

基于贝叶斯方法的流量数据的不确定性分析*

顾西辉, 张 强

(中山大学 水资源与环境系//华南地区水循环与水安全广东省普通高校重点实验室, 广东 广州 510275)

摘要: 用贝叶斯方法估计水位流量幂律关系式中参数, 并计算东江流域平均低流量、中流量和高流量的不确定性, 并分析影响流量不确定性的因素。结果表明: ① 东江流域流量不确定性总体水平在中等偏上水平, 其中高流量不确定性最大, 中、低流量次之; ② 人类活动比如兴建水库、河道整治、围堰筑坝以及大规模的河道采沙对流量不确定性的大小及趋势有着显著的影响; 大量级的洪水对河道水力特性以及几何形状的影响比较复杂, 变化没有明显统一的趋势。

关键词: 贝叶斯; 水位流量关系; 不确定性; 东江

中图分类号: P333 **文献标志码:** A **文章编号:** 0529-6579 (2014) 02-0131-06

Uncertainty Analysis of Flow Data Based on Bayesian Methods

GU Xihui, ZHANG Qiang

(Department of Water Resources and Environment//Guangdong University Key Laboratory of Water Cycle and Security in South China, Sun Yat-sen University, Guangzhou 510275, China)

Abstract: In this paper, Bayesian methods are used to estimate parameters of the stage-discharge power law model in the Dongjiang River, the average low flow, medium flow and high flow uncertainty are calculated and factors affecting the flow uncertainty are analyzed. The results show that: (1) The overall uncertainty of flow data is at a middle-high level, in which high flow uncertainty is maximum, followed by medium and low flow. (2) Human activities, such as construction of reservoirs, river regulation, cofferdams damming, and large-scale mining river sand, have a significant impact on the size and trends of flow uncertainty; the impact of large river floods on hydraulic characteristics and geometric shape is complex, without obvious change in uniform trend.

Key words: Bayesian method; stage-discharge relationship; uncertainty; Dongjiang river

水文资料整编规范(2000)指出^[1]: 水位流量关系受到冲淤、变动回水、洪水涨落、水生植物、结冰等因素的影响, 会产生不确定性。国内外对于这种不确定性的分析进行了大量研究。程伟等^[2]对三峡水库蓄水后下游近坝段水位流量关系中的研究指出: 河道侵蚀基点条件是影响水位流量关系的主要因素; 戴凌全等^[3]用多次幂多项式函数作为水位流量关系的拟合公式, 用最小二乘法进行拟合; 夏军强等^[4]提出一维水动力模型对黄河下游平滩流量进行分析; 董晓华等^[5]基于最小二乘法

对绳套型水位流量关系曲线进行拟合。此外采用改进的BP神经网络、混合禁忌搜索算法、简化的人工鱼群算法等算法进行水位流量关系曲线的拟合^[6-8]。文献[9]对意大利Po河流域下游河段进行上述各个来源不确定性定量分析, 得出流量总的不确定性在6.2%~42.8%之间; José-Luis Guerrero等^[10]用广义似然估计和蒙特卡洛方法分析水位流量关系的时间变异性。上述研究对于水位流量关系的时空变异和流量数据的不确定性大小及原因有着重要的科学意义。

* 收稿日期: 2013-09-04

基金项目: 国家自然科学基金资助项目(41071020), 新世纪优秀人才支持计划基金资助项目

作者简介: 顾西辉(1990年生), 男, 研究方向: 气象水文学; 通讯作者: 张强; E-mail: zhangq68@mail.sysu.edu.cn

但是以往的研究往往局限于对水位流量关系曲线的拟合方法、拟合线型的研究,并且在流量不确定性方面局限于冲淤、变动回水、洪水涨落、水生植物、结冰等因素对拟合精度的影响,对水位流量关系的变异性引起的不确定性研究较少,而水位流量关系变异是多种因素影响下的结果,同时也是流域水资源管理等诸多实践应用的基础。基于此,本次采用贝叶斯方法来估计水位流量幂律关系式的参数,定量界定水位流量关系变异性引起的平均低流量、平均中流量和平均高流量的不确定性^[11]。东江流域是珠江流域主要支流之一,主要担负珠三角地区大城市,如广州、东莞等,以及香港 80% 的用水,对于保障区域水资源安全与社会稳定具有极其重要的作用和意义。研究东江流域流量不确定性对于东江流域水资源管理具有重要理论与现实意义。

