# 高关水库入库径流状态预测

胡小梅[1], 陈向东[2], 黄云章[1], 丁俊芝[1]
(1. 湖北省漳河工程管理局, 湖北 荆门 448156; 2. 湖北省高关水库管理局, 湖北 荆门 431806)

**摘　要**: 依据高关水库1971－2010年入库径流量资料,应用均值标准差法建立5级分级标准。针对径流量为相依随机变量的特点,以各阶自相关系数为权重,运用马尔可夫链模型预测未来一年的入库径流量状态。结果表明:该方法直观、预测准确、计算简便,为区域径流量的中长期预测提供了新的分析途径。

**关键词**: 自相关系数; 马尔可夫链; 径流量变化预测; 高关水库流域
**中图分类号**:P332.4　　**文献标识码**: A　　**文章编号**: 1672-643X(2012)05-0123-04

# Forcast of runoff state inflow Gaoguan Reservoir

**HU Xiaomei**[1], **CHEN Xiangdong**[2], **HUANG Yunzhang**[1], **DING Junzhi**[1]
(1. *Zhanghe Project Administration Bureau*, *Jingmen* 448156, *China*;
2. *Gaoguan Reservoir Bureau*, *Jingmen* 431806, *China*)

**Abstract**: According to runoff stroage data from 1971 to 2010, five level standards were established by using mean and standard deviation method. Appionted at the runoff characteristics of dependent random variables, taking all orders autocorrelation coefficients as weights, Markov chain model was used to predict the runoff state in the next year. The results show that the method can provide a new feasible way for regional and long－term runoff prediction, and is direct, accurate and simple.

**Key words**: auto correlation coefficient; Markov chain; forecast of runoff change; Gaoguan reservoir watershed

## 0　引　言

高关水库位于湖北省京山县北部的大富水河上游,水库流域面积303 km$^2$,总库容2.12亿m$^3$,其中兴利库容1.85亿m$^3$,设计灌溉面积2.56万hm$^2$,是汉北河水利规划建设中的一座以灌溉为主,兼有防洪、发电、养殖、旅游等综合效益的大(2)型水利工程。从运用实践来看,水库多年平均缺水量为0.56亿m$^3$,倘若遭遇干旱或特大干旱年,则需水缺口更大。因此,水库每年的来水量和来水量变化趋势是水库调度人员最为关心的问题,它不仅直接关系到能否满足当年水库的兴利运用需要,而且也关系到水库调度运用安全与否。对于多年调节运用的水库,还关系到下一年水库的调度运用问题[1]。径流过程是一随机变量,存在许多不确定因素,且错综复杂[2]。到目前为止,还难以用物理成因分析来确定出未来某一时段(如年、季、月等)流域径流量的准确数值。但实际调度运用中,在有些情况下,仅预测出未来某时段径流量适当的变化区间(即丰枯状况)即可[3]。基于此,用加权的马尔可夫链对未来流域径流丰枯变化趋势进行分析,用以指导水库的调度运用。

## 1　马尔可夫链预测的方法

马尔可夫过程是随机过程的一个分支,它的最基本特征是"无后效性",即在已知某一随机过程"现在"的条件下,其"将来"与"过去"是独立的,它是一个时间离散、状态离散的时间序列,其数学表达如下[4-6]:

定义在概率空间($\Omega$、$F$、$P$)上的随机序列$\{\chi_t \in T\}$,其中参数集$T=\{0,1,2,\cdots\}$,状态空间$E=\{0,1,2,\cdots\}$,称为马尔可夫链,如果对任意正整数$l$、$m$、$k$及任意的非负整数有$j_l > \cdots > j_2 > j_1 (m > j_1)$, $i_{m+k}$, $i_m$, $i_{jl}$, $\cdots$, $i_{j2}$, $i_{j1} \in E$有:

$$P\{\chi_{m+k}=i_{m+k} | \chi_m = i_m, \chi_{jl}=i_{jl}, \cdots, \chi_{j2}=i_{j2}, \chi_{j1}=i_{j1}\} = P\{\chi_{m+k}=i_{m+k} | \chi_m = i_m\} \tag{1}$$

