

DOI:10.3969/j.issn.1672-1144.2021.06.016

广义极值分布函数参数估计与设计洪水计算方法分析

袁超,王盼,李幼凤

(中国电建集团西北勘测设计研究院有限公司, 陕西 西安 710065)

摘要: 广义极值分布是国际上水文频率计算普遍使用的分布线型。为分析广义极值分布推求设计洪水的适用性,总结了广义极值分布函数分布参数、统计特征参数、线性矩、离均系数等相互关系与计算方法,并与国内常用皮尔逊Ⅲ型分布在统计参数、设计成果等方面的差异性进行了对比分析。结果表明利用广义极值分布推求设计洪水具有一定的适用性,但推求校核洪水的合理性仍有待进一步研究。

关键词: 广义极值分布;参数估计;线性矩法;历史洪水

中图分类号: TV122

文献标识码: A

文章编号: 1672—1144(2021)06—0100—05

Parameter Estimation and Design Flood Method in General Extreme Value Distribution

YUAN Chao, WANG Pan, LI Youfeng

(Power China Northwest Engineering Corporation Limited, Xi'an, Shaanxi 710065, China)

Abstract: General extreme value distribution (GEV) is commonly used in international hydrology frequency analysis. In order to assess the feasibility of GEV in design flood, this paper summarized the relationship among distribution parameters, statistic character parameters, L-moment and dispersion coefficient of general extreme value distribution. Meanwhile, compares with P-III distribution in difference of statistic character parameters and design value. The results show that GEV is feasible for design flood, but the rationality for check flood calculation still need more research.

Keywords: general extreme value distribution; parameter estimation; L-moment; historical flood

水利水电工程设计洪水计算常用统计分析方法,主要涉及频率分布模型的选择和参数估计方法的采用两方面的工作^[1]。目前,人们还无法从水文机理上证明水文事件的概率(频率)分布函数。实际中,计算者选用现有的随机变量分布函数,根据收集的水文数据进行审查,通过选用合理的参数估计方法进行分布函数参数计算,经分布函数拟合度检验后,依据水文数据的经验频率与选用分布函数理论频率之间的拟合效果来评估水文序列频率分布函数的拟合效果^[2]。

洪水频率分布模型大致可分为正态分布类、 Γ 分布类、极值分布类、指数分布类等^[1],其中广义极

值分布是国际上水文频率计算采用较多的分布线型^[1]。1998年,金光炎^[3]教授对广义极值分布的统计性质、离均系数计算进行了系统分析研究。2006年,刘聪等^[4]基于广义极值分布研究了最大风速频率分布模型。2011年,谢欣荣^[5]采用均值、变差系数和偏态系数等三个统计参数计算广义极值分布曲线,并给出了离均系数表。金连根^[6],李扬^[7],许弘喆等^[8]先后利用极值分布在设计风速、降水、潮水位推算中的应用。

洪水概率分布参数估计常见方法有矩法、极大似然法、概率权重矩法、线性矩法、权函数法及优化适线法等,除矩法外,上述各法均与所选定的分布模

收稿日期:2021-08-29

修稿日期:2021-09-25

基金项目:国家自然科学基金(52079110)

作者简介:袁超(1983—),男,硕士,高级工程师,主要从事水利水电工程水文及水资源工作。E-mail:121681618@qq.com

通讯作者:王盼(1989—),女,硕士,工程师,主要从事水文泥沙与水动力计算工作。E-mail:hhu_wp@163.com

型有关^[1]。Hosking(1990)在常规矩的基础上提出了线性距(L-moments)的概念,定义次序统计量线性组合的期望值为线性距。线性矩是概率加权矩的线性组合,是对概率权重矩的一种改进,与常规矩法相比,估计量具有良好的不偏性,变差系数和偏态系数的误差明显要小些^[5]。2003年,陈元芳等^[9]提出了两种可考虑历史洪水的样本线性距估计公式。2004年,段忠东等^[10]比较了极值概率分布参数估计方法。2005年,金光炎^[11]给出了矩、概率权重矩与线性矩之间的关系,分析了计算参数的近似公式及其精度。2007年,陈民等^[12]分析了极值分布统计参数与线性矩的关系。2008年,陈元芳等^[13]给出了具有历史洪水系列的线性矩公式,并对参数 k 值作了改进。国内学者对广义极值分布参数估计方法进行了比较^[14-18]。

