

台风形成库区降雨的影响因子挖掘

李子阳^{①*}, 张湛^②, 马福恒^①, 李成才^③, 潘树军^④^① 南京水利科学研究院水文水资源与水利工程科学国家重点实验室, 南京 210029;^② 河海大学水利水电学院, 南京 210098;^③ 河海大学土木与交通学院, 南京 210098;^④ 河海大学设计研究院有限公司, 南京 210098

* E-mail: zyli@nhri.cn

收稿日期: 2014-05-22; 接受日期: 2014-06-29

国家自然科学基金(批准号: 51209144, 51209145, 51209143, 51139001, 51109143)、水利部公益性行业科研专项(编号: 201401022)和中央级公益性科研院所基本科研业务费专项资金(批准号: Y713008, Y714014)资助项目

摘要 考虑影响台风降雨因素的复杂性以及现有水库台风遥测资料的不完备性, 采用基于信息熵与条件信息熵建立起来的不完备信息系统的约简方法, 对台风形成库区降雨的影响因子进行挖掘. 以典型水库中的台风特性参数及库区降雨实时监测信息为依据, 忽略环境因素影响, 重点研究包括台风风速、风力、7 级/10 级风圈半径、中心气压、台风中心与库区距离等典型台风特性因子对库区降雨量的影响, 经信息约简认为台风风速、中心气压、与库区距离等台风特性因子是库区降雨的主要影响因子. 分析结果对水库台风监测及库区降雨强度分析具有一定指导作用.

关键词台风
库区降雨
影响因子
信息约简

1 引言

台风往往带来特大暴雨和爆发性洪水, 对水库大坝安全有着重要影响. 由于台风带来的强降雨主要是短时程的连续性集中降水, 常形成突发性洪水^[1], 造成库水位的急剧上涨, 甚至超过校核洪水位, 引起大坝漫顶破坏. 而水位的急剧上升对于大坝而言, 相当于受到突然加荷的过程, 巨大浪涌冲击力往往也比设计荷载大很多且多为突变荷载, 大坝在这种瞬态超强作用下极易造成结构性态疲劳损伤, 引发裂缝、渗漏、滑坡等问题^[2]. 除上述直接作用外, 强降雨还可能诱发水库地质环境脆弱区的崩塌、滑坡和泥石流等, 也可能携带库区植被堵塞输泄水建筑物等, 从而威胁大坝安全. 因此, 台风形成的库区降雨

对水工结构性态的影响和致灾效应不容忽视.

但台风形成的强降雨不仅与台风本身的强度、范围及结构有关, 还与环境因素密切相关. 研究表明, 台风本身的强度对台风暴雨有重要影响, 强度越强的台风, 其产生的暴雨往往也越强, 但也并不尽然. 因此, 台风暴雨的准确预报仍然比较困难, 尤其是台风与环境的相互作用导致预报难度更大^[3-5]. 如何根据台风对水库致灾的特征和规律, 从诸如台风强度、风速、中心气压等众多的台风特性因子中寻找影响库区降雨的主要因素, 即台风形成库区降雨的影响因子挖掘, 是台风对水工结构致灾影响分析的前提, 也是水库大坝面向极端气候条件下安全监控和管理研究的热点问题. 本文基于特定水库工程, 考虑水库现有台风遥感监测技术及监测资料的有限性, 忽略环

引用格式: 李子阳, 张湛, 马福恒, 等. 台风形成库区降雨的影响因子挖掘. 中国科学: 技术科学, 2014, 44: 1020–1026
Li Z Y, Zhang Z, Ma F H, et al. Impact factors mining of reservoir rainfall caused by typhoon (in Chinese). Sci Sin Tech, 2014, 44: 1020–1026, doi: 10.1360/N092014-00179

境因素影响, 尝试采用基于条件信息熵建立起来的不完备信息系统的约简方法, 进行台风形成库区降雨的影响因子挖掘, 重点对包括台风风速、风力、7 级/10 级风圈半径、中心气压、台风中心与影响库区距离等典型台风特性因子与库区降雨的关系进行信息约简和分析.

