

DOI: 10.19797/j.cnki.1000-0852.20210479

基于多元线性回归的雷达在线测流系统研究应用

郑凯¹, 赵淑饶², 吉俊峰², 贺俊华¹, 张莉¹

(1.黄河水利委员会 中游水文水资源局, 山西 晋中 030600;

2.黄河水利委员会水文局, 河南 郑州 450000)

摘要:为提高雷达在线测流系统的测验精度,通过拉伊达准则剔除粗大误差,以相对误差平方和最小为目标函数,利用多元线性回归分析方法,建立了多探头水面流速与断面平均流速的多元一次拟合公式。通过 t 检验,分量级、分类型探讨了雷达在线测流系统探头安装位置与最优数目问题。最后,以白家川站和丁家沟站为例,比较了几种雷达在线测流系统系数率定方法。结果表明,以相对误差平方和最小为目标函数的多元线性回归方法提高了拟合精度,降低了建设维护成本。

关键词:雷达在线测流系统;多元线性回归;相对误差; t 检验

中图分类号:P335

文献标识码:A

文章编号:1000-0852(2023)02-0030-05

0 引言

雷达在线测流系统(以下简称RG-30)利用声学多普勒原理^[1-2]进行流量测验,根据断面宽度,将多个RG-30探头固定于已建成的缆道上,通过流速面积法推算断面流量,从而实现流量在线监测。按照规范的比测要求^[3],测站多以铅鱼^[4-5]、测船^[6]或吊箱^[7]流速仪法、走航式ADCP法^[8]实测流量作为标准值,进行比测率定。但黄土高原山区性河流洪水暴涨暴落,漂浮物多,流速仪法或走航式ADCP法需接触水面进行流量测验,对测验人员及仪器设备均存在不安全因素^[9-10],资料收集困难。参照规范关于浮标法系数的确定方法^[3],可采用水位~流量关系曲线分析RG-30系数^[8,10-11]。因RG-30每6 min即可得到一个流量数据,此种方法具有数据量大的优势,解决了实测流量相应水位与RG-30相应水位无法对应的问题,在绘制水位~流量关系曲线的过程中也消除了一部分实测流量引起的误差。

为使RG-30虚流量转换为断面流量,朱志雄等^[11]建立了缆道测流结果与RG-30的相关关系;黄剑等^[12]建立了LS25-3A流速仪和RG-30的相关关系。但在建立相关关系过程中,绘制流速横向分布图存在一

定主观性,探头流速测验误差会发生传递,目标函数为绝对误差平方和最小,流量较小时会导致相对误差较大。

针对上述问题,本文以相对误差平方和最小为目标函数,建立了多探头水面流速与断面平均流速的关系,与传统方法进行了精度比较。

1 多元线性回归模型

1.1 概述

探头流速与断面平均流速的多元线性回归模型建立如下:

$$V'_i = a_0 + a_1 V_{i1} + \dots + a_n V_{in} \quad (1)$$

$$V_i = Q_{\text{实}i} / A_i \quad (2)$$

式中: V'_i 、 V_i 分别为第 i 次断面平均流速拟合值、实测值(m/s); V_{in} 为第 n 个探头第 i 次水面流速值(m/s); a_n 为系数; $Q_{\text{实}i}$ 为第 i 次实测流量或线查流量(m³/s); A_i 为第 i 次实测断面面积或借用断面面积(m²)。

上式的实际意义即在天然河流中,通过赋予代表垂线水面流速不同的权重,找出其与断面平均流速的关系,物理意义明确。

为求得式(1)的参数,需运用最小二乘法,对其进行参数估计,求绝对误差平方和的最小值点,即

收稿日期:2021-11-16

网络首发地址:https://kns.cnki.net/kcms/detail/11.1814.P.20230404.1704.003.html

作者简介:郑凯(1989—),男,广东揭阳人,本科,工程师,主要从事水文测验方面的研究工作。E-mail:771558153@qq.com

网络首发日期:2023-04-06

$$L = \min \left\{ \sum (V'_i - V_i)^2 \right\} \quad (3)$$

式中: V'_i 、 V_i 分别为第 i 次断面平均流速拟合值、实测值 (m/s)。

1.2 t检验

运用逐步回归法,通过逐步地从模型中剔除或加入变量(每次至多一个),使模型在有限步内既不能再从中剔除变量,也不能再向模型中增加变量,得到“最优”模型。

1.3 相对误差平方和最小的目标函数

在模型优化的基础上,按规范要求^[3]以相对误差平方和最小为目标函数,即

$$L = \min \left\{ \sum \left(\frac{V'_i - V_i}{V_i} \right)^2 \right\} \quad (4)$$

