· 水文地质 ·

# 马尔可夫链模型在地下水水位预测中的应用

宋印胜
(山东省鲁南地质工程勘察院)

**提要**　作者将鲁南地区一水文观测孔的 92 次地下水水位月平均监测数据, 划分了 10 种状态范围, 运用马尔可夫链模型, 对未来地下水水位进行状态范围预测。在与实际监测资料进行对比的基础上, 验证了马尔可夫链理论在地下水水位预测中的可行性及可靠性, 并对运用中的具体问题进行了探讨。

**关键词**　马尔可夫链　地下水水位预测　状态划分　状态概率　转移概率

## 0　引言

马尔可夫过程(Markov Process)是研究事物状态及状态转移的理论。它通过对事物所处不同状态的初始概率以及状态之间的转移概率关系, 来确定事物所处状态的变化趋势, 从而达到预测的目的。马尔可夫(Markov. A. A)于 1907 年用数学的方法分析了布朗运动的随机过程, 后来称之为马尔可夫过程。40 年代A. N. 卡尔曼哥隆等人又发展了马尔可夫理论, 到目前, 这个理论已广泛用于气象、生态及经济学等领域的预测研究[1,2,3]。

## 1　马尔可夫链预测模型

马尔可夫过程是较普遍随机过程的一种, 该过程考虑了以前事件对后来事件的影响, 即从一种状态转移到另一种状态, 随时间变化所作的状态转移, 且状态转移具有概率性质。时间离散、状态离散的马尔可夫过程称为马尔可夫链。对一重(一阶)平稳的马尔可夫链, 系统每次转移, 仅依赖于前一次的状态 i, 与更前一次的状态 i-1 无关, 且这个概率与几次转移无关[4]。

在马尔可夫链中, 系统状态的转移可用概率矩阵 $P$ 表示:

$$P=\begin{pmatrix} P_{11}, P_{12}, \cdots P_{1m} \\ P_{21}, P_{22}, \cdots P_{2m} \\ \vdots \quad \vdots \quad \vdots \\ P_{m1}, P_{m2}, \cdots P_{mm} \end{pmatrix}$$

$t$ 到 $t+1$ 时刻, 状态从 $S_i$ 转移为 $S_j$ 的频数 $n_{ij}$ 与总频数 $n$ 之比(频率), 则为状态 $S_i$ 转

本文 1997 年 5 月收到, 1998 年 1 月改回, 张天祯编辑。

移为 $S_i$ 的转移概率:

$$P_{ij}=\frac{n_{ij}}{n} \tag{1}$$

从任何一个状态出发, 经过一次转移, 必然出现该系统所有状态中的一个, 故 $\sum_{j=1}^{m}P_{ij}=1$, 其中系统停留状态 i 的概率 $P_{ii}$ 包含在内。由于 $P_{ij}$ 是概率, 则 $0\leq P_{ij}\leq 1$, i, j= 1, 2, …m。

设初始(0 步)处于状态 $S_i$ 的概率为 $A_i(0)$, 从状态 $S_i$ 转移到状态 $S_j$ 的概率为 $P_{ij}$, 在第一步处于的状态 $S_j$ 的概率为 $A_j(1)$, 按照贝叶斯(Bayes)公式:

$$P(A,B)=P(A)P(B/A) \tag{2}$$

则有:

$$A_j(1)=A_j(0)P_{ij} \tag{3}$$

若从第 $n$ 步推广到 $n+1$ 步便有:

$$A_j(n+1)=A_j(n)P_{ij} \tag{4}$$

若研究的事物在 $n$ 步有 $m$ 种状态, 则在 $n+1$ 步可能转入状态 $S_j$ 的概率为:

$$\begin{aligned}A_j(n+1)&=A_1(n)P_{1j}+A_2(n)P_{2j}+\cdots+A_m(n)P_{mj}\\&=\sum_{j=1}^{m}A_j(n)P_{ij}\end{aligned} \tag{5}$$

如果考虑到从零步到第一步转移过程, j= 1, 2, …$m$; 由(5)式得:

$$\begin{aligned}&j=1, A_1(1)=A_1(0)P_{11}+A_2(0)P_{21}+\cdots+A_m(0)P_{m1}\\&j=2, A_2(1)=A_1(0)P_{12}+A_2(0)P_{22}+\cdots+A_m(0)P_{m2}\\&\quad\vdots\\&j=m, A_m(1)=A_1(0)P_{1m}+A_2(0)P_{2m}+\cdots+A_m(0)P_{mm}\end{aligned}$$

