

吴文鑫, 高杰, 于会来, 等. 融合凝聚层次聚类 and 互相关函数的城市排水管网监测点优化布置研究[J]. 水利水电技术(中英文), 2025, 56(6): 101-110. DOI: 10.13928/j.cnki.wrahe.2025.06.009

WU Wenxin, GAO Jie, YU Huilai, et al. Optimizing the layout of urban drainage pipeline monitoring points using agglomerative hierarchical clustering and cross-correlation functions[J]. Water Resources and Hydropower Engineering, 2025, 56(6): 101-110. DOI: 10.13928/j.cnki.wrahe.2025.06.009

融合凝聚层次聚类和互相关函数的城市排水管网监测点优化布置研究

吴文鑫¹, 高杰², 于会来¹, 林洁¹, 蒋俊豪¹, 李传奇^{2,3}

(1. 深圳市水务工程检测有限公司, 广东深圳 518110; 2. 山东大学 土建与水利学院, 山东济南 250061; 3. 山东大学 深圳研究院, 广东深圳 518057)

摘要:【目的】随着城市排水管网在洪水管理和水污染控制中的重要性日益凸显, 对其监测网络的优化布置提出了迫切需求。【方法】提出一种高效、准确的城市排水管网监测点优化布置方法, 以监测和识别管网污水非法排放。采用暴雨管理模型(SWMM)模拟城市排水管网中污染物的运移过程, 获取污染物时间序列数据; 利用凝聚层次聚类算法对数据进行分类处理, 确定排水管网中最优监测点的数量; 通过互相关函数评估节点间的相关性, 选择互相关系数最大的节点作为监测点的位置, 以确保全面反映簇内的污染情况。【结果】使用SWMM手册中的排水管网案例验证了所提方法的有效性, 选择部署3个监测点为最优选择。监测点布置效果评价表明, 监测点数为3时, 监测可靠度达到91.86%, 监测平均响应时间为3.26 min。凝聚层次聚类在监测点布置效果上优于K-means算法, 并且在全面覆盖排水管网上游、中游和下游方面表现更为出色。【结论】提出的城市排水管网监测点布局优化方法为城市排水管网监测与管理提供了新的技术手段, 为相关领域的研究提供了新思路。

关键词: 城市排水管网; 监测点优化布置; 凝聚层次聚类算法; 互相关函数; SWMM模型

DOI: 10.13928/j.cnki.wrahe.2025.06.009

开放科学(资源服务)标志码(OSID):

中图分类号: TU992

文献标志码: A

文章编号: 1000-0860(2025)06-0101-10



Optimizing the layout of urban drainage pipeline monitoring points using agglomerative hierarchical clustering and cross-correlation functions

WU Wenxin¹, GAO Jie², YU Huilai¹, LIN Jie¹, JIANG Junhao¹, LI Chuanqi^{2,3}

(1. Shenzhen Water Engineering Testing Co., Ltd., Shenzhen 518110, Guangdong, China; 2. School of Civil Engineering, Shandong University, Jinan 250061, Shandong, China; 3. Shenzhen Research Institute of Shandong University, Shenzhen 518057, Guangdong, China)

收稿日期: 2024-03-07; 修回日期: 2024-07-03; 录用日期: 2024-07-09; 网络出版日期: 2024-08-14

基金项目: 深圳市可持续发展科技专项项目(KCXFZ20201221173407021); 国家自然科学基金项目(52109025); 山东省自然科学基金项目(ZR2021ME030)

作者简介: 吴文鑫(1982—), 男, 高级工程师, 学士, 主要从事城市水环境治理研究。E-mail: 429701634@qq.com

通信作者: 李传奇(1963—), 男, 教授, 博士, 主要从事防洪减灾、智慧水利研究。E-mail: lichuanqi@sdu.edu.cn

©Editorial Department of Water Resources and Hydropower Engineering. This is an open access article under the CC BY-NC-ND license.

Abstract: [Objective] With the increasing importance of urban drainage pipelines in flood management and water pollution control, there is an urgent need for optimized monitoring networks. [Methods] An efficient and accurate method for optimizing the layout of monitoring points in urban drainage pipelines to detect and identify illegal sewage discharge was proposed. A Storm Water Management Model (SWMM) was utilized to simulate pollutant transport within urban drainage pipelines and generate time-series data. An agglomerative hierarchical clustering algorithm was applied to classify the data and determine the optimal number of monitoring points in the drainage network. Node correlation was assessed through cross-correlation functions, selecting nodes with the highest coefficients to ensure comprehensive pollution conditions within the clusters. [Results] The proposed method was validated using a drainage network case from the SWMM manual, and it was determined that deploying three monitoring points is the optimal choice. The evaluation revealed a reliability of 91.86% and an average response time of 3.26 minutes at three monitoring points. Agglomerative hierarchical clustering outperformed the K-means algorithm in terms of monitoring point layout effectiveness, especially in covering the upstream, midstream, and downstream sections of the drainage network. [Conclusion] This method offers a new technical means for optimizing the layout of monitoring points in urban drainage pipelines, enhancing urban drainage monitoring and management and offering new perspectives for related research areas.

Keywords: urban drainage pipelines; monitoring point optimization; agglomerative hierarchical clustering algorithm; cross-correlation function; SWMM model

0 引言

城市排水管网是城市基础设施的关键组成部分,对减轻洪水威胁和控制水污染具有重要作用。随着气候变化引发的极端天气事件增多,迫切需要采用先进的监测和管理技术来增强城市排水管网的适应性和韧性。构建高效、准确的监测网络,对于及时应对极端天气、降低洪水风险及控制水污染至关重要^[1]。城市排水管网监测主要包括三个方面:一是水位和流量监测,用于提供排水能力和溢流风险的信息,以减少城市洪水带来的经济损失和公共安全威胁^[2-3];二是常规监测,通过选择具有代表性的监测节点,持续监测这些节点的关键参数(如流量、水位等),以全面了解排水管网的运行状态;通过对长期监测数据的分析,及时发现管网混接、入流渗入等异常情况并采取相应修复措施,以确保排水管网的正常运行并提高其效率^[4-5];三是水质监测,通过监测城市排水管网的的水质参数,及时发现水质恶化趋势和潜在污染源,采取有效措施减少污染对环境和公众健康的影响^[6-7]。尽管理论上监测点越多越好,但在实际操作中,过多的监测点会增加成本并可能降低数据传输效率。因此,合理优化监测点的布置,以提升监测效率和数据质量,成为当前重要的研究课题。