运用最小二乘法原理, n 元一次拟合公式:

$$V'_i = a_0 + a_1 V_{1i} + \dots + a_{n-1} V_{(n-1)i} + a_n V_{ni} \quad (5)$$

的回归方程组为:

$$\begin{pmatrix} \sum \frac{1}{V_i} & \sum \frac{V_{1i}}{V_i} & \dots & \sum \frac{V_{(n-1)i}}{V_i} & \sum \frac{V_{ni}}{V_i} \\ \sum \frac{V_{1i}}{V_i} & \sum \frac{V_{1i}^2}{V_i} & \dots & \sum \frac{V_{1i} V_{(n-1)i}}{V_i} & \sum \frac{V_{1i} V_{ni}}{V_i} \\ \dots & \dots & \dots & \dots & \dots \\ \sum \frac{V_{(n-1)i}}{V_i} & \sum \frac{V_{(n-1)i} V_{1i}}{V_i} & \dots & \sum \frac{V_{(n-1)i}^2}{V_i} & \sum \frac{V_{(n-1)i} V_{ni}}{V_i} \\ \sum \frac{V_{ni}}{V_i} & \sum \frac{V_{ni} V_{1i}}{V_i} & \dots & \sum \frac{V_{ni} V_{(n-1)i}}{V_i} & \sum \frac{V_{ni}^2}{V_i} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_{n-1} \\ a_n \end{pmatrix} = \begin{pmatrix} \sum V_{ii} \\ \dots \\ \sum V_{(n-1)i} \\ \dots \\ \sum V_{ni} \end{pmatrix} \quad (6)$$

式中: V'_i 、 V_i 分别为第 i 次断面平均流速拟合值、实测值 (m/s); V_{ni} 为第 n 个探头第 i 次水面流速 (m/s); a_n 为系数; m 为样本数。

通过 excel 的 minverse 函数求出系数矩阵的逆矩阵,再运用 mmult 函数即可求得解向量。

2 白家川站 RG-30 应用情况

白家川站为无定河出口控制站。洪水多集中在 6—9 月,一般由暴雨形成,涨落较快,历时较短,漂浮物较多,含沙量较大,一般沙峰滞后水峰。测验断面左岸为陡石坎,右岸为耕地,河床主要由岩石、砂卵石、细沙组成,冲淤变化不大。中高水流向顺直,主流稳定,水位~流量关系一般为单一线。

该系统建设在基本断面下游 4 m 处,由 5 个 RG-30 探头组成,分别位于起点距 30 m、40 m、56 m、73 m 和 85 m 处(见图 1)。

因洪水暴涨暴落、漂浮物较多等客观因素影响,流速仪法实测流量资料收集较为困难,故以线查流量为标准,选取 2020 年 8 月 5—6 日和 24—26 日全年最

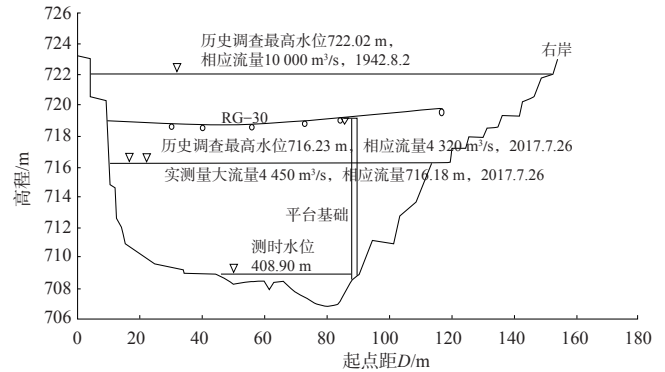


图1 RG-30 安装位置示意
Fig.1 The installation location of RG-30

大的 2 场洪水,分析了 5 探头均能测得水面流速的 112 组数据。文献[11]在应用过程中发现 4 级以上的风对 RG-30 测得结果有影响,故本文在分析过程中选取的样本数据风力均在 2 级以下。

