降水预测的对数马尔可夫模型及其在盐池县沙地中的应用

张维江，李 娟，齐化龙

（宁夏大学 西北退化生态系统恢复与重建省部共建教育部重点实验室，土木与水利工程学院，宁夏 银川 750021）

摘 要：中长期降水量的预测是气象科学的一个难点问题，也是水文学中的一个重要问题。建立对数马尔可夫模型预测降水量，弥补了传统的马尔可夫模型降水预测中峰值的不准确性，提高了预测精度，并用宁夏盐池县气象站 43 a 的降水资料进行了验证。结果表明，模型预测精度较高，为半干旱风沙区中长期降水量预报提供了一条简便可行的途径。

关键词：降水；对数函数；马尔可夫链；盐池县沙地

文献标识码：B　　**文章编号**：1000—288X(2007)05—0096—05　　**中图分类号**：P457.6

Logarithmic Markov Chain Model for Precipitation Forecast and Its Application to Yanchi Sandy Land

ZHANG Wei-jiang, LI Juan, QI Hua-long

(*Key Laboratory of Degraded Ecosystem Restoration and Rehabilitation in Northwest China, Ministry of Education & Ningxia University, and College of Civil and Water Conservancy Engineering, Ningxia University, Ningxia, Yinchuan* 750021, *China*)

Abstract: The medium-and-long-term prediction of precipitation is not only difficult in meteorology, but also an important issue in hydrology. The logarithmic Markov chain model is established to predict precipitation, which can compensate the uncertainty of peak point in precipitation prediction by the traditional Markov model and raise the prediction accuracy. At last, precipitation data of 43 years from the hydrological station in Yanchi County are used to verify the model, and prediction accuracy is found to be satisfied. So, the logarithmic Markov chain can provide an oversimplified channel to the medium-and-long-term precipitation prediction in the semi-arid sandy region.

Keywords: **precipitation; logarithm function; Markov chain; sandy land of Yanchi County**

“马尔可夫性”是由俄国数学家 A. A. Markov 在 1906 年最早提出的。经过几十年不断的发展，Markov 过程已成为随机过程的一个重要分支，该方法利用变量的状态转移概率矩阵预报变幅较大的随机波动，在生物学、物理学、天文学领域中已有广泛的应用[1-2]；马尔可夫链可以描绘一个随机变化的动态系统，它根据状态之间的转移概率来推测一个系统未来的发展变化，而转移概率反映了各随机因素的影响程度，反映了各状态之间转移的内在规律性，适合描述随机波动性较大的预测问题[3]。

大气降水是自然界水循环的一个重要环节。尤其在干旱半干旱地区，降水是水资源的主要补给来源，降水量的大小，决定着该地区水资源的丰富程度。因此，在水资源预测、水文预报中经常需要对降水量进行预报。然而，由于气象条件的变异性、多样性和复杂性，降水过程存在着大量的不确定性与随机性，因此，到目前为止，还难以通过物理成因来确定出未来某一时段降水量的准确数值。

1 马尔可夫过程及马尔可夫链

(1) 定义 1。设 $X(t)$ 是一随机过程，当过程在时刻 t_0 所处的状态已知的条件下，过程在时刻 $t(t>t_0)$ 所处状态与过程在时刻 t_0 之前的状态无关，这个特性称为无后效性。无后效性的随机过程称为马尔可夫过程。(2) 定义 2。时间离散、状态离散的马尔可夫过程称为马尔可夫链[4]。

收稿日期：2006-12-09　**修回日期**：2007-02-25

基金项目：国家科技攻关项目“宁夏河东沙地退化草场植被恢复与利用技术研究与示范”(2005BA517A10)

作者简介：张维江(1963—)，男(汉族)，宁夏海原人，教授，博士，博士生导师，主要从事旱区水资源和水土保持与荒漠化防治方面的教学与研究工作。E-mail:zwjang@263.net。

