

识别合适的降雨强度-历时-频率模型的方法

陆宝宏¹, 汤有光², 陆晓明³, 李 伟¹, 邱绍伟¹

(1. 河海大学水文水资源及环境学院, 江苏 南京 210098; 2. 香港科技大学, 香港 九龙清水湾;
3. 南京水文及水资源分局, 江苏 南京 210000)

摘要 提出一种基于 IDF 模型对未来事件预测能力来选择合适的 IDF 模型的方法. 通过对香港气象台 13 个不同历时的年最大降雨资料的数据分析, 发现 IDF 模型的预测能力取决于它的数学形式、降雨类型以及模型参数估计所采用的方法.

关键词 雨强-历时-频率模型 交叉检验 参数估计 预测能力 重现期

中图分类号 P33 **文献标识码** A **文章编号** 1000-1980(2001)04-0109-07

在水利工程设计中, 一个特定历时和频率的降雨强度(雨深)常作为设计暴雨^[1~3]. 实际上, 设计暴雨强度(或雨深)可以从降雨强度-历时-频率关系中获得. 为了克服从图或曲线上摘取降雨信息的缺点, 水利工作者推导出了许多 IDF 数学模型^[4,5]. 开发一个 IDF 模型常用的方法是采用一个 IDF 模型, 通过估计它的参数, 使计算的降雨强度(或雨深)和‘观测降雨’拟合得最好. 在参数估计中所谓的‘观测降雨’来自于降雨频率分析. 然而, 一个拟合观测资料很好的模型, 不一定能保证其对未来事件有很好的预测能力. 本文旨在采用统计学中的交叉检验方法对常用的 IDF 模型的预测能力及几种参数估计方法进行识别.

1 降雨强度-历时-频率(IDF)模型

1.1 IDF 模型的形式

在水利工程中, 常用的 IDF 模型可以归纳为三类(见表 1). 模型中, $y(t, T)$ 是重现期为 T (年)和历时为 t (小时或分钟)的降雨量; a, b, c, d, g 和 h 是 IDF 模型的参数. 降雨量 $y(t, T)$ 可以是降雨强度 $I(t, T)$ 、雨深 $d(t, T)$ 和雨强比 $I(t, T)/I(t^*, T^*)$, $I(t^*, T^*)$ 是由事先确定的基本重现期 T^* 和历时 t^* 所对应的降雨强度.

表 1 三种 IDF 模型的分布函数和概率密度函数

Table 1 CDF and PDF of three IDF Models

模型类型	模型形式	分布函数 $F(y_i)$	密度函数 $f(y_i)$
模型-1	$y(t, T) = [g + h \times \log(T)] \frac{a}{(t + c)^d}$	$F_1(y_i) = 1 - \frac{1}{10^{-\frac{1}{h} [\frac{y_i(t+c)^d}{a} - g]}}$	$f_1(y_i) = \frac{\ln(10) \times (t + c)^d}{h \times a \times 10^{\frac{1}{h} [\frac{y_i(t+c)^d}{a} - g]}}$
模型-2	$y(t, T) = \frac{a \times T^b}{(t + c)^d}$	$F_2(y_i) = 1 - \left[\frac{(t + c)^d}{a} y_i \right]^{-\frac{1}{b}}$	$f_2(y_i) = \frac{\left[\frac{(t + c)^d}{a} y_i \right]^{-\frac{1}{b}}}{b \times y_i}$
模型-3	$y(t, T) = \frac{a \times T^b}{t^d + c}$	$F_3(y_i) = 1 - \left[\frac{t^d + c}{a} y_i \right]^{-\frac{1}{b}}$	$f_3(y_i) = \frac{\left[\frac{t^d + c}{a} y_i \right]^{-\frac{1}{b}}}{b \times y_i}$

1.2 IDF 模型参数的确定

估计 IDF 模型参数的传统方法是首先对不同历时的降雨进行频率分析, 然后根据事先选择的重现期和

暴雨历时得到的降雨量(雨深、雨强和雨强比)拟合选择的 IDF 模型.采用最小二乘方准则,其目标函数如下:

$$\min D(\theta) = \sum_T \sum_t [y(t, T) - \hat{y}(t, T/\theta)]^2 \tag{1}$$

其中 $y(t, T)$ 来自于频率分析, $\theta = (g, h, a, c, d)$ 是 IDF 模型参数, $\hat{y}(t, T/\theta)$ 是重现期为 T 、历时为 t 的降雨量的估计.下面将介绍文中采用的方法.

