



# 基于隐马氏性分析的Bayes网络面雨量模型及其应用

王红瑞<sup>①\*</sup>, 叶乐天<sup>②</sup>, 许新宜<sup>①</sup>, 冯启磊<sup>③</sup>, 江燕<sup>①</sup>, 刘琼<sup>①</sup>, 唐奇<sup>①</sup>

① 北京师范大学水科学研究院, 水沙科学教育部重点实验室, 北京 100875;

② 北京大学数学科学学院, 北京 100089;

③ 北京教育学院教师教育数理学院, 北京 100011

\* E-mail: henryzsr@bnu.edu.cn

收稿日期: 2009-03-16; 接受日期: 2009-09-15

国家科技支撑计划(批准号: 2006BAB04A08)资助项目

**摘要** 面雨量估算是水文预报及各类流域水文模型中的基础工作. 基于任意网格上的降水量只与其四周网格的降水量具有不可否定的相关性, 我们提出将 Bayes 网络和 Markov 随机场应用于面雨量估算中. 采用球面坐标系、EM 算法对区域进行插值, 从而可以估算出区域内任一点的降水量. 以青海湖地区的 7 个降水站点为例进行了面雨量估算的实例研究, 通过与泰森多边形法、距离平方倒数法和克里金法等常用的面雨量估算方法对比发现, 本文的方法具备统计学的意义, 并能解决复杂环境下判断不同点降水量之间关系的问题, 模拟结果更为合理、准确.

**关键词**

面雨量  
隐马氏性  
Markov 随机场  
Bayes 网络  
EM 算法  
青海湖

流域面雨量的估算, 是水文预报及各类流域水文模型最基础的一个重要环节, 在分析水情变化时应用非常广泛<sup>[1]</sup>, 其估算的精度直接影响到水文预报等的准确性. 目前, 面雨量常用的方法有距离平方倒数法、等雨量线法、泰森多边形法和克里金插值法等<sup>[2-5]</sup>.

这些算法中距离平方倒数法计算简单, 但仅适用于面积较小, 地形起伏不大, 且站点多而分布又较均匀的区域, 否则, 计算结果粗糙, 误差大. 泰森多边形法的不足之处是将平面分为多个多边形后, 每一个多边形里各个点的降水量值均相同, 这一点不易被人接受, 另外当某测站降水量缺报时将影响面雨量的估算精度, 因而也缺乏弹性. 等雨量线法精度

较高, 但其精度较多地依赖于专业分析技能, 而且操作复杂, 不便于日常业务使用. 而克里金插值法是对距离平方倒数法的改进, 但其假定整个区域的降水量平稳或在每个漂移方向有已知的漂移函数(泛克里金插值法), 但实际中漂移函数比较复杂且不易确定. 文献[6]将离散的雨量值用距离平方倒数法插值到网格点上计算得出流域平均面雨量. 文献[7, 8]在比较了一般的雨量处理方法, 如距离平方倒数法(IDW)、克里金插值法(Kriging)及引入面积权重的泰森多边形雨量处理方法等的基础上, 认为面积权重的泰森多边形雨量处理方法较其它三种方法有更好的精度, 在雨量处理上具有一定的优势. 文献[9]对平均雨量估算方法进行研究, 提出了三种新方法——

引用格式: Wang H R, Ye L T, Xu X Y, et al. Bayesian networks precipitation model based on hidden Markov analysis and its application. Sci China Tech Sci, 2010, 53: 539-547, doi: 10.1007/s11431-010-0034-3

回归算法(regression algorithms)、基于卫星映像的雨量计算方法(satellite image-based precipitation approaches)和双光谱方法的应用(application of the bi-spectral method). 文献[10~12]建立了地区减少因子曲线(areal reduction factor, ARF), 由点雨量的雷达数据来推测面雨量的方法. Daly 等人[13, 14]认为高程是影响降雨空间分布最重要的因素, 以数字高程模型(DEM)为平台, 结合 GIS 的空间插补技术, 提出了针对复杂山地的面雨量计算模型(precipitation-elevation regressions on independent slopes model), 简称为 PRISM.

本文基于任意网格上的降水量只与其四周网格的降水量具有不可否定的相关性这一假设(降水量的隐马氏性), 构建 Bayes 网络, 使整个面上的降水量形成一个 Markov 随机场. 并采用球面坐标系、EM(expectation maximum)算法对区域进行插值, 来估计区域内任一点的降水量, 进而精确估计出面雨量. 以青海湖周边 7 个雨量站为例进行实例分析, 并对模型的精度加以检验.

## 1 Markov随机场(Markov Random Field, 以下简称MRF)

二维图像上的一个函数  $X$  能被看作一个 2 维随机场, 如果这个随机场满足下列条件概率分布:

$$P(X_s = x_s | X_r = x_r, s \neq r) = P(X_s = x_s | X_r = x_r, r \in N_s) > 0,$$

其中  $N_s$  表示像素  $s$  的邻居, 则其就叫做随机场. 即  $s$  的性质完全由它的邻居所决定, 并与除邻居外的其它点条件独立. 在这样的定义下, 一个 MRF 能由条件分布来表示, 这个分布叫做 MRF 的局部性质. MRF 是一个条件独立模型, 能被用来降低极大化的复杂性.