对于马尔可夫链,用 P_{ij} 表示系统由状态 E_i 经过一次转移到达状态 E_j 的转移概率。由转移概率构成的矩阵 $\boldsymbol{P}$,即

$$\boldsymbol{P}=\begin{Bmatrix} P_{11} & P_{12} & P_{13} & P_{14} & \cdots \\ P_{21} & P_{22} & P_{23} & P_{24} & \cdots \\ P_{31} & P_{32} & P_{33} & P_{34} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ P_{n1} & P_{n2} & \cdots & \cdots & P_{nn} \end{Bmatrix}$$

称为马尔可夫链的状态转移概率矩阵。

由于从任何一个状态 E_i 出发,经过转移后,必然出现状态 E_1 或 E_2,或 $E_3 \ldots$ 或 E_n,因此

$$\sum_{k=1}^{n} P_{ik}=1 \quad (P_{ik} \geqslant 0) \tag{1}$$

式中:n——研究序列的状态数。

2　模糊集理论中的级别特征值

传统的马尔可夫链模型采用最大隶属原则来确定预报对象的状态,它具有两大缺点。(1)只考虑了最大概率,忽略了其它概率的影响;(2)不清楚预报对象在区间内的大致位置,无法达到预报对象具体值的要求,而模糊集理论中的级别特征值可以有效地解决这个问题[5]。

首先给各状态赋以相应的权重,构成权重集 $D=\{d_1,d_2,d_3,d_4,\ldots,d_n\}$,其中权重的大小取决于各状态概率的大小,可以通过式(2)进行计算。

$$d_i=P_i^{\eta} / \sum_{i=1}^{n} P_i^{\eta} \tag{2}$$

式中:i——研究序列的状态;η——最大概率的作用系数,通常取4[5]。

级别特征值 H 可以通过式(3)进行计算:

$$H=\sum_{i=1}^{n} i \times d_i \tag{3}$$

确定最大概率的状态 i 后,可以根据式(4)确定系统在预报时段的预报值。

$$\begin{cases} X_{预报}=T_i H /(i+0.5) & H>i \\ X_{预报}=B_i H /(i-0.5) & H<i \end{cases} \tag{4}$$

式中:T_i,B_i——分别为状态 i 区间值的上限与下限。

3　对数马尔可夫模型及其预测实现的基本步骤

在干旱半干旱地区,降水量较小,且时间分布差异较大,极不均匀。传统的马尔可夫过程只考虑了降水序列概率分布的影响,对于预报与多年平均水平相仿的降水量时精度相对较高。

但是对于时间分布极不均匀的降水序列来讲,利用传统的马尔可夫模型进行预测,在峰值(即丰水年与枯水年)降水的预测中,误差较大,不能满足降水预报工作的精度要求。

本文将降水序列进行对数换算,使得序列的大小差异降低,峰值减小,然后利用马尔可夫模型预测未来某时段的降水量对数,推算该时段的降水量,误差降低,精度相对提高。基于上述思路,对数马尔可夫模型预测实现的基本步骤为:

(1)将降水序列进行对数换算;

(2)计算降水量对数序列的各阶自相关系数。

由物理成因的定性分析及大量的降水序列资料的统计分析得知,降水量为一相依随机变量,通常用各阶自相关系数来表示各种滞时的降水量之间的相关关系及其强弱。各阶自相关系数可以通过式(5)进行计算。

$$r_k=\sum_{t=1}^{n-K}(x_t-\bar{x})(x_{t+k}-\bar{x}) / \sum_{t=1}^{n-K}(x_t-\bar{x})^2 \tag{5}$$

式中:x_t——t 时段的年降水量对数;$\bar{x}$——降水量对数序列的均值;n——降水量对数序列的样本数;k——自相关系数的阶数。

对于降水序列,一般计算5阶自相关系数即可。

(3)对各阶自相关系数进行归一化处理,作为各种滞时(步长)对数马尔可夫链的权重,具体计算见公式(6)。

$$\omega_k=\frac{|r_k|}{\sum_{k=1}^{m}|r_k|} \tag{6}$$

式中:m——按预测需要计算到的最大阶数,本文中取值为5。

(4)将降水量对数序列进行聚类,生成降水量对数的分级标准,确定对数马尔可夫链状态空间。

(5)根据第(4)步所生成的分级标准,确定各时段降水量对数所处的状态 i。

(6)根据第(5)步得到的状态序列,生成不同滞时(步长)的马尔可夫链的状态转移概率矩阵。

(7)分别以前面若干时段的降水量对数为初始状态,综合其相应的状态转移概率矩阵,预测某时段降水量对数的状态概率 $P_i^{(k)}$,其中 i 为状态,k 为滞时(步长)。