这里要求(1)式的左端有意义,即假定

$$P\{\chi_{m+k}=i_m, \chi_{jl}=i_{jl}, \cdots, \chi_{j2}, \chi_{j1}=i_{j1}\} > 0$$

**收稿日期**:2012-05-11;　**修回日期**:2012-06-10
**作者简介**:胡小梅(1980-), 女, 湖北仙桃市人, 工程师, 研究方向:水利工程管理。

实际应用中，常记(1)式右端

$$P\{\chi_{m+k}=i_{m+k}|\chi_m=i_m=P\{\chi_{m+k}=j\mid\chi_m=i\}$$
$$=P_{ij}\{m;k\};\ i,j,\ \in E$$

一般考虑齐次马尔可夫链，即对任意的 $m,k\in T$ 有：

$$P_{ij}(m;k)=P_{ij}(k)\qquad i,j\in E \tag{2}$$

式中：$P_{ij}\{m;k\}$ 为系统在 $m$ 时刻处在状态 $i$，经 $k$ 步转移至状态 $j$ 的概率；$P_{ij}(k)$ 为系统从状态 $i$，经 $k$ 步转移至状态 $j$ 的概率，此时转移概率与初始时刻无关，$k$ 取 1 时，$P_{ij}(1)$ 记为 $P_{ij}$。

齐次马尔可夫链完全由其初始分布 $\{P_0(i_0),i_0\in E\}$ 及其一步状态转移概率矩阵 $P=(P_{ij}),i,j\in E$ 所决定。若已知时刻 $n$ 时的绝对分布 $P(n)=\{P_n(j),j\in E\}$，则时刻 $n+1$ 的绝对分布为：

$$P(n+1)=P(n),\ P=\sum_{i\in E}^{n}P_n(i)P_{ij} \tag{3}$$

## 2 马尔可夫链预测的步骤[4-10]

(1) 计算均值 $\bar{\chi}$。设时间序列 $\{\chi(t)\}$，$t=1,2,\cdots n$，定义均值序列：

$$\bar{\chi}=\frac{1}{n}\sum_{t=1}^{n}\chi(t) \tag{4}$$

(2) 计算标准差 $s$ 。

$$s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\chi_i-\bar{\chi})^2} \tag{5}$$

(3) 分组。一般将数据序列分为5组：$(-\infty,\bar{\chi}-s)$、$(\bar{\chi}-s,\bar{\chi}-0.5s)$、$(\bar{\chi}-0.5s,\bar{\chi}+0.5s)$、$(\bar{\chi}+0.5s,\bar{\chi}+s)$、$(\bar{\chi}+s,+\infty)$。利用这种方法对指标值分类，不考虑物理成因对指标值的影响，仅仅是在统计的角度上，简单地把样本均值作为指标值的中心，此方法操作较为方便，因此得到了广泛地应用。

(4) 转移概率矩阵的估算。

设 $\chi_1,\chi_2,\cdots,\chi_n$ 是马尔可夫链的一个指标值序列，它包含 $m$ 个状态，即状态空间 $E=\{1,2,\cdots,m\}$，用 $f_{ij}$ 表示指标值序列中从状态 $i$ 经过 $n$ 步转移到达状态 $j$ 的频数，$i,j\in E$，由 $f_{ij}(i,j\in E)$ 组成的矩阵 $(f_{i,j})$，$i,j\in E$(称为转移频数矩阵。将转移频数矩阵的第 $i$ 行第 $j$ 列元素 $f_{i,j}$ 除以各行的总和所得的值称为转移概率，记为 $f_{i,j},i,j\in E$ 即：