自 1974 年 Besag 首次将 MRF 引入图像处理领域之后, MRF 模型得到了广泛的应用, 在计算机视觉, 图像识别、图像分类、图像存储等方面均有深入的应用[15~17]. 但目前为止其主要应用仍集中在图像处理方面.

## 2 基于隐马氏性的Bayes网络面降水量数学模型的设计

### 2.1 问题的数学描述

已知某流域中  $p$  个观测点的  $q$  个时刻的降水量

$Y = (Y_1, Y_2, \dots, Y_p)$ , 欲估计流域上任意一点  $C$  的任一时刻的降水量  $y$ .

### 2.2 构建Bayes网络

首先对需要估计的区域进行细分, 设置插值点, 其次构建 Bayes 网络, 算法设计如下.

① 区域划分: 考虑到区域的大小, 采用球面坐标系把区域分成了  $m \times n$  个网格, 从左到右, 从上到下依次记为  $D_1, D_2, \dots, D_{m \times n}$ .

② 设置插值点: 将每个网格的 4 个顶点设置为插值点, 从左至右, 从上至下依次记为  $B_1, B_2, \dots, B_{(m+1)(n+1)}$ , 相应的降水量依次记为  $x_1, x_2, \dots, x_{(m+1)(n+1)}$ .

③ 构建 Bayes 网络: 由于每个网格周边有 4 个插值点, 因此每个网格内点的降水量可由本网格的四个插值点估计得到, 即每个网格内点的降水量只与本网格的四个插值点降水量相关. 进一步的, 由于网格较小, 每个网格内的降水量可认为是平稳变化的, 因此在每个网格内部运用距离平方倒数法是合适的, 由此得到假设 1.

**假设 1.** 每个网格内的点降水量可以由本网格的四个插值点的降水量通过距离平方倒数法得到. 即若  $C \in D_i$ ,  $D_i$  的四个顶点分别为  $B_{i1}, B_{i2}, B_{i3}, B_{i4}$ , 相应的降水量分别为  $x_{i1}, x_{i2}, x_{i3}, x_{i4}$ ,  $C$  点到  $B_{i1}, B_{i2}, B_{i3}, B_{i4}$  的距离平方分别为  $d_1, d_2, d_3, d_4$ , 则

$$y_i = \left( \sum_{j=1}^4 \frac{1}{d_j} \right)^{-1} \sum_{j=1}^4 \frac{x_{ij}}{d_j}. \quad (1)$$

由于网格较小, 且每个插值点的四周各有一个相邻插值点, 可以认为每个插值点的降水量在其相邻插值点降水量已知的情况下, 与其他插值点的降水量条件独立, 即如下的条件独立性成立:

$P(X_s = x_s | X_r = x_r, s \neq r) = P(X_s = x_s | X_r = x_r, r \in N_s)$ , 也可以表示为  $X_s = f_s(X_r, r \in N)$ . 这样降水量的分布就是一个 MRF.

进一步, 可以认为  $f_s$  为一个线性函数, 即得到假设 2.

**假设 2.** 每个插值点的降水量可以由其相邻插值点降水量线性表示.

$$\text{即: } X_s = k_r + \sum_{r \in N_s} a_{sr} X_r + \varepsilon_s,$$

其中  $k_r$  和  $a_{sr}$  均为待定系数,  $\varepsilon_s$  为白噪声,  $\varepsilon_s \sim N(0, \sigma^2)$ .

记  $K = (k_1, k_2, \dots, k_{m \times n})$ ,  $X = (X_1, X_2, \dots, X_{m \times n})$ ,  $A = (a_{sr})_{m \times m}$  ( $a_{sr} = 0$  当  $r \notin N_s$  时),  $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{m \times n})$ , 则上式可以表示成

$$X = K + AX + \varepsilon. \quad (2)$$

### 3 模型的计算

#### 3.1 方程组的确立.

由上可知, 每个网格内的点降水量可由本网格的四个插值点的降水量来决定, 所以只需所有插值点的降水量即可. 由构建的 Bayes 网络可得如下方程组:

$$y_i = \left( \sum_{j=1}^4 \frac{1}{d_j} \right)^{-1} \sum_{j=1}^4 \frac{x_{ij}}{d_j}, \quad i=1, 2, \dots, p, \quad (3)$$

其中  $y_i$  为水文站  $Y_i$  的降水量,  $x_{i1}, x_{i2}, x_{i3}, x_{i4}$  为  $Y_i$  所在网格的四个插值点的降水量,  $d_1, d_2, d_3, d_4$  为  $A$  点到这四个插值点的距离.

令  $W = (w_{i,j}), i=1, 2, \dots, p$ , 其中

$$w_{i,j} = \frac{1}{d_{i,j}} \left( \sum_{\substack{x_k \text{ 为 } Y_i \text{ 所在} \\ \text{网格的插值点}}} \frac{1}{d_{i,k}} \right)^{-1},$$

当  $X_j$  为  $Y_i$  所在网格的插值点时, 否则  $w_{i,j} = 0$ .

方程组(3)可以写成

$$Y = WX. \quad (4)$$

方程组(2), (4)联立可得:

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} A \\ W \end{pmatrix} X + \begin{pmatrix} K \\ 0 \end{pmatrix} + \varepsilon. \quad (5)$$

由于此方程是以  $X, A, K$  为变量的非线性方程 ( $Y$  为水文站的降水量数据,  $w_{ij}$  为距离加权系数, 均为已知), 直接求解几乎不可能实现, 因此考虑采用 EM 算法求解.

