

DOI: 10.13476/j.cnki.nsbdtqk.2015.04.003

加权马尔可夫链模型在密云水库入库流量中的应用

贺娟, 王晓松, 王彩云

(中国水利水电科学研究院 水力学所, 北京 100038)

摘要: 根据1960年-2011年的实测入库流量资料,以河川径流量为相依随机变量,介绍加权马尔可夫链模型的相关概念及预测未来一年入库流量的步骤,采用均值-标准差分级法把入库流量序列划分成枯、偏枯、偏丰、丰4种状态。以各阶自相关系数为权重,预测2010年-2011年的入库流量,将其所在状态区间与实测值进行对比。结果表明,加权马尔可夫链模型对密云水库入库流量预测精度较高,以此又对2012年-2013年的入库流量进行了预测。最后对其遍历性和平稳分布进行分析,计算入流丰、枯状态在实测序列中的重现期,其中出现偏枯状态的概率最大,由此预测密云水库未来的入库流量处于偏枯状态。

关键词: 加权马尔可夫链模型; 密云水库; 入库流量; 转移概率矩阵; 马氏性检验; 自相关系数; 偏枯

中图分类号: TV213.4 **文献标志码:** A **文章编号:** 1672-1683(2015)04-0618-04

Application of the weighted Markov chain model in the inflow prediction of the Miyun Reservoir

HE Juan, WANG Xiaosong, WANG Caiyun

(Department of Hydraulics, China Institute of Water Resources and Hydropower Research, Beijing 100038, China)

Abstract: According to the actual inflow data of the Miyun Reservoir from 1960 to 2011, river runoff was selected as the random variable, and the related concept of the weighted Markov chain model and the steps for the inflow prediction in the incoming one year were introduced. The classification method of average standard was used to divide the inflow sequence into four conditions, including drought, lean drought, lean wet, and wet. The autocorrelation was regarded as weight coefficient to predict inflow between 2010 and 2011, which were compared with the measured data. The results showed that the weighted Markov chain model can predict inflow of the Miyun Reservoir with high precision. Therefore, the model was used to predict inflow between 2012 and 2013. Finally, the ergodicity and stationary distribution of Markov chain were analyzed, and the return periods of observed sequence under the wet and dry conditions were calculated, which suggested that the occurrence probability of lean drought is the largest. The inflow of the Miyun Reservoir was predicted to be lean drought in the future.

Key words: weighted Markov chain model; Miyun Reservoir; reservoir inflow; transition probability matrix; Markov property testing; autocorrelation coefficient; lean drought

密云水库是京津唐地区第一大水库,为北京最重要的地表饮用水源。但是,近年来,在“自然-人工”二元因素共同作用的影响下,其入库流量呈减少态势,1999年以后减少的趋势尤为明显^[1]。因此,对水库未来的来水量和变化趋势做出准确的预测关系到水库的调度应用及北京市的用水安全。夏乐天^[2-3]等人研究了加权马尔可夫链在降水状况预测中的应用和马尔可夫链预测方法的统计试验研究;王永兵^[4]等人利用马尔可夫链对水库入库径流状态进行了预测;冯利华,陈雄^[5]利用马尔

可夫链研究了区域干旱的变化趋势;冯耀龙^[6]等利用马尔可夫链对河流丰枯状况进行了预测;杨国范,刘冰等^[7]利用加权马尔可夫链对河流水质进行了预测;冯小明,刘桂清等^[8]对灌溉用水量进行了预测;Zekai Sen等^[9]对洪水来流量的预测也采用了马尔可夫链。目前的预测结果表明,加权马尔可夫链的预测精度较高,且在物理成因上也较为合理^[10]。本文采用加权马尔可夫链模型对密云水库的来水趋势进行预测,预测趋势跟实测资料吻合较好。

收稿日期: 2014-11-14 修回日期: 2015-04-30 网络出版时间: 2015-07-23

网络出版地址: <http://www.cnki.net/kcms/detail/13.1334.TV.20150723.1120.004.html>

基金项目: 北京市科技计划课题(Z141100006014049); 国家科技重大专项课题(2012ZX07205-005)

