

DOI:10.19853/j.zgjsps.1000-4602.2021.09.021

基于情景模拟的城市排水管网监测点自动识别

郭效琛¹, 李萌², 赵冬泉², 杜鹏飞¹

(1. 清华大学 环境学院, 北京 100084; 2. 北京清环智慧水务科技有限公司, 北京 100083)

摘要: 城市排水管网监测对于城市水环境治理及雨洪管理具有重要意义, 建立排水管网的在线监测体系是智慧排水的重要组成部分, 监测点位置的合理选择与布置则是监测有效性的重要保障。为此, 基于情景模拟, 建立了排水管网监测点的自动识别方法, 并以我国华北地区 Q 市作为案例, 建立管网模型, 对监测点进行自动识别。该方法可避免人为制定方案时对管网运行规律认识不足的局限性, 降低人工制定和调整监测方案的工作量, 并且可对监测点进行定量化评估, 确保了监测点的代表性。基于系统整体监测的概念进行点位的识别与优化, 利用有限数量的监测设备尽可能全面反映排水管网的整体运行状况, 是实现智慧排水的基础与必要条件。

关键词: 智慧排水; 管网监测点; 自动识别; 情景模拟

中图分类号: TU992 **文献标识码:** A **文章编号:** 1000-4602(2021)09-0130-07

Automatic Identification of Urban Drainage Pipe Monitoring Points Based on Scenario Simulation

GUO Xiao-chen¹, LI Meng², ZHAO Dong-quan², DU Peng-fei¹

(1. School of Environment, Tsinghua University, Beijing 100084, China; 2. Beijing Tsinghuan Smart Water Tech. Co. Ltd., Beijing 100083, China)

Abstract: Urban drainage network monitoring is meaningful for urban water environment governance and stormwater management. Establishing the on-line monitoring system of drainage pipes is an important part of smart drainage, and suitable selection and arrangement of monitoring points is an important guarantee for the reliability of the monitoring data. Based on scenario simulation, an automatic identification method for monitoring points of drainage network was established. The method was applied to Q city in North China, and a drainage pipe model was established for automatic identification of monitoring points. The method could avoid the limitation of insufficient understanding of the operation of the pipeline network when formulating the program manually, reduce the workload of making and adjusting the monitoring program, and make quantitative evaluation of the monitoring points, so as to ensure the representativeness of the monitoring points. The identification and optimization of monitoring points based on the concept of overall monitoring of the system, and the use of a limited number of monitoring equipment to fully reflect the overall operation status of the drainage network as far as possible, are the basis and requisite for the realization of smart drainage.

Key words: smart drainage; monitoring point of drainage pipe; automatic identification; scenario simulation

通信作者: 杜鹏飞 E-mail: dupf@tsinghua.edu.cn

加强排水管理智能化、建立信息化管理模式,是我国排水管网管理工作的重点之一。近年来,推进智慧排水建设在国家政策层面有明确的工作要求、在标准规范层面有明确的技术要求、在各城市行业管理层面有明确的现实需求。构建完善的在线监测网络,全面获取管网运行动态信息是智慧排水的重要组成部分,而监测点的准确识别可在一定经济投入下获取最有效的信息。目前,有关排水管网监测点识别与优化布置的研究非常有限,且主要依赖少量实测数据和人工筛选,无法掌握管网全部节点信息和规律,节点设置存在很强的随机性,不利于管网信息化建设。因此,笔者提出一种基于模型情景模拟的排水管网监测点自动识别方法,同时在我国华北地区Q市进行应用研究。该方法以各节点时间序列为研究对象,构建衡量节点间距离的综合指标,利用聚类的思想,对监测点进行自动识别,可降低人工筛选监测点带来的不确定性和主观性,保障排水管网监测信息的有效性,推进排水管理的智能化。