1 研究区域和数据

东江属于珠江三大水系之一,位于中国南方湿润地区,长度 523 km,流域总面积 35 636 km²,平均坡降为 0.35‰。本文以东江流域河源、岭下和博罗 3 个测站为研究对象(图 1)。采用河源站 1981-2009 年水位流量数据,岭下站 1956-2009 年水位流量数据,博罗站 1956-2005 年水位流量

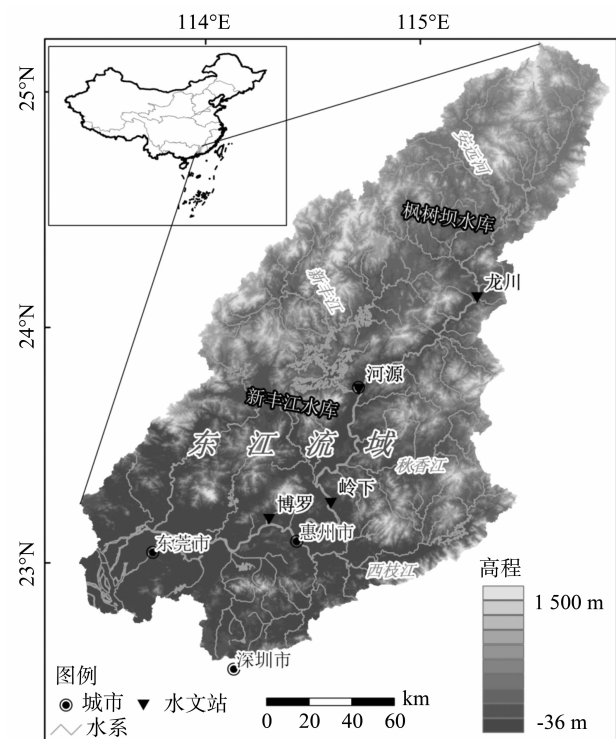


图 1 河源、岭下和博罗 3 个测站的地理分布图

Fig. 1 The geographic distribution of three stations

Heyuan, Lingxia and Boluo

数据进行研究。

2 研究方法

2.1 水位流量曲线拟合的贝叶斯公式

本文采用有着位置参数的幂律关系式,进行单一曲线法推流^[1]:

$$Q = \begin{cases} A_1(h-h_0)^{b_1} & h \leq h_1 \\ A_2(h-h_0)^{b_2} & h_1 < h \leq h_2 \\ \vdots & \\ A_n(h-h_0)^{b_n} & h_{n-1} < h \end{cases} \quad (1)$$

公式中, Q 是流量, h 是水位。 (A_1, A_2, \dots, A_n) 是尺度参数, (b_1, b_2, \dots, b_n) 是形状参数, h_0 是位置参数,代表河道横断面最低点的高程, $(h_1, h_2, \dots, h_{n-1})$ 是不同片段水位流量曲线的分割点,例如:河道的深槽、边滩、低滩、高滩等由于水力性质、径流条件等因素的差异,公式(1)中的各种参数是不同的。由于各个片段在拟合时是相互独立的,所以本文只考虑一条单一的水位流量曲线进行研究。文献[12]以假设河道横断面宽度-深度呈指数关系作为基础,结合曼宁公式推导上述满足幂律关系的水位流量关系式,给出了公式(1)详细的水力学方程式推导过程。单一片段的对数回归模型为:

$$q_i = a + b \lg(h_i - h_0) + \varepsilon_i \quad (2)$$

公式(2)中 $q_i = \lg(Q_i)$, $a = \lg(A)$,为尺度参数 A 的自然对数值。 ε_i 为模型残差,假设 $\varepsilon_i \sim N(0, \sigma^2)$ 。构造向量 $\theta = (a, b, h_0)$,根据贝叶斯概率公式,参数 (θ, σ^2) 的后验概率密度函数为:

$$f(\theta, \sigma^2 | D) = \frac{L(D | \theta, \sigma^2) \pi(\theta, \sigma^2)}{\int_{\theta} L(D | \theta, \sigma^2) \pi(\theta, \sigma^2) d(\theta, \sigma^2)} \quad (3)$$