在河流^[8-9]和城市供水管网^[10-11]的监测点优化方面已有广泛的研究成果,但对于城市排水管网监测点优化研究还相对较少^[12-13]。在高监测成本与全面监测之间需要找到平衡。为解决这一问题,提出了多种优化方法。例如, FATTORUSO 等^[14]提出了一种监测位置优化方法,用于城市内涝预警; BANIK 等^[15]

提出了针对排水系统的多目标和单目标优化方法,以优化监测点的位置,从而有效监测水质并探测非法排放行为; YAZDI^[16]结合熵理论和差分进化算法设计了城市排水系统水质监测网络,以最小化监测点数量的同时最大化信息覆盖; SAMBITO 等^[17]利用贝叶斯决策网络(BDN)优化了水质传感器的位置,结合网络拓扑、流量等信息,以便及时发现排水系统的非法排放; SIMONE 等^[18]提出了一种确定传感器最佳位置的全局优化方法,增加排水管网污染排放探测的可靠性。

近年来,数据驱动的机器学习技术,尤其是聚类分析,在处理大量在线监测数据方面表现出独特优势^[19]。聚类分析作为一种无监督学习技术,主要通过分析样本间的相似度来对数据进行分类^[20]。例如, LIU 等^[21]通过空间聚类方法将城市排水系统划分为多个区域,提高了洪水监测与预报的准确性; LI^[22]提出将无监督机器学习(聚类分析)与方差分析(ANOVA)相结合以优化在不确定降雨条件下的监测点布置。赵文涓等^[23]基于改进的 K-means 算法,优化了排水管网监测点位。

目前,城市排水管网的监测点布置常常依赖于主观经验,这增加了监测布局的随机性^[13]。水质监测点布置通常基于关键设施的临近性、可访问性和空间密度等因素,而非优化技术或聚类技术。水质异常监测的准确性与传感器的数量和位置密切相关,增加传感器的数量有助于及时发现污染事件并准确定位^[24-25]。如何在有限的监测点中进行最优布置,以及及时发现违法排放或意外泄漏成为水质监测的核心问

题。为此,本研究提出了一种结合凝聚层次聚类算法和互相关函数的监测点优化方法。首先,采用SWMM模型模拟城市排水管网污染物运移过程,生成水质特征数据;其次,通过凝聚层次聚类算法处理这些数据,识别出关键监测点;最后,应用互相关函数分析这些点间的相关性,优化监测点的布置。案例研究表明,该方法显著提升了监测网络的准确性和响应速度,扩大了监测网络的覆盖范围,提高了数据质量,并有效降低了监测成本。本研究展示了数据驱动方法和机器学习技术在环境监测中的应用潜力,为未来城市排水管理提供了重要的技术支持。

1 研究方法

本节介绍了SWMM模型用于生成污染物的时间序列数据;采用凝聚层次聚类算法对模拟得到的时间序列数据进行分类,以识别具有相似水文响应特征的节点;利用互相关函数分析不同节点间的相关性,从而优化监测点的选择和布局。

1.1 水力模拟模型

为了准确模拟城市排水系统中污染物的对流扩散过程,本研究采用了美国环保署(EPA)开发的暴雨管理模型(SWMM)。SWMM是一个综合性的模型,它集成了地表径流(Runoff)、输送(Transport)、扩充输送(Extran)和储存/处理(Storage/Treatment)四个主要计算模块^[26]。这些模块共同支持城市排水和雨水管理的模拟需求,使SWMM不仅可以精确模拟地表径流过程,还可以处理管网流动、污染物输送以及低影响开发(LID)措施的效果,展现了其多功能性和广泛的适用范围^[27]。

SWMM模型主要利用城市的污水和降雨数据作为输入,这些数据分别输入到输送模块和径流模块中。排水管网的汇流演算主要通过输送模块和扩展输送模块实现,采用圣维南方程组来描述管网中水流的运动状态,支持恒定流法、运动波法和动力波法等多种水动力学求解方法。此外,SWMM的水质模块通过求解一维污染物对流扩散方程,能够动态模拟保守物质及非保守物质在排水管网中的迁移与转化过程。由于其开源性质,SWMM已在全球城市排水系统管理中得到了广泛的应用,并有效地支持了监测点的布置优化和污染传播路径的识别。

1.2 凝聚层次聚类算法

1.2.1 算法概述

聚类分析是一种探索性数据分析技术,目标是将数据集划分为由相似对象构成的多个“簇”^[28]。层次

聚类方法,作为聚类分析的一种重要方法,按照其构建簇的方向可以分为两类:“自顶向下”的分裂层次聚类和“自底向上”的凝聚层次聚类(AHC)。AHC以其在各数据分析场景中的广泛应用,成为一种重要的无监督学习算法。通过计算数据点之间的相似度,逐步构建出一种层次化的聚类结构。具体来说,该算法开始时把每个数据点作为一个独立的簇,然后根据一定的距离度量(如欧氏距离)逐步合并最相似的簇,直至形成单一簇或达到其他预设条件。AHC的优势在于不仅能提供单一的聚类结果,还能通过树状图(dendrogram)揭示数据之间复杂的层次结构,为研究人员提供了丰富的数据结构信息。AHC算法的实施步骤如下。

步骤1:初始化,将每个观测数据视为一个独立的聚类;

步骤2:计算各聚类之间的距离;

步骤3:选择距离最近的两个聚类合并成一个新聚类,并更新距离矩阵以反映这一合并过程;

步骤4:重复第2和第3步,直到所有数据点归入一个聚类或达到其他终止条件。

在相似度计算上,除了常用的欧式距离,AHC算法支持多样化的距离度量方法,如曼哈顿距离和最大距离,提高了其适应性。此外,AHC算法的合并策略也多样化,包括单链接(最短距离)、全链接(最长距离)、平均链接、质心链接和Ward链接等,每种方法都适用于特定的应用场景,并有其独到的优势。本研究采用Ward链接作为聚类合并规则,该规则通过每次合并过程中最小化总方差的增加,保证簇内部的一致性。

在排水系统监测点布置优化方面,AHC算法通过分析污染物浓度的时序数据,能够有效识别污染物传输路径上的关键监测点。这一能力不仅帮助确定监测网络中的重要节点,指导传感器的优化布置,保障关键区域的有效监控,也通过减少非关键监测点的数量,助力资源的优化分配,提升监测网络的整体运行效率。