1 研究背景

1.1 排水管网在线监测的重要性

随着城市化进程的加快,大多数城市原有的排水设施已无法满足城市排水需求,导致城市局部内涝频发,黑臭水体等问题严重,因此需要对城市排水管网进行改造和管理,充分发挥其作用。在排水管网的规划、设计、管理、评估与诊断等环节,均需以真实的管网结构数据与在线监测数据为支撑^[1];在城市雨洪管理方面,监测和模型技术也是提高方案有效性、保障城市水安全的关键环节^[2];同时,在线监测技术在排水管网运行管理中应发挥更为积极与重要的作用^[3]。2014年,住房和城乡建设部提出建设海绵城市,2015年又进一步提出了海绵城市考核评估办法,在此背景下,对城市排水管网监测工作提出了进一步的需求。近年来,国家先后出台了《国务院办公厅关于做好城市排水防涝设施建设工作的通知》(国办发[2013]23号)、《国务院关于加强城市基础设施建设的意见》(国发[2013]36号)等一系列文件,都对城市排水信息系统的建设提出了明确要求,智慧排水应当做到信息化、实时监控、智能化,其基础是进行排水管网的有效监测。

1.2 监测点识别布置的研究不足

目前已有不少研究着眼于排水管网监测方面,主要集中于排水管网信息管理及系统的设计,但针

对排水管网监测点布置及优化的研究非常有限^[4-5]。目前已有的研究,主要依赖于实际监测数据,如以某天的监测数据为依据,选择最大流量、最小流量及COD作为指标,再进行一定的数据处理,作为衡量节点间相关性的依据^[6],这样得到的优化结果只能对某一天的排水或降雨情景适用,无法全面掌握排水管网的运行规律。在监测点设置方法上主要依赖于人工筛选,而没有一定的技术体系,对方案更缺少量化的评估,存在很强的主观性和不确定性,无法保障排水管网信息获取的有效性^[7]。因此,需要建立一套科学的方法体系,对排水管网监测点进行自动识别与筛选,并能够对监测方案进行量化评估,为排水管网在线监测体系的建立提供基础,进一步建立排水管网动态诊断模式,改变现有的粗放式规划管理模式,大幅提高排水管网建设的规划、管理与决策水平,推进排水系统的智能化管理。

2 监测点自动识别方法

本研究构建监测点自动识别方法的整体技术路线如图1所示。首先建立研究区域排水管网模型,模拟得到各节点监测指标的时间序列曲线,然后根据相似性进行节点的聚类分析,最后根据聚类结果筛选代表性监测点,确定监测方案。

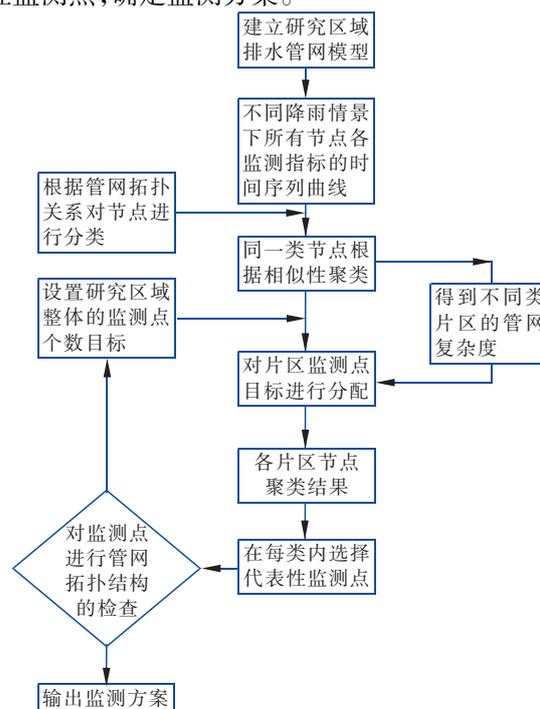


图1 排水管网监测点自动识别技术路线

Fig.1 Methodology framework of automatic identification of drainage pipe monitoring points

2.1 时间序列相似性衡量

对时间序列进行聚类,首先需要对序列间的相似/不相似性进行衡量^[8],其中皮尔逊相关系数和欧氏距离是常用的两个指标^[9]。

两个变量 X, Y , 则变量间的皮尔逊相关系数可通过公式(1)计算:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y} = \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sigma_x \sigma_y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)} \sqrt{E(Y^2) - E^2(Y)}} \quad (1)$$