作者简介: 贺娟(1988),女,陕西延安人,主要从事城市暴雨及洪水模拟方面研究。E-mail: 18910232716@163.com

通讯作者: 王晓松(1962),男,河北人,教授级高级工程师,博士,主要从事高坝泄洪消能、河流生态修复与治理方面研究。E-mail: wangxs@iwhr.com

1 马尔可夫链的介绍

1.1 马尔可夫链的定义及分类

设有随机过程 $\{X_n, n \in T\}$ 若对于任意整数 $n \in T$ 和任意 $i_0, i_1, i_2, \dots, i_{n-1} \in I$ 条件概率满足

$$P\{X_{n+1} = i_{n+1} | X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\} = P\{X_{n+1} = i_{n+1} | X_n = i_n\} \quad (1)$$

则称 $\{X_n, n \in T\}$ 为马尔可夫链,简称马氏链^[11-12]。式中: i_n 表示随机过程 X_n 在 n 时刻的状态, P 表示事件发生的概率。

马尔可夫链预测方法可分为3种^[2]:基于绝对分布的马尔可夫链预测方法、叠加马尔可夫链预测方法和加权马尔可夫链预测方法。其中预测精度最高的是加权马尔可夫链预测方法,最低的是基于绝对分布的马尔可夫链预测方法。

1.2 转移概率矩阵^[11]

一步转移概率为

$$P_{ij}(n, n+1) = P(X_{n+1} = j | X_n = i) \quad (2)$$

K 步转移概率表示为

$$P_{ij}(n, n+k) = P(X_{n+k} = j | X_n = i) \quad (3)$$

所有转移概率构成的转移概率矩阵具有的性质如下:

$$0 \leq P_{ij} \leq 1; \sum_{j=1}^m P_{ij} = 1。$$

若转移概率与时刻 n 无关,则称为齐次马尔可夫链。用 $P^{(1)}$ 表示一步转移概率矩阵, $P^{(k)}$ 表示 k 步转移概率矩阵,则有:

$$P^{(k)} = (P^{(1)})^k \quad (4)$$

应用上遇到的马氏链一般不满足“时齐”条件,因此仅讨论一步转移概率^[10]。

1.3 加权马尔可夫链的预测步骤^[3-8,10]

(1) 计算均值 \bar{x} 。

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x(t) \quad (5)$$

(2) 计算标准差 s 。

$$s = \sqrt{\frac{1}{n-1} \sum_{t=1}^n (x(t) - \bar{x})^2} \quad (6)$$

(3) 指标值的样本均值 标准差分级。

根据均值及标准差,一般将实测序列分为5个状态,分级标准见表1。

(4) 转移概率矩阵的估算。

(5) 马氏性检验。

表1 分级标准

Tab. 1 Classification standard

状态	级别	分级标准
1	枯	$[0, x - s]$
2	偏枯	$[x - s, x - 0.5s]$
3	平	$[x - 0.5s, x + 0.5s]$
4	偏丰	$[x + 0.5s, x + s]$
5	丰	$[x + s, + \infty]$

可用 χ^2 统计量来检验离散序列的马氏链,设研究序列包含 m 个可能的状态,用 $f_{ij}(i, j \in E)$ 记为转移频数 概率矩阵,边际概率 $P_{.j}$ 的计算公式如下。

$$P_{.j} = \frac{\sum_{i=1}^m f_{ij}}{\sum_{i=1}^m \sum_{j=1}^m f_{ij}} \quad (7)$$

用 $P_{ij}(i, j \in E)$ 表示转移概率矩阵元素。当 m 较大时, χ^2 统计量如下:

$$\chi^2 = 2 \sum_{i=1}^m \sum_{j=1}^m f_{ij} |\ln(P_{ij}/P_{.j})| \quad (8)$$

服从自由度为 $(m-1)^2$ 的 χ^2 分布。给定置信度 α ,查表可 $\chi_{\alpha}^2((m-1)^2)$ 的值, $\chi^2 > \chi_{\alpha}^2((m-1)^2)$,则零假设被拒绝,即认为该序列具有“马氏性”。

