

基于线性可加模型的渡槽变形预测对比

邓树森¹, 朱赵辉^{1,2}, 吴浩^{1,2}, 王子文^{1,2}

(1. 中国水利水电科学研究院, 北京 100048; 2. 北京中水科工程集团有限公司, 北京 100048)

摘要: 渡槽是引调水工程中常见的输水建筑物, 准确预测渡槽变形对于确保水利工程稳定运行至关重要。为此, 以南水北调工程中的潦河渡槽为例, 基于渡槽长期变形监测数据建立弹性网回归、多元线性回归、逐步回归、岭回归和 LASSO 回归 5 种不同线性可加模型, 并采用 5 种不同线性可加模型对渡槽变形行为的预测结果进行比较。结果表明, 随着预测时间增加, 不同线性可加模型预测精度均呈现逐渐下降的趋势; LASSO 模型通过交叉验证选择最优正则化参数, 实现变量选择精简, 模型复杂度降至最低; 同时验证了训练长度会对多元线性回归和逐步回归的预测效果产生影响。研究结果可为选择渡槽变形预测模型提供一些有益的参考。

关键词: 渡槽; 变形; 预测; 线性可加模型

中图分类号: TV672⁺.3

文献标志码: A

文章编号: 1000-7709(2025)09-0123-04

1 引言

南水北调工程是我国解决水资源分布不均匀问题的世纪工程^[1], 该工程包含大量的渡槽结构, 具备输水、分水、调水等重要功能。受强震、异常工况、环境侵蚀等多方面因素影响, 渡槽结构会发生变形, 从而影响工程安全运行^[2]。因此选择合适的预测模型分析渡槽变形的发展趋势对于保障渡槽的安全运行具有重要意义。已有研究表明, 目前的监测模型主要运用于大坝工程中^[3], 如基于坝工理论和数学力学建立的大坝统计模型、基于优化变分模态分解与门控循环单元的预测模型、考虑库水压力作用滞后效应的渗漏模型、基于稳健估计的统计模型、分形—混沌混合预测模型等一系列大坝模型^[4-8]。SALAZAR F 等^[9]采用机器学习预测大坝变形, 并与统计模型进行比较; 顾冲时等^[10]指出统计模型影响因子大于 7 个, 易出现病态问题。而解决此类问题的常用方法是采用逐步回归, 但逐步回归分析是在非共线情况下选择变量的有用工具^[11]; 罗璐等^[12]采用 LASSO (Least Absolute Shrinkage and Selection Operator) 算法剔除不显著影响因子, 有效降低了

变形预测模型的复杂程度; ZOU H 等^[13]研究出弹性网解决变量之间相关性高的选择问题; 田果等^[14]采用弹性网进行面板数据变量选择, 结果显示弹性网可筛选出强相关变量。针对渡槽的安全监控模型较少且大多停留在定性分析阶段。为此, 本文以南水北调工程中的潦河渡槽为例, 基于渡槽长期变形监测数据, 通过对比弹性网回归、多元线性回归 (MLR)、逐步回归、岭回归和 LASSO 回归在预测渡槽变形方面的特性, 以期对监测渡槽变形提供理论依据。

2 基本原理及方法

2.1 基本原理

弹性网^[15]结合岭回归和 LASSO 的正则化来增强模型的预测准确性和解释性, 具体形式为:

$$\hat{\beta}_{EN} = \operatorname{argmin} \left\{ \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right\} \quad (1)$$

式中, n 为变量总数; y_i 为因变量; β_0 为截距项; x_{ij} 为自变量; β_j 为系数; λ_1, λ_2 均为非负调节参数; p 为自变量个数。

令 $\alpha = \lambda_1 / (\lambda_1 + \lambda_2)$, $\lambda = \lambda_1 + \lambda_2$, 可简化为:

收稿日期: 2024-10-23, 修回日期: 2024-11-25

基金项目: 基于信息驱动的水库大坝安全风险态势管控平台研究的项目 (ZS1003062024)