1.2.2 算法构建

(1)特征矩阵构建。构建特征矩阵是AHC算法实施的关键步骤。本研究利用SWMM模拟城市排水管网中污染物的传输过程,生成了包括多个时间步的各节点污染物浓度数据。这些数据用于构建特征矩阵,为AHC算法的实施提供结构化的输入。

特征矩阵的构建步骤为:首先,通过SWMM模拟获得各节点在不同时间步长下的污染物浓度数

据;其次,对获取的数据进行标准化处理(归一化);最后,构建特征矩阵 Y_t , 该矩阵包括多个节点在连续时间步上的污染物浓度数据。矩阵的每行代表一个节点, 每列代表不同的时间步。构建的特征矩阵 Y_t 为

$$Y_t = \begin{bmatrix} x_{(0,0)} & x_{(0,1)} & \cdots & x_{(0,m)} \\ x_{(1,0)} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{(n,0)} & \cdots & \cdots & x_{(n,m)} \end{bmatrix} \quad (1)$$

式中, Y_t 为 t 时刻所有节点的污染物浓度组成的输入矩阵; n 为排水管网中节点的总数; m 为模拟期间考虑的时间步长总数; $x_{(n,m)}$ 为第 n 个节点在第 m 时间步的污染浓度。

这种数据结构不仅方便进行污染浓度的时间序列分析, 而且为基于数据点间相似性的聚类算法提供了理想的输入格式。通过聚类分析, 能够有效识别出具有相似污染动态特性的节点群体, 从而为监测网络的优化布置提供坚实的科学基础。

(2) 基于 Ward 法的聚类分析。Ward 方法通过最小化每次合并簇后的误差平方和 (Sum of Squares due to Errors, SSE) 来优化聚类过程^[29]。相比其他方法, Ward 方法的核心优势在于其能够在增加簇类间差异的同时, 有效降低簇类内部的差异, 使之成为监控城市排水系统中污染分布模式的理想选择。在每一次迭代中, Ward 方法选择两个最相似的聚类进行合并, 确保此操作导致的总方差增加最小。通过不断地合并相似的聚类, 最终形成一个层次化的聚类结构。SSE 的计算公式为

$$SSE = \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2)$$

式中, x_i 为簇内第 i 个样本的值; \bar{x} 为簇内样本的平均值; n 为簇内样本总数。

Ward 距离是衡量不同簇间差异的关键指标, 基于两簇合并后 SSE 的增量计算。Ward 距离的计算公式为

$$d_{\text{ward}}(C_i - C_j) = SSE(C_i \cup C_j) - SSE(C_i) - SSE(C_j) \quad (3)$$

式中, C_i 、 C_j 为待合并的簇; $d_{\text{ward}}(C_i - C_j)$ 为簇 C_i 与 C_j 之间的 Ward 距离; $SSE(C_i \cup C_j)$ 、 $SSE(C_i)$ 、 $SSE(C_j)$ 为合并后的簇与合并前各自簇的误差平方和。

通过最小化 Ward 距离, 确保每一步聚类过程最大限度地保持簇间的相似性。

(3) 聚类性能评估。评估聚类效果是聚类分析的

重要环节。本研究采用轮廓系数 (Silhouette Coefficient, SC) 和戴维森堡丁指数 (Davies Bouldin Index, DB) 两种指标对聚类结果进行了全面评估。SC 衡量的是簇内的紧密性与簇之间的分离度^[30], 其取值范围为 $-1 \sim 1$ 。当 SC 值接近 1 时, 说明簇内样本具有较高的相似性; 而值接近 -1 时, 则可能暗示有样本被错误地归类到其他簇中。SC 为每个样本 i 的计算公式为

$$s(i) = \frac{h(i) - a(i)}{\max\{a(i), b(i)\}} \quad (4)$$

式中, $a(i)$ 为样本 i 与其所属簇内其他样本的平均距离, 反映了簇内的紧密性; $b(i)$ 为样本 i 到最近一个其他簇中所有样本的平均距离, 表示了簇间的分离度。

一个较高的 SC 意味着簇与簇之间的界限清晰, 对于确定监测点的准确位置非常有帮助。

DB 是通过簇内紧密性与簇间分割性的比较来评估聚类的质量^[31]。计算 DB 时, 对于每个簇, 找出与其最相似的另一簇, 通过比较它们内部的平均距离与簇中心点之间的距离来进行评分。DB 值越小, 意味着簇内部紧密度更高, 而不同簇之间的区别更加明显, 从而指示出更好的聚类效果。DB 的计算公式为

$$DB = \frac{1}{N} \sum_{i=1}^N \max_{j \neq i} \left(\frac{W_i + W_j}{C_{ij}} \right) \quad (5)$$

式中, N 为聚类簇数; W_i 为簇 C_i 内所有样本到该簇中心的平均距离; C_{ij} 为簇 C_i 与簇 C_j 中心之间的距离。一个较小的 DB 值意味着簇内部结构紧密而与其他簇之间界限分明, 指示了较好的聚类性能。

通过综合应用 SC 和 DB 这两个指标, 能够全面评价不同聚类方案的性能, 从而在设计监测点布置策略时, 达到聚类质量和监测成本之间的最优平衡。

1.3 互相关函数

在通过 AHC 明确了监测点的数量之后, 互相关函数可以帮助精确确定监测点的具体位置。互相关函数 (Cross-correlation Function), 作为一种用于评估两个时间序列在不同时间滞后水平上相互关系的统计方法, 为城市排水系统监测点布局的优化提供了重要的支持。通过分析时间序列间的相关性, 互相关函数揭示了监测点之间的相互作用模式, 特别是污染物浓度变化这一关键环节。

具体而言, 如果在某一特定的时间滞后水平 τ 下, 两个监测点之间的互相关系数显著高, 这表明一个监测点处的污染物浓度变化能够预示另一个监测点处的污染物浓度变化。这种显著的相关性反映了污染物在水体中的传播方向与速度, 从而为监测点的最优

位置布局提供了理论依据。因此, 监测点应该被策略性地布置在污染物传播的关键路径上, 以便能够更有效地捕捉污染事件的发展过程, 优化整个监测网络的配置和性能。

设 X_t 和 Y_t 分别表示两个不同监测点或不同时间点上的污染物浓度的时间序列。在考虑到时间滞后 τ 时, 互相关函数 $\rho_{X, Y}(\tau)$ 可表示为^[32]

$$\rho_{X, Y}(\tau) = \frac{Y_{X, Y}(\tau)}{S_X S_Y} \quad (6)$$

$$Y_{X, Y}(\tau) = \frac{1}{n} \sqrt{\sum_{t=1}^{n-\tau} (X_t - \bar{X})(Y_{t+\tau} - \bar{Y})} \quad (7)$$