式中: E 表示数学期望; cov 表示协方差。

欧氏距离则可以反映在 m 维空间中两个点之间的真实距离。

点 $A = (a_1, a_2, \dots, a_n)$ 和点 $B = (b_1, b_2, \dots, b_n)$ 的距离为 $d(A, B)$, 可通过公式(2)进行计算:

$$d(A, B) = \sqrt{\sum (a_i - b_i)^2} \quad (i=1, 2, \dots, n) \quad (2)$$

皮尔逊相关系数衡量时间序列相似性关注的是时间序列变化趋势,而欧氏距离则衡量时间序列各时间点所对应的差值。对于排水管网而言,相同的变化趋势说明有着类似的汇水过程,而变化范围或峰值的大小则反映了该节点水量的多少。考虑排水管网监测的实际需求,聚为一类的监测点,既需要有相似的变化趋势,也需要有相似的变化范围。因此将皮尔逊系数与欧氏距离综合起来,共同衡量节点间的相似性。

2.2 综合距离指标的构建

将皮尔逊系数和欧氏距离进行结合,构建综合距离指标。具体步骤如下:

① 求取 m 个节点时间序列彼此间的皮尔逊相关系数,构建相似性矩阵:

$$P_0 = (r_{ij})_{m \times m} \quad (3)$$

该相似性矩阵对角线元素为 1, 取值范围为 $[-1, 1]$, 即相似性越大, 变化趋势越相似, 则越接近 1, 变化趋势上关联性小则接近 0, 若呈现相反趋势则小于 0, 表现出一定的负相关。

② 求取 m 个节点时间序列彼此间的欧氏距离, 构建距离矩阵:

$$D_0 = (q_{ij})_{m \times m} \quad (4)$$

该距离矩阵对角线元素为 0, 两个节点间差异越大, 则距离越远, 取值范围为 $[0, +\infty)$ 。

③ 矩阵的处理。为了综合两个指标, 对相似

性矩阵 P_0 中的元素进行如下变化得到矩阵 P_1 :

$$P_1 = (1 - r_{ij})_{m \times m} = (s_{ij})_{m \times m} \quad (5)$$

矩阵 P_1 的对角线元素为 0, 取值范围为 $[0, 2]$, 即节点越相似则数值越小, 变化趋势差异越大则数值越大, 与距离矩阵相一致。

对新的相似性矩阵 P_1 和距离矩阵 D_0 进行线性比例变换:

$$P' = \left(\frac{s_{ij}}{\max s_{ij}} \right)_{m \times m} = (p_{ij})_{m \times m} \quad (6)$$

$$D' = \left(\frac{q_{ij}}{\max q_{ij}} \right)_{m \times m} = (d_{ij})_{m \times m} \quad (7)$$

矩阵 P' 和 D' 的取值范围都为 $[0, 1]$, 其中对角线上元素为 0, 节点 i 与节点 j 间的差异越大, 则 p_{ij} 和 d_{ij} 的值越接近 1。

④ 构建综合距离矩阵:

$$M = \alpha P' + \beta D' = (m_{ij})_{m \times m} \quad (8)$$

式中: α 和 β 分别表征了皮尔逊相关性系数和欧氏距离对综合距离矩阵的影响程度, 其中 $\alpha + \beta = 1$ 。当更关注节点变化趋势是否一致时则可增大 α 所占比例; 反之, 若更关注绝对数值的变化, 则需要增大 β 所占比例。本研究假定两者具有同样的重要性, 即 $\alpha = \beta = 0.5$ 。

2.3 聚类分析与代表性节点的确定

以综合距离矩阵 M 作为聚类的依据, 为在不同片区进行横向比较与监测点的筛选, 在进行矩阵的线性变化处理时, 分母为所有类别中的最大值, 但聚类仍在每一片区节点内进行。利用 Matlab 中的 `pdist` 和 `linkage` 函数, 选择相对具有较强鲁棒性的未加权平均距离法进行聚类^[10]。

对各片区节点进行聚类分析后, 可根据聚类过程对不同片区的管网复杂程度进行对比, 并作为整体监测点个数在各片区进行分配的依据。从而得到在整体监测目标下, 各片区的聚类数目和每一类中所包含的节点。

在每一类节点中, 计算每个节点与其他节点的平均综合距离, 选择平均距离最小的节点作为该类的代表性节点。对自动识别的监测点进行管网拓扑关系的检查, 得到最终监测方案。

