

基于门限回归的稳定水位流量关系曲线定线方法

程银才, 范世香

(山东农业大学水利土木工程学院, 山东 泰安 271018)

摘要: 本文采用基于时间序列与动态数据处理的门限回归方法改进现有的单一线的定义方法, 通过实例研究表明, 新方法较现有的方法精度更高、稳定性更好。

关键词: 门限回归; 稳定水位流量关系; 定线; 方法

中图分类号: TV122 文献标识码: A 文章编号: 1000-0852(2013)04-0022-03

1 现有的稳定水位流量关系曲线定线方法的缺陷

测站水位流量关系包括稳定、不稳定水位流量关系两类。针对稳定的水位流量关系, 现有的定线方法为单一曲线法, 它分为人工图解法、计算机定线法^[1]。人工图解法方法简单, 但随意性大、计算量大, 且计算精度难以控制。计算机定线法, 因采用统一标准, 随意性较小, 计算效率高, 但在整个水位、流量实测范围内, 采用具有一个函数关系一条单一线, 这对于多数具有测站控制的断面或河道, 可以接受。但出现以下情况时, 这种处理方法, 就不符合测站水位流量关系固有的特点, 进而影响到定线的精度, 以及它在工程实践中的应用。

a) 高水、中低水时水位流量关系均单一, 但高水与中低水的水位流量关系规律不一致, 如图 1。

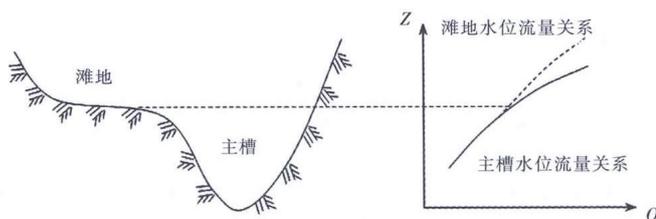


图 1 具有边滩的水位流量关系
Fig.1 Stage discharge relation at flood plain

涨水漫滩时, 断面形状出现突变, 使得水位面积、水位流速关系曲线在漫滩水位处出现转折, 最终导致水位流量关系曲线在高、中低水位下出现差异。

b) 高水控制较好的测站, 受局部冲淤的影响, 水位流量关系的高、低水规律不同, 如图 2。

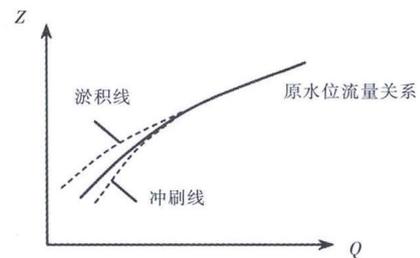


图 2 受局部冲淤影响的水位流量关系
Fig.2 Stage discharge relation effected by local scouring and silting-up

因断面冲淤多发生在河床底部, 局部冲刷、淤积增减的面积占高水面积的比重较小, 又测站高水控制良好, 因此, 高、中低水下水位流量仅在中低水位以下呈现差异。

c) 具有支流的河段, 受到支流来水性质、比例不同的影响, 测站水位流量关系规律不同, 如图 3。



图 3 水位流量关系受支流来水影响
Fig.3 Stage discharge relation effected by tributary inflow

测站断面处的流量由该断面以上不同流域的降水经产汇流汇集而成,而不同的流域降雨径流规律不一致;且断面以上流域降雨不同步及降水量大小差异,导致不同支流在该断面处的来水比例不同。以上两个原因综合,使得测站断面处的水位流量关系因支流影响规律不同。

事实上,测站的水位流量关系受到不只上述(当然还包括没有提及到的)一种因素的影响,多种影响因素往往同时存在,只不过有主次之分。此时,测站水位流量关系可表示为

$$Q=f(z, \text{因子 } 1, \text{因子 } 2, \dots) \quad (1)$$

根据测站实际情况,如何从式(1)中辨识出最主要的影响因子,并建立多因素的分段水位流量关系,这可以借助多元门限回归模型来解决。

2 多元门限回归的数学模型

多元门限回归是时间序列与动态数据处理的一种方法^[2],它用于研究一个因变量与多个影响因子之间的相关关系,已应用到工程实际的多个方面^[3-5]。

门限回归建模的基本思想是,对给定预报对象和预报因子资料,根据门限变量(即影响因子)和门限值,决定在不同的情况下使用不同的相关关系方程,以此解释各种类似于突变的现象。门限回归的实质是将预报问题按状态空间的取值进行分类,用分段线性回归模型来描述总体的非线性规律^[3]。

设因变量为 y , 它与多个影响因子 x_1, x_2, \dots, x_m 存在相关关系,以两段门限回归为例,模型为:

$$y = \begin{cases} a_0 + \sum_{i=1}^{m_1} a_i^1 x_i & , x_j < b \\ a_0 + \sum_{i=1}^{m_2} a_i^2 x_i & , x_j \geq b \end{cases} \quad (2)$$