式中, $Y_{X, Y}(\tau)$ 为随时间滞后 τ 变化的互协方差函数; n 为样本量; \bar{X} 和 \bar{Y} 分别为时间序列 X 和 Y 的样本均值; S_X 和 S_Y 分别为序列 X 和 Y 的标准差。

通过计算不同时间滞后 τ 下的互相关值, 可以揭示监测点之间的动态相互影响关系。

2 案例研究

本研究选取了 SWMM 用户手册示例 8 中的一个典型排水管网系统作为研究对象^[27], 该系统覆盖面积为 0.12 km²。所构建的模型包括 7 个集水区、28 个节点、2 个排放口、1 个蓄水装置、29 条管道和 1 个泵站(见图 1)。排水系统设有两个排水出口: 污水处理厂(O2)和溢流口(O1)。所有的几何数据以及干旱天气下的入流量(表示为日均值)均引用自 SWMM 用户手册。

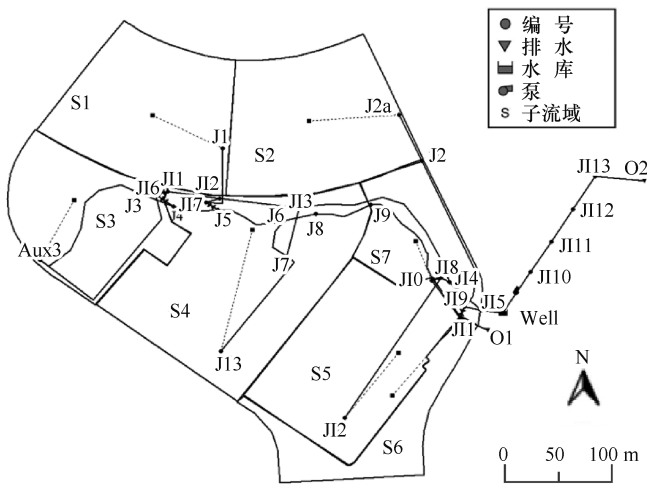


图 1 SWMM 示例 8 系统的结构

Fig. 1 Schematic diagram of SWMM Example 8

2.1 污染事件模拟

本研究利用 SWMM 软件的水质模块并结合 Python 语言, 通过随机生成大量污染事件来动态模拟排水管网中污染物的传播过程。通过提取必要的污染

物浓度时间序列数据, 为后续的聚类分析提供输入数据。

每个模拟的污染事件包括三个要素: 污染物的排放节点、污染物浓度以及排放持续时间。污染物浓度随机设定在 1~10 mg/L 之间, 持续时间则是基于污染物在排水系统中从一个最远点移动到另一个最远点所需的时间, 随机设定在 10~120 min 之间。此外, 假设水质监测可以测量到的最小浓度阈值为 0.001 mg/L。研究仅考虑旱季天气条件下的污染事件, 因为在这种情况下非法排放对排水系统的影响更为显著。为模拟管道水质变化, 选用了连续搅拌槽反应器(CSTR)模型。

考虑到实际应用场景, 假定污染物为保守物质, 不参与化学反应, 并以固定浓度进入排水系统。模拟时长定为 2 h, 以 10 s 为时间步长进行模拟, 报告的时间步长为 1 min, 共生成了 5 000 个污染事件的数据集。

为提高模拟数据的真实性和可靠性, 在数据集中引入了随机性因素, 以模拟水质监测中固有的不确定性噪声。该噪声反映了实际环境监测中可能出现的观测误差, 其公式为

$$\rho'(n, t) = \rho(n, t) \times U(1 - \delta, 1 + \delta) \quad (8)$$

式中, $\rho'(n, t)$ 为在 t 时刻第 n 个节点受噪声影响的污染物浓度; $\rho(n, t)$ 为无噪声的模拟值; u 为均匀分布; δ 设定为 0.05, 用以界定加入噪声的幅度。

在模拟数据基础上, 构建了一个特征矩阵, 为聚类分析提供基础数据。特征矩阵为二维数组, 每一行对应研究区域内的一个节点, 共有 28 个节点。每一列对应不同的时间点, 时间点的间隔为模拟报告的时间步长, 即 1 min。矩阵中的每个值代表对应节点在对应时间点的污染物浓度。

2.2 污染数据聚类

利用层次聚类算法对模拟生成的污染数据进行分析, 以识别具有相似污染特征的节点群体。选择 Ward 方法进行层次聚类, 通过逐渐合并距离最近的簇构建出树状图(见图 2)。在树状图中, 水平线的高度代表了相应节点合并为同一簇时的差异性大小, 高度较高意味着合并时的相似度越低, 导致形成更少的簇群体。通过在垂直轴上选择一个特定高度的截断阈值, 可以确定理想的簇数量来优化监测网络布局。

通过 0.85 的截断阈值, 将节点划分为三个主要簇: 第一簇包括 Aux3、J1、J2 等多个节点, 它们之间显示出显著的污染参数相似性; 第二簇由 J5、J6 等节点组成, 这些节点在污染特性上具有较高的一致

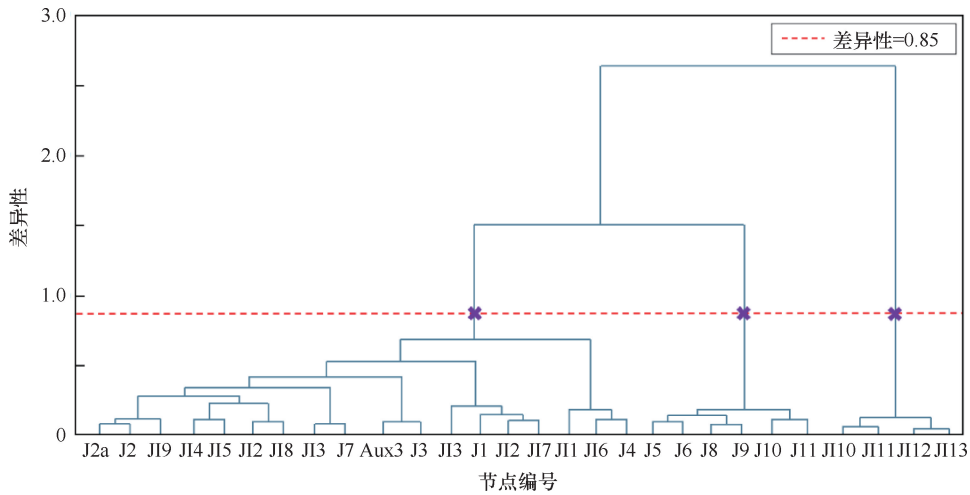


图2 聚类分析树状图

Fig. 2 Dendrogram of cluster analysis

性;第三簇则由 J110、J111 等节点组成。这些聚类结果不仅揭示了各节点之间在污染特征上的相似性和差异性,而且还提供了有力的依据来优化监测网络,确保监测点能够有效地捕获关键的污染事件,及时反映污染模式的变化。

