

Analysis on Saline Water Intrusion in Qiantang River Based on Bayesian Network*

Xichao Gao, Chong Ma, Xujie Zhang, Yueping Xu, Yongqiang Zhang

Institute of Hydrology and Water Resources, Civil Engineering, Zhejiang University, Hangzhou
Email: pandagxc@gmail.com

Received: Jul. 12th, 2012; revised: Aug. 1st, 2012; accepted: Aug. 14th, 2012

Abstract: Saline water intrusion in Qiantang River exerts a severe impact on residential life, environment and economic development. In this article, a simple Bayesian Network model based on water levels of Tonglu, Zhakou and Kanpu stations is established and the parameters are analyzed through the observation data. Whether the saline water intrusion will occur or not is determined by the relative water level of Tonglu, Zhakou and Kanpu stations. Based on the model, the probability of saline water intrusion at Tonglu and Zhakou stations is determined. The reasoning ability of Bayesian Network is utilized to simulate the potential saline water intrusion in some certain situations.

Keywords: Bayesian Network; Saline Water Intrusion; Estuary of Qiantang River; Joint Risk

基于贝叶斯网络的钱塘江河口地区咸潮入侵研究*

高希超, 马 冲, 张徐杰, 许月萍, 张永强

浙江大学建筑工程学院水利工程学系, 杭州
Email: pandagxc@gmail.com

收稿日期: 2012年7月12日; 修回日期: 2012年8月1日; 录用日期: 2012年8月14日

摘 要: 本文建立了钱塘江下游桐庐、闸口及瞰浦三站水(潮)位之间的贝叶斯网络模型, 并对网络模型的参数进行了估计和分析。通过三站高、低水位的不同遭遇情况来确定是否有咸潮入侵的情况发生, 给出了桐庐、闸口两站咸潮入侵的概率, 且通过贝叶斯网络的推理功能对所研究地区的咸潮入侵情况进行仿真分析, 给出在某些特定情况下各站发生咸潮入侵的概率。

关键词: 贝叶斯网络; 咸潮入侵; 钱塘江河口; 联合风险

1. 研究背景

钱塘江是浙江省的第一大河, 由西往东穿越浙江省全境, 在杭州湾向东注入东海。钱塘江主流有两条, 兰江为其南源, 北源为新安江, 南北两源在新安江交汇是为富春江, 东江嘴以下称为钱塘江。钱塘江河口

地形条件特殊浪潮十分汹涌, 自桐庐以下皆为潮水影响区间。然据资料所述^[1], 钱塘江下游杭州地区 80% 的工业及人民生活用水, 绍兴、萧山、上塘河平原及南沙近 20 余万 hm^2 的农田灌溉用水皆取自钱塘江。因此, 钱塘江的咸潮入侵现象将会给下游工农业生活带来严重的影响。据调查, 1979 年夏, 因上游供水匮乏, 咸潮大量入侵, 影响杭州供水 40 余天, 1994 年夏季的咸潮造成杭州市区大面积停水。鉴于以上所述情况, 对咸潮入侵进行进一步的研究便显得十分必要。

*基金项目: 科技部国际科技合作计划(2010DFA24320); 国家自然科学基金(50809058)。

作者简介: 高希超(1987-), 男, 山东滨州, 硕士研究生, 主要从事贝叶斯网络在水文学中的应用研究。

贝叶斯网络起源于20世纪80年代中期对人工智能中的不确定性问题研究, 近来其影响力不断扩大, 已成为将概率、统计应用于复杂系统的不确定性推理和数据分析的一种有效工具。随着贝叶斯网络理论的越来越成熟, 其在水文学中的应用也逐渐广泛起来。如袁江南等^[2]利用贝叶斯网络预测台风等突发事件, 舍敦先等^[3]利用贝叶斯网络对南水北调工程运行风险进行了综合分析, 罗建男^[4]等利用贝叶斯网络进行水环境管理的研究, 卢文喜^[5]等将贝叶斯网络应用于水资源管理中。但是, 利用贝叶斯网络进行咸潮入侵风险的研究尚还少见。