2 信息熵和条件信息熵

在信息论中通常把 Shannon 定义的熵称为信息熵, 其定义为

$$H(x) = -\sum_{i=1}^n p_i \ln p_i, \quad (1)$$

其中, p_i 为信息源中信号 x_i 出现的概率, 有 $\sum p_i = 1$; $H(x)$ 表征了信息量的大小, 是一个系统状态不确定性的量度. 信息熵度量了信源提供的平均信息量的大小, 关系的分辨能力越强, 信息越丰富, 信息熵就越大, 反之亦然.

基于等价关系的知识信息熵利用等价关系对论域的划分进行定义: U 为论域, 等价关系 P 对 U 的划分为 $P = \{X_1, X_2, \dots, X_k\}$, 则 P 的信息熵可定义为^[6]

$$H(P) = -\sum_{i=1}^k \frac{|X_i|}{|U|} \ln \frac{|X_i|}{|U|}, \quad (2)$$

其中, $|X_i|, |U|$ 分别为 X_i, U 的基数.

在现实中, 等价关系要求过于苛刻, 信息的不完备现象是广泛存在的. 于是将等价关系放宽为相似关系(自反性、传递性)、相容关系(自反性、对称性)、甚至为一般二元关系(一般要求满足自反性). 此时, 人们利用论域中任一对象在一般二元关系构成的所有对象的邻域中出现的次数来定义知识的信息熵^[7].

对信息系统 $S = (U, A, V, f)$ (其中 U 为论域对象空间, $A = C \cup D$ 是属性(条件属性 C 和决策属性 D)的有限非空集合, V 是属性值的集合; $f: U \times (C \cup D) \rightarrow V$ 为信息函数), $P \subseteq A$ 是论域 $U = \{x_1, x_2, \dots, x_u\}$ 上的某二元关系, n_p 表示在 P 下的邻域算子; $n_p(x_i)$ 表示在知识 P 下 x_i 的一般二元关系邻域, 且

$$x_i \in n_p(x_j), i = 1, 2, \dots, u;$$

$$|x_i|_p = \left| \left\{ n_p(x_j) \mid x_i \in n_p(x_j), 1 \leq j \leq u \right\} \right|,$$

即 x_i 在所有 $x_j (1 \leq j \leq u)$ 的邻域中出现的次数, 则知识 P 在论域 U 上的信息熵定义为

$$H(n_p) = -\frac{1}{|U|} \sum_{i=1}^u \ln \frac{|x_i|_p}{|U|}. \quad (3)$$

同样的, 在一般二元关系下定义的条件信息熵如下: $P, Q \subseteq A$ 是论域 U 上的 2 个二元关系, $P \cup Q$ (表示同时满足 P 与 Q) 仍为 U 上的二元关系, 对 U 中任一元 $x_i (1 \leq i \leq u)$, 在 $P \cup Q$ 下的邻域为 $n_{P \cup Q}(x_i) = n_p(x_i) \cap n_q(x_i)$. 相应可定义 $P \cup Q$ 的信息熵为

$$H(P \cup Q) = -\frac{1}{|U|} \sum_{i=1}^u \ln \frac{|x_i|_{P \cup Q}}{|U|}. \quad (4)$$

关系 Q 相对于关系 P 的条件信息熵定义为

$$H(Q|P) = H(P \cup Q) - H(P). \quad (5)$$

3 不完备信息系统的约简方法

台风降雨的影响因子挖掘, 实际上就是从台风环境量与影响水库降雨量等的监测信息中, 获取典型台风特性因子与水库降雨效应量之间的逻辑关系的过程. 台风形成水库降雨作用机理和台风遥测的复杂性, 决定了监测信息的不完备性, 等价关系不再成立, 因此, 基于一般二元关系的信息系统知识约简就具有重要的理论和实际意义. 这里通过引入信息熵和条件信息熵, 对信息系统中属性的必要性进行定义; 继而引入一种基于条件信息熵的知识约简启发式算法^[6], 该算法可得到信息表的约简和决策表的最小相对约简.