2.3 聚类算法效果评估

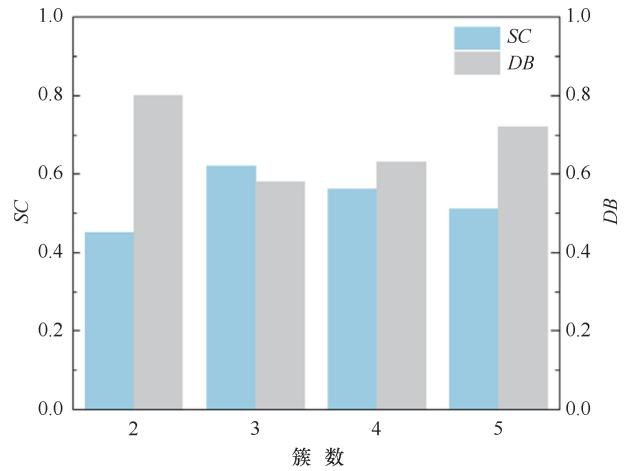
确定最优簇数是制定有效监测点布置策略的关键步骤。本研究评估了簇数 k 从 2 到 5 变化时对聚类质量的影响,通过多次迭代验证了算法的稳定性。聚类质量通过轮廓系数 (SC) 和戴维森堡丁 (DB) 进行评估,高 SC 值反映了簇类内部的高相似性,而较低的 DB 值则表示簇间清晰的分隔。

不同 k 值的聚类效果评估结果如图 3 所示。当簇数为 3 时,聚类效果达到最优,此时 SC 值为 0.62, DB 指数为 0.58。当簇数增至 4 或 5 时,观察到 SC 值有所下降, DB 指数略有上升,表明聚类质量未能达到簇数为 3 时的效果。不同簇数下排水系统节点的分类结果如图 4 所示,显示簇间区分度随簇数增加而提高,但簇内节点数减少。因此,选择簇数需在监测成本与聚类精度间寻找最佳平衡。

2.4 监测点位置确定

本研究通过层次聚类分析确定了排水管网系统中最优监测点的位置。在每个生成的簇内选择一个节点作为潜在的监测点,通过互相关函数评估节点间的相似性,选择互相关系数最大的节点作为监测点,以确保所选监测点能够全面反映簇内的污染情况。监测点的具体布置如图 5 所示。

本研究采用污染平均响应时间 (ART) 和监测可靠度 (R) 两个关键指标来评价监测点的布置效果。 ART

图3 不同 k 值的聚类效果评估Fig. 3 Evaluation of clustering performance for different k values

衡量从污染发生到被监测网络首次探测到所需的平均时间。较短的 ART 表示监测网络能够迅速响应污染事件。监测可靠度 R 反映了监测网络在多种情况下连续探测特定污染事件的概率,较高的 R 值表示监测系统稳定,且对污染事件的探测更可靠。

监测点数量对 ART 和 R 值的影响如表 1 所列。结果显示,监测点增多可以提高对污染的响应速度和监测的可靠性,但当监测点从 3 个增加到 5 个时,边际效益逐渐降低。监测点数量为 3 时, R 值达到 91.86%,能探测到大多数污染事件。根据聚类结果,这 3 个监测点分别布置在 J10、J13 和 J112 节点。因此,综合考虑成本和效益,选择部署 3 个监测点为最优选择。