a. 模型-1 的参数估计.文中采用两种方法来估计 IDF 模型-1 的参数, $\theta = (g, h, a, c, d)$.采用降雨强度比,模型-1 可被分解成为 $\frac{I(t, T)}{I(t^*, T^*)} = r(T) \times s(t)$.因此, $r(T)$ 和 $s(t)$ 可表示为

$$r(T) = \frac{I(t, T)}{I(t, T^*)} = g + h^* \log(T) \tag{2}$$

对于所有历时 t

$$s(t) = \frac{I(t, T^*)}{I(t^*, T^*)} = \frac{a}{(t + c)^d} \tag{3}$$

通过简单的线性回归,可以得到参数 g, h .显然,选定基本重现期 T^* ,参数 g 和 h 必须满足 $g + h^* \log(T^*) = 1$ 的限制.因此,参数 h 的约束最小二乘估计为

$$\hat{h} = \frac{\overline{r_i - 1}}{\log(T_i/T^*)} \tag{4}$$

其中 $\overline{r_i - 1}$ 和 $\log(T_i/T^*)$ 分别是 $r_i - 1$ 和 $\log(T_i/T^*)$ 的均值, $r_i - 1 = \frac{y(t, T)}{y(t^*, T^*)} - 1$.参数 g 的最小二乘估计为

$$\hat{g} = 1 - \hat{h}^* \log(T^*)$$

估计方程(3)中的参数 a, c 和 d .假定一个 c 值,方程(3)可转化为

$$\log\left[\frac{I(t, T^*)}{I(t^*, T^*)}\right] = \log(a) - d^* \log(t + c^*) \tag{5}$$

可以用简单的线性回归估计方程(5)中的参数 a 和 d 结合黄金分割法来反复估计参数 a, d 和 c 的最优值.

模型-1 中的参数 $\theta = (g, h, a, c, d)$ 还可以采用最小二乘方准则估计出,目标函数如下:

$$\min D_1(g, h, a, d, c) = \sum_T \sum_t \left\{ y(t, T) - [g + h^* \log(T)] \frac{a}{(t + c)^d} \right\}^2 \tag{6}$$

可用共轭梯度及其它非线性算法估计参数.由于共轭梯度法及一般非线性算法的参数估计常收敛于局部最优值,所以很难找到全局最优.主要原因是目标函数非线性较强和方程(6)中指数“ d ”的目标函数梯度远远大于其它参数的梯度,导致沿着参数“ d ”的梯度方向搜索最优值.为改善解的最优性,文中采用模拟退火算法^[6].模拟退火算法是一种随机搜索方法,它避免搜索时停留在一个局部最优值,是根据概率接受准则,通过接受劣解而实现的.在不断搜索的过程中,接受劣解的可能性降为零.根据经验,交替使用模拟退火算法和共轭梯度法可以得到模型的最优参数值.

b. 模型-2 的参数估计 模型-2 中的参数 $\theta = (a, b, c, d)$ 可以通过式(7)求解:

$$\min D_{2A}(\theta) = \sum_T \sum_t \{ \log[y(t, T)] - [\log(a) + b^* \log(T) - d^* \log(t + c)] \}^2 \tag{7}$$

其中参数 a, b, c, d 用多元线性回归结合黄金分割法估计,或直接采用共轭梯度法和模拟退火法相结合来估计模型参数,目标函数如下:

$$\min D_{2B}(\theta) = \sum_T \sum_t \left\{ y(t, T) - \frac{a^* T^b}{(t + c)^d} \right\}^2 \tag{8}$$

c. 模型-3 的参数估计 模型-3 无法进行对数变换,可交替使用模拟退火算法和共轭梯度法,目标函数如下:

$$\min D_3(a, b, d, c) = \sum_T \sum_t \left\{ y(t, T) - \frac{a^* T^b}{t^d + c} \right\}^2 \tag{9}$$

2 IDF 模型的校验

可以理解一个与观测资料拟合很好的模型不一定有令人满意的预测能力.本文旨在识别有着令人满意

的预测未来降雨能力的 IDF 模型. 评价一个模型预测能力常用的方法是进行参数分析,分析时要随机生成资料. 由于参数分析得到的结论是建立在随机生成理论上的,而不同历时的年最大降雨的真实分布又是未知的.

