

滑动平均-马尔可夫模型在降水预测中的应用

李娟, 张维江, 马轶

(宁夏大学土木与水利工程学院, 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。根据降水过程存在大量不确定性的特点, 通过聚类分析建立降水序列的分级标准, 采用规范化的各阶自相关系数为权重, 用滑动平均的马尔可夫链模型, 通过状态转移概率矩阵预测未来时段的降水状态, 并根据模糊集理论中的级别特征值计算具体的降水量, 最后以隆德县水文站 54 年的降水资料为实例, 对该方法进行了具体的应用, 预测精度较高, 为提高中长期降水量预报的精度提供了一条值得探索的途径。

关键词: 降水; 聚类分析; 滑动平均; 马尔可夫链; 预测

中图分类号: P426.614

文献标识码: A

文章编号: 1005-3409(2005)06-0196-03

Application of Moving Average- Markov Chain in the Prediction of Precipitation

LI Juan, ZHANG Wei-jiang, MA Yi

(College of Civil and Water Conservancy Engineering, Ningxia University, Yinchuan 750021, China)

Abstract: The prediction of the medium-and-long-range precipitation is a difficult problem of meteorology. And it is also an important one in hydrology. According to the uncertain characteristic in precipitation procession, the graduation standard of the precipitation serial was set through cluster analysis, the normalized different autocorrelation coefficient was used as weight, the moving average- Markov chain was used to predict precipitation state in a future period with state transition probability matrix, then the particular precipitation was calculated according to the class designated value in unresolved set theory. At last, the precipitation material in 54 years of the hydrological station in Longde county was used as a model and found out that the precipitation accuracy was satisfied. So the moving average Markov chain can be used in the prediction of the medium-and-long-range precipitation, which provides a channel to be searched.

Key words: precipitation; cluster analysis; moving average; Markov chain; prediction

“马尔可夫模型”是由俄国数学家 A. A. Markov 在 1996 年最早提出, 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可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即变化的动态系统, 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, 适合描述随机波动行较大的预测问题^[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。

在实际的水资源工作中, 有时不必预测出某一年的降水

量, 仅需预测出某个时段内降水的状况即可满足工作需要, 因此, 预测的范围相应扩大, 精度相应提高。因此, 本文考虑到降水序列的随机性和连续性, 首先对降水序列进行滑动平均处理, 并通过聚类分析对降水的滑动平均值进行状态分级, 并采用规范化的各阶自相关系数作为权重, 用滑动平均-马尔可夫模型对未来时段的降水进行预测, 并用隆德县 1951~2004 年的降水资料进行验证, 预测精度较高。

1 马尔可夫过程及马尔可夫链

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 $t(t > t_0)$ 所处的状态与过程在时刻 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。

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]。

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

* 收稿日期: 2005-07-08

基金项目: 由国家“十五”重大科技攻关项目“宁夏河东沙地退化草场植被恢复与风蚀沙化防治技术示范区(盐池)(2003BA517A)”资助
作者简介: 李娟(1982-), 女, 在读硕士研究生, 研究方向: 水文与水资源, 曾获宁夏回族自治区科学技术进步三等奖。

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

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, E_2, E_3, \dots , 因此

$$\sum_k P_{ik} = (P_{ik} = 0) \tag{1}$$

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, 具有两大缺点: 第一, 只考虑了最大概率, 忽略了其它概率的影响; 第二, 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。模糊集理论中的级别特征值可以有效地解决这个问题, 具体计算方法如下:
首先给各状态赋以相应的权重, 构成权重集 $D = \{d_1, d_2, d_3, d_4, \dots, d_m\}$, 其中 m 为研究系统的状态数。其中权重的大小取决于各状态概率的大小, 即

$$d_i = P_i^\eta / \sum_i P_i^\eta \tag{2}$$

式中: η ——最大概率的作用系数, 通常取 2。
级别特征值 H 可以通过式 (2- 3) 进行计算:

$$H = \sum_i i \times d_i \tag{3}$$

确定最大概率的状态 i 后, 可以根据式 (2~ 4) 确定系统在预报时段的预报值^[5]。

$$\begin{cases} X_{\text{预报}} = T H \ (i + 0.5) & H > i \\ X_{\text{预报}} = B H \ (i - 0.5) & H < i \end{cases} \tag{4}$$