进一步比较了凝聚层次聚类与 K-means 算法在

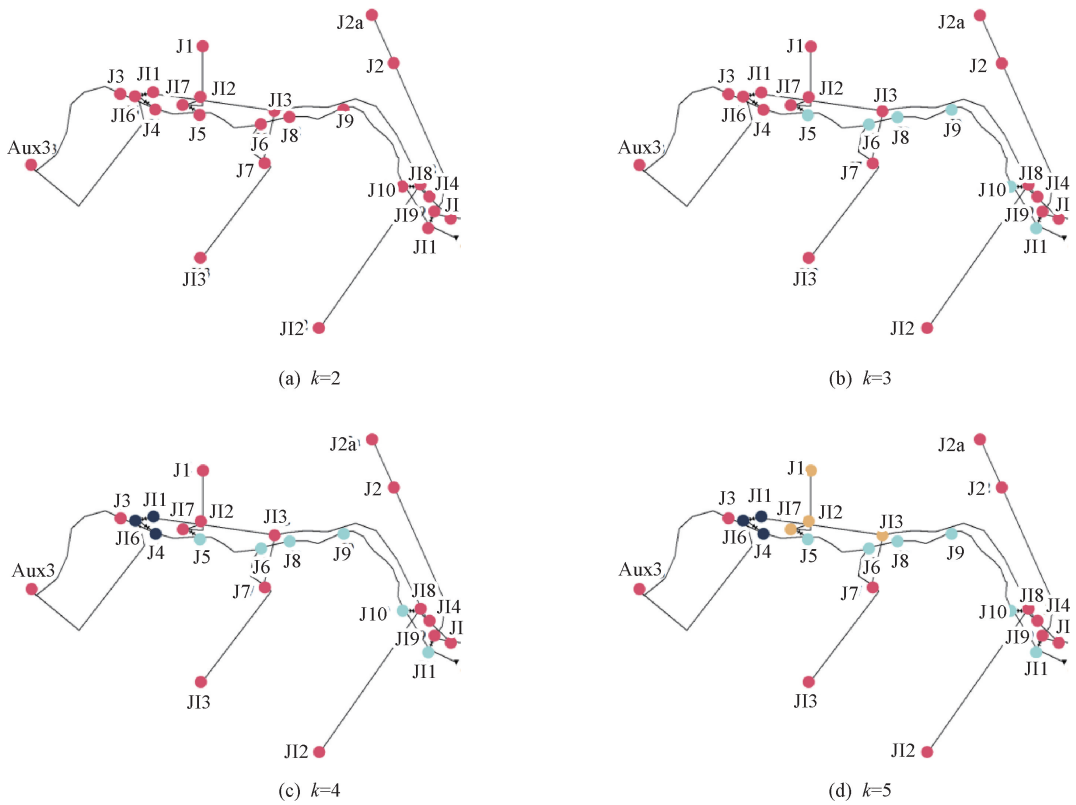


图 4 不同 k 值排水管网节点的聚类结果

Fig. 4 Clustering results of drainage network nodes for different k values

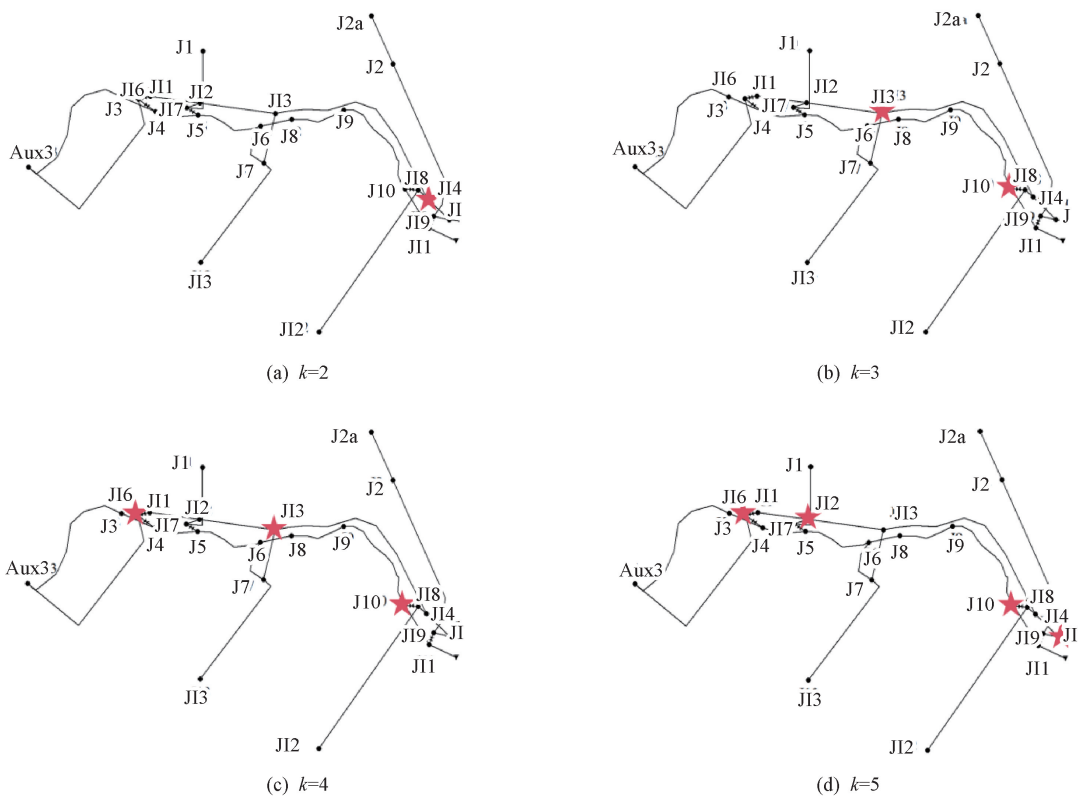


图 5 不同 k 值时监测点的布置方案

Fig. 5 Layout schemes of monitoring points for different k values

表 1 监测点布置效果评价

Table 1 Evaluation of monitoring point layout effectiveness

| 监测点数量/个 | 平均响应时间/min | 时间减少率/% | 可靠度/% | 可靠度增量/% |
|---------|------------|---------|-------|---------|
| 2 | 4.43 | — | 81.43 | — |
| 3 | 3.26 | 35.89 | 91.86 | 10.43 |
| 4 | 2.90 | 12.41 | 95.04 | 3.18 |
| 5 | 2.75 | 5.45 | 96.29 | 1.25 |

确定监测点位置时的性能, 如表 2 所列。凝聚层次聚类在平均响应时间 (3.26 min) 和可靠度 (91.86%) 上均优于 K-means 算法 (3.55 min 和 87.14%)。

表 2 聚类算法性能对比

Table 2 Performance comparison of clustering algorithms

| 聚类方法 | 监测点数量/个 | 平均响应时间/min | 可靠度/% |
|------------|---------|------------|-------|
| 凝聚层次聚类 | 3 | 3.26 | 91.86 |
| K-means 聚类 | 3 | 3.55 | 87.14 |

采用层次聚类分析 (AHC)、贝叶斯方法 (Bayesian)、以及多目标遗传算法 (NSGA-II) 获得的监测点布置方案的比较结果如图 6 所示。

3 结果讨论

与 BANIK 等^[15] 的 NSGA-II 算法和 SAMBITO 等^[18] 的贝叶斯方法 (Bayesian) 相比, 本研究的监测点布置策略实现了对排水系统上游、中游和下游的全面覆盖, 增强了整个系统的监控能力。相对于以往研究中主要关注下游区域的布置, 本方法显著提高了对突

发水污染事件的识别与响应效率。

层次聚类算法通过树状图对数据进行可视化, 特别适用于阐明排水管网中不同节点之间的层次关系, 这种可视化方式有助于直观理解数据结构。相较于需要详细定义问题的 NSGA-II 和贝叶斯方法, 层次聚类算法具有更好的决策支持能力和灵活性。NSGA-II 和贝叶斯方法通常需要较为丰富的先验知识, 这在一定程度上限制了它们的灵活性和适应性。通过结合层次聚类和交叉相关函数, 本方法在优化城市排水管网监测点布置方面展示了更高的适应性。采用 Ward 方法最小化了合并簇后的误差平方和, 确保每个簇内高度同质性同时与其他簇保持差异。通过应用交叉相关函数, 优化了监测点的布置, 确保每个监测点位置都能全面反映其所在簇的污染状况。

在本案例研究中, 本方法显著提升了监测网络的准确性和响应速度。优化后的监测点平均响应时间 (ART) 为 3.26 min, 相比于 K-means 算法的 3.55 min, 减少了污染事件的响应时间。此外, 本方法的监测可靠度 (R) 达到 91.86%, 高于 K-means 算法的 87.14%, 显示出更高的污染事件监测稳定性。

监测点优化结果在实际用于排水管网监测时需要综合考虑以下因素。

首先, 数据质量与可用性至关重要。尽管本方法在优化城市排水系统监测点布置方面表现出色, 但数据的缺失或质量不佳可能影响监测网络的准确性和响应速度。为应对此问题, 需进一步研究如何整合多源数据, 包括实时监测数据和历史数据, 以提高监测点布置的精确度和适应性。

