

降雨观测误差抗差估计方法比较研究

瞿思敏,包为民,石朋,赵超

(河海大学水文水资源与水利工程科学国家重点实验室、河海大学水资源环境学院,江苏 南京 210098)

摘要 根据遥测系统雨量观测资料误差的类型和特点,在抗差最小二乘法的基础上提出了雨量观测误差的三步修正方法.其中,每一步都根据雨量资料误差的不同特点,用不同的权函数进行修正.权函数设计包括保权区、降权区、拒绝区的权函数设计 3 个部分.降权区降权函数的选择无统一标准,所以采用了线性降权函数和抛物线降权函数.同时,根据面平均雨量是否分级 2 种情况和权函数的类型,组成了 4 种方案并对其抗差计算结果进行了比较分析.此外,还选用危水水库乌溪沟以上流域 1974~1999 年间 28 场洪水的暴雨资料进行了抗差分析计算.结果表明,面平均雨量是否分级对抗差效果的影响较大,而降权函数的形式对抗差效果的影响并不显著.

关键词 降雨 观测误差 抗差最小二乘法 权函数

中图分类号 P333

文献标识码 A

文章编号 1000-1980(2006)01-0051-05

降雨资料是洪水预报模型最重要的输入项,其精度将直接影响到洪水预报的精度.文献[1]研究了雨量观测系统误差和随机误差对洪水计算结果的影响,但没有讨论粗差^[2]对洪水计算的影响.

目前,很多流域都利用遥测系统进行雨量观测.水文遥测系统,除存在一般性随机观测误差外,还存在以下几种误差(即粗差):一是信号接收误差.这类误差是在信号接收过程中产生的.如信号碰撞,使信号丢失,会导致系统误认为无降雨;信号干扰,收到误码,会导致降雨观测误差,常使降雨量出现异常情况,有时出现负值,有时出现很大的值(即常说的冒大数现象).如果获得的时段降雨量为负值,遥测系统会进行一些修正;如果获得的值偏大或很大,遥测系统就无法进行实时修正.二是机械误差.例如,翻斗式雨量计的翻转频率低会导致雨量观测值偏小.

抗差估计在水文学中的应用研究已取得了一定的成果^[3-6],但关于如何消除遥测系统中出现的降雨冒大数现象以保证降雨资料的合理性的研究还不多见,尤其是以流域尺度为研究对象的遥测系统的降雨观测误差的修正方法还很缺乏.基于此,本文根据降雨资料误差的类型和特点,设计了用于降雨观测误差修正的多种抗差方案.

1 降雨资料的误差分布

一个流域系统的雨量和观测误差是随时间和空间变化的,这种变化的统计特征对误差估计方法有着直接的影响^[7].

虽然降雨分布受气流、地理位置、地形和高程等因素的影响,变化十分复杂,但其频率分布具有相似性.降雨量的分布主要集中在 $[0, P_{\max}]$ 区间,流域不同,上限值 $P_{\max}$ 也会不同.一般情况下,降雨分布具有特定的高频区和低频区,范围随流域特征而变.

雨量站每个时段的降雨量相对于流域时段平均值的离差绝对值,也随时间和空间而变,其频率分布也具有相似性.离差绝对值的分布主要集中在 $[0, P_{P_{\max}}]$ 区间,流域不同, $P_{P_{\max}}$ 也不同.离差绝对值的分布也有特定的高频区和低频区,范围也随流域特征而变.

雨量时空变化有一定的规律,把雨量作为时空变化的函数,则有

$$P_{\text{est}} = f(\text{经度, 纬度, 面平均雨量}) \quad (1)$$

其中 P_{est} 为估计雨量.用估计雨量减去观测雨量则可以得到雨量模拟误差系列.模拟误差系列的频率分布也

具有相似性. 误差分布具有单峰特征, 峰值一般出现在均值附近.

2 误差估计方法

从上面的叙述可以看出, 流域的时段降雨量、时段雨量离差绝对值及各雨量站模拟误差分布都限制在 $[0, Z_{\max}]$ 的范围内 (Z 可表示时段降雨量、时段雨量离差绝对值、时段雨量模拟误差绝对值), $Z_{\max}$ 从时段降雨量、时段雨量离差绝对值到模拟误差绝对值依次变小. 笔者根据这些特征, 提出了适用于水文学雨量资料观测误差的各种修正方法.