作者简介: 邓树森 (2001-), 男, 硕士研究生, 研究方向为大坝安全监测理论与方法, E-mail: dengshusen@edu.iwhr.com

通讯作者: 朱赵辉 (1981-), 男, 正高级工程师, 研究方向为大坝安全监测理论与方法, E-mail: zzhbb@163.com

$$\hat{\beta}_{EN} = \operatorname{argmin} \left\{ \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j)^2 + \lambda \left[\alpha \sum_{j=1}^p |\beta_j| + (1 - \alpha) \sum_{j=1}^p \beta_j^2 \right] \right\} \quad (2)$$

为避免量纲影响,对数据进行标准化处理,标准化公式为:

$$x_{ij} = \frac{X_{ij} - \bar{x}_i}{\sqrt{\frac{1}{p} \sum_{j=1}^p (X_{ij} - \bar{x}_i)^2}} \quad (3)$$

式中, X_{ij} 为原始测值; \bar{x}_i 为第 i 个变量的平均值。

根据式(3)进行数据标准化,再代入式(2)中求解即可得到 $\hat{\beta}_{EN}$ 。

2.2 评价指标

为定量分析预测结果,采用平均绝对误差(M_{MAE})、均方根误差(R_{RMSE})来验证模型的有效性。

2.3 影响因素

渡槽受水压、环境温度及时效因素影响^[16],得到变形表达式为:

$$y = y_H + y_T + y_\theta \quad (4)$$

式中, y 为渡槽变形量; y_H 为水压引起的渡槽变形分量; y_T 为渡槽结构混凝土温度变化引起的渡槽变形分量; y_θ 为时间效应引起的渡槽变形分量。

(1)水位因子。 y_H 与水压呈线性关系,一般使用 H 的幂次方表示:

$$\begin{cases} H_1 = H \\ H_2 = H^2 \\ H_3 = H^3 \end{cases} \quad (5)$$

(2)温度因子。温度随季节变化发生周期性波动引起混凝土胀缩,导致渡槽变形。因此可使用下式计算温度因子:

$$y_T = \sin \frac{2\pi it}{365} - \sin \frac{2\pi it_0}{365} \quad (6)$$

$$y_T = \cos \frac{2\pi it}{365} - \cos \frac{2\pi it_0}{365} \quad (7)$$

式中, t 为当前时刻; t_0 为监测基准时刻。

基于式(6)、(7)及实际环境温度得到计算温

度因子的公式:

$$\begin{cases} T_1 = \sin \frac{2\pi t}{365} - \sin \frac{2\pi t_0}{365} \\ T_2 = \sin \frac{4\pi t}{365} - \sin \frac{4\pi t_0}{365} \\ T_3 = \cos \frac{2\pi t}{365} - \cos \frac{2\pi t_0}{365} \\ T_4 = \cos \frac{4\pi t}{365} - \cos \frac{4\pi t_0}{365} \\ T_5 = T^2 \end{cases} \quad (8)$$

式中, $i=1,2$; T_1, T_2, T_3, T_4, T_5 均为温度因子; T 为实测环境温度。

(3)时效因子。通常考虑的时效因子有如下4种形式:

$$\begin{cases} I_1 = \ln(t - t_0 + 1) \\ I_2 = (t - t_0)/(1 + t - t_0) \\ I_3 = (t - t_0)^{0.5} \\ I_4 = (t - t_0)^{-0.5} \end{cases} \quad (9)$$

式中, I_i 为时效因子, i 取 1,2,3,4。

2.4 方法

本文回归方法均在 R 语言软件中实现,其中弹性网、LASSO、岭回归均基于 R 语言中的 `glmnet`^[17] 包实现,逐步回归和多元线性回归则采用 R 语言软件中的内置函数实现。逐步回归依据 AIC(Akaike Information Criterion)准则进行变量选择。

3 沉降变形监测

3.1 工程概况

潦河渡槽为涵洞式渡槽,采用双线双槽,布置两个矩形槽,单槽净宽 11 m,侧墙高 8.4 m,槽身侧墙顶端之间设拉杆,拉杆设置间距为 2.65 m,共 78 个拉杆。渡槽表面布置 72 个水准点,人工监测渡槽沉降变化,监测频率 1 次/月,监测点布置见图 1,渡槽结构见图 2。由图 2 可知,下部涵洞共 26 孔,共 9 联,前 8 联为三孔一联,第 9 联为