3.1 信息约简及约简的核

设信息系统 $S = (U, A, V, f)$, 对 $c \in B \subseteq C$, 若 $H(\{c\} | (B - \{c\})) = 0$, 则称 c 在 B 中是不必要的, 即如果从属性集 B 其余属性能完全确定属 c , 那么该属性是多余的(不必要的), 否则称 c 在 B 中是必要的, 所有 B 中必要的元素称为 B 的核.

设 P 为定义在信息系统 $S = (U, A, V, f)$ 上的一个关系族(包含决策属性集). 若 $H(D|B) = H(D|(B - \{c\}))$, 则称属性 $c \in B \subseteq C$ 是 B 相对于 D 不必要的, 否则称属性 c 是 B 相对于 D 必要的. 即如

果将属性 c 从属性子集 B 中去掉前后对决策属性集的不确定性度量不变, 则可将该属性约去.

根据上述定义, 如果决策信息表是一致的, 则表约简计算的代数观点和信息论观点是等价的. 如果一个属性不能为另一个属性的集合分类增加任何信息, 就可以将它进行约简. 基于这种考虑, 可以用条件信息熵作为启发知识来寻找决策信息表的启发式知识约简算法.

3.2 基于条件信息熵的属性约简

对于无决策信息表, 论域上的二元关系只要满足自反性, 即可定义此二元关系的信息熵, 通过保持信息熵不变的标准, 逐次检查各个属性是否必要, 从而确定核, 再根据各非必要属性对核的条件熵从小到大排序, 逐个加入, 直到信息熵等于约简前的信息熵, 即得到信息系统的属性约简, 算法如下.

输入: 1 个信息系统 $S = (U, C, V, f)$, $U = \{x_1, x_1, \dots, x_n\}$ 为论域, C 为属性集.

输出: 信息系统的一个属性约简 B .

Step 1, 计算 S 的信息熵 $H(C)$;

Step 2, 对每个 $c \in C$, 计算 $H(C|(C - \{c\}))$, 令 $C_0 = \{c | H(C|(C - \{c\})) > 0\}$, 则 C_0 为核;

Step 3, 令 $B = C_0$, 对属性集 $C - B$ 重复如下运算:

Step 3.1, 若 $H(B) = H(C)$ 则终止运算, 转到 Step 4;

Step 3.2, 对每个 $p \in C - B$, 计算 $H(\{p\}|B)$;

Step 3.3, 选择使 $H(\{p\}|B)$ 最大的属性 p (若有多个属性同时达到最大, 则任选 1 个), 作 $B = B \cup \{p\}$, 转到 Step 3.

Step 4, 输出 B , 即为信息系统的 1 个约简.

3.3 基于条件信息熵的知识约简

对于决策表 $S = (U, A, V, f)$ (包含决策属性), 可以利用每个条件属性对决策属性的条件信息熵大小来判断该条件属性与决策属性的影响程度: 条件属性对决策属性的条件信息熵越大, 它对决策属性的影响也就越小, 该属性就越可能被约简. 基于此, 给出决策表的知识约简算法, 算法如下.

输入: 1 个决策表 $S = (U, C \cup D, V, f)$, 其中 C 是

条件属性集, D 是决策属性集.

输出: 该决策表的一个相对约简 B .

Step 1, 计算决策属性 D 相对于条件属性 C 的条件信息熵 $H(D|C)$;

Step 2, 计算决策属性相对于每个条件属性的条件信息熵 $H(D|\{a_i\}) (a_i \in C)$;

Step 3, 令 $B = C$ 按 $H(D|\{a_i\}) (a_i \in C)$ 递减的顺序对每个 a_i 重复作如下运算:

Step 3.1, 计算决策属性集相对条件属性集 B 在删除 a_i 后的条件信息熵 $H(D|B - a_i)$;

Step 3.2, 判断 $H(D|C) = H(D|B - \{a_i\})$, 若成立, 则属性 a_i 是 B 相对于 D 不必要的; 否则, 属性 a_i 不能被约简, B 不变.