3 案例分析

3.1 模型概化和管网节点分类

Q 市位于我国华北地区, 研究区域覆盖面积为 21.5 km^2 。以已有的雨污水管网信息为基础, 利用

SWMM 构建排水管网模型,共包含 187 个节点、192 根连接管、17 个排口,排水管网拓扑关系如图 2 所示。根据节点排口、地理分区及管网拓扑结构等,对示范区内 187 个节点进行分类,每类节点的位置分布如图 2 所示。

节点分类详细情况如表 1 所示。在 187 个主要节点中,源头入流点 89 个,非源头点 98 个。因为源头入流点主要受汇水区汇水情况影响,彼此间差异较大,且源头入流点一般不作为监测点,故对 98 个非源头点进行聚类分析。

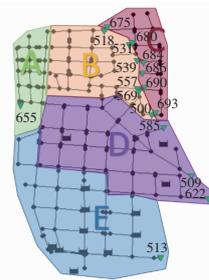


图 2 研究区域排水管网拓扑关系和节点分类情况
Fig. 2 Topology diagram of drainage pipes and category of nodes in study area

表 1 节点分类详细情况汇总

Tab. 1 Detailed information of nodes in different categories

类别	排口	上游节点		个数	位置
A	655	源头入流点	647、652、645、640、642、649、639、634、635、630、632、624、625、626、637、627、629	17	西北片区
		非源头点	654、651、650、646、653、644、641、648、638、633、643、631、435、636、628	15	
B	590、569、539、531、557、518	源头入流点	586、588、527、533、522、525、532、516、520、521、548、550、543、696、541、545、566、568、695、534、540、546	22	北部片区
		非源头点	530、529、528、526、523、556、554、553、542、552、551、549、544、567、589、587、538、537、547、536、517、519、535	23	
C	693、690、686、684、680、675	源头入流点	691、687、685、678、681、674、670、682	8	东北片区
		非源头点	679、689、688、692、672、671、683	7	
D	585、609、622	源头入流点	564、565、581、577、580、558、490、493、606、601、602、599、597、592、595、591、593、615、611、613、623、617、619	23	中部片区
		非源头点	308、492、621、491、620、618、616、614、612、584、187、583、582、562、579、561、608、600、594、559、578、596、598	23	
E	513	源头入流点	720、718、719、697、514、721、507、498、725、475、476、726、462、469、467、460、659、656、704	19	南部片区
		非源头点	472、506、477、515、499、511、510、509、502、508、500、501、707、706、705、724、471、474、505、727、470、466、463、465、468、461、464、657、660、667	30	

3.2 监测点的自动识别

在 2017 年 7 月 14 日的降雨条件下,对研究区管网运行情况进行模拟,降雨随时间的变化如图 3 所示。降雨持续时间 84 min,累积降雨量 8.6 mm。

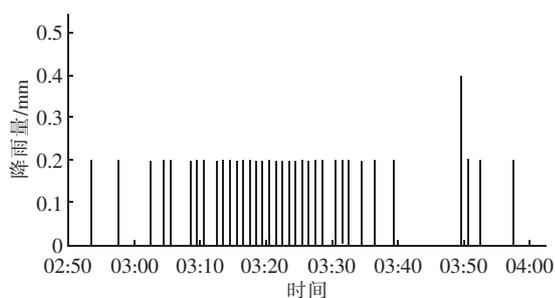


图 3 2017 年 7 月 14 日降雨情况

Fig. 3 Rainfall event on 14th July, 2017

以液位监测为例,以综合距离指标为聚类依据,在 Matlab 中利用未加权平均距离法,对所有节点进行聚类分析,所得聚类结果如图 4 所示。可以看出,根据拓扑结果将节点分类后,每个类别内部节点间的综合距离有明显差异。B 类和 C 类节点彼此间距离较小,相似性强;A 类和 E 类节点彼此间的距离相对较大,相似性弱;D 类节点,除 608 和 308 节点与其他节点距离远、相似性差以外,其余节点彼此间距离相近,具有很强的相关性。

不同片区的管网拓扑关系、地形地势、基础设施及社会环境等因素存在差异,即使在相似的降雨条件下,各片区汇水规律、管网节点彼此间的相似性也存在差异。综合距离可作为衡量管网复杂度的量化指标,指导监测点的分配。假设研究区内需设置

n 个监测点 ($n \geq 5$, 每个分类中至少有 1 个监测点), 将所有 5 类聚类的综合距离由大到小排序, 记为 d_i (第 i 个距离值), 则 D_n 的要求为 $d_{n-4} < D_n < d_{n-5}$ 。