其次, 经济成本的限制不容忽视。全面覆盖城市排水管网的监测点布置成本高昂, 可能超出预算限制。因此, 应在预算范围内优化监测点的选择, 以尽可能少的监测点覆盖更多的信息。

此外, 技术可行性和环境因素也是关键考虑因素。一些监测点可能因地理位置偏僻或环境复杂, 导致设备安装和维护困难。同时, 自然环境如地形和气候条件也可能影响监测设备的性能, 特别是在极端天气下。因此, 应选择适应不同环境条件且便于安

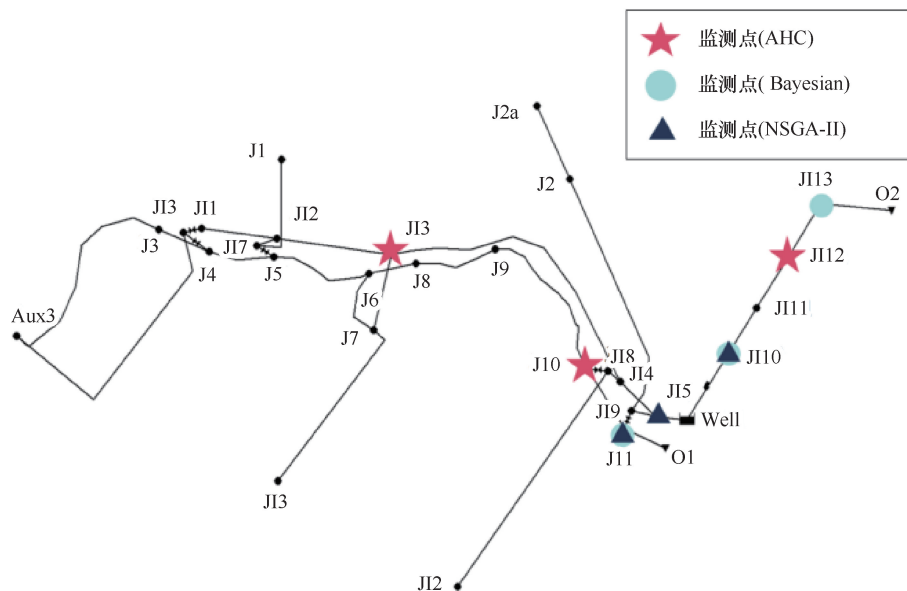


图 6 监测点布置方案对比

Fig. 6 Comparison of monitoring point layout schemes

装和维护的监测设备, 并采取必要的防护措施。

最后, 动态调整和优化机制至关重要。由于实时监测数据和环境条件的变化, 固定的监测点可能无法满足所有需求。因此, 应采用动态调整机制, 根据实时数据优化监测点位置, 保持监测网络的有效性和灵活性。此外, 需定期评估监测点布置效果, 进行必要的优化和调整。

通过上述措施, 可以更好地应对实际应用中的挑战, 确保监测点布局的科学性和合理性, 提高监测效果和管理水平。

4 结论

本研究针对城市排水系统监测点布置的优化问题, 提出了一种结合凝聚层次聚类算法和互相关函数的新方法。主要结论如下。

(1) 通过将凝聚层次聚类分析与互相关函数相结合, 优化了城市排水系统监测点布置。通过 SWMM 模型模拟城市排水管网中污染物的运移过程, 生成了具有代表性的污染物时间序列数据。凝聚层次聚类算法有效地对排水管网的节点进行了分类, 确定了最优的监测点数量, 并利用互相关函数分析节点间的关系, 明确了最优的监测点位置。

(2) 案例分析表明, 优化后的监测点布局能够全面监测排水管网的污染动态, 扩大了监测网络的覆盖范围和数据收集的质量。优化后的监测点平均响应时间(*ART*)为 3.26 min, 优于 K-means 算法的 3.55 min, 显示了更高的响应速度。同时, 监测可靠度(*R*)达到 91.86%, 高于 K-means 算法的 87.14%, 进一步证明了该方法的稳定性。

(3) 提出的监测点布局优化方案为城市排水系统的监测与管理提供了实用工具, 减少了排水管网监测点布局的主观性与不确定性。未来研究应继续探索更多先进的数据分析和机器学习技术, 以提高监测设计的灵活性和适应性, 更有效地应对城市排水系统的挑战。

参考文献 (References):

[1] YAZDI J. Optimization of hydrometric monitoring network in urban drainage systems using information theory [J]. *Water Science and Technology*, 2017, 76(7): 1603-1613.

[2] 李小宁, 俞悦, 王船海, 等. 南方平原河网区城市内涝的全过程监测与特征分析[J]. *水利学报*, 2022, 53(7): 845-853.

LI Xiaoning, YU Yue, WANG Chuanhai, et al. Monitoring and characterization of the entire urban flooding process in the typical southern plain river network area [J]. *Journal of Hydraulic*

Engineering. 2022, 53(7): 845-853.