2.1 三步抗差方案权函数的选取

由测量误差理论可知, 界限 $k\sigma$ 之 k 可取 1.5 (根据正态分布规律, 误差在 $\pm 1.5\sigma$ 之外的概率仅为 0.13). 此区间以外的观测值既不能舍弃, 又要限制其有害影响. 当余差超过 $\pm 2.5\sigma$ 时 (根据正态分布规律, 概率为 0.01), 在观测值模式可用的情况下, 这部分数据不应作为有用的观测信息, 故观测值的权为 0.

根据测量误差理论, 可以对 1 h 时段降雨量、1 h 雨量离差绝对值、1 h 模拟误差系列进行选择 and 排频计算, 找出 $p = 1\%$ 和 $p = 13\%$ 的临界值, 并以此来确定每一步的权函数.

三步权函数设计的共同点是 如果时段雨量、降雨离差绝对值、模拟误差超过某一定值 $Z_{1\%}$, 则该观测值的权取为 0 (权值取为 0 的区称为拒绝区) 如果时段雨量、降雨离差绝对值、模拟误差小于某一定值 $Z_{13\%}$, 则该观测值的权取为 1 (权值取为 1 的区称为保权区) 如果时段雨量、降雨离差绝对值、模拟误差介于 $Z_{13\%}$ 与 $Z_{1\%}$ 之间, 则该观测值的权为 $[0, 1]$ 之间的数值 (权值取为 $[0, 1]$ 之间数值的区称为降权区). Z 可取 P, P_D, P_P 系列 (其中 $\{P_i\}$ 为某站降雨量系列, $\{P_{D_i}\}$ 为某站降雨量离差绝对值系列, $\{P_{P_i}\}$ 为某站降雨量模拟误差系列, $i = 1, 2, \dots, n, n$ 为观测值个数). 降权区函数采用了线性函数 $y = ax + b$ (记为 I 型) 和抛物线函数 $y = ax^2 + bx + c$ (记为 II 型) 2 种不同的函数.

a. 雨量分布权函数. 先用超定量法^[8]对暴雨资料进行取样, 然后进行排频计算, 找出 $p = 1\%$ 和 $p = 13\%$ 对应的雨量值 $P_{1\%}$ 和 $P_{13\%}$, 再参照测量学中常用的 IGG 抗差方案^[2]设计出雨量权函数

$$\omega(P_i) = \begin{cases} 1 & (P_i \leq P_{13\%}) \\ (P_{1\%} - P_i) / (P_{1\%} - P_{13\%}) & (P_{13\%} < P_i \leq P_{1\%}) \\ 0 & (P_i > P_{1\%}) \end{cases} \quad (\text{I 型}) \quad (2a)$$

$$\omega(P_i) = \begin{cases} 1 & (P_i \leq P_{13\%}) \\ \frac{P_i^2 - 2P_{1\%}P_i + P_{1\%}^2}{(P_{13\%} - P_{1\%})^2} & (P_{13\%} < P_i \leq P_{1\%}) \\ 0 & (P_i > P_{1\%}) \end{cases} \quad (\text{II 型}) \quad (2b)$$

b. 降雨离差绝对值分布权函数. 同样, 先用超定量法对降雨离差绝对值资料进行取样, 然后进行排频计算, 找出 $p = 1\%$ 和 $p = 13\%$ 对应的雨量离差绝对值 $P_{D_{1\%}}$ 和 $P_{D_{13\%}}$, 再参照 IGG 抗差方案设计出雨量离差绝对值的权函数

$$\omega(P_{D_i}) = \begin{cases} 1 & (P_{D_i} \leq P_{D_{13\%}}) \\ \frac{P_{D_{1\%}} - P_{D_i}}{P_{D_{1\%}} - P_{D_{13\%}}} & (P_{D_{13\%}} < P_{D_i} \leq P_{D_{1\%}}) \\ 0 & (P_{D_i} > P_{D_{1\%}}) \end{cases} \quad (\text{I 型}) \quad (3a)$$

$$\omega(P_{D_i}) = \begin{cases} 1 & (P_{D_i} \leq P_{D_{13\%}}) \\ \frac{P_{D_i}^2 - 2P_{D_{1\%}}P_{D_i} + P_{D_{1\%}}^2}{(P_{D_{13\%}} - P_{D_{1\%}})^2} & (P_{D_{13\%}} < P_{D_i} \leq P_{D_{1\%}}) \\ 0 & (P_{D_i} > P_{D_{1\%}}) \end{cases} \quad (\text{II 型}) \quad (3b)$$