上述方程组表示为:

$$\begin{aligned}&[A_1(1),A_2(1),\cdots A_m(1)]\\&=[A_1(0),A_2(0),\cdots A_m(0)]\cdot\begin{pmatrix}P_{11},P_{12},\cdots P_{1m}\\P_{21},P_{22},\cdots P_{2m}\\ \vdots\\P_{m1},P_{m2}\cdots P_{mm}\end{pmatrix}\end{aligned}$$

令 $A_1(1)=[A_1(1),A_2(1),\cdots A_m(1)]$, $A(0)=[A_1(0),A_2(0),\cdots A_m(0)]$

则

$$A_1(1)=A(0)P \tag{6}$$

从 $n$ 步推广到 $n+1$ 步得:

$$A(n+1)=A(n)P \tag{7}$$

可以导出:

$$\begin{aligned}&n=0\text{ 时}, A(1)=A(0)P\\&n=1\text{ 时}, A(2)=A(1)P=A(0)P^2\\&n=2\text{ 时}, A(3)=A(2)P=A(0)P^3\\&\cdots\\&n=n-1\text{ 时}, A(n)=A(0)P^n\end{aligned} \tag{8}$$

(8)式称为马尔可夫链预测模型, 它表示只要知道状态转移概率矩阵, 就可以根据初始时刻处于各状态的概率来预测以后任一时刻各状态的概率。

随着过程的持续发展, 初始阶段的影响将逐步消失, 系统在 $n$ 时刻处于 $S_j$ 的概率与

初始状态无关,仅决定于转移概率矩阵。当 $n\to\infty$,绝对概率分布 $P^{(n)}$ 收敛于一个独立的初始分布 $P^{(0)}$ 的极限概率 $P$,$\lim\limits_{n\to\infty}P^{(0)}=P$,这时系统状态趋于一个稳定状态。

## 2 独立假设检验

运用马尔可夫链理论,首先要判断转移概率矩阵,是否具有马尔可夫链性质,因而需要从统计上进行假设检验。假如研究的事物之间是相互独立的,设计一个检验这种假设的统计量 λ。

$$\lambda=\prod_{ij}^{m}\left(\frac{P_{\mathrm{i}}}{P_{\mathrm{ij}}}\right)^{n_{\mathrm{ij}}}$$

又可写成便于计算的形式:

$$-2\mathrm{Ln}\lambda=2\sum_{i=1}^{m}\sum_{j=1}^{m}n_{\mathrm{ij}}\mathrm{Ln}\left(\frac{P_{\mathrm{ij}}}{P_{\mathrm{j}}}\right)\tag{9}$$

这里 $-2\mathrm{Ln}\lambda$ 近似于带有 $(m-1)^2$ 自由度的 $\chi^2$ 分布,$n_{\mathrm{ij}}$ 为转移频数矩阵中 i 行 j 列元素,$P_{\mathrm{ij}}$ 为转移概率矩阵中 i 行 j 列元素,其中:

$$P_{\mathrm{j}}=\frac{\sum\limits_{i=1}^{m}n_{\mathrm{ij}}}{\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{m}n_{\mathrm{ij}}}\tag{10}$$

若 $-2\mathrm{Ln}$ 计算值大于根据自由度 $(m-1)^2$ 适当选取置信度下 $\chi^2$ 分布表值,则假设被拒绝,接受相反的假设,说明研究事物状态之间具有马尔可夫链性质。

## 3 地下水水位动态预测实例

鲁南地区一山前冲洪积平原下游地下水水位监测孔(217 号),孔深 150m,开采并观测 50~150m,微承压水。50m 以上弱透水层岩性为粉土、粉质粘土和粘土,含水层岩性以细砂和中细砂为主。50~150m 段,含水层由 7 至 9 层中、粗砂组成,总厚度为 36~42m,相对隔水层为粉质粘土和粘土;地下水水位动态属入渗—开采型;单井涌水量为 1800~2500m³/d;地下水靠大气降水、河流侧渗及来自山前倾斜平原的侧向径流补给;工业及生活用水为主要排泄方式。潜水与微承压水之间水力联系密切,存在"天窗式"补给,即潜水补给微承压水。根据地下水水位监测资料(表 1),在 1985 年 1 月至 1992 年 8 月份的监测资料中,地下水月平均水位最高为 34.333m(黄海高程,下同),最低为 24.574m。因地下水在补给和开采状态的时空上有很大离散性,从而显示出地下水水位有较强的随机特征。

