

李虎, 王丽晶, 杨思敏, 等. 基于多距离融合的场次降雨时空相似性比较[J]. 水利水电技术(中英文), 2024, 55(6): 106-119.
LI Hu, WANG Lijing, YANG Simin, et al. Spatio-temporal similarity comparison of rainfall events based on multi-distance fusion [J]. Water Resources and Hydropower Engineering, 2024, 55(6): 106-119.

基于多距离融合的场次降雨时空相似性比较

李 虎¹, 王丽晶², 杨思敏², 赵 飞², 王加虎¹

(1. 河海大学 水文水资源学院, 江苏 南京 210098; 2. 北京市水科学技术研究院, 北京 100048)

摘 要:【目的】降雨相似性分析是一种用于比较和评估不同降雨事件之间相似性的一种方法, 可用于降雨径流过程的预测。针对降雨时空分布特征, 以北京市中心城区 2008—2021 年为数据样本, 提出了一种基于多距离融合的场次降雨时空相似性比较方法。【方法】首先通过设置雨量占比阈值, 进行降雨场次划分, 共划分位于汛期(6—9 月)降雨 225 场。随后提取降雨量、降雨历时、降雨强度、降雨中心范围等特征要素, 并计算 3 个方面的降雨特征距离: 降雨要素距离、降雨中心距离、降雨总量距离。同时为了更好综合特征距离, 采用主成分分析方法进行距离融合计算权重, 得出最终的距离, 并根据计算结果进行相似度排序。【结果】通过对 33 号和 120 号两场降雨的比较, 从降雨量分布图、降雨中心范围图及降雨特征要素的直观对比, 展示了较高的相似度, 经过计算的综合相似度为 78.46%。【结论】结果表明: 该方法综合考虑了降雨的强度、空间范围和降雨总量等方面的差异, 能够提供准确可靠的场次降雨相似性分析结果, 并为城市洪涝预警和灾害防治提供有益工具。

关键词: 多距离融合; 场次降雨; 降雨特征; 相似性分析; 主成分分析

DOI: 10.13928/j.cnki.wrahe.2024.06.009

开放科学(资源服务)标志码(OSID):

中图分类号: TV125

文献标志码: A

文章编号: 1000-0860(2024)06-0106-14



Spatio-temporal similarity comparison of rainfall events based on multi-distance fusion

LI Hu¹, WANG Lijing², YANG Simin², ZHAO Fei², WANG Jiahu¹

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

2. Beijing Water Science and Technology Institute, Beijing 100048, China)

Abstract: [Objective] Rainfall similarity analysis is a method used to compare and evaluate the similarity between different rainfall events, which can be used for rainfall-runoff process prediction. Focusing on the spatio-temporal distribution characteristics of rainfall, this paper takes the central urban area of Beijing from 2008 to 2021 as the data sample, and proposes a multi-distance fusion-based rainfall event spatio-temporal similarity comparison method. [Methods] First, rainfall events were divided by setting a rainfall percentage threshold, dividing 225 rainfall events in the flood season (June-September). Then, feature elements such as rainfall, rainfall duration, rainfall intensity, and rainfall center range were extracted, and rainfall feature distances were calculated in three aspects: rainfall element distance, rainfall center distance, and total rainfall distance. At the same time, in order

收稿日期: 2023-07-22; 修回日期: 2023-08-31; 录用日期: 2023-09-28; 网络出版日期: 2023-10-23

基金项目: 北京市科技计划项目“暴雨洪涝灾害智能感知及多源数据驱动风险预警技术研究与示范”(Z231100003823005); 北京市科技计划课题“北京城市洪涝淹水实时预测预警关键技术与示范”(Z201100008220005)

