

# 秃尾河流域降水量权马尔可夫链模型预测研究

李小冰, 蔡焕杰, 张鑫, 王健

(西北农林科技大学旱区农业水土工程教育部重点实验室, 陕西 杨凌 712100)

**摘要:** 针对降水过程存在着大量不确定性、不精确性的特点, 采用均值-一标准差分级法将降水量分为雨涝、偏涝、正常、偏旱和干旱 5 个状态; 针对降水量为相依随机变量的特点, 采取规范化的各阶自相关系数为权重, 用加权马尔可夫链模型预测未来降水的变化状况; 最后以黄河中游秃尾河流域为例对该方法进行了检验, 分析了其未来情况下降水的可能变化趋势。结果表明, 2002 年秃尾河流域降水量预测状态与实际相吻合。

**关键词:** 降水量; 权; 马尔可夫链; 预测; 秃尾河

**中图分类号:** S161.6<sup>+1</sup> **文献标识码:** A **文章编号:** 1000-7601(2009)06-0252-05

大气降水是水资源的总补给源, 降水量的多少很大程度上反映了水资源量的大小, 直接影响水资源的开发和利用, 而且与干旱、洪涝等自然灾害的发生密切相关<sup>[1]</sup>。因此, 准确地预测降水量, 既可为制定水资源开发利用策略提供科学依据, 又可为防灾、抗灾、救灾提供有效的指导。降水是一种复杂的自然现象, 呈现出不确定性与随机性。马尔可夫链<sup>[6]</sup>是一种以概率论和随机过程理论为基础, 应用数学模型分析客观对象发展变化过程中的数理关系的一种统计方法。它通过把事物不同阶段联系起来分析其变化的规律, 实质属于动态分析法。我们把具有离散状态和时间序列的某一个过程可视为马尔可夫链。根据  $n$  时刻的状态就可预测  $n+1$  时刻的状态, 这就是应用马尔可夫模型解决各种预测问题的基本的思想。降水量是一个随时间变动的过程, 可视为一相依的随机变量序列, 其年际相依程度的强弱, 可采用自相关系数作为其定量的测度, 基于此可建立马尔可夫链模型进行预测分析。

## 1 马尔可夫链

马尔可夫链是状态和时间均离散的马尔可夫过程, 简称“马氏链”。它的最基本特征是“马氏性”, 也称“无后效性”, 即在系统“现在”的状态已知的条件下, 其“将来”的状态与“过去”的状态无关。马尔可夫链的数学表述如下<sup>[5]</sup>: 定义在概率空间  $(\Omega, F, P)$  上的随机序列  $\{X(t), t \in T\}$ , 其中参数集  $T = \{0, 1, 2, \dots\}$ , 状态空间  $E = \{0, 1, 2, \dots\}$ , 称为马尔可夫链, 如果对任意的正整数  $l, m, k$ , 及任意的非负整数  $j_l > \dots > j_2 > j_1 (m > j_l), i_{m+k}, i_m, i_{j_l}, \dots, i_{j_2}, i_{j_1} \in$

$E$  有式(1)成立(这里要求式(1)的左端有意义)。  
$$P\{X(m+k) = i_{m+k} | X(m) = i_m, X(j_l) = j_l, \dots, X(j_2) = j_2, X(j_1) = j_1\}$$
$$= P\{X(m+k) = i_{m+k} | X(m) = i_m\} \quad (1)$$

实际应用中, 常记式(1)的右端  
$$P\{X(m+k) = i_{m+k} | X(m) = i_m\} = P\{X(m+k) = j | X(m) = i\} = p_{ij}(m; k) (i, j \in E)$$