#### 3.2 EM算法<sup>[18-20]</sup>

极大似然估计是参数估计的一种常用方法. 其计算方法大体可以分为两类: 一类是直接运用于后验分布以得到似然函数极大值的估计, 另一类算法可以总称为数据添加算法, 是近年来发展很快且应用很广的一种算法, 它不是直接对复杂的似然函数进行极大化或模拟, 而是在观察数据的基础上添加一

些“潜在数据”, 从而通过把一个复杂的极大化转化为一列简单的极大化来简化计算, EM 算法正是这样一种通过添加“潜在数据”来间接求解模型参数的极大似然估计的方法.

由于添加的“潜在数据”是未知的, 因此该类添加数据是不完全的. 用 EM 算法处理不完全数据时, 先根据资料对缺失的数据进行估计和推断来补全数据, 然后再根据补全后所得的完全数据集进行迭代计算. 如果  $X$  表示观察到的数据集,  $Y$  表示缺失数据集, 则可假设存在一完全数据集  $Z = (X, Y)$  它的概率密度函数为

$$p(z|\theta) = p(x, y|\theta) = p(x|\theta)p(y|x, \theta),$$

其中  $\theta \in R^p$  为一  $p$  维参数向量. 基于上述概率密度函数, 定义完全数据的似然函数为

$$L(\theta|Z) = L(\theta|X, Y) := p(X, Y|\theta),$$

于是 EM 算法设计如下, 记  $\theta^{(i)}$  为第  $i+1$  次迭代开始时  $\theta$  的估计值, 则  $i+1$  步迭代的两步设计为

① E步: 将  $p(X, Y|\theta)$  或  $\log p(X, Y|\theta)$  关于  $Y$  的条件分布求期望, 即:

$$\begin{aligned} Q(\theta, \theta^{(i)}) &:= E[\log p(X, Y|\theta) | \theta^{(i)}, X] \\ &= \int \log [p(X, Y|\theta)] p(Y|\theta^{(i)}, X) dY. \end{aligned}$$

② M步: 将  $Q(\theta, \theta^{(i)})$  极大化, 即找到一点  $\theta^{(i+1)}$ , 使得

$$Q(\theta^{(i+1)}, \theta^{(i)}) = \max_{\theta} Q(\theta, \theta^{(i)}).$$

如此形成一次迭代  $\theta^{(i)} \rightarrow \theta^{(i+1)}$ , 将上述 E 步和 M 步进行迭代直到  $\|\theta^{(i+1)} - \theta^{(i)}\|$  充分小时停止, 这种迭代是收敛的, 而且必定会收敛到  $\theta$  的极大似然估计.

#### 3.3 方程组的求解

令  $\theta^{(i)} = (A^{(i)}, K^{(i)})$ .

初始值选取:

令  $A^{(0)} = (a_{i,j})$ , 其中  $a_{i,j} = 0$ , 若在 Bayes 网络中  $B_i, B_j$  不相邻, 则取  $a_{i,j} = 1$ , 否则令  $K^{(0)} = (0, 0, \dots, 0)^T$ .

① E步:

$$\begin{aligned} Q(\theta, \theta^{(i)}) &:= E[\log p(X, Y|\theta) | \theta^{(i)}, Y] \\ &= \log p(E(X|\theta^{(i)}, Y), Y|\theta), \end{aligned}$$

而  $E(X | \theta^{(i)}, Y)$  即满足方程(3), (4)的  $X$  的最小二乘估计. 由方程组(3), (4)可得

$$\begin{pmatrix} W \\ I - A \end{pmatrix} X = \begin{pmatrix} Y \\ K \end{pmatrix} + \varepsilon,$$

因此  $X$  的最小二乘估计即

$$X^{(i)} = \left( \begin{pmatrix} W \\ I - A^{(i)} \end{pmatrix}^T \begin{pmatrix} W \\ I - A^{(i)} \end{pmatrix} \right)^{-1} \begin{pmatrix} W \\ I - A^{(i)} \end{pmatrix}^T \begin{pmatrix} Y \\ K^{(i)} \end{pmatrix}.$$

② M 步:

将  $Q(\theta, \theta^{(i)})$  极大化, 即找到一点  $\theta^{(i+1)}$ , 使得  $Q(\theta^{(i+1)}, \theta^{(i)}) = \max_{\theta} Q(\theta, \theta^{(i)})$ , 在这里即求满足方程(3)

的  $\theta^{(i+1)}$  的最小二乘估计. 由方程组(3)可得  $X^{(i)} = (A, K) \begin{pmatrix} X^{(i)} & 0 \\ 0 & I \end{pmatrix} + \varepsilon$ , 所以  $\theta^{(i+1)}$  的最小二乘估计  $\theta^{(i+1)} =$

$$\left( \begin{pmatrix} X^{(i)} & 0 \\ 0 & I \end{pmatrix}^T \begin{pmatrix} X^{(i)} & 0 \\ 0 & I \end{pmatrix} \right)^{-1} \begin{pmatrix} X^{(i)} & 0 \\ 0 & I \end{pmatrix}^T X^{(i)}$$

迭代直至收敛, 即可得到  $X, A, K$  的极大似然估计, 然后根据(1)式即可求出  $y$  的值.

