

# 马尔科夫链 在溧阳市降水量预测中的应用

林小明<sup>1</sup>, 华 晨<sup>2</sup>, 胡尊乐<sup>2</sup>, 李 丹<sup>2</sup>, 闫 浩<sup>2</sup>, 纪小敏<sup>3</sup>

(1. 溧阳市水利局, 江苏 常州 213300; 2. 江苏省水文水资源勘测局常州分局, 江苏 常州 213022;  
3. 江苏省水文水资源勘测局, 江苏 南京 210029)

**摘要:**在采用均方差分级法对溧阳市1951—2020年的降水量进行分级的基础上,运用马尔科夫链理论和模糊集理论,建立加权马尔科夫链降水量预测模型,并运用于溧阳市年降水量和丰枯水重现期的预测,预测结果良好,为区域水旱灾害防御、水资源优化配置、河湖生态保护提供了技术支撑。

**关键词:**马尔科夫链; 模糊集理论; 降水量; 预测

中图分类号:TV125

文献标识码:B

文章编号:1007-7839(2022)07-0041-0005

## Application of Markov chain in precipitation prediction and analysis of Liyang City

LIN Xiaoming<sup>1</sup>, HUA Chen<sup>2</sup>, HU Zunle<sup>2</sup>, LI Dan<sup>2</sup>, YAN Hao<sup>2</sup>, JI Xiaomin<sup>3</sup>

(1. Liyang Water Resources Bureau, Liyang 213300, China; 2. Changzhou Branch of Jiangsu Province Hydrology and Water Resources Survey Bureau, Changzhou 213022, China;  
3. Jiangsu Province Hydrology and Water Resources Survey Bureau, Nanjing 210029, China)

**Abstract:** Based on the classification of the precipitation in Liyang City from 1951 to 2020 using the mean square error classification method, using Markov chain theory and fuzzy set theory, a weighted Markov chain precipitation prediction model is established and applied to Liyang City Prediction of annual precipitation and return period of high and low water. The prediction results are good, providing technical support for regional flood and drought disaster prevention, optimal allocation of water resources, and ecological protection of rivers and lakes.

**Key words:** Markov chain; fuzzy set theory; precipitation; forecast

溧阳市地处北亚热带边缘、太湖流域上游,境内地形变化较大,山圩紧依,河流源短流急,水情盈枯强弱多变。进入20世纪80年代以后,一方面随着溧阳市工业化和城镇化的迅速发展,水资源短缺问题日趋突出;另一方面因流域下垫面及上下游工情变化较大,区域防洪压力有加重之势。准确预测区域年度降水量,合理配置水资源,减缓水资源供需矛盾,为区域防洪提供可靠的技术支撑显得较为迫切且具有重要的社会经

济效益<sup>[1]</sup>。

本文对溧阳市1951—2020年的降水量进行分级,将其分为枯水年、偏枯年、平水年、偏丰年、丰水年5个状态,以马尔科夫链理论<sup>[2-3]</sup>和模糊集理论<sup>[4]</sup>为基础,以规范化的各阶自相关系数为权重,建立适用于溧阳市降水量的马尔科夫链加权预测模型,对溧阳市2021年的降水量、丰枯水重现期进行了预测,旨在为溧阳市水旱灾害防御、水资源优化配置以及河湖生态流量的调度提供技术支撑。

收稿日期: 2022-04-21

作者简介:林小明(1980—),男,工程师,本科,主要从事水文水资源工作。E-mail: 379314953@qq.com

# 1 基本原理与步骤

## 1.1 马尔科夫链

马尔科夫链是研究事物状态及状态转移规律的理论,通过不同状态的初始概率及状态的转移概率来确定状态之间的变化趋势,从而预测其未来变化趋势的一种技术,其基本原理如下。

如果对时间  $t$  的任意  $n$  个数值 ( $n \geq 3$ ), 在条件  $X(t_i)=x_i (i=1, 2, \dots, n-1)$  下,  $X(t_n)$  的分布函数恰好等于在条件  $X(t_{n-1})=x_{n-1}$  下  $X(t_n)$  的分布函数, 即:

$$F(x_n; t_n | x_{n-1}, x_{n-2}, \dots, x_1; t_{n-1}, t_{n-2}, \dots, t_1) = F(x_n; t_n | x_{n-1}, t_{n-1}) \quad (1)$$