c. 模拟误差分布权函数. 对模拟误差系列进行排频计算 (近似正态分布), 找出 $p = 1\%$ 和 $p = 13\%$ 对应的模拟误差的绝对值 $P_{P_{1\%}}$ 和 $P_{P_{13\%}}$, 再参照 IGG 抗差方案设计出模拟误差的权函数

$$\omega(P_{P_i}) = \begin{cases} 1 & (|P_{P_i}| \leq P_{P_{13\%}}) \\ \frac{P_{P_{1\%}} - |P_{P_i}|}{P_{P_{1\%}} - P_{P_{13\%}}} & (P_{P_{13\%}} < |P_{P_i}| \leq P_{P_{1\%}}) \\ 0 & (|P_{P_i}| > P_{P_{1\%}}) \end{cases} \quad (\text{I 型}) \quad (4a)$$

$$\omega(P_{P_i}) = \begin{cases} 1 & (|P_{P_i}| \leq P_{P_{13\%}}) \\ \frac{P_{P_i}^2 - 2P_{P_{1\%}}|P_{P_i}| + P_{P_{1\%}}^2}{(P_{P_{13\%}} - P_{P_{1\%}})^2} & (P_{P_{13\%}} < |P_{P_i}| \leq P_{P_{1\%}}) \\ 0 & (|P_{P_i}| > P_{P_{1\%}}) \end{cases} \quad (\text{II 型}) \quad (4b)$$

用 2 种方法进行权函数计算. 第 1 种方法(设为方法 1),通过对所有雨量资料进行频率分析和统计计算,得出三步权函数;第 2 种方法(设为方法 2),先对面平均雨量进行分级,然后通过对不同区间雨量资料、雨量离差绝对值资料及模拟误差资料的统计分析求取权函数.

将方法 1、方法 2 与 I 型和 II 型权函数进行组合,得到 4 种不同的计算方案,见表 1.

表 1 方案的不同组合

Table 1 Different combinations of schemes

方案类型	面平均雨量是否分级	权函数类型
A	不分级(方法 1)	I 型
B	分级(方法 2)	I 型
C	不分级(方法 1)	II 型
D	分级(方法 2)	II 型

2.2 计算步骤

2.2.1 方法 1

- a. 对所有站的雨量资料进行第 1 步抗差计算;
- b. 先计算修正过的雨量离差绝对值系列,再进行第 2 步抗差计算;
- c. 先用式(1)计算估计降雨,再计算模拟误差系列,最后进行第 3 步抗差计算.

2.2.2 方法 2

- a. 根据流域特征对面平均雨量进行分级;
- b. 计算每个时段的面平均雨量;
- c. 对每个面平均雨量段的所有雨量资料进行频率分析和统计计算,求出每个分段内的雨量三步修正权函数;
- d. 先计算面平均雨量,再判断其落在哪个区间,最后选用相应的雨量权函数、雨量离差绝对值权函数、模拟误差权函数进行抗差计算.

3 实 例

选用危水水库乌溪沟以上流域 1974 ~ 1999 年间共 28 场洪水的时段降雨量资料进行抗差计算. 该流域面积为 884.2 km²,流域设川山(0.077 4)、清水湾(0.134 3)、大栗树(0.175 1)、仁和坪(0.154 1)、太平街(0.111 8)、熊家岭(0.175 3)、子良坪(0.118 1)和乌溪沟(0.053 9)8 个雨量站(括号中数字为雨量站权重).方法 1 和方法 2 的权函数临界值列于表 2. 参照式(2a) ~ (4b),可得各种抗差方案的雨量三步抗差权函数.

选用 1999 年的 4 场洪水 990625, 990707, 990715 和 990722 来检验三步雨量抗差估计方法的合理性和有效性. 在 4 场洪水的雨量资料中随机地加上冒大数误差,用 A, B, C 和 D 4 种方案进行抗差计算,以比较各方案的优缺点. 该方法考虑了当 1 个站、2 个站、3 个站和 4 个站同时发生粗差的情况. 表 3 是 990625 次洪水加上冒大数误差后采用 4 种抗差估计方案进行抗差估计后的结果. 表 4 是 4 场洪水的误差由冒大数引起时采用 4 种抗差估计方案进行抗差估计后的平均值.