模型校验分为两个部分,因此需要估计及校验资料集. 估计资料集用来估计模型的参数,校验资料集用来预测模型的输出,确定模型的预测能力. 获得校验资料集的方法之一是收集新资料,但获取新资料需要时间及额外的费用. 对于某些工程问题来说,无实用价值,有时甚至会导致决策延误. 另一种可行的方法是把现有资料进行分割,如何分割数据是一个重要的统计问题,统计及水利工作者已研究出了多种分割方法. 为了评价不同 IDF 模型的预测性,文中采用 leave-half-out 交叉检验方法^[7~10]. 假定不同历时的年最大降雨强度(或降雨强度比)实际观测资料独立且分布一致,则可以推导出对应于每个模型的不同降雨强度的概率分布及密度函数(见表 1).

由于每个观测值的出现都受到 IDF 模型自身的统计规律支配,因此文中评价各种 IDF 模型预测能力以校验降雨量发生的似然值为准则. 三个模型 K 个随机观测值的对数似然函数分别为

$$\ln \xi_1[y_{1,v}(t), y_{2,v}(t), \dots, y_{k,v}(t)] = K[d \ln(t+c) + \ln \ln(10) - \ln(h) - \ln(a) + \frac{g}{h} \ln(10)] - \left[\frac{\ln(10)}{ha}\right] (t+c)^d \sum_{k=1}^k y_{k,v}(t) \tag{10}$$

$$\ln \xi_2[y_{1,v}(t), y_{2,v}(t), \dots, y_{k,v}(t)] = K \left\{ \left(\frac{-1}{b}\right) [d \ln(t+c) - \ln(a)] - \ln(b) \right\} - \left(1 + \frac{1}{b}\right) \sum_{k=1}^k \ln[y_{k,v}(t)] \tag{11}$$

$$\ln \xi_3[y_{1,v}(t), y_{2,v}(t), \dots, y_{k,v}(t)] = K \left\{ \left(\frac{-1}{b}\right) [d \ln(t^d+c) - \ln(a)] - \ln(b) \right\} - \left(1 + \frac{1}{b}\right) \sum_{k=1}^k \ln[y_{k,v}(t)] \tag{12}$$

式中 $y_{1,v}(t), y_{2,v}(t), \dots, y_{k,v}(t)$ 是校验集合中历时为 t 的 K 个随机观测值.

假定有 M 个不同历时的 N 年年最大降雨,文中采用的 Leave-half-out 交叉检验算法步骤概括如下:

- (a) 从不同历时的 N 年降雨系列中,随机、不重复地抽取 $N/2$ 年的资料,作为估计资料集,剩余的 $N/2$ 年的数据,作为检验资料集.
- (b) 根据估计资料集中的年最大降雨进行频率分析,得到对应于一定的重现期和历时的降雨(雨深、雨强或雨强比)的估计值.
- (c) 根据第(b)步的结果,选择一个优化方法来估计 IDF 模型的参数.
- (d) 用估计的参数及校验资料计算 IDF 模型的似然函数值.
- (e) 不断地重复步骤(a)~(d),计算每一个 IDF 模型的似然函数的均值.
- (f) 选择具有最大似然函数值的模型.

3 应 用

应用由香港气象台观测的 1884~1939 年期间和 1947~1990 年期间(共 100 年)13 个历时($t=1\text{ h}, 2\text{ h}, 4\text{ h}, 6\text{ h}, 8\text{ h}, 12\text{ h}, 16\text{ h}, 24\text{ h}, 2\text{ d}, 3\text{ d}, 4\text{ d}, 5\text{ d}, 7\text{ d}$)的年最大降雨资料^[12]来识别一个合适的 IDF 模型,文中选用的重现期 T 分别为 5,10,100,200,500,1 000,2 000,5 000 及 10 000 年. 本研究涉及两个主要问题:(a) IDF 模型的形式,包括是采用降雨强度还是采用降雨强度比;(b) 模型参数估计方法. 文中采用 L-moments 方法^[12]进行频率分析,为了考虑分布模型的不确定性,文中采用 9 种分布(正态,对数正态, P-3, 对数 P-3, 概化极值, 概化 logistic, 概化 Pareto, gumbel, gamma 分布)的坐标加权法建立最终的降雨强度-频率关系.

