

基于多元线性回归模型的测流垂线精简分析

唐伟霄, 程虎成, 申迪, 方圆

(江苏省水文水资源勘测局淮安分局, 江苏 淮安 223001)

摘要: 为确保在出现特殊水情时,最大限度地缩短测流历时或需要大量增加测流次数,可将断面测验垂线进行精简,从而提高测验效率,而一般对垂线精简方法为收集足够测次的、有代表性的、精度高的实测资料,然后制作垂线平均流速、水深横向分布图及全断面流速分布图等图表,最终拟定精简方案。利用多元线性回归分析方法对朱码闸2016~2017年实测资料进行精简分析,能够快速有效得到精简方案,且符合相关关系线检验及精度评定等规范要求,为开展水文站测流精简分析做参考,成为自动测流系统的理论依据。

关键词: 多元线性回归;精简分析;流量测验;自动测流系统

中图分类号:TV123

文献标识码:B

文章编号:1007-7839(2019)11-0041-05

Streamlined analysis of flow measurement vertical line based on multiple linear regression model

TANG Weixiao, CHENG Hucheng, SHEN Di, FANG Yuan

(Huai'an Hydrology and Water Resources Investigation Bureau of Jiangsu Province, Huai'an 223001, Jiangsu)

Abstract: In order to minimizing the flow measurement duration and the number of flow measurement when a special water condition occurs, the cross-section test vertical line can be simplified to improve the test efficiency. Generally, the vertical line simplified method is collecting enough times, representative and high-precision measured data, and then making a graph of the average vertical flow velocity, the water depth lateral distribution map and the full-section flow velocity distribution map, and finally formulating a streamlined scheme. In this paper, the multivariate linear regression analysis method is used to streamline the two-year actual measurement data of Zhuma gate for 2016-2017, which can quickly and effectively obtain the streamlined plan, and the program meets the requirements of relevant relationship line test and accuracy evaluation. For reference analysis of hydrological station flow measurement, it can also provide a theoretical basis for automatic flow measurement systems and contribute to streamlined flow analysis of hydrological stations.

Key words: multiple linear regression; streamlined analysis; flow test; automatic flow measurement system

0 引言

在发生大洪水期间,为了缩短测流时间,并且得到准确的流量资料,我们常将水文站测流断面上的垂线进行精简,精简分析也是流量测验的一个重要工作,目的是在保证流量测验成果具有规定精度

的前提下,减少测验工作量,缩短测验历时,以适应不同需求时测流工作的需要。一般精简方案计算量较大、耗时久,本文以多元线性回归模型为基础,以朱码闸水文站2016~2017年资料做精简分析,并最终完成了对断面测速垂线的精简,且符合相关规范要求。

收稿日期:2019-03-17

作者简介:唐伟霄(1989—),男,本科,工程师,主要从事水文水资源、水雨情遥测工作。

1 精简方案分析比较

1.1 一般精简方法介绍

对断面测速垂线精简的方法很多,一般方法是收集足够测次的、有代表性的、精度高的实测资料,以此资料作为分析的基础,然后制作垂线平均流速、水深横向分布图及全断面流速分布图等图表,以拟定精简方案。在具体计算各方案精简前后的相对误差,统计各种误差是否超过规范要求的精度质量指标,以确定精简方案的可行性,确定是否修改方案,即试错法试验并选取最佳的测速垂线,最后绘制断面平均流速与相关垂线流速点分析结果的关系图,通过点群中心绘出关系线,并对关系线进行检验。经检验合格后,再比选出精确度最佳的方案作为最终流量测验精简结果,最终提供断面平均流速的计算公式^[1]。此方法虽能够最终得到断面流速与单流速尤其是多根垂线流速的关系,但是需要绘图、制表等工作,不断试错来确定最优方案,其工作量大,耗时久。

