

灰色马尔可夫链模型的改进及其应用

赵琳琳 夏乐天

(河海大学理学院,江苏 南京 210098)

摘要 将灰色模型与马尔可夫链方法相结合,建立了灰色马尔可夫链预测模型,并对该模型进行了改进.采用滑动转移概率矩阵,并依据广西合面狮水库 10 a 的年最大洪水流量预测以后 2 a 的年最大洪水流量范围.结果表明,采用改进的灰色马尔可夫链模型,可比较准确地预测随机波动性较大的水文数据序列的范围.

关键词 马尔可夫链;灰色模型;滑动转移概率矩阵;年最大洪水流量;合面狮水库

中图分类号 :O211.62 **文献标识码** :A **文章编号** :1000-198X(2007)04-0487-04

灰色预测^[1]的主要特点是模型不采用原始数据序列,而采用生成的数据序列.对于序列较短并具有明确上升趋势的数据来说,这种预测可以取得较好的预测效果^[2].

本文首先将灰色模型与马尔可夫链方法相结合,建立灰色马尔可夫链预测模型,并对该模型进行改进,然后采用滑动转移概率矩阵,并依据广西合面狮水库 10 a 的年最大洪水流量预测以后 2 a 的年最大洪水流量范围^[3-4].

1 灰色模型 GM(1,1)

令 $X^{(0)} = \{X^{(0)}(k) | X^{(0)}(k) \geq 0, k = 1, 2, \dots, n\}$ 其中 $X^{(0)}(k)$ 为某水文序列的第 k 个实测值.为构造一个满意的系统模型,对 $X^{(0)}$ 做一次累加生成数列 $X^{(1)} = \{X^{(1)}(k) | X^{(1)}(k) \geq 0, k = 1, 2, \dots, n\}$, 其中 $X^{(1)}(k) = \sum_{m=1}^k X^{(0)}(m)$, 并建立相应模型

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = u \quad (1)$$

用最小二乘法求解 a, u , 其白化形式的微分方程离散解为

$$\hat{X}^{(1)}(i+1) = \left(X^{(0)}(1) - \frac{u}{a} \right) e^{-ai} + \frac{u}{a} \quad (2)$$

累减还原即可得到灰色模型 GM(1,1):

$$\hat{X}^{(0)}(i+1) = \hat{X}^{(1)}(i+1) - \hat{X}^{(1)}(i) = (1 - e^a) \left(X^{(0)}(1) - \frac{u}{a} \right) e^{-ai} \quad (3)$$

2 马尔可夫链方法

$P_{ij}(k)$ 表示系统从状态 i 经 k 步状态转移到达状态 j 的概率,此时转移概率与初始时刻无关, P_k 表示 k 步转移概率矩阵.状态的 k 步转移概率 $p_{ij} = \frac{n_{ij}}{N_i}$, 其中 n_{ij} 表示状态 i 经过 k 步转移到状态 j 的次数.从而可求得 k 步转移概率矩阵^[5].设有限状态齐次马尔可夫链的初始分布为 $P(0) = (p_0(1), p_0(2), \dots, p_0(n))$, 则 k 时刻的绝对分布为

$$P(k) = (p_0(1), p_0(2), \dots, p_0(n)) P_k \quad (4)$$

3 改进的灰色马尔可夫链模型

3.1 一般灰色马尔可夫链模型

首先根据GM(1,1)模型求出水文序列相应的拟和值 $\hat{X}^{(0)}(k)$,然后求出残差 $\Delta(k)=X^{(0)}(k)-\hat{X}^{(0)}(k)$,残差相对值为

$$\epsilon(k)=\frac{X^{(0)}(k)-\hat{X}^{(0)}(k)}{X^{(0)}(k)}\times 100\%$$

(5)

再对残差相对值序列进行状态划分,求出残差相对值序列各值所处状态.最后针对残差相对值序列,建立马尔可夫链模型^[6].此模型即为灰色马尔可夫链模型.

3.2 一般灰色马尔可夫链模型的改进

一般的灰色马尔可夫链模型采用1步或2步转移矩阵,转移矩阵的数据来源固定不变.由于预测较远时刻对象所处的状态时,很久以前的数据对预测的影响不大,本文在预测1998年的状态时采用滑动转移概率矩阵,去掉了老的数据并补充了新的数据.对于状态序列,去掉第1个状态值并加上预测对象1997年的状态值,得出新的1步转移概率矩阵,再以预测对象1997年状态分布作为初始分布,求1998年所处状态.由于预测对象1997年所处各种状态有各自的概率,本文把各种状态分别看做是1997年的状态值,以这种状态值构建新的1步转移概率矩阵,把处于这种状态的概率看做是权重 $p_i(i)$. $P^i(1)$, $P^i(i=1,2,\dots,n)$ 分别表示预测对象时刻1以概率1处于状态*i*时的初始分布和所处状态为*i*时的滑动1步转移概率矩阵,则预测对象时刻2所处的状态概率分布为

$$P(2)=p_1(1)P^1(1)P^{(1)}+\dots+p_i(i)P^i(1)P^{(i)}+\dots+p_n(n)P^n(1)P^{(n)}$$

(6)