一般考虑齐次马尔可夫链, 即对任意的  $m, k \in T$  有

$$p_{ij}(m; k) = p_{ij}(k) \quad (i, j \in E) \quad (2)$$

式中,  $p_{ij}(m, k)$  为系统在时刻  $m$  时处在状态  $i$ 、经  $k$  步状态转移到达状态  $j$  的概率;  $p_{ij}(k)$  为系统从状态  $i$ 、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,  $k$  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  $n+1$  时的绝对分布

$$P(n+1) = P(n)P = P(0)P^n \quad (3)$$

## 2 权马尔可夫链预测模型

### 2.1 建模思路

权马尔可夫链预测的基本思路: 一系列相依的随机变量, 其各阶自相关系数刻画了各种滞时的状态间相关关系的强弱, 考虑先分别依其前面若干时段的指标值状态对该时段指标值的状态进行预测, 然后, 按前面各年与该年相依关系的强弱加权求和, 即可充分、合理地利用信息进行预测。

收稿日期: 2009-09-08

基金项目: 教育部高校青年教师奖资助项目; 国家自然科学基金资助项目(50479051)

作者简介: 李小冰(1986-), 女, 河南许昌人, 在读硕士, 主要从事水资源高效利用方面的研究。E-mail: lixiaobing2005@yahoo.com.cn。

通讯作者: 蔡焕杰(1962-), 男, 教授, 博导, 从事节水农业与水资源高效利用研究。E-mail: huanjie@yahoo.com。

## 2.2 权马尔可夫链预测的方法和步骤

基于以上思路,权马尔可夫链预测具体方法和步骤如下:

(1) 计算指标值序列均值  $\bar{x}$  和均方差  $S =$

$\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ , 建立指标值的分级标准(相当于确定马尔可夫链的状态空间),可根据资料序列的长短及具体问题的要求进行分级。本文采用样本均方差为标准确定马尔可夫链的状态空间  $E = \{1, 2, \dots, 5\}$ 。

(2) 按步骤(1)所建立的分级标准,确定资料序列中各时段指标值所对应的降水量状态。

(3) 对步骤(2)所得的结果进行统计,可得不同滞时(步长)马尔可夫链的转移概率矩阵,它决定了指标值状态转移过程的概率法则。将转移频数矩阵的第  $i$  行第  $j$  列元素  $f_{ij}$  除以各行的总和所得的值称为转移概率,记为  $p_{ij} (i, j \in E)$ , 即

$$p_{ij} = f_{ij} / \sum_{j=1}^m f_{ij} \quad (4)$$

(4) 进行“马氏性”检验。用  $f_{ij}$  表示指标值序列  $x_1, x_2, \dots, x_n$  中从状态  $i$  经过一步转移到达状态  $j$  的频数 ( $i, j \in E$ )。将转移频数矩阵的第  $j$  列之和除以各行各列的总和所得的值称为边际概率,记为  $p \cdot j$ , 即

$$p \cdot j = \sum_{i=1}^m f_{ij} / \sum_{i=1}^m \sum_{j=1}^m f_{ij} \quad (5)$$

当  $n$  充分大时,统计量

$$\chi^2 = 2 \sum_{i=1}^m \sum_{j=1}^m f_{ij} \left| \lg \frac{p_{ij}}{p \cdot j} \right| \quad (6)$$

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(4)定义。

给定显著性水平  $\alpha$ , 查表可得分位点  $\chi_{\alpha}^2((m-1)^2)$  的值,计算后得统计量  $\chi^2$  的值。若  $\chi^2 > \chi_{\alpha}^2((m-1)^2)$ , 则可以认为序列  $\{x_i\}$  符合马氏性,否则认为该序列不可作为马尔可夫链来处理。

(5) 计算各阶自相关系数  $r_k (k \in E)$

$$r_k = \frac{\sum_{l=1}^{n-k} (x_l - \bar{x})(x_{l+k} - \bar{x})}{\sum_{l=1}^n (x_l - \bar{x})^2} \quad (7)$$

式中,  $r_k$  为第  $k$  阶(滞时为  $k$  年的)自相关系数;  $x_l$  为第  $l$  时段的指标值;  $\bar{x}$  为指标值均值;  $n$  为指标值序列长度。