公式(3)中, D 代表实测的水位-流量值, L 是似然函数, π 是先验分布。

2.2 流量不确定性计算公式

统计每一个时间序列水位值的最小值和最大值,分别为 h_{\min} 和 h_{\max} ,定义 $H = h_{\max} - h_{\min}$ 。 H 的 25% 和 75% 分别表示为 $H_{0.25}$ 和 $H_{0.75}$ 。时间序列水位值在 $(h_{\min}, h_{\min} + H_{0.25})$, $(h_{\min} + H_{0.25}, h_{\min} + H_{0.75})$ 和 $(h_{\min} + H_{0.75}, h_{\max})$ 估计的流量分别为低流量,中流量和高流量,用 $\bar{Q}(h)$ 表示,同时计算相关的 95% 置信区间的上下界,分别用 $\hat{Q}_U^{95}(h)$ 和 $\hat{Q}_L^{95}(h)$ 表示。所以平均的低流量、中流量和高流量不确定性为^[11]:

$$\begin{cases} \int_{h_{\min}}^{h_{\min}+H_{0.25}} \frac{\widehat{Q}_U^{95}(h) - \widehat{Q}_L^{95}(h)}{\widehat{Q}(h)} dh \\ \int_{h_{\min}+H_{0.25}}^{h_{\min}+H_{0.75}} \frac{\widehat{Q}_U^{95}(h) - \widehat{Q}_L^{95}(h)}{\widehat{Q}(h)} dh \\ \int_{h_{\min}+H_{0.75}}^{h_{\max}} \frac{\widehat{Q}_U^{95}(h) - \widehat{Q}_L^{95}(h)}{\widehat{Q}(h)} dh \end{cases} \quad (4)$$

3 分析结果

3.1 贝叶斯方法的拟合优度检验

参数 a 和 b 分别是河道的尺度参数和形状参数，共同为河道水力性质和横断面几何形状所决

定，所以给予一个二维联合正态分布作为先验分布。位置参数 h_0 给予一个无信息先验分布，噪声（公式（2）对数回归模型残差）参数 σ^2 给予一个逆伽马分布。作者在另一篇文章详细介绍了参数集 (a, b, h_0, σ^2) 的先验分布、后验分布以及推断方法。

选取东江流域 2004 年干旱年、2001 年平水年和 2005 年洪水年的河源、岭下和博罗 3 个测站的水位 - 流量数据进行水位流量曲线的拟合以及求出水位流量曲线的 95% 置信区间。并选取 PPCC (probability plot correlation coefficient)、RMSE (root mean square error)^[13] 和随机不确定度 X'_Q ^[1] 作为拟合优度检验指标，见表 1。

表 1 贝叶斯方法拟合的参数值和拟合优度指标

Table 1 The parameter values fitted by Bayesian method and goodness of fitted index

站点	年份	a	b	h_0	PPCC	RMSE/%	X'_Q /%
河源	2001	4.419	1.827	28.79	0.997 5	0.082	0.16
	2004	5.534	1.366	29.43	0.993 8	0.25	0.46
	2005	3.749	2.078	27.69	0.999 2	0.055	0.11
岭下	2001	3.81	2.221	11.17	0.999	0.059	0.12
	2004	5.037	1.655	12.09	0.998 6	0.02	0.04
	2005	4.513	1.988	11.77	1	0.034	0.068
博罗	2001	3.605	2.356	1.677	0.998 9	0.68	1.52
	2004	5.198	1.706	1.311	0.985 2	0.62	1.45
	2005	5.139	1.742	0.797	0.998 4	0.35	0.75