Step 4, 输出 B , 即为决策表的 1 个知识约简.

4 某水库台风降雨影响因子挖掘

某水库位于浙江省鳌江最大支流横阳支江上游, 总库容 $8433 \times 10^4 \text{ m}^3$, 是一座以防洪、灌溉为主, 结合发电的综合利用中型水利枢纽. 水库所处苍南县是受台风影响与登陆的主要区域 (据统计, 仅 1980~1990 年 11 年间, 浙江省发生的较大的洪涝灾害中有 19 次是由台风引起), 水库自 1989 年投入蓄水运行以来, 在历次台风暴雨中, 拦蓄上游洪水, 充分发挥了拦洪错峰等综合效益, 减轻下游损失, 保证了下游县城和乡镇百万人民群众生命财产安全. 但同时由于台汛期库水位的剧烈变化, 大坝每次都相当于受到突然加荷和迅速卸荷的突变过程, 每年汛期后, 大坝防渗结构均有新的裂缝、变形及渗漏等情况发生, 需要加强管理和不断加固处理. 因此, 挖掘台风形成库区降雨的影响因子, 对水库台风监测及库区降雨强度分析, 应对台汛期洪水、保障水库安全和加强防灾减灾管理具有指导意义.

选取监测资料较为齐全的 2005 年该水库遭受“海棠”台风时的库区降雨进行分析^[8]. 考虑影响台风降雨的因素较多且复杂, 以及获取台风遥测资料的有限性, 这里对于特定工程而言, 忽略环境因素影响, 重点研究台风风速、风力、7 级风圈半径、10 级风圈半径、中心气压、台风中心与影响库区距离等典型台

风特性因子对库区降雨的影响.

4.1 台风降雨监测决策信息表构建

2005 年 5 号台风“海棠”7 月 19 日 17 时在福建省连江县登陆, 自 18 日上午开始影响苍南县. 图 1 为“海棠”登陆路径图, 可见海棠距离库区较近, 最近时不到 200 km, 对形成库区降雨有着重要影响. 据库区雨量遥测站的监测资料, 这次台风带来的暴雨强度大, 降雨集中, 平均过程降雨量 552 mm, 超过台风汛期 20 年一遇的设计暴雨. 部分“海棠”台风特性参数与库区降雨量监测数据见表 1.

对所获得的监测信息, 构建台风特性参数相关属性及决策信息系统 $S=(U,CUD,V,f)$, 其中: 条件属性 c_1 —最大风速; c_2 —风力; c_3, c_4 —7 级, 10 级风圈半径; c_5 —中心气压; c_6 —中心与库区距离; 决策属性 d —工程区降雨. 其属性值分别为: $V_{c1}=\{1,2,3,4\}$, 其中 1—风速 <25 m/s; 2—风速为 25~35 m/s; 3—风速为 35~45 m/s; 4—风速 ≥ 45 m/s. $V_{c2}=\{1,2,3,4\}$, 其中 1—风力 <10 级; 2—风力为 10, 11 级; 3—风力为 12, 13

级; 4—风力 ≥ 14 级. $V_{c3}=\{1,2,3,4,5\}$, 其中 1—7 级风圈半径 <350 km; 2—7 级风圈半径为 350~400 km; 3—7 级风圈半径为 400~450 km; 4—7 级风圈半径为 450~500 km; 5—7 级风圈半径 ≥ 500 km. $V_{c4}=\{1,2,3,4\}$, 其中 1—10 级风圈半径 <50 km; 2—10 级风圈半径为 50~100 km; 3—10 级风圈半径为 100~150 km; 4—10 级风圈半径为 150~200 km; 5—10 级风圈半径 ≥ 200 km. $V_{c5}=\{1,2,3,4,5\}$, 其中 1—中心气压 <940 百帕; 2—中心气压为 940~955 百帕; 3—中心气压为 955~970 百帕; 4—中心气压为 970~985 百帕; 5—中心气压 ≥ 985 百帕. $V_{c6}=\{1,2,3,4\}$, 其中 1—距离 <250 km; 2—距离为 250~350 km; 3—距离为 350~450 km; 4—距离 ≥ 450 km. $d=\{1,2,3,4,5\}$, 其中 1—降雨量 <5 mm; 2—降雨量为 5~15 mm; 3—降雨量为 15~25 mm; 4—降雨量为 25~35 mm; 5—降雨量 ≥ 35 mm.

