

城市雨水管理

DOI: 10. 19853/j. zgjsps. 1000-4602. 2024. 03. 017

划分聚类法在短历时雨型设计中的应用

金 灿¹, 袁文秀², 周 宏¹, 刘 俊¹

(1. 河海大学 水文水资源学院, 江苏 南京 210098; 2. 江苏省水利工程规划办公室, 江苏 南京 210029)

摘 要: 为更好地反映短历时暴雨特征,对雨型进行分类,基于四川省乐山国家基准气候站1981年—2021年逐分钟降雨数据资料,使用多种方法对90场历时60 min、78场历时120 min和86场历时180 min降雨选取最佳聚类数,采用K-means聚类和PAM聚类两种方法对雨型进行分析。结果表明,K-means聚类法对城市短历时暴雨雨型分类的结果比PAM聚类法更直观有效;乐山气候站历时60 min的雨型分为3类、历时120 min的雨型分为2类、历时180 min的雨型分为3类,单峰靠前的雨型最为常见。研究结果可为乐山市海绵城市建设、排水防涝规划等提供参考,同时可为机器学习在设计暴雨雨型研究中的应用提供新思路。

关键词: 设计暴雨雨型; 划分聚类; K-means聚类; PAM聚类; 乐山市

中图分类号: TU992 **文献标识码:** A **文章编号:** 1000-4602(2024)03-0113-07

Application of Partitioning Clustering in Short-duration Storm Pattern Design

JIN Can¹, YUAN Wen-xiu², ZHOU Hong¹, LIU Jun¹

(1. College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China;

2. Jiangsu Provincial Water Resources Engineering Planning Office, Nanjing 210029, China)

Abstract: In order to better reflect the characteristics of short-duration rainstorms and classify the storm patterns, based on the minute-by-minute rainfall data from 1981 to 2021 at the National Benchmark Climate Station in Leshan, Sichuan Province, the best clustering numbers were selected for 90 rainstorms with 60 min duration, 78 rainstorms with 120 min duration and 86 rainstorms with 180 min duration using various methods. K-means and PAM partitioning clustering methods were used to analyze the storm pattern. The results showed that K-means clustering was more intuitive and effective than PAM clustering in classifying urban short-duration storm patterns. The storm patterns of Leshan Climate Station with duration of 60 min, 120 min and 180 min were divided into three categories, two categories and three categories respectively, and the storm patterns with a single peak in front were the most common. The results are intended to provide references for sponge city construction, drainage and waterlogging control planning of Leshan City, and provide new ideas for the application of machine learning in the research of design storm patterns.

基金项目: 国家自然科学基金资助项目(41471015); 国家重点研发计划项目(2018YFC0407201); 中英联合科学创新基金资助项目(UUFRIP_100051)

通信作者: 刘俊 E-mail: lj@hhu.edu.cn

Key words: design storm pattern; partitioning clustering; K-means clustering; PAM clustering; Leshan City

由于城镇化发展,城市不透水下垫面显著增加,改变了产汇流机制^[1],城市热岛效应加剧^[2]。IPCC AR6报告指出,全球水文循环受气候变化影响而加快,全球降水的强度呈上升态势,且时空分布不均^[3]。在城市化与气候变化的双重影响下,我国城市极端暴雨发生的强度和频率都有所增加,暴雨引起的洪涝灾害频发,给国民经济和人民生命安全带来了巨大威胁。