1.2 多元线性回归分析

1.2.1 概述

多元线性回归分析是指在相关变量中将一个变量视为因变量,其他多个变量视为自变量,建立多个变量之间线性数学模型的数量关系式并利用样本数据进行分析的统计分析方法。同样道理,断面平均流速变化的趋势可以由断面各垂线流速来确定,本文以朱码闸 2016~2017 年常测法资料做分析,并利用 Excel 数据分析工具进行运算,将断面平均流速作为因变量,断面上所有垂线的流速做自变量,从而进行多元线性回归分析,能够快速计算出各垂线的系数,并得到断面平均流速与各垂线流速的理论回归线,其原理是使得观测的实测数据对经验回归直线的离差平方和达到最小,即最小二乘原理^[2]。在回归分析过程中,可合理利用 Excel 统计分析出的 R^2 值、F 检验、t 检验等统计检验值结果来判断取舍相关垂线^[3],从而快速得到垂线精简方案。

1.2.2 数据回归分析

根据朱码闸测站任务书要求测得的 2016~2017 年实测资料显示,所有测次用到的测速垂线为 10 根,起点距(m)分别为:16.7、17.5、20.0、25.0、30.0、35.0、40.0、45.0、50.0、52.0(垂线流速分别记作 V1、V2、V3...V10,下同),共 36 测次包含高中低 3 个水位级的测次数据。我们以 10 根垂线流速做

自变量,断面平均流速作为因变量,利用 Excel 数据分析工具中的回归分析功能,得到下列回归统计、方差分析以及相关系数表格,见表 1。

以上分析结果显示,在 10 根垂线流速全部参与断面平均流速回归计算的情况下,效果非常好。其中 Multiple R(复相关系数 R)和 R Square(复相关系数 R 平方)分别达到了 0.999773948、0.999547947,且 Adjusted R Square(调整后的 R 平方)为 0.999367126。这充分说明了参与计算的各垂线流速和断面平均流速相关性非常高,所算的回归方程非常显著。表 1 第 2 行中 F 检验中 Significance F 的值几乎为 0,远远小于显著性水平 0.05,可以认定该回归方程回归效果显著,方程中至少有 1 个回归系数显著不为 0。其中第 3 行 P-value 为回归系数 t 统计量的 P 值,其值越小说明此根垂线流速与断面平均流速相关性越大,所以为了减少垂线数目,将此值较大的垂线剔除,不参与下一步回归分析;同时,分析结果显示 V2、V10 垂线系数显示为负数,即垂线流速与断面流速成负相关,所以也将此垂线剔除。为了达到精简的目的,分析过程中需要剔除明显不符的垂线继续做回归分析(只剩下第 4 根~第 9 根垂线)。

通过以上初步精简后,可以迅速的去除明显不符要求的垂线,然后根据 t 检验的分析结果选取最小 P-value(表 1 中加粗值)的 3 根垂线继续做精简。从表 1 观测到最小 P-value 值分别为第 6、第 7、第 8 根垂线,但是此 3 根垂线在断面的一边,为了使得精简垂线分析分布合理,最终确定方案一(3 组最小值)选取第 6、第 7、第 8 根,方案二(垂线分布均匀)选取第 4、第 6、第 7 根,并用同样的方法来做回归分析。如果任意一个方案满足,将继续减少垂线数并按照 P-value 最小和在断面上分布均匀的原则做回归分析,直到 1 根垂线或者不满足条件结束。

2 成果分析

2.1 方案回归成果分析

通过以上确立的 2 种方案,将根据分析成果进行逐一取舍。通过对各方案做回归分析,得到的结果显示:虽然调整后的 R^2 值都达到了 98% 以上,但是其 P-value 值有明显的差异,见表 2。

根据表 2 中方案一、方案二分析结果显示,方案一的 V7、V8 这 2 根垂线 P-value 比方案二明显较大,舍弃此方案,取用方案二,并算得方案二断面

表 1 多元线性回归分析成果表

回归统计结果	Multiple R	0.999773948	观测值	36
	R Square	0.999547947		
	Adjusted R Square	0.999367126		
	标准误差	0.008004994		

		df	SS	MS	F	Significance F
方差分析结果	回归分析	10	3.54222856	0.354223	5527.82832	2.73E-39
	残差	25	0.001602	6.41E-05		
	总计	35	3.54383056			