通常称  $X(t_n)$  为马尔可夫链(过程), 马尔科夫链法可用来预测未来发生状态的水文时间序列, 广泛用于一个区域中长期降水量的预测预报。

从应用的实践来看, 传统的马尔科夫模型采用最大隶属原则来确定预报对象的状态, 它有两点缺点: 一是只考虑最大概率, 忽略了其他概率的影响; 二是无法预报对象的具体值, 只能预报对象所在的区间。而加权马尔科夫链预测模型<sup>[5-7]</sup>先依据前面若干时段的预报对象对后面某一时段的预测对象进行预测, 再考虑前面各个时段与所求时段的相关关系进行加权求和, 其最大优点是既考虑最大概率的影响, 也考虑其他概率的影响, 弥补了传统的马尔科夫模型的不足, 提高了预测的可靠性。

## 1.2 模糊集理论中的级别特征值

模糊集理论是一种描述模糊现象的方法, 它把待考察的对象及反映它的模糊概念作为一定的模糊集合, 建立适当的隶属函数, 通过模糊集合的有关运算和变换, 对模糊对象进行分析。具体到前述马尔科夫链中的预测对象, 就是通过模糊集理论中的级别特征值<sup>[8]</sup>解决传统马尔科夫链中“最大隶属原则”只考虑最大概率影响的问题, 而兼顾考虑其他概率的影响, 也就是通过级别特征值对马尔科夫链中的各状态赋予相应权重, 亦即加权马尔科夫链。

(1) 对各状态赋予相应权重, 得权重集  $w=\{w_1, w_2, \dots, w_i\}$ , 其中  $i$  为研究系统的状态。权重公式为

$$w_i = \frac{p_i^\eta}{\sum_{i=1}^s p_i^\eta} \quad (2)$$

式中:  $\eta$  为最大概率作用指数, 指数越大越能够突出最大概率的作用, 本文取 2;  $p_i$  为状态  $i$  的概率。

(2) 对各级别特征值进行计算, 级别特征值  $H$  计算式为

$$H = \sum_{i=1}^s i \times w_i \quad (3)$$

若级别特征值  $H > i$ , 则预测对象的特征值  $Q$  为

$$Q = \frac{T_i H}{i + 0.5} \quad (4)$$

若级别特征值  $H < i$ , 则预测对象的特征值  $Q$  为

$$Q = \frac{B_i H}{i - 0.5} \quad (5)$$

式中:  $T_i, B_i$  分别为状态区间的上限和下限。

## 1.3 加权马尔科夫链法计算步骤

以均值、方差作为选取预测对象( $Q$ )组合的准则, 即既定组合方差下的最大均值或既定均值下的最小方差。

(1) 计算预测对象的均值和方差  $S$ , 建立指标值分级标准, 确定序列中各个时段的指标值所处的状态。计算式为

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (6)$$

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \quad (7)$$

(2) 对预测对象的序列进行马氏性检验。检验随机过程是否具有马尔科夫性<sup>[9]</sup>是应用马尔科夫预测模型的必要前提。一般情况下, 选用离散序列的马尔可夫链来对变量具有随机性的序列进行“马氏性”检验, 检验常用  $\chi^2$  (卡方分布) 统计量。计算方法为:

设研究的序列状态个数为  $m$ , 使用  $(f_{ij}), i, j \in I$  表示转移频数概率矩阵, 把各个列之和去除以全部元素之和, 就会得到“边际概率”, 用字母  $p_j$  表示, 其中:

$$p_j = \frac{\sum_{i=1}^m f_{ij}}{\sum_{i=1}^m \sum_{j=1}^m f_{ij}} \quad (8)$$

当  $m$  很大时,  $\chi^2$  (卡方分布) 统计量为

$$\chi^2 = 2 \sum_{i=1}^m \sum_{j=1}^m f_{ij} \left| \log \frac{p_{ij}}{p_j} \right| \quad (9)$$

如果给定显著性水平为  $\alpha$ , 经查表可得  $\chi_{\alpha}^2((m-1)^2)$  值。如果  $\chi^2 > \chi_{\alpha}^2((m-1)^2)$ , 则拒绝零假设, 可以认为序列具备“马氏性”, 反之则这个序列不能当作马尔可夫链来对待。

(3) 计算各阶自相关系数。计算式为

$$r_k = \frac{\sum_{i=1}^{n-k} (X_i - \bar{X})(X_{i+k} - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (10)$$

(4) 考虑到各阶相关系数可能会出现负值, 需对各阶自相关系数进行规范化处理, 即:

$$W_k = \frac{|r_k|}{\sum_{k=1}^m |r_k|} \quad (11)$$

式中,  $m$  为预测时需要计算的最大阶数。

(5) 对上述步骤的结果进行统计, 得到不同步长的状态转移概率矩阵。以前面各年预测对象为初始状态, 结合其相应的状态转移概率矩阵即可预测出该时段预测对象状态  $P_i^{(k)}$ , 其中  $k$  为步长。

(6) 指标值处于某状态的预测概率可用处于此状态各预测概率的加权和  $P_i$  表示,  $\max(P_i, i \in I)$  ( $I$  为状态空间) 对应的状态即为此时段指标值的预测状态。然后, 利用模糊集理论中的级别特征值即可计算出此时段的预测值 ( $Q$ )。重复以上步骤可预测下一时段的预测值。

(7) 进一步对加权马尔科夫链的遍历性、平稳分布进行预测。其中, 遍历性是指对于有限状态马氏链, 如果存在正整数  $s$ , 并且对所有的  $i, j$ , 都有  $j=1, 2, \dots, k$  对于  $P_{ij}^{(s)}$  成立, 则该马氏链必具有遍历性。平稳分布是指对于有限状态马氏链, 这个链的所有状态都是正常返的, 且这是极限分布  $P_i$  是唯一的平稳分布<sup>[10]</sup>。马尔科夫链的遍历性、平稳分布预测是采用加权马尔科夫链法的必要条件。

## 2 应用实例

本文选取江苏省溧阳市 1951—2020 年降水量作为对象, 采用上述加权马尔科夫链法对该市 2021 年降水量及丰枯年重现期进行预测分析。

### 2.1 均值与方差

利用式(6)、式(7), 计算出溧阳市 1951—2020 年平均降水量  $\bar{X}=1\ 165.6$  mm, 均方差  $S=246.1$  mm。

### 2.2 分级与状态

(1) 根据水文上传统的年型分级定义, 用均方差法把溧阳市 1951—2020 年降水量分成干旱年、偏旱年、平水年、偏丰年、丰水年等 5 级(状态)。其中, 干旱年定义为  $x < (\bar{X}-S)$ , 偏旱年定义为  $(\bar{X}-S) < x < (\bar{X}-0.5S)$ , 平水年定义为  $(\bar{X}-0.5S) < x < (\bar{X}+0.5S)$ , 偏丰年定义为  $(\bar{X}+0.5S) < x < (\bar{X}+S)$ , 丰水年定义为  $x > (\bar{X}+S)$ , 见表 1。