对各阶自相关系数规范化,即

$$w_k = |r_k| / \sum_{k=1}^m |r_k| \quad (8)$$

作为各种滞时(步长)的马尔可夫链的权重( $m$

为按预测需要计算到的最大阶数)。

(6) 分别以前面若干时段的价值结合其相应的各阶转移概率矩阵即可预测出该时段指标值的状态概率  $P_i^{(k)} (i \in E), k = 1, 2, \dots, m$ 。

(7) 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

$$P_i = \sum_{k=1}^m w_k P_i^{(k)} \quad (i \in E) \quad (9)$$

$\max\{P_i, i \in E\}$  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) 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 秃尾河流域降水预测

### 3.1 研究区概况

秃尾河流域位于黄土高原北部,地理坐标为位于东经  $109^{\circ}57' \sim 110^{\circ}31'$ , 北纬  $38^{\circ}31' \sim 39^{\circ}01'$  之间, 总面积  $3\,294 \text{ km}^2$ 。秃尾河流域地处中纬度大陆性季风气候区,流域内光热资源丰富、水热组合较好,但灾害性天气经常发生,旱灾是发生频率最高、对农业危害最大的自然灾害,“十年九旱、旱涝交错”是流域气候的主要特点。全流域年平均日照  $2\,871 \text{ h}$ , 平均气温  $8.5^{\circ}$ , 最低气温  $-28.1^{\circ}$ , 极端最高气温  $38.9^{\circ}$ , 平均无霜期  $175 \text{ d}$ , 多年平均降雨量为  $400$  多毫米。降水量自东南向北和西北递减。夏季流域降雨量较多,  $6 \sim 9$  月降水量占全年降水量的  $60\%$  以上。秃尾河流域处于半湿润气候与半干旱气候的过渡地带,夏季干燥炎热蒸发能力较强,多年平均水面蒸发量为  $1\,120 \text{ mm}$ 。

### 3.2 模型应用

以陕北秃尾河流域旱涝状态的预测为例对该方法进行检验。秃尾河流域  $1956 \sim 2000$  年降水量资料如图 1 所示。

(1) 由计算得年降水量序列均值  $\bar{x} = 407.5 \text{ mm}$ , 标准差  $S = 124.2 \text{ mm}$ , 经分析本文可取  $\alpha_1 = 1.1, \alpha_2 = 0.5$ , 将秃尾河流域年降水量划分为雨涝、偏涝、正常、偏旱和干旱 5 个状态(见表 2), 作为年旱涝状态的标准, 据此确定年降水量的分级标准(相当于确定了马尔可夫链的状态空间  $E = \{1, 2, 3, 4, 5\}$ )。

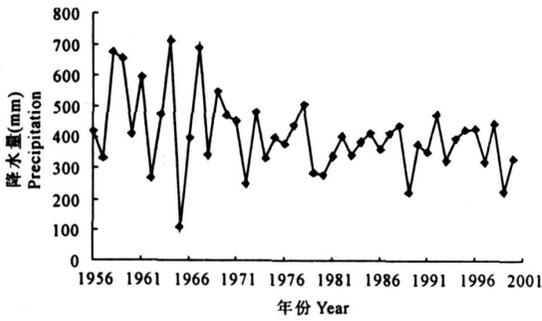


图 1 秃尾河流域 1956~2000 年降水量

Fig.1 The precipitation of Tuwei River reaches from 1956~2000

(3) 对所得结果进行统计计算, 可得不同滞时 (步长从 1 到 5) 的一步转移概率矩阵如下, 它决定了指标值状态转移过程的概率法则。

$$P^{(1)} = \begin{bmatrix} 0 & 1/5 & 2/5 & 2/5 & 0 \\ 0 & 2/9 & 5/9 & 0 & 2/9 \\ 3/18 & 2/18 & 9/18 & 2/18 & 2/18 \\ 0 & 3/5 & 1/5 & 0 & 1/5 \\ 2/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{bmatrix}$$

$$P^{(2)} = \begin{bmatrix} 0 & 1/4 & 1/4 & 0 & 2/4 \\ 1/9 & 2/9 & 4/9 & 1/9 & 1/9 \\ 2/19 & 5/19 & 8/19 & 3/19 & 1/19 \\ 2/5 & 1/5 & 2/5 & 0 & 0 \\ 0 & 0 & 3/6 & 1/6 & 2/6 \end{bmatrix}$$