本文对桐庐、闸口以及瞰浦三站的水(潮)位数据建立了贝叶斯网络模型, 并对模型进行了参数学习, 获得了网络模型中的各个参数。然后, 对该网络模型进行进一步分析, 得出各站发生咸潮入侵的概率, 为采取措施避免咸潮入侵提供一定的参考依据。

2. 研究方法

2.1. 贝叶斯网络理论简介

贝叶斯网络是一种将概率统计应用于复杂领域、进行不确定性推理和数据分析的工具, 是一种基于网络结构的有向无环图。贝叶斯网络是概率论与图论相结合的产物, 它运用图论的语言来揭示问题的结构, 然后按照概率论的原则对问题的结构加以利用, 降低推理计算的复杂度。

贝叶斯网络是由节点及连接节点的线段构成的有向无环图。其基本原理是^[6]: 利用变量之间的相互关系及链规则来减少求联合分布时的复杂度。一般的, 考虑一个包含 n 个变量的联合分布 $P(X_1, \dots, X_n)$ 。利用链规则, 可以把它写为

$$P(X_1, \dots, X_n) = P(X_1)P(X_2|X_1)\dots P(X_n|X_1, X_2, \dots, X_{i-1}) \quad (1)$$

对于任意 X_i , 如果存在 $\pi(X_i) \subseteq \{X_1, \dots, X_{i-1}\}$, 使得给定 $\pi(X_i)$, X_i 与 $\{X_1, \dots, X_{i-1}\}$ 中的其他变量条件独立, 即 $P(X_i|X_1, \dots, X_{i-1}) = P(X_i|\pi(X_i))$, 则有

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i|\pi(X_i)) \quad (2)$$

这样, 得到了联合分布的一个分解, 其中当 $\pi(X_i) = \emptyset$ 时, $P(X_i|\pi(X_i))$ 为边缘分布律 $P(X_i)$ 。

由贝叶斯公式可知, 条件概率分布的求解公式为

$$P(X_1|X_2) = \frac{P(X_1X_2)}{P(X_2)} \quad (3)$$

2.2. 贝叶斯网络的参数学习

贝叶斯网络参数学习^[7]是指已知贝叶斯网络结构通过分析数据而获得贝叶斯网络参数的过程。参数学习的方法主要有两种: 最大似然估计与贝叶斯估计。这里仅介绍贝叶斯估计。

在贝叶斯估计中, 视参数 θ 为随机变量, 对它进行估计就是计算其后验概率分布。为此, 首先选用一个概率分布 $P(\theta)$ 来总结关于 θ 的先验知识, 然后把数据 D 的影响用似然函数 $L(\theta|D) = P(D|\theta)$ 来归纳, 最后使用贝叶斯公式将先验分布和似然函数结合, 得到 θ 的后验分布, 即为

$$P(\theta|D) \propto P(\theta)L(\theta|D) \quad (4)$$

3. 模型建立

3.1. 数据资料处理

本文研究的桐庐、闸口及瞰浦三站的水(潮)位数据资料为1959~2008年的实测水位数据。根据建模需要, 本文将水(潮)位数据离散为三个阶段即高水(潮)位、平水(潮)位、低水(潮)位。确定以某一站的水位数据作为标准, 采用PIII型频率曲线^[8]通过线性矩方法^[9]来确定计算统计参数以及各个频率的设计值。然后根据得出的频率曲线求得频率为0.375、0.625处的值, 来划分水位的等级标准。本文采用两组数据分别训练模型, 并根据所训练模型得出的结果来确定判别咸潮入侵的控制因素。本文采用的数据组合为: 1) 上游桐庐站的年最低水(潮)位及其对应的下游闸口、瞰浦两站的水(潮)位, 2) 下游瞰浦站的年最高水(潮)位及其对应的上游桐庐、闸口两站的水(潮)位。由于缺少钱塘江各站点的具体盐度数据, 因此对于咸潮入侵的判定只能依靠水位的相对高低来决定, 可能出现咸潮入侵的不利组合情况如下表所示(仅列桐庐站)(见表1)。