表 2 危水水库权函数临界值

Table 2 Critical values of weight function for Weishui Reservoir

		mm					
方 法	面平均雨量	$P_{1\%}$	$P_{13\%}$	$P_{D_{1\%}}$	$P_{D_{13\%}}$	$P_{P_{1\%}}$	$P_{P_{13\%}}$
方法 1	不分级	76.1	49.7	65.0	39.4	12.5	7.5
	0 ~ 2	10.2	4.3	9.1	3.4	1.25	0.75
	2 ~ 4	17.5	6.6	15.2	4.9	2.5	1.5
方法 2	4 ~ 6	25.8	9.1	22.8	6.4	4.5	2.7
	6 ~ 8	36.6	13.2	32.1	9.6	6.0	3.6
	8 ~ 10	33.9	16.2	27.2	10.8	7.5	4.5
	> 10	52.0	27.0	37.2	15.8	14.0	8.4

表中： $e_{\max}$ 为冒大数误差（实际计算时， $e_{\max}$ 分别取为 5 mm, 10 mm, 15 mm, 20 mm, 30 mm, 40 mm, 50 mm, 60 mm, 70 mm, 80 mm, 90 mm, 100 mm, 200 mm, 300 mm, 400 mm, 500 mm, 表中仅列出了有代表性的几个 $e_{\max}$ ）； r_{11} 、 r_{13} 和 r_{11} 分别为 1 个雨量站冒大数误差估计的第 1 步有效性、第 3 步有效性和总有效性； r_{14} 为 4 个雨量站冒大数误差估计的总有效性。误差估计有效性按下式进行计算。

$$\begin{cases} r_{11} = 1 - \frac{\sum |\hat{e}_{1i}|}{\sum |\hat{e}_{0i}|} \\ r_{13} = 1 - \frac{\sum |\hat{e}_{3i}|}{\sum |\hat{e}_{2i}|} \\ r_{11} = 1 - \frac{\sum |\hat{e}_{3i}|}{\sum |\hat{e}_{0i}|} \end{cases} \quad (5)$$

式中 $\hat{e}_{0i}$ 、 $\hat{e}_{1i}$ 、 $\hat{e}_{2i}$ 和 $\hat{e}_{3i}$ 分别为初始、第 1 步、第 2 步和第 3 步修正后的雨量误差。 r_{14} 的计算式同 r_{11} 。

表 3 990625 次洪水雨量观测冒大数误差估计

Table 3 Estimation of observed extreme value error of rainfall in flood 990625

$e_{\max}$ /mm	r_{11}				r_{13}				r_{11}				r_{14}			
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
5	0	0.144	0	0.252	0	-0.140	0	-0.350	0	0.035	0	0.005	0.058	-0.550	0.058	-0.560
10	0	0.396	0	0.549	0.217	0.281	0.235	-0.040	0.217	0.572	0.235	0.532	0.076	0.350	0.076	0.345
20	0	0.730	0	0.798	0.631	0.424	0.653	0.167	0.631	0.847	0.653	0.833	0.375	0.696	0.375	0.704
50	0.241	0.913	0.283	0.926	0.896	0.395	0.884	0.311	0.921	0.948	0.917	0.949	0.686	0.890	0.699	0.910
100	0.834	0.974	0.849	0.976	0.859	0.432	0.840	0.383	0.977	0.985	0.976	0.985	0.934	0.979	0.942	0.981
200	0.891	0.980	0.898	0.981	0.846	0.437	0.826	0.409	0.983	0.989	0.983	0.989	0.956	0.985	0.961	0.986
500	0.979	0.993	0.979	0.993	0.672	0.421	0.667	0.419	0.993	0.996	0.993	0.996	0.988	0.995	0.988	0.995

表 4 雨量观测冒大数误差估计平均值

Table 4 Average for error estimation of observed extreme value of rainfall

$e_{\max}$ /mm	r_{11}				r_{13}				r_{11}				r_{14}			
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
5	0	0.119	0	0.120	0	-0.350	0	-0.410	0	-0.180	0	-0.230	0.070	-0.800	0.070	-0.800
10	0	0.379	0	0.477	0.073	0.206	0.092	0	0.073	0.524	0.092	0.480	0.134	0.162	0.134	0.171
20	0	0.672	0	0.748	0.492	0.346	0.525	0.125	0.492	0.793	0.525	0.780	0.411	0.606	0.411	0.628
50	0.140	0.881	0.200	0.904	0.876	0.385	0.868	0.260	0.894	0.933	0.896	0.928	0.703	0.855	0.709	0.876
100	0.819	0.967	0.846	0.970	0.832	0.385	0.786	0.351	0.970	0.981	0.968	0.982	0.936	0.973	0.939	0.976
200	0.883	0.976	0.897	0.978	0.813	0.391	0.768	0.359	0.979	0.987	0.977	0.987	0.960	0.983	0.963	0.984
500	0.974	0.991	0.977	0.991	0.668	0.412	0.594	0.383	0.992	0.995	0.992	0.995	0.988	0.994	0.989	0.994