为了符合马尔可夫链预测要求,把地下水水位由低到高划分为 10 个状态级别标准(表 2),并依此把每月的监测资料进行状态划分和统计,先统计各状态级别出现的总次数,然后统计每个状态级别后面出现各状态级别的次数。显然在 92 次监测数据中,出现水位< 26m 的有 4 次,水位在 26~27m 的为 9 次,……。在 1 级状态(< 26m)后转为 1 级状态的有2次,转为2级状态的有1次,无转移1次(1992年8月);2级状态9次中,其

表2 状态划分标准表

Table 2 Standard values of state division

| 状态级别 | 水位特征(m) | 出现频数 | 注 |
|---|---|---|---|
| 1 | < 26 | 4 | 无转移1个 |
| 2 | 26~27 | 9 | |
| 3 | 27~28 | 8 | |
| 4 | 28~29 | 14 | |
| 5 | 29~30 | 13 | |
| 6 | 30~31 | 10 | |
| 7 | 31~32 | 12 | |
| 8 | 32~33 | 6 | |
| 9 | 33~34 | 14 | |
| 10 | > 34 | 2 | |

表3 地下水水位各状态转移统计表

Table 3 Statistics of state transfer of underground water levels

| 级别 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Σ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| 2 | 2 | 6 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 9 |
| 3 | 0 | 2 | 4 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 8 |
| 4 | 0 | 0 | 4 | 8 | 1 | 0 | 0 | 1 | 0 | 0 | 14 |
| 5 | 0 | 0 | 0 | 4 | 6 | 2 | 0 | 0 | 1 | 0 | 13 |
| 6 | 0 | 0 | 0 | 0 | 4 | 4 | 2 | 0 | 0 | 0 | 10 |
| 7 | 0 | 0 | 0 | 0 | 1 | 4 | 6 | 1 | 0 | 0 | 12 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 2 | 0 | 0 | 6 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 11 | 1 | 14 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 2 |

由表3及(1)式得转移概率矩阵。

$$P=\begin{pmatrix} 0.667 & 0.333 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.222 & 0.667 & 0 & 0 & 0.111 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.250 & 0.500 & 0.250 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.286 & 0.571 & 0.071 & 0 & 0 & 0.071 & 0 & 0 \\ 0 & 0 & 0 & 0.308 & 0.462 & 0.154 & 0 & 0 & 0.077 & 0 \\ 0 & 0 & 0 & 0 & 0.400 & 0.400 & 0.200 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.083 & 0.333 & 0.500 & 0.083 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.667 & 0.333 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.143 & 0.786 & 0.071 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.500 & 0.500 \end{pmatrix}$$

中转为1级状态的2次, 转为2级状态的为6次, 转为5级状态的为1次, ……。仿此便得到地下水水位状态转移频数矩阵(表3)。统计检验, $\sum_{i=1}^{m}\sum_{j=1}^{m}n_{ij}=91$, 则:

$P_j=(0.033, 0.099, 0.088, 0.154, 0.143, 0.110, 0.132, 0.066, 0.154, 0.022)$

由计算得$-2Ln\lambda=226.506$, 地下水水位为10个状态, 序列自由度为81, 当选择置信水平$\alpha=0.05$时, 查$x^2$分布表得103, 显然$226.506>103$, 可以拒绝独立假设, 说明地下水水位各状态之间不独立, 具有马尔可夫链性质。

$$\frac{P_{ij}}{P_j}=\begin{pmatrix} 20.182 & 3.364 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 6.727 & 6.737 & 0 & 0 & 0.776 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2.525 & 5.682 & 1.623 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3.250 & 3.708 & 0.497 & 0 & 0 & 1.076 & 0 & 0 \\ 0 & 0 & 0 & 2.000 & 3.231 & 1.400 & 0 & 0 & 0.500 & 0 \\ 0 & 0 & 0 & 0 & 2.797 & 3.636 & 1.515 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.580 & 3.027 & 3.788 & 1.258 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 5.053 & 5.046 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.167 & 5.104 & 3.227 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3.247 & 22.727 \end{pmatrix}$$