3.2. 建立贝叶斯网络咸潮入侵风险模型

利用贝叶斯网络建立模型通常有以下几个步骤:

Table 1. Combination of salt water invasion at Tonglu station
表 1. 桐庐站咸潮入侵的组合情况

序号	桐庐	闸口	瞰浦
1	低	低	平
2	低	低	高
3	低	平	平
4	低	平	高
5	低	高	高
6	平	平	高
7	平	高	高

1) 选取模型变量; 2) 确定网络结构; 3) 计算网络参数; 4) 利用网络模型进行推理。根据研究目的, 本文确定桐庐、闸口及瞰浦三站的水位数据分别作为模型的变量。模型结构通常有两种方法确定^[10]: 1) 通过大量的实测数据由贝叶斯网络的结构学习的功能学习获得。2) 经由专家知识获得。即由行业内专业人士通过对变量间的因果依存关系的认知, 直接得出变量与变量之间的关系, 从而得出网络模型的结构。本文采用第二种方法来确定钱塘江河口地区咸潮入侵风险贝叶斯网络模型的结构。模型结构如下图 1 所示。

网络建立之后, 计算网络参数。即计算网络中根节点的先验概率分布和非根节点的条件概率分布表(CPT)。由于本文是用两类数据组分别训练的模型, 因而会得到两组网络参数, 其结果如表 2~4 所示, 由数据(1)所得网络参数为:

网络各节点组合的联合概率分布(以数据(1)为例)如表 5 所示, 表里仅列举了最不利的几个组合。

利用网络模型进行推理。贝叶斯网络可以利用变量间的条件独立关系对联合分布进行分解, 减少模型参数的个数, 从而简化了知识的获取。对由两组不同的数据组分别训练来的网络参数进行咸潮入侵组合的联合概率分布求解可得: 对于表 4 所示网络参数,

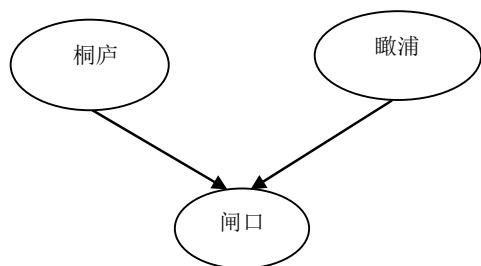


Figure 1. Structure of Bayesian Network
图 1. 贝叶斯网络结构

Table 2. Marginal probability of water level at Tonglu station
表 2. 桐庐站水位边缘概率分布

桐庐	低	平	高
概率	0.2667(0.0667)	0.4303(0.1030)	0.3030(0.8303)

Table 3. Marginal probability of water level at Kanpu station
表 3. 瞰浦站水位边缘概率分布

瞰浦	低	平	高
概率	0.5758(0.3939)	0.2121(0.2667)	0.2121(0.3394)

Table 4. Conditional probability of water level at Zhakou station
表 4. 闸口站水位条件概率分布

桐庐	瞰浦	闸口站水位概率		
		低	平	高
低	低	0.4859(0.3448)	0.0282(0.3050)	0.4859(0.3502)
平	低	0.4563(0.3333)	0.4563(0.0521)	0.0874(0.6146)
高	低	0.1891(0.1245)	0.4487(0.1245)	0.3622(0.7511)
低	平	0.6992(0.5339)	0.0407(0.1327)	0.2602(0.3335)
平	平	0.8958(0.7619)	0.0521(0.1190)	0.0521(0.1190)
高	平	0.0521(0.0147)	0.3333(0.0944)	0.6146(0.8909)
低	高	0.6145(0.8551)	0.3333(0.0725)	0.0521(0.0725)
平	高	0.4859(0.1255)	0.3333(0.3897)	0.1808(0.4848)
高	高	0.1190(0.0119)	0.1190(0.0762)	0.7619(0.9119)