比测期间雷达水位计水位资料、流速仪法流量资料均符合规范要求^[3,14](见表 1)。

表 1 RG-30 比测情况

项目	水位/m	相应线查流量/ (m³·s⁻¹)	含沙量/ (kg·m)	风速/ (m·s⁻¹)
范围	709.51 ~ 710.51	97.2 ~ 325	26.2 ~ 126	0 ~ 3.50

水位~流量关系曲线采用当年全部实测流量资料定线,定线精度符合规范要求^[14](见图 2)。

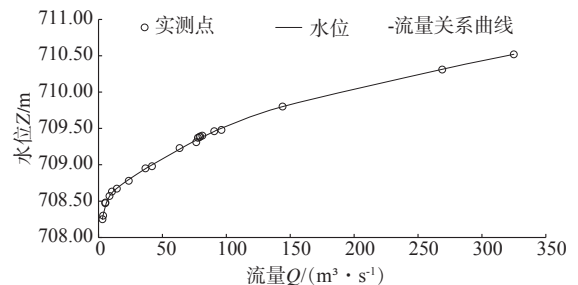


图 2 水位~流量关系曲线
Fig.2 The curve of stage-discharge relation

3 模型运用

3.1 方案制定与比较

为找出白家川站 5 探头流速与断面平均流速的最

优拟合关系,制定了以下5种方案。

方案1:传统方法。利用5探头水面流速数据,通过借用断面,绘制流速横向分布图和水道断面图,运用流速面积法计算虚流量,与相应水位对应的线查流量进行一元一次直线拟合。

方案2:剔除粗差后拟合。方法同方案1,但在确定拟合公式后,采用拉伊达准则^[15]分析剔除粗差,然后重新进行关系拟合。

方案3:多元线性回归分析。在方案2的基础上,采用多元线性回归分析方法,用5探头流速与断面平均流速建立5元一次方程。借用断面测次同方案2,断面平均流速为线查流量与借用断面面积之商。

方案4:模型优化。在方案3的基础上,通过t检验优化探头数量,剔除与断面平均流速关系不大的73 m、85 m探头流速,重新建立3元一次方程。

方案5:改变目标函数。在方案4的基础上,建立3元一次方程。

表2 不同拟合方案误差统计

序号	方案	公式	系统误差/%	随机不确定度/%	R ²
1	一元一次方程	$Q=0.9624Q_{虚}-23.0787$	0.10	9.69	0.9799
2	一元一次方程+去粗差	$Q'=0.96Q_{虚}-23.1476$	0.10	8.34	0.9841
3	去粗差+5元1次方程	$V'=0.3359+0.1563V_1+0.2948V_2+0.229V_3+0.0106V_4+0.0499V_5$	0.09	6.77	0.9891
4	去粗差+3元1次方程	$V'=0.3612+0.1487V_1+0.3152V_2+0.2614V_3$	0.09	6.76	0.9890
5	去粗差+相对误差平方和最小+3元1次方程	$V'=0.3458+0.1567V_1+0.2537V_2+0.3143V_3$	0	6.63	0.9886

注:V₁、V₂、V₃、V₄、V₅分别表示起点距30 m、40 m、56 m、73 m、85 m的探头流速,Q'表示断面拟合流量,V'表示断面平均拟合流速。本文随机不确定度均取2倍标准差。

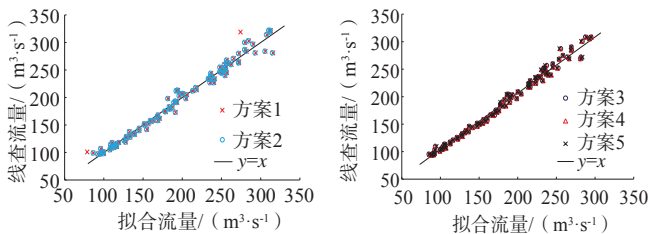


图3 各方案拟合流量与线查流量相关
Fig.3 The correlation of the fitting discharge and the curve discharge in different schemes

从表2、图3可以看出:

(1)方案1→.....→方案5,每步都能提高拟合精

度,系统误差由0.10%降到0%,随机不确定度由9.69%降到6.63%。

(2)方案1~2中,剔除2个粗差样本后,随机不确定度略有下降,由9.44%降到8.34%,R²由0.9799提高至0.9841,说明个别粗差的存在对拟合精度有一定影响。