(8)将同一状态的各预测概率加权作为降水量对数处于该状态的预测概率,即,$P_i=\sum_{i=1}^{n}\omega_k P_i^{(k)}$。

(9)取 $\max\{P_i, i\in 1,2,3,\cdots,n\}$ 所对应的 i 即为预测时段降水量对数所处的状态,应用级别特征值求出该时段具体的降水量对数,预测该时段的降水量。再将预测的降水量对数加入原序列,再次预测下一时段的降水量对数,并预测降水量。

4 实例分析

利用宁夏盐池县气象站 1961—2003 年的降水资料,讨论对数马尔可夫模型在降水预测中的应用。

4.1 确定降水量对数的状态

考虑水文现象的本身特性及序列数据的结构合理性,将降水量对数序列聚类成 5 类,即将降水量对数划分为 5 个区间,具体结果见表 1。根据以上状态划分标准,确定年降水量对数的状态,见表 2。

表 1 降水量对数状态划分

状 态	等 级	降水量对数
1	枯水年	$X<2.206$
2	偏枯年	$2.206\leqslant X<2.261$
3	平水年	$2.261\leqslant X<2.295$
4	偏丰年	$2.295\leqslant X<2.589$
5	丰水年	$X\geqslant 2.589$

表 2 宁夏盐池县气象站降水量对数序列及状态

时 段	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970
$\overline{X}$/mm	420.8	182.5	245.5	464.0	144.5	147.5	445.4	284.0	200.0	231.8
X/mm	2.624	2.261	2.390	2.667	2.160	2.169	2.649	2.453	2.301	2.365
状 态	5	3	4	5	1	1	5	4	4	4
时 段	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
$\overline{X}$mm	323.6	199.9	276.7	166.3	218.7	203.4	253.4	263.2	206.8	145.3
X/mm	2.510	2.301	2.442	2.221	2.340	2.308	2.404	2.420	2.316	2.162
状 态	4	4	4	2	4	4	4	4	4	1
时 段	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
$\overline{X}$/mm	276.9	171.7	227.6	316.6	399	236.8	197.2	276.2	296.4	333.2
X/mm	2.442	2.235	2.357	2.501	2.601	2.374	2.295	2.441	2.472	2.523
状 态	4	2	4	4	5	4	4	4	4	4
时 段	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
$\overline{X}$/mm	276.8	299.3	230.1	392.4	303.4	348.5	256.4	365.1	294.2	160.8
X/mm	2.442	2.476	2.362	2.594	2.482	2.542	2.409	2.562	2.469	2.206
状 态	4	4	4	5	4	4	4	4	4	2
时 段	2001	2002	2003							
$\overline{X}$/mm	387.7	399.1	293.7							
X/mm	2.588	2.601	2.468							
状 态	5	5	4							

注:$\overline{X}$——年降水量;X——年降水量对数。

4.2 对年降水量对数进行预测,确定年降水量

4.2.1 2000 年降水量预测

(1) 根据表 2 中的状态,计算 1961—1999 年降水量对数序列各种步长的状态转移概率矩阵如下:

$$\boldsymbol{P}_1=\begin{vmatrix} 1/3 & 0/3 & 0/3 & 1/3 & 1/3 \\ 0/1 & 0/1 & 0/1 & 1/1 & 0/1 \\ 0/1 & 0/1 & 0/1 & 1/1 & 0/1 \\ 1/27 & 2/27 & 0/27 & 21/27 & 3/27 \\ 1/5 & 0/5 & 1/5 & 3/5 & 0/5 \end{vmatrix}$$