4 结果分析

- 根据表 3 和表 4,可得出如下结论：
- a. 本文提出的雨量误差三步权函数修正方法是有效的。
 - b. 不同量级误差修正效果不一样。小误差修正效果较差,大误差修正效果较好。这说明雨量误差三步权函数修正法比较适用于大误差估计。
 - c. 雨量误差三步权函数修正法,对大误差来说,第 1 步修正效果最好,后 2 步次之。而对小误差来说,第 1 步、第 2 步修正基本不起作用,只有第 3 步修正效果较好。第 2 步修正效果不明显,有时还会出现负效果,这可能是由于抗差过度(即把真实的观测值当作误差处理)造成的。
 - d. 雨量误差三步权函数修正法的效果与冒大数误差产生的站数有一定的关系。产生误差的站数增加时,抗差效果会有一定的降低,但并不是很明显。
 - e. 当误差由冒大数引起时,不论采用的是线性降权函数还是抛物线降权函数,面平均雨量分级比不分级的抗差效果要好。当误差由冒大数引起且面平均雨量不分级时,采用抛物线降权函数比采用线性降权函数

抗差效果要好一些,但并不显著,当误差由冒大数引起且面平均雨量分级时,采用抛物线降权函数不一定比采用线性降权函数抗差的效果差。

f. 面平均雨量是否分级对抗差效果的影响较大,而降权函数的形式对抗差效果的影响并不显著。

参考文献:

- [1] 张正康,胡凤彬,廖劲红,等. 小河配套雨量站网密度分析[J]. 水文,1997(增刊):20-23.
- [2] 周江文,黄幼才,杨元喜,等. 抗差最小二乘法[M]. 武汉:华中理工大学出版社,1997:105-119.
- [3] 包为民,林跃,黄贤庆,等. 水库入库河段洪水汇流参数抗差估计研究[J]. 武汉大学学报:工学版,2004(6):13-16.
- [4] 包为民,瞿思敏,黄贤庆,等. 水文系统抗差权函数分析与检验[J]. 清华大学学报:自然科学版,2003(8):1127-1129.
- [5] 瞿思敏,包为民,石朋,等. AR 模式误差修正方程参数抗差估计[J]. 河海大学学报:自然科学版,2003,31(5):497-500.
- [6] 包为民,嵇海祥,胡其美,等. 抗差理论及在水文学中的应用[J]. 水科学进展,2003(4):528-532.
- [7] 包为民,瞿思敏,李清生,等. 遥测系统降雨观测误差估计方法研究[J]. 水利学报,2003(4):30-34.
- [8] 金光炎. 城市设计暴雨频率计算问题[J]. 水文,2000(2):14-18.

Comparison of robust estimation method for correction of rainfall observation error

QU Si-min, BAO Wei-min, SHI Peng, ZHAO Chao

(State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering,
College of Water Resources and Environment, Hohai University, Nanjing 210098, China)

Abstract: The types and characteristics of rainfall observation error of hydrological remote monitoring system were analyzed, and a three-step weight function correction method for observation error of the system was developed on the basis of the theory of robustified least square method. In each step of the method, different kinds of weight functions were adopted according to the characteristics of observation error, therefore, the weight function design for weight-keeping area, weight-declining area, and weight-zero area was involved. Without a uniform format for determination of weight-declining function in weight-declining area, the linear function and parabolic function were adopted in this area. Four schemes for correction of rainfall observation error were formed with different weight functions adopted under the condition of classification or without classification of mean areal precipitation, and the calculated results of different schemes were compared. The analysis and calculation of rainfall data of 28 floods from 1974 to 1999 of the upper Wuxigou Basin of the Weishui Reservoir show that whether the mean areal precipitation is classified or not is of great influences on robust estimation, however, the influence of format of weight-declining function is not notable.

Key words: precipitation; observation error; robustified least square method; weight function