计算程序图如图 1 所示.

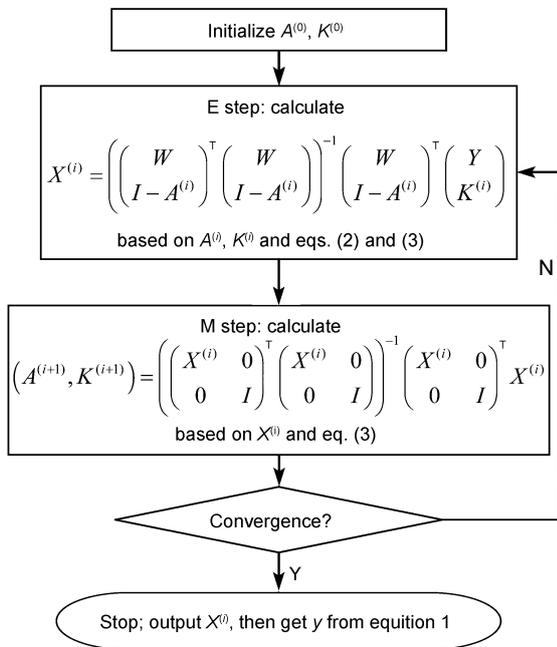


图 1 方程组求解流程图

## 4 应用分析

### 4.1 基本情况

选取青海湖周边的七个水文站的降水数据, 运用本文设计的模型对该地区进行面雨量估算, 7 个水文站的分布见图 2.

青海湖是我国最大的内陆咸水湖, 其流域位于青海省东北部, 地处北纬  $36^{\circ}15' \sim 38^{\circ}20'$ , 东经  $97^{\circ}50' \sim 101^{\circ}20'$ , 海拔范围  $3194 \sim 5174 \text{ m}$ , 总面积  $29623.4 \text{ km}^2$ . 青海湖流域属半干旱地区, 处于我国东部季风区、西北部干旱区和西南部高寒区的交汇地带, 干旱少雨, 蒸发量大, 由于受地形及湖水体影响较大, 降水分布很不均匀. 因此, 青海湖及环湖地区的水文、生态等问题一直是研究热点<sup>[21-24]</sup>.

### 4.2 训练

考虑到区域的大小, 采用球面坐标系把区域分成了  $3 \times 5 = 15$  个网格, 从左到右, 从上到下依次记为  $D_1, D_2, \dots, D_{15}$ , 在每个网格的四个顶点设置为插值点, 从左到右, 从上到下依次记为  $B_1, B_2, \dots, B_{24}$ , 相应的降水量依次记为  $x_1, x_2, \dots, x_{24}$ . 如图 3 所示.

从 1959~2000 年中 7 个水文站的 504 组月降水记录中选取 300 组数据(1959 年 1 月~1983 年 12 月)作为训练样本, 以另外 204 组(1984 年 1 月~2000 年 12 月)为测试样本.

根据上面的分析, 利用训练数据以及方程(1), (2), 采取 EM 算法, 经过 121 次迭代,  $\|\theta^{(i+1)} - \theta^{(i)}\| < 0.001$ , 停止迭代, 训练结束.

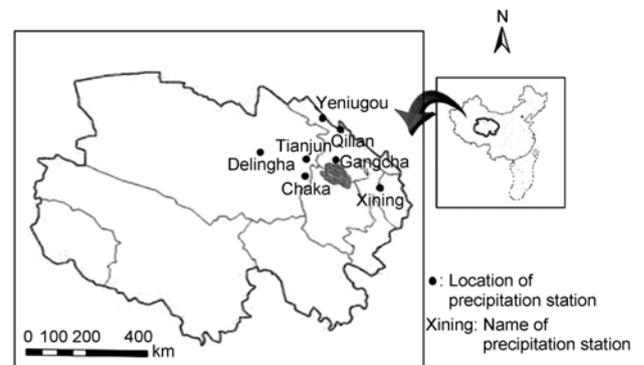


图 2 青海湖地区 7 个降水站的分布图

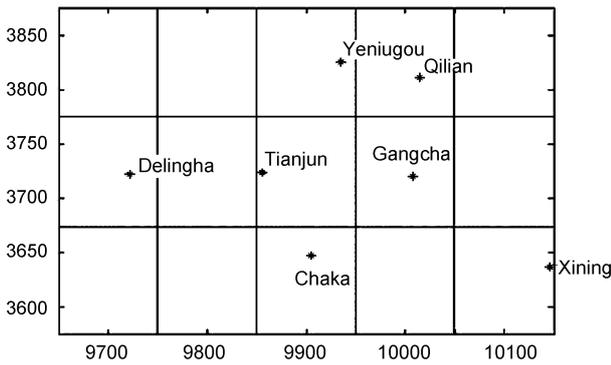


图3 网格分割图

### 4.3 测试

① 方法: 利用测试样本中的6个水文站的数据, 估计出另一个水文站的降水量, 通过估计值和真实值的比较来判断模型是否准确合理.

② 步骤: 依次利用测试样本中的其中6个水文站的数据, 根据训练得到的所有插值点的值  $X$  以及2中的假设1, 用距离平方倒数法得到另一个水文站降水量的估计值.