其中: T, B 分别为状态区间值的上限与下限。

3 实例分析

以宁夏隆德县 1951~ 2004 年的降水序列为例, 讨论滑动平均- 马尔可夫链模型在降水预测中的应用, 具体过程如下:

3.1 对降水序列进行滑动平均计算, 并确定各时段滑动平均降水量状态

考虑到降水序列的连续性及降水特性, 本文取降水序列的 3 年滑动平均值进行降水预测。计算出降水序列的 3 年滑动平均值, 作为一个新的序列, 对该序列进行聚类。考虑水文现象的本身特性及序列数据的结构合理性, 将滑动平均降水量序列分为 5 类, 即将滑动平均降水量划分为 5 个区间, 具体结果见表 1。

表 1 滑动平均降水量等级划分表		
状态	等级	数值区间/mm
1	枯水年	$X < 449.3$
2	偏枯年	$449.3 \leq X < 494.3$
3	平水年	$494.3 \leq X < 584.4$
4	偏丰年	$584.4 \leq X < 629.9$
5	丰水年	$X \geq 629.9$

根据以上分级标准, 确定各时段滑动平均降水量的状态, 见表 2。

表 2 隆德县水文站滑动平均降水序列及状态表

时段	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
X /mm	452.6	355.3	405.3	559.7	493.8	633.4	448.2	491.3	566.9	515.3
X /mm	404.4	440.1	486.3	562.3	525.1	524.3	502.1	524.5	650.9	646.0
状态	1	1	2	3	3	3	3	3	5	5
时段	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970
X /mm	870.5	552.3	467.0	775.9	427.9	617.2	550.7	585.2	503.4	510.4
X /mm	629.9	598.4	556.9	607.0	531.9	584.4	546.4	533.0	464.6	454.4
状态	5	4	3	4	3	4	3	3	2	2
时段	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
X /mm	380.0	472.9	602.8	463.3	489.5	529.5	624.6	645.7	486.6	510.7
X /mm	485.2	513.0	518.5	494.1	547.9	599.9	585.6	547.7	510.0	454.8
状态	2	3	3	3	3	4	4	3	3	2
时段	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
X /mm	532.8	320.9	433.4	676.6	539.0	419.0	431.4	536.0	484.6	685.9
X /mm	429.0	477.0	549.7	544.9	463.1	462.1	484.0	568.8	510.1	555.1
状态	1	2	3	3	2	2	2	3	3	3
时段	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
X /mm	359.7	619.7	553.1	461.3	423.2	529.4	395.4	534.9	473.1	400.5
X /mm	510.8	544.7	479.2	471.3	449.3	486.6	467.8	469.5	467.1	449.6
状态	3	3	2	2	2	2	2	2	2	2
时段	2001	2002	2003	2004						
X /mm	527.7	420.7	674.4	543.0						
X /mm	540.9	546.0	/	/						
状态	3	3	/	/						

注: X 指时段的年降水量, X 指包括当前时段的 3 年滑动平均降水量

- 3.2 对各时段的滑动平均降水量进行预测
3.2.1 利用 1951~ 1997 年的滑动平均降水量对数预测 1998 年滑动平均降水量
(1) 根据表 2 中的状态, 计算 1951~ 1997 年的滑动平均降水量的各种步长的状态转移概率矩阵如下:

$P_1 =$	1/3	2/3	0/3	0/3	0/3
	1/13	8/13	4/13	0/13	0/13
	0/22	4/22	14/22	3/22	1/22
	0/5	0/5	4/5	1/5	0/5
	0/3	0/3	0/3	1/3	2/3
$P_2 =$	1/3	2/3	0/3	0/3	0/3
	6/12	6/12	6/12	0/12	0/12
	0/22	4/22	14/22	3/22	1/22
	0/5	0/5	4/5	1/5	0/5
	0/3	0/3	0/3	1/3	2/3
$P_3 =$	1/3	2/3	0/3	0/3	0/3
	0/11	3/11	8/11	0/11	0/11
	0/22	4/22	14/22	3/22	1/22
	0/5	0/5	4/5	1/5	0/5
	0/3	0/3	0/3	1/3	2/3
$P_4 =$	1/3	2/3	0/3	0/3	0/3
	1/10	2/10	8/10	0/10	0/10
	0/22	4/22	14/22	3/22	1/22
	0/5	0/5	4/5	1/5	0/5
	0/3	0/3	0/3	1/3	2/3
$P_5 =$	1/3	2/3	0/3	0/3	0/3
	0/9	3/9	5/9	1/9	0/9
	0/22	4/22	14/22	3/22	1/22
	0/5	0/5	4/5	1/5	0/5
	0/3	0/3	0/3	1/3	2/3

(2) 确定序列的各阶自相关系数并进行归一化作为权重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(1)进行计算。

$$r_k = \frac{\sum_{i=1}^{n-k} (x_{i+k} - \bar{x})(x_i - \bar{x})}{\sum_{i=1}^{n-k} (x_i - \bar{x})^2} \quad (5)$$

其中: x_i —— t 时段的滑动平均降水量; \bar{x} ——滑动平均降水量序列的均值; n ——滑动平均降水序列长度。对于降水序列, 一般计算5阶自相关系数即可。

根据式(5)确定1951~1997年的滑动平均降水量序列的各阶自相关系数如下: $r_1 = 0.633$, $r_2 = 0.327$, $r_3 = 0.023$, $r_4 = 0.081$, $r_5 = 0.071$ 。

对各阶自相关系数进行归一化, 作为各种滞时(步长)的滑动平均-马尔可夫链的权重, 即 $\omega_k = |r_k| / \sum_{k=1}^m |r_k|$ (m 为按预测需要计算到的最大阶数, 本文中为5)。

计算得, $\omega_1 = 0.558$, $\omega_2 = 0.288$, $\omega_3 = 0.020$, $\omega_4 = 0.071$, $\omega_5 = 0.063$ 。

根据1993~1997年的滑动平均降水量及其相应的状态转移概率矩阵对1998年的滑动平均降水量进行预测, 计算结果见表3。

表3 1998年滑动平均降水量预测表

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
1997	1	0.558	1/13	8/13	4/13	0/13	0/13
1996	2	0.288	0/12	6/12	6/12	0/12	0/12
1995	3	0.020	0/11	3/11	8/11	0/11	0/11
1994	4	0.071	0/10	2/10	8/10	0/10	0/10
1993	5	0.063	0/9	3/9	5/9	1/9	0/9
P_i 加权求和			0.043	0.528	0.422	0.007	0.000

由表3可知: 当 $i = 2$ 时, $P_2 = 0.528$ 为最大值, 说明1998年的滑动平均降水量状态为2, 即数值区间为(449.3, 494.3)。根据式(3)进行计算, 级别特征值为2.240, 根据式(4)计算, 求出1998的滑动平均降水量为442.38 mm, 而实测值为469.5 mm, 相对误差为5.69%。

3.2.2 利用1951~1998年的滑动平均降水量预测1999年滑动平均降水量

将1998年实际的滑动平均降水量加入到序列中, 用1994~1998年的滑动平均降水序列预测1999年的滑动平均降水量, 计算结果见表4。

表4 1999年滑动平均降水量预测表

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
1998	1	0.572	1/14	9/14	4/14	0/14	0/14
1997	2	0.287	0/13	7/13	6/13	0/13	0/13
1996	3	0.006	0/12	4/12	8/12	0/12	0/12
1995	4	0.067	0/11	3/11	8/11	0/11	0/11
1994	5	0.067	0/10	4/10	5/10	1/10	0/10
P_i 加权求和			0.041	0.564	0.387	0.007	0.000