随着海绵城市建设的不断推进,对城市雨洪精细化模拟及管理提出了越来越高的要求^[4-5]。在此背景下,国内外学者对降雨强度和降雨量^[6]、时空分布^[7]等开展了相关研究,但对城市短历时设计暴雨雨型的相关研究不够深入。目前常用的城市短历时设计暴雨雨型有芝加哥雨型^[8]、Pilgrim&Cordery雨型^[9]、Huff雨型^[10]、通过模糊识别法确定的雨型^[11]等。随着机器学习的发展,聚类分析等算法也被应用于雨型研究中。聚类分析是统计学中常用的数据挖掘方法,其目的是在大量数据中识别相似对象,以便将其分成不同的组别,同组中的对象有较大的相似性,不同组的对象有较大的相异性。在机器学习中,聚类分析属于无监督学习。相比于人工雨型分类,聚类分析方法能更客观、更充分地呈现降雨过程。目前常用的聚类方法有层次聚类、划分聚类、人工神经网络聚类、核聚类、序列数据聚类、复杂网络聚类等^[12]。其中,划分聚类方法具有聚类明确、类别以及数据之间的距离清晰等特点,但其需要预先指定要生成的聚类数量。目前国内关于K-means聚类划分雨型的研究^[13-15]主要集中在长历时设计雨型,聚类数的选定方法主要为试算或枚举,方法单一低效。划分聚类方法在城市短历时设计雨型中的研究有待开展,雨型聚类数的选定需要更加高效可靠的方法。鉴于此,笔者基于四川乐山国家基准气候站的逐分钟降雨数据,划分不同历时的降雨后进行归一化,使用多种方法选取雨型聚类数,采用K-means聚类和PAM (Partitioning Around Medoids)聚类两种方法对雨型进行分析,比较不同聚类方法的合理性,得到不同历时下的雨型特征和分类,旨在为后期城市规划和排水防涝设计提供

参考。

1 数据来源与处理

本研究采用乐山国家基准气候站1981年—2021年共计41年的逐分钟降雨数据,以短历时60、120、180 min划分降雨场次。由于聚类的样本是多维空间中的点,维度与计算的复杂程度相关,为方便统计与计算,逐分钟降雨以5 min为间隔进行分段,最终获得90场历时60 min的降雨、78场历时120 min的降雨和86场历时180 min的降雨。由于乐山市当地没有制定短历时降雨等级划分的标准,因此参考《短时气象服务降雨量等级》(T/CMSA 0013—2019)和《短时临近降雨强度等级划分》(DB 34/T 1592—2012)标准划分不同历时降雨场次的降雨等级,如表1和表2所示。

表1 短历时降雨等级划分

Tab.1 Classification of short-duration rainfall

等级	各历时降雨量		
	历时 60 min	历时 120 min	历时 180 min
小雨	<2	<2.5	<3
中雨	2~3.9	2.5~5.9	3~7.9
大雨	4~7.9	6~12.9	8~15.9
暴雨	8~19.9	13~29.9	16~39.9
大暴雨	20~50	30~70	40~90
特大暴雨	>50	>70	>90

表2 乐山降雨场次的降雨量等级统计

Tab.2 Rainfall rating of rainfall fields in Leshan

等级	各历时降雨场次		
	历时 60 min	历时 120 min	历时 180 min
小雨	0	0	0
中雨	3	2	0
大雨	19	2	1
暴雨	27	11	11
大暴雨	33	49	57
特大暴雨	8	14	17
总计	90	78	86

由表2可知,本次统计的降雨样本主要为暴雨至特大暴雨,但也有部分场次为中到大雨,降雨量分布跨度较大。由于本研究是利用降雨过程的相

似性对其进行分类,不同范围的降雨量对距离计算的贡献也不同,尤其是降雨量为特大暴雨的场次权重会明显偏大。为消除样本的雨量等级差异带来的影响,使每场降雨的权重公平,对每场降雨的各时段雨量都进行归一化处理。每场降雨各时段归一化后的数值范围均为 $[0,1]$,但仍保留降雨过程中的峰值位置和相对比例。

归一化公式为:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

式中: x 为各时段的降雨量; x' 为归一化后各时段的降雨量比例; $\max(x)$ 、 $\min(x)$ 分别为一场降雨中各时段的最大降雨量和最小降雨量。