③ 结果: 依次用该方法对每个降水站的降水量均进行了估算, 同时为了进一步检验此方法, 将基于隐马氏性分析的 Bayes 网络方法与泰森多边形法及距离平方倒数法等进行了比较, 同样每次用其余的6个站点来估计1个站点的降水量, 在用泰森法估算祁连站的降水量时, 考虑到祁连站和野牛沟站非常接近, 在划分块时, 两站在同一块内, 而天峻站被划分

在茶卡站所在块内, 刚察站被划分在祁连站所在块内. 实验结果比较见表1和表2(其中表2中的平均相对误差是指(估计值-真实值)绝对值的均值/真实值的均值).

由表1和表2可知, Bayes 网络法的平均误差一直非常稳定的低于10 mm, 并且其平均相对误差不超过25%. 其显著优于泰森法、距离平方倒数法和克里金法, 可见 Bayes 网络法的计算结果较其它三种常用的算法能够更加接近实际降水量的值.

其中祁连站、刚察站和天峻站的估计结果分别如图4, 5, 6所示, 图7为利用其他方法对天峻站的降水量估计结果.

### 4.4 比较结果讨论

因为泰森法是对整个区域进行分块, 对每个块内的任意一点的降水量均认为与该块内的已知站点相同, 因此泰森法得到的点降水量实际上是很粗糙的. 而本文的方法是在细分网格的基础上, 对局部应用了距离平方倒数法, 所以可认为这也是对距离平方倒数法的一种改进, 可以预料本文的方法会比距离平方倒数法要更好一些. 而简单的算术平均法是对整个平面都采用了一个估计, 效果自然不会很好.

本文的方法最主要的特点在于有效的利用了以前的降水数据, 采用化整体为局部的方法, 更加有效的解决了资料缺失地区上不同点降水量之间关系的问题. 尤其在对天峻站的降水量估计中, 该方法的劣势显得更加明显.

表1 三种方法比较精度对照表(月降水量估计平均误差, 单位: mm)

	野牛沟站	祁连站	德令哈站	天峻站	刚察站	茶卡站	西宁站
Bayes 网络法	7.7284	2.8889	5.0760	4.5513	5.3067	7.2247	9.9662
泰森法	8.0335	5.9643	9.5647	41.8991	8.9759	12.7508	9.4184
距离平方倒数法	12.3611	14.2213	15.5853	32.9878	18.7825	13.2595	10.3305
克里金法	8.0682	6.9643	7.5485	40.4109	7.6038	8.6084	9.0856

表2 三种方法比较精度对照表(月降水量估计平均相对误差, %)

	野牛沟站	祁连站	德令哈站	天峻站	刚察站	茶卡站	西宁站
Bayes 网络法	19.01	3.94	10.56	9.26	11.34	17.79	23.18
泰森法	18.92	16.08	26.25	103.70	22.54	32.76	22.43
距离平方倒数法	33.80	37.93	42.43	83.04	48.01	31.91	26.03
克里金法	17.45	18.85	22.63	98.91	15.23	23.35	20.79