(6) 各阶自相关系数 $r_k, (k \in E)$ 计算。

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2} \quad (9)$$

式中: x_t 为第 t 时段的指标值; \bar{x} 为指标值均值; n 为指标值序列长度; r_k 为第 k 阶(滞时为 k 年)自相关系数。对各阶相关系数规范化如下:

$$w_k = |r_k| / \sum_{k=1}^n |r_k| \quad (10)$$

式中: w_k 为各种滞时的马尔可夫链的权重。

(7) 预测。

将同一状态的各预测概率加权和作为处于该状态的预测概率,即

$$P_i = \sum_{k=1}^m w_k P_i^k \quad (11)$$

$\text{Max}\{P_i, i \in E\}$ 对应的状态为该时段的预测状态,确定该时段的指标值,并将其加入到原序列之中,再重复上述步骤,就可预测下时段的指标值状态。

2 工程实例

密云水库1960年-2009年入库流量资料^[13]见表2,利用加权马尔可夫链模型对2010年的入库流量进行预测。

表2 1960年-2009年入库流量序列及状态

Tab. 2 Inflow sequence and status from 1960 to 2011

年份	入库流量/($m^3 \cdot s^{-1}$)	状态	年份	入库流量/($m^3 \cdot s^{-1}$)	状态	年份	入库流量/($m^3 \cdot s^{-1}$)	状态	年份	入库流量/($m^3 \cdot s^{-1}$)	状态
1960	12.509	4	1973	24.192	4	1986	8.929	2	1999	2.401	1
1961	8.754	2	1974	23.443	4	1987	9.682	3	2000	1.735	1
1962	10.588	3	1975	8.727	2	1988	7.296	2	2001	5.346	1
1963	10.269	3	1976	11.975	3	1989	5.142	1	2002	1.976	1
1964	20.440	4	1977	13.887	4	1990	11.063	3	2003	2.875	1
1965	10.273	3	1978	14.099	4	1991	12.266	4	2004	4.697	1
1966	8.740	2	1979	15.992	4	1992	8.159	2	2005	5.133	1
1967	13.886	4	1980	6.038	1	1993	5.280	1	2006	4.329	1
1968	8.176	2	1981	5.551	1	1994	17.751	4	2007	2.408	1
1969	19.673	4	1982	13.811	4	1995	7.326	2	2008	5.033	1
1970	10.717	3	1983	5.161	1	1996	13.047	4	2009	2.105	1
1971	7.026	2	1984	3.231	1	1997	6.490	1			
1972	7.191	2	1985	6.822	2	1998	11.315	3			

2.1 均值、标准差状态分级计算

依据表 2, 得到该序列的均值 $\bar{x} = 9.259$ 亿 m^3 , 标准差 $s = 5.443$, 由于样本序列不多, 所以将其划分为枯、偏枯、偏丰、丰 4 个状态^[8,14], 径流状态分级标准见表 3, 序列状态见表 2。

表 3 径流状态分级标准

Tab. 3 Classification standard of runoff

状态	级别	分级标准	数值区间
1	枯	$[0, x - 0.5s]$	$[0, 6.538]$
2	偏枯	$[x - 0.5s, x]$	$[6.538, 9.259]$
3	偏丰	$[x, x + 0.5s]$	$[9.259, 11.981]$
4	丰	$[x + 0.5s, +\infty]$	$[11.981, +\infty]$

2.2 建立转移概率矩阵

对步骤 1 得到的结果进行统计, 可得不同滞时的转移概率矩阵如下:

步长为 1 的一步转移概率矩阵

$$P^{(1)} = \begin{bmatrix} 12/17 & 1/17 & 2/17 & 2/17 \\ 2/11 & 2/11 & 3/11 & 4/11 \\ 1/8 & 3/8 & 1/8 & 3/8 \\ 3/13 & 5/13 & 2/13 & 3/13 \end{bmatrix}$$