2 雨型划分聚类及设计

2.1 K-means 聚类

K-means 聚类^[16]是最常用的聚类分析算法,它需要预先指定初始聚类数目 k 以及 k 个初始聚类中心,通过计算数据对象与聚类中心之间的欧氏距离,以新的均值更新聚类中心(即质心)的位置并降低各个簇的簇内平方和(WSS),当WSS实现收敛时,即可输出最终结果。欧氏距离与数据的相似度成反比,其计算公式为:

$$d_{\text{euc}}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

式中: x 为数据对象; y 为第 i 个聚类中心; n 为数据对象的维度; x_i 、 y_i 分别为 x 和 y 的第 i 个属性值。

WSS的计算公式为:

$$\text{WSS} = \sum_{k=1}^K \sum_{x_i \in C_k} (x_i - y_i)^2 \quad (3)$$

式中: C_k 为第 k 个簇。WSS的大小表示聚类结果的好坏。

K-mean 聚类的步骤如下:①预先指定要创建的集群数量 K ;②从数据集中随机选择 k 个对象作为初始聚类中心;③根据数据和聚类中心之间的欧氏距离,将每个数据分配到最近的中心所在的聚类;④对于每个聚类,通过计算簇中所有数据点的新平均值来更新质心,并计算所有簇的WSS;⑤迭代步骤③和④,直到集群分配停止更改,并且各簇的WSS尽可能小。

2.2 PAM 聚类

PAM 聚类属于 K-medoids 聚类,是一种比 K-means 聚类更具鲁棒性的替代方法。与 K-means

聚类相比,该算法对噪声和异常值的敏感性较低,它使用实际数据对象而非计算得到的均值作为聚类中心,可避免某个极大值对质心的偏移影响。

PAM 聚类的算法类似于 K-means 聚类的算法:

①预先指定要创建的集群数量 K ;②从数据集中随机选择 k 个对象作为初始聚类中心;③根据数据和聚类中心之间的欧氏距离,将每个数据分配到最近的中心所在的聚类;④对各个聚类交换非聚类中心和聚类中心,再次计算各个数据与新的质心的距离,如果距离减小,则选择使距离减幅最大的数据作为该簇的质心;⑤如果至少有一个聚类已改变,转至步骤③,否则结束算法。

2.3 最佳聚类数 K 值的选择

在上述两种聚类方法中,集群数量 K 均由人工预先指定,如何选择最佳 K 值一直是算法研究中的难点。最佳聚类数在某种程度上是主观的,取决于测量相似性的方法和用于划分的参数。手肘法是常用的聚类数选定方法,R语言中的NbCluster程序包是多种聚类数确定方法的集合。本研究使用手肘法和NbCluster程序包确定最佳 K 值。

手肘法是一种视WSS为 K 值的函数,利用二者的关系图来确定最佳聚类数^[17]。其基本原理是:随着聚类数的增加,聚类会被分割得更详细,WSS会缩小。当 K 值小于最佳聚类数并增加时,WSS因每个聚类的聚合程度增加而骤降;当 K 值大于最佳聚类数时,WSS下降缓慢。因此WSS与 K 值的关系图呈“手肘”型,其中“肘部”点即为最佳 K 值。

NbCluster程序包^[18]提供30种确定相关聚类数量的方法,其中包括2种图解法和28种指标计算方法,这些方法通过改变聚类数量、距离度量和聚类方法的所有组合得到不同的结果。本研究参考7种典型雨型的划分并结合实际情况,设置最小 K 值为2、最大 K 值为7作为初值,使用NbCluster程序包,最终以众数原则取最佳 K 值。

2.4 聚类结果比较方法

本研究使用聚类散点图和轮廓系数对比两种方法的聚类结果。其中,聚类散点图是一种形象展示各数据间欧氏距离的方法,图中每一个点代表一场降雨。由于降雨样本被划分为多个时段,是多维度数据,不能以散点图表现,因此使用主成分分析法将其降维成二维数据,以便绘制散点图。