在交叉校验方法中,对于 100 年而言,随机不重复地抽取 50 年($N_{\text{est}}=50$)不同历时的年最大降雨来组成估计资料系列. 这个步骤重复 1 000 次. 然后根据由校验资料系列似然函数值的均值来比较三种 IDF 模型在不同条件下的预测能力.

4 结果和讨论

文中采用基准历时 $t^* = 1\text{ h}$,基准重现期 $T^* = 25\text{ a}$,用前面介绍的方法估计参数,用极大似然函数准则校验模型的输出成果.

应用所有可得到的数据,采用不同的方法估计的模型参数值分别列在表 2~3 中.如果基于拟合优度准则采用传统的方法模型,将选择以降雨强度比表示的模型-1 和模型-2,以及选择以降雨强度表示的模型-3.正如下面所显示的,预测性准则的运用导致了与传统方法截然不同的选择结果.

表 2 降雨强度比的 IDF 模型参数估计值
Table 2 Estimated IDF model parameters based on rainfall intensity ratio

模型	参数估计方法	g	h	a	b	c	d	均方差
模型-1	分解法	0.4946	0.5054	33.00		47.20	0.7196	0.291
	最小二乘法	(0.4329)*	(0.5349)					
	模拟退火算法	0.8752	0.328					
	共轭梯度法	0.8745	0.3268					
模型-2	最小二乘法			38.80	0.1434	103.90	0.7711	0.185
	模拟退火算法			50.80	0.0970	109.60	0.7761	0.129
	共轭梯度法			50.80	0.0924	109.60	0.7678	0.120
模型-3	模拟退火算法			132.00	0.0733	83.60	0.8745	0.135
	共轭梯度法			132.00	0.0776	83.60	0.8868	0.133

* ()中的参数值是根据无约束的最小二乘法计算得来.

表 3 降雨强度的 IDF 模型参数估计值
Table 3 Estimated IDF model parameters based on rainfall intensity

模型	参数估计方法	G	h	a	b	c	d	均方差
模型-1	模拟退火算法	3.5040	1.4162	88.9		2.31	0.3939	11.785
	共轭梯度法	3.5262	1.4853					
模型-2	最小二乘法			3181.4	0.1434	103.90	0.7711	15.121
	模拟退火算法			2836.7	0.1296	123.30	0.7326	13.070
	共轭梯度法			2836.7	0.0994	123.30	0.6938	11.480
模型-3	模拟退火算法			1086.2	0.1079	13.20	0.6268	10.121
	共轭梯度法			1086.2	0.1072	13.20	0.6268	9.910

a. 基于 IDF 模型的降雨强度比:冷降雨量为降雨强度比,即, $y(t,T) = I(t,T)/I(t^*,T^*)$,用不同的最优化方法来估计 IDF 模型的参数.在 $T^* = 25\text{ a}$ 和 $N_{\text{est}} = 50$ 的条件下,各个模型似然函数均值随历时的变化见图 1.显然,各模型似然函数均值一般随着暴雨历时的增加而增加.IDF 模型-2 和模型-3 的预测能力要明显高于模型-1 的,而且随着暴雨历时的增加它们之间的差距也逐渐增加.模型-2 和模型-3 的预测能力非常接近,差别在于模型-3 的预测性要稍微好于模型-2.如果根据均方差最小(见表 2)选择模型,会选择模型-1 或模型-2,而不是模型-3.