$$\boldsymbol{P}_2=\begin{vmatrix} 0/3 & 1/3 & 0/3 & 1/3 & 1/3 \\ 0/1 & 0/1 & 0/1 & 1/1 & 0/1 \\ 0/1 & 0/1 & 0/1 & 0/1 & 1/1 \\ 3/26 & 0/26 & 0/26 & 21/26 & 2/26 \\ 1/5 & 0/5 & 0/5 & 4/5 & 0/5 \end{vmatrix}$$

$$\boldsymbol{P}_3=\begin{vmatrix} 0/1 & 0/1 & 0/1 & 1/1 & 0/1 \\ 0/1 & 0/1 & 0/1 & 1/2 & 1/2 \\ 1/1 & 0/1 & 0/1 & 0/1 & 0/1 \\ 2/25 & 2/25 & 0/25 & 20/25 & 1/25 \\ 0/5 & 0/5 & 0/5 & 3/5 & 2/5 \end{vmatrix}$$

$$\boldsymbol{P}_4=\begin{vmatrix} 0/1 & 0/1 & 0/1 & 1/1 & 0/1 \\ 0/1 & 0/1 & 0/1 & 2/2 & 0/2 \\ 1/1 & 0/1 & 0/1 & 0/1 & 0/1 \\ 1/24 & 2/24 & 0/24 & 19/24 & 2/24 \\ 1/5 & 0/5 & 0/5 & 4/5 & 0/5 \end{vmatrix}$$

$$\boldsymbol{P}_5=\begin{vmatrix} 0/1 & 0/1 & 0/1 & 1/1 & 0/1 \\ 0/1 & 0/1 & 0/1 & 2/2 & 0/2 \\ 1/1 & 0/1 & 0/1 & 0/1 & 1/1 \\ 1/24 & 2/24 & 0/24 & 19/24 & 2/24 \\ 1/4 & 0/4 & 0/4 & 3/4 & 0/4 \end{vmatrix}$$

(2) 确定序列的各阶自相关系数并进行归一化作为权重。对于降水序列,一般计算 5 阶自相关系数即可[5]。

根据式(5)确定 1961—1999 年的降水量对数序列的各阶自相关系数如下:$r_1=-0.045$,$r_2=-0.187$,$r_3=0.277$,$r_4=0.245$,$r_5=-0.160$。

利用式(6)对各阶自相关系数进行归一化,作为各种滞时(步长)的对数马尔可夫链的权重。计算得:$\omega_1=0.049$,$\omega_2=0.205$,$\omega_3=0.303$,$\omega_4=0.268$,$\omega_5=0.175$

根据 1961—1999 年的降水量对数及其相应的状态转移概率矩阵对 2000 年的降水量对数进行预测,计算结果见表 3。

由表 3 可知:当 $i=4$ 时,$P_4=0.514$ 为最大值,说明 2000 年的降水量对数状态为 4,即数值区间为(2.295,2.589)。

根据式(3)计算级别特征值为4.011,根据式(4),求出 2000 年的降水量对数为2.300,预测 2000 年降水量为 199.5 mm。

4.2.2　2001—2003 年降水量预测　将前一年降水量对数加入到序列中,用同样的方法预测 2001—2003 年的降水量对数,计算结果见表 4—6。

根据表 4—6,预测 2001,2002 年及 2003 年的降水量对数分别为 2.623,2.621 及 2.420,预测 2001,2002 年及 2003 年的降水量分别为 419.1 mm,417.7 mm 及 263.3 mm。

表 3　2000 年降水量对数预测

初始年	滞时(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
1999	1	0.049	1/27	2/27	0	21/27	3/27
1998	2	0.205	2/26	0/26	0	21/26	2/26
1997	3	0.303	2/25	2/25	0	20/25	1/25
1996	4	0.268	1/24	2/24	0	19/24	2/24
1995	5	0.175	1/23	2/23	0	19/23	1/23
	P_i 加权求和		0.045	0.043	0	0.514	0.048

表 4　2001 年降水量对数预测

初始年	滞时(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
2000	1	0.066	0	0	0	1	0
1999	2	0.240	2/26	2/26	0	20/26	2/26
1998	3	0.293	2/25	3/25	0	19/25	1/25
1997	4	0.208	1/24	3/24	0	18/24	2/24
1996	5	0.193	1/24	3/24	0	19/24	1/24
	P_i 加权求和		0.059	0.104	0.000	0.782	0.056