从表 1 中可以看出，贝叶斯方法对 3 个站点干旱年、平水年和洪水年的水位流量关系的拟合效果都很好。PPCC 最小达到 0.985 2，RMSE 最大达到 0.68%，随机不确定度 X'_Q 最大为 1.52%，达到了一类精度的水文站要求^[7]。贝叶斯方法对东江流域水位流量关系参数推断效果很好。岭下站 2005 年贝叶斯方法推断的水位流量关系曲线及 95% 置信区间见图 2。

3.2 水位流量关系的时间变异性

公式（2）中的尺度参数 a 、形状参数 b 和位置参数 h_0 反映出了河道水力特性和横断面几何形状的变化情况。用贝叶斯方法分别分析 1981 - 2009 年河源站，1956 - 2009 年岭下站和 1956 - 2005 年博罗站逐年的水位 - 流量数据，研究尺度参数 a 、形状参数 b 和位置参数 h_0 的变化情况（图 3，表 2）。

从图 3 中可以看出，以 1980 年为分界点，岭下站的尺度参数 a 呈下降趋势，形状参数 b 呈上升趋势，博罗站正好相反。河源站在 1980 年之后，变化情况规律性不明显，总体来说，尺度参数 a 和形状参数 b 均呈下降趋势。对于位置参数 h_0 ，3 个测站均呈下降趋势，这种趋势在 1980 年之后表现的更加显著。并且博罗站在 3 个测站中的下降趋势最显著。这些变化可能受到 1980 年代东江流域开始兴起的大量采砂的影响。Luo 等^[14] 指出：从 1986 - 2003 年珠江流域共采掘沙子超过 $8.7 \times 10^8 \text{ m}^3$ ，其中导致东江流域河道平均下切 1.77 ~ 6.48 m。河道采沙的另一后果是导致河道形状向窄

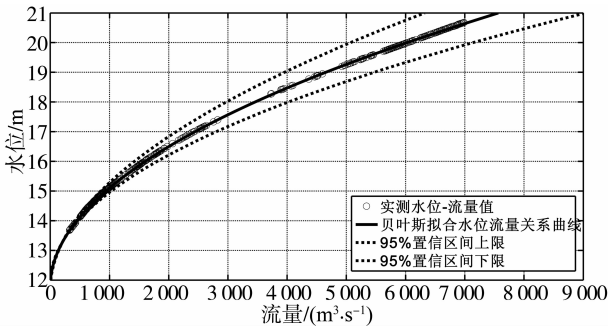


图 2 岭下站 2005 年水位流量关系的贝叶斯拟合效果及置信区间

Fig.2 The fitting effect and confidence interval of stage-discharge relationship fitted by Bayesian method in Lingxia station