注: 括号内为由数据(2)计算得来的网络参数。

Table 5. Joint probabilities
表 5. 联合概率分布表

桐庐	瞰浦	闸口	概率
低	高	低	0.0348
平	高	低	0.0443
高	高	低	0.0077
低	高	平	0.0189
平	高	平	0.0304
低	平	高	0.0147
低	高	高	0.0029

考虑所有可能遭遇咸潮入侵的不利组合: 桐庐站的咸潮入侵的概率为 14.53%, 闸口站咸潮入侵的概率为 29.01%; 对于表 5 所示网络参数: 桐庐站的咸潮入侵的概率 6.51%, 闸口站的咸潮入侵的概率为 11.85%。由以上结果可知对于判定是否有咸潮入侵现象发生, 数据组(1)起着控制作用, 即决定是否有咸潮入侵现象发生的主要因素是上游河道的水位, 下游海面的潮水位起次要作用。对于桐庐站, 如果仅仅考虑最不利组合, 即桐庐与闸口站水位为低, 瞰浦站潮位为高时,

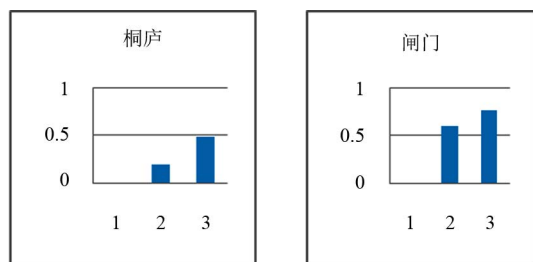
则桐庐发生咸潮入侵的概率为 3.48%, 远低于闸口站遭遇咸潮入侵的概率。我们对由第一种数据组合得来的网络模型进行更进一步的推理分析, 以下分析采用所有可能的不利组合。

经贝叶斯网络模型推理可知, 当桐庐站的水位测得为低水位时, 桐庐站遭遇咸潮入侵的概率为 36.91%, 闸口站咸潮入侵的概率为 36.91%; 当桐庐站的水位为平水位时, 桐庐站遭遇咸潮入侵的概率为 10.91%, 闸口站遭遇咸潮入侵的概率为 40.21%; 当桐庐站的水位测得为高水位时, 桐庐站遭遇咸潮入侵的概率为 0%, 闸口站遭遇咸潮入侵的概率为 6.16%。当闸口站的水位已知为低水位时, 桐庐站遭遇咸潮入侵的概率为 11.31%, 闸口站遭遇咸潮入侵的概率为 42.42%; 当闸口站的水位数据已知为平水位时, 桐庐站遭遇咸潮入侵的概率为 20.44%, 闸口站遭遇咸潮入侵的概率为 26.87%; 当闸口站的水位已知为低水位时, 桐庐站遭遇咸潮入侵的概率为 14.78%, 闸口站遭遇咸潮入侵的概率为 14.78%。当瞰浦站的潮位已知为低水位时, 桐庐站和闸口站遭遇咸潮入侵的概率都为 0; 当瞰浦站的潮位已知为平水位时, 桐庐站遭遇咸潮入侵的概率为 19.73%, 闸口站遭遇咸潮入侵的概率为 59.86%; 当已知瞰浦站的潮位为高水位时, 桐庐站和闸口站遭遇咸潮入侵的概率分别为 48.79%、76.91%。

图 2 所示为已知瞰浦水位时桐庐、闸口两站遭遇咸潮入侵的边缘概率。

4. 结论

本文针对钱塘江河口地区的咸潮入侵情况, 利用贝叶斯网络理论建立了钱塘江河口地区的咸潮



(图中 1、2、3 分别代表瞰浦站低、平、高水位, 纵坐标为发生咸水入侵的概率)。

Figure 2. Marginal probability of salt water intrusion at Tonglu and Zhakou stations with known water level at Kanpu station
图 2. 已知瞰浦水位时桐庐、闸口站咸潮入侵的概率