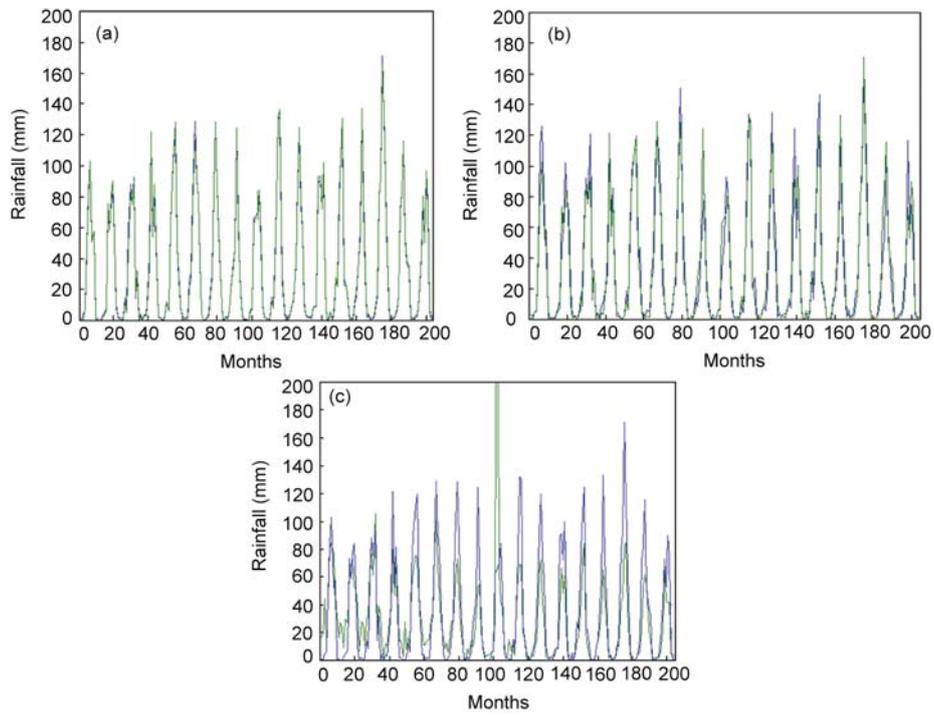


图 4 三种方法对祁连站的降水量估计结果比较(蓝线为实际值, 绿线为估计值)  
(a) Bayes 网络法; (b) 泰森法; (c) 距离平方倒数法

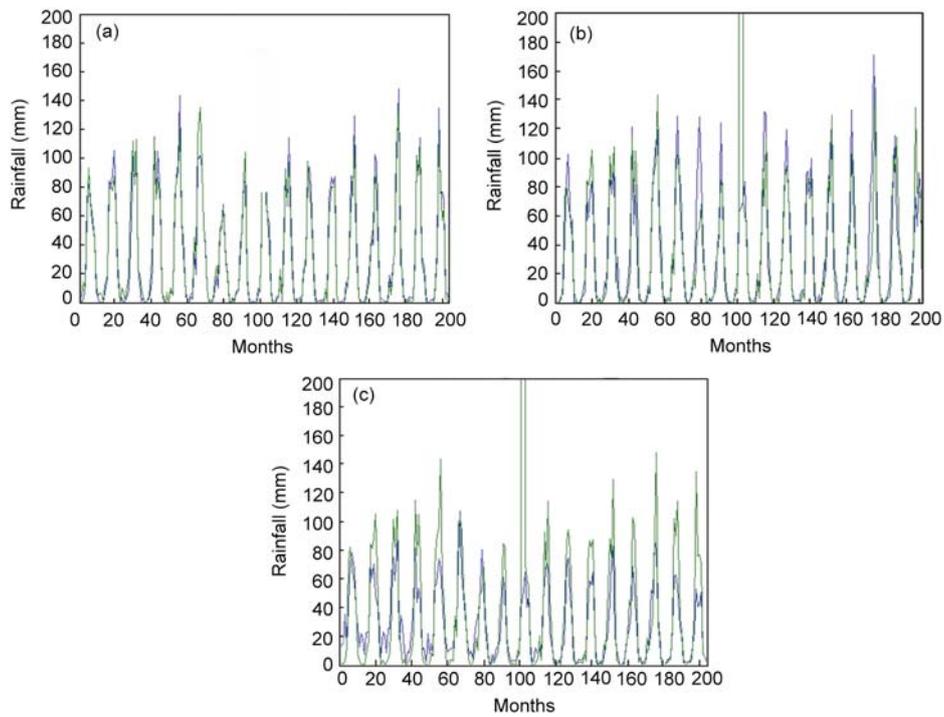


图 5 三种方法对刚察站的降水量估计结果比较(蓝线为实际值, 绿线为估计值)  
(a) Bayes 网络法; (b) 泰森法; (c) 距离平方倒数法

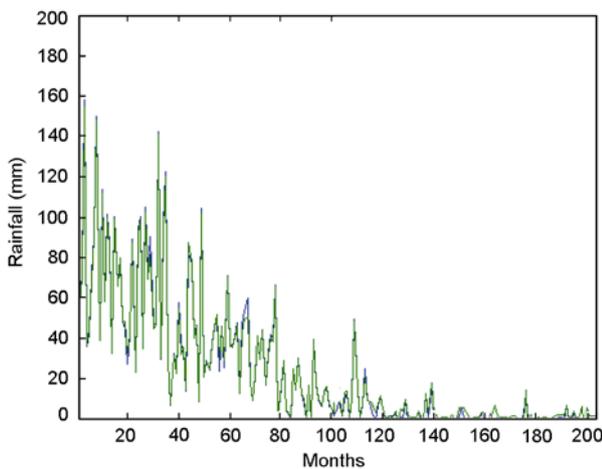


图6 利用 Bayes 网络法对天峻站的降水量估计结果 (蓝线为实际值, 绿线为估计值)

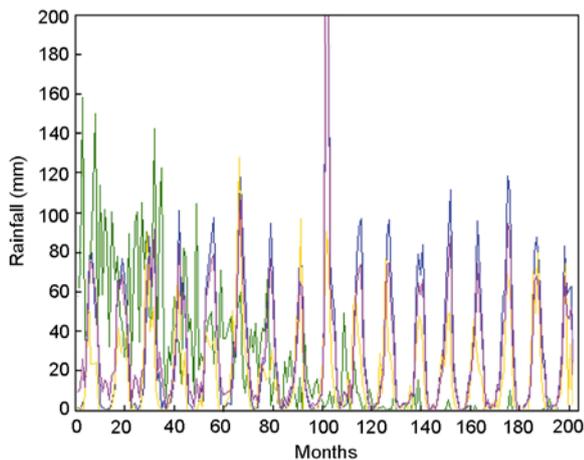


图7 利用其他方法对天峻站的降水量估计结果(绿线为实际值, 蓝线为泰森法估计值, 黄线为克里金法估计值, 红线为距离平方倒数法估计值)

## 5 模型的进一步讨论与改进

从测试的结果可以很明显的看到, 仅仅通过 7 个雨量站的数据, 就很好的估算了整个面上的降水量, 与泰森法等常用的估算方法相比, 这种估算方法要精确和细致的多, 更为的突出是其对突变点捕捉的很好(见图 6 和图 7), 而在水文过程中的突变往往意味着趋势的变化, 所以这是非常重要和有意义的。