$$P_{i,j}=f_{ij}/\sum_{j=1}^{m}f_{ij} \tag{6}$$

(5)"马氏性" 检验。离散序列的马氏链可用 $\chi^2$ 统计量来检验。设所研究的序列包含 $m$ 个可能的状态，用 $(f_{i,j})$，$i,j\in E$ 记为转移频数概率矩阵，将转移频数矩阵的各列之和除以各行各列的总和，所得到的值称为"边际概率" 记为 $P_{\bullet j}$。

$$P_{\bullet j}=\sum_{i=1}^{m}f_{ij}/\sum_{i=1}^{m}\sum_{i=1}^{m}f_{ij} \tag{7}$$

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

$$\chi^2=2\sum_{i=1}^{m}\sum_{j=1}^{m}f_{ij}\left|\ln\frac{P_{ij}}{P_{\bullet j}}\right| \tag{8}$$

服从自由度为 $(m-1)^2$ 的 $\chi^2$ 分布，其中 $P_{ij}$ 为转移概率，由式(6) 定义。给定置信度 $a$，查表可得 $\chi_a^2((m-1)^2)$ 的值，如 $\chi^2>\chi_a^2((m-1)^2)$，则零假设被拒绝，即认为该序列具备"马氏性"。

(6) 计算各阶自相关系数 $r_k$。

$$r_k=\sum_{t=1}^{n-k}(\chi_t-\bar{\chi})(\chi_{t+k}-\bar{\chi})/\sum_{t=1}^{n}(\chi_t-\bar{\chi})^2$$
$$(k\in E) \tag{9}$$

各种步长的马尔科夫链的权重按照下式计算：

$$\omega_k=|r_k|/\sum_{k=1}^{m}|r_k| \tag{10}$$

(7) 预测。将同一状态的各预测概率加权和作为指标值处于该状态的预测概率，即：

$$P_i=\sum_{k=1}^{m}\omega_kP_i^k\qquad(i\in E) \tag{11}$$

$\max\{P_i,i\in E\}$ 所对应的 $i$ 即为该时段指标值的预测状态。待该时段的指标值确定之后，将其加入到原序列之中，再重复步骤(1)～(7)，即可进行下时段指标值状态的预测。

(8)可进一步对该马尔可夫链的特征(遍历性、平稳分布等)进行分析。

## 3 实例分析

利用高关水库 1971－2008 年共 38 年的入库径流量资料(见表 1，表 1 中径流量单位为万 m³)，按上述方法对 2009、2010 年入库径流量进行预测，以说明该方法的预测效果。

(1)均值与标准差计算。依据表 1 中 1971－2008 年的数据，得到该径流序列的均值 $\bar{\chi}=10\ 643$ 万 m³，径流序列标准差 $s=5\ 577$。划分为枯、偏枯、平、偏丰、丰等 5 个状态，径流状态分级标准见表 2，序列状态见表 1。

(2)建立矩阵。经计算，得到各种步长的状态转移概率矩阵 P。

表 1　高关水库入库径流量序列及状态　　万 $m^3$

| 年份 | 1971 | 1972 | 1973 | 1974 | 1975 | 1976 | 1977 | 1978 | 1979 | 1980 | 1981 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 径流量 | 7692 | 7081 | 17615 | 8200 | 11700 | 3752 | 9721 | 3129 | 7005 | 20278 | 4696 |
| 状态 | 2 | 2 | 5 | 3 | 3 | 1 | 3 | 1 | 2 | 5 | 1 |
| 年份 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 |
| 径流量 | 11126 | 16320 | 8282 | 7076 | 4473 | 9792 | 6259 | 10677 | 9866 | 13322 | 5550 |
| 状态 | 3 | 5 | 3 | 2 | 1 | 3 | 2 | 3 | 3 | 3 | 2 |
| 年份 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 |
| 径流量 | 7647 | 7158 | 11208 | 19971 | 12503 | 27514 | 5445 | 9762 | 6982 | 12910 | 13440 |
| 状态 | 2 | 2 | 3 | 5 | 3 | 5 | 2 | 3 | 2 | 3 | 4 |
| 年份 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | | | | |
| 径流量 | 10441 | 8728 | 7312 | 24795 | 15030 | 8371 | 9509 | | | | |
| 状态 | 3 | 3 | 2 | 5 | 4 | 3 | 3 | | | | |