- [3] 王先伟, 宁亚洲, 方勇军, 等. 暴雨洪涝灾情快速监测评估技术框架体系与应用: 以 2022 年 6 月北江流域洪水灾情分析为例 [J]. *水利水电技术(中英文)*, 2023, 54(3): 1-20.
- WANG Xianwei, NING Yazhou, FANG Yongjun, et al. A technique framework and implementation for rapid survey and assessment of flood disasters: A case study in the North River Basin in June 2022 [J]. *Water Resources and Hydropower Engineering*, 2023, 54(3): 1-20.
- [4] 赵丛丛, 赵云飞, 任紫嫣. 水务物联监测技术在排水管网问题诊断中的应用 [J]. *环境工程*, 2023, 41(11): 34-38.
- ZHAO Congcong, ZHAO Yunfei, REN Ziyun. Application of water iot monitoring technology in problem diagnosis of drainage network [J]. *Environmental Engineering*, 2023, 41(11): 34-38.
- [5] WANG S, ZHANG X, WANG J, et al. Optimal sensor placement for the routine monitoring of urban drainage systems: A re-clustering method [J]. *Journal of Environmental Management*, 2023, 335: 117579.
- [6] 赵振东, 刘健, 崔华峰, 等. 厂网一体运行视角下汛期入河污染物监测分析及控制策略研究 [J]. *给水排水*, 2023, 59(8): 44-51.
- ZHAO Zhendong, LIU Jian, CUI Huafeng, et al. Monitoring analysis and control strategies of pollutants discharged into rivers during rainfall from the perspective of integrated operation of drainage system [J]. *Water & Wastewater Engineering*, 2023, 59(8): 44-51.
- [7] SAMBITO M, PIAZZA S, FRENI G. Optimal deployment of the water quality sensors in urban drainage systems. *Environmental Sciences Proceedings*, 2022, 21(1): 42.
- [8] BARTOS M, KERKEZ B. Observability-based sensor placement improves contaminant tracing in river networks [J]. *Water Resources Research*, 2021, 57(7): e2020WR029551.
- [9] 陈安琪, 刘玉玉, 高艺雄, 等. 面向洪水预报的雨量站网合理性评估与优化研究(英文) [J]. *水利水电技术(中英文)*, 2024, 55(5): 1-14.
- CHEN Anqi, LIU Yuyu, GAO Yixiong, et al. Evaluation and optimization of rainfall station network rationality for flood forecasting [J]. *Water Resources Research*, 2024, 55(5): 1-14.
- [10] 解子藤, 王圃, 王峰青, 等. 基于安全综合评价的管网监测点选址优化及应用 [J]. *中国给水排水*, 2023, 39(23): 59-67.
- XIE Ziteng, WANG Pu, WANG Fengqing, et al. Optimization and application of pipe network monitoring point location based on comprehensive safety valuation [J]. *China Water & Wastewater*, 2023, 39(23): 59-67.
- [11] 冯鹏, 吕谋, 董深, 等. 基于不确定性分析的供水管网压力监测点优化布置 [J]. *水电能源科学*, 2024, 42(3): 147-151.
- FENG Peng, LYU Mou, DONG Shen, et al. Optimization arrangement of pressure monitoring points in water supply network based on uncertainty analysis [J]. *Water Resource and Power*, 2024, 42(3): 147-151.
- [12] 郭效琛, 李萌, 赵冬泉, 等. 城市排水管网监测点优化布置的研究与进展 [J]. *中国给水排水*, 2018, 34(4): 26-31.
- GUO Xiaochen, LI Meng, ZHAO Dongquan, et al. Research and progress on optimal layout of monitoring points in urban drainage

- networks[J]. *China Water & Wastewater*, 2018, 34(4): 26-31.
- [13] 范立柱,樊博,李游坤. 城镇河网水系排涝优化方案研究[J]. *水利发展研究*, 2023, 23(2): 65-69.
FAN Lizhu, FAN Bo, LI Youkun. Research on optimal drainage scheme of urban river system [J]. *Water Resources Development Research*, 2023, 23(2): 65-69.
- [14] FATTORUSO G, AGRESTA A, GUARNIERI G, et al. Optimal sensors placement for flood forecasting modelling [J]. *Procedia Engineering*, 2015, 119: 927-936.
- [15] BANIK B K, ALFONSO L, DI CRISTO C, et al. Evaluation of different formulations to optimally locate sensors in sewer systems[J]. *Journal of Water Resources Planning and Management*, 2017, 143(7): 04017026.
- [16] YAZDI J. Water quality monitoring network design for urban drainage systems, an entropy method [J]. *Urban Water Journal*, 2018, 15(3): 227-233.
- [17] SAMBITO M, DI CRISTO C, FRENI G, et al. Optimal water quality sensor positioning in urban drainage systems for illicit intrusion identification[J]. *Journal of Hydroinformatics*, 2020, 22(1): 46-60.
- [18] SIMONE A, CESARO A, DI CRISTO C, et al. Two different approaches for monitoring planning in sewer networks: Topological vs. deterministic optimization[J]. *Journal of Hydroinformatics*, 2023, 25(6): 2323-2337.
- [19] ZHAO Y, SCHWARTZ R, SALOMONS E, et al. New formulation and optimization methods for water sensor placement [J]. *Environmental Modelling & Software*, 2016, 76: 128-136.
- [20] JAIN A K, M N MURTY, P J FIYNN. Data clustering: A review [J]. *ACM computing surveys (CSUR)*, 1999, 31(3): 264-323.
- [21] LIU L, LI X, XIA G, et al. Spatial fuzzy clustering approach to characterize flood risk in urban storm water drainage systems [J]. *Natural Hazards*, 2016, 83: 1469-1483.
- [22] LI J. Exploring the potential of utilizing unsupervised machine learning for urban drainage sensor placement under future rainfall uncertainty[J]. *Journal of Environmental Management*, 2021, 296: 113191.
- [23] 赵文涓,程雨涵,李梅. 基于改进 K-means 算法的排水管网监测点位优化[J]. *环境监测管理与技术*, 2024, 36(1): 79-83.
ZHAO Wenjuan, CHENG Yuhan, LI Mei. Optimization of monitoring points in drainage pipe network based on Improved K-means algorithm [J]. *The Administration and Technique of Environmental Monitoring*, 2024, 36(1): 79-83.
- [24] CARDOSO S M, BARROS D B, OLIVEIRA E, et al. Optimal sensor placement for contamination detection: A multi-objective and probabilistic approach [J]. *Environmental Modelling & Software*, 2021, 135: 104896.
- [25] OHAR Z, LAHAV O, OSTFELD A. Optimal sensor placement for detecting organophosphate intrusions into water distribution systems [J]. *Water Research*, 2015, 73: 193-203.
- [26] GIRONÁS J, ROESNER L A, DAVIS J, et al. Storm water management model applications manual [R]. Cincinnati: National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency, 2009.
- [27] ROSSMAN L A, HUBER W C. Storm water management model reference manual Volume III—Water Quality [R]. Cincinnati: US Environmental Protection Agency, Office of Research and Development, National Risk Management Laboratory, 2016.
- [28] 韩家炜. 数据挖掘: 概念与技术: 第二版[M]. 北京: 机械工业出版社, 2007.
HAN Jiawei. *Data mining: concepts and techniques: Vol. 2* [M]. Beijing: China Machine Press, 2007.
- [29] WARD J R J H. Hierarchical grouping to optimize an objective function[J]. *Journal of the American Statistical Association*, 1963, 58(301): 236-244.
- [30] ROUSSEEUW P J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis[J]. *Journal of Computational and Applied Mathematics*, 1987, 20: 53-65.
- [31] DAVIES D L, BOULDIN D W. A cluster separation measure[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1979(2): 224-227.
- [32] BOX G E P. *Time series analysis: Forecasting and control*[M]. New Jersey: John Wiley and Sons Inc, 2015.

(责任编辑 王璐)