

DOI: 10. 19853/j. zgjsps. 1000-4602. 2025. 13. 018

基于支持向量机回归的城市排水窨井水位预测

赖泽辉¹, 彭进双¹, 包世泰², 曹 兵¹

(1. 奥格科技股份有限公司, 广东 广州 510663; 2. 华南农业大学 资源环境学院, 广东 广州 510642)

摘 要: 为实时掌握城市排水管网水位状态,减少水位监测设备长期运行的维护成本,通过获取排水窨井水位及降雨量短期监测数据,分析水位站点水文特征因子,制作样本数据集,并对支持向量机回归算法进行优化训练,构建城市排水窨井水位预测模型,最后利用典型天气数据对模型进行测试,模型预测结果精度良好。与长期布设窨井水位传感器开展长期监测相比,该方法通过获取短期水位数据即可对窨井实时水位进行预测,具有部署方便、运维成本低的特点,在城市排水日常提质增效运维、内涝应急调度等业务应用中具有较高的实用价值。

关键词: 城市排水系统; 水位预测; 支持向量机回归; 提质增效

中图分类号: TU992 **文献标识码:** A **文章编号:** 1000-4602(2025)13-0126-05

Prediction of Urban Drainage Manhole Water Level Using Support Vector Machine Regression

LAI Ze-hui¹, PENG Jin-shuang¹, BAO Shi-tai², CAO Bing¹

(1. Augur Intelligent Technology Co. Ltd., Guangzhou 510663, China; 2. College of Resource and Environment, South China Agricultural University, Guangzhou 510642, China)

Abstract: A model for predicting urban drainage manhole water level was constructed to monitor the water level status of the urban drainage network in real time and reduce the long-term operational maintenance costs of water level monitoring equipment. This involved collecting short-term monitoring data on manhole water levels and rainfall, analyzing hydrological characteristic factors of water level stations, creating a sample dataset, and optimizing and training a support vector machine regression algorithm. The model was subsequently tested using typical weather data, and the results demonstrated high prediction accuracy. Compared with the long-term installation of manhole water level sensors for continuous monitoring, this approach enabled the prediction of real-time manhole water levels by acquiring short-term water level data. It was characterized by ease of deployment and low operational and maintenance costs, making it highly practical for business applications such as daily quality and efficiency enhancement in urban drainage operations, and emergency dispatching during urban waterlogging events.

Key words: urban drainage system; water level prediction; support vector machine regression; quality and efficiency improvement

城市排水管网高水位运行是外水入侵排水系统最直观的表现。排水管网高水位运行,导致管网水

体流速降低,污染物沉积管道,受降雨二次冲刷影响,污染物随溢流进入受纳水体,易引起水体黑臭,二次

危害巨大。对排水管网水位运行状态进行实时监测,识别排水管网高水位风险,能有效辅助排查排水管网缺陷,为排水运维调度、系统设计改造提供决策依据,助力排水系统提质增效^[1]。

在排水系统中布设在线监测设备是获取地下管井实时水位、识别高水位风险的有效手段。长期在线监测条件下,由于管井内部环境恶劣,监测设备易腐蚀损坏,且需定期更换电池及零件,后期运维成本高;短期监测能够辅助问题分析,但不能满足在线实时监测及预警需求^[2]。针对上述问题,笔者考虑降雨因素,收集短期窨井在线水位监测数据及雨量监测数据,应用支持向量机(SVM)回归算法,构建城市排水系统窨井水位预测模型;基于该预测模型,根据地面降雨监测数据实时预测地下窨井水位,识别管井高水位风险运行状态,以期为排水系统运维和监测提供低成本解决方案。

1 支持向量机回归

近年来,机器学习方法在城市排水系统中的应用越来越广泛^[3-4]。支持向量机是在统计学习理论基础上发展起来的一种新的、通用的机器学习方法,能够解决小样本、非线性和高维数等实际难题,可以使在小样本下建立的分类器具有很强的推广能力。为支持回归问题,Drucker等将其拓展到支持向量机回归算法(SVMR),管井水位预测属于非线性回归预测问题,因此,选择SVMR进行模型构建^[5]。