4 实例分析

依据合面狮水库10a的年最大洪水流量预测以后2a的年最大洪水流量范围.根据表1中的 $X^{(0)}(k)$ 求出年最大洪水流量GM(1,1)模型中的*a*,*u*,即*a*=-0.058264,*u*=1763.9490.由式(3)求得

$$\hat{X}^{(0)}(i+1)=(1-0.9434)(2098+30274)e^{0.058264i}=1832e^{0.058264i}$$

GM(1,1)模型的预测值 $\hat{X}^{(0)}(k)$ 与实际值 $X^{(0)}(k)$ 的残差相对值序列 $\epsilon(k)$ 的范围为(-45%,30%],且相对值 $\epsilon(k)$ 分布在0左右的数据比较多,根据实际情况(数据不多)分为3个状态,为了使每一状态的数据一样多,按 $\epsilon(k)$ 的值从小到大排列,取前3个值为第1状态,中间4个值为第2状态,后3个值为第3状态.3个状态分别用1,2,3表示.划分出的3个状态的区间分别为 $E_1=(-45\%,-10\%]$, $E_2=(-10\%,5\%]$, $E_3=(5\%,30\%]$,如表1所示.

求出的1步状态转移概率矩阵为 P_1 .预测对象1997年处于状态2,那么下一年的绝对分布为

表1 合面狮水库1988~1997年最大洪水流量相关数据

Table 1 Correlative data of annual maximum flood discharge of Hemianshi Reservoir from 1988 to 1997

序号	年份	$X^{(0)}(k)$ ($\text{m}^3\cdot\text{s}^{-1}$)	$\hat{X}^{(0)}(k)$ ($\text{m}^3\cdot\text{s}^{-1}$)	$\Delta(k)$ ($\text{m}^3\cdot\text{s}^{-1}$)	$\epsilon(k)\%$	状态
1	1988	2098	2098	0	0	2
2	1989	2316	1942	364	15.70	3
3	1990	1423	2058	-635	-44.60	1
4	1991	2396	2182	214	8.93	3
5	1992	2117	2313	-196	-9.26	2
6	1993	2225	2451	-226	-10.10	1
7	1994	3579	2599	980	27.30	3
8	1995	2517	2754	-237	-9.40	2
9	1996	2496	2920	-424	-16.90	1
10	1997	3254	3095	159	4.88	2

$$P(1)=(p_0(1),p_0(2),\dots,p_0(n))P_1=(0,1,0)\begin{bmatrix}0&\frac{1}{3}&\frac{2}{3}\\\frac{2}{3}&0&\frac{1}{3}\\\frac{1}{3}&\frac{2}{3}&0\end{bmatrix}=\left(\frac{2}{3},0,\frac{1}{3}\right)$$

所以下一年预测对象最有可能处于状态1,即 $\epsilon(k)\in E_1=(-45\%,-10\%]$.由 $\epsilon=\frac{X^{(0)}-\hat{X}^{(0)}}{X^{(0)}}$ 得 $X^{(0)}(k)=\frac{\hat{X}^{(0)}(k)}{1-\epsilon}$,所以 $X^{(0)}(11)\in(2262.7,2982.7)$.由于要预测的1998年最大洪水流量的实测值为 $2425\text{ m}^3/\text{s}$,在此

预测区间内.

现预测 1999 年所处状态.

a. 用传统的灰色马尔可夫链模型进行预测.应用马尔可夫链模型预测第 2 时刻的状态时,采用 2 步状态转移矩阵:

$$P(2) = (p_0(1), p_0(2), \dots, p_0(n))P_2 = (0, 1, 0) \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & 0 & \frac{1}{3} \end{bmatrix} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

结果表明,预测对象时刻 2 所处的各种状态概率都是 $\frac{1}{3}$,预测范围为

$$X^{(0)}(12) \in (2637.9\ 3\ 161.8) \cup (3\ 161.8\ 4\ 026.3) \cup (4\ 026.3\ 4\ 186.6)$$

b. 用改进的马尔可夫链模型预测时刻 2 的状态.假设 1998 年以 $\frac{2}{3}$ 概率处于状态 1,以 $\frac{1}{3}$ 概率处于状态 3,则采用滑动转移概率,去掉状态序列的第 1 个值即 1988 年的状态值,加上刚预测出的 1998 年的状态值,得出新的 1 步转移矩阵为

$$P^{(1)} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ 1 & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 \end{bmatrix}$$
$$P^{(3)} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & 0 \end{bmatrix}$$