步长为 2 的一步转移概率矩阵

$$P^{(2)} = \begin{bmatrix} 11/16 & 3/16 & 0 & 2/16 \\ 1/11 & 2/11 & 4/11 & 4/11 \\ 2/8 & 2/8 & 1/8 & 3/8 \\ 4/13 & 3/13 & 3/13 & 3/13 \end{bmatrix}$$

步长为 3 的一步转移概率矩阵

$$P^{(3)} = \begin{bmatrix} 11/15 & 2/15 & 1/15 & 1/15 \\ 1/11 & 4/11 & 1/11 & 5/11 \\ 2/8 & 2/8 & 2/8 & 2/8 \\ 4/13 & 3/13 & 2/13 & 2/13 \end{bmatrix}$$

步长为 4 的一步转移概率矩阵

$$P^{(4)} = \begin{bmatrix} 11/14 & 2/14 & 1/14 & 0 \\ 2/11 & 3/11 & 4/11 & 2/11 \\ 2/8 & 1/8 & 0 & 5/8 \\ 3/13 & 4/13 & 1/13 & 5/13 \end{bmatrix}$$

2.3 马氏性检验

对步长为 1 的一步转移频数矩阵及转移概率矩阵进行马氏性检验, 计算结果见表 4 及表 5。

表 4 边际概率

Tab. 4 The marginal probability

状态	1	2	3	4
$P_{.j}$	0.367	0.224	0.163	0.245

表 5 统计量 $\sum_{i=1}^m \sum_{j=1}^m f_{ij} |\ln(P_{i,j}/P_{.j})|$ 计算

Tab. 5 Calculation of $\sum_{i=1}^m \sum_{j=1}^m f_{ij} |\ln(P_{i,j}/P_{.j})|$

状态	$f_{i1} \ln \frac{P_{i1}}{P_{.1}} $	$f_{i2} \ln \frac{P_{i2}}{P_{.2}} $	$f_{i3} \ln \frac{P_{i3}}{P_{.3}} $	$f_{i4} \ln \frac{P_{i4}}{P_{.4}} $	合计
1	7.838	1.339	0.655	1.466	11.299
2	1.407	0.422	1.539	1.581	4.949
3	1.078	1.539	0.267	1.278	4.163
4	1.395	2.692	0.119	0.178	4.384
合计	11.717	5.992	2.581	4.504	24.794

χ^2 的值为 49.588, 选取显著性水平 $\alpha = 0.05$, 查表可得 $\chi^2_{\alpha}(4-1)^2 = 16.919$, $\chi^2 > \chi^2_{\alpha}((m-1)^2)$, 可以验证该序列具有马氏性。

2.4 自相关系数及权重计算

由式(9)、式(10)可得各阶自相关系数及各种滞时的权重, 见表 6。

表 6 各阶自相关系数和各种滞时的权重

Tab. 6 The weights of Markov chain of various steps and various autocorrelation coefficients

k	1	2	3	4
r_k	0.337	0.204	0.274	0.418
w_k	0.273	0.165	0.222	0.339

2.5 预测

依据 2006 年-2009 年的入库流量及所对应的状态转移矩阵对 2010 年径流量状态进行预测, 结果见表 7。

表 7 2010 年入库流量状态预测

Tab. 7 Inflow status prediction in 2010

初始年	状态	滞时 /a	权重	状态				概率来源
				1	2	3	4	
2009	1	1	0.273	0.706	0.059	0.118	0.118	$P^{(1)}$
2008	1	2	0.165	0.688	0.188	0.000	0.125	$P^{(2)}$
2007	1	3	0.222	0.733	0.133	0.067	0.067	$P^{(3)}$
2006	1	4	0.339	0.786	0.143	0.071	0.000	$P^{(4)}$
P_i 权重和				0.736	0.125	0.071	0.068	

