

洪水预报误差分布的极大熵法

周惠成*, 李丽琴, 王本德

(大连理工大学 土木水利学院, 辽宁 大连 116024)

摘要: 首先根据实际洪水预报误差出现在有限区域的特点,应用极大熵原理,建立了洪水总量预报误差分布的极大熵模型;通过几个不同流域的计算,得出随着降雨量的增大,产流预报误差趋于一个稳定值的结论.同时将该模型计算的分布与正态分布进行了比较,结果表明用极大熵法求得的误差概率分布能更好地描述洪水总量预报误差的分布特性,可以根据实际降雨量的大小确定不同的最大不确定性的误差分布,为分析不同量级洪水预报的风险提供依据.

关键词: 极大熵法; 洪水预报; 误差分布
中图分类号: P333.9 **文献标识码:** A

0 引言

随着信息与预报技术的发展,在实时洪水调度与修改水库洪水调度规划中,人们越来越重视发挥洪水预报的作用^[1].应用洪水预报,既可提高水库防洪系统的防洪标准,减少洪灾损失,又可在保证系统防洪安全的前提下,使水库为兴利多蓄水,提高洪水资源的利用率^[2].可见,洪水预报在充分发挥水库防洪与兴利的双重作用中,起着非常重要的作用.

然而,洪水预报误差是不可避免的.引起洪水预报误差的原因有多种,如模型误差、资料误差、降雨预报误差等.在给定雨量站及其实际降雨信息的条件下,由降雨径流预报模型而引起的预报误差(以下简称产流预报误差),是洪水预报误差的最主要来源.洪水调节能力强的水库,主要是洪量对水库洪水调度起控制作用.因此,本文重点研究产流预报误差对水库防洪调度的影响.通常,产流预报误差的大小由产流预报方案的精度来评定^[3].产流预报方案精度达到85%为甲级方案,表明预报误差较小,预报方案可在防洪调度中应用.不过,只有产流预报方案的精度评定是不够的,还必须分析预报误差的分布特性,以更全面地分析预报误差对水库及其防护对象防洪

安全的影响.

目前洪水预报模型和方法已经比较成熟,对洪水预报方案的精度评定也有相应规范,但对洪水预报误差分布特性的研究却很少,已有的方法主要用正态分布和对数正态分布来描述^[4].对于一次降雨径流过程,尽管产流预报误差是随机的,但至少应有一个误差上限值,例如预报误差不会大于流域平均降雨量.而这一降雨径流的明显特性,已有的方法未能考虑,是欠合理的.为此,本文应用多个流域产流预报方案的模拟预报成果,尝试应用极大熵原理与方法,将预报误差限制在一有限区域内,建立求解洪水总量预报误差的分布模型,并给出具体的求解方法.

1 极大熵原理 (POME)

熵的概念源于热力学,是用来说明热运动过程不可逆的一个物理量,反映了自然界出现的热变化过程是有方向性的,是不可逆的.1929年, Szilard首先提出了熵与信息不确定性的关系,使信息科学引用熵的概念成为可能.1948年, Shannon创立了信息论,把通讯过程中信源讯号的平均信息量称为熵,实现了信息熵的实际应用. Shannon对离散无记忆信源 X 的信息量,定义为 n

收稿日期: 2005-11-25; 修回日期: 2007-03-25.
作者简介: 周惠成* (1958-), 男, 教授, 博士生导师.

维概率矢量 $\mathbf{P} = (p_1 \ p_2 \ \cdots \ p_n)$ 的函数:

$$H(\mathbf{X}) = - \sum_{i=1}^n p_i \log_2 p_i = H(\mathbf{P})$$

熵恒增原理在信息熵领域则叫做“极大熵原理”(principle of maximum entropy, POME).

1957年, Jaynes 在统计力学中首先提出了极大熵原理,随后将之引入到信息论中^[5]. 按照极大熵准则,在不确定性问题的所有的可行解中,应该选择在一定约束下使得熵(或条件熵)达到极大化的一个解,它是 Laplace 著名的“充足理由律”的重要发展. 熵最大意味着对因为数据不足而作的人为假定(人为添加信息)最少,从而所获得的解最合乎自然,最为超然、偏差最小.