各垂线有关参数

	截距	V1	V2	V3	V4	V5
系数	0.005804	0.019584	-0.05408	0.072826	0.144685	0.134815
t Stat	1.618323	0.600846	-1.92834	2.720471	5.594842	4.913059
P-value	0.11814184	0.55335175	0.06524353	0.01169014	8.07E-06	4.67E-05
	V6	V7	V8	V9	V10	
系数	0.229924	0.173899	0.152476	0.083904	-0.09895	
t Stat	7.861021	5.789104	5.816369	4.857552	-2.449	
P-value	3.23E-08	4.92E-06	4.59E-06	5.39E-05	0.02167903	

表 2 几种方案回归 P-value 分析结果

垂线号	方案一			方案二		方案三		方案四		
	V6	V7	V8	V4	V6	V7	V4	V7	V6	V7
P-value	2.60E-11	0.048914	0.029077	5.09E-08	3.65E-09	8.34E-11	1.41E-09	1.6E-16	9.40E-11	9.28E-07

平均流速公式为:

$$V_{\text{二}} = 0.00003335 + 0.230053 \times V4 + 0.347368 \times V6 + V7 \times 0.318312$$

为了继续达到精简目的, 将方案二中的 3 根垂线继续进行精简, 将垂线缩减为 2 根, 确定方案三: 第 4、第 7 根(垂线在断面上分布均匀); 或者方案四: 第 6、第 7 根(P-value 最小的 2 根垂线)。根据表 2 中方案三、方案四 t 检验的分析结果显示, 统计学规律无明显差异, 算得方案三、方案四回归公式为:

$$V_{\text{三}} = 0.021703 + 0.378923 \times V4 + 0.03487 \times V7$$

$$V_{\text{四}} = 0.000861 + 0.521821 \times V6 + 0.318442 \times V7$$

因为此河道断面较宽, 将不再进一步缩减垂线数目, 即减至 2 根后即做水文定线相关检验分析。

2.2 水文成果定线检验

从上文得到了基本符合统计学规律的 3 个方案: 方案二、方案三和方案四。

根据《水文测验手册》^[4] 中规定, 首先对 2 根垂

线结果的方案三、方案四进行误差分析,得到表 3 结果,方案三、方案四不符合规范要求,我们将 3 根垂线即方案二进行误差界限分析,符合要求。

3 结论与建议

(1) 经过 Excel 多元线性回归分析精简的 3 根

表 3 简测法的误差界限(以常测法资料做精简)

规范要求	累计频率达 75% 以上误差	累计频率达 95% 以上误差
	不超过 $\pm 4\%$	不超过 $\pm 8\%$
方案三	58.33%	88.89%
方案四	52.78%	88.89%
方案二	83.33%	100.00%

根据《水文资料整编规范》定线要求,对方案二继续做关系线的符号检验、适线检验和偏离数值检验以及精度评定,结果见表 4。结果显示,经过精简后的 3 根垂线 V4、V6 和 V7 能够满足定线要求,可以认为定线正确,且定线的精度达到了一类精度的标准。