轮廓系数(Si)是将某个对象与自己所在聚类的

相似程度和与其他聚类的相似程度进行比较, S_i 值范围为 $[-1, 1]$, S_i 值接近1表示对象聚类良好, S_i 值接近-1表示对象聚类不佳,应将 $S_i < 0$ 的对象分配给其他更邻近的聚类。

2.5 聚类设计雨型构建

获得聚类分组后,计算同历时同分组的降雨样本各时段均值,将其相加得到某一雨型的总雨量。由于聚类时归一化后的样本时段数值是每个时段雨量相对于峰值时段雨量的相对值,而雨型是每个时段雨量相对于总雨量的相对值,故将总雨量除以各时段雨量均值,获得雨型每个时段的雨量占比。

3 结果与分析

3.1 最佳聚类数 K 值的确定

对3种历时的归一化降雨样本分别使用肘法和NbCluster程序包进行最佳聚类数 K 值的选定。其中肘法对3种历时的降雨样本最佳聚类数均为4。NbCluster程序包中的Frey指数对历时60 min和180 min的数据计算结果均显示为NaN,表明该方法

对本研究不适用,因此除去该方法。使用共计30种方法确定最佳聚类数 K 值,3种降雨历时计算的聚类数统计如表3所示。可以看出,对历时60 min的降雨样本,最佳聚类数 K 值为3;对历时120 min的降雨样本,最佳聚类数 K 值为2;对历时180 min的降雨样本,最佳聚类数 K 值为3。

表3 不同降雨历时下聚类数计算结果统计

Tab.3 Calculation statistics of the clusters number for different rainfall durations

历时/ min	各方法聚类数统计/种					
	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$
60	2	16	7	1	1	3
120	14	11	2	1	0	3
180	6	14	1	0	5	4

3.2 聚类结果与分析

使用K-means聚类和PAM聚类方法分别对降雨样本进行聚类分析,结果如图1所示。可以看出,K-means聚类的结果相互重合更少,说明聚类之间的相似性更低,聚类效果更好。