地下水水位高低程度用 10 种状态(等级)刻划, 各状态出现的可能性大小用状态概率 $A_i(t)$ 表示, $i=1,2\cdots10$; $t$ 为预测月份刻度。各状态对应的状态概率构成状态概率向量, $A_t=(A_1(t), A_2(t), \cdots A_{10}(t))$。该水位预测基准($t=0$)的水位处于 1 级, 其它各级状态概率均为 0, 则初始状态概率向量 $A(0)=(1,0,0,0,0,0,0,0,0,0)$。为了预测 1992 年 9 月至 12 月份该监测孔可能出现的水位状态, 则需要计算各预测月的状态概率向量 $A(t)$, $t=1,2,3,4$。由(8)式可递推 1992 年 9 月份的水位状态概率向量为:

$$\begin{aligned} A(1)=A(0)P &= (1,0,0,0,0,0,0,0,0,0)\cdot P \\ &= (0.667, 0.333, 0, 0, 0, 0, 0, 0, 0, 0) \end{aligned}$$

仿此可计算出其它各月的月平均水位状态概率向量, 即:

$$A(2)=A(1)P=(0.519, 0.444, 0, 0, 0.037, 0, 0, 0, 0, 0)$$

$$A(3)=A(2)P=(0.445, 0.461, 0, 0.011, 0.066, 0.006, 0, 0, 0.003, 0)$$

$$A(4)=A(3)P=(0.401, 0.461, 0.003, 0.027, 0.086, 0.013, 0.001, 0.001, 0.007, 0)$$

根据上述预测, 1992 年 9 月、10 月水位属于 1 级状态概率占绝对优势; 1992 年 11 月至 12 月份水位属 1 级状态的概率与属 2 级状态的概率较接近, 属 1 级状态的概率略大 1%～6%。从地下水监测获得的信息可知, 1992 年 9 月至 12 月份实际月平均水位分别为 25.481m、25.534m、24.964m 和 24.682m, 这与用马尔可夫链模型预测的状态基本吻合。

# 4　应用探讨

马尔可夫链模型, 是以概率论为基础, 对平稳随机现象用自回归过程方法进行定量预测的模型[5]。基于事物未来状况出现的概率不是恒定的, 而是随时间或状态遵循某一规律变化。而后一阶段(一阶)的客观状况的概率, 仅由它相邻前一阶段的概率所决定, 与其它阶段的概率无关为建模基础; 而事物历史监测数据、状态划分和状态转移概率是建模预测的必备条件。

从理论上讲, 历史数据愈长, 状态划分愈多, 预测精度愈高。但目前对状态划分数目和预测步长无统一标准, 状态划分过多, 导致各状态样本点少; 转移概率规律性不强; 预测步长过长, 未把新近状态转移信息加进去, 会使预测精度降低。笔者认为状态划分数目, 应满足各状态平均出现频数大于二分之一状态数目; 预测步长应小于状态数目。

本文应用马尔可夫链模型预测未来地下水水位状态,显示了较高的可信度及可行性。随着监测数据的增多,可把新信息加入,不断修正转移概率矩阵,建立地下水水位预测模型群,通过有效的预测,以期达到管理、监督开发地下水资源和提供早做科学决策依据的目的,对保障和促进经济的发展无疑有着现实意义。

地下水水位是地下水补、径、排条件信息的综合反映,对未来突变事件难以做出真切的预测,这也是所有预测模型共同的弱点。该模型已把历史信息(含突变事件)考虑进去,在一般情况下,能够获得较高精度的预测。随着地质科学研究精度提高和资料的积累,马尔可夫链模型在水工环地质及地层和古生物等方面会有广阔的应用前景。

## 参 考 文 献


[1] 中国科学院地质研究所. 数学地质引论. 北京: 地质出版社, 1977. 257—271

[2] 邓聚龙. 灰色预测与决策. 武汉: 华中理工大学出版社, 1988. 71—79

[3] 徐一飞, 周斯富. 系统工程应用手册. 北京: 煤炭工业出版社, 1991. 193—194

[4] 潘汉军. 马尔可夫过程在基性-超基性含矿性评价中的应用. 见: 地学探索编辑部. 地学探索(2)武汉: 中国地质大学出版社, 1989. 19—26

[5] 何勇, 鲍一丹. 灰色马尔柯夫预测模型及其应用. 系统工程理论与实践, 1992, (4): 59—63


# APPLICATION OF MARKOVIAN CHAIN MODEL TO THE PREDICATION OF UNDERGROUND WATER LEVEL


Song Yinsheng

(*Lunan Geologic-engineering Institute*)



**Abstract**

According to 92 times monitoring datas of underground water levels in Lunan which are monthly averaged, 10 types states were divided. The underground water levels in the future can be predicated by Markovian chain model. On the basis of comparasion with the real monitoring datas, the practicableness and reliability of Markovian chain model are examined, and individual issues related are disscused as well.



**Key words**: Markovian chain, predication of underground water level, state division, state ratio, variation ratio