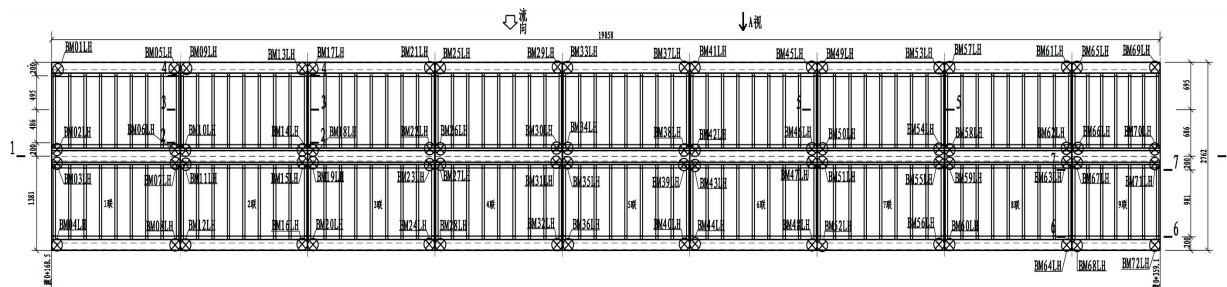


图 1 渡槽监测点布置图(单位:mm)

Fig. 1 Aqueduct monitoring point layout diagram

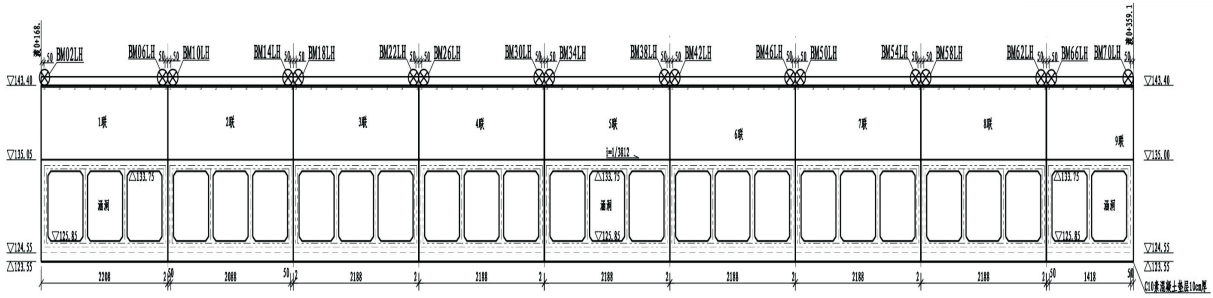


图 2 渡槽结构图(单位:mm)

Fig. 2 Structural diagram of the aqueduct

两孔一联,孔口尺寸为 6.1 m×7.9 m(宽×高)。涵洞地基主要由粘土岩、砂岩组成,局部有薄层全新统砾质粗砂、砂砾石层。

3.2 变量选择结果对比

选择 BM13~BM16 测点 2016 年 6 月~2023 年 12 月的监测数据为样本,其中 2016 年 6 月~2020 年 9 月的数据作为训练集,剩余数据作为预测集。采取式(5)、(8)、(9)计算影响变量,得到测点的变量集合为 $\{H_1, H_2, H_3, T, T_1, T_2, T_3, T_4, T_5, I_1, I_2, I_3, I_4\}$ 。

以测点 BM15 为具体案例说明变量的选择过程,由于多元线性回归和岭回归不具备变量选择的功能,因此所有的影响因子均作为输入变量,逐步回归则基于 AIC 准则选择。在利用 LASSO 和弹性网进行选择时,通过 10 折交叉验证方法确定各自的最优正则化参数 λ 。随着参数 λ 的对数值变化时均方误差的变化趋势为:当 λ 值较小时,均方误差也较小。表明在低正则化强度下,模型的拟合能力较强,可能存在过拟合的风险。随着 λ 值的增加,均方误差逐渐增大,正则化开始发挥作用,限制模型的复杂度,减少过拟合的可能性。在某个特定的 λ 值处,模型实现最佳平衡,得到相应的变量数量。据此,LASSO 方法选择 5 个变量作为模型的最优组成,弹性网选择 6 个变量。