基于上述研究成果,本文对广义极值分布函数分布参数、统计特征参数、线性矩、离均系数等相互关系与计算方法进行了全面总结,并与国内常用皮尔逊Ⅲ型分布在统计参数、设计成果等方面差异性进行了对比分析。

1 广义极值分布

1.1 函数形式

1928年,Fisher和Tippett证明极值具有渐近分布函数(即其极限概率分布),并概括了与原始分布对应的通常有三种类型的极限概率分布(渐近的极值分布)模型。其中,第Ⅰ型为指数原始分布,又称Gumbel分布(耿贝尔分布);第Ⅱ型为柯西型原始分布,亦称Fréchet分布(弗雷歇分布);第Ⅲ型为有界型分布,即Weibull分布(威布尔分布)^[5]。

$$F(x) = \exp[-(1 - ky)^{1/k}] \quad k \neq 0 \quad (1)$$

$$F(x) = \exp[-\exp(-y)] \quad k = 0 \quad (2)$$

式中: $y = (x - \xi)/\alpha$, k 、 α 和 ξ 分别为形状参数、尺度参数和位置参数。 $k = 0$ 时为第Ⅰ型分布, $k < 0$ 时为第Ⅱ型分布, $k > 0$ 时为第Ⅲ型分布。其概率密度函数为:

$$f(x) = (1 - ky)^{1/k-1} \exp[-(1 - ky)^{1/k}]/\alpha \quad k \neq 0 \quad (3)$$

$$f(x) = \exp[-y - \exp(-y)]/\alpha \quad k = 0 \quad (4)$$

1.2 分布参数与统计特征参数的关系

(1) 当 $k \neq 0$ 且 $k > -1/3$ 时,分布参数与统计特征参数关系式为^[5,12]:

$$\alpha = |k| EXC_v / [\Gamma(1 + 2k) - \Gamma^2(1 + k)]^{1/2} \quad (5)$$

$$\xi = EX \{1 - \text{sign}(k) * [1 - \Gamma(1 + k)] C_v\} / [\Gamma(1 +$$

$$2k) - \Gamma^2(1 + k)]^{1/2} \quad (6)$$

$$EX = \xi + \alpha [1 - \Gamma(1 + k)] / k \quad (7)$$

$$C_v = \alpha [\Gamma(1 + 2k) - \Gamma^2(1 + k)]^{1/2} / (|k| EX) \quad (8)$$

$$C_s = [\Gamma(1 + 3k) - 3\Gamma(1 + 2k)\Gamma(1 + k) + 2\Gamma^3(1 + k)] / [\Gamma(1 + 2k) - \Gamma^2(1 + k)]^{3/2} \quad (9)$$

式中: EX 为均值、 C_v 为离差系数、 C_s 为偏态系数,sign为符号函数, $\Gamma(*)$ 为gamma函数。

(2) 当 $k = 0$ 时,分布参数与统计特征参数关系式为^[5]:

$$\alpha = \sqrt{6} EX C_v / \pi \quad (10)$$

$$\xi = EX(1 - \sqrt{6}\gamma C_v / \pi) \quad (11)$$

$$EX = \xi + \alpha\gamma \quad (12)$$

$$C_v = \alpha\pi / (\sqrt{6} EX) \quad (13)$$

$$C_s = 1.139 547 1 \quad (14)$$

式中: γ 为欧拉常数,约等于0.577 215 7。

2 线性矩与参数估计

2.1 分布函数的线性矩

假定序列 X 服从特定分布函数,从序列中挑选出一组容量为 n 的样本,其次序统计量为 $x_{1:n} \leq x_{2:n} \leq \dots x_{n:n}$,则变量所服从分布函数的前四阶线性矩可以表示为^[18]:

$$\begin{cases} \lambda_1 = E(X_{1:1}) \\ \lambda_2 = \frac{1}{2} E(X_{2:2} - X_{3:3}) \\ \lambda_3 = \frac{1}{3} E(X_{3:3} - 2X_{2:3} + X_{1:3}) \\ \lambda_4 = \frac{1}{4} E(4X_{4:4} - 3X_{3:3} + 3X_{2:3} - X_{1:3}) \end{cases} \quad (15)$$

式中: $E(\cdot)$ 表示为期望, λ_1 为分布的线性位置(L-location)或为分布函数的均值; λ_2 为线性尺度(L-scale), λ_3 和 λ_4 为衡量分布函数的偏度和峰度的三阶、四阶线性矩。