(2) 根据已知的溧阳市年降水量序列, 结合表 1 的分级标准, 判断每年降水量的状态, 见表 2。

### 2.3 马氏性检验

利用式(8)、式(9), 对溧阳市年降水量序列进行马氏性检验。其中, 状态空间  $m=5$ , 显著性水平  $\alpha=0.0$ ,  $\chi^2_{\alpha}((m-1)^2)=26.3$ ,  $\chi^2=187.9 > \chi^2_{\alpha}((m-1)^2)$ , 则拒绝零假设, 可以认为溧阳市年降水量序列具备“马氏性”。

### 2.4 成果预测

(1) 利用式(10)和式(11), 计算出前 5 阶自相关系数  $r_k$  与相应的权重  $W_k$ , 见表 3。

(2) 根据表 2, 计算溧阳市 1951—2020 年降水量各种步长的状态概率矩阵, 见式(12)。

$$\left\{ \begin{array}{l} P_1 = \begin{bmatrix} 1/7 & 1/7 & 1/7 & 3/7 & 1/7 \\ 1/14 & 5/14 & 5/14 & 2/14 & 1/14 \\ 3/30 & 5/30 & 15/30 & 3/30 & 4/30 \\ 1/11 & 1/11 & 7/11 & 2/11 & 0 \\ 1/7 & 2/7 & 2/7 & 1/7 & 1/7 \end{bmatrix} \\ P_2 = \begin{bmatrix} 1/7 & 2/7 & 3/7 & 1/7 & 0 \\ 2/14 & 6/14 & 5/14 & 1/14 & 0 \\ 3/29 & 8/29 & 12/29 & 4/29 & 2/29 \\ 3/11 & 1/11 & 5/11 & 2/11 & 0 \\ 0 & 0 & 4/7 & 1/7 & 2/7 \end{bmatrix} \\ P_3 = \begin{bmatrix} 2/7 & 0 & 3/7 & 0 & 2/7 \\ 1/13 & 2/13 & 6/13 & 3/13 & 3/13 \\ 3/29 & 8/29 & 13/29 & 4/29 & 2/29 \\ 0 & 2/11 & 4/11 & 4/11 & 1/11 \\ 1/7 & 1/7 & 3/7 & 2/7 & 0 \end{bmatrix} \\ P_4 = \begin{bmatrix} 0 & 1/7 & 3/7 & 2/7 & 1/7 \\ 1/13 & 2/13 & 5/13 & 3/13 & 2/13 \\ 5/29 & 7/29 & 11/29 & 4/29 & 2/29 \\ 1/11 & 3/11 & 5/11 & 1/11 & 1/11 \\ 0 & 0 & 4/6 & 1/6 & 2/6 \end{bmatrix} \\ P_5 = \begin{bmatrix} 0 & 1/7 & 4/7 & 1/7 & 1/7 \\ 1/13 & 2/13 & 5/13 & 4/13 & 1/13 \\ 3/28 & 7/28 & 11/28 & 4/28 & 3/28 \\ 3/11 & 1/11 & 5/11 & 1/11 & 1/11 \\ 0 & 2/6 & 3/6 & 1/6 & 0 \end{bmatrix} \end{array} \right. \quad (12)$$

(3) 根据溧阳市 1951—2020 年降水量及其相应

表 1 溧阳市降水量分级

状态	级别	分级标准	降水量区间/mm
1	干旱年	$x < (\bar{X}-S)$	$x < 919.5$
2	偏旱年	$(\bar{X}-S) < x \leq (\bar{X}-0.5S)$	$919.5 < x \leq 1\ 042.6$
3	平水年	$(\bar{X}-0.5S) < x \leq (\bar{X}+0.5S)$	$1\ 042.6 < x \leq 1\ 288.7$
4	偏丰年	$(\bar{X}+0.5S) < x \leq (\bar{X}+S)$	$1\ 288.7 < x \leq 1\ 411.7$
5	干旱年	$x > (\bar{X}+S)$	$x > 1\ 411.7$