聚类综合距离排序如表 2 所示。根据 Q 市的经济投入及监测目的, 将管网节点上的监测点布设目标定为 12 个, 则 D 的要求为 $D_8 < D < D_7$, 即 $0.1468 < D < 0.1608$, 根据表 2 的数据可知, A、B、C、D 和 E 类片区的监测点设置个数分配应为 2、2、1、3 和 4 个。

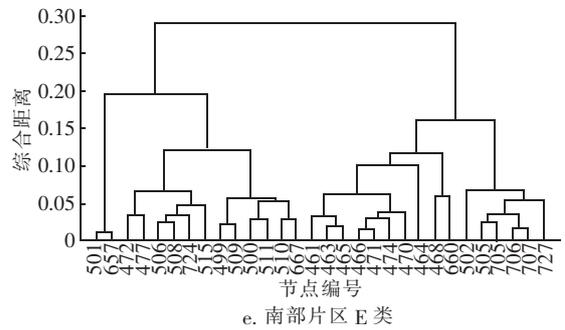
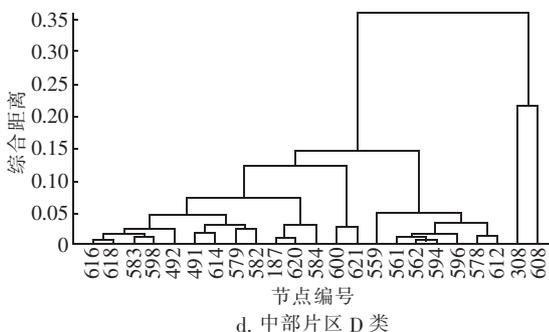
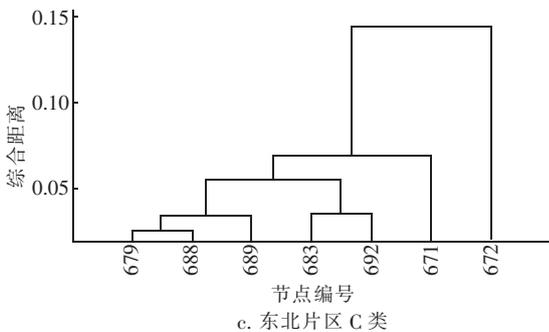
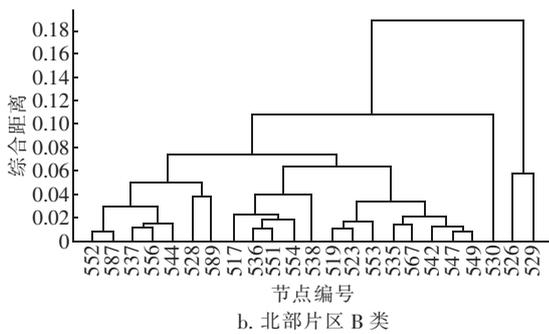
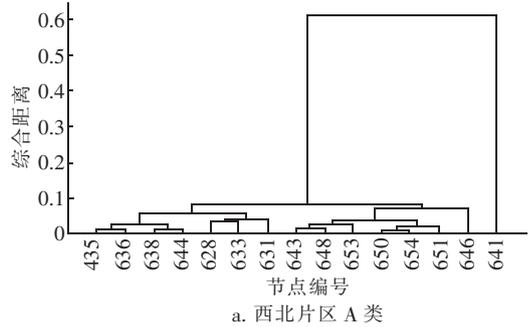


图 4 不同片区的液位数据聚类结果

Fig. 4 Clustering results of water level data among different categories

表 2 聚类综合距离排序

Fig. 2 Sorted list of comprehensive clustering distance

序号	数值	所属类别
1	0.613 1	A
2	0.359 7	D
3	0.289 0	E
4	0.215 2	D
5	0.194 3	E
6	0.188 9	B
7	0.160 8	E
8	0.146 8	D
9	0.144 7	C
10	0.123 0	D
11	0.119 9	E
12	0.115 5	E
13	0.107 5	B
14	0.100 7	E