作者简介: 李 虎(1999—), 男, 硕士研究生, 主要从事水文与水资源研究。E-mail: tigerlee@hhu.edu.cn

通信作者: 赵 飞(1977—), 男, 正高级工程师, 硕士, 主要从事水文与水资源研究。E-mail: zf@bwsti.com

王加虎(1975—), 男, 副教授, 博士, 主要从事水文物理规律模拟及水文预报研究。E-mail: TigerLLy@126.com

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

to better integrate the feature distances, the principal component analysis method was used to calculate the weights for distance fusion, and the final distance was obtained, and similarity ranking was performed according to the calculation result. [Results] By comparing rainfall events No. 33 and No. 120, it can be seen from the rainfall distribution map, rainfall center range map and intuitive comparison of rainfall feature elements that they have a high degree of similarity. The calculated comprehensive similarity is 78.46%. [Conclusion] The result show that: this method comprehensively considers the differences in intensity, spatial range, total rainfall amount and other aspects of rainfall, and can provide accurate and reliable rainfall event similarity analysis result, and provide useful tools for urban flood warning and disaster prevention.

Keywords: multi-distance fusion; rainfall events; rainfall characteristics; similarity analysis; principal component analysis

0 引言

随着城市化进程的加速和全球气候变化的影响,城市洪涝问题日益凸显。大量人口和基础设施集中在城市地区,暴雨来袭时,由于排水系统的局限性,雨水难以及时排除,从而导致道路积水、建筑物受损甚至威胁人民生命安全^[1-3]。因此,提前预测和预判城市内涝风险对于提高城市抗洪能力、保障人民生命财产安全具有重要意义^[4]。

水文时空相似性分析作为数据挖掘技术在水文领域中的应用,其关键技术包括特征提取及相似性度量两个方面,被广泛应用于城市降雨径流过程的预测^[5]。其主要思路是依托历史降雨数据,通过提取降雨特征并进行分类分析。通过对不同降雨事件之间的特征参数如降雨强度、降雨持续时间等进行相似性比较,以判断它们是否具有相似的洪涝风险。这种分析方法为城市洪涝预测预警提供了可靠的依据,帮助决策者采取相应的措施来减轻洪涝灾害的影响。

降雨-径流过程是一个复杂的系统,涉及多个变量和因素之间的相互作用。传统的降雨-径流相似性比较方法通常只考虑少数降雨指标进行距离计算,例如汤静静等^[6]选择8个子流域的面平均雨量作为特征指标,基于集对分析原理建立了降雨相似性评价方法-集对分析法。陈建等^[7]通过降雨量、降雨历时等4个判别指标,在寻找历史相似洪水方面进行了试验应用。XIAO等^[8]以降雨强度、覆盖率和持续时间等作为等级划分的指标因子,采用K均值法分析了降水过程之间的相似性;或者仅通过计算降雨时间序列相似性来进行分析,例如AUNG等^[9]运用动态规划方法匹配两个降雨序列的相似情况。OUYANG等^[10]采用动态时间扭曲距离法对沙里桂兰克水文站220场洪水流量过程进行相似性搜索。赵圆元等^[11]构建了动态权重下的相似性分析方法,为实时洪水预报中不同研究问题的相似洪水识别提供了新途径。

但是,这些方法容易忽略降雨时间空间位置上的

对应关系,导致结果准确度不高。同时,由于降雨数据本身的复杂性,单一的距离计算方法也难以完全满足综合分析的需求。因此,国内外学者在这方面做了大量的工作,肖章玲等^[12]提出了基于面雨量及暴雨空间分布两个方面的暴雨相似性度量方法,应用于黄河中游新庙站某场暴雨相似性分析中,效果良好。钱磊等^[13]通过建立华东区域历史降雨个例库,寻找相似个例,实现了对实时多模式降雨预报的最优集成。YU等^[14]基于嘉陵江流域逐日降雨量图进行相似性搜索,为水文预报提供了新的思路。

基于此,本文提出了一种基于多距离融合的场次降雨相似性比较方法,旨在通过将多个降雨特征因素进行综合考虑,提高相似性比较的准确性。本文将研究区降雨数据进行场次划分,基于栅格编码进行降雨特征提取,原理简单易操作。提取并计算三个方面的降雨特征距离:降雨要素距离、降雨中心距离、总降雨量距离,综合考虑了降雨时空分布特征,能够更好的反映场次降雨的真实情况。同时为了融合特征距离,采用主成分分析方法进行距离权重计算,并根据最终距离进行相似度排序。