表2 溧阳市年降水量序列及状态

年份	降水量/mm	状态	年份	降水量/mm	状态	年份	降水量/mm	状态
1951	1 231.4	3	1975	1 217.9	3	1999	1 396.5	4
1952	1 151.8	3	1976	1 187.7	3	2000	1 021.7	2
1953	1 024.5	2	1977	1 373.2	4	2001	1 020.5	2
1954	1 708.7	5	1978	582.0	1	2002	1 163.1	3
1955	1 113.9	3	1979	970.5	2	2003	1 087.9	3
1956	1 399.4	4	1980	1 376.8	4	2004	941.9	2
1957	1 320.0	4	1981	1 185.2	3	2005	1 027.3	2
1958	1 000.6	2	1982	1 113.9	3	2006	1 052.9	3
1959	968.0	2	1983	1 296.8	4	2007	983.6	2
1960	1 031.1	2	1984	1 337.6	4	2008	1 078.4	3
1961	1 098.4	3	1985	1 224.7	3	2009	1 262.6	3
1962	1 368.0	4	1986	925.8	2	2010	1 076.3	3
1963	1 130.2	3	1987	1 575.6	5	2011	1 131.9	3
1964	1 143.3	3	1988	850.1	1	2012	1 242.4	3
1965	1 418.9	5	1989	1 486.8	5	2013	1 116.6	3
1966	933.9	2	1990	1 137.7	3	2014	1 328.1	4
1967	954.8	2	1991	1 627.4	5	2015	1 486.5	5
1968	856.4	1	1992	1 122.6	3	2016	2 307.6	5
1969	1 068.5	3	1993	1 408.5	4	2017	1 123.4	3
1970	1 235.5	3	1994	846.9	1	2018	1 211.4	3
1971	877.1	1	1995	917.9	1	2019	956.1	2
1972	1 092.6	3	1996	1 294.3	4	2020	1 245.3	3
1973	1 056.7	3	1997	811.9	1	平均	1 165.6	
1974	1 178.7	3	1998	1 097.1	3	方差	246.1	

表3 溧阳市年降水量1~5阶各阶自相关系数及权重

项目	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$r_k$	-0.0342	0.1958	-0.0484	0.0356	-0.4150
$W_k$	0.0469	0.2686	0.0664	0.0488	0.5693

的状态转移概率矩阵,采用1~5阶权重系数对溧阳市2021年的降水量状态进行预测,结果见表4。

各状态对应的加权和分别为0.0469、0.2686、0.0664、0.0488、0.5693。由表4可知,  $\max\{p_i, i \in E\} = 0.4715$ , 对应的  $i=3$ , 即溧阳市2021年降水量预测状态为3(平水年), 状态3的数值区间为[1 042.6, 1 288.7]。由模糊集理论计算得级别特征值  $H=3.825$ ,  $H>i$ , 故由式(4)计算得2021年降水量为1 408.4 mm, 实测值为1 464.5 mm, 相对误差为-3.8%, 符合相关标准( $\leq \pm 20\%$ )要求。

## 2.5 丰枯水重现期预测

丰枯水重现期预测是溧阳市防汛防旱和水资

源科学开发利用的重要依据。本文根据溧阳市1951—2020年的滑动平均降水量序列,以相依性最强的步长为5的马尔科夫链进行分析。降水量马尔科夫链的5个状态是互通的,且是非周期的,因此该马氏链具有遍历性,且存在唯一的平稳分布。

步长为5的状态转移概率矩阵为

$$P_5 = \begin{bmatrix} 0 & 1/7 & 4/7 & 1/7 & 1/7 \\ 1/13 & 2/13 & 5/13 & 4/13 & 1/13 \\ 3/28 & 7/28 & 11/28 & 4/28 & 3/28 \\ 3/11 & 1/11 & 5/11 & 1/11 & 1/11 \\ 0 & 2/6 & 3/6 & 1/6 & 0 \end{bmatrix} \quad (13)$$

设 $\pi_i$ 、 $\pi_j$ 分别为状态 $i$ 、 $j$ 的平稳分布,则得到方程组:



表4 溧阳市2021年降水量预测

初始年	状态	步长	权重	状态转移概率矩阵				
				状态1	状态2	状态3	状态4	状态5
2020	3	1	0.0469	1/10	1/6	1/2	1/10	2/15
2019	2	2	0.2686	0	1/7	3/7	5/14	1/14
2018	3	3	0.0664	3/29	8/29	13/29	2/29	3/29
2017	3	4	0.0488	5/29	7/29	11/29	4/29	2/29
2016	5	5	0.5693	0	1/3	1/2	1/6	0
$\sum p_i$				0.0200	0.2661	0.4715	0.2068	0.0357

$$\begin{cases} \sum_{j=1}^5 \pi_j = 1 \\ \pi_j = \sum_{i=1}^5 \pi_i P_{ij} \end{cases} \quad (14)$$

由上述状态转移概率矩阵计算出极限分布( $y$ )与各状态的重现期( $T$ )见表5。

由表5可知,溧阳市出现平水年可能性较大,重现期2.17 a。同时溧阳市近年来降水量总体呈增加趋势,降水变化存在5 a和11 a这2个主周期<sup>[11]</sup>。

### 3 结 语

通过对2021年溧阳市年降水量预测实践,可得

表5 溧阳市丰枯水重现期预测

项目	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$y_i$	0.0914	0.1942	0.4607	0.1702	0.0836
$T_j$	10.95	5.15	2.17	5.88	11.97

出以下结论。

(1)采用均方差分级法对溧阳市1951—2020年的降水量进行分级,建立的加权马尔科夫链预测模型较可靠,对2021年溧阳市年降水量预测结果相对误差仅为-3.8%。

(2)以马尔科夫链理论和模糊集理论为基础,以规范化的各阶自相关系数为权重,建立的溧阳市降水量的加权马尔科夫链预测模型,对丰枯水重现期的预测结果较合理,预测结果与实际基本吻合。

(3)运用加权马尔科夫链预测模型,得出的溧阳市降水量处于平水年的概率最大,平均周期为2.17 a,降水年际变化存在5 a和11 a这2个主周期,为区域防汛防旱、水资源综合利用规划提供可靠的依据。

#### 参考文献:

- [1] LI Z J, LI X P. Analysis on variation of rainfall trend and agriculture dryness in Benxi City of Liaoning Province[J]. Meteorological and Environmental Research, 2010, 1(7): 82-85.
- [2] 王涛,钱会,李培月. 加权马尔科夫链在银川地区降水量预测中的应用[J]. 南水北调与水利科技, 2010, 8

(1):78-81.

- [3] 孙才志,张戈,林学钰. 加权马尔科夫模型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): 103-105.
- [4] 季小兵,王志坚,魏光辉. 加权马尔科夫链在开都河径流预测中的应用[J]. 人民黄河, 2009, 31(4): 42-44.
- [5] 孙才志,林学钰. 降水预测的模糊权马尔科夫模型及应用[J]. 系统工程学报, 2003, 8(4): 295-298.
- [6] 李娟,张维生,马轶. 滑动平均——马尔科夫模型在降水预测中的应用[J]. 水土保持研究, 2005, 12(6): 196-198.
- [7] 冯耀龙,韩文秀. 马尔科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): 89-98.
- [8] 孙才志,林学钰. 降水预测的模糊权马尔科夫模型及应用[J]. 系统工程学报, 2003, 18(4): 294-299.
- [9] 仲远见,李靖,王龙. 改进马尔科夫链降水量预测模型的应用[J]. 济南大学学报(自然科学版), 2009, 23(4): 402-405.
- [10] 周秀平,黄伟军,王文圣,等. 模糊马尔科夫链预测模型在桂江流域径流预测中的应用[J]. 广西水利水电, 2006(1): 29-31.
- [11] 潘杰. 常州地区1978—2012年降水量时空分布特征[J]. 南水北调与水利科技, 2015(2): 245-248.