线性矩法中,定义线性离差系数 τ_2 (即 $L-C_v$)、线性偏态系数 τ_3 (即 $L-C_s$)、线性峰度系数 τ_4 (即 $L-C_e$)分别表示为^[11,18]:

$$\begin{cases} \tau_2 = \lambda_2 / \lambda_1 \\ \tau_3 = \lambda_3 / \lambda_2 \\ \tau_4 = \lambda_4 / \lambda_2 \end{cases} \quad (16)$$

2.2 样本线性矩的计算公式

当样本容量有限,变量 X 的 n 观测值为 $x_1 \leq x_2 \leq \dots \leq x_n$,则 $\lambda_1, \lambda_2, \lambda_3$ 对应的样本矩 l_1, l_2, l_3 及线性

离差系数 τ_2 、线性偏态系数 τ_3 、线性峰度系数 τ_4 计算公式如下^[12-13]:

$$\begin{cases} l_1 = b_0 \\ l_2 = 2b_1 - b_0 \\ l_3 = 6b_2 - 6b_1 + b_0 \\ \tau_2 = l_2/l_1 \\ \tau_3 = l_3/l_2 \\ \tau_4 = l_4/l_2 \end{cases} \quad (17)$$

当样本为连续序列时, b_0 、 b_1 、 b_2 的计算公式如下^[13]:

$$b_0 = \frac{1}{n} \sum_{j=0}^n x_{jn} \quad (18)$$

$$b_1 = \frac{1}{n} \sum_{j=2}^n \frac{(j-1)}{(n-1)} x_{jn} \quad (19)$$

$$b_2 = \frac{1}{n} \sum_{j=3}^n \frac{(j-1)(j-2)}{(n-1)(n-2)} x_{jn} \quad (20)$$

当样本为具有历史洪水的非连续序列时, b_0 、 b_1 、 b_2 的计算公式如下^[13]:

$$b_0 = \frac{1}{N} \left[\sum_{m=n_0-a+1}^{n_0} x_m + \frac{N-a}{n_0-a} \sum_{m=a+1}^{n_0-a} x_m \right] \quad (21)$$

$$b_1 = \frac{1}{N} \left[\frac{N-a}{n_0-a} \sum_{m=2}^{n_0-a} \frac{m-1}{n_0-a-1} \frac{N-a+1}{N-1} x_m + \sum_{m=n_0-a+1}^{n_0} \frac{N-n_0+m-1}{N-1} x_m \right] \quad (22)$$

$$b_2 = \frac{1}{N} \left[\frac{N-a}{n_0-a} \sum_{m=3}^{n_0-a} \frac{(m-1)(m-2)}{(n_0-a-1)(n_0-a-2)} \frac{(N-a-1)(N-a-2)}{(N-1)(N-2)} x_m + \sum_{m=n_0-a+1}^{n_0} \frac{(N-n_0+m-1)(N-n_0+m-2)}{(N-1)(N-2)} x_m \right] \quad (23)$$

式中: N 为历史洪水考证期; a 为历史洪水个数; l 为实测洪水系列中特大洪水个数; n 为实测洪水系列长度, $n_0 = n - l + a$ 为洪水系列总个数。在此指出, 文献[7]中 b_1 、 b_2 中第一项求和公式 m 初值有误, 应分别为 2、3。

2.3 线性矩与分布参数的关系

(1) 当 $k \neq 0$ 且 $k > -1/3$ 时, 线性矩、分布参数及统计特征参数关系式为^[12]:

$$\alpha = \lambda_2 k / [(1 - 2^{-k}) \Gamma(1 + k)] \quad (24)$$

$$\xi = \lambda_1 - \alpha [1 - \Gamma(1 + k)] / k \quad (25)$$

$$EX = \lambda_1 \quad (26)$$

$$C_v = \alpha [\Gamma(1 + 2k) - \Gamma^2(1 + k)]^{1/2} / (|k| \lambda_1) \quad (27)$$