(3)方案3~5中,方案3将一元回归优化为多元回归,将间接计算值Q_虚与线查值建关系优化为由直接测量值V_i与实测断面平均流速建关系。方案4运用t检验,得出V₄、V₅的尾概率P大于0.05(见表3),说明该两点处的探头流速与断面平均流速无明显相关,去掉上述两点数据后用剩余的V₁、V₂、V₃重新建立3元一次方程。方案5则将目标函数由绝对误差平方和改为相对误差平方和。通过上述步骤,系统误差由0.09%降到0%,随机不确定度由6.77%降到6.63%。虽然精度指标提升并不显著,但当样本容量增大、比测资料流量变幅增大时,上述方法对拟合精度的提高将会比较明显。

表3 各探头流速的t统计量

变量	V ₁	V ₂	V ₃	V ₄	V ₅
t值	4.76	4.44	3.18	0.13	0.76
尾概率 P{t ≥ t _α }	0	0	0.002	0.90	0.45

3.2 方案确定与检验

通过3.1节,确定方案5为最终方案,拟合公式如下:

$$V'=0.3458+0.1567V_1+0.2537V_2+0.3143V_3 \quad (7)$$

式中:V'为断面平均流速拟合值(m/s);V₁、V₂、V₃分别为起点距30 m、40 m、56 m处探头测得水面流速值(m/s)。

所有样本的相对误差范围-9.21%~8.35%,均未超过±10%,拟合精度较好。

方案5比方案1的实测点更集中于水位~流量关系曲线,但流量超过270 m/s时点据略有散乱,可能为含沙量影响水面流速系数或者借用断面面积非实测断面面积所致(见图4)。

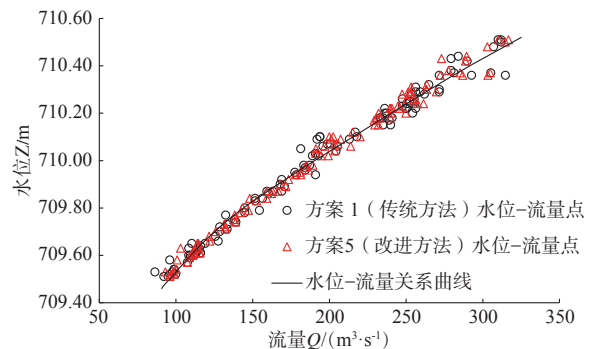


图4 RG-30实测点与水位~流量关系曲线
Fig.4 The RG-30 observed dots and the stage-discharge relation curve

通过与适用范围内的3次流速仪法实测流量比较,相对误差分别为-0.69%、-1.85%、8.18%,拟合精度较好。RG-30瞬时流速数据以流速仪法的相应水位为标准,在RG-30流速原始资料中找出相应的探头流速值。

3.3 模型推广

3.3.1 白家川站4探头流速

按照上述方法,对白家川站4探头能测得流速的2020年8月24日6:00至26日18:00数据与断面平均流速进行了拟合与优化,公式如下:

$$V' = 0.665 + 0.3226V_1 \quad (8)$$

式中: V' 为断面平均流速拟合值(m/s); V_1 为起点距73 m处探头测得水面流速值(m/s)。式(8)适用范围见表4。

表4 式(8)适用范围

项目	水位/m	相应线查流量/ ($m^3 \cdot s^{-1}$)	含沙量/ ($kg \cdot m^{-3}$)	风速/ ($m \cdot s^{-1}$)
适用范围	709.06 ~ 709.48	47.8 ~ 93.0	0 ~ 26.0	0 ~ 2.80

以线查流量为标准,计算得系统误差为0.06%,随机不确定度为4.73%,较方案1的0.31%、9.56%均有较大提升。

通过与适用范围内的10次流速仪法实测流量比较(其中4次未参与拟合计算),相对误差在-4.07% ~ 5.56%范围内,拟合精度较好。

为进一步验证式(8)的可行性,运用未参与拟合计算的2020年9月14日10:48 ~ 17日23:54的818个RG-30探头流速数据计算流量,与线查流量比较,相对误差范围-5.24% ~ 12.6%,系统误差1.38%,随机不确定度6.74%,期间水量相对误差1.30%,拟合精度良好。流量过程线对照见图5。