Silviu 认为,虽然极大熵准则有其主观性,不过构造一个随机分布时,应当把它看成最客观的主观准则. 实际上,只要承认熵是计量不确定性的最合适的标尺,就完全可以在给定约束下选择不确定性最大的那种分布作为随机变量的分布,从其含有的不确定性的角度看,这种随机分布是最为随机的分布. 依照极大熵准则,就要在给定的那些约束下选择不确定性最大的那种概率分布.

2 洪水总量预报误差分布的极大熵模型

水文水资源学科从本质上看,是一门有关水信息(采集、传输、整理、分析、研究)的学科,其中存在着很多的不确定性问题^[6],因此 POME 是一种解决问题的新途径. Sonuga 首次将 POME 应用到水文频率分析中,推导了基于有限数据的小偏差的正态分布^[7]. 他还针对降雨-径流过程,应用 POME 推导出径流对降雨的条件分布^[8]. 随后,国内外许多学者应用 POME 推导了许多概率密度函数^[9-20]. 总之,基于 POME 推导概率分布,人为偏差最小,所得结果客观,合乎自然.

对于洪水总量的预报误差,其分布特性是与已有的误差信息有关的. 由于无法通过概率理论直接推导出误差变量的先验分布,而且误差变量的数据十分有限,是一个典型的不确定性问题,只能依据现有的数据,推求一种概率分布,使它和已经预报的洪水误差的信息一致,并且没有太大的

误差. 极大熵准则正好可以解决这个问题.

洪水总量预报误差虽然有正有负,但不是没有界限,而是在有限的区域 $(-a, a)$ 内. 设产流预报误差的概率密度函数为 $f(x)$, 建立其最大熵模型:

$$\max H = - \max \int_{-a}^a f(x) \ln f(x) dx \quad (1)$$

$$\text{s. t. } \int_{-a}^a f(x) dx = 1 \quad (2)$$

$$\int_{-a}^a x f(x) dx = \bar{x} \quad (3)$$

$$\int_{-a}^a x^2 f(x) dx = \sigma^2 \quad (4)$$

式中 $\bar{x}$ 为误差的均值, σ^2 为方差.

分布首先应满足一般概率分布的约束式(2),同时,由于误差在有限区域内,其数学期望和方差应为有限值,即满足约束式(3)和式(4).

3 模型的求解

上述模型的求解是一个泛函条件极值问题. 为了求得 $f(x)$ 的表达式,根据变分法引入拉格朗日乘子 λ_1 , λ_2 和 λ_3 , 则

$$J = - \int_{-a}^a f(x) \ln f(x) dx + \lambda_1 \left(\int_{-a}^a f(x) dx - 1 \right) + \lambda_2 \left(\int_{-a}^a x f(x) dx - \bar{x} \right) + \lambda_3 \left(\int_{-a}^a x^2 f(x) dx - \sigma^2 \right) \quad (5)$$

由欧拉方程得到

$$- \ln f(x) - 1 + \lambda_1 + \lambda_2 x + \lambda_3 x^2 = 0 \quad (6)$$

解之得

$$f(x) = \exp(\lambda_1 - 1 + \lambda_2 x + \lambda_3 x^2) \quad (7)$$

式(7)即为极大熵的洪水总量预报误差概率密度函数的解析形式,只要确定其中的参数 λ 就可以完全确定 $f(x)$, 文献[21]中也给出了具体的求解方法. 将式(7)代入约束条件式(2)~(4),并令 $\lambda_0 = \lambda_1 - 1$, 可得

$$\lambda_0 = - \ln \left[\int_{-a}^a \exp(\lambda_2 x + \lambda_3 x^2) dx \right] \quad (8)$$

$$\int_{-a}^a x \exp(\lambda_2 x + \lambda_3 x^2) dx \bigg/ \int_{-a}^a \exp(\lambda_2 x + \lambda_3 x^2) dx = \bar{x} \quad (9)$$

$$\int_{-a}^a x^2 \exp(\lambda_2 x + \lambda_3 x^2) dx \bigg/ \int_{-a}^a \exp(\lambda_2 x + \lambda_3 x^2) dx = \sigma^2 \quad (10)$$