表 2　入库径流量分级

| 状态 | 级别 | 分级标准 | 数值区间 |
|---|---|---|---|
| 1 | 枯 | $[0, \bar{x}-s]$ | [0,5067] |
| 2 | 偏枯 | $(\bar{x}-s, \bar{x}-0.5s]$ | (5067,7855] |
| 3 | 平 | $(\bar{x}-0.5s, \bar{x}+0.5s]$ | (7855, 13432] |
| 4 | 偏丰 | $(\bar{x}+0.5s, \bar{x}+s]$ | (13432,16221] |
| 5 | 丰 | $(\bar{x}+s, +\infty]$ | (16221, +∞] |

$$P^{(1)}=\begin{bmatrix}0.000 & 0.250 & 0.750 & 0.000 & 0.000\\0.091 & 0.273 & 0.364 & 0.000 & 0.272\\0.200 & 0.267 & 0.267 & 0.066 & 0.200\\0.000 & 0.000 & 1.000 & 0.000 & 0.000\\0.166 & 0.167 & 0.500 & 0.167 & 0.000\end{bmatrix}$$

$$P^{(2)}=\begin{bmatrix}0.173 & 0.268 & 0.291 & 0.050 & 0.218\\0.143 & 0.240 & 0.401 & 0.070 & 0.146\\0.111 & 0.227 & 0.485 & 0.051 & 0.126\\0.200 & 0.267 & 0.267 & 0.066 & 0.200\\0.115 & 0.220 & 0.486 & 0.033 & 0.146\end{bmatrix}$$

$$P^{(3)}=\begin{bmatrix}0.119 & 0.230 & 0.464 & 0.056 & 0.131\\0.127 & 0.232 & 0.444 & 0.052 & 0.145\\0.139 & 0.240 & 0.409 & 0.053 & 0.159\\0.111 & 0.227 & 0.485 & 0.051 & 0.126\\0.141 & 0.243 & 0.402 & 0.057 & 0.157\end{bmatrix}$$

$$P^{(4)}=\begin{bmatrix}0.136 & 0.238 & 0.418 & 0.053 & 0.155\\0.134 & 0.238 & 0.422 & 0.054 & 0.152\\0.130 & 0.236 & 0.433 & 0.054 & 0.147\\0.139 & 0.240 & 0.409 & 0.053 & 0.159\\0.129 & 0.235 & 0.436 & 0.053 & 0.147\end{bmatrix}$$

$$P^{(5)}=\begin{bmatrix}0.131 & 0.236 & 0.430 & 0.054 & 0.149\\0.131 & 0.236 & 0.430 & 0.054 & 0.149\\0.133 & 0.237 & 0.426 & 0.053 & 0.151\\0.130 & 0.236 & 0.433 & 0.054 & 0.147\\0.133 & 0.237 & 0.425 & 0.054 & 0.151\end{bmatrix}$$

(3) 随机变量序列得马氏检验。对步长为 1 的一步转移概率矩阵 $P^{(1)}$ 进行统计分析,$\chi^2 = 38.99$,给定显著性水平 $a = 0.05$,查表得分位点 $\chi_a^2(16) = 26.296$。由于 $\chi^2 > \chi_a^2((m-1)^2)$,故漳河水库流域径流量序列满足马氏性。