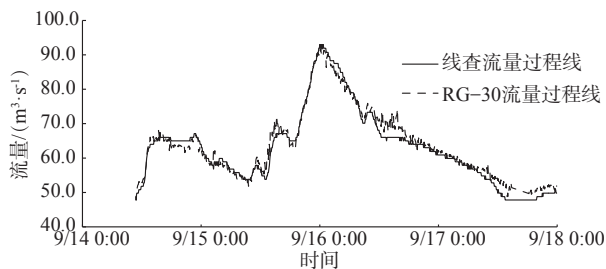


图5 流量过程线对照
Fig.5 the comparison of discharge hydrograph

3.3.2 丁家沟站

丁家沟站为无定河重要控制站。洪水一般由暴雨形成,涨落较快,历时较短,含沙量较大。水沙峰峰型多为复式峰,一般沙峰滞后水峰。中、高水水位~流量关系受附加比降和断面冲淤变化等因素综合影

响,多呈逆时针绳套。在起点距65 m、80 m、95 m、105 m、150 m处安装了5探头RG-30。

按照上述方法,对丁家沟站3探头(起点距80 m、95 m、105 m)能测得流速数据的31组样本与断面平均流速进行了拟合与优化,公式如下:

$$V' = 0.5487 + 0.2574V_1 + 0.3969V_2 \quad (9)$$

式中: V' 为断面平均流速拟合值(m/s); V_1 、 V_2 分别为起点距80 m、105 m处探头测得水面流速值(m/s)。式(9)适用范围见表5。

表5 式(9)适用范围

项目	水位/m	相应线查流量/ ($m^3 \cdot s^{-1}$)	含沙量/ ($kg \cdot m^{-3}$)	风速/ ($m \cdot s^{-1}$)
适用范围	815.17 ~ 815.57	81.9 ~ 144	38.5 ~ 75.2	0 ~ 3.90

以线查流量为标准,计算得系统误差为0%,随机不确定度为6.74%,较方案1的0.69%、13.02%均有较大提升(见图6)。图中9号线与12号线为根据流速仪法实测流量点据所定的临时水位~流量关系曲线。

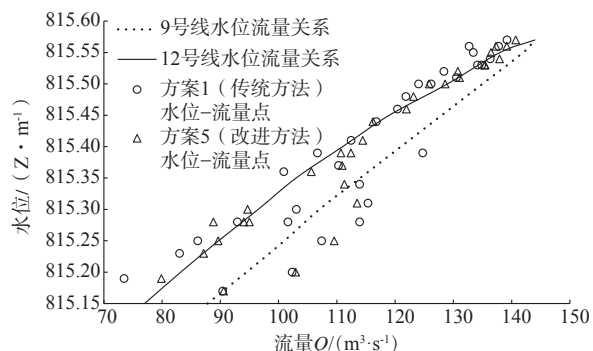


图6 RG-30实测点与水位~流量关系曲线
Fig.6 The RG-30 observed dots and the stage-discharge relation curve

通过与适用范围内的5次流速仪法实测流量比较,相对误差分别为-0.18%、0.99%、4.47%、5.21%、7.31%,拟合精度较好。

4 RG-30建设与应用启示

(1)当白家川站流量在47.8 ~ 93.0 m^3/s 时,起点距40 m、56 m、85 m探头流速与断面平均流速无明显相关,可用73 m探头流速拟合断面平均流速。73 m为该流量级的中泓点。故当水面流速横向分布为∧型时,只需在中泓点处安装RG-30探头,并用此处水面流速拟合断面平均流速。

(2)当白家川站流量在97.2 ~ 322 m^3/s 时,起点距73 m、85 m探头流速与断面平均流速无明显相关,可用起点距30 m、40 m、56 m探头流速的线性组合来拟

合断面平均流速。该流量级下主流呈一带状,位于起点距55~85 m之间。故当水面流速横向分布为—型时,除需在左侧主流边安装1个探头外,还需在流速缓慢变化得左侧区间等距离安装2个探头,并用此3处水面流速拟合断面平均流速。

(3)当丁家沟站流量在81.9~144 m³/s时,起点距95 m探头流速与断面平均流速无明显相关,可用起点距80 m、105 m探头流速的线性组合来拟合断面平均流速。受断面冲淤变化影响,该流量级下主流呈带状,位于起点距80~105 m之间,中泓点位于80 m、105 m。故当水面流速横向分布为马鞍型时,需在两个中泓点位置各安装1个RG-30探头,并用此2处水面流速拟合断面平均流速。