(2) 当 $k = 0$ 时, 分布参数与统计特征参数关系式为^[12]:

$$\alpha = \ln 2 / \lambda_2 \quad (28)$$

$$\xi = \lambda_1 - \frac{\gamma}{\alpha} \quad (29)$$

$$EX = \lambda_1 \quad (30)$$

$$C_v = \pi \lambda_2 / (\sqrt{6} \lambda_1 \ln 2) \quad (31)$$

(3) k 值改进公式。由于 $k = f(C_s)$ 无显式表达式, 常用试算或近似公式计算 k 值。当 $-0.5 \leq \tau_3 \leq 0.5$ 时, k 值近似公式为^[12-13]:

$$k \approx 7.8590c + 2.9554c^2 \quad (32)$$

式中: $c = 2/(3 + \tau_3) - \ln 2 / \ln 3$, 上式可将计算误差控制在 9×10^{-4} 以下^[13]。

(1) 当 $k < 0$ 且 $k > -1/3$ 时, k 值改进公式如下^[12-13]:

$$k = A_0 + A_1 \tau_3 + A_2 \tau_3^2 + A_3 \tau_3^3 \quad (33)$$

式中: $A_0 = 0.28194810$, $A_1 = -1.77112869$, $A_2 = 0.69440903$, $A_3 = -0.21940001$ 。

(2) 当 $k < 0$ 时, k 值改进公式如下^[12-13]:

$$k = B_0 + B_1 \tau_3 + B_2 \tau_3^2 + B_3 \tau_3^3 + B_4 \tau_3^4 + B_5 \tau_3^5 + B_6 \tau_3^6 \quad (34)$$

式中: $B_0 = 0.28382048$, $B_1 = -1.79626958$, $B_2 = 0.81065621$, $B_3 = -0.54534441$, $B_4 = 0.73859957$, $B_5 = 1.12411925$, $B_6 = 1.96961941$ 。

改进后的 k 值计算精度有较明显的提高^[13], 因此, 本文采用改进后的 k 值计算公式。

3 设计值推求

频率 P 的设计值 X_p 一般式如下:

$$X_p = EX(1 + \Phi_p C_v) \quad (35)$$

(1) 当 $k \neq 0$ 且 $k > -1/3$ 时, 离均系数 Φ_p 的计算公式为^[5]:

$$\Phi_p = \text{sign}(k) \times \{ \Gamma(1 + k) - [-\ln(1 - P)]^k \} / [\Gamma(1 + 2k) - \Gamma^2(1 + k)]^{1/2} \quad (36)$$

(2) 当 $k = 0$ 时, 离均系数 Φ_p 的计算公式为^[5]:

$$\Phi_p = -\sqrt{6} \{ \gamma + \ln[-\ln(1 - P)] \} / \pi \quad (37)$$

至此, 利用式(18)一式(19)或式(20)一式(22)与式(17)可求得样本矩 l_1 、 l_2 、 l_3 及线性偏态系数 τ_3 。利用式(33)或式(34)求得 k 值, 代入式(36)或式(37)求得离均系数 Φ_p , 代入式(9)求得偏态系数 C_s ; 利用式(24)一式(27)或式(28)一式(27), 求得均值 EX 、离差系数 C_v 。

4 应用实例

某水文站有 1940 年—2019 年年最大洪峰流量系列,调查历史洪水有 1864 年、1867 年、1931 年,1931 年洪水量级与 1980 年实测洪水相当,不作特大值处理,加入连续系列进行频率计算。 $N = 156$, $a = 2; l = 0, n = 81, n_0 = n - l + a = 83$ 。由式(21)一式(23)求得参数 b_0, b_1, b_2 ,按线性矩法估计统计特征参数,结果见表 1。为验证 GEV 分布的适用性,采用文献[11]中线性矩法估计 P-III 型频率曲线统

计特征参数,结果见表 1。GEV 分布与 P-III 分布频率曲线见图 1。

表 1 GEV 分布与 P-III 分布统计特征参数估计成果表

线型	参数估计方法	洪水均值 $/(m^3 \cdot s^{-1})$	C_v	C_s
GEV	矩法	1315	0.73	2.20
	线性矩法	1315	0.97	4.23
P-III	矩法	1315	0.73	2.20
	线性矩法	1315	0.92	3.29