确定每类监测点的聚类个数后, 可根据树状图得到聚为一类的节点, 以每个节点与其余节点综合距离的平均值为确定代表性节点的指标, 在每一类内, 选择平均距离最小的节点作为代表性节点。

以南部片区 E 类监测点为例, 在布置 4 个监测点的目标下, 30 个节点被分为 4 类, 其中节点 502、505、705、706、707 和 727 分为一类, 节点间彼此的综合距离如表 3 所示。选择平均距离最小的节点 505 作为该类的代表性节点进行监测。

按照相同的方法, 根据监测点目标的分配, 对示范区范围内 5 个片区的节点分别进行聚类, 并确定每类的代表性节点, 结果如表 4 所示。由此可知, 研究区内 12 个监测点分别为 308、470、499、501、505、523、526、583、608、636、672 和 679, 监测点位置如图 5 所示。

表3 E类监测点分为4类后其中1类的综合距离

Tab.3 Comprehensive distance of one cluster in E category after divided into four clusters

节点	502	505	705	706	707	727	平均
502	0	0.029 7	0.063 2	0.084 3	0.071 8	0.086 5	0.055 9
505	0.029 7	0	0.023 3	0.049 4	0.035 6	0.038 8	0.029 5
705	0.063 2	0.023 3	0	0.034 4	0.017 2	0.039 1	0.029 5
706	0.084 3	0.049 4	0.034 4	0	0.016 1	0.074 5	0.043 1
707	0.071 8	0.035 6	0.017 2	0.016 1	0	0.057 9	0.033 1
727	0.086 5	0.038 8	0.039 1	0.074 5	0.057 9	0	0.049 5

表4 聚类及代表性节点结果汇总

Tab.4 Summary of clustering and representative nodes

类别	分类序号	包含节点	代表性节点	平均距离
A	1	435,628,631,633,636,638,643,644,646,648,650,651,653,654,641	636	0.039 4
B	1	517,519,523,528,530,535,536,537,538,542,544,547,549,551,552,553,554,556,567,587,589	523	0.039 7
	2	526,529	526/529	0.028 1
C	1	671,679,683,688,689,692	679	0.020 3
	2	672	672	0
D	1	308	308	0
	2	608	608	0
	3	187,491,492,559,561,562,578,579,582,583,584,594,596,598,600,612,614,616,618,620,621	583	0.061 3
E	1	502,505,705,706,707,727	505	0.029 5
	2	461,463,464,465,466,468,470,471,474,660	470	0.050 0
	3	501,657	501/657	0.005 6
	4	472,477,499,500,506,508,509,510,511,515,667,724	499	0.055 6

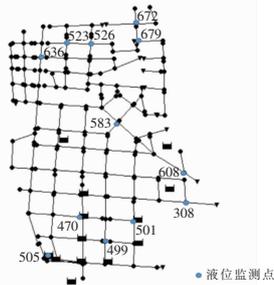


图5 研究区水位监测点分布示意

Fig.5 Layout of water level monitoring points in study area

3.3 监测方案的评估

从表4可以看出,所选代表性节点都有一个平均距离,该数值反映了代表性节点与所代表节点的相似程度。平均距离越小,该点与其他点的相似性就越强,作为监测点的代表性也就越好,监测方案则更优。在设置12个液位监测点的目标下,用自动识别的方法所得方案,12个点的平均综合距离为0.027。

根据监测目的的不同,可在不同降雨情景下进

行模拟,甚至可模拟实际难以监测到的50年一遇、100年一遇等极端降雨条件下的管网运行情况,从而更加全面地获取所有节点的有效信息,支持监测方案的制定。也可以选择不同监测目标,如液位、流量、悬浮物等,使监测点的布置更具有针对性。

在合理的布点方案下,当监测点设置数量增加时,分类增加,每类所包含的节点数减少,每一类内代表性节点与其他节点的平均距离减小。如果监测点的布置不合理,一方面,可能在同一类节点内重复筛选监测点,即监测点彼此具有很强的相关性,造成经济浪费;另一方面,即使监测点分属不同类别,所选代表性监测点的代表性也存在差异,而无法有效获取管网信息。