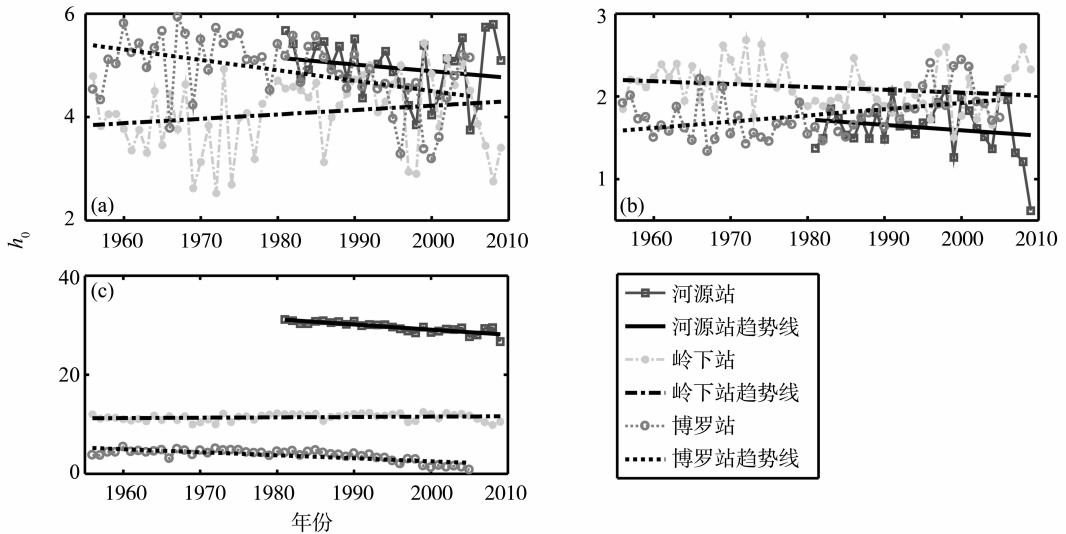


图 3 河源站、岭下站和博罗站贝叶斯估计的系数历年变化及趋势图

Fig. 3 The coefficients changing with time estimated by Bayesian method and their trends of Heyuan, Lingxia and Boluo stations

表 2 三个站点的参数统计值

Table 2 The parametric statistical values of three stations

站点	参数	均值	离势系数	最大差值
河源	a	4.952	0.113	2.042
	b	1.622	0.19	1.467
	h_0	29.611	0.035	4.44
岭下	a	4.072	0.177	2.903
	b	2.106	0.133	1.177
	h_0	11.413	0.059	2.615
博罗	a	4.895	0.134	2.745
	b	1.773	0.148	1.107
	h_0	3.713	0.312	4.638

深方向发展, 即河道宽深比下降^[15]。参数 h_0 持续下降反应了东江流域河床下切的现象, 参数 a 、 b 反映了河道形状—宽深比的变化。

从河源、岭下和博罗站各个参数变化情况来看, 洪水对尺度参数 a 和形状参数 b 的趋势也起到了显著的影响。比如 1959 年的洪水, 岭下站尺度参数 a 由下降趋势变成上升趋势, 同样的情况也出现在 1966 年洪水。1994、2005 年 2 场洪水则把尺度参数 a 由上升趋势变成下降趋势。所以洪水对河道水力特性以及几何形状的影响比较复杂, 变化没有明显统一的趋势。

从表 2 中可以看出, 地理位置相对靠近河口的岭下站和博罗站比相对远离河口的河源站, 其离势系数相对较大, 这表明东江流域河道水力特性和几何形状的变化越靠近河口, 变化越剧烈, 其水位流量关系的不确定性也应该相应的增加。这种性质也可以从参数的最大差值中进一步的反映出来。

3.3 水位流量关系的变异性引起的不确定性

东江流域河源、岭下、博罗 3 个测站每一年的水位数据作为一个时间序列。使用这个时间序列的水位值, 用贝叶斯推断的水位流量关系曲线来估计流量值, 同时计算相关的 95% 置信区间的上界和下界。公式 (4) 用近似的求和公式来代替, 求取平均的低流量、中流量和高流量的不确定性 (图 4)。图 4 中可以看出 3 个测站 1956 - 1974 年, 低流量、中流量和高流量的不确定性总体趋势均在减小。珠江流域 20 世纪 60 - 70 年代大规模的联围筑闸、整治河道使河道形状趋于稳定, 同时河道内流量趋于稳定, 这一时期水位流量关系受到干扰较小, 相对稳定。90 年代末期, 低流量、中流量和高流量不确定性总体趋势趋于稳定 (2009 年除外, 2009 年珠江流域遇到大旱, 流量的急剧减小是不确定性升高的主要原因); 河源站、岭下站和博罗站, 低流量、中流量和高流量的不确定性趋于吻合, 个别年份有较大差异。主要因为 90 年代末期东江流域主要水利工程已经修建完毕、河道大量采沙得到有效管理和治理, 河道形状趋于稳定, 并且东江流域 3 大水库 (枫树坝、新丰江、白盆珠) 对东江流域径流量的有效调节, 综合影响下, 东江流域低流量、中流量和高流量流量大小差距显著变小, 水位流量关系趋于稳定。