SVMR的基本思想是通过一个非线性映射 Φ ,将数据 x_i 映射到某一高维空间 F 中进行线性回归。给定数据集 $\{x_i, y_i\}, i = 1, 2, \dots, l, x_i \in R^n$ 为 n 维的输入向量, $y_i \in R$ 为输出向量,构建SVMR的决策函数。为保证决策函数平坦,假设所有训练数据都可以在一定精度下用线性函数拟合,考虑到允许拟合误差情况,引入松弛变量,将寻优问题表示为凸优化问题,采用Lagrange乘子得到原始问题的对偶优化问题,求解该问题可获得SVM的回归函数。

2 案例分析

2.1 研究区域概况

研究区域选择南方某市某河涌中游区域,该区域污水汇入沿河涌敷设的一条污水主干管,主干管下游连接污水处理厂。由于管网破损、雨污分流不彻底等因素,雨季条件下存在雨水入流入渗问题,易导致主干管满管运行、下游污水厂进水量增大、

进厂污染物浓度偏低等问题。为缓解污水厂压力、降低主干管淤积风险,当管道水位达到一定高度时,则实行人工调度措施管理。为实时掌握主干管水位运行情况,辅助运行调度,主干管沿途窨井内部布设水位监测设备用于监测管网水位,并在区域布设多个雨量站点用于监测雨情。但由于主干管来水环境复杂,窨井监测设备经常受垃圾遮挡、酸性液体腐蚀等影响,需定期维护,增加了运维成本。

研究区域概化见图1。主干管收集周边排水单元产生的污水;从上游到下游,依次布设水位站点1~4用于监测窨井水位数据;区域内布设雨量站点1和2用于监测降雨量;河涌内布设河道水位站点1用于监测河道实时水位。水位站点及雨量站点的测量频率均为5 min。该研究选取水位站点3作为窨井水位建模站点(下称“站3”),雨量站点1作为站3的主关联雨量站,其他站点数据用于建模辅助。



图1 研究区域概化示意

Fig.1 Generalization of the study area

2.2 水文特征分析

根据站3周边环境的初步分析,污水系统外水来源主要包括地下水入渗、降雨入流入渗、河湖水倒灌等。通过对站3与河道水位站点1常水位分析,河涌水位远低于排水管网敷设高度,可排除河

涌水位倒灌影响。为提高SVMR算法的学习效率,保证城市排水管网井口水位预测模型的建设质量,通过对站3旱季、雨季水文特征的进一步分析,提取水位影响因子,为特征样本提取提供依据。

2.2.1 旱季水位规律分析

选取典型旱季天气进行旱季水位曲线规律分析,旱季天气选取标准:①日累计降雨量在5 mm以下;②旱季时间跨度不宜少于连续7 d;③降雨结束后至少48 h后,以排除前一场降雨的影响。

对选取的旱季数据中的无效数据进行人工剔除,以确保数据精度,剔除标准:①为确保雨量站点1监测数据的可靠性,对雨量站点1与站点2日累计降雨量进行校验,若两者之间的偏差大于5 mm,则剔除该日旱季数据;②为确保站3监测数据的可靠性,剔除数据中存在数据缺失、零值情况。

采用以上标准,选取7个工作日、6个非工作日旱季水位监测数据集,并分别进行叠加,绘制旱季水位变化曲线,如图2所示。可以看出,站3水位高峰出现在每日凌晨01:00左右,水位低谷出现在每日07:00左右;随着每日时间的推移,整体变化规律一致性较强,不同日同一时刻水位波动差异较小,呈现以日为单位的周期变化规律,满足一般旱季规律。分析表明,站点来水环境相对稳定,存在突发异常排放的可能性较小。

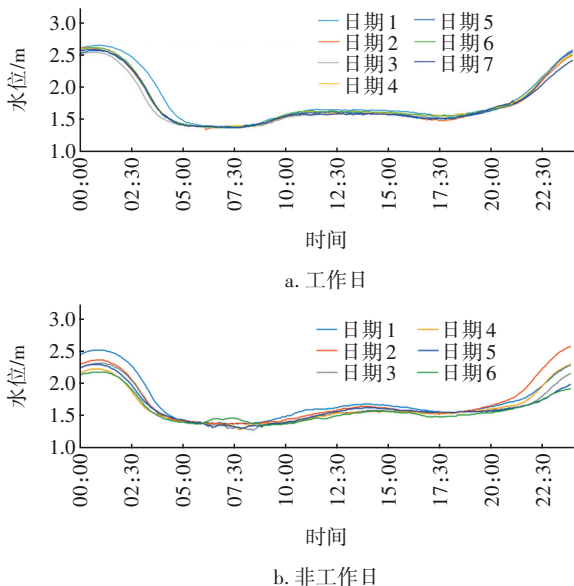


图2 工作日和非工作日旱季水位监测数据叠加曲线

Fig.2 Water level monitoring data superimposed curve of dry season on weekdays and non-working days