(2) 按所建立的分级标准确定资料序列中各时段降水量所对应的状态(见表 1)。

表 1 秃尾河流域 1956~2000 年降水量分级

Table 1 Classification of precipitation from 1956~2000 in Tuwei River reaches

状态 State	级别 Grade	分级标准 Grading standard	降水量区间(mm) Interval of precipitation
1	干旱 Drought year	$x < x - 1.1s$	$x < 270.9$
2	偏旱 Weak drought year	$x - 1.1s \leq x < x - 0.5s$	$270.9 \leq x < 345.4$
3	正常 Normal year	$x - 0.5s \leq x < x + 0.5s$	$345.4 \leq x < 469.7$
4	偏涝 Weak water-logging year	$x + 0.5s \leq x < x + 1.1s$	$469.7 \leq x < 544.2$
5	雨涝 Water-logging year	$x \geq x + 1.1s$	$x \geq 544.2$

$$P^{(3)} = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 & 0 \\ 0 & 2/8 & 6/8 & 0 & 0 \\ 2/18 & 5/18 & 7/18 & 2/18 & 2/18 \\ 0 & 1/4 & 2/4 & 1/4 & 0 \\ 2/6 & 0 & 0 & 1/6 & 3/6 \end{bmatrix}$$

$$P^{(4)} = \begin{bmatrix} 0 & 1/4 & 2/4 & 0 & 1/4 \\ 1/8 & 2/8 & 3/13 & 1/8 & 1/8 \\ 2/18 & 4/18 & 9/18 & 2/18 & 1/18 \\ 0 & 1/5 & 3/5 & 0 & 1/5 \\ 2/5 & 1/5 & 0 & 1/5 & 0 \end{bmatrix}$$

$$P^{(5)} = \begin{bmatrix} 0 & 0 & 2/4 & 1/4 & 1/4 \\ 1/7 & 1/7 & 4/7 & 1/7 & 0 \\ 3/17 & 4/17 & 8/17 & 1/17 & 1/17 \\ 0 & 3/5 & 1/5 & 1/5 & 0 \\ 1/6 & 1/6 & 1/6 & 1/6 & 2/6 \end{bmatrix}$$

$\chi^2_{\alpha}(16) = 26.296$ 。由于  $\chi^2 > \chi^2_{\alpha}((m-1)^2)$ , 故秃尾河流域年降水量序列满足马氏性。

(5) 按式(7)和式(8)可计算得各阶自相关系数:  $r_1 = -0.128, r_2 = -0.066, r_3 = 0.322, r_4 = -0.107, r_5 = -0.044$ , 各步长马尔可夫链的权重  $w_1 = 0.192, w_2 = 0.098, w_3 = 0.483, w_4 = 0.161, w_5 = 0.066$ 。

(6) 依 2000, 1999, 1998, 1997, 1996 年的年降水量及其相应的状态转移概率矩阵对 2001 年的年降水量状态进行预测, 结果如表 2 所示。

(7) 由表 2 可得  $P_1 = 0.085, P_2 = 0.257, P_3 = 0.410, P_4 = 0.078, P_5 = 0.170$ 。将同一状态的各预测概率加权求和之后得  $\max \{P_i, i \in E\} = 0.410 = P_3$ , 可见 2001 年秃尾河流域年降水量的预测状态是 3, 即该年的降水量约在 345.4~469.7 mm 之间, 属于正常水文年。2001 年秃尾河流域的实际降水量为 442 mm, 该结果与预测结果相吻合。同理, 以 1997~2001 年的资料预测 2002 年降水量强度状态, 预测过程重复(1)~(7), (3)可不作, 限于篇幅仅将结果列于表 3。