按照上述说明对相关属性进行离散化, 删除重复属性条件和不能分辨的对象, 由此建立台风环境量相关属性及决策信息系统 $S=(U,CUD,V,f)$ 如表 2

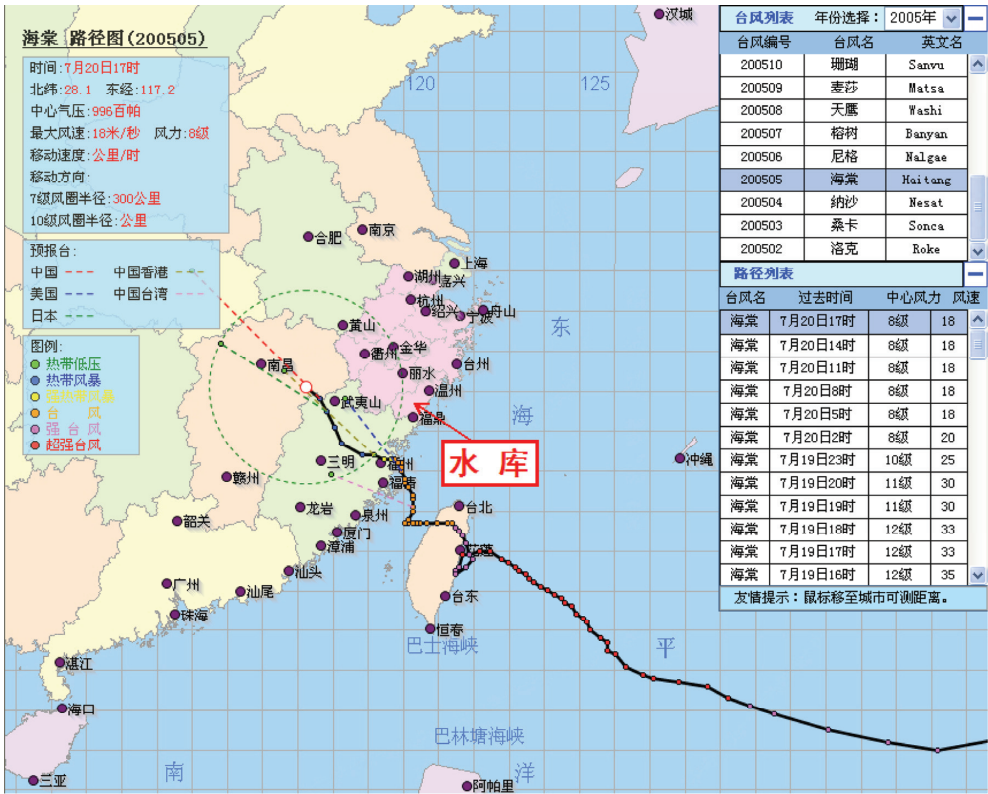


图 1 2005 年 5 号台风“海棠”登陆路径图

表 1 部分“海棠”台风特性参数与库区降雨量监测数据

时间	台风特性参数						库区降雨量 (mm)
	最大风速 (m/s)	中心风力	风圈半径 (km)		中心气压 (百帕)	中心与库区距离 (km)	
			7 级	10 级			
05-7-18-09:00	50	15	520	200	935	520	11
05-7-18-11:00	45	14	480	180	940	540	4
05-7-18-16:00	45	14	480	180	940	430	3
05-7-18-18:00	45	14	480	180	940	400	13
05-7-18-19:00	40	13	450	150	955	380	2
05-7-18-22:00	40	13	450	150	955	370	7
05-7-18-23:00	40	13	450	150	955	360	15
05-7-19-02:00	40	13	450	150	960	330	6
05-7-19-05:00	40	13	400	120	960	330	19
05-7-19-06:00	40	13	400	120	960	330	33
05-7-19-07:00	40	13	400	120	960	300	19
05-7-19-10:00	40	13	400	120	960	270	45
05-7-19-11:00	40	13	400	120	960	260	38
05-7-19-14:00	35	12	400	100	970	190	27
05-7-19-18:00	33	12	400	100	975	190	14
05-7-19-19:00	30	11	380	80	980	190	9
05-7-19-20:00	30	11	380	70	980	190	2
05-7-19-23:00	25	10	350	50	985	200	
05-7-20-04:00	20	8	350		985	250	5
05-7-20-08:00	18	8	300		990	260	
05-7-20-10:00	18	8	300		990	260	15

表 2 台风相关属性及决策信息约简表

序号	属性集 c						决策集
	c_1	c_2	c_3	c_4	c_5	c_6	d
1	4	4	5	5	1	4	2
2	4	4	4	4	1	4	2
3	4	4	4	4	1	3	1
4	4	4	4	4	1	3	2
5	3	3	4	4	2	3	2
6	3	3	4	4	2	3	3