其中 $P^{(1)}, P^{(3)}$ 分别表示 1998 年处于状态 1 或 3 时的 1 步转移概率矩阵.于是

$$P(2) = \frac{2}{3}(1\ 0\ 0) \begin{bmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ 1 & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 \end{bmatrix} + \frac{1}{3}(0\ 0\ 1) \begin{bmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & 0 \end{bmatrix} = \left(\frac{1}{9}, \frac{4}{9}, \frac{4}{9}\right)$$

可见 1999 年指标值的状态为 2 或 3,此时 $\varepsilon(k) \in E_2 \cup E_3 = (-10\%, 5\%] \cup (5\%, 30\%]$,由 $X^{(0)}(k) = \frac{\hat{X}^{(0)}(k)}{1 - \varepsilon}$ 得 $X^{(0)}(12) \in (3\ 161.8\ 4\ 026.3) \cup (4\ 026.3\ 4\ 186.6)$. 1999 年的年最大洪水流量实测值为 $3\ 165\text{ m}^3/\text{s}$,在此区间内.

由上可知,改进的灰色马尔可夫链模型比一般的灰色马尔可夫链模型预测的范围更准确,如表 2 所示.

表 2 1998 年和 1999 年合面狮水库年最大洪水流量预测区间

Table 2 Prediction range of annual maximum flood discharge of Hemianshi Reservoir for 1998 and 1999

序号	年份	年最大洪水流量实测值 $X^{(0)}(k)(\text{m}^3 \cdot \text{s}^{-1})$	GM(1,1)预测值 $\hat{X}^{(0)}(k)(\text{m}^3 \cdot \text{s}^{-1})$	一般灰色马尔可夫链模型预测的区间	改进的灰色马尔可夫链模型预测的区间
11	1998	2425	3281	(2262.7 2982.7)	(2262.7 2982.7)
12	1999	3165	3478	(2637.9 4186.6)	(3161.8 4186.6)

5 结 语

建立灰色模型时数据不能过多,5~10 个数据的处理效果比较理想,而且只能处理单调序列,随机波动

较大的数据列的预测是不可行的^[7].用灰色模型预测曲线来反映数据列的发展规律,用马尔可夫链模型来反映数据列的波动规律,则可以给出预测值的大致范围^[3-4].从理论上讲,历史数据愈长,预测精度愈高^[8].由于应用这种模型时采用 5~10 个数据比较理想,本文选用了 10 个数据.考虑到预测步长过长以及未把新的状态信息加进去,会使精度降低.本文采用滑动转移概率 1 步转移矩阵,去掉老的信息,加入新的状态信息,提高了预测精度.另外,划分的状态数愈多,预测精度愈高.但状态数目过多,会导致各状态样本点过少,转移概率规律性不强^[8],故本文只划分为 3 种状态.总之,用改进的灰色马尔可夫链模型来预测随机波动较大的数据列,可提高预测范围的准确度.

参考文献:

[1] 刘思峰,邓聚龙.GM(1,1)灰色模型的适用范围[J].系统工程理论与实践,2000(5):122-124.
[2] 赖红松,祝国瑞,董品杰.基于灰色预测和神经网络的人口预测[J].经济地理,2004,24(2):197-201.
[3] 梅振宇,王炜,李铭,等.高速公路交通生成预测的改进灰色马尔可夫模型[J].公路交通科技,2004(12):89-92.
[4] 赵佳妮,吴兆麟.基于灰色马尔可夫模型的水上交通事故预测[J].大连海事大学学报,2005,31(4):15-18.
[5] 夏乐天.梅雨强度的指数权马尔可夫链预测[J].水利学报,2005,36(8):988-993.
[6] 张克中,毛树华,袁卫红.马尔科夫残差修正灰色模型及其在公路网规划中的应用[J].武汉理工大学学报:交通科学与工程版,2005,29(4):503-505.
[7] 张超,马存宝,许家栋.基于灰色马尔可夫 SCGM(1,1)_c 模型的空难人数预测[J].系统工程理论与实践,2000(5):135-138.
[8] 张宸,林启太.模糊马尔可夫链状模型在矿区降水灾害预测中的应用[J].国外建材科技,2004,25(1):56-58.

Improved gray Markov SCGM(1,1) model and its application

ZHAO Lin-lin , XIA Le-tian

(College of Sciences , Hohai University , Nanjing 210098 , China)

Abstract :By combination of gray prediction model with Markov chain theory , a gray Markov SCGM(1,1) model was proposed and improved. Based on sliding transition probability matrix and the annual maximum flood discharge of 10 years of Hemianshi Reservoir , the range of the annual maximum flood discharge for the next 2 years was predicted. The result shows that the present model can accurately predict the range of hydrologic data sequences with large random fluctuation.

Key words :Markov chain ; gray model ; sliding transition probability matrix ; annual maximum flood discharge ; Hemianshi Reservoir