由于式 (9) 和式 (10) 是一个二元非线性方程组, 目前没有解析解法, 而且存在多解现象, 几何意义是几张曲面在空间相交的公共点. 因此本文将其转换为一个非线性优化问题, 计算中令

$$R_1 = \int_{-a}^a x \exp(\lambda_2 x + \lambda_3 x^2) dx \left| \int_{-a}^a \exp(\lambda_2 x + \lambda_3 x^2) dx - e \right| \quad (11)$$

$$R_2 = \left[\int_{-a}^a x^2 \exp(\lambda_2 x + \lambda_3 x^2) dx \left| \int_{-a}^a \exp(\lambda_2 x + \lambda_3 x^2) dx \right|^{\frac{1}{2}} - e \right] \quad (12)$$

为了求得全局最优解, 本文通过遗传算法求上式残差平方和 $R = R_1^2 + R_2^2$ 的最小值来求得方程组的近似解.

4 应用实例

本文选取以洪量对水库洪水调度起控制作用

的 3 个水库为应用实例, 对流域产流预报方案的模拟预报成果进行误差分布特性分析. 现以蓁窝水库 23 场历史洪水产流模拟预报的误差为例, 说明洪水预报误差分布的具体求解过程. 其中产流模型采用大伙房模型^[1].

蓁窝水库位于辽宁省辽阳市安平乡太子河干流中游, 于 1972 年建成, 是一座以防洪和灌溉为主, 兼顾发电和工业用水的大型综合性水利枢纽工程. 坝址以上流域面积 6 175 km², 最大库容 7.91×10⁸ m³. 1994 年汛后, 上游建成了观音阁水库, 它是一座对洪水调节能力很强的水库, 因此洪量对水库洪水调度起控制作用.

根据表 1 计算该系列的均值 $\bar{x} = -0.173$, 方差 $\sigma^2 = 9.493$, 将其代入式 (3) 和 (4), 联立式 (2), 对于给定的 a 值, 通过遗传算法求得概率密度函数, 本文给出几个不同上限值 a 的计算结果, 见表 2

表 1 蓁窝水库 23 场历史洪水产流模拟预报结果

Tab. 1 Runoff simulation results of 23 history flood of Shenwo reservoir

洪号	实际降雨量 /mm	实际净雨量 /mm	预报净雨量 /mm	预报误差 /mm
19600731	359.8	322.1	309.5	-12.6
19610715	54.3	34.8	30.8	-4.1
19620806	120.1	67.9	74.3	6.3
19640806	120.4	106.7	100.3	-6.4
19650801	133.2	56.6	67.9	11.3
19660728	45.9	28.6	31.2	2.6
19670801	38.2	30.1	33.8	3.7
19670810	39.2	22.2	26.1	3.9
19690726	133.8	74.0	73.8	-0.2
19700801	63.6	55.3	50.6	-4.7
19710731	171.3	151.9	134.3	-17.6
19730715	117.8	96.1	76.9	-19.2
19740810	98.2	65.5	70.3	4.8
19750728	204.7	156.5	168.2	11.8
19770802	76.8	58.1	63.1	5.0
19790725	27.0	12.0	13.4	1.4
19810720	111.1	68.9	77.0	8.1
19840813	33.6	22.2	19.6	-2.6
19850717	156.7	97.0	116.4	19.3
19850801	127.4	108.0	94.6	-13.3
19850813	71.4	45.3	45.3	0
19850818	128.2	108.5	112.8	4.3
19860729	146.3	123.1	117.2	-5.9

表 2 ^茅 ^窝 水库不同误差上限对应的结果

Tab. 2 Calculation results between different error up range value in Shenw o reservoir