关于该模型的复杂度以及模型对数据量的需求作如下说明。

### 5.1 模型计算的复杂度

从方程(1), (2)可以看出模型计算的本质是求最小二乘估计, 这已经有了很成熟的算法, 其复杂度完全由网格的数目  $m \times n$  决定, 复杂度为  $O((m \times n)^{3/2})$ , 而且由于迭代的收敛速度很快, 所以实际计算速度是非常快的。

### 5.2 模型对数据量的需求

方程共有未知量  $X, A, K$ , 因为  $A=(a_{i,j})$  (其中  $a_{i,j}=0$  若在 Bayes 网络中  $i, j$  之间没有边), 因此  $A$  共有  $4 \times 2 + 2(m+n-2) \times 3 + (m-1)(n-1) \times 4 = 4mn + 2m + 2n$  个未知量, 而  $X$  是  $(m+1)(n+1)q$  维未知向量,  $K$  为  $(m+1)(n+1)$  维未知向量, 因此共有  $4mn + 2m + 2n + (m+1)(n+1) + (m+1)(n+1)q = 5mn + 3m + 3n + 1 + (m+1)(n+1)q$  个未知量, 而方程组(2)共有  $p \times q$  组方程, 方程组(3)共有  $(m+1)(n+1)q$  组方程, 当  $p \times q + (m+1)(n+1)q > 5mn + 3m + 3n + 1 + (m+1)(n+1)q$  时, 即  $p \times q > 5mn + 3m + 3n + 1$  时, 参数是可估的, 因此当水文站点或观测次数足够多时, 参数可估. 这表明模型所需要的数据量不大, 且和网格数相关. 在这个实例中, 最少只需 2 个雨量站, 5 组观测数据即可进行估计, 而 5 个雨量站, 10 组观测数据就能达到很好的估计效果. 所以该方法很适合于资料缺失地区的面雨量的估算。

### 5.3 网络分割

为了进一步减少算法的复杂度, 由图 3 可知网格实际上被中间的 3 个水文站分成了左右两边, 这可以认为左边的降水量和右边的降水量在水文站的降水量已知的前提下是条件独立的, 则可以分别考虑左边的网络和右边的网络, 若所要求的点在左边则只需求解左边的参数, 如果是在右边则只需求解右边的参数, 不仅大大降低了算法的复杂度(如果将整个 Bayes 网络等分成  $N$  个小网络, 则计算量将是原来的  $N^{-1/2}$ ), 而且因为网络变小准确度也会提高. 在降水站更多的情况下, 分割所带来的好处将会更加明显。

### 5.4 网格细分

为了更加细致的估计面上的降水量, 可将网格进行加细, 这样估计将会更加准确, 却不需要额外的观测数据, 这是非常实用的, 尤其是在对复杂度不敏感的情况下, 可以得到很精细的估计。

## 6 结论与讨论

从上面的研究和对模型的应用及讨论中可知, 基于隐马氏性分析的 Bayes 网络面雨量模型实际上是在球面网格划分的基础上对泰森多边形法、算术平均法、距离平方倒数法的改进, 后三种方法均对降水量的分布做了很强的假设, 如泰森多边形法和算术平均法均假设各降水站权重为定值, 距离平方倒数法假设权重为距离平方倒数. 克里金插值法则假定整个区域的降水量平稳或在每个漂移方向有已知的漂移函数(泛克里金插值法), 该法需要更多的先验知识, 比如方向漂移函数的确定即纯粹的经验确定, 而在没有任何信息的情况下, 克里金法就是完全由到各个观测点的距离决定. 这些都缺乏统计学上的基础.

此外, 克里金法的计算过程主要是解一个线性方程组, 将降水量表示成距离的线性函数, 不同点的降水量之间的关系仅和它们的距离相关. 由 5 节的讨论可知我们设计的 Bayes 方法计算过程是一个迭代过程, 每一步迭代也是解一个线性方程组. 因此算法复杂度是同一个量级的. 在实际应用中, 由于克里金法一次只能确定一个格点一个时点的降水量, 而我们的方法在训练中确定好参数后, 每次都能一次性确定所有格点的降水量, 在格点较多时我们的方法在估算面雨量上会有更好的效率.

PRISM 是针对复杂山地的回归统计面雨量计算模型, 如果山地降水站点分布少, 其分布状态不足以

表征地形地貌特征时, 此时该方法有着比较显著的优越性. PRISM 是在高程与降水呈线性相关、同一坡面类型的网格具有相同降水分布特征的假设基础上对克里金插值法的进一步改进. 由于降水受很多因素的综合影响, 是一个十分复杂的过程. 近年来, 也有不少学者针对 PRISM 做了不少改进[25~28], 推动了 PRISM 的应用和发展.

基于隐马氏性分析的 Bayes 网络面雨量模型仅假设任意网格上的降水量只与其四周网格的降水量具有不可否定的相关性, 其很好的避免了泰森多边形法、克里金法等其他面雨量法的人为假定. 模型在很大程度上避免了大区域内环境的复杂性对降水量的影响, 化整体为局部, 并充分利用了时序数据, 得到了很好的结果, 将 Markov 随机场模型成功应用到了面雨量的估算中, 并采用 EM 算法给出了数值解法. 尤其适合于对资料缺失地区的面雨量的估算, 能够非常稳定的保持较高的准确性, 并且能够采用网格细分方法自由的达到所需要的精度.