式中: $a_0^1, a_1^1, \dots, a_{m_1}^1$ 以及 $a_0^2, a_1^2, \dots, a_{m_2}^2$ 为回归系数; x_j 为门限变量; b 为门限值。

建立门限回归模型的关键是选取合适的门限变量及其门限值,其原理如下。

选取的门限变量不同,相关关系也不同,导致因变量有显著差异。当某个影响因子大于等于某一门限值时,它所分成的两个子样本组间的方差有显著差异,据此可以选取使因变量分组产生最大差异的影响因子作为门限变量,并同时确定门限值。其具体步骤为:

Step1 获取因变量 y 和 m 个影响因子 x_1, x_2, \dots, x_m 共 n 个实测样本。依次选取 $x_i (i=1, 2, \dots, m)$, 将 x_i 的 n 个

实测值从大到小排序,相应地,将 y 的 n 个实测值按 x_i 的大小顺序重新排列,得到新序列,记为 $y^i = \{y_1^i, y_2^i, \dots, y_n^i\}$ 。

Step2 将 y^i 进行 $L=n-1$ 次二分割,分割成的两段分别记为 $y^{i1} = \{y_1^i, y_2^i, \dots, y_k^i\}$, $y^{i2} = \{y_{k+1}^i, y_{k+2}^i, \dots, y_n^i\}$ ($k=1, \dots, L$),再计算两段的综合标准差 s :

$$s = \sqrt{\frac{1}{n-2} \left(\sum_{j=1}^k (y_j^i - \bar{y}_1)^2 + \sum_{j=k+1}^n (y_j^i - \bar{y}_2)^2 \right)} \quad (3)$$

其中, $\bar{y}_1 = \frac{1}{k} \sum_{j=1}^k y_j^i$, $\bar{y}_2 = \frac{1}{n-k} \sum_{j=k+1}^n y_j^i$ 。

Step3 构造统计量 $t = \frac{|\bar{y}_1 - \bar{y}_2|}{s \sqrt{1/k + 1/(n-k)}}$, 采用 t 检

验,以判断组间是否存在显著性差异。对某个 y^i 的 L 次二分割,计算各次分割的 $t_j (j=1, 2, \dots, L)$ 值,取 $t_i^* = \max(t_1, t_2, \dots, t_L)$ 。

Step4 采用 Step3 的方法,计算其余 $m-1$ 个 y^i 的 t_i^* 值,取 $t^* = \max(t_1^*, t_2^*, \dots, t_m^*)$ 。根据 t^* 对应的 t_i^* , 确定最优分割影响因子 x_i 及最优分割点 j , 则 x_i 为门限变量,与 y_j^i 对应的 y_j^i 为门限值。

Step5 根据已确定的门限变量与门限值,采用正交多项式法,建立两段门限回归模型,拟合水位流量关系。

3 应用实例

本文选用黄山市呈村水文站 1998 年实测水位流量资料。先不考虑测站控制条件的变化,在整个实测水位、流量范围内,用现有的稳定水位流量关系曲线单一曲线定线方法,采用正交多项式拟合水位流量关系,用计算机定线,经编程计算,结果见表 1 中的计算流量 Q^1 及误差¹ 栏。

对现有方法计算结果分析,在整个实测水位、流量范围内,采用一个函数关系定线,计算精度不高,且出现了系统偏离的现象,总体定线效果不好。结合测站实际情况分析,导致该结果的原因是测站的控制条件发生了变化,导致高、中低水水位流量关系不一致。采用本文提出的方法,使用两段门限回归模型,采用正交多项式拟合水位流量关系,经编程计算,最优分割点 $j=22$, 水位门限值为 $z=5.0\text{m}$, 统计量 t 最优值 $t^*=10.32$ 。计算结果见表 1 中的计算流量 Q^2 及误差² 栏。