7	3	3	4	4	3	2	2
8	3	3	3	3	3	2	3
9	3	3	3	3	3	2	4
10	3	3	3	3	3	2	5
11	3	3	3	3	4	1	2
12	3	3	3	3	4	1	4
13	2	3	3	3	4	1	2
14	2	2	2	2	4	1	1
15	2	2	2	2	4	1	2
16	2	2	2	2	5	1	1
17	1	1	2	1	5	2	2
18	1	1	1	1	5	2	1
19	1	1	1	1	5	2	3

所示.

4.2 基于条件信息熵的属性约简与知识约简

不同的条件属性得到的不同的等价类 $U/\{c_1\}$, $U/\{c_2\}$, $U/\{c_3\}$, $U/\{c_4\}$, $U/\{c_5\}$, $U/\{c_6\}$, $U/\{d\}$ 的信息熵按(3)式计算如下:

$$\begin{aligned} H(c_1) &= 1.312, H(c_2) = 1.265, H(c_3) = 1.448, \\ H(c_4) &= 1.466, H(c_5) = 1.572, \\ H(c_6) &= 1.548, H(d) = 1.365. \end{aligned}$$

由信息熵计算结果可以看出, 划分越细, 则信息熵的值越大.

不考虑决策属性 D , 利用节 3.2 算法进行属性约简, 可得最终的信息表的一个约简为

$$C_0 = \{c_1, c_3, c_5, c_6\}. \quad (6)$$

按(5)式各条件属性熵计算值如下:

$$\begin{aligned} H(D|C) &= 1.579, H(D|\{c_1\}) = H(D|\{c_2\}) \\ &= 0.941, H(D|\{c_3\}) = 0.996, \\ H(D|\{c_4\}) &= 1.069, H(D|\{c_5\}) \\ &= 1.160, H(D|\{c_6\}) = 1.260. \end{aligned}$$

考虑决策属性 D , 利用节 3.3 算法对决策表进行约简, 由此得到决策表的 1 个相对约简为

$$B = \{c_1, c_3, c_6\}. \quad (7)$$

4.3 成果分析

对“海棠”台风所获得的特性参数监测数据进行属性约简和知识约简, 可获得台风的条件属性分类描述约简集和形成库区降雨的决策约简集, 结论认

为: 台风的风速、风力、7 级/10 级风圈半径、中心气压、与水工结构距离等典型台风特性因子中, 经过熵理论分类描述约简, 主要特性因子为风速、7 级风圈半径、中心气压、与水工结构距离等 4 项; 从形成库区降雨效应量角度考虑, 对台风的特性因子和库区降雨量进行决策约简, 认为台风风速、中心气压、与水工结构距离等 3 项对库区降雨起主要作用, 是台风形成库区降雨的主要影响因子.