1 数据与方法

1.1 数据来源

本文研究区域为北京市中心城区,包括东城、西城、朝阳、海淀、丰台及石景山大部分地域,面积1 085 km²。中心城区地形地势西高东低,地面高程介于9.11~743.69 m区间。中心城区属于温带大陆性季风气候,受区域气候变化和城市热岛效应的影响,局部地区极端降雨频次和强度频繁发生,易造成城市道路积水、区域交通瘫痪,对城市安全运行造成极大威胁^[15]。

北京市降水量年内分配不均,降雨主要集中在汛期6—9月,汛期多年平均降雨量为488 mm,占全年降水量的80%以上^[16],2008—2021年全市汛期降雨量如图1所示。根据北京市汛期相关雨情资料^[17],

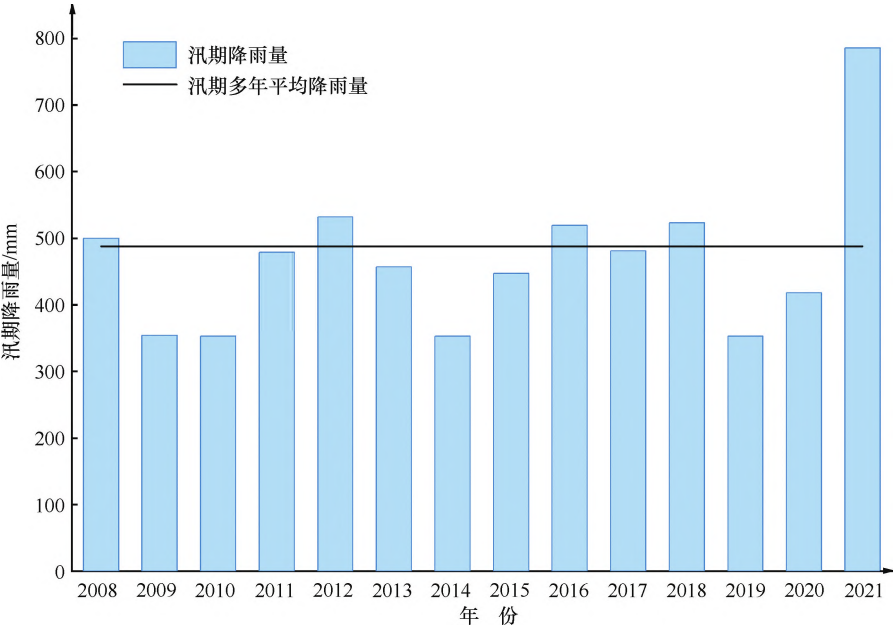


图1 2008—2021年全市汛期降雨量统计

Fig. 1 Rainfall statistics for the flood season across the city from 2008 to 2021

按照降雨量级划分天数,表1统计了2008—2021年汛期全市降雨量及小雨以上降雨日数。其中2021年6—9月累计降雨量786 mm,超多年平均同期降雨量61%;汛期降雨日数为78 d,超汛期平均降雨日数23 d。

本文所用资料为北京市155个雨量站6月1日至9月31日降雨量数据,降雨资料时间间隔为5 min,系列长度为2008—2021年。研究区范围及雨量站点分布如图2所示。

1.2 技术路线

使用计算机编程语言将历史降雨信息进行场次划

分,并提取相关降雨特征,进行聚类分析,构建历史场次降雨特征库。再提取待比较降雨特征信息,与特征库中已有的降雨特征信息进行距离计算,最终根据相似度量融合结果排序输出。具体流程如图3所示。

1.3 场次降雨的划分

强降雨往往会带来较大的径流,从而增加城市内涝的风险。降雨场次的划分是强降雨特性研究的基础,对于精准地分析降雨特征极为重要^[18]。但是,如何对降雨进行场次划分还没有统一的方法。在时间上,降雨场次可以根据研究区实际情况按照不同小时进行划分;在空间上,降雨场次可以按照地理位置、流域、行政区划等不同空间尺度进行划分。