(4) 自相关系数计算。系列各阶自相关系数分别为:$r_1 = -0.0573$、$r_2 = 0.0753$、$r_3 = 0.0323$、$r_4 = -0.0637$、$r_5 = -0.0066$。将各阶自相关系数归一化后作为各种滞时的马尔可夫链权重,分别为:$\omega_1 = 0.2437$、$\omega_2 = 0.3202$、$\omega_3 = 0.1373$、$\omega_4 = 0.2709$、$\omega_5 = 0.0280$。

(5) 预测。依据 2004 - 2008 年径流量及其对应的状态转移矩阵对 2009 年径流量状态进行预测,结果见表 3。

表 3　2009 年径流量预测

| 初始年 | 状态 | 滞时/a | 权重 | 状态 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|---|
| 2008 | 4 | 1 | 0.2437 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 |
| 2007 | 5 | 2 | 0.3202 | 0.115 | 0.220 | 0.486 | 0.033 | 0.146 |
| 2006 | 2 | 3 | 0.1373 | 0.127 | 0.232 | 0.444 | 0.052 | 0.145 |
| 2005 | 3 | 4 | 0.2709 | 0.130 | 0.236 | 0.433 | 0.054 | 0.147 |
| 2004 | 3 | 5 | 0.0280 | 0.133 | 0.237 | 0.426 | 0.053 | 0.151 |
| $P_i$(权重和) | | | | 0.093 | 0.173 | 0.589 | 0.034 | 0.111 |

由表 3 可知,$\max\{P_i, i \in E\} = 0.589$,所对应的 $i = 3$,即 2009 年径流量为平水年状态。2009 年实测值为 8371 万 $m^3 \in (7855, 13432]$,与实际情况吻合。同理,以 2005 -2009 年的径流量资料预测 2010 年的径流量状态,重复步骤(1) ~ (7),(3)可以不做,预测结果见表 4。由表 4 可知,$\max\{P_i, i \in E\} = 0.339$,所对应的 $i = 3$,即 2010 年径流量为平水年状态。2010 年实测值为 9 509 万 $m^3$(7855, 13432],与实际情况吻合。说明用马尔可夫链方法对高关水库流域径流量丰枯状态进行预测是可行的。

表4 2010年径流量预测

| 初始年 | 状态 | 滞时/a | 权重 | 状态 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|---|
| 2009 | 3 | 1 | 0.3020 | 0.200 | 0.267 | 0.267 | 0.066 | 0.200 |
| 2008 | 4 | 2 | 0.2178 | 0.200 | 0.267 | 0.267 | 0.066 | 0.200 |
| 2007 | 5 | 3 | 0.1327 | 0.141 | 0.243 | 0.402 | 0.057 | 0.157 |
| 2006 | 2 | 4 | 0.3049 | 0.134 | 0.238 | 0.422 | 0.054 | 0.152 |
| 2005 | 3 | 5 | 0.0427 | 0.133 | 0.237 | 0.426 | 0.053 | 0.151 |
| $P_i$(权重和) | | | | 0.169 | 0.253 | 0.339 | 0.061 | 0.178 |

(6) 分析马尔可夫链的特征(遍历性、平稳分布等)。步长为3的马尔可夫链的相依性较强,此时的转移概率矩阵是$P^{(3)}$,由$P^{(3)}$所决定的马尔可夫链的5个状态是互通的,即对任意的$i,j \in E, i \leftrightarrow j, (i \neq j)$,且为非周期的,其全部状态构成的状态空间是一个闭集,因此,此链是不可约的,是一个正常返链,是一个遍历链。根据遍历性定理,可以求出此链的极限分布(此时极限分布即平稳分布)。极限分布求解方程为:

$$\sum_{j=1}^{5} \pi_j = 1, \quad \pi_j = \sum_{j=1}^{5} \pi_j P_{ij} \quad (j \in E) \tag{12}$$

