,拟合效果很好.

表 1 应用实例数据:新疆大山口水文站月平均流量值

Tab.1 Average monthly runoff at the Dashankou Hydrological Station in Xinjiang, China $m^3 \cdot s^{-1}$

年份	1 月份	2 月份	3 月份	4 月份	5 月份	6 月份	7 月份	8 月份	9 月份	10 月份	11 月份	12 月份
1978	39.99	37.81	41.66	89.03	99.88	240.50	152.13	154.82	113.73	73.41	54.84	43.43
1979	38.70	38.60	39.40	96.00	95.00	187.00	175.00	155.00	92.00	88.80	54.30	46.50
1980	44.10	44.50	44.60	158.00	290.00	165.00	150.00	150.00	113.00	92.70	70.30	52.10
1981	48.00	42.50	46.50	92.00	137.00	148.00	198.00	152.00	91.20	76.40	52.30	44.20
1982	44.30	44.60	43.80	112.00	151.00	215.00	177.00	164.00	125.00	89.00	64.60	46.80
1983	43.60	42.30	44.70	78.60	119.00	173.00	146.00	155.00	110.00	74.40	52.40	48.80
1984	41.60	43.10	42.60	92.90	154.00	193.00	141.00	120.00	82.00	69.50	51.50	36.50
1985	33.20	35.60	40.70	115.00	175.00	193.00	131.00	141.00	78.90	66.80	48.60	41.00
1986	39.60	37.00	38.30	75.00	86.70	145.00	154.00	127.00	76.90	65.00	51.20	38.00
1987	39.10	36.30	39.10	85.00	115.00	220.00	318.00	155.00	90.60	72.00	52.40	47.90
1988	46.30	44.20	42.30	96.70	180.00	180.00	153.00	177.00	110.00	78.30	57.30	45.30
1989	43.00	40.10	40.90	77.10	142.00	128.00	183.00	186.00	211.00	101.00	63.00	51.20
1990	48.10	41.90	45.45	107.01	198.00	188.90	188.19	131.87	100.00	75.48	57.79	48.62
1991	42.17	41.23	43.40	86.90	124.00	141.00	238.00	260.00	109.00	80.26	58.09	45.20
1992	42.22	41.06	42.50	95.80	155.00	241.00	226.00	140.00	99.40	78.76	57.21	44.82
1993	42.08	40.64	50.50	115.00	102.00	181.00	180.00	159.00	103.00	79.50	56.75	44.74
1994	50.90	45.10	49.60	85.10	149.00	268.00	284.00	193.00	152.00	96.70	83.10	61.20
1995	56.80	53.60	52.00	95.20	106.00	139.00	161.00	138.00	114.00	75.50	66.00	64.90
1996	61.60	48.00	48.70	113.00	206.00	195.00	245.00	245.00	117.00	90.70	73.90	57.90
1997	58.10	47.20	52.30	181.00	121.00	166.00	220.00	172.00	130.00	79.40	58.00	60.60

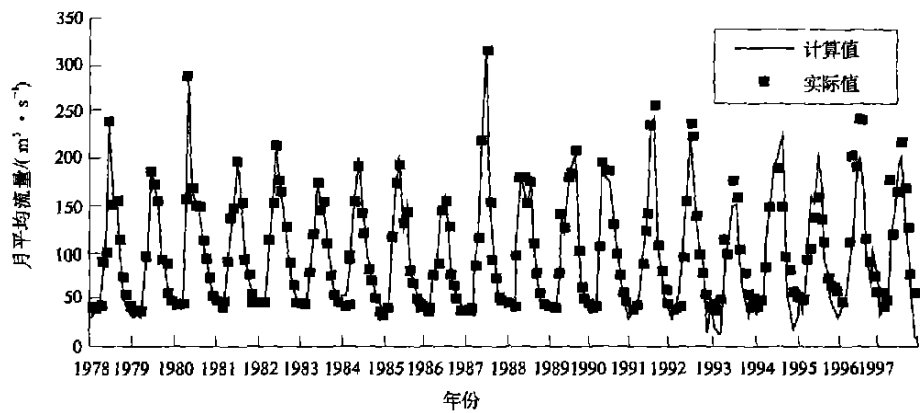


图 1 月平均流量POL 计算值与实际值对比图

Fig. 1 Comparison chart of the calculated values with the POL model and the gauged values of the average monthly runoff

2 趋势—周期叠加组合预测模型(TPOLC)

水文时间序列中,除暗含周期变化外,有时还存在增大或减小的趋势.表 2 是某观测井地下

水位观测值序列,该序列反映了由降水直接补给(呈周期性)且受人类用水量增加(导致水位逐渐下降)影响的地下水位变化.