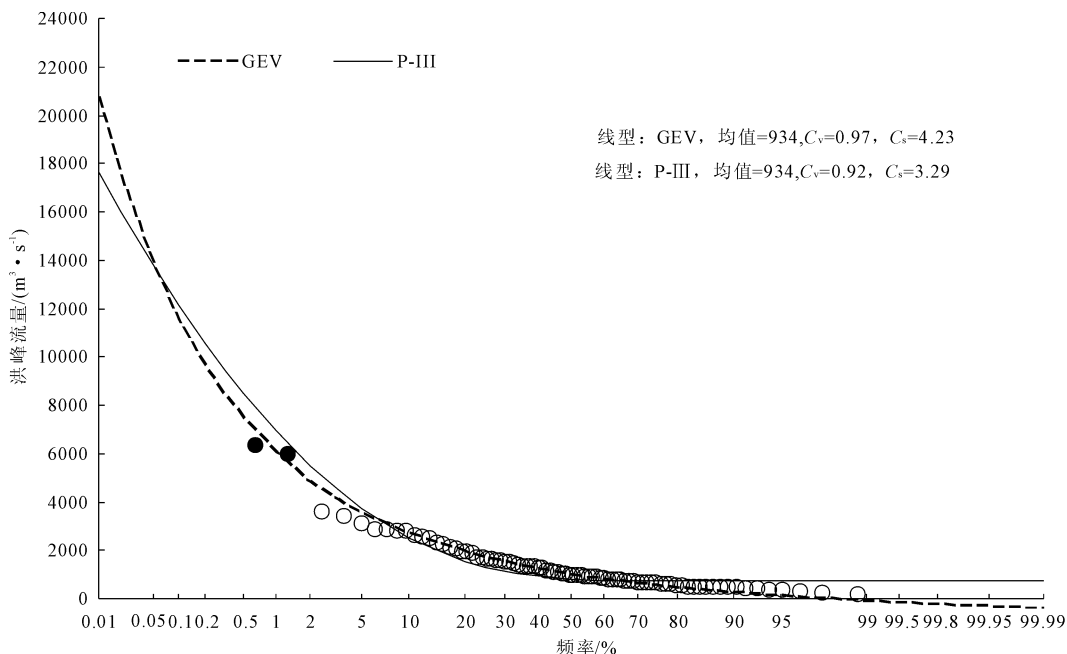


图 1 GEV 分布与 P-III 分布频率曲线图(线性矩法)

从参数估计看(见表 1),线性矩法与矩法估计均值相同,线性矩法估计 C_v 与 C_s 值偏大。对于 GEV 分布与 P-III 分布,线性矩法估计均值相同与 C_v 值相近, C_s 值相差较大。从计算结果看(见图 1),线性矩法求得的 GEV 分布与 P-III 分布频率曲线交叉,GEV 分布频率曲线上部上翘,中部下凹,尾部可能会出现负值的不合理情况,即当重现期大于 200 年时,GEV 分布计算设计成果偏大,且随着重现期的增大,两种线型偏离程度越大;当重现期在 10 年~200 年时,GEV 分布计算成果偏小,但两者相差不大,因此,采用 GEV 分布推求设计洪水具有一定的合理性,但推求校核洪水的合理性仍有待进一步研究。

5 结 论

本文对广义极值分布函数分布参数、统计特征

参数、线性矩、离均系数等相互关系与计算方法进行了全面总结,并与国内常用皮尔逊 III 型分布在统计参数、设计成果等方面的差异性进行了对比分析,得出以下几点结论:

(1) 线性矩法与矩法估计均值相同,线性矩法估计 C_v 与 C_s 值偏大。对于 GEV 分布与 P-III 分布,线性矩法估计均值相同与 C_v 值相近, C_s 值相差较大。

(2) 与 P-III 分布相比,GEV 分布曲线上部上翘,中部下凹,尾部可能会出现负值的不合理情况,即随着重现期的增大,两种线型偏离程度越大;在曲线中部,GEV 分布计算成果偏小,但两者相差不大。

(3) 广义极值分布推求设计洪水具有一定的适用性,但推求校核洪水的可靠性仍有待进一步研究。

参考文献:

- [1] 金光炎. 洪水频率分布模型和参数估计的研究进展 [C]//2004 年全国水文学学术讨论会/2004 年全国水文学学术讨论会论文集. 南京, 2004-12-11, 北京: 中国水利水电出版社, 2004: 43-48.
- [2] 宋松柏. 水文频率计算研究面临的挑战与建议 [J]. 水利与建筑工程学报, 2019, 17(2): 12-18.
- [3] 金光炎. 广义极值分布及其在水文中的应用 [J]. 水文, 1998(2): 10-16.
- [4] 刘 聪, 秦伟良, 江志红. 基于广义极值分布的设计基本风速及其置信限计算 [J]. 东南大学学报(自然科学版), 2006(2): 331-334.
- [5] 谢欣荣. 广义极值分布的适线法应用 [J]. 水文, 2011, 31(2): 27-31.
- [6] 金连根. 复合极值分布及其在台风多发海域设计风速推算中的应用 [J]. 水利与建筑工程学报, 2014, 12(3): 138-141, 168.
- [7] 李 扬. 极值指数估计在太原站月降水频率分析中的应用研究 [J]. 水利与建筑工程学报, 2015, 13(4): 55-59.
- [8] 许弘喆, 程素珍. 基于 MATLAB 的年最高潮水位极值 I 型分布频率分析计算 [J]. 水利与建筑工程学报, 2020, 18(4): 76-80.
- [9] 陈元芳, 沙志贵, 顾圣华, 等. 可考虑历史洪水对数正态分布线性矩的研究 [J]. 河海大学学报, 2003, 31(1): 80-83.
- [10] 段忠东, 周道成. 极值概率分布参数估计方法的比较研究 [J]. 哈尔滨工业大学学报, 2004, 36(12): 1605-1609.
- [11] 金光炎. 矩、概率权重矩与线性矩的关系分析 [J]. 水文, 2005(5): 3-8.
- [12] 陈 民, 陈元芳. 历史洪水变化时矩法与线性矩法统计性能对比分析 [J]. 水力发电, 2007(7): 20-22, 26.
- [13] 陈元芳, 李兴凯, 陈 民, 等. 可考虑历史洪水信息的广义极值分布线性矩法的研究 [J]. 水文, 2008, 28(3): 8-13.
- [14] 陈元芳, 李兴凯, 陈 民, 等. 虑历史洪水时 Gumbel 分布线性矩法的研究 [J]. 水能源科学, 2008, 38(3): 3-4.
- [15] 史景钊, 杨星钊, 陈新昌. 3 参数威布尔分布参数估计方法的比较研究 [J]. 河南农业大学学报, 2009, 43(4): 405-409.
- [16] 陈子燊, 刘曾美, 路剑飞. 广义极值分布参数估计方法的对比分析 [J]. 中山大学学报(自然科学版), 2010, 49(6): 105-109.
- [17] 马海波, 赵东亮, 祝薄丽. 几种水文频率曲线参数估计方法的比较 [J]. 人民黄河, 2016, 38(3): 9-11, 14.
- [18] 许文涛. 考虑日径流特征的非一致性洪水频率分析方法研究 [D]. 武汉: 武汉大学, 2019: 47-48.

(上接第 49 页)

- [6] 严雷鸣, 祝建中, 朱晓强, 等. 影响滩涂淤泥早期固化强度的多因子正交分析 [J]. 环境工程学报, 2017, 11(8): 4766-4774.
- [7] 袁飞飞. 高含水率滩涂淤泥固化土的工程特性研究 [D]. 杭州: 浙江大学, 2017.
- [8] 王功忠, 叶玉康. 滩涂围垦区河道开挖边坡失稳成因及治理方案探讨 [J]. 福建水力发电, 2020(2): 63-65.
- [9] 吴王意, 应乘风, 李娜, 等. 纤维改性滨海水泥土的直剪和无侧限特性试验研究 [J]. 绍兴文理学院学报(自然科学), 2018, 38(02): 45-50.
- [10] Wang W, Zhang C, Li N, et al. Characterisation of nano magnesia-cement-reinforced seashore soft soil by direct-shear test [J]. Marine Georesources and Geotechnology, 2019, 37(8): 1-10.
- [11] 梁仕华, 何 超, 龚 星, 等. 掺预处理垃圾焚烧飞灰水泥土的抗剪强度和浸出毒性试验研究 [J]. 硅酸盐通报, 2020, 39(2): 507-512.
- [12] 解邦龙, 张吾渝, 张丙印, 等. 不同龄期下粉煤灰水泥土的 UU 抗剪强度试验研究 [J]. 工程地质学报, 2021, 29(4): 1216-1223.
- [13] 邱英勤. 镍铁渣粉水泥土的力学性能及固化机理研究 [D]. 福州: 福州大学, 2018.
- [14] 陈金辉. 镍铁渣粉水泥土的固化机理试验研究 [J]. 水利与建筑工程学报, 2020, 18(1): 122-127.
- [15] 柯开展. 早龄期镍铁渣粉水泥土强度试验研究 [J]. 水利与建筑工程学报, 2018, 16(4): 192-195.
- [16] 通用硅酸盐水泥: GB 175—2007 [S]. 北京: 中国标准出版社, 2007.
- [17] 水泥制品用矿渣粉应用技术规程: JC/T 2238—2014 [S]. 北京: 中国建材工业出版社, 2014.