由表4可知: 当 $i = 2$ 时, $P_2 = 0.564$ 为最大值, 说明1999年的滑动平均降水量状态为2, 即数值区间为(449.3, 494.3)。根据式(3)进行计算, 级别特征值为2.315, 根据式(4)计算, 求出1999的滑动平均降水量为457.6 mm。

3.2.3 预测2000年、2001年、2002年年的滑动平均降水量

同理, 写出2000年、2001年、2002年年的滑动平均降水

量预测表, 见表5、表6、表7。

表5 2000年滑动平均降水量预测表

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	0.480	1/15	10/15	4/15	0/15	0/15
1998	2	0.256	0/14	8/14	6/14	0/14	0/14
1997	3	0.108	0/13	5/13	8/13	0/13	0/13
1996	4	0.082	0/12	4/12	8/12	0/12	0/12
1995	5	0.074	0/11	5/11	5/11	1/11	0/11
P_i 加权求和			0.032	0.569	0.392	0.007	0.000

表6 2001年滑动平均降水量预测表

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	0.4473	1/16	11/16	4/16	0/16	0/16
1999	2	0.26	0/15	9/18	6/15	0/15	0/15
1998	3	0.079	0/14	6/14	8/14	0/14	0/14
1997	4	0.109	0/13	5/13	8/13	0/13	0/13
1996	5	0.103	0/12	6/12	5/12	1/12	0/12
P_i 加权求和			0.028	0.592	0.372	0.009	0.000

表7 2002年滑动平均降水量预测表

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	0.505	0/22	4/22	14/22	3/22	1/22
2000	2	0.259	0/16	9/16	7/16	0/16	0/16
1999	3	0.057	0/15	6/15	9/15	0/15	0/15
1998	4	0.091	0/14	5/14	9/14	0/14	0/14
1997	5	0.089	0/13	6/13	6/13	1/13	0/13
P_i 加权求和			0.000	0.334	0.568	0.076	0.023

3.3 模型检验

利用1998~2002年的实测滑动平均降水量对预测滑动平均降水量进行检验, 检验结果见表8。

表8 模型检验表

时段	预测状态	预测区间	实际状态	级别特征值	预测值	实测值	相对误差
	i	/mm	i	H	/mm	/mm	/%
1998	2	(449.3, 494.3)	2	2.243	442.38	469.5	5.69
1999	2	(449.3, 494.3)	2	2.315	457.60	467.1	2.04
2000	2	(449.3, 494.3)	2	2.293	453.32	449.6	0.80
2001	2	(449.3, 494.3)	3	2.290	452.3	540.9	16.38
2002	3	(494.3, 584.4)	2	2.763	546.21	546.0	0.38

由表8中可以看出, 在1998~2002年共计5年的预测中, 除了2001年的预测状态与实际状态不符外, 其余4年的预测状态皆与实际相符。尽管2001年的预测状态与实际有偏差, 但是实际的滑动平均降水量与实际相差不大, 与实际情况比较吻合; 从数值上讲, 1998年、1999年、2000年、2002年的预测相对误差都较小, 在0.38%~5.69%, 只有2001年的预测相对误差较大, 但是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^[6-8], 因此预测误差满足预测精度需要, 以此证明本文的方法进行降水量的中长期预报是有效可行的。根据以上预测模型, 可以预测2003年的滑动平均降水量为567.2 mm。

4 结论

(1) 与传统的马尔可夫过程相比, 滑动平均-马尔可夫模型具有以下特点:

(下转第205页)

场的围封和人工灌溉草场产草暂时较低,经济效益虽然能维持现状,但草地生态还未恢复。在建设期的第2~3年,灌溉人工草地的全额产草,使得载畜量上升到4 050个羊单位,相应的净效益也会上达到60 67万元,草地生态还未达到最大恢复潜力,在项目实施3年后,载畜量可上升到4 329个羊单位,净效益会上升到64 93万元,草地生态达到了最大恢复潜力(草地植被盖度75%,干草产量达1 200 kg/hm²),研究区域人均纯收入将由现在的2 210元上升到4 637元,真正实