由表 7 可知, $\text{Max}\{P_i, i \in E\} = 0.736$, 所对应的 $i = 1$, 处于区间 $[0, 6.538]$, 即 2010 年入库流量状态为枯水年状态。2010 年的实测资料是 3.196 亿 $m^3 \in [0, 6.538]$, 与实测结果相吻合。同理, 用 1960 年-2010 年的资料预测出 2011 年的水库入流状态处于区间 $[0, 6.538]$, 实测值为 4.123 亿 m^3 , 预测结果较为合理。最后预测出 2012 年和 2013 年的入库流量分别为 5.102 亿 m^3 、4.380 亿 m^3 , 处于枯水年状态。

2.6 马尔可夫链的特征^[10,14,15]分析

具有遍历性的马尔可夫链, 当转移的步数 n 足够大时, 从系统的任何一个状态 i 转移到状态 j 的概率都近似等于 $\pi(j)$ 。通过分析得出此链是遍历的(不可约、非周期、正常反链), 因此, 此链存在唯一的平稳分布即为它的极限分布。设此链的平稳分布为 $\{\pi_j, j \in E\}$, 平稳分布、极限分布与各状态的重现期的计算公式如下:

$$\sum_{j \in E} \pi_j = 1; \pi_j = \sum_{i \in E} \pi_i P_{ij} \quad (12)$$

利用密云水库 1960 年-2009 年的资料, 以相依性较强的步长为 4 的马氏的特征分析为例。计算所得的值见表 8。

表 8 平稳分布、极限分布与各状态重现期

Tab. 8 Stationary distribution, limiting distribution, and return periods of various states

状态(j)	1	2	3	4
π_j	0.5022	0.1959	0.1210	0.1808
$T_j = u_j$	1.991	5.105	8.264	5.531

$u_j (\pi_j = 1/u_j)$ 表示系统从状态 j 出发, 首次返回状态 j 的平均时间, 同时也是各状态的极限分布。由表 8 可知, 各

状态的重现期为 $T_1 = 1.991$ 年; $T_2 = 5.105$ 年; $T_3 = 8.264$ 年; $T_4 = 5.531$ 年。根据现有的实测资料,由本文确定的分级标准,枯水年出现的次数最多,偏枯年出现的次数次之,这两种状态出现的概率为 0.698 1,从而可以说明密云水库来水长期处于枯水和偏枯状态的可能性比较大。

3 结语

根据马尔可夫链的预测理论,马尔可夫链针对的是一组离散的数据序列,其最基本的特征是:“马氏性”,也称“无后效性”。预测结果为入库流量的某一个状态,是一个区间,而不是一个具体的数值,在满足工程要求的前提下,可以较准确的预测出入库流量的变化趋势。通过对 2010 年-2013 年入库流量的预测,可以看出马尔可夫链模型能较好的预测来水状况。根据马尔可夫链的遍历性和平稳分布,由实测入库流量序列中出现枯、偏枯状态的年份比较多和来水呈逐渐下降的趋势预测出密云水库未来来水处于比较短缺的状态。这对水库管理者和流域规划者具有一定的参考价值。

参考文献(References):

[1] 高迎春,姚治君,刘宝勤,等.密云水库入库径流变化趋势及动因分析[J].地理科学进展,2002,21(6):546-553.(GAO Yingchun, YAO Zhijun, LIU Baogin, et al. Evolution trend of miyun reservoir inflow and its motivating factors analysis[J]. Progress in Geography, 2002, 21(6): 546-553. (in Chinese))

[2] 夏乐天,朱元牲,沈永梅.加权马尔可夫链在降水状况预测中的应用[J].水利水电科技进展,2006,26(6):20-24.(XIA Letian, ZHU Yuan sheng, SHEN Yong mei. Application of weighted Markov Chain to prediction of precipitation[J]. Advances in Science And Technology of Water Resources, 2006, 26(6): 20-24. (in Chinese))