2.2.2 雨季降雨入流入渗分析

为分析站3水位受降雨入流入渗的影响,选取12月和5月的监测数据,分别对雨量站点1的降雨数据及站3水位监测数据进行叠加分析,结果如图3所示。可以看出,旱季流呈周期规律性;降雨情况下,站3受降雨影响水位明显上涨,水位与降雨量呈正相关;降雨结束后,水位慢慢恢复到正常水平。这表明,站3为典型的受降雨入流入渗影响的站点,上游排水管网可能存在破损、雨污分流不彻底等问题。

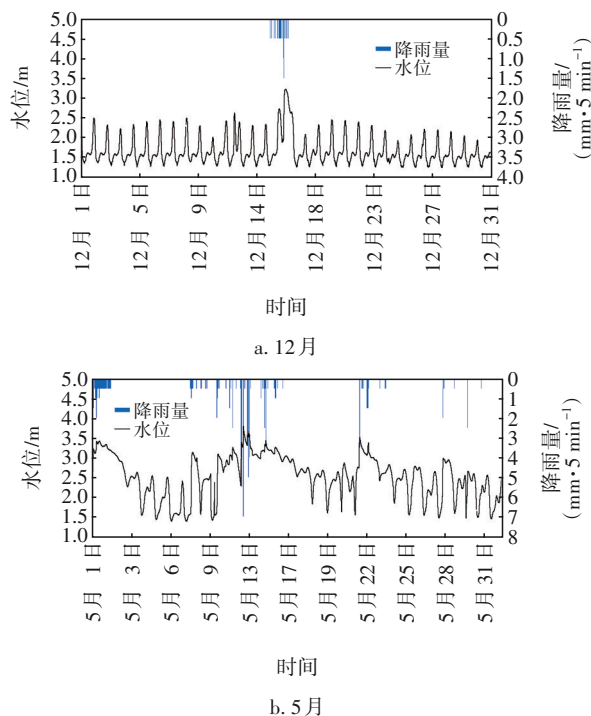


图3 12月和5月的降雨与水位叠加过程线

Fig.3 Rainfall and water level superposition process line in December and May

综上所述,站3的水位时序过程线与日时刻、降雨量均有极大相关性。为了能够准确描述站3的水位变化规律,充分表达地表径流、土壤渗流、管网汇流对水位曲线变化的影响,选取日小时数及雨量站点1的5 min、10 min、15 min、30 min、1 h、2 h、3 h、4 h、6 h、12 h、24 h、48 h累计降雨量作为特征因子。

2.3 模型构建

2.3.1 样本制作

考虑兼顾雨季及旱季水位特征,选取2022年4月—7月以及11月—12月等6个月的监测数据制作样本集。对数据进行预处理,剔除存在缺失、零值

等异常值的数据,具体原则如下:①剔除雨量站点1日监测数据缺失率大于5%的天数;②剔除水位站3日监测数据缺失率或零值大于5%的天数;③剔除雨量站点1与站点2日累计降雨量偏差大于10 mm的天数。最后得到48 426组样本数据。

从整体样本数据中选择6 d典型旱季及雨季天气的数据作为测试数据,包括:2022年11月11日旱季工作日;2022年11月13日旱季非工作日;2022年11月7日,累计降雨量为9.5 mm的小雨;2022年6月1日,累计降雨量为23.5 mm的中雨;2022年4月24日,累计降雨量为36 mm的大雨;2022年5月1日,累计降雨量为69 mm的暴雨。