不同模型的变量选择结果见表 1。由表 1 可看出,4 个测点均是 LASSO 选择的变量因子数量最少,模型复杂度最低。逐步回归和弹性网选择变量因子数量接近,略大于 LASSO 选择的变量因子数量。渡槽的变形应综合考虑水位、温度及时效等多种因素的影响。针对 BM13 测点,逐步回归、LASSO 及弹性网回归均合理纳入水位变量、温度变量和时效变量。BM14、BM15、BM16 测点变量选择结果表明不同方法所选择的变量类型呈现出出不一致性,且存在变量类型缺失的情况,这与理论预期相悖。

表 1 不同模型变量选择结果

Tab. 1 Results of variable selection for different models

测点	模型	选择结果	个数
BM13	多元线性回归	$H, H_2, H_3, T, T_1, T_2, T_3, T_4, T_5, I_1, I_2, I_3, I_4$	13
	逐步回归	$H, H_2, H_3, T_1, T_3, I_1, I_2, I_4$	8
	弹性网	$H_3, T_5, T_1, I_1, I_2, I_3, I_4$	7
	LASSO	H_3, T_5, T_1, I_2, I_3	5
	岭回归	$H, H_2, H_3, T, T_1, T_2, T_3, T_4, T_5, I_1, I_2, I_3, I_4$	13
BM14	多元线性回归	$H, H_2, H_3, T, T_1, T_2, T_3, T_4, T_5, I_1, I_2, I_3, I_4$	13
	逐步回归	$T, T_1, T_2, T_5, I_1, I_2, I_3, I_4$	8
	弹性网	$H, H_3, T, T_1, T_2, T_3, T_4, T_5, I_2, I_3$	10
	LASSO	T_1, T_5, I_2, I_3	4
	岭回归	$H, H_2, H_3, T, T_1, T_2, T_3, T_4, T_5, I_1, I_2, I_3, I_4$	13
BM15	多元线性回归	$H, H_2, H_3, T, T_1, T_2, T_3, T_4, T_5, I_1, I_2, I_3, I_4$	13
	逐步回归	$T_1, T_5, I_1, I_2, I_3, I_4$	6
	弹性网	$T_1, T_5, I_1, I_2, I_3, I_4$	6
	LASSO	T_5, T_1, I_1, I_2, I_3	5
	岭回归	$H, H_2, H_3, T, T_1, T_2, T_3, T_4, T_5, I_1, I_2, I_3, I_4$	13
BM16	多元线性回归	$H, H_2, H_3, T, T_1, T_2, T_3, T_4, T_5, I_1, I_2, I_3, I_4$	13
	逐步回归	$H_2, H_3, T_1, I_1, I_2, I_3, I_4$	7
	弹性网	$T_1, T_5, I_1, I_2, I_3, I_4$	6
	LASSO	T_1, T_5, I_2, I_3	4
	岭回归	$H_2, H_3, T, T_1, T_2, T_3, T_4, T_5, I_1, I_2, I_3, I_4$	13

3.3 结果与讨论

采用 5 种回归方法对 4 个测点进行预测效果比较,结果见图 3。由图 3 可看出,基于多元线性回归框架并依据 AIC 准则进行变量选择,逐步回归与线性回归的预测趋势呈现出一致性;同时通过施加不同程度的惩罚因子进行变量选择与模型