表 2 应用实例数据:河北省一观测井地下水位观测值

Tab. 2 Measured values of the groundwater level from an observation well

月份	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
2 月	27.86	27.38	26.88	26.17	25.64	25.22	25.10	24.84	24.75	24.61	24.50
6 月	25.67	25.55	25.42	24.86	24.33	24.00	23.70	23.61	23.49	23.40	23.48
10 月	27.33	26.91	26.40	25.87	25.42	25.12	24.93	24.83	24.73	24.56	24.60

对于这种序列,仅采用POL 模型建立的模型,不能很好地反映整体序列的增加或减小趋势,其拟合效果很差,主要原因是周期叠加预测模型不能反映序列总体变小的趋势(模拟图此略).根据这一特征,可以先选择一个能表征序列趋势的模型进行建模,再对序列的残差建立周期叠加模型.这样组合处理,既能反映序列变化趋势,又能表述序列周期变化特征.本文把这一改进模型称为趋势—周期叠加组合预测模型,其基本思路是如下.

第一步:建立序列 $x(k)$, $k=1,2,\dots,N$ 的趋势预测模型,设为 $\hat{x}(k)$.

第二步:计算原序列的残差序列 $x'(k)=x(k)-\hat{x}(k)$.

第三步:建立残差序列 $x'(k)$ 的周期叠加预测模型 $\tilde{x}(k)$,方法同上.

第四步:叠加.把 $\tilde{x}(k)=\hat{x}(k)+\tilde{x}(k)$ 作为 $x(k)$ 的一个近似,称为趋势—周期叠加组合

预测模型(TPOLC).

关于TPOLC 模型中趋势预测,可以选用的方法很多,比如灰色GM(1,1) 模型、自回归模型.本文针对实例选用灰色GM(1,1) 模型^[9],根据建模步骤,建立本实例GM(1,1) 模型,其趋势线和建立的TPOLC 模型拟合曲线如图 2 所示,模型拟合效果明显改善.

3 压缩振幅的周期叠加预测模型(AC—POL)

当水文时间序列的低谷值接近于 0 且峰值较大时,周期叠加预测模型拟合效果不好,且常出现低谷计算值小于 0 的情况.这在实际应用中很容易遇到,表 3 所示的是新疆博湖水文站观测的月蒸发量,对其建立POL 模型(模拟图略),计算得到的平均相对误差为 19.6%,且在某些点上出现负值.现实中,月蒸发量出现负值是不可能的,且拟合效果也不好.

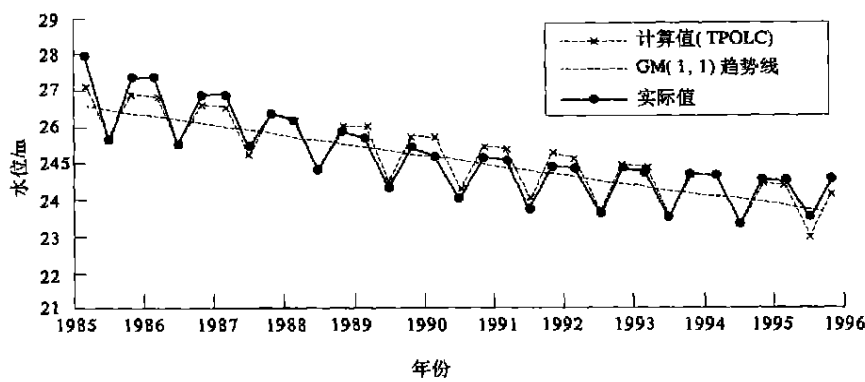


图 2 地下水位 TPOLC 计算值与实际值对比图

Fig. 2 Comparison curve of the calculated values with GM(1, 1) and the TPOLC model and the gauged values of the groundwater level

表 3 应用实例数据:新疆博湖水文站月蒸发量值

Tab. 3 The monthly evaporation at the Bosteng Lake Hydrological Station in Xinjiang, China