2.3.2 参数优化

选择径向基函数作为SVMR核函数。SVMR性能与超参数的选取有很大关系,研究表明,超参数主要包括核参数 σ 、惩罚参数 C 、损失参数 ε 。为选取合适的超参数,防止模型过拟合及欠拟合,充分评价模型超参数 $[\sigma, C, \varepsilon]$ 组合的优劣程度,使用遗传算法优化参数选择,采用 K 折交叉方法及均方根误差(RMSE)衡量指标验证模型优劣程度,具体步骤如下^[6]:①将实际问题数值化,转化成SVMR可以处理的数据格式;②为消除样本在数值上相差较大的影响,提高模型的精度,对样本进行归一化处理;③根据数据特征,预设超参数的取值范围,其中 σ 取值范围为0~5, C 取值范围为0~10, ε 取值范围为0~0.1,设置种群规模为50,最大迭代次数为30次;④采用10折交叉验证方法评估参数组合优劣;⑤利用遗传算法对超参数进行优化,求RMSE最小化问题。通过上述方法,最终得到最优参数 σ 为2.561、 C 为0.898、 ε 为0.076 8。

2.3.3 模型测试

为验证模型可靠性,利用最优参数 $[\sigma, C, \varepsilon]$ 建立SVMR预测模型,分别利用6组测试数据对窖井水位进行预测,计算预测值与实测值的RMSE和决定系数 R^2 ,对模型进行评价。结果表明,旱季非工作日、旱季工作日、小雨、中雨、大雨、暴雨测试集的RMSE分别为0.053 7、0.045 0、0.057 0、0.046 0、0.045 8、0.051 3, R^2 分别为0.845 3、0.886 5、0.732 2、0.969 5、0.544 2、0.752 8。

预测数据与实测数据曲线如图4所示。

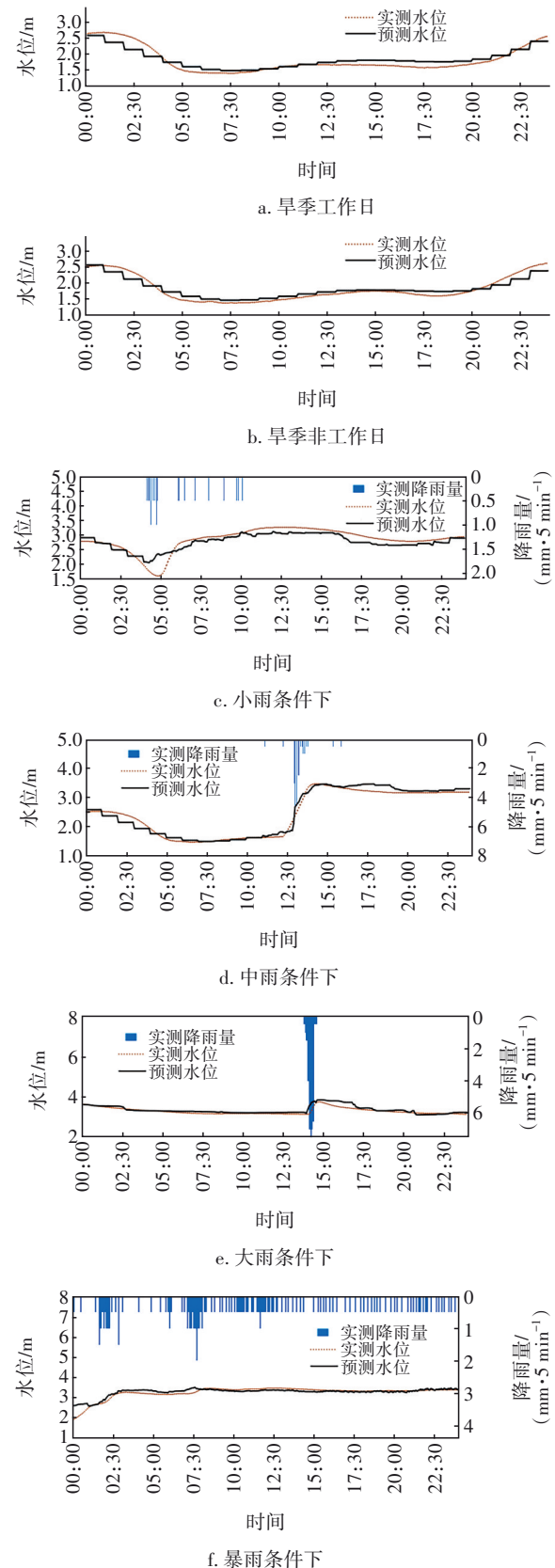


图4 实测与预测水位对比

Fig.4 Comparison of measured and predicted water levels

通过分析模拟结果可知,模型对6组典型旱季及雨季数据模拟的RMSE均小于0.06、决定系数 R^2 均大于0.5,预测值与实测值的差异较小,拟合效果较好,精度能够满足一般基于管网水位预测的业务应用要求。