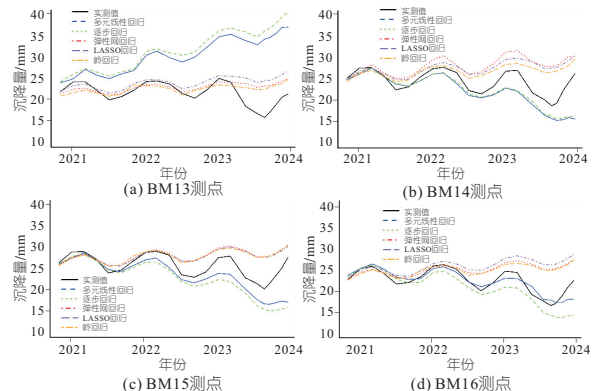


图 3 BM13~BM16 测点各模型预测性能对比图

Fig. 3 Comparison of prediction performance for each model at BM13-BM16 measurement point

构建,LASSO 回归、岭回归及弹性网回归的预测趋势也相互吻合。分析结果表明,采用不同的变量选择与模型构建后,特定回归方法间的预测趋势仍能展现出一致性。此外,不同模型预测规律表明,回归模型的预测准确度均随着时间的推移而逐渐下降。

表 2 为预测集的评价指标, R_{RMSE} 是衡量预测值与观测值之间差异的一个统计量,反映样本标准差的大小。在变量选择不受限制的情况下,随着纳入模型变量数量的增加, R_{RMSE} 值倾向于减小。由表 2 可知,岭回归和多元线性回归方法所得到的 R_{RMSE} 值达到最小,但模型复杂度较高。平均绝对误差(M_{MAE})表明,逐步回归方法展现出相对优势。

表 2 预测集的评价指标

Tab. 2 Evaluation metrics for the prediction set

测点	模型	R_{RMSE}	M_{MAE}	测点	模型	R_{RMSE}	M_{MAE}
BM13	多元线性回归	11.16	9.48	BM15	多元线性回归	3.72	2.78
	逐步回归	12.34	10.55		逐步回归	2.85	2.02
	弹性网	3.02	2.26		弹性网	3.62	2.78
	LASSO	3.51	2.48		LASSO	4.57	3.77
	岭回归	2.94	2.30		岭回归	3.61	2.80
BM14	多元线性回归	2.52	2.77	BM16	多元线性回归	1.72	1.38
	逐步回归	3.72	2.70		逐步回归	1.93	1.51
	弹性网	4.81	4.00		弹性网	4.25	3.12
	LASSO	4.77	3.98		LASSO	5.06	3.89
	岭回归	3.67	2.84		岭回归	4.26	3.14

根据图 3(a)所示,逐步回归与多元线性回归模型在预测 BM13 时表现出一定的局限性。为验证是否由于训练长度导致预测效果降低,将训练集的时间范围扩展至 2022 年 7 月。图 4 为 BM13 扩展训练集后各模型预测性能对比图。由图 4 可知,训练集的扩展对逐步回归与多元线性回归的预测精度产生了显著提升效果。证明训练样本的长度会影响多元线性回归和逐步回归预测精度,分析结果与江守燕等^[16]研究结论一致。

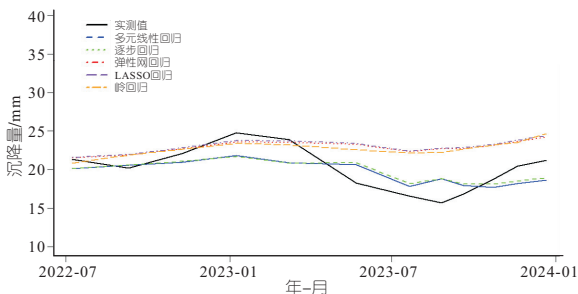


图 4 BM13 扩展训练集后各模型预测性能对比图

Fig. 4 Comparison of prediction performance among models after the BM13 training set expansion

4 结论

a. 本文以潦河渡槽为例,基于长期变形监测

数据建立 5 种不同线性可加模型,并对比分析了渡槽的变形预测结果。

b. 5 种线性可加模型从 13 个变量中筛选出合适变量进行预测,LASSO 选择的变量数量最少,模型复杂度最低,逐步回归和弹性网选择的变量数量接近,多元线性回归和岭回归覆盖所有变量,模型复杂度最高。

c. 纳入变量最多的岭回归和多元线性回归的 R_{RMSE} 最小,逐步回归的 M_{MAE} 指标占据优势。表明不同指标会产生不同的模型选择结果,实际应用时应根据工程需要合理选择模型。此外,线性可加模型预测精度随着时间延长会逐渐下降。

d. 训练集长度影响多元线性回归和逐步回归的预测精度,不断将最新的监测数据补充至训练集中建立新的预测模型,可得到更为精确的预测值。

参考文献:

- [1] 朱伟,付柯锦,李红艳,等. 南水北调中线水源地水生态与经济发展适配性——以南阳市为例[J/OL]. 南水北调与水利科技(中英文). <https://link.cnki.net/urlid/13.1430.tv.20241010.1623.006>.
- [2] 张建伟,曹克磊,赵瑜,等. 基于流固耦合模型的 U 型渡槽模态分析及验证[J]. 农业工程学报, 2016, 32(18): 98-104.
- [3] 王崧,黄耀英,刘钰,等. 大型渡槽实测钢筋应力定量与可靠性分析[J]. 长江科学院院报, 2019, 36(11): 21-26.
- [4] 吴中如,沈长松,阮焕祥. 论混凝土坝变形统计模型的因子选择[J]. 河海大学学报, 1988, 16(6): 1-9.
- [5] 张建中,顾冲时,袁冬阳,等. 基于优化 VMD 与 GRU 的混凝土坝变形预测模型[J]. 水利水电科技进展, 2023, 43(5): 38-44.
- [6] 程乐群,陈龙,顾冲时. 碾压混凝土坝变形监控模型研究[J]. 水电能源科学, 2008, 26(1): 107-110.
- [7] 朱赵辉,刘健,李新,等. 基于稳健估计的大坝变形监测统计模型分析[J]. 中国水利水电科学研究院学报, 2018, 16(2): 105-112.
- [8] 何鲜峰,顾冲时,谷艳昌. 分形-混沌混合预测模型在大坝安全监测中的应用[J]. 武汉大学学报(工学版), 2008, 41(1): 45-49.
- [9] SALAZAR F, TOLEDO M A, OÑATE E, et al. An empirical comparison of machine learning techniques for dam behaviour modelling[J]. Structural safety, 2015, 56: 9-17.
- [10] 顾冲时,汪亚超,彭妍,等. 大坝安全监控模型的病态问题及其处理方法[J]. 中国科学: 技术科学, 2011, 41(12): 1574-1579.
- [11] 吴子安. 大坝变形监测数据回归分析中的因子选择[J]. 武汉测绘科技大学学报, 1993, 18(1): 20-26.

(下转第 118 页)

孔周围区域网格密度大、计算效率低及不同网格划分方式导致的计算结果差异问题,提出了等效(E-FEM)模型。相同计算精度下对比了E-FEM模型和原始(Ref-FEM)模型的计算效率,发现主要模型网格数量降低了25.2%,同时TCT减少75.1%,WCT减少了71.9%,显著提高计算效率。

参考文献:

- [1] FOLIAS E S. On the three-dimensional theory of cracked plates[J]. *Journal of applied mechanics*, 1975, 42(3): 663-674.
- [2] 杜平安. 有限元网格划分的基本原则[J]. *机械设计与制造*, 2000(1): 34-36.
- [3] 李海峰, 吴冀川, 刘建波, 等. 有限元网格剖分与网格质量判定指标[J]. *中国机械工程*, 2012, 23(3): 368-377.

- [4] 范崇仁. 水工钢结构[M]. 5版. 北京: 中国水利水电出版社, 2019.
- [5] 许中武, 蔡伟, 周建方. 平面钢闸门三维有限元线性分析的若干问题[J]. *中国农村水利水电*, 2023(8): 222-228.
- [6] HAJIZADEH A, KHATIBINIA M. Fundamental mode shape-based damage detection in arbitrary shaped plates using a hybrid of contourlet transform and Delaunay triangulation[J]. *Structures*, 2023, 58: 105631.
- [7] 任林娟, 周平章. 基于规则背景网格的数值积分[J/OL]. *计算力学学报*. <https://link.cnki.net/urlid/121.1373.03.20240718.1644.004>.
- [8] KIRAN U, SHARMA D, GAUTAM S S. A GPU-based framework for finite element analysis of elastoplastic problems[J]. *Computing*, 2023, 105(8): 1673-1696.