对比两种方法的计算结果可知,新方法较现有方法计算精度显著提高,系统偏离情况明显改善,总的定线效果更好。

表1 两种定线方法的计算成果
Table1 The calculated results of the two methods

水位 Z/m	实测流量 $Q/\text{m}^3\cdot\text{s}^{-1}$	计算流量 $Q^1/\text{m}^3\cdot\text{s}^{-1}$	计算流量 $Q^2/\text{m}^3\cdot\text{s}^{-1}$	误差 $^1/\%$	误差 $^2/\%$	水位 Z/m	实测流量 $Q/\text{m}^3\cdot\text{s}^{-1}$	计算流量 $Q^1/\text{m}^3\cdot\text{s}^{-1}$	计算流量 $Q^2/\text{m}^3\cdot\text{s}^{-1}$	误差 $^1/\%$	误差 $^2/\%$
2.09	0.20	0.21	0.21	4.29	4.29	4.07	139	141.2	142.9	1.60	2.79
2.12	0.31	0.28	0.29	-8.49	-1.01	4.16	143	153.6	153.5	7.38	7.34
2.21	0.74	0.71	0.70	-4.52	-5.41	4.32	163	176.0	172.0	8.00	5.49
2.53	4.04	5.73	5.31	41.82	31.44	4.53	190	206.4	194.9	8.61	2.60
2.64	7.40	9.15	8.98	23.64	21.31	4.73	208	235.7	215.1	13.31	3.43
2.68	10.6	10.6	10.6	0.39	-0.34	5.00	242	275.4	239.2	13.79	-1.15
2.83	16.7	17.5	18.0	4.59	7.54	5.20	276	304.5	274.1	10.31	-0.69
2.91	22.2	21.9	22.8	-1.27	2.89	5.50	307	346.9	311.7	13.01	1.52
2.96	27.3	25.0	26.2	-8.48	-3.96	5.68	338	371.5	337.4	9.92	-0.17
3.04	35.7	30.4	32.1	-14.98	-10.02	5.98	384	410.6	384.8	6.94	0.20
3.22	54.3	44.4	47.5	-18.19	-12.52	6.21	419	438.8	424.1	4.73	1.22
3.30	59.5	51.5	55.2	-13.40	-7.29	6.36	471	456.3	450.8	-3.13	-4.29
3.34	64.0	55.3	59.2	-13.65	-7.56	6.67	506	489.9	507.8	-3.19	0.35
3.42	74.7	63.1	67.5	-15.53	-9.69	6.82	516	504.9	535.8	-2.15	3.83
3.53	80.5	74.6	79.4	-7.31	-1.33	6.87	553	509.8	545.1	-7.82	-1.42
3.72	105	96.3	101.2	-8.26	-3.59	7.74	703	579.6	700.6	-17.56	-0.35

4 结语

采用门限回归可以辨识决定水位流量关系规律的主要影响因子,以及高、中水规律变化的转折点,这意味着,根据门限回归,可以由实测的水位、流量这一结果反思引起测站水位流量关系规律发生变化的原因以及变化的转折点。解决了简单多段回归方法拟合水位流量关系时,只能识别主要的影响因子,不能很好地确定变化转折点的困扰(这时只能依靠测站具体情况,大致确定转折点)。

将门限回归与现有稳定水位流量关系单一曲线的计算机定线方法结合起来,不仅能大大提高水位流量关系的定线精度,而且能增强定线的稳定性,不因个别突出点或少量实测数据的增减使得定线的结果发生大的偏离。

限于资料所限,本文引用的水位、流量实测资料的影响因子只有水位变量。若能加入其它影响因子,如落差、水面比降、相邻测站的水位,多元门限回归的稳定

水位流量关系定线的方法更能显示其优势。

参考文献:

- [1] 谢悦波. 水信息技术 [M]. 北京: 中国水利水电出版社, 2009.(XIE Yuebo. Hydrology Information Technology[M]. Beijing:China Water-Power Press, 2009.(in Chinese))
- [2] 杨位钦,顾岚. 时间序列分析与动态数据建模[M]. 北京:北京工业学院出版社,1986.(YANG Weiqin, GU Lan. Time Series Analysis and Dynamic Data Modeling [M].Beijing:Beijing Industry University Press,1986. (in Chinese))
- [3] 严华生, 曹杰. 多元门限回归模型的一种建模方法 [J]. 大气科学, 1994,18(2):194-199. (YAN Huasheng,CAO Jie. Method of building a multivariate threshold regression model [J].Scientia Atmospherica Sinica, 1994,18(2):194-199.(in Chinese))
- [4] 朱亦鹏,何海林,张国祥. 门限回归预报模型在秋汛预报调度中的可行性分析 [J]. 水文,1996, 4:25-28.(ZHU Yipeng, HE Hailin, ZHANG Guoxiang. Feasibility analysis of threshold regressive model in autumn flood forecast operation [J].Hydrology, 1996, 4: 25-28. (in Chinese))
- [5] 金菊良,丁晶. 水资源系统工程[M].成都:四川科学技术出版社,2002. (JIN Juliang, DING Jing. Water Resources Systems Engineering [M]. Chengdu:Sichuan Science and Technology Publishing House, 2002. (in Chinese))

Determination Method of Stage Discharge Relationship Curve Based on Threshold Regression

CHENG Yincai , FAN Shixiang

(Water Conservancy and Civil Engineering College, Shandong Agricultural University , Taian 271018, China)

Abstract: Threshold regression is a method in time series and dynamic data processing, based on thresholds variables and threshold values. Threshold regression was used to improve existing single curve method. The case studies show that the new method has higher accuracy and better stability than the existing method.

Key words: threshold regression;stage discharge relation;determination of relation curve;method