(4) 对秃尾河流域 1956~2000 年共 45 a 的降水量序列作马氏性检验。由图 1 提供的资料可算得年降水量的频数转移矩阵, 再结合步长为 1 的一步转移概率矩阵  $P^{(1)}$ , 可算得统计量  $\chi^2 = 51.03$ , 给定显著性水平  $\alpha = 0.05$ , 查  $\chi^2$  分布分位数表可得分位点

表 2 2001 年年降水量状态预测  
Table 2 Forecast of precipitation condition in 2001

初始年 Initial year	初始状态 Initial state	滞时(a) Residence time	权重 Weight	状态 1 State 1	状态 2 State 2	状态 3 State 3	状态 4 State 4	状态 5 State 5	概率来源 Probability source
2000	2	1	0.192	0	0.222	0.556	0	0.222	P(1)
1999	1	2	0.098	0	0.250	0.250	0	0.500	P(2)
1998	3	3	0.483	0.111	0.278	0.389	0.111	0.111	P(3)
1997	2	4	0.161	0.125	0.250	0.375	0.125	0.125	P(4)
1996	3	5	0.066	0.176	0.235	0.471	0.059	0.059	P(5)
	$P_i$			0.085	0.257	0.410	0.078	0.170	—

表 3 2002 年年降水量状态预测  
Table 3 Forecast of precipitation condition in 2002

初始年 Initial year	初始状态 Initial state	滞时(a) Residence time	权重 Weight	状态 1 State 1	状态 2 State 2	状态 3 State 3	状态 4 State 4	状态 5 State 5	概率来源 Probability source
2001	3	1	0.192	0.158	0.105	0.526	0.105	0.105	P(1)
2000	2	2	0.109	0.100	0.200	0.500	0.100	0.100	P(2)
1999	1	3	0.471	0.200	0.200	0.400	0.200	0.000	P(3)
1998	3	4	0.162	0.105	0.211	0.526	0.105	0.053	P(4)
1997	2	5	0.065	0.125	0.125	0.625	0.125	0.000	P(5)
	$P_i$			0.161	0.179	0.470	0.151	0.040	—

由表 3 知,2002 年秃尾河流域降水量预测状态为 3,其实际降水量为 440 mm,因此预测与实际相吻合。

(8) 由步骤(5)知,步长为 3 的马尔可夫链的相依性较强,因此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。此时的转移概率矩阵是  $P^{(3)}$ ,由于由  $P^{(3)}$  所决定的马尔可夫链的 5 个状态是互通的,即对  $i, j \in E \in E, i, j(i \neq j)$ ,且为非周期的,故这是一个不可约的正常返链。由马尔可夫链理论还可知道此链是遍历的(不可约、非周期、正常返),因此,此链存在唯一的平稳分布,且此时的平稳分布即为它的极限分布。极限分布求解方程为:

$$\sum_{j \in E} \pi_j = 1 \quad \pi_j = \sum_{i \in E} \pi_i P_{ij} \quad (j \in E) \quad (11)$$

由此可解得平稳分布(极限分布)与各状态的重现期。如表 4 所示。

表 4 平稳分布、极限分布与各状态重现期  
Table 4 Stationary distribution, limit distribution and recurrence interval of various conditions

状态 $j$	1	2	3	4	5
$\pi_j$	0.1076	0.2379	0.4357	0.1219	0.0968
$T_i = \mu_i$	9.29	4.20	2.29	8.20	10.33