通过阅读相关文献[19—20],本文根据研究区域内雨量站点逐5 min的降雨数据,通过限定阈值 RS_i 、 RS_m 、 RS_l 、 RS_n 来进行降雨场次划分,其中阈值 RS_i 为降雨数值超过0 mm的站点数量,阈值 RS_m 和 RS_n 分别为所有雨量站的降雨量总和的最小值和最大值,阈值 RS_l 为降雨事件最小时间间隔。具体划分标准分为两个阶段:一是判断降雨开始时刻。在某一时刻降雨数值超过0的站点数量大于 RS_i ,或某一时刻所有雨量站雨量总和大于 RS_m ,即认为一场降雨事件开始;二是判断降雨结束时刻。在某一时刻之后,

表1 2008—2021年汛期北京市小雨以上降雨日数统计

Table 1 Statistics on number of rainy days during the flood season from 2008 to 2021 in Beijing

年 份	汛期降雨量/mm	降雨日数/d	10 mm 以下日数/d	10~25 mm 日数/d	25~50 mm 日数/d	50 mm 以上日数/d
2008	500	62	49	7	5	1
2009	354	49	32	17	0	0
2010	353	49	37	9	3	0
2011	479	45	28	13	2	2
2012	532	50	38	8	2	2
2013	457	54	35	16	3	0
2014	353	49	41	6	2	0
2015	447	66	52	11	3	0
2016	519	50	36	10	3	1
2017	481	48	35	8	2	3
2018	523	59	39	14	6	0
2019	353	54	44	6	4	0
2020	418	64	54	7	2	1
2021	786	78	52	17	7	2