4 结论

① 本研究提出了排水管网在线监测点的自动识别方法:以模型模拟为基础,得到全部节点的时间序列曲线,综合皮尔逊系数和欧氏距离,构建综合距离指标作为聚类依据。

② 将该方法在Q市进行应用,以液位作为代

表性监测指标,利用模型模拟得到2017年7月14日降雨条件下全部节点的液位变化曲线,在设置12个监测点的整体约束下,对监测点个数在不同片区进行了合理分配,并最终得到监测方案。借助平均综合距离,可对监测布点方案进行定量化评估,平均综合距离为0.027。

③ 利用自动化的识别方法,减少了人工筛选监测点时的主观性与不确定性;通过模型的应用,避免了以少量实测数据为依据时对管网规律认识的不全面性,可以科学地指导监测点的增加或调整;基于系统整体监测的概念进行点位的优化,是构建排水管网在线监测体系的基础,确保了在线数据的有效性和代表性。

参考文献:

- [1] 段军,宋鹏飞. MSDF在城市排水管网监测系统中的应用研究[J]. 土木建筑工程信息技术,2010,2(2): 73-77.
DUAN Jun, SONG Pengfei. The research and application of multi-sensor data fusion in urban drainage pipe network monitoring system[J]. Journal of Information Technology in Civil Engineering and Architecture, 2010, 2(2): 73-77 (in Chinese).
- [2] 赵冬泉,邢薇,佟庆远,等. 基于数字排水技术的城市雨洪控制方案设计与评估[J]. 中国给水排水,2010,26(16): 74-77.
ZHAO Dongquan, XING Wei, TONG Qingyuan, et al. Design and evaluation of urban stormwater management projects based on DigitalWater drainage system[J]. China Water & Wastewater, 2010, 26(16): 74-77 (in Chinese).
- [3] 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 and management of urban drainage network[J]. China Water & Wastewater, 2012, 28(8): 11-14 (in Chinese).
- [4] 王言之. 基于物联网的排水监控系统规划与设计[D]. 青岛:山东科技大学,2014.
WANG Yanzhi. The Plan and Design of Urban Drainage Monitoring and Controlling System Based on Internet of Things[D]. Qingdao: Shandong University of Science and Technology, 2014 (in Chinese).
- [5] 袁景冬. 城市排水管网水力建模及其监测点优化布置研究[D]. 长沙:湖南大学,2010.
YUAN Jingdong. Research on Hydraulic Modeling and Optimization of Monitoring Points of Urban Drainage System[D]. Changsha: Hunan University, 2010 (in Chinese).
- [6] 秦礼琦. 聚类分析在优化排水管网监测点中的应用[J]. 贵州大学学报(自然科学版),2013,30(2): 123-125,134.
QIN Liqi. The application of cluster analysis in monitoring point optimal of sewer system[J]. Journal of Guizhou University (Natural Sciences), 2013, 30(2): 123-125,134 (in Chinese).
- [7] LEE J H. Determination of optimal water quality monitoring points in sewer systems using entropy theory[J]. Entropy, 2013, 15(9): 3419-3434.
- [8] 李海林,郭崇慧. 时间序列数据挖掘中特征表示与相似性度量研究综述[J]. 计算机应用研究,2013,30(5): 1285-1291.
LI Hailin, GUO Chonghui. Survey of feature representations and similarity measurements in time series data mining[J]. Application Research of Computers, 2013, 30(5): 1285-1291 (in Chinese).
- [9] 陈海燕,刘晨晖,孙博. 时间序列数据挖掘的相似性度量综述[J]. 控制与决策,2017,32(1): 1-11.
CHEN Haiyan, LIU Chenhui, SUN Bo. Research on similarity measurement of time series data mining[J]. Control and Decision, 2017, 32(1): 1-11 (in Chinese).
- [10] 胡雷芳. 五种常用系统聚类分析方法及其比较[J]. 浙江统计,2007(4): 11-13.
HU Leifang. Five common methods for system clustering and their comparison[J]. Zhejiang Statistics, 2007(4): 11-13 (in Chinese).

作者简介:郭效琛(1992-),女,山西太原人,博士,研究方向为城市排水管理。

E-mail: guoxiaochen_92@163.com

收稿日期:2020-02-27

修回日期:2020-03-20

(编辑:刘贵春)