<i>a</i> /mm	λ_0	λ_2	λ_3	<i>f</i> (<i>x</i>)	<i>R</i> ₁	<i>R</i> ₂	<i>R</i>
30	- 3. 181 4	- 0. 001 9	- 0. 005 4	$e^{-3.1814-0.0019x-0.0054x^2}$	3.3×10^{-7}	1.2×10^{-6}	1.1×10^{-13}
40	- 3. 174 0	- 0. 001 9	- 0. 005 5	$e^{-3.1740-0.0019x-0.0055x^2}$	2.1×10^{-6}	4.3×10^{-8}	4.4×10^{-12}
50~ 300	- 3. 165 0	- 0. 001 9	- 0. 005 6	$e^{-3.1650-0.0019x-0.0056x^2}$	3.6×10^{-7}	1.5×10^{-7}	1.3×10^{-13}

 目前对洪水预报误差的分布特性大都用正态分布描述，因此本文用数理统计中的Kolmogorov-Smirnov 检验法 (简称 K-S 检验) ^[22] 对表 1 中的系列进行了正态分布检验，检验结果表明上述系列服从正态分布。为了将本文的方法和正态分布做比较，寻找一种适合描述洪水总量预报误差分布特性的分布，本文将各分布的计算结果列于表 3。

 表 4 各参数的计算与表 2 相同，将各参数值代入式 (7)，即可求得不同误差上限值 *a* 对应的概

率密度函数。从表 3 和 4 初步可以得出以下结论：
 (1) 用极大熵模型求得的概率分布存在误差的上下限，而正态分布的误差上下限是趋于无穷大的。对于一场洪水，降雨量总是一定的而非无穷大，由此产流预报误差也应有上下限，因此用极大熵模型求得的概率分布来描述产流预报误差的分布特性更符合实际。同样对于非正态分布的系列，在约束条件中增加偏态系数的约束，用上述的极大熵模型也可以求得其概率分布，由于资料和篇幅的限制，本文不再赘述。

表 3 ^茅 ^窝 水库两种方法计算结果比较

Tab. 3 Comparison between different distribution in Shenw o reservoir

概率 <i>P</i> /%	预报偏大 /mm				预报偏小 /mm			
	正态	<i>a</i> = 30 mm	<i>a</i> = 40 mm	<i>a</i> = 50~ 300 mm	正态	<i>a</i> = 30 mm	<i>a</i> = 40 mm	<i>a</i> = 50~ 300 mm
0. 01	35. 1	29. 6	35. 0	34. 9	35. 5	29. 7	35. 3	35. 3
0. 02	33. 4	29. 4	33. 4	33. 3	33. 8	29. 4	33. 7	33. 6
0. 05	31. 1	28. 7	31. 1	30. 9	31. 4	28. 8	31. 5	31. 3
0. 1	29. 2	27. 7	29. 2	29. 0	29. 5	27. 7	29. 5	29. 3
0. 2	27. 2	26. 4	27. 3	27. 1	27. 5	26. 5	27. 6	27. 3
0. 5	24. 3	24. 1	24. 4	24. 2	24. 6	24. 3	24. 7	24. 5
1	21. 9	21. 9	22. 0	21. 8	22. 3	22. 2	22. 3	22. 2
2	19. 3	19. 4	19. 4	19. 2	19. 7	19. 8	19. 8	19. 6
5	15. 5	15. 6	15. 5	15. 4	15. 8	15. 9	15. 8	15. 7
10	12. 0	12. 1	12. 0	11. 9	12. 3	12. 5	12. 4	12. 3
20	7. 8	7. 9	7. 8	7. 8	8. 2	8. 2	8. 2	8. 1
50	- 0. 173	- 0. 173	- 0. 173	- 0. 173	0. 173	0. 173	0. 173	0. 173

表 4 各水库不同误差上限对应的结果

Tab. 4 Calculation results of different forecast error up range value in other reservoir