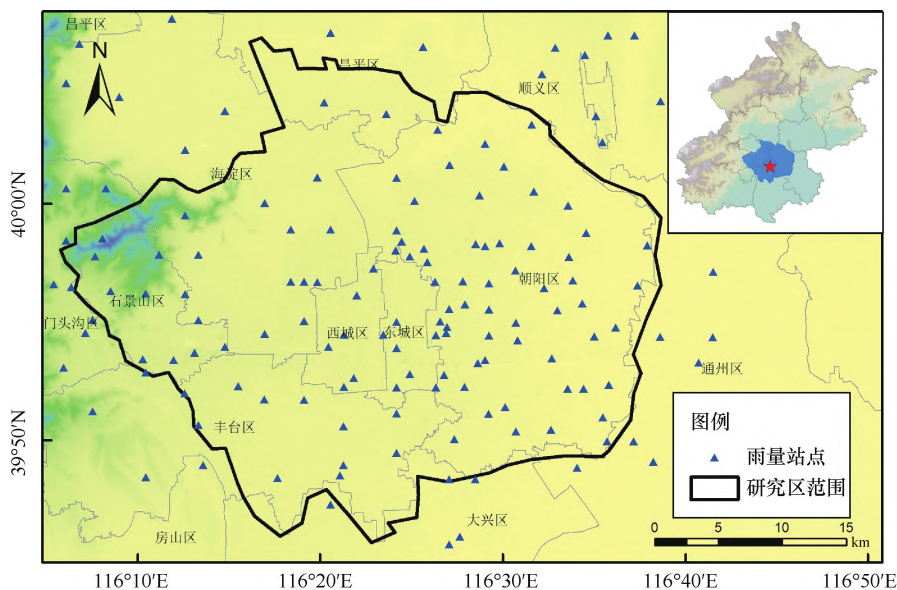


图2 研究区范围及雨量站点分布

Fig. 2 Rain station distribution in study area

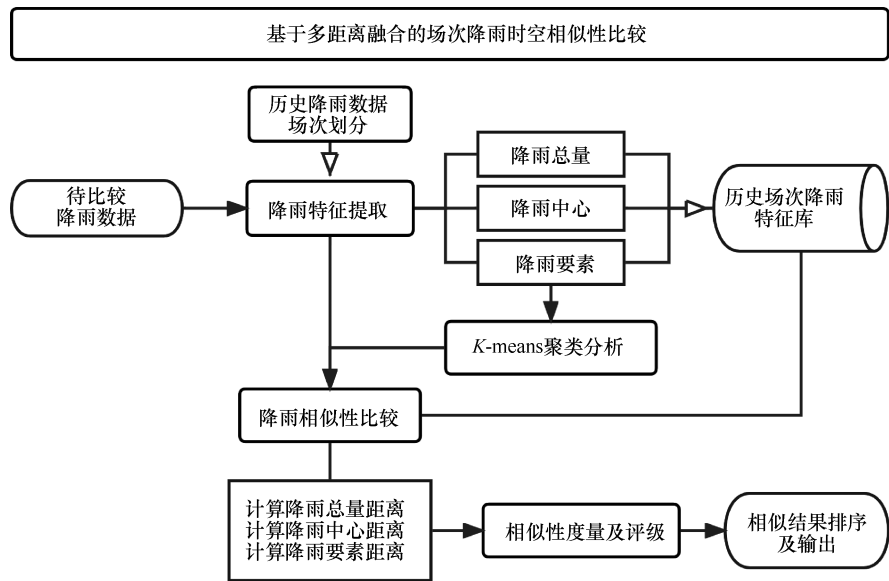


图3 技术流程

Fig. 3 Technical and operation coordination

所有雨量站在 RS_i 时段内降雨量和小于 RS_n 时, 即认为一场降雨事件结束。这样可以更好把握每个降雨过程, 提取出的降雨特征信息也更加准确。相关流程如图4所示。

本文基于已有数据进行降雨场次划分, 结合研究区实际情况及设定阈值标准, 经过整理并剔除不合理划分场次, 共划分 2008—2021 年汛期北京市中心城降雨场次 225 场。表 2 统计了 225 场次降雨起止时间。

1.4 距离计算方法

距离计算是指通过一定的数学模型和算法来衡量

两个样本之间的差异程度或相似程度。常见的距离计算方法包括欧几里得距离、曼哈顿距离、闵可夫斯基距离^[21]等。

其中, 欧几里得距离被广泛应用于判断两组数据之间的相似程度。欧几里得距离可以定义为

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

式中, x 和 y 分别表示两个样本; n 表示样本维度。

除了欧几里得距离, 曼哈顿距离也常用于计算两个样本之间的差异。曼哈顿距离可以定义为

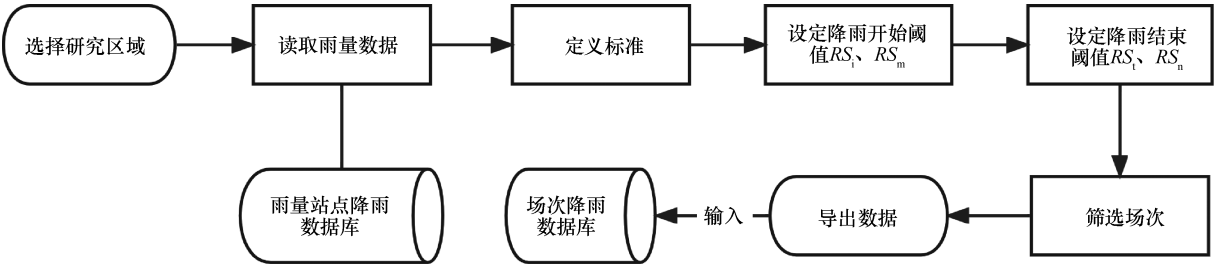


图 4 降雨场次划分流程

Fig. 4 Rainfall event division process

表 2 场次降雨起止时间

Table 2 Start and end time of rainfall events

场次降雨序号	起始时间	结束时间
1	2008-06-03T03 : 30	2008-06-03T13 : 00
2	2008-06-13T16 : 55	2008-06-14T15 : 35
3	2008-06-16T21 : 30	2008-06-17T06 : 05
4	2008-06-20T18 : 35	2008-06-20T21 : 45
5	2008-06-22T19 : 15	2008-06-23T05 : 45
⋮	⋮	⋮
221	2021-08-14T16 : 30	2021-08-14T23 : 50
222	2021-08-16T00 : 30	2021-08-16T06 : 25
223	2021-08-16T20 : 05	2021-08-17T01 : 25
224	2021-08-23T20 : 25	2021-08-24T01 : 30
225	2021-09-04T04 : 55	2021-09-04T17 : 30