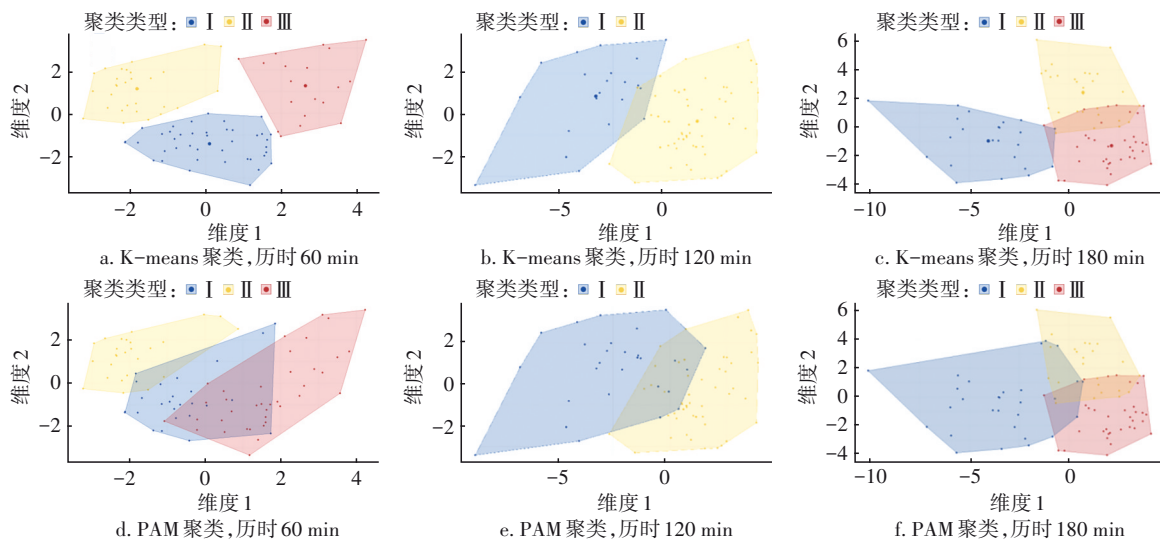


图1 不同降雨历时下两种方法的聚类情况

Fig.1 Clustering results of two methods for different rainfall durations

另外,针对历时60、120、180 min的降雨样本,采用K-means聚类法时 S_i 平均值分别为0.233、0.215、0.160,采用PAM聚类法时 S_i 平均值分别为0.207、0.165、0.153。可以看出,所有结果的 S_i 值均大于0,聚类结果良好;同一历时下K-means聚类法的 S_i 值均大于PAM聚类法,说明K-means聚类结果优于PAM聚类结果。基于以上分析,本研究选择K-means聚类的结果作为雨型分类的依据。

3.3 雨型分类结果

K-means聚类法对历时60 min降雨样本的划分结果如图2所示。第I类雨型包含41场降雨,其降雨量分布基本对称,雨峰占比为15.64%;第II类雨型包含30场降雨,其雨峰靠前且占比较大,为19.32%,降雨量在末期略有增加;第III类雨型包含19场降雨,其雨峰靠后且占比较大,为19.64%,而且在前期降雨量先增加后减少,呈双峰型分布。

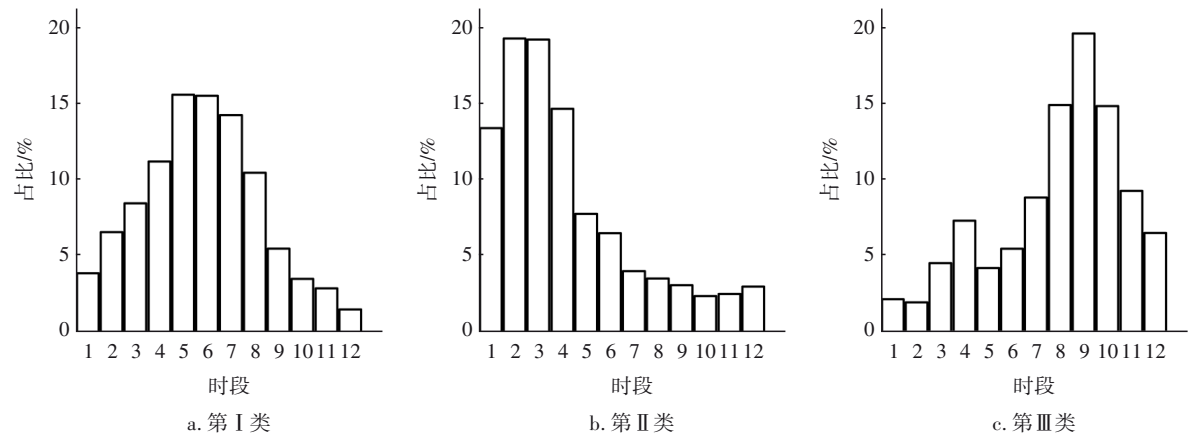


图 2 历时 60 min 降雨样本的 K-means 聚类雨型结果

Fig.2 Results of K-means clustering storm pattern for 60 min duration

K-means 聚类法对历时 120 min 降雨样本的划分结果见图 3。聚类得到的两种雨型的雨量分布都相对均匀,呈单峰型分布。第 I 类雨型包含 21 场降雨,雨峰靠后,雨峰占比为 8. 34%;第 II 类雨型包含 57 场降雨,雨峰靠前,雨峰占比为 7. 87%。

