

基于加权的滑动平均—马尔科夫预测模型及其应用

秦 松

(中煤科工集团 重庆设计研究院, 重庆 400016)

摘 要: 降水量预测是雨水潜力化计算的首要环节。但由于降水过程存在高度的不确定性和随机性, 很难用物理成因方法来确定某一时段降水量的准确值。因此可采用基于概率论和随机过程理论的马尔科夫模型进行预测^[13], 采用加权的方法, 并通过对降水序列的滑动平均处理, 降低序列的随机性, 提高预测的准确性。以简阳 1953—2004 年年降水资料为分析对象进行实例分析, 其中应用 2006—2009 年降水资料作模型检验, 并对 2010 年—2005 年年降水量作模型预测。结果表明: 应用加权的滑动平均马尔科夫预测模型进行降水量预测是可行的, 而且意义明确, 计算简便, 预测精度较高。

关键词: 降水量; 马尔科夫预测模型; 滑动平均; 降雨量预测

中图分类号: TV125

文献标识码: A

文章编号: 1672-643X(2013)01-0185-04

Markov prediction model and its application based on the weighted sliding average

QIN Song

(Chongqing Design and Research Institute, China Coal Technology & Engineering Group, Chongqing 400016, China)

Abstract: The precipitation forecast is the primary link in rainwater potential calculation. However, there is a high degree of uncertainty and randomness in precipitation process, it is difficult to determine the exact value of the precipitation in certain period of time with physical methods. Therefore Markov model can be used to predict based on probability theory and random process theory through using the method of weighted and sliding average treatment of precipitation sequence so as to reduce the randomness of the sequence and improve the prediction accuracy. Case study made the analysis of Jianyang precipitation data from 1953 to 2004 in which application of precipitation data from 2006 to 2009 was used to do model checking. Therefore, model predicts applying this model of annual precipitation from 2010 to 2015 was done. The results shows that precipitation forecast is feasible applying weighted moving average Markov prediction model. And the method is meaningful with simple calculation and high prediction accuracy.

Key words: precipitation; Markov prediction model; sliding average; precipitation prediction

0 引 言

降水预测是水资源预测和水文预报中的一项经常性工作, 也是雨水资源化潜力计算的首要环节。由于降水的变化是多种影响因素促动的结果, 而这些自然因素又存在着变异性、多样性和复杂性, 因此使降水过程存在高度不确定性和随机性, 并且很难用物理成因来确定出将来某一时段内的降水量准确值。目前预测降水量的方法很多^[1-5], 但各有其优势和不足。马氏链预测则以马尔科夫过程为理论基础, 根据状态之间的转移概率来推测系统未来发展变化, 较适合于随机波动性较大的预测问题。

考虑到降水序列的连续性和随机性, 可采用基

于概率论和随机过程理论的马尔可夫模型进行预测^[6-9], 并通过对降水序列的滑动平均处理, 降低序列的随机性, 提高预测的准确性。本文以简阳 1953—2009 年年降水资料^[10]为分析对象, 应用加权的滑动平均马尔科夫预测模型进行降水量预测。

1 加权滑动平均马尔科夫模型构建

1.1 原始数据序列的滑动平均处理

设有原始数据序列 $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$, 对原始数据序列进行滑动平均处理: $X_i^{(0)} = (x_1^{(0)}(1), x_1^{(0)}(2), \dots, x_1^{(0)}(n))$, 其中:

$$x_1^{(0)}(1) = \frac{1}{4}(3x^{(0)}(1) + x^{(0)}(2))$$

收稿日期: 2012-10-16; 修回日期: 2012-12-01

作者简介: 秦松(1983-), 男, 重庆人, 硕士, 工程师, 主要从事环境影响评价与资源环境保护研究。

$$x_1^{(0)}(i) = \frac{1}{4}(x^{(0)}(i-1) + 2x^{(0)}(i) + x^{(0)}(i+1))$$

$$x_1^{(0)}(n) = \frac{1}{4}(x^{(0)}(n-1) + 3x^{(0)}(n)) \quad (1)$$

其中, $i=1, 2, \dots, n$; $X_1^{(0)}$ 是对 $X^{(0)}$ 的一次滑动平均处理。

1.2 加权滑动平均-马尔可夫模型的建立过程

步骤 1: 对时间序列值进行滑动平均处理, 并计算出新序列的平均值和均方差。

步骤 2: 以样本均方差和均值为划分依据, 将指标值进行分级, 即确定马尔可夫状态空间, 并对新序列进行聚类。

步骤 3: 按照第 2 步所述分级标准, 明确新序列各时段指标值的对应状态, 并对所得结果进行统计, 得出不同步长马尔可夫链的转移概率矩阵。

步骤 4: 计算各阶步长下的自相关系数, 计算公式如下:

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

式中: x_i 为第 i 时段的降水量, mm; $\bar{x}$ 为新序列的均值; n 代表降水量序列长度。

在计算出各阶自相关系数后进行归一化处理, 并将处理结果值作为各种步长的马尔可夫链的权重。归一化公式如下:

$$w_k = \left| r_k / \sum_{k=1}^n |r_k| \right| \quad (3)$$

步骤 5: 以前面若干时段的降水量值作为初始状态, 再联系其相应的状态转移概率, 计算出当前时段降水量的状态概率 $P_i^{(k)}$, i 为状态, $i \in R$, k 为步长, $k=1, 2, \dots, m$ 。

步骤 6: 按照(4)式将同一状态各预测概率加权作为降水量处于该状态的预测概率。

$$P_i = \sum_{i=1}^m w_k P_i^{(k)} \quad (4)$$

依据最大隶属度原则, 当存在 $\max\{P_i, i \in I\} > 0.5$ 时, 则 $\max\{P_i, i \in I\}$ 所对应的状态 i 定为该时段降水量所处的状态; 否则就用级别特征值法来确定其状态, 当有 $-0.5 < H - i < 0.5$ 存在时, 则判定该时段的降水量所处的状态为 i 。

步骤 7: 计算级别特征值和降水量预测值。可按式(5)计算级别特征值:

$$D_k = P_k^\eta / \sum_{k=1}^5 P_k^\eta \quad (5)$$

式中: η 为最大概率作用系数, 在实际应用中通常取为 2。

定义 H 为级别特征值, 其值可由(6)式确定:

$$H = \sum_{i=1}^5 i \times D_i \quad (6)$$

预测值确定: 当依据最大概率原则确定的状态为 i , 同时存在 $H > i$ 时, 则按 $T_i H_i / (i + 0.5)$ 确定预测值; 若 $H < i$ 存在时, 则按 $B_i H_i / (i - 0.5)$ 确定预测值。其中, T_i, B_i 为状态 i 的区间值的上下限值。

1.3 系统新信息的优先应用

系统在变化过程中经常会受一些随机扰动和驱动的影响, 从而改变预测精度。而通常离原点数据最近的 1~2 个数据其受扰动最小, 预测精度较高。因此, 应用灰色新陈代谢预测模型, 即将最新的模型预测值补充到原始数据序列, 同时去掉一个起始数据, 这样依次递补, 直至预测完毕。

2 实例应用: 四川省简阳市降水量预测

依据简阳市 1953-2009 年气象统计资料得到表 1 中的年降雨量数据。

按照本文 1.1 和 1.2 节内容的基于滑动平均-马尔可夫模型的降水量预测步骤:

步骤 1: 计算年降水系列三年滑动平均值。

依据年降水系列的连续性, 应用简阳市 1953 年-2009 年共计 57 年的降水序列的三年滑动平均值作为一个新的降水序列, 并对该序列进行聚类。

步骤 2: 计算新序列的均值和均方差。依据以下两式计算新序列的均值和均方差:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (7)$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (8)$$

通过相关计算, 新序列的均值 $\bar{x} = 826.37$ mm, 标准差为 $s = 114.15$ mm。

步骤 3: 应用均值均方差法对降水序列进行分级聚类。

依据水文学原理, 降水量大小一般分成 5 个等级状态, 即 1. 枯水年, 2. 偏枯年, 3. 平水年, 4. 偏丰年, 5. 丰水年。应用均值均方差法就可以将新序列分成 5 个区间, 而每一个区间又对应着一个区间状态, 即马尔可夫状态, 分级结果参见表 2。依据表 2 的分级结果, 可区分出新降水序列所处的状态。

表 1 四川省简阳市 1953 - 2009 年年降雨量数据 mm

序号	年份	Q	序号	年份	Q
1	1953	793.60	30	1982	937.80
2	1954	978.25	31	1983	715.72
3	1955	732.76	32	1984	1023.40
4	1956	1275.12	33	1985	628.08
5	1957	732.05	34	1986	611.41
6	1958	783.02	35	1987	804.59
7	1959	876.31	36	1988	1041.07
8	1960	822.76	37	1989	763.24
9	1961	1210.33	38	1990	766.79
10	1962	1106.60	39	1991	719.20
11	1963	987.81	40	1992	850.27
12	1964	880.43	41	1993	680.23
13	1965	1057.27	42	1994	496.65
14	1966	1026.70	43	1995	774.63
15	1967	916.71	44	1996	498.46
16	1968	729.64	45	1997	613.34
17	1969	713.77	46	1998	1023.27
18	1970	1166.32	47	1999	706.70
19	1971	839.55	48	2000	559.70
20	1972	726.20	49	2001	1100.30
21	1973	568.93	50	2002	625.20
22	1974	952.31	51	2003	591.50
23	1975	1158.85	52	2004	752.50
24	1976	994.46	53	2005	1046.80
25	1977	765.07	54	2006	567.80
26	1978	910.11	55	2007	659.50
27	1979	720.74	56	2008	889.80
28	1980	685.50	57	2009	785.40
29	1981	772.39			