$$d(x,y)=\sum_{i=1}^n|x_i-y_i| \tag{2}$$

此外，闵可夫斯基距离是欧几里得距离和曼哈顿距离的一般化形式。闵可夫斯基距离可以定义为

$$d(x,y)=\left(\sum_{i=1}^n|x_i-y_i|^p\right)^{\frac{1}{p}} \tag{3}$$

式中， p 是一个自然数，当 $p=1$ 时，是曼哈顿距离；当 $p=2$ 时，是欧几里得距离。

1.5 K-means 聚类分析

K-means 聚类分析^[22]是一种常用的数据分析方法，其基本思想是将数据分成 K 个簇，使得每个数据点都属于与之最近的簇的中心。在场次降雨相似性分析中，可用于将降雨场次划分为不同的类别，以探索不同降雨事件之间的相似性。具体步骤如下。

(1) 根据研究目的选择合适的聚类指标。

(2) 对样本数据进行标准化处理。设有 n 场降雨，每场降雨都有 p 个降雨特征，这 $m \times p$ 个数据构成一个降雨特征观测矩阵

$$X=\begin{bmatrix}x_{11}&x_{12}&\cdots&x_{1p}\\x_{21}&x_{22}&\cdots&x_{2p}\\\vdots&\vdots&\vdots&\vdots\\x_{n1}&x_{n2}&\cdots&x_{np}\end{bmatrix} \tag{4}$$

式中， x_{ij} 为第 i 场降雨的第 j 个降雨特征值。

(3) 将所有样本分成 K 簇，进而将这 K 个簇的重心(均值)作为初始的簇中心点(质心)。

(4) 依次计算各样本数据点到 K 个簇中心点的欧氏距离，将每个数据点分配到离其最近的质心所在的簇。

(5) 更新每个簇的质心，计算簇内数据点的平均值。

(6) 重复步骤(4)和(5)，直到质心不再发生明显变化或达到预定迭代次数，并统计不同样本的归属类别。

1.6 主成分分析

主成分分析是一种常用的统计方法，可以将高维数据映射到低维空间中，并保留原始数据集的大部分信息^[23]。在本文中，使用主成分分析法对三个距离矩阵进行处理，得出各主成分对应的方差和权重系数，并将权重系数归一化。具体步骤如下。

(1) 设有 n 场降雨，每场降雨与其他所有降雨都能计算出 3 个距离特征

$$X_{1,2,3}=\begin{bmatrix}X_{11}&\cdots&X_{1n}\\\vdots&\ddots&\vdots\\X_{n1}&\cdots&X_{nn}\end{bmatrix} \tag{5}$$

式中， X_1 中 X_{ij} 表示第 i 场降雨与第 j 场降雨的第一个距离矩阵。最终构成一个 $n \times n \times 3$ 的三维距离矩阵。

(2) 为了方便分析，我们将其展开成一个二维矩阵，用 $X(n^2 \times 3)$ 表示，其中每行代表一个样本，每列代表一个特征值。

(3) 将每个特征进行中心化处理，得到新矩阵 X'

$$X'_{p,q}=X_{p,q}-\frac{\sum_{k=1}^{n^2}X_{k,q}}{n^2} \tag{6}$$

(4) 计算协方差矩阵

$$C=\begin{bmatrix}c_{1,1}&c_{1,2}&c_{1,3}\\c_{2,1}&c_{2,2}&c_{2,3}\\c_{3,1}&c_{3,2}&c_{3,3}\end{bmatrix} \tag{7}$$

式中, $c_{i,j}$ 表示第 i 个特征和第 j 个特征的协方差

$$c_{i,j} = \frac{\sum_{k=1}^n X'_{k,i} X'_{k,j}}{n^2 - 1} \quad (8)$$

(5) 对协方差矩阵进行特征值分解, 得到他的特征值 $\lambda_1, \lambda_2, \lambda_3$ 和对应的特征向量 $\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3$ 。特征向量组成的新矩阵记为 $\mathbf{V}_{3 \times 3}$, 其中每一列表示一个主成分。