下面的两条曲线显示的是用不同的参数估计方法得到的模型-1 的结果.模型-1A 的参数应用式(4)和(5)估计而得.而模型-1B 的参数来自式(6)的估计.显然,后一种方法得到的参数值的预测能力要高于前者,与表 2 和表 3 中拟合的结果一致.显然,IDF 模型的参数估计应该用原始的降雨资料而不该用经过

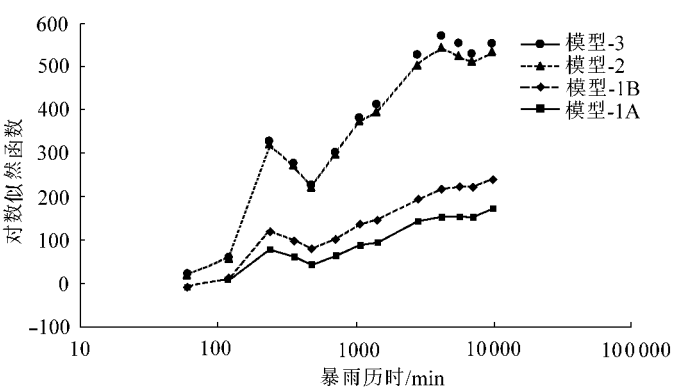


图 1 基于降雨强度比的参数,雨强比似然值随历时的变化

Fig.1 Temporal variation of log-likelihood values of rainfall intensity ratio based on ratio parameters

对数变换后的资料.

为了比较降雨强度比与其它两种形式之间的区别,文中用降雨强度比得到的模型参数分别计算校验资料的似然函数值,结果分别显示在图 2 和图 3 中.图 2 中的曲线趋势与图 1 中的很相似,只不过强度的似然函数值要低于强度比的值.在图 3 中可以发现采用雨深得到的模型-2 和模型-3 的曲线变化趋势与图 1 中的基本一致.值得注意的是在图 3 中模型-1 的似然函数值和历时的关系曲线呈下降趋势.

b. 基于 IDF 模型的降雨强度 在这里需要引起注意的是不同形式的 IDF 模型的预测能力受到降雨量 $y(t,T)$ 的影响,结果如图 4~6 所示,图中的极大似然值是根据拟合降雨强度的参数计算而来的.

从图 4~6 中可以看出根据拟合雨强的参数计算模型-1 的预测能力要高于其它两种模型.模型-2 和模型-3 的预测能力还是差不多,只不过模型-3 要好一些.同时可以看出,三个模型的似然值随历时变化趋势一致,且这三种模型预测能力的差异在所研究的历时范围内保持不变.不管预测的是哪一种降雨量,这里的极大似然值与历时关系仍然不变,而且,如果用所有可得到的数据进行参数估计,将会选择模型-3.

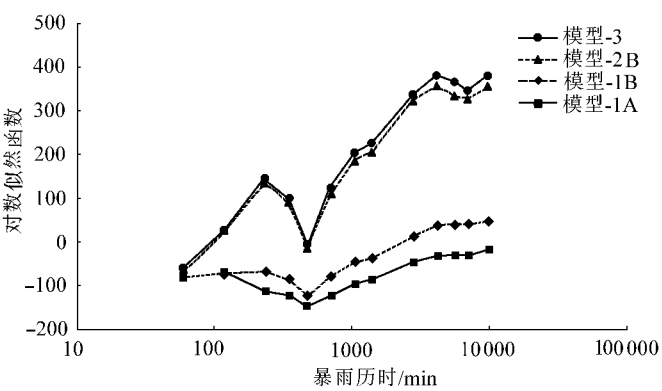


图 2 基于降雨强度比的参数,雨强似然值随历时的变化

Fig.2 Temporal variation of log-likelihood values of rainfall intensity based on ratio parameters

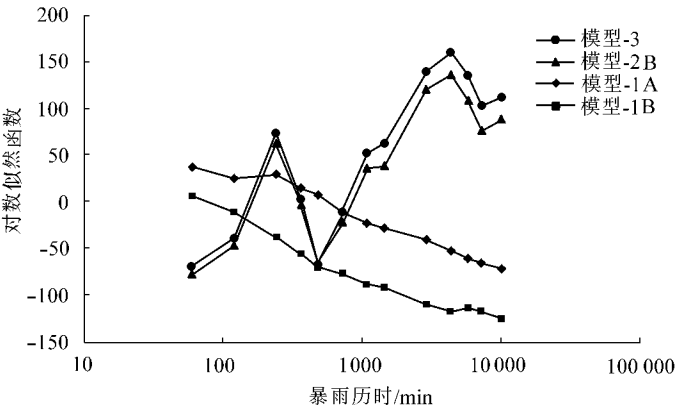


图 3 基于降雨强度比的参数,雨深似然值随历时的变化

Fig.3 Temporal variation of log-likelihood values for rainfall depth based on ratio parameters