年份	1 月份	2 月份	3 月份	4 月份	5 月份	6 月份	7 月份	8 月份	9 月份	10 月份	11 月份	12 月份
1981	20.0	44.3	164.7	297.9	433.5	446.5	354.6	309.5	211.5	174.9	56.5	25.0
1982	24.1	46.0	161.5	322.6	426.1	402.5	392.7	344.6	196.0	150.5	47.7	19.2
1983	16.6	47.4	133.2	327.6	374.6	373.7	365.8	324.8	205.3	161.4	56.6	11.4
1984	16.8	43.4	135.5	294.5	392.2	388.9	368.9	320.0	222.7	149.4	59.5	20.2
1985	13.0	41.9	107.0	298.9	397.4	477.2	422.2	362.7	242.4	156.6	60.3	20.3
1986	23.9	52.6	106.7	250.9	330.5	342.4	381.6	321.4	211.1	141.6	61.6	9.8
1987	9.2	36.2	154.3	352.6	454.7	354.6	404.8	367.2	240.3	148.9	51.4	16.0
1988	16.4	44.2	124.3	309.4	298.7	415.3	359.6	290.9	216.6	127.2	48.6	24.1
1989	22.0	45.9	147.8	249.4	365.8	361.5	342.8	321.3	208.8	161.1	49.1	14.1
1990	26.6	53.4	142.8	276.9	394.6	431.5	318.9	309.1	240.3	112.6	78.5	19.8
1991	10.9	23.6	120.0	261.9	379.2	378.1	372.6	270.6	207.6	134.0	59.5	23.7
1992	12.7	42.7	105.3	321.4	372.3	356.2	298.6	246.2	190.9	121.8	60.2	24.9
1993	20.7	42.3	140.9	256.2	381.8	377.9	443.1	476.0	388.0	170.9	58.5	13.1
1994	12.2	39.3	115.6	265.5	455.3	454.6	310.1	305.3	320.3	191.6	77.0	14.6
1995	12.8	34.4	132.0	295.8	318.6	369.5	240.1	248.4	221.6	160.0	72.3	27.4
1996	14.8	43.5	154.5	261.6	329.0	329.6	404.9	273.2	245.0	155.6	55.5	17.2
1997	9.6	39.5	153.9	312.6	398.6	342.5	289.9	424.4	196.5	223.5	58.5	29.9
1998	14.0	34.2	108.0	263.0	263.5	294.1	270.8	208.7	215.1	137.4	67.5	21.3
1999	11.0	40.3	128.3	211.1	300.2	286.3	242.1	234.6	180.8	110.8	43.2	18.6
2000	11.5	29.3	97.2	230.0	277.5	234.2	258.6	228.1	168.8	87.3	32.9	10.7

为了解决这一问题,我们采取压缩振幅(即缩小值与值之间的差距)的方法.这种处理减少了序列波动的幅度,等于增强了周期性.另外,为了避免上文中模型计算值出现负值的现象,采取:先对原始数据取对数(即压缩);再模拟预测;最后再还原.本方法适用于原始数据都大于1的情况.如果有小于1的,需同时转化使其满足要求.当然,也可以采取其它处理方法来压缩振幅.本文只介绍取对数进行压缩的方法.

设原始数据列为 $x(k)$, $k=1,2,\dots,N$.

- 第一步:对原始数据取对数,得到新的序列
 $y(k)=\ln x(k)$, $k=1,2,\dots,N$.
- 第二步:对新序列 $y(k)$ 建立周期叠加预测模型,设为 $\hat{y}(k)$ (具体步骤同上).
- 第三步:还原,得原序列的预测模型
 $\hat{x}(k)=e^{\hat{y}(k)}$.
- 作为 $x(k)$ 的一个近似,称为压缩振幅的周期叠加预测模型(AC-POL).
- 对上文月蒸发量重新建立 AC-POL 模型,计算值与实际值对比如图 3 所示,计算的平均相对

误差为 8.8%. 模拟效果有较大提高, 并避免了出现负值的可能性, 结果比较符合实际.

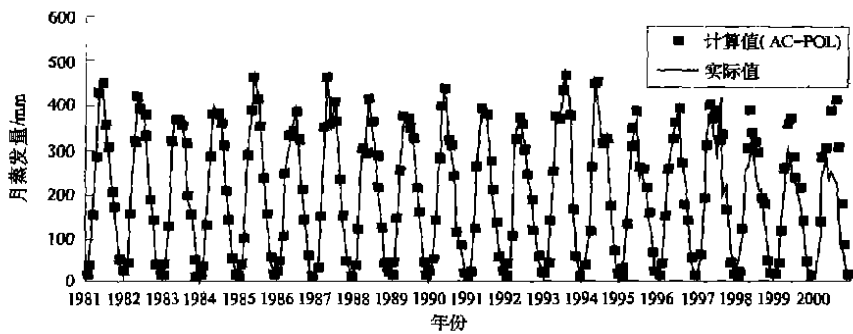


图 3 月蒸发量 AC-POL 计算值与实际值对比图
Fig. 3 Comparison curves of the calculated values with the AC-POL model and the gauged values of monthly evaporation