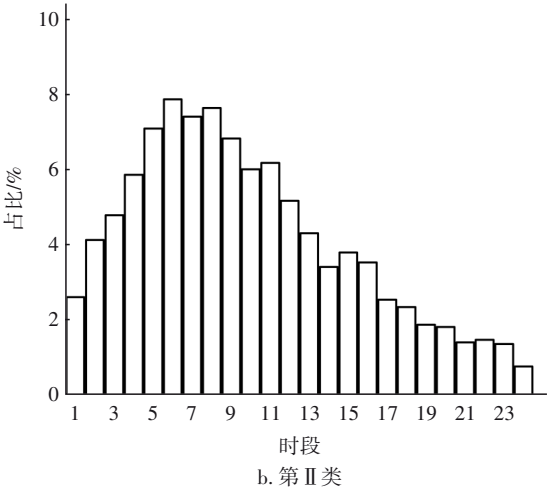
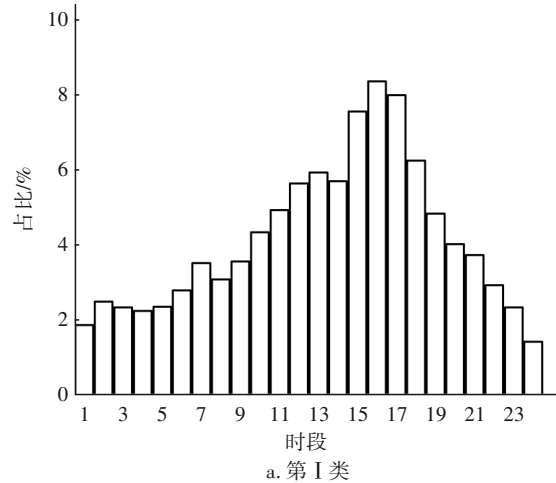


图 3 历时 120 min 降雨样本的 K-means 聚类雨型结果
Fig.3 Results of K-means clustering storm pattern for 120 min duration

K-means 聚类法对历时 180 min 降雨样本的划分结果如图 4 所示。

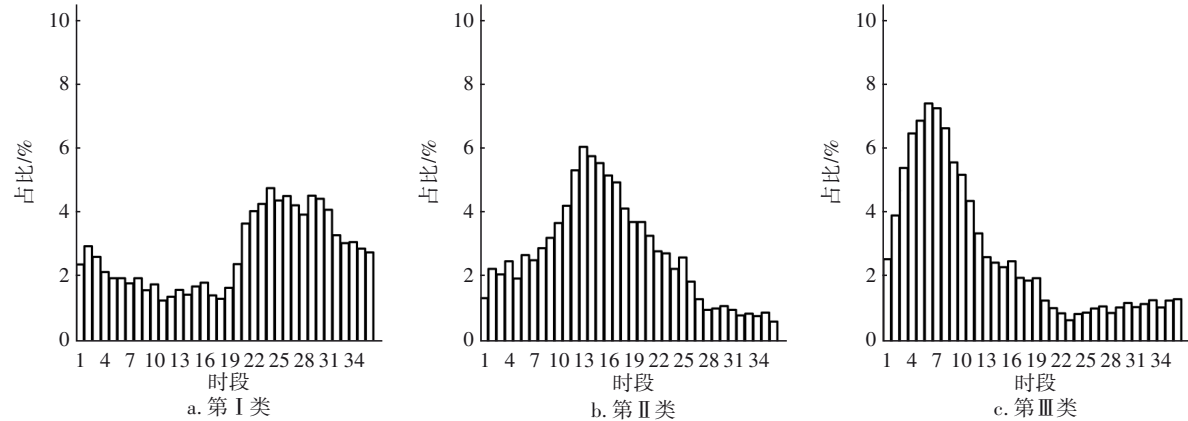


图 4 历时 180 min 降雨样本的 K-means 聚类雨型结果

Fig.4 Results of K-means clustering strom pattern for 180 min duration

由图 4 可知,第 I 类雨型包含 23 场降雨,其呈 双峰型分布,降雨量先减少后增加再减少,最大雨

峰占比为4.78%,总体分布较为均匀;第Ⅱ类雨型包含28场降雨,其呈单峰型分布且基本对称,雨峰占比为6.08%;第Ⅲ类雨型包含35场降雨,其雨量分布不均,雨峰靠前,峰值占比为7.43%,雨峰过后雨量快速减少,后期雨量略有增加,但占比均在2%以下。