Equivalent Simplification of Gate with Perforated Structure Based on Superposition Method

HUANG Yu-xi, LI Cheng, XUE Qing-chen, CHEN Dong

(School of Mechanical and Power Engineering, Zhengzhou University, Zhengzhou 450001, China)

Abstract: Perforated structures are commonly found in engineering applications, where stress concentration around holes significantly affects structural load-bearing capacity and safety. Therefore, accurate mechanical analysis of perforated structures is essential. However, the traditional finite element analysis methods face several challenges, including complex mesh generation, high computational resource requirements, and poor convergence when dealing with perforated structures. To overcome these challenges, a finite element analysis method based on the superposition principle is proposed. This approach replaces the original model with a simplified equivalent model featuring a less complex mesh for simulation. The stress field of sub-models is subsequently employed to correct the stress distribution around holes in the equivalent model. Comparative results indicate that the discrepancy between the equivalent model and the original model is within 5%. The equivalent model achieves a 25.2% reduction in mesh elements and a 75.1% reduction in simulation time, substantially improving simulation efficiency.

Key words: planar steel gate; stress concentration; mesh generation; superposition method; finite element simulation

(上接第126页)

- [12] 罗璐, 李志, 张启灵. 大坝变形预测的最优因子长短期记忆网络模型[J]. *水力发电学报*, 2023, 42(2): 24-35.
- [13] ZOU H, HASTIE T. Regularization and variable selection via the elastic net[J]. *Journal of the royal statistical society series B: Statistical methodology*, 2005, 67(2): 301-320.
- [14] 田果, 杨宜平. 面板数据固定效应线性回归模型的变量选择[J/OL]. *山西大学学报(自然科学版)*. <https://doi.org/10.13451/j.sxu.ns.2024023>.

- [15] 魏巍, 王星惠, 陈晓星. 基于弹性网的两阶段模型平均方法及应用研究[J]. *哈尔滨师范大学自然科学学报*, 2022, 38(6): 47-53.
- [16] 江守燕, 赵林鑫, 杜成斌, 等. 基于多元线性回归的渡槽变形预测模型研究[J]. *三峡大学学报(自然科学版)*, 2020, 42(2): 12-17.
- [17] KENNETH TAY J, NARASIMHAN B, HASTIE T. Elastic net regularization paths for all generalized linear models[J]. *Journal of statistical software*, 2023, 106: 1.

Comparison of Aqueduct Deformation Prediction Based on Linear Additive Models

DENG Shu-sen¹, ZHU Zhao-hui^{1,2}, WU Hao^{1,2}, WANG Zi-wen^{1,2}

(1. China Institute of Water Resources and Hydropower Research, Beijing 100048, China;

2. Beijing IWHR Technology Co., Ltd., Beijing 100048, China)

Abstract: Aqueducts are common water conveyance structures in water diversion projects, and accurate prediction of aqueduct deformation is crucial for ensuring the stable operation of water conservancy projects. For this purpose, taking the Liaohe Aqueduct in the South-to-North Water Diversion Project as an example, five different linear additive models, namely elastic net regression, multiple linear regression, stepwise regression, ridge regression and LASSO regression, were established based on the long-term deformation monitoring data of the aqueduct. The prediction results of the aqueduct's deformation behavior by the five different linear additive models were compared. The results indicate that as the prediction time increases, the prediction accuracy of different linear additive models gradually decreases. The LASSO model selects the optimal regularization parameter through cross-validation, achieving variable selection simplification and minimizing model complexity. Additionally, it is verified that the training length affects the prediction performance of multiple linear regression and stepwise regression. The findings of this study provide valuable references for selecting prediction model of aqueduct deformation.

Key words: aqueduct; deformation; prediction; linear additive model