表 2 简阳市 3 年滑动平均年降水量状态表

年份	状态	年份	状态	年份	状态	年份	状态	年份	状态
1953	3	1965	5	1977	3	1989	2	2001	3
1954	5	1966	4	1978	3	1990	3	2002	1
1955	4	1967	3	1979	2	1991	2	2003	3
1956	4	1968	3	1980	3	1992	1	2004	3
1957	3	1969	4	1981	3	1993	1	2005	2
1958	3	1970	4	1982	4	1994	1	2006	1
1959	5	1971	1	1983	3	1995	1	2007	3
1960	5	1972	2	1984	2	1996	1	2008	
1961	5	1973	4	1985	1	1997	3	2009	
1962	5	1974	5	1986	3	1998	2		
1963	5	1975	5	1987	3	1999	3		
1964	5	1976	4	1988	3	2000	2		

注:1. 枯水年; 2. 偏枯年; 3. 平水年; 4. 偏丰年; 5. 丰水年。

依据表 2 的年降水量状态得出相应的状态转移概率矩阵:

$$P_1 = \begin{bmatrix} 4/9 & 1/9 & 4/9 & 0 & 0 \\ 3/8 & 0 & 1/2 & 1/8 & 0 \\ 1/20 & 7/20 & 7/20 & 2/20 & 2/20 \\ 1/8 & 0 & 1/2 & 1/4 & 1/8 \\ 0 & 0 & 0 & 3/10 & 6/10 \end{bmatrix}$$

$$P_2 = \begin{bmatrix} 1/3 & 1/9 & 2/9 & 1/9 & 0 \\ 1/4 & 1/4 & 3/8 & 0 & 1/8 \\ 1/10 & 1/10 & 3/10 & 1/5 & 1/10 \\ 1/8 & 1/4 & 1/2 & 0 & 1/8 \\ 0 & 0 & 1/5 & 3/10 & 1/2 \end{bmatrix}$$

$$P_3 = \begin{bmatrix} 2/9 & 2/9 & 1/3 & 0 & 1/9 \\ 1/4 & 0 & 1/4 & 1/8 & 1/8 \\ 1/10 & 1/5 & 1/4 & 1/10 & 1/10 \\ 1/8 & 1/4 & 1/8 & 3/8 & 1/8 \\ 0 & 0 & 2/5 & 1/10 & 2/5 \end{bmatrix}$$

$$P_4 = \begin{bmatrix} 2/9 & 2/9 & 2/9 & 0 & 1/9 \\ 3/8 & 0 & 1/4 & 1/8 & 0 \\ 1/5 & 1/5 & 7/20 & 1/20 & 1/10 \\ 0 & 0 & 3/8 & 1/4 & 3/8 \\ 0 & 1/10 & 2/5 & 1/5 & 3/10 \end{bmatrix}$$

$$P_5 = \begin{bmatrix} 0 & 2/9 & 5/9 & 1/9 & 0 \\ 1/4 & 3/8 & 1/4 & 0 & 0 \\ 3/10 & 1/10 & 3/10 & 1/10 & 1/10 \\ 1/8 & 0 & 3/8 & 0 & 1/2 \\ 0 & 1/10 & 3/10 & 3/10 & 3/10 \end{bmatrix}$$

步骤 4: 计算新序列的自相关系数并归一化。

按照(2) 式,计算新降水序列的自相关系数,其中 $r_1 = 0.744, r_2 = 0.492, r_3 = 0.321, r_4 = 0.418, r_5 = 0.443$, 归一化处理后得 $\bar{\omega}_1 = 0.308, \bar{\omega}_2 = 0.203, \bar{\omega}_3 = 0.133, \bar{\omega}_4 = 0.173, \bar{\omega}_5 = 0.183$ 。

第 5 步: 确定年降水状态。以判断 2006 年的年降水状态为例,依据 2001 - 2005 年滑动平均降水量及其所对应的状态转移概率矩阵为依据,对 2006 年的年降水状态进行判断见表 3,判断结果参见表 2。其中表 3 中的步长是根据预测年份与数据资料所处年份的间隔确定的,并且将步长与相应的状态概率转移矩阵相对应。