水库	<i>a</i> /mm	λ_0	λ_2	λ_3	分布函数及参数
白龟山	15	- 3. 100 7	- 0. 014 5	- 0. 004 7	正态分布
	20	- 2. 969 7	- 0. 014 5	- 0. 008 2	$\mu = - 0. 815$
	30~ 300	- 2. 939 1	- 0. 014 6	- 0. 008 9	$\sigma = 7. 526$
丰满	10	- 2. 423 7	- 0. 046 7	- 0. 024 2	正态分布
	15	- 2. 363 0	- 0. 047 0	- 0. 028 9	$\mu = - 0. 808$
	20~ 300	- 2. 359 8	- 0. 047 0	- 0. 029 1	$\sigma = 4. 224$
大伙房	15	- 2. 845 4	0. 165 8	- 0. 020 3	正态分布
	20	- 2. 817 3	0. 172 2	- 0. 022 0	$\mu = 3. 913$
	30~ 300	- 2. 815 8	0. 172 7	- 0. 022 1	$\sigma = 6. 162$

(2) 不同的误差上限得到不同的概率分布. 因此对于不同量级的洪水应选取对应的概率分布, 而正态分布不论是大小洪水采用的都是同一个误差分布. 例如对于蓼窝水库 1979 年的场次洪水, 降雨量不到 30 mm, 可以选取最大误差为 30 mm 或更小的最大误差对应的概率分布. 而对于像 1960 年的场次洪水, 则可选取最大误差为 50 ~ 300 mm 对应的概率分布. 但是正态分布对于像 1979 年的场次洪水, 发生 0.01% 概率的误差也是 35.1 mm, 显然不符合物理意义. 实际中误差上限的选取需要根据降雨量、前期土壤含水量、流域的径流系数等条件来确定.

(3) 在满足给定精度要求的前提下, 蓼窝水库误差上限为 40 和 50 ~ 300 mm 与正态分布的结果基本相同, 而正态分布也是通过极大熵模型推导出的一种分布, 从另一个角度证明了这种方法的严密性. 另外, 通过极大熵模型的求解可以发现, 产流预报误差不是随着降雨量的增大而线性增大的, 因为对于一个流域, 当下垫面达到饱和以后, 所有的降雨基本上都产流, 因此预报误差也会趋于一个稳定的数值, 而这个数值通过正态分布或者非正态分布都是无法求得的. 当然, 随着系列的不断延长, 这个数值可能会发生变化, 因此需要在实时预报中不断加以修正.

(4) 不同的水库, 求得的概率分布不同, 对应的误差上限也不同. 因此洪水总量预报误差的分布特性主要与误差系列的样本有关, 样本越多, 越能反映总体的特性, 求解的分布函数和误差上限越具有代表性.

5 结 论

本文根据实际洪水预报误差出现在有限区域的特点, 建立了洪水总量预报误差分布的极大熵模型, 通过 4 个水库的实例进行了计算, 并与正态分布进行了比较, 结果表明用极大熵法求得的误差概率分布能更好地描述洪水总量预报误差的分布特性, 可以根据实际洪水降雨量的大小确定不同的最大不确定性的误差分布. 计算结果表明不同的误差系列存在不同的误差分布和误差上限, 为分析不同量级洪水预报的风险提供了依据, 尽

管模型在实际应用中如何根据降雨量、前期土壤含水量、流域的径流系数等条件确定误差的上限有一定难度, 但总可以从防洪安全的角度综合多种因素确定. 例如: 如果流域前期降雨较多, 下垫面比较饱和, 则可以将实时降雨量作为净雨计算预报误差的上限, 对防洪来说肯定是偏安全的. 但为了便于实际的工程应用, 仍需进一步研究定量的确定方法.