4 逆推优势周期的周期叠加预测模型 (DPI-POL)

前面介绍的周期叠加预测模型计算, 均是从序列的第一项开始向后计算提取优势周期和周期延拓. 这样, 可能会对序列的最后几项数据考虑不够, 特别是当最后几项数据异常变化时, 这种误差会更大. 比如, 原始序列共有 100 个数据, 假如提取到优势周期为 12, 在作均值生成函数计算时, 实际上只计算到 96 个数据 (即 8 个周期), 剩余的最后 4 个数据根本没有考虑. 下面还是引用一个实例来说明, 这个实例是本文表 1 实例的延续. 在表 1 中, 列举了新疆大山口水文站 1978~1997 年月平均流量值序列. 实际上, 该水文站观测的 1998~2000 年月平均流量值与历史数据相比异常增大 (对比表 1 和表 4). 在上文介绍 POL 模型应用实例时没有包括这 3 年数据. 如果用 POL 模型对 1978~2000 年月平均流量值序列进行建模 (对比图此略), 结果是最后几年拟合效果很差, 1998~2000 年水文站出现特殊丰水期, 但在模型计算中没有科学反映.

针对这类情况, 可以调整计算顺序, 从序列中的某一项开始计算到最后一项, 把第项作为计算的起点, 其中, $N-i+1$ 肯定是周期的倍数. 或者说, 从系列的最后往前逆推, 寻找优势周期. 这就是逆推优势周期的周期叠加预测模型 (DPI-POL) 基本思路, 步骤如下.

第一步: 逆推计算均值生成函数

$$\bar{x}_l(i) = \frac{1}{n_l} \sum_{j=0}^{n_l-1} x(R_l + i + jl),$$
$$i = 1, 2, \dots, l; 1 \leq l \leq M,$$

式中: $n_l = [N/l]$, $M = [N/2]$, $R_l = N - n_l \cdot l$, 方括号表示取整. R_l 是序列总项为 N 、周期为 l 的剩余项数.

对均值生成函数 $\bar{x}_l(i)$ 作周期延拓, 即令 $f_l(k) = \bar{x}_l(i), k+l-R_l \equiv i[\text{mod}(l)], k = 1, 2, \dots, N$.

第二步: 提取优势周期. 计算方法同 POL 模型, 只有 S 计算式改为

$$S = \sum_{i=1}^l \sum_{j=1}^{n_i} \{x[R_l + i + (j-1)l] - \bar{x}_l(i)\}^2.$$

第三步: 生成新序列 $x'(k)$. 计算式同 POL 模

表 4 新疆大山口水文站 1998~2000 年月平均流量值

Tab. 4 Average monthly runoff measured at the Dashankou Hydrological Station in Xinjiang, China during the period from 1998 to 2000

年份	1 月份	2 月份	3 月份	4 月份	5 月份	6 月份	7 月份	8 月份	9 月份	10 月份	11 月份	12 月份
1998	66.0	67.1	63.2	105.0	169.0	268.0	200.0	208.0	138.0	107.0	91.0	79.5
1999	70.5	65.4	65.9	107.0	139.0	188.0	356.0	380.0	175.0	120.0	93.6	73.4
2000	76.6	58.1	61.1	132.0	158.0	333.0	342.0	293.0	165.0	122.0	83.6	66.8

型

$$x'(k) = x(k) - f_{l_i}(k),$$

再对此新序列 $x'(k)$ 重复第二、第三步,可以进一步提取其它优势周期.

第四步:叠加.计算式同POL 模型

$$\tilde{x}(k) = f(k) = \sum_{i=1}^l f_{l_i}(k)$$

作为 $x(k)$ 的一个近似.这就是所建的逆推优势周

期的周期叠加预测模型(DPI-POL).

这种逆推优势周期的处理也可以使用于趋势一周期叠加组合预测模型(TPOLC)和压缩振幅的周期叠加预测模型(AC-POL).针对应用实例——新疆大山水文站 1978~2000 年月平均流量值序列,采用 DPI-POL 模型方法,建立模型,结果如图 4,最后几年的拟合效果有很大提高.