5 结论与展望

(1)RG-30流速系数率定时,粗大误差对拟合精度有一定影响,需在分析原因的基础上加以剔除。

(2)与传统的流速面积法相比,多元一次方程计算简便,省去了流速横向分布图及断面图的绘制,降低了人为主观误差和计算错误,使RG-30流量测验的物理意义更加明晰。同时断面平均拟合流速直接由探头流速计算,降低了计算过程中误差传递的风险。

(3)以相对误差平方和最小为目标函数,应用多元线性回归分析方法,可在一定程度上提高拟合精度,降低系统误差和随机不确定度,更符合水文实际。

(4)运用 t 检验,在基本不降低精度的情况下,分不同水面流速横向分布类型找出了最优探头安装位置和数量,节约了建设与维护成本。

本文虽然得到了RG-30系数拟合公式,但比测范围较小,在下一步工作中应继续补充中高水资料,注意分析新旧资料的一致性,若为同一系列,可继续扩

大拟合公式的使用范围,若不为同一系列,则按不同水位级分别予以处理。另外,分析过程中发现风和含沙量是影响RG-30测验精度的重要因素,可参照规范关于浮标系数的确定,推求考虑空气阻力对RG-30系数的影响公式,建立含沙量、水面流速、RG-30系数的关系^[3],以提高RG-30拟合精度。

参考文献:

- [1] 董向. RG-30非接触式雷达监测系统在水南站应用分析[J]. 甘肃水利水电技术,2016,52(4):7-9.
- [2] 江虹,桂笑,许柳青. 雷达波测流仪在中小河流的应用[J]. 水资源研究,2019,8(1):93-99.
- [3] GB 50179—2015,河流流量测验规范[S].
- [4] 詹戈.RG-30型雷达波流速仪在后峡水文站适用性分析[J]. 地下水,2021,43(3):193-194.
- [5] 陈甜,徐建秋,杨旭昌. 街口水文站自动监测系统设计与建设[J]. 水利信息化,2020(5):32-36.
- [6] 刘华,刘毅,王红彬. 黄河小浪底站RG30分析研究[J]. 东北水利水电,2020,38(11):24-25.
- [7] 杨丽萍. 雷达流量在线监测系统(RG-30)在内蒙古东居延海水文站的应用研究[J]. 内蒙古水利,2020(9):41-44.
- [8] 周波. 基于RG30雷达流速仪的流量在线监测系统设计与实现[C]//2015(第二届)城市防洪排涝国际论坛论文集. 广州:中国土木工程学会,2015.
- [9] 潘晓斌,胡辉华,陈端丹等. 移动雷达波测流系统在松柏水文站的应用[J]. 水利信息化,2021(4):64-67,76.
- [10] 薛荣正. 水吉(塔下)水文站缆道雷达波流量系数分析[J]. 水利科技,2021(2):15-17.
- [11] 朱治雄,谢永勇,高夏阳等. 雷达(RG-30)流量在线监测系统应用研究[J]. 人民黄河,2018,40(1):12-14,22.
- [12] 黄剑,刘铁林,王贞荣等. 非接触式雷达测流系统在吉安地区中小河流的应用研究[J]. 珠江水运,2018(18):53-54.
- [13] GB/T 50138—2010,水位观测标准[S].
- [14] SL/T 247—2020,水文资料整编规范[S].
- [15] 张留柱. 水文测量误差研究[D]. 南京:河海大学,2005.

Research and Application of RG-30 Based on Multiple Linear Regression

ZHENG Kai¹, ZHAO Shurao², JI Junfeng², HE Junhua¹, ZHANG Li¹

(1. Bureau of Midstream Hydrology and Water Resources, YRCC, Jinzhong 030600, China;

2. Hydrological Bureau, YRCC, Zhengzhou 450000, China)

Abstract: In order to improve the measuring accuracy of RG-30, the gross error was canceled by Rayda's criterion. Taking the minimum relative error's sum of squares as the target function, this paper established the multiple linear fitting formula by applying multiple linear regression, and discussed the problem of RG-30's probe installation position and optimal number on different discharge stages and categories by t test. At last, several methods of RG-30 coefficient calibration were compared at the Baijiachuan and Dingjiagou gauging stations. The results show that the method of multiple linear regression can improve the measuring accuracy of RG-30. Besides, it can reduce the construction maintenance costs.

Keywords: RG-30; multiple linear regression; relative error; t test