实例分析表明本文设计的基于隐马氏性分析的 BAYES 网络面雨量模型, 其平均相对误差不超过 25%, 显著优于泰森法、距离平方倒数法和克里金法等, 其后的研究和算法复杂度分析进一步表明了该模型的适用性, 这为资料缺失地区或无资料地区提供了一种新的并具有统计学意义的面雨量估算模型, 如将该模型嵌入新安江、SWAT 或 HIMS 等水循环综合模拟系统<sup>[29]</sup>、也将有助于流域水文模型预报精度的提高.

## 参考文献

- 1 芮孝芳. 水文学原理. 中国水利水电出版社, 2004. 76—82
- 2 彭广. 湖北省流域面雨量预报技术研究和应用. 2000 年全国七大江河流域面雨量交流会文集. 北京: 气象出版社, 2000. 13—16
- 3 徐晶, 林建, 姚学祥, 等. 七大江河流域面雨量计算方法及应用. 气象, 2001, 11: 14—17
- 4 潘永地, 方庆文. 有关面雨量研究综述. 贵州气象, 2004, 28: 3—5
- 5 Yang C, Chandler R E, Isham V S, et al. Spatial—temporal rainfall simulation using Generalized Linear Models. Water Resour Res, 2005, 41: 1—13
- 6 梁树献, 刘小虎, 程兴元, 等. 距离平方倒数法计算流域平均面雨量. 新世纪气象科技创新与大气科学发展. 武汉: 测绘出版社, 2004: 315—317
- 7 朱求安, 张万昌. 流域水文模型中面雨量的空间插值. 水土保持研究, 2005, 12: 11—14
- 8 Hanefi B F, Sezer T, Xen Z K. The estimation of average areal rainfall by percentage weighting polygon method in Southeastern Anatolia Region, Turkey. Atmos Res, 2005, 73: 149—160
- 9 Tsintikidisa D, Georgakakosa K P, Artana G A, et al. A feasibility study on mean areal rainfall estimation and hydrologic response in the Blue Nile region using METEOSAT images. J Hydrol, 1999, 221: 97—116
- 10 Robert J A, Arthur T D. Considerations for the use of radar-derived precipitation estimates in determining return intervals for extreme areal

- precipitation amounts. *J Hydrol*, 2005, 315: 203—219
- 11 Grimes D I F, Pardo-Igúzquiza E, Bonifacio R. Optimal areal rainfall estimation using raingauges and satellite data. *J Hydrol*, 1999, 222: 93—108
  - 12 Eulogio P. Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. *J Hydrol*, 1998, 210: 206—220
  - 13 Daly C, Nelson R P, Phillips D L. Statistical-topographic model for mapping climatological Precipitation over mountainous terrain. *J Appl Meteorol*, 1994, 33: 140—158
  - 14 Daly C, Gibson W P, Taylor G H, et al. A knowledge-based approach to the statistical mapping of climate. *Climate Res*, 2002, 22: 99—113
  - 15 Rue H, Held L. *Gaussian Markov Random Fields: Theory and Applications*. London: Chapman, 2004
  - 16 Stan Z L. *Markov Random Field Modeling in Image Analysis*. New York: Springer-Verlag Tokyo, 2001
  - 17 侯一民, 郭雷. 一种基于马尔可夫随机场的 SAR 图像分割新方法. *电子与信息学报*, 2007, 5: 1069—1072
  - 18 赵海兵, 程依明. 无失效数据的 EM 算法. *华东师范大学学报(自然科学版)*, 2006, 5: 94—98
  - 19 房祥忠, 陈家鼎. EM 算法在假设检验中的应用. *中国科学 A 辑: 数学*, 2003, 33: 180—184
  - 20 郑术蓉, 史宁中, 郭建华. 含缺失数据线性模型的线性不等式约束 EM 算法. *中国科学 A 辑: 数学*, 2005, 35: 231—240
  - 21 李凤霞, 伏洋, 杨琼, 等. 环青海湖地区气候变化及其环境效应. *资源科学*, 2008, 30: 348—353
  - 22 俞文政, 常庆瑞, 王锐, 等. 青海湖地区耕地演变过程与影响机制分析. *水土保持通报*, 2006, 26: 19—22
  - 23 孙永亮, 李小雁, 汤佳, 等. 青海湖流域气候变化及其水文效应. *资源科学*, 2008, 30: 354—362
  - 24 时兴合, 李凤霞. 海西东部及环青海湖地区 40 多年的气候变化研究. *干旱地区农业研究*, 2005, 23: 215—222
  - 25 Luzio M D, Johnson G L, Daly C, et al. Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. *J Appl Meteorol Climatol*, 2008, 47: 475—497
  - 26 Daly C, Halbleib M, Smith J I, et al. Physiographically-sensitive mapping of temperature and precipitation across the conterminous United States. *Int J Climatol*, 2008, 28: 2031—2064
  - 27 朱求安, 张万昌, 赵登忠. 基于 PRISM 和泰森多边形的地形要素日降水量空间插值研究. *地理科学*, 2005, 25: 233—238
  - 28 赵登忠, 张万昌, 刘三超. 基于 DEM 的地理要素 PRISM 空间内插研究. *地理科学*, 2004, 24: 205—211
  - 29 刘昌明, 王中根, 郑红星. HIMS 系统及其定制模型的开发与应用. *中国科学 E 辑: 技术科学*, 2008, 38: 350—360