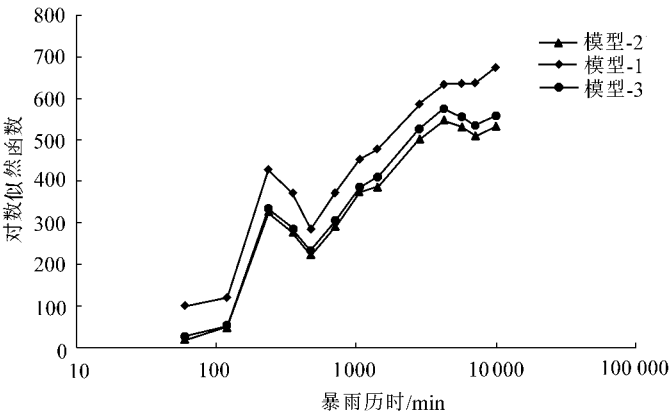


图 4 基于降雨强度的参数,雨强比似然值随历时的变化

Fig.4 Temporal variation of log-likelihood values of rainfall intensity ratio based on intensity parameters

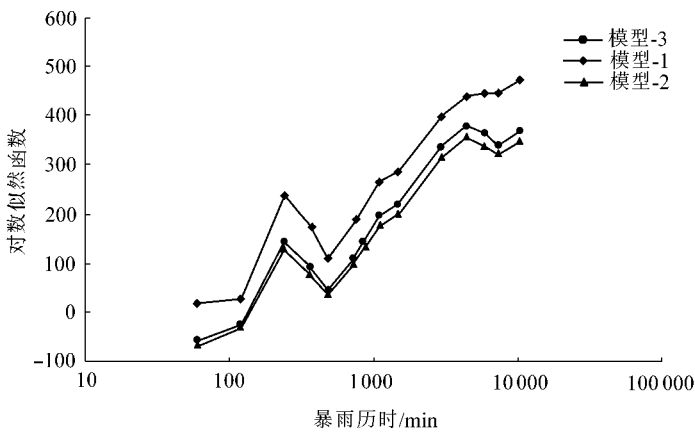


图 5 基于降雨强度的参数,雨强似然值随历时的变化

Fig.5 Temporal variation of log-likelihood values of rainfall intensity based on intensity parameters

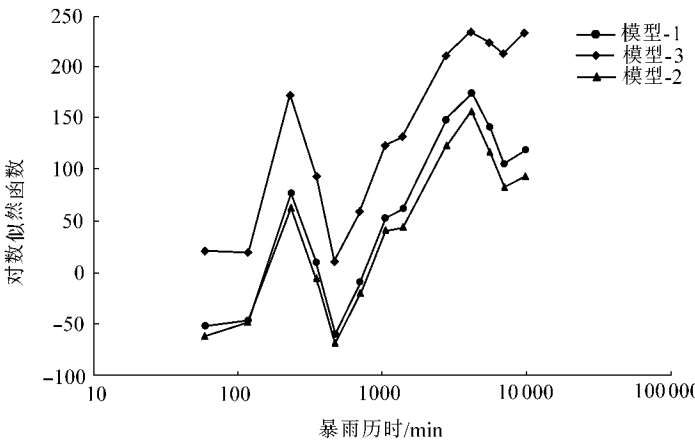


图 6 基于降雨强度的参数,降雨深度似然值随历时的变化

Fig.6 Temporal variation of log-likelihood values for rainfall depth based on intensity parameters

c. 参数对模型预测性的影响 随着模型参数的增加,模型的自由度也在增加,这样会导致模型的输出结果更接近实测值.然而,却未必会得到如表 2 所示的较高的预测性.根据极大似然函数值,以降雨强度比参数计算的模型-3 的预测能力要高于模型-1 的.

5 结 论

本文建议采用交叉校验的方法评价 IDF 模型的预测能力,其预测能力受到模型形式和参数估计方法的影响.这种用交叉检验方法来评价模型预测性的主要优点在于它直接使用已有的数据进行检验而不必等待新的观测数据或依赖人工生成的数据.