5 结论

台风形成的强降雨对水库大坝性态安全的影响和致灾效应不容忽视. 本文考虑影响台风降雨因素的复杂性以及现有水库台风遥测资料的不完备性, 采用基于信息熵与条件信息熵建立起来的不完备信息系统的约简方法, 对台风形成库区降雨的影响因子进行挖掘. 以典型水库中的台风特性参数及效应量监测信息为依据, 忽略环境因素影响, 重点分析了包括台风风速、风力、7 级/10 级风圈半径、中心气压、中心与影响工程区距离等典型台风特性因子对水库大坝效应量(库区降雨量)的影响. 分析结论认为: 不考虑对库区降雨, 经属性约简后台风风速、7 级风圈半径、中心气压、与水工结构距离等 4 项为台风主要特性因子; 考虑库区降雨, 经知识约简台风风速、中心气压、与库区距离等 3 项台风特性因子对库区降雨起主要作用, 是主要的降雨影响因子. 分析结果对水库台风监测及库区降雨强度分析具有一定指导作用.

由于台风降雨问题的复杂性, 本文根据特定工程典型台风特性因子遥测数据所获取的分析成果还有待进一步验证和完善.

参考文献

- 崔星, 袁丽侠, 陆彦俊. 台风诱发滑坡灾害的机理. 自然灾害学报, 2010, 4: 80-84
- 施齐欢, 潘仲力, 童霄. 丁香坞山塘大坝背水坡滑坡原因分析. 浙江水利科技, 2007, 149: 29-30, 42
- 钮学新, 杜惠良, 滕代高, 等. 影响登陆台风降水量的主要因素分析. 暴雨灾害, 2010, 29: 76-80
- Rodgers E B, Pierce H F. Environmental influence on typhoon bobbie's precipitation distribution. J Appl Meteorol, 1995, 34: 2513-2532
- 程正泉, 陈联寿, 刘燕, 等. 1960-2003 年我国热带气旋降水的时空分布特征. 应用气象学报, 2007, 18: 427-434
- 仪垂祥. 非线性科学及其在地学中的应用. 北京: 气象出版社, 1995
- 副昂, 王国胤, 胡军. 基于信息熵的不完备信息系统属性约简算法. 重庆邮电大学学报(自然科学版), 2008, 20: 586-592
- Li Z Y, Ma F H, Yuan H, et al. Nonlinear analysis of reservoir rainfall caused by typhoon. In: Proceedings of Earth and Space 2012, ASCE, 2012. 920-926

Impact factors mining of reservoir rainfall caused by typhoon

LI ZiYang¹, ZHANG Zhan², MA FuHeng¹, LI ChengCai³ & PAN ShuJun⁴

¹ *Nanjing Hydraulic Research Institute, State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing 210029, China;*

² *College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China;*

³ *College of Civil and Transportation Engineering, Hohai University, Nanjing 210098, China;*

⁴ *Hohai University Design Institute CO, Ltd, Nanjing 210098, China*

Considering the complexity of the factors affecting typhoon rainfall and incompleteness of typhoon monitoring data, based on incomplete information system reduction method set up by information entropy and information entropy conditions, hazard factors of typhoon causing reservoir rainfall are excavated. Taking the monitoring data of typhoon parameters and reservoir rainfall of some typical engineering as an example, the relationships between the amount of reservoir rainfall and typhoon characteristic factors are analyzed, including wind speed, wind force, 7/10 wind circle radius, central pressure, the distance between the typhoon center and engineering area without the environment factors. With information reduction, the conclusion are that the typhoon wind speed, central pressure and the distance are the major typhoon impact factors. The results have a certain guiding role for the typhoon monitoring and analysis of the rainfall intensity.

typhoon, reservoir rainfall, impact factor, information reduction

doi: 10.1360/N092014-00179