入侵风险计算模型。该模型简单、直观, 并定量表达了桐庐、闸口、瞰浦三站水位之间的内在关系, 并藉此得出各站遭遇咸潮入侵风险的概率。通过后验推理, 给出了在各种情况下各站遭遇咸潮入侵的概率。该模型可以根据随后获得的数据实时更新、修正, 使得模型更加完善, 能够正确的反应真实情况, 为钱塘江河口地区咸潮入侵现象的治理提供科学依据。但是, 该模型仅仅依据水位数据来判定咸潮入侵是否发生, 模型过于简单, 存在一定的局限性。尤其是对于距离河口较远的桐庐站所获得的结果误差可能较大, 仍需进一步改进。

参考文献 (References)

- [1] 徐小燕. 新安江水库的压咸作用初探[J]. 浙江水利水电专科学校学报, 2001, 13(1): 42-43.
XU Xiaoyan. Discussion on desalinity function in Xi'anjiang reservoir. Journal of Zhejiang Water Conservancy and Hydropower College, 2001, 13(1): 42-43. (in Chinese)
- [2] 裘江南, 王延章, 董磊磊, 等. 基于贝叶斯网络的突发事件预测模型[J]. 系统管理学报, 2011, 20(1): 98-102.
QIU Jiangnan, WANG Yanzhang, DONG Leilei, et al. A model for predicting emergency event based on Bayesian Networks. Journal of Systems & Management, 2011, 20(1): 98-102. (in Chinese)
- [3] SHE Dunxian, YANG Xiaohua and XIA Jun. The evaluation of the integrated risk for the South-to-North Water Transfer Project using the Bayesian Network theory. Journal of Resources and Ecology, 2010, 1(3): 259-267.
- [4] 罗建男, 卢文喜, 陈社明, 等. 基于贝叶斯网络的水环境管理研究[J]. 节水灌溉, 2011, 3: 68-72.
LUO Jiannan, LU Wenxi, CHEN Sheming, et al. Study on water environment management based on Bayesian Networks. Water Saving Irrigation, 2011, 3: 68-72. (in Chinese)
- [5] 卢文喜, 罗建南, 鲍新华. 贝叶斯网络在水资源管理中的应用[J]. 吉林大学学报, 2011, 41(1): 153-158.
LU Wenxi, LUO Jiannan and BAO Xinhua. Application of Bayesian Networks in water resource management. Journal of Jilin University, 2011, 41(1): 153-158. (in Chinese)
- [6] 张连文, 郭海鹏. 贝叶斯网引论[M]. 北京: 科学出版社, 2006: 33-34.
ZHAN Lianwen, GUO Haipeng. Introduction of Bayesian Networks. Beijing: Science Press, 2006: 33-34. (in Chinese)
- [7] 张连文, 郭海鹏. 贝叶斯网引论[M]. 北京: 科学出版社, 2006: 143-171.
ZHAN Lianwen, GUO Haipeng. Introduction of Bayesian Networks. Beijing: Science Press, 2006: 143-171. (in Chinese)
- [8] 戴昌军. 沿海地区年最高潮位频率分析研究[J]. 中国农村水利水电, 2005, 11: 37-39.
DAI Changjun. Analysis of annual design highest tide level frequency in littoral area. China Rural Water and Hydropower, 2005, 11: 37-39. (in Chinese)
- [9] 许月萍, 童杨斌, 楼章华. 无资料小流域设计洪水不确定性研究[J]. 中山大学学报(自然科学版), 2008, 47(6): 61-69.
XU Yueping, TONG Yangbin and LOU Zhanghua. Estimating uncertainty in design flood for ungauged catchments. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2008, 47(6): 61-69. (in Chinese)

- [10] 康玲, 何小聪, 熊其玲. 基于贝叶斯网络理论的南水北调中线工程水源区与受水区降雨丰枯遭遇风险分析[J]. 水力学报, 2010, 41(8): 908-913.
KANG Lin, HE Xiaocong and XIONG Qilin. Risk analysis for precipitation rich-poor encounter between source area and receiving area of the Middle Route of South-to-North Water Transfer Project based on Bayesian theory. Journal of Hydraulic Engineering, 2010, 41(8): 908-913. (in Chinese)