[3] 夏乐天,朱元牲.马尔可夫链预测方法的统计试验研究[J].水利学报,2007(增刊):372-378.(XIA Letian, ZHU Yuan sheng. Study on statistical experiments of Markov Chain prediction methods[J]. Journal of Hydraulic Engineering, 2007 (supplement): 372-378. (in Chinese))

[4] 王永兵,胡小梅,彭丹芬,等.马尔可夫链在水库入库径流状态预测中的应用[J].水电与新能源,2011(4):18-21.(WANG Yongbing, HU Xiaomei, PENG Danfen, et al. The application of Markov Chain to forecasting reservoir inflow state[J]. Hydropower and Energy, 2011(4): 18-21. (in Chinese))

[5] 冯利华,陈雄.浙江干旱的变化趋势[J].农业系统科学与综合研究,2001(3):177-179.(FENG Lihua, CHEN Xiong. Change tendency of drought in zhejiang province[J]. System Sciences and Comprehensive studies in Agriculture, 2001(3): 177-179. (in Chinese))

[6] 冯耀龙,韩文秀.加权马尔可夫链在河流丰枯状况预测中的应用[J].系统工程理论与实践,1999(10):89-93.(FENG Yaolong, HAN Wenxiu. The application of weighted Markov Chain to the prediction of river runoff state[J]. Systems Engineering theory and Practice, 1999(10): 89-93. (in Chinese))

[7] 杨国范,刘冰,金鑫,等.加权马尔可夫链在河流水质预测中的应用[J].节水灌溉,2008,(6):16-18.(YANG Guofan, LIU Bing, JIN Xin et al. Application of Weighted Markov Chain method in water quality forecast[J]. Water Saving Irrigation, 2008, (6): 16-18. (in Chinese))

[8] 冯小明,刘桂清,周莉,等.马尔可夫链在灌溉用水量预测中的应用[J].水利与建筑工程学报,2011,9(2):98-101.(FENG Xiaoming, LIU Guqing, ZHOU Li, et al. Application of Markov Chain in prediction for irrigation water consumption[J]. Journal of Water Resources and Architectural Engineering, 2011, 9(2): 98-101. (in Chinese))

[9] Zekai Sen. Critical Drought Analysis by Second Order Markov Chain[J]. Journal of Hydrology, 1990, 120(1-4): 183-202.

[10] 夏乐天.马尔可夫链预测方法及其在水文序列中的应用[D].南京:河海大学,2005.(XIA Letian. Research of Markov Chain prediction method and its application on hydrology series[D]. Nanjing: Hohai University, 2005. (in Chinese))

[11] 鲁帆,严登华,王勇,等.中长期径流预报技术与方法[M].北京:中国水利水电出版社,2012.(LU Fan, YAN Denghua, WANG Yong, et al. Medium and long term runoff forecasting techniques and approaches[M]. Beijing: China Water Power Press, 2012. (in Chinese))

[12] Daniel P, Loucks and Eelco van Bee. Water Resources Systems Planning and Management[M]. The United Nations Educational, Scientific and Cultural Organization, 2005.

[13] 段新光,郝丽娟,栾芳芳.密云水库流域降水量与径流量特征分析[J].北京水务,2013(1):38-41.(DUAN Xirguang, HAO Lirjuan, LU AN Fangfang. Analysis of the characteristics of the rainfall and runoff in the basin of the Miyun reservoir[J]. Beijing Water, 2013(1): 38-41. (in Chinese))

[14] 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.(XIA Letian. Prediction of plum rain intensity based on index Weighted Markov Chain[J]. Journal of Hydraulic Engineering, 2005, 36(8): 988-993. (in Chinese))

[15] 潘刚,芦冰,邹兵,等.马尔可夫链在水库主汛期降雨状态预测中的应用[J].水利科技与经济,2011,17(6):33-36.(PAN Gang, LU Bing, ZOU Bing, et al. Markov Chain state in the reservoir the main flood season rainfall forecast[J]. Water Conservancy Science and Technology and Economy, 2011, 17(6): 33-36. (in Chinese))