- 对香港气象台提供的 100 年不同历时的年最大降雨所得到的结果揭示了以下几点：
- a. 从预测性的角度,本文建议香港气象台采用以降雨强度比-历时-频率关系表示的模型-2 和模型-3 并且取基准重现期为 25 年、基准历时为 1 h 的基准数据来计算降雨强度比.如果选用降雨强度-历时-频率模型,建议采用模型-1.
 - b. 如果在原始数据中用最小误差标准来估计模型参数而不是用经过对数变换的,那么所选择的 IDF 模型就可能得到较好的预测效果.这也暗示要采用最优化方法来估计参数.
 - c. 不同 IDF 模型预测成果的好坏取决于所采用的降雨量的形式.在香港,当选用降雨强度比模型时应选择模型-2 和模型-3;当选用降雨强度模型时,应选择模型-1.

参考文献:

- [1] Huff F A. Time distribution of rainfall in heavy storms [J]. Water Resources Research, 1967, 3(4): 1007 ~ 1019.
- [2] Stedinger J R, Vogel R M, Foufoula-Georgiou E. Handbook of Hydrology [M]. New York: McGraw-Hill Book Company, 1993. 988 ~ 1113.
- [3] LU Bao-hong, TUNG Yeou-Kuong. Investigating inherent probabilistic characteristics of rainfall IDF models [A]. In WANG Zhi-yin, eds. Proceedings of the 8-th International Symposium (2000) on Stochastic Hydraulics [C]. Nether Land: Balkema AA/Rotterdam/Book Field, 501 ~ 505.
- [4] Chen C L. Rainfall intensity-duration-frequency formulas [J]. J of Hydraul Eng, ASCE, 1983, 109(12): 1603 ~ 1621.
- [5] Pilgrim D H, Cordery I. Rainfall temporal patterns for design flood [J]. Hydraul Eng, ASCE, 1975, 101(HY): 81 ~ 95.
- [6] Guus C, Boender E, Romeijn H E. Stochastics methods in handbook of global optimization [M]. London: Kluwer Academic Publishers, 1995. 829 ~ 865.
- [7] Zhao B, Tung Yeou-Kuong, Yang J C. Estimation of unit hydrographs by ridge least square method [J]. Irrigation and Drainage Engineering, ASCE, 1995, 121(3): 253 ~ 259.
- [8] Shih S D, Tung Yeou-Kuong. Internal layers of a transient convection-diffusion problem by perturbation methods [J]. Advances in Water Resources, 1995, 18(2): 111 ~ 120.
- [9] Zhao B, Tung Yeou-Kuong, Yeh K C, et al. Statistical validation methods: applications to unit hydrographs [J]. Hydraulic Engineering, ASCE, 1995, 121(8): 618 ~ 625.
- [10] Tung Yeou-Kuong, Yeh K C, Yang J C. Regionalization of unit hydrograph parameters: 1. Comparison of regression analysis techniques [J], Stochastic Hydrology and Hydraulics, 1997, 11(2): 145 ~ 172.
- [11] Lam C C, Leung Y K. Extreme rainfall statistics and design rainstorm profiles at selected locations in Hong Kong [Z]. Technical Note No. 86, Hong Kong Observatory, 1994. 31 ~ 38.
- [12] Hosking J R M, Wallis J R. Regional frequency analysis—an approach based on L-moment [M]. United Kingdom: Cambridge University Press, 1997. 8 ~ 31.

Method for Identifying Appropriate Rainfall Intensity-Duration-Frequency Models

LU Bao-hong¹, TUNG Yeou-Kuong², LU Xiao-ming³, LI Wei¹, QIU Shao-wei¹

(1. College of Water Resources and Environment, Hohai Univ., Nanjing 210098, China;

2. The Hong Kong University of Sciences and Technology, Hong Kong;

3. Hydrology and Water Resources Investigation of Nanjing City, Nanjing 210000, China)

Abstract Rainfall intensity-duration-frequency (IDF) relationships are often used to determine the design storms in water resources projects. Several mathematical models have been used in practice to describe such relationships. Based on the predicting capability of models for future events, a procedure for the selection of an appropriate rainfall IDF is presented in this paper. Numerical analysis of annual maximum rainfall data for different periods at the Hong Kong Observatory shows that the predicting capability of an IDF model depends on its mathematical forms, the types of rainfall, and the method for the estimation of model parameters.

Key words IDF model; cross-validation; parameter estimation; predicting capability; return period