参考文献:

- [1] 大连理工大学, 国家防汛抗旱总指挥部办公室. 水库防洪预报调度方法及应用 [M]. 北京: 中国水利水电出版社, 1996 5-15
- [2] 邱瑞田, 王本德, 周惠成. 水库汛期限制水位控制理论与观念的更新探讨 [J]. 水科学进展, 2004, 15(1): 68-72
- [3] 水文情报预报规范, SL250-2000 [S]. 北京: 中国水利水电出版社, 2000 9-14
- [4] 左保河. 实时洪水预报过程中的误差分布特性研究 [J]. 华北水利水电学院学报, 1997, 18(2): 41-46
- [5] JAYNES E T. Information theory and statistical mechanics [J]. **Phys Rev**, 1957, 106 620-630
- [6] 王 栋, 朱元生. 最大熵原理在水文水资源科学中的应用 [J]. 水科学进展, 2001, 12(3): 75-80
- [7] SONUGA J O. Principle of maximum entropy in hydrologic frequency analysis [J]. **J Hydrol**, 1972, 17 177-191
- [8] SONUGA J O. Entropy principle applied to the rainfall-runoff process [J]. **J Hydrol**, 1976, 30 81-94
- [9] JOWITT P W. The extreme - value type - I distribution and the principle of maximum entropy [J]. **J Hydrol**, 1979, 42 23-38
- [10] SINGH V P, SINGH K. Derivation of the gamma distribution by using the principle of maximum entropy (POME) [J]. **Water Resour Bull**, 1985, 21(6): 941-952
- [11] SINGH V P, SINGH K. Derivation of the Pearson type (PT) III distribution by using the principle of maximum entropy [J]. **J Hydrol**, 1985, 80 197-214
- [12] 邱宛华. 管理决策与应用熵学 [M]. 北京: 机械工业

- 出版社, 2001: 140-181
- [13] 李元章, 丛树铮. 熵及其在水文频率计算中的应用 [J]. 水文, 1985(1): 22-26
- [14] 马力, 张学文. 最大熵原理与概率分布函数 [J]. 数理统计与应用概率, 1993, 8(4): 78-84
- [15] SINGH V P, SINGH K. Parameter estimation for log-Pearson type III distribution by POME [J]. *J Hydraul Eng*, 1988, 114(11): 112-122
- [16] SINGH V P, GUO H. Parameter estimation for 2-parameter generalized Pareto distribution by POME [J]. *Stoch Hydrol Hydraul*, 1997, 11(3): 211-227
- [17] 黄克中, 张金阳. 水文频率线型选优的最小信息熵准则 [J]. 中山大学学报 (自然科学版), 1996, 35(增刊): 8-14
- [18] 陈丽英. 信息论在确定误差分布律中的应用 [J]. 宇航计测技术, 1999, 19(2): 6-8
- [19] 张明. 最大信息熵理论在地貌瞬时单位线中的应用 [J]. 水文, 2000, 20(3): 13-14
- [20] 乔卫平, 王宇青. 信息熵与误差分布密度函数 [J]. 物理与工程, 2003, 13(1): 26-28
- [21] 李继清, 张玉山, 王丽萍, 等. 应用最大熵原理分析水利工程经济效益的风险 [J]. 水科学进展, 2003, 14(5): 626-630
- [22] 藤素珍. 数理统计 [M]. 大连: 大连理工大学出版社, 1996

Maximum entropy method used in error distribution of flood forecast

ZHOU Hui-cheng*, LI Li-qin, WANG Ben-de

(School of Civil and Hydraul. Eng., Dalian Univ. of Technol., Dalian 116024, China)

Abstract Based on the characteristics of error distribution of flood forecast presented in the limited area, the error distribution of flood volume forecast is established using the principles of maximum entropy, and the maximum entropy model of error distribution of flood volume forecast is put forward. The corresponding solving method is also given in details. The computation results show that the runoff forecast error is always inclined to a constant value with the rainfall volume increasing. And the constant values of several different basins are obtained using this method. After that, the comparison analysis between the obtained model and the normal distribution is done in details. The results show that the probability distribution of error obtained by the method of maximum entropy can describe the characteristics of error distribution of flood volume forecast. The method can be used to determine the different maximum likelihood probability distribution of error according to the different real-time precipitation. At the same time, it can also be used to provide the basic reference for risk analysis on different levels of flood forecast.

Key words maximum entropy method; flood forecast; error distribution