对比旱季与雨季的预测结果可知,雨季预测效果稳定性相对较差,主要原因是:①雨季降雨雨型复杂、产汇流环境不确定因素较多,导致管网液位特征不规律,加上雨季样本占整体样本量的比例较小,雨季模型训练欠佳。可通过增加雨季样本量进一步提高模型对雨季预测的稳定性。②该研究使用单一雨量计数据制作样本进行模型训练,而污水管网上游汇水区实际降雨一般存在空间分布差异特征,模型输入降雨量与实际降雨量不一致,导致雨季模型预测结果变差。可通过管网上游汇水区布设多个雨量站点,再根据每个站点降雨量插值模拟空间分布差异的降雨特征解决该问题。

3 结论

① 窨井水位站点3旱季水位曲线整体变化规律一致性较强,不同日同一时刻水位波动差异较小,呈现以日为单位的周期变化规律,来水环境相对稳定,异常排放的可能性较小;雨季的水位受降雨影响较大,与降雨量呈正相关,上游排水管网可能存在管道破损、雨污分流不彻底等问题。

② 选取日时刻及不同时段累计降雨量作为建模特征因子、窨井水位作为建模目标变量,对模型进行训练验证、测试,测试结果与实际值拟合程度较高,模型预测可靠性良好。基于水位预测模型,在不布设管井水位传感器的条件下,通过地面降雨监测数据,能够预测地下窨井实时水位,与长期布设窨井水位物联网监测设备相比,具有部署方便、运维成本低的优点,对于城市排水日常调度、内涝应急决策等具有较高的使用价值。

③ 与长期实时监测方法相比,该方法也存在一定的不足。由于模型建立需依赖流域历史水文样本数据,若流域发生城市更新、管网改造等对流域产汇流环境产生影响的活动,已建立的模型模拟结果可能与实际监测数据具有较大的偏差,模型适用性变差。为提高模型的可靠性,需重新收集流域样本数据、分析特征因子,并对模型进行训练及验证。

参考文献:

- [1] 张旭东,马振华,王海玲,等.基于排水系统提质增效的集约治滇探索与实践[J].中国给水排水,2022,38(4):125-132.
ZHANG Xudong, MA Zhenhua, WANG Hailing, *et al.* Exploration and practice of integrative management of Dianchi Lake based on the quality and efficiency improvement of drainage system [J]. China Water & Wastewater, 2022, 38(4): 125-132 (in Chinese).
- [2] 赵冬泉,王浩正,陈吉宁,等.监测技术在排水管网运行管理中的应用及分析[J].中国给水排水,2012,28(8):11-14.
ZHAO Dongquan, WANG Haozheng, CHEN Jining, *et al.* Application and analysis of monitoring technology in operation management of urban drainage network [J]. China Water & Wastewater, 2012, 28(8): 11-14 (in Chinese).
- [3] YAN J, JIN J, CHEN F, *et al.* Urban flash flood forecast using support vector machine and numerical simulation [J]. Journal of Hydroinformatics, 2018, 20(1/2): 221-231.
- [4] 黄华兵,王先伟,柳林.城市暴雨内涝综述:特征、机理、数据与方法[J].地理科学进展,2021,40(6):1048-1059.
HUANG Huabing, WANG Xianwei, LIU Lin. A review on urban pluvial floods: characteristics, mechanisms, data, and research methods [J]. Progress in Geography, 2021, 40(6): 1048-1059 (in Chinese).
- [5] 李航.统计学习方法[M].北京:清华大学出版社,2012:95-134.
LI Hang. Statistical Learning Methods [M]. Beijing: Tsinghua University Press, 2012: 95-134 (in Chinese).
- [6] 杜长海.基于磷虾群算法的SVR参数选取方法及其应用[J].自动化技术与应用,2016,35(5):10-14,19.
DU Changhai. Parameters selection method for support vector regression based on Krill Herd algorithm and its application [J]. Techniques of Automation and Applications, 2016, 35(5): 10-14, 19 (in Chinese).

作者简介:赖泽辉(1989—),男,湖南株洲人,硕士,工程师,研究方向为水务信息化及模型技术。

E-mail: laizh@augurit.com

收稿日期:2024-02-10

修回日期:2024-03-05

(编辑:刘贵春)