由式(12)解得平稳分布(极限分布)与各状态的重现期,见表5。表中$\pi_j = 1/\mu_j$。

表5 平稳分布(极限分布)与各状态的重现期

| 状态 | $\pi_j$ | $\mu_j$ | 状态 | $\pi_j$ | $\mu_j$ |
|---|---|---|---|---|---|
| 1 | 0.1321 | 7.57 | 4 | 0.0535 | 18.68 |
| 2 | 0.2366 | 4.23 | 5 | 0.1501 | 6.66 |
| 3 | 0.4277 | 2.34 | | | |

状态$i$的重现期$T_i = \mu_i (i \in E)$,对应的概率$P_i = 1/T_i = \pi_i (i \in E)$。各状态的重现期分别是$T_1 = \mu_1 = 7.57a$、$T_2 = \mu_2 = 4.23a$、$T_3 = \mu_3 = 2.34a$、$T_4 = \mu_4 = 18.68a$、$T_5 = \mu_5 = 6.66a$。即在38 a中,正常年出现的机会最大,平均每隔2.34a出现1次,出现的概率为42.77%;偏丰年出现的概率最小,平均每隔18.68a就会出现1次,出现的概率为5.35%。

# 4 结 语

(1)利用马尔可夫链预测的结果为径流量的某一个状态,是一个区间值,而不是一个具体的数值,从而使预测的允许范围扩大,在一定程度上可以满足实际工作需要,说明运用该方法对径流量进行状态预测是有效可行的[5-9]。

(2)由于以各种步长的自相关系数为权重,用各种步长的马尔可夫链加权来预测径流量状态,所以与普通的马尔可夫链预测方法相比较,更能充分、合理地利用信息,是对建立马尔可夫链与相关分析有效结合的预测方法的进一步尝试[10]。

(3)随着预报对象序列的逐年增加,资料的代表性也日益增强,自相关系数、状态转移概率矩阵、权重将会发生某些变化,这种变化是预报理论模式不断完善的过程,预报方案不是固定不变的。因此,应将每年预报对象的新的实测值加入到资料系列中去,实现动态调整预报对象的自相关系数、状态转移概率矩阵、权重,以期进一步提高预报精度[8]。

(4)将马尔可夫链预测方法与序列的自相关分析[11]、多步长一步转移概率分析等有机地结合起来,物理概念清晰,计算简便,为提高中短期径流量预测的精度提供了一条值得探索的途径[12]。但如何根据最后计算出的状态概率分布求出径流量的具体值(在某些情况下,仍需要一个具体数值)仍是一个有待解决的问题。

**参考文献:**


[1] 陈崇德,牛爱军. R/S分析在水库年来水趋势预测中的应用[J]. 水资源与水工程学报. 2010,21(3):174-176.

[2] 陈崇德,邵春玲. 漳河水库区域水资源利用评价指标的研究应用[J]. 水利水电工程设计,2009,28(3):23-25.

[3] 张笑天,陈崇德. 漳河水库灌区水资源脆弱性评价研究[J]. 华北水利水电学院学报,2010,31(2):12-15.

[4] 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):1-8.

[5] 刘德地,陈晓宏. 一种北江流域年降雨量的权马尔可夫链预测模型[J]. 水文,2006,26(6):23-26.

[6] 安润秋,郝玉芹. 基于随机过程的中长期降水预测模型[J]. 唐山学院学报,2007,20(2):7-9.

[7] 孙天青,张 鑫,梁学玉. 基于加权马尔柯夫链的汛期降水丰枯预测[J]. 中国农村水利水电,2010(4):1-3.

[8] 崔锦龙,邓姝杰. 基于马尔可夫模型的降水预测及其利用[J]. 资源开发与市场,2008,24(2):115-117.

[9] 夏乐天,朱元生生,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):20-23.

[10] 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.

[11] 丁 晶,刘权授. 随机水文学[M]. 北京:中国水利水电出版社,1997.

[12] 宋松柏,蔡焕杰,粟晓玲. 专门水文学概论[M]. 杨凌:西北农林科技大学出版社,2005.