4 结论

① 分别采用K-means聚类和PAM聚类两种方法对乐山气候站的降雨数据进行聚类分析,绘制两种方法的聚类散点图并计算轮廓系数,结果表明,K-means聚类法对雨型的聚类分析效果更好,各聚类之间相似性更低。

② 乐山气候站历时60 min的降雨雨型分3类、历时120 min的降雨雨型分2类、历时180 min的降雨雨型分3类,3种历时的降雨雨型均是雨峰靠前的单峰型雨型最为常见。本研究的相关成果可为乐山市及其他城市短历时设计暴雨雨型的推求提供借鉴,并可为机器学习在城市设计暴雨研究中的应用提供新思路。

参考文献:

- [1] 刘俊,周宏,鲁春辉,等.城市暴雨强度公式研究进展与述评[J].水科学进展,2018,29(6):898-910.
LIU Jun, ZHOU Hong, LU Chunhui, *et al.* A review on recent advances of urban rainfall intensity-duration-frequency relationships [J]. *Advances in Water Science*, 2018, 29(6): 898-910 (in Chinese).
- [2] 杨龙.城市下垫面对夏季暴雨及洪水的影响研究[D].北京:清华大学,2014.
YANG Long. Urban Impacts on Summer Heavy Rainfall and Flooding [D]. Beijing: Tsinghua University, 2014 (in Chinese).
- [3] 刘俊国,陈鹤,田展. IPCC AR6报告解读:气候变化与水安全[J].气候变化研究进展,2022,18(4):405-413.
LIU Junguo, CHEN He, TIAN Zhan. Interpretation of IPCC AR6: climate change and water security [J]. *Climate Change Research*, 2022, 18(4): 405-413 (in Chinese).
- [4] 吕恒.城市复杂条件对精细水文过程的影响研究[D].北京:清华大学,2018.
LÜ Heng. Impacts of Urban Complex Conditions on Detailed Hydrological Processes [D]. Beijing: Tsinghua University, 2018 (in Chinese).
- [5] 周宏,刘俊,高成,等.考虑有效不透水下垫面的城市雨洪模拟模型——Ⅱ.雨洪模拟及水文响应分析[J].水科学进展,2022,33(3):485-494.
ZHOU Hong, LIU Jun, GAO Cheng, *et al.* Development of an urban stormwater model considering effective impervious surface: Ⅱ. urban stormwater simulation and anaysis of hydrological response [J]. *Advances in Water Science*, 2022, 33(3): 485-494 (in Chinese).
- [6] 司波,余锦华,丁裕国.四川盆地短历时强降水极值分布的研究[J].气象科学,2012,32(4):403-410.
SI Bo, YU Jinhua, DING Yuguo. Research on extreme value distribution of short-duration heavy precipitation in the Sichuan Basin [J]. *Journal of the Meteorological Sciences*, 2012, 32(4): 403-410 (in Chinese).
- [7] 周琳,许武成,尹义星.近61年四川省降雨侵蚀力的时空变化趋势[J].水土保持研究,2018,25(4):8-14.
ZHOU Lin, XU Wucheng, YIN Yixing. Temporal and spatial variations of rainfall erosivity in Sichuan Province in recent 61 years [J]. *Research of Soil and Water Conservation*, 2018, 25(4): 8-14 (in Chinese).
- [8] ROSBJERG D, MADSEN H. Initial design of urban drainage systems for extreme rainfall events using intensity-duration-area (IDA) curves and Chicago design storms (CDS) [J]. *Hydrological Sciences Journal*, 2019, 64(9/12): 1397-1403.
- [9] 欧淑芳,叶兴成,王飞,等. P&C雨型在城市排水计算中的适用性分析[J].水电能源科学,2018,36(2):32-35.
OU Shufang, YE Xingcheng, WANG Fei, *et al.* Applicability of Pilgrim & Cordery rainstorm pattern in urban drainage calculation [J]. *Water Resources and Power*, 2018, 36(2): 32-35 (in Chinese).
- [10] 黄津辉,王超,范泽华.天津市设计暴雨雨型的演变[J].水资源保护,2020,36(1):38-43.
HUANG Jinhui, WANG Chao, FAN Zehua. Evolution of design rainfall pattern in Tianjin [J]. *Water Resources Protection*, 2020, 36(1): 38-43 (in Chinese).
- [11] CHEN Z H, YIN L, CHEN X H, *et al.* Research on the characteristics of urban rainstorm pattern in the humid area of Southern China: a case study of Guangzhou City [J]. *International Journal of Climatology*, 2015, 35(14): 4370-4386.