(6) 获取最终权重系数, 即将每个主成分中的权重系数取绝对值后, 选取其中绝对值最大的那个作为该主成分的权重系数 ω_i , 最终归一化权重系数 ω'_i , 使它们之和为 1, 计算公式为

$$\omega'_i = \frac{\omega_i}{\sum_{j=1}^k \omega_j} \quad (9)$$

2 处理与分析

2.1 特征提取及距离计算

本文采取反距离权重法对历史降雨数据进行二维插值, 得到场次降雨的逐小时及总降雨量栅格矩阵。插值生成的像元大小需要与后续用来匹配的卫星降雨预报精度保持一致。基于目前所掌握的资料, 本文将像元大小设定为 $500 \text{ m} \times 500 \text{ m}$, 则研究区内共包含 4 506 个栅格像元。为便于特征提取和对比, 通过顺序编码形式, 将二维降雨数据降低为一维序列数据, 如图 5 所示。这种方法实现了将二维空间特征映射到一维序列的线性转换, 降低数据维度的同时保留了空间特征。经过编码后的一维降雨栅格序列 R 即作为本文的降雨特征提取和相似性分析的直接输入数据

$$R = (r_1, \dots, r_n) \quad (10)$$

2.1.1 降雨总量

将每场降雨与其他所有场次降雨的总降雨量信息进行点对点的距离计算, 计算方法采用欧氏距离。将 D_R 定义为两场降雨的总降雨量距离, x_i, y_i 表示两场降雨相同编码位置降雨信息, 计算方法为

$$D_R = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (11)$$

最终得到降雨总量距离矩阵

$$\mathbf{M}_R = \begin{bmatrix} D_{1,1}^R & D_{1,2}^R & \dots & D_{1,n}^R \\ \vdots & \ddots & \ddots & \vdots \\ D_{n,1}^R & D_{n,2}^R & \dots & D_{n,n}^R \end{bmatrix} \quad (12)$$

式中, $D_{i,j}^R$ 表示第 i 场降雨与第 j 场降雨的降雨总量距离。

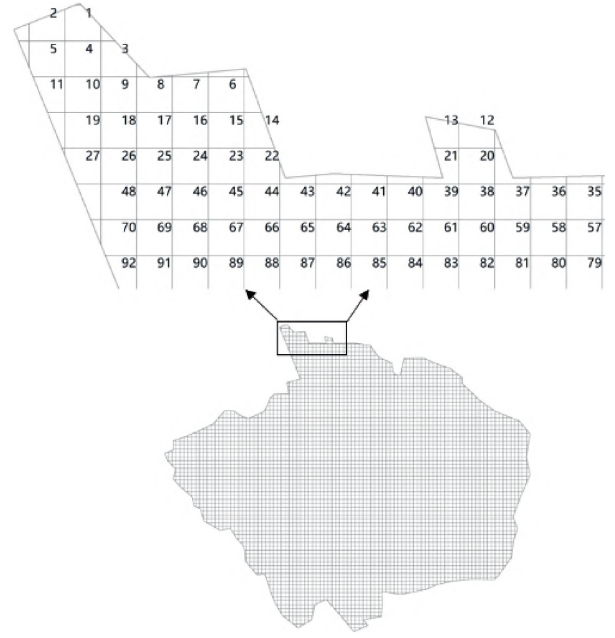


图 5 研究区栅格编码

Fig. 5 Grid coding of the research area

2.1.2 降雨中心

取场次总降雨量靠前的 10% (450 个) 格点, 即认定为此次降雨的降雨中心范围, 用格点编码序号表示其降雨中心范围特征。计算每场降雨与其他场次降雨中心范围特征的重叠情况, 记录特征重叠个数 m , 得出降雨中心范围相似性, 计算方法为