70 年代中期, 岭下站低流量不确定性有一个明显增加的跳跃性变化。这一时期枫树坝和新丰江两座水库的建成和运行使得岭下站河道枯水期流量增加, 低流量水位流量关系产生明显异变。70 年代中期到 90 年代初, 岭下站低、中、高流量普遍

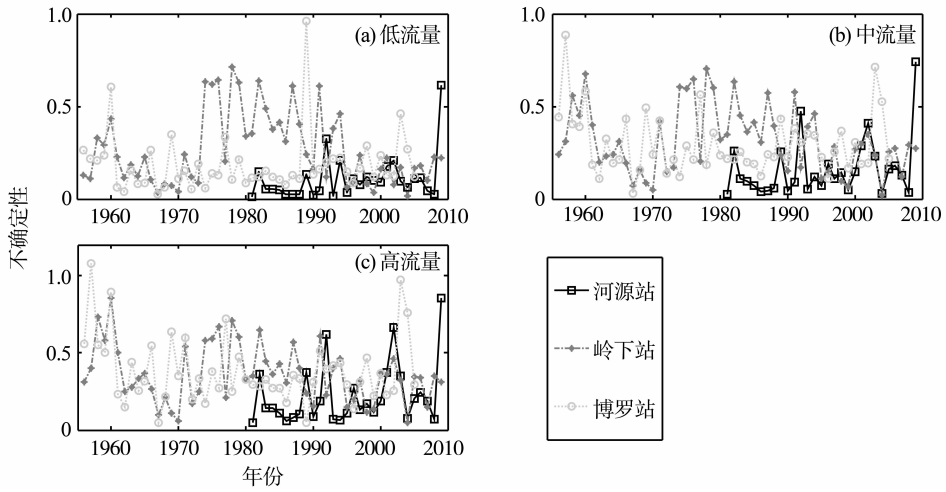


图 4 河源站、岭下站和博罗站低流量、中流量和高流量不确定性

Fig. 4 The uncertainty in low flow、medium flow and high flow of Heyuan、Lingxia and Boluo stations

高于其他时期 (图 5(b))。这一时期河道采沙大量增加, 对河道形状变化影响剧烈, 水位流量关系非常不稳定。

单个测站来说, 河源站、岭下站和博罗站, 表现的共同规律是: 高流量不确定性最大, 中流量不确定性次之, 低流量不确定性最小 (个别年份除外)。以博罗站为例 (图 5c), 绝大部分年份, 高流量要比中流量、低流量的不确定性要大, 最高达到 108%。由于博罗站位于东江流域的干流, 上有

秋香江等支流的汇流, 因此水量充沛, 洪峰流量曾经达到 $12\ 800\ m^3/s$ (1959 年)。高流量条件下, 洪水对河道的冲刷、河道断面几何形状以及河道水力特性的改变更加剧烈, 相比低、中流量, 高流量还涉及到河漫滩的形状、水力性质以及粗糙度的变化。基于上述原因, 高流量不确定性相对较大。但是, 图 5(c) 显示, 70 年代初期之后, 高流量和低、中流量不确定性的差异显著在减小。1974 年新丰江水库和枫树坝水库的建成和使用, 对博罗站的消峰作用很明显, 有效的调节了博罗站高、低流量之间的差距。两大水库对稳定博罗站流量的不确定性起到了很大作用。

河源站、岭下站和博罗站每一年的水位 - 流量数据作为一个时间序列, 共有 133 个时间序列。文献 [11] 提出了一个质量等级系统: 不确定性在 0 ~ 9%, 10% ~ 19%, 20% ~ 39%, 40% ~ 79% 和 >80% 被分别评为优、良、中、差和极差。统计这 133 个时间序列不确定性在各个等级中的比例 (图 6)。可以看出, 东江流域低流量不确定性集中在 0 ~ 20%, 占到 63%, 表现最好。这一点与文献 [11] 在挪威测站的研究是不同的: 他认为挪威测站低流量不确定性表现最差。而东江流域高流量不确定性相对较大, 集中在 20% ~ 40%, 属于中等水平。中流量不确定性居于低、高流量之间, 不确定性低于 40% 的比例达到 78%。总体来说, 东江流域流量不确定性在中等偏上水平。