- [12] 章永来,周耀鉴. 聚类算法综述[J]. 计算机应用, 2019, 39(7): 1869-1882.
ZHANG Yonglai, ZHOU Yaojian. Review of clustering algorithms [J]. Journal of Computer Applications, 2019, 39(7): 1869-1882 (in Chinese).
- [13] 刘媛媛,王毅,刘洪伟,等. 基于动态聚类分析和模糊模式识别法的北京城区汛期降雨时空分布规律研究[J]. 水文, 2019, 39(1): 74-77.
LIU Yuanyuan, WANG Yi, LIU Hongwei, *et al.* Study on temporal distribution of precipitation in Beijing City during flood period based on dynamic cluster analysis and fuzzy pattern recognition [J]. Journal of China Hydrology, 2019, 39(1): 74-77 (in Chinese).
- [14] 李雨欣,王瑛,马庆媛,等. 基于DTW与K-means算法的河北场雨及雨型分区特征研究[J]. 地球信息科学学报, 2021, 23(5): 860-868.
LI Yuxin, WANG Ying, MA Qingyuan, *et al.* Research on the characteristics of rainfall events and rain pattern zoning in Hebei based on data mining technology [J]. Journal of Geo-information Science, 2021, 23(5): 860-868 (in Chinese).
- [15] 杨友健,郭林,林乐曼,等. 基于K-means雨型分类方法的温州市台风暴雨时空演变特征研究[J]. 水利水电技术(中英文), 2022, 53(4): 55-70.
YANG Youjian, GUO Lin, LIN Leman, *et al.* K-means classification method of rain pattern-based study on spatio-temporal evolution characteristics of typhoon rainstorm in Wenzhou [J]. Water Resources and Hydropower Engineering, 2022, 53(4): 55-70 (in Chinese).
- [16] 杨俊闯,赵超. K-Means聚类算法研究综述[J]. 计算机工程与应用, 2019, 55(23): 7-14.
YANG Junchuang, ZHAO Chao. Survey on K-means clustering algorithm [J]. Computer Engineering and Applications, 2019, 55(23): 7-14 (in Chinese).
- [17] 王建仁,马鑫,段刚龙. 改进的K-means聚类 k 值选择算法[J]. 计算机工程与应用, 2019, 55(8): 27-33.
WANG Jianren, MA Xin, DUAN Ganglong. Improved K-means clustering k -value selection algorithm [J]. Computer Engineering and Applications, 2019, 55(8): 27-33 (in Chinese).
- [18] CHARRAD M, GHAZZALI N, BOITEAU V, *et al.* NbClust: an R package for determining the relevant number of clusters in a data set[J]. Journal of Statistical Software, 2014, 61(6): 1-36.
- 作者简介:金灿(1997-),女,河南郑州人,硕士研究生,研究方向为城市内涝防治。
E-mail:jincan@hhu.edu.cn
收稿日期:2022-09-26
修回日期:2023-01-03

(编辑:刘贵春)

加强湖泊管理保护
改善湖泊生态环境
维护湖泊健康生命