注:表中  $\pi_j = 1/\mu_j$ 。

由表 4 可知,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  $T_1 = \mu_1 = 9.29 \text{ a}$ ,  $T_2 = \mu_2 = 4.20 \text{ a}$ ,  $T_3 = \mu_3 = 2.29 \text{ a}$ ,  $T_4 = \mu_4 = 8.20 \text{ a}$ ,  $T_5 = \mu_5 = 10.33 \text{ a}$ 。由上面的分析可知,按照本文确定的分级标准,根据现有的资料信息,在 40 多年的降水过程中正常年出现的机会最大,平均每隔 2.29 a 出现 1 次,出现的概率为 43.57%;雨涝年出现的机会最小,出现的概率仅为 9.68%,平均每隔 10.33 a 才出现 1 次;偏涝年出现的概率比干旱年和雨涝年的大,平均每隔 8.20 a 就会出现 1 次。总体来说,正常年份出现的机会最多,偏涝和偏旱出现的机会偏少,干旱和雨涝出现的机会最少。这与秃尾河流域地处中纬度大陆性季风气候区和处于半湿润气候与半干旱气候的过渡地带有关,同时也与降雨的物理成因有关。

### 4 结 语

1) 通过对秃尾河流域 1956~2000 年降水量系列进行干旱、偏旱、正常、偏涝、雨涝年 5 种类型的划分,利用权马尔可夫链模型对 2001、2002 年降水量处于某一区间进行预测,结果较为满意。

2) 结合马尔可夫链的遍历性和平稳分布进行分析,得出秃尾河流域在年降水量长期丰枯变化中,出现正常年份的状态占优势,概率为 43.57%;出现干旱和雨涝的概率较小;出现偏旱年的概率相对较大为 23.79%,出现偏涝年的概率较小为 12.19%。

3) 结合资料情况对流域未来降水变化特点进

行了预测,预测的结果是一个状态,而非具体值,仅适合非点值的状态预测,如何根据预测的状态求出具体数值将有待于进一步研究。

#### 参考文献:

- [1] 陈守煜. 中长期水文预报综合分析理论模式与方法[J]. 水利学报, 1997, (4): 15—21.
- [2] 王家生, 刘嘉. 随机过程基础[M]. 天津: 天津大学出版社, 2003.
- [3] 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.
- [4] 罗积玉, 邢瑛. 经济统计分析方法及预测[M]. 北京: 清华大学出版社, 1987: 347—348.
- [5] 夏乐天, 朱永忠. 工程随机过程[M]. 南京: 河海大学出版社, 2000.
- [6] 夏乐天, 朱元铎, 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—27.
- [7] 冯耀龙, 韩文秀. 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.
- [8] 丁晶, 邓育红. 随机水文学[M]. 成都: 成都科技大学出版社, 1988.
- [9] 孙才志, 张戈, 林学钰. 加权马尔可夫链在降水丰枯状况预测中的应用[J]. 系统工程理论与实践, 2003, (4): 100—105.

## Prediction on changes of precipitation in Tuwei River reaches based on weighted Markov Chain

LI Xiao-bing, CAI Huan-jie, ZHANG Xin, WANG Jian

(Key Laboratory of Education Ministry for Agricultural Soil and Water Engineering  
in Arid Area, Northwest A & F University, Yangling 712100, China)

**Abstract:** Mean-standard deviation classification method is applied to classify precipitation of five states, i.e. water-logging year, weak water-logging year, normal year, weak drought year, and drought year based on the fact that there are much uncertainty and imprecise characteristics in the precipitation course. Then a method called Markov chain with weights is presented to predict future precipitation state by regarding the standardized self-coefficients as weights based on the special characteristics of precipitation being a dependent stochastic variable. Finally the middle reach of the Tuwei River of the Huanghe River basin is taken as an example to carry out the method and analyze the possibility and change patterns of precipitation in the future condition. The results indicate that the precipitation prediction state of Tuwei River consistent with the practical situation in 2002.

**Keywords:** precipitation; weight; Markov chain; prediction; Tuwei River