垂线方案通过检验且符合相关规范要求,可以作为简测法使用,且求算过程简单快捷,减少相关做图表时间,其中 F 检验、t 检验等统计检验值能够直观地得到结果。

(2) 进行回归分析精简时,可“从小到大”进行回归分析,即选取最小 P-value 值,做 1 根向几根

表 4 关系线检验及精度评定

测次	断面			P_i	$P_i - \bar{p}$	$(P_i - \bar{p})^2$	测次	断面			P_i	$P_i - \bar{p}$	$(P_i - \bar{p})^2$
	平均流速 (m/s)	线上断面 平均流速 (m/s)	相对误差 (%)					平均流速 (m/s)	线上断面 平均流速 (m/s)	相对误差 (%)			
1	0.47	0.46	2.42	2.40	5.74	23	1.02	1.05	-2.54	-2.56	6.55		
2	0.45	0.44	1.17	1.15	1.31	24	0.72	0.75	-3.81	-3.83	14.69		
3	0.46	0.45	2.07	2.04	4.18	25	0.53	0.53	-0.63	-0.65	0.42		
4	0.46	0.47	-2.02	-2.04	4.17	26	1.20	1.20	0.29	0.27	0.07		
5	0.24	0.23	6.64	6.62	43.85	27	0.48	0.49	-3.02	-3.04	9.23		
6	0.44	0.44	1.03	1.01	1.02	28	1.21	1.19	1.47	1.45	2.11		
7	0.45	0.47	-4.32	-4.34	18.88	29	0.76	0.76	0.48	0.46	0.21		
8	0.62	0.61	1.39	1.37	1.88	30	0.64	0.63	1.25	1.23	1.51		
9	0.38	0.37	3.12	3.10	9.60	31	1.32	1.28	2.98	2.95	8.72		
10	0.41	0.41	0.46	0.44	0.19	32	0.74	0.73	1.28	1.26	1.59		
11	0.69	0.69	-0.44	-0.46	0.22	33	1.32	1.36	-2.96	-2.99	8.91		
12	0.22	0.21	5.79	5.77	33.30	34	1.00	1.01	-0.94	-0.96	0.93		
13	0.38	0.38	-1.17	-1.19	1.43	35	0.23	0.25	-6.83	-6.85	46.97		
14	0.69	0.69	-0.03	-0.06	0.00	36	0.47	0.47	0.01	-0.01	0.00		
15	0.90	0.89	1.33	1.31	1.71	三项检验及精度评定							

(续表 4)

测次	断面	线上断面	相对误差			测次	断面	线上断面	相对误差		
	平均流速 (m/s)	平均流速 (m/s)	P_i (%)	$P_i - \bar{p}$	$(P_i - \bar{p})^2$		平均流速 (m/s)	平均流速 (m/s)	P_i (%)	$P_i - \bar{p}$	$(P_i - \bar{p})^2$
16	0.41	0.41	1.17	1.14	1.31		计算值	允许值	结果评定		
17	0.38	0.38	0.53	0.51	0.26	符号检验	0.50	1.15	合格		
18	0.36	0.36	-0.84	-0.87	0.75	适线检验	0.34	1.28	合格		
19	0.15	0.15	-1.51	-1.53	2.34	偏离数值检验	0.05	1.31	合格		
20	0.74	0.79	-6.61	-6.63	43.93	系统误差	0.09				
21	0.72	0.71	1.09	1.07	1.15	标准差	3%	达到一类定性精度			
22	1.07	1.02	4.89	4.86	23.66	随机不确定度	6%				

垂线的分析过程, 适合窄且规则的断面。当断面较宽, 明显需要几个垂线才能满足精简要求的, 建议从多根垂线着手进行精简。

(3) 进过垂线精简后的测流方案可以用于自动测流系统 V-ADCP 的应用, 作为现代化水文测报的依据。

(4) 每年应开展一定次数的测验工作用来比测, 并验证和完善精简分析成果。

参考文献:

- [1] 曹杰, 周洪顺, 方圆. 省界水资源监测站盱眙水文站流量测验精简分析[J]. 江苏水利, 2016(11):19-22.
- [2] 黄振平, 等. 水文统计学[M]. 北京: 中国水利水电出版社, 2017.
- [3] 赵明. 多元线性回归预测及其检验在 EXCEL 中的实现[J]. 吉林化工学院学报, 2003(2):85-87.
- [4] 水利电力部水利司. 水文测验手册[M]. 北京: 水利电力出版社, 1975.