表 5　2002 年降水量对数预测

初始年	滞时(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
2001	1	0.113	1/5	0	1/5	3/5	0
2000	2	0.218	0	0	0	1	0
1999	3	0.319	2/26	3/26	0	19/26	2/26
1998	4	0.203	1/25	3/25	0	18/25	3/25
1997	5	0.147	1/25	3/25	0	19/25	2/25
	P_i 加权求和		0.061	0.079	0.023	0.777	0.061

表6 2003年降水量对数预测

初始年	滞时（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
2002	1	0.059	1/6	0	1/6	3/6	1/6
2001	2	0.250	1/5	0	0	4/5	0
2000	3	0.319	0	0	0	1/2	1/2
1999	4	0.230	1/26	3/26	0	18/26	4/26
1998	5	0.141	1/5	0	0	4/5	0
	P_i 加权求和		0.097	0.027	0.010	0.661	0.205

4.3 模型检验及未来盐池降水量预报

2000—2003年降水量实测资料和预测资料及其检验结果见表7。

由表7可以看出，在2000—2003年降水量预测中，预测相对误差分别为24.3%，8.09%，4.66%，10.3%，平均相对误差为11.8%。由于中长期水文预报的复杂性，一般认为相对误差小于20%即可[6-7]。因此预测误差满足预测精度需要，同时这4年的实际降水量反映了丰、平、枯3个状态，故此模型可以用于盐池沙地不同频率降水的预报。

利用对数马尔可夫模型，对盐池县未来各年份的降水量进行预测，预测结果见表8。

表7 模型检验结果

时段	级别特征值 H	降水量对数预测值	降水量预测值/mm	实测值/mm	相对误差/%
2000	4.011	2.300	199.5	160.8	24.3
2001	3.999	2.623	419.1	387.7	8.09
2002	3.997	2.621	417.7	399.1	4.66
2003	4.202	2.420	263.3	293.7	10.3
平均	—	—	—	—	11.8

表8 盐池县未来时段降水量预测

年份	2004	2005	2006	2007	2008	2009	2010
降水量对数	2.302	2.240	2.610	2.302	2.557	4.001	2.303
年降水量/mm	200.5	208.9	407.1	200.5	360.8	200.4	200.9

5 结论

(1) 本文将对数函数与马尔可夫链相结合，利用对数马尔可夫模型进行降水预测，弥补了传统的马尔可夫模型预测中峰值的不准确性，提高了预测精度。本文提出的对数马尔可夫模型通过宁夏盐池县气象站2000—2003年降水资料进行验证，可满足降水预报精度的要求。

(2) 本文利用对数马尔可夫模型，预测了盐池县2004—2010年的降水量情况，其中2006年和2008年分别为407.1 mm和360.8 mm。从趋势上看，不论是农业生产还是生态工程建设，2006—2008年的降水情况值得重视。

(3) 盐池县多年平均降水量为270 mm，对数马尔可夫模型能否应用于其它地区，尚需要进一步的研究与论证。

[参考文献]

[1] 邓聚贤，许刘俊．随机过程[M]．北京：高等教育出版社，1992．

[2] Yakowitz S J. Markov flow models and the flood warning problem[J]. Water Resources Research, 1985, 21(1): 81-83.

[3] 张曙红，曹让会，陈绵云．灰色马尔可夫SCGM(1,1)预测模型[J]．佛山科学技术学院学报（自然科学版），2004，3(1)：16—19．

[4] 沈永欢，梁在中．实用数学手册[M]．北京：科学出版社，2003．

[5] 孙才志，林学钰．降水预测的模糊权马尔可夫模型及应用[J]．系统工程学报，2003，8(4)：295—298．

[6] 王本德．水文中长期预报模糊数学方法[M]．大连：大连理工大学出版社，1993．

[7] 陈守煜．水文水资源系统模糊识别理论[M]．大连：大连理工大学出版社，1992．