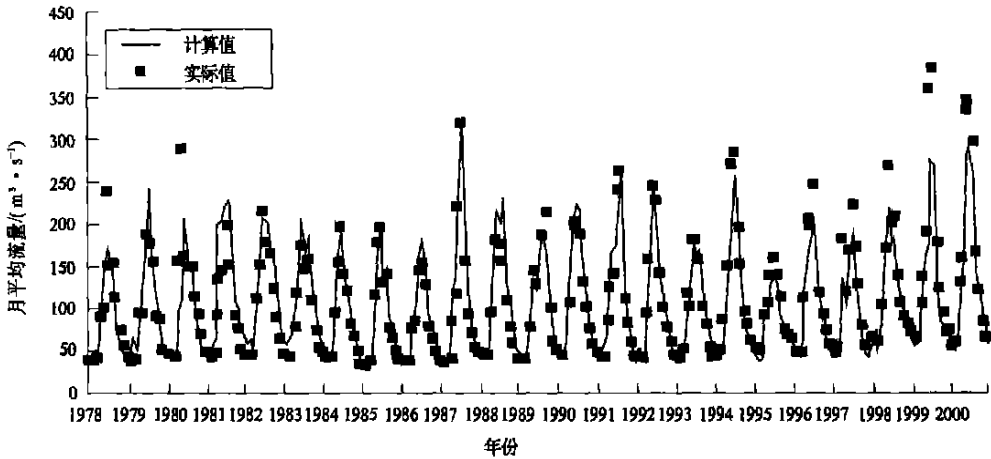


图 4 月平均流量 DPI-POL 计算值与实际值对比图

Fig. 4 Comparison curves of the calculated values with the DPI-POL model and the gauged values of the average monthly runoff

5 结论

本文针对水文时间序列的不同情形,介绍周期叠加预测模型(POL)及其 3 种改进模型.

趋势一周期叠加组合预测模型(TPOLC),主要针对水文时间序列暗含增加或减小趋势的情况;压缩振幅的周期叠加预测模型(AC-POL),主要针对水文时间序列中低谷数据接近于 0、振幅比较大的情况;逆推优势周期的周期叠加预测模型(DPI-POL),主要是在计算方法上的改进,是从序列的最后项开始提取周期和叠加计算.

各种类型模型还可以根据需要组合建模,这些模型可广泛应用于对应类型水文时间序列建模.

参考文献:

[1] 魏风英,曹鸿兴.长期预测的数学模型及其应用[M].北京:气象出版社,1990.

[2] 左其亭.地下水动态预测的灰色一周期外延组合预测模型[J].水文地质工程地质,1996,(5):16~19.
[3] 左其亭,王中根.现代水文学[M].郑州:黄河水利出版社,2002.
[4] PETER J Brockwell, RICHARD A Davis. Time Series Theory and Methods[M]. New York:Springer-Verlag, 1991.
[5] MAURICE Kendall. Time Series[M]. London: Charles Giffin and Company Ltd. 1976.
[6] ZUO Q T, MA J X, ZHANG N K. The Grey-periodic extensional combination forecasting model and its application[J]. Advances in Systems Science and Applications, 1997,(Special Issue):592~595.
[7] XIA J, CHEN Z, HUANG G H et al. An Integrated hydro-ecological modeling approach applied to the lake bositeng basin in China[J]. Water International, 2001, 26(1): 105~118.
[8] XIA J. A grey system method applied to forecasting of lake water quality[J]. J. Grey System, 1990, 2(3): 257~266.

Periodic Overlapping Prediction Model and Its Three Improved
Models of Hydrological Time Series

ZUO Qiting¹, GAO Feng²

(¹College of Environmental & Hydraulic Engineering, Zhengzhou University, Zhengzhou 450002, China; ²Suixian Water Resources Bureau, Suixian 476000, China)

Abstract : The periodicity is usually implied in the hydrological phenomenon in the nature . Based on the characteristics of hydrological time series , periodic overlapping prediction model is proposed . In this paper , periodic overlapping prediction model is introduced at first ; then the insufficiency of periodic overlapping prediction model under certain conditions are pointed out based on the analysis of examples , and three improved models are put forward . The first improved model is trend — periodic overlapping combinatorial prediction model (TPOLC) ; the second improved model is amplitude — compacted periodic overlapping prediction model (AC — POL) ; the third improved model is dominant period inversely — deduced periodic overlapping prediction model (DPI — POL) . These models are all improved based on the practical work , which can be extensively used to build the modeling of corresponding type of hydrological time series .

Key words : hydrological time series ; prediction ; periodic overlapping prediction model (POL) ; trend — periodic overlapping combinatorial prediction model (TPOLC) ; amplitude — compacted periodic overlapping prediction model (AC — POL) ; dominant period inversely — deduced periodic overlapping prediction model (DPI — POL) .