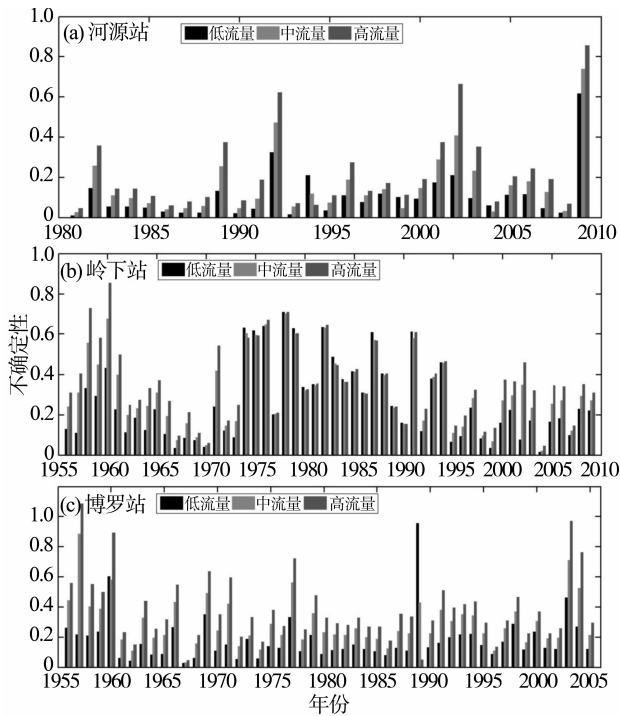


图 5 河源、岭下和博罗 3 站历年流量不确定性

Fig. 5 The flow uncertainty over the years of Heyuan, Lingxia and Boluo stations

从均值和离势系数来看 (表 3), 三个测站共同表现的规律是: 流量越大, 其不确定性的均值越大, 离势系数越小; 同一类别的流量, 越靠近河口, 不确定性的均值相对较大, 离势系数相对较小。

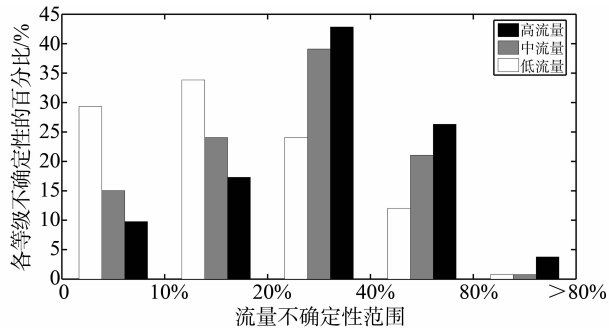


图 6 133 个时间序列流量不确定性各等级统计比例

Fig. 6 The statistical proportions of each grade of 133 time series flow uncertainty

表 3 河源站、岭下站和博罗站不确定性统计值

Table 3 The statistical values of uncertainty in Heyuan, Lingxia and Boluo stations

站点	不确定性	均值	离势系数	最大差值
河源	低流量	0.108	1.103	0.605
	中流量	0.16	0.968	0.712
	高流量	0.223	0.876	0.811
岭下	低流量	0.266	0.727	0.696
	中流量	0.317	0.563	0.679
	高流量	0.363	0.508	0.809
博罗	低流量	0.184	0.829	0.958
	中流量	0.296	0.542	0.857
	高流量	0.385	0.559	1.036

这说明: 同一个测站, 低流量的不确定性相对较小, 变化幅度相对较大; 同一类别流量, 越靠近河口, 流量的不确定性相对较大, 变化幅度相对较小。

4 结论

本文通过用贝叶斯方法推断水位流量幂律关系式的参数, 继而求出水位流量关系的变异性引起的不确定性, 得出以下结论:

1) 用贝叶斯方法推断东江流域水位流量幂律关系式的参数是适合的, 通过 PPCC、RMSE 和 X'_0 等拟合优度指标的检验, 水位流量幂律关系式的拟合精度达到一类水文测站的要求。

2) 东江流域水位流量关系时间变异性显著, 河源站和博罗站河道尺度参数呈下降趋势, 形状参数呈上升趋势, 岭下站则相反。3 个测站的位置参数均呈下降趋势。3 个测站参数的变化趋势以 1980 年为分界点, 前后变化显著。

3) 东江流域 20 世纪 60-70 年代大规模的联围筑闸、整治河道对减小 3 个测站的流量不确定性有显著的作用; 20 世纪 70 年代中期到 90 年代初

期, 大规模的河道采沙对岭下站流量不确定性影响最显著。90 年代之后, 河源站和岭下站流量不确定性在稳定减少, 博罗站流量不确定性呈增加趋势; 3 个测站低、中、高流量不确定性之间的差距减小。

4) 东江流域高流量不确定性相对较大, 集中在 20%~40%, 属于中等水平。中流量不确定性居于低、高流量之间, 不确定性低于 40% 的比例达到 78%。总体来说, 东江流域流量不确定性在中等偏上水平。

参考文献:

- [1] 水利部长江水利委员会水文局. 水文资料整编规范 [S], 2000.
- [2] 程伟, 陈立, 许文盛, 等. 三峡水库蓄水后下游近坝段水位流量关系 [J]. 武汉大学学报: 工学版, 2011, 44 (4): 434-444.
- [3] 戴凌全, 戴会超, 蒋定国, 等. 基于最小二乘法的河流水位流量关系曲线推算 [J]. 人民黄河, 2010, 32 (9): 37-39.
- [4] 夏军强, 吴保生, 李文文. 黄河下游平滩流量不同确定方法的比较 [J]. 泥沙研究, 2009 (3): 20-29.
- [5] 董晓华, 薄会娟, 邓霞. 基于最小二乘法的绳套型水位流量关系最优定线研究 [J]. 中国农村水利水电, 2010 (1): 51-55.
- [6] 李中志. 基于改进 BP 神经网络的水位流量关系拟合 [J]. 中国农村水利水电, 2008 (10): 30-35.
- [7] 汪嘉杨, 李祚泳, 倪长健, 等. 基于混合禁忌搜索算法的水位流量关系拟合 [J]. 系统工程, 2006, 24 (6): 107-110.
- [8] 施秋红, 王联合国. 基于简化的人工鱼群算法的水位流量关系拟合 [J]. 甘肃农业大学学报, 2010, 45 (2): 147-151.
- [9] BALDASSARRE G D, MONTANARI A. Uncertainty in river discharge observations: a quantitative analysis [J]. Hydrology and Earth System Sciences Discussions, 2009 (6): 39-61.
- [10] GUERRERO J L, WESTERBERG I K, HALLDIN S, et al. Temporal variability in stage-discharge relationships [J]. Journal of Hydrology, 2012, 446-447: 90-102.
- [11] PETERSEN-ØVERLEIR A, SOOT A, REITAN T. Bayesian rating curve inference as a streamflow data quality assessment tool [J]. Water Resour Manage, 2009, 23 (9): 1835-1842.
- [12] REITAN T, PETERSEN-ØVERLEIR A. Dynamic rating curve assessment in unstable rivers using Ornstein-Uhlenbeck processes [J]. Water Resources Research, 2011, 47 (2): 1-14.
- [13] 曹深西, 陈子燊. 广东沿海的极值风速概率分布研究 [J]. 海洋通报, 2013, 32 (1): 12-18.
- [14] LUO X L, ZENG E Y, JI R Y, et al. Effects of in-channel sand excavation on the hydrology of the Pearl River Delta, China [J]. Journal of Hydrology, 2007, 343 (3/4): 230-239.
- [15] 季荣耀, 陆永军, 左利钦. 东江下游博罗河段人类活动影响下的河床演变 [J]. 泥沙研究, 2010 (5): 48-54.