现生态效益与经济效益的双赢。

3 小 结

本研究是在对沙、退化草地生态恢复与重建的各单项技术(草地封育、补播改良,灌溉人工牧草,灌溉青贮玉米等)进行了综合效益试验研究的基础上,应用目标规划优化方法求解得到,在确定措施和参数方面有待调整完善,模式尚需进行实地验证,提高其实用性。

(上接第195页)

从表2可以看出,对同一场洪水模拟来说,两种DEM 情况下模拟结果相似,也就说明了在皎口水库流域 SRTM DEM 作为基础地形数据用于TOPMODEL 模型时可获得较好的模拟结果。对单一流域的SRTM DEM 在TOPMODEL 模型中的可用性评价不能得出一般性的规律,应对若干典型流域进行分析并得出相应的结论。但Raaflaub(2002)⁶研究发现DEM 中的误差对TOPMODEL 影响甚微,Franchini等^[7](1996)发现TOPMODEL 模型的模拟结果对地形指数-面积曲线的形状并不敏感,以及距流域出口距离-累积面积数据的不同可用模型中参数来调节或弥补,因此可以得出SRTM DEM (水平分辨率约90 m)数据可用于TOPMODEL

模型模拟,并能取得较好的模拟效果。

3 结论与讨论

从以上可以得出,SRTM DEM 水平分辨率约90 m 的数据能够用于数字高程模型难以获取的大中型流域的TOPMODEL 模型的水文模拟,并能取得较好的模拟效果。SRTM DEM 的出现将会拓宽TOPMODEL 的应用范围和区域,由于受其水平分辨率的限制,其不适用于小型流域。由于不同的水文模型DEM 的作用不同,SRTM DEM 对其它基于DEM 的水文模型来说可用性如何需进一步研究。

参考文献:

- [1] 熊立华,郭生练 分布式流域水文模型[M] 北京:中国水利水电出版社,2004
- [2] Band L E,Moore ID. Scale: landscape attributes and geographical information system s[J] Hydrological process, 1995, 9: 401- 422
- [3] 郭方,刘新仁,任立良 以地形为基础的流域水文模型- TOPMODEL 及其拓宽应用[J] 水科学进展,2000,11(3): 296 - 301.
- [4] 陈仁升,康尔泗,杨建平,等 TOPMODEL 模型在黑河干流出山径流模拟中的应用[J] 中国沙漠,2003,23(4): 428- 434
- [5] Quinn P, Beven K J, Lamb R. The $\ln(\alpha/\tan\beta)$ index: how to calculate it and how to use it in the TOPMODEL framework[J] Hydrologic Process, 1994, 9: 161- 185
- [6] L D Raaflaub The effect of error in gridded digital elevation models on topographic analysis and on the distributed hydrological model Topmodel[D] Department of Geomatics Engineering, The University of Calgary, 2002
- [7] Franchini M, Wendling J, Obled C, et al Physical interpretation and sensitivity analysis of the TOPMODEL [J] Journal of Hydrology, 1996, 175: 293- 338

(上接第198页)

采用了滑动平均的思想,预测未来某时段的滑动平均降水量,预测范围扩大,预测精度提高。

采用了聚类的方法对滑动平均降水量进行等级划分,更加充分地考虑了序列中数据结构的合理性,更加有效地显示了降水量序列的内在分布规律,使得区间的划分更加合理。

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] Sidney, Yakowitz Markov flow models and the flood warning problem [J] Water Resources Research, 1985, 21(1): 81 - 83
- [3] 张曙红,曹建会,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
- [8] 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.

计算序列的各阶自相关系数,归一化后作为马尔可夫模型的权重,充分、合理地利用已有信息,使得预测结果更为精确。

(2)根据滑动平均-马尔可夫模型,预测2003年的3年滑动平均降水量为567.2 mm。