判定规则:

(1) 若有 $\max \{P_i, i \in I\} > 0.5$ 存在时,则按照最大隶属度原则取 $\max \{P_i, i \in I\}$ 对应着的 i 作为判断时段降水量所处的状态,否则采用级别特征值法来判断。

(2) 级别特征值判断法:若有 $-0.5 < H - i < 0.5$ 存在时,则判定该时段降水量所处状态为 i 。通过上表中的数值计算可知,2006 年的降水状态为 1,这与表中的结果是一致的,可知 2006 年的降雨状态是 i 。

表 3 依据 2001 - 2005 年数据判断 2006 年
降水状态转移概率计算表

年份	对应 状态	步长	权重	状态转移概率				
				1	2	3	4	5
2005	2	1	0.308	5/8	0	1/2	1/8	0
2004	3	2	0.203	1/10	1/10	3/10	1/5	1/10
2003	3	3	0.133	1/10	1/5	1/4	1/10	1/10
2002	1	4	0.173	2/9	2/9	2/9	0	1/9
2001	3	5	0.183	3/10	1/10	3/10	1/10	1/10
加权求和				0.372	0.104	0.342	0.111	0.071

第 6 步:年降水量预测。应用级别特征值求解公式,求出预测时间段的降水级别特征值,并根据预测值计算公式计算出预测值。将预测值与实测值进行比较,求解出相对误差值,并对模型精度进行评估。例如,2006 年的预测值为 559.9 mm,相对误差为 1.4%。按上述方法同理也可预测出 2006 年以后的年降水量值。

3 模型检验

用基于加权的滑动平均 - 马尔科夫预测模型对 2006 - 2009 年的年降水量进行预测,得到表 4 的预测结果,和实测值相比,其预测精度参见表 4。

表 4 简阳市 2006 - 2009 年年降水量预测值
及相对误差 mm, %

年份	实测值	预测值	相对误差
2006	567.80	559.90	1.40
2007	659.50	665.76	0.95
2008	889.80	942.30	5.90
2009	785.40	748.48	4.70

从预测结果和实测值的相对误差来看,还是能够满足实际需要的。继续对简阳 2010 - 2015 年年降水量作出预测,预测结果参见表 5。

表 5 简阳市 2011 - 2015 年年降水量预测值 mm

年 份	2010	2011	2012	2013	2014	2015
降水量	1089.4	802.6	793.6	776.7	912.5	698.5

4 结 语

基于加权的滑动平均 - 马尔科夫预测模型是一个计算简便、合理、预测精度较高的预测方法。用该方法预测四川简阳市 2006 - 2009 年的年降水量,并和实际值对照进行模型精度检验,其预测精度完全满足需要,同时本文应用该方法对简阳市 2010 - 2015 年年降水量进行了预测,其预测精度完全满足实际需要的。但应用该方法只能确定预测年所处的状态,不能确定确定预测年的准确预测值。要想获得准确预测值,还需结合其他方法进行准却确定。

参考文献:

[1] 刘玉邦. 农业水资源高效利用理论及其在川中丘陵区的应用研究[D]. 成都:四川大学,2011.

[2] 韦 庆,卢文喜,田竹君. 运用蒙特卡罗方法预报年降水量研究[J]. 干旱区资源与环境,2004,18(4):144 - 146.

[3] 马占青,徐明仙,俞卫阳,等. 年降水量统计马尔科夫预测模型及其应用[J]. 自然资源学报,2010,25(6):1033 - 1041.

[4] 冯利华. 应用灰色聚类分析做降水趋势预报的探讨[J]. 地域研究与开发,2003,22(1):10 - 13.

[5] 仲远见,李 靖,王 龙. 改进马尔可夫链降雨量预测模型的应用[J]. 济南大学学报(自然科学版),2009,23(4):402 - 405.

[6] 杨国巍. 基于小波分析预测月径流量的贝叶斯(BPF)预报方法[J]. 水利科技与经济,2011,17(8):53 - 55.

[7] 王 蓓,刘玉甫. 滑动平均 - 马尔科夫模型在降水预测中应用[J]. 水资源研究,2009,30(2):25 - 27.

[8] 钱 会,李培月,王 涛. 基于滑动平均 - 加权马尔科夫链的宁夏石嘴山市年降雨量预测[J]. 华北水利水电学院学报,2010,31(1):6 - 9.

[9] 张国帅. 基于累积法的灰色马尔科夫预测模型及其应用[J]. 统计与决策,2011(8):157 - 158.

[10] 四川省农业科学院节水农业课题组. 四川节水农业课题 2009 年工作总结[Z]. 成都:四川农业节水农业课题 2009 年工作总结会议,2009.