$$S_{RC} = \frac{m}{450} \quad (13)$$

式中, S_{RC} 表示降雨中心范围相似性; m 表示降雨中心范围特征重叠个数。

将 D_C 定义为两场降雨的降雨中心范围距离, 最终得到降雨中心距离矩阵

$$D_C = 1 - S_{RC} \quad (14)$$

$$\mathbf{M}_C = \begin{bmatrix} D_{1,1}^C & D_{1,2}^C & \dots & D_{1,n}^C \\ \vdots & \ddots & \ddots & \vdots \\ D_{n,1}^C & D_{n,2}^C & \dots & D_{n,n}^C \end{bmatrix} \quad (15)$$

式中, $D_{i,j}^C$ 表示第 i 场降雨与第 j 场降雨的降雨中心距离。

2.1.3 降雨要素

降雨特征要素的选取对于降雨相似性分析十分重要, 应选择对洪涝灾害的发生有显著影响的要素, 并在洪涝发生前容易获取的指标要素^[24]。为了全面、准确反映研究区域内的降雨特征要素, 本文选取降雨历时 T 、强降雨历时 T_i 、降雨量平均值 R_a 、降雨量最大值 R_m 、1 小时最大雨强 I 、降雨中心最大雨强 I_c 共

6个特征要素。

特征要素的计算依赖于式(10),各特征要素释义为:(1)降雨历时 T ,单位为h,表示降雨过程的时间要素,以整点计算。(2)强降雨历时 T_l ,单位为h,表示降雨中心范围内雨强大于7 mm/h的小时数。(3)降雨量平均值 R_a ,单位为mm,表示场次降雨总量中所有格点降雨信息的平均值。(4)降雨量最大值 R_m 。单位为mm,表示场次降雨总量中前10%格点降雨信息的平均值。(5)1 h最大雨强 I ,单位为mm/h,表示先求出逐小时雨量中所有格点降雨信息的平均值,再选出其中的最大值。(6)降雨中心最大雨强 I_c 。单位为mm/h,表示先求出逐小时雨量中降雨中心范围内降雨信息的平均值,再选出其中的最大值。

使用机器学习方法提取每场降雨的降雨要素使其与其余所有降雨进行欧氏距离计算。计算前进行标准化操作,某特征要素初始值为 X ,该要素均值 μ ,标准差为 σ ,计算方法为

$$X^* = \frac{X - \mu}{\sigma} \tag{16}$$

$$D_e = \sqrt{\sum_{i=1}^6 (x_i - y_i)^2} \tag{17}$$

将 D_e 定义为两场降雨的降雨要素距离,最终得到降雨要素距离矩阵

$$M_E = \begin{bmatrix} D_{1,1}^e & D_{1,2}^e & \cdots & D_{1,n}^e \\ \vdots & \ddots & \ddots & \vdots \\ D_{n,1}^e & D_{n,2}^e & \cdots & D_{n,n}^e \end{bmatrix} \tag{18}$$

式中, $D_{i,j}^e$ 表示第 i 场降雨与第 j 场降雨的降雨要素距离。

2.2 聚类分析

本文采用 K -means聚类方法对降雨场次进行分类,聚类变量选择上文提及的6个降雨要素。使用手肘法^[25]确定最佳的聚类数目为6类。统计每一类别具体数目及聚类中心(见表3),可以看出6类降雨在

降雨特征上存在显著差异。具体而言,类别1主要为短历时小雨,类别2为中雨,类别3为大雨,类别4为长历时强降雨,类别5和类别6为极端暴雨。值得注意的是,类别5和类别6仅包含一场降雨,均为北京市罕见的极端暴雨(2012年“7·21”和2016年“7·20”),考虑到其代表性及影响,本文将这两场异常降雨作为“极端事件类别^[26]”(研究区内任意一个区县代表站24 h内降水量 ≥ 100 mm,且至少有一个自动气象站降水强度 ≥ 40 mm/h)单独保留其统计特征,则最终聚类结果为4类。通过绘制所有类别样本在聚类变量上的箱型图(见图6),可以发现随着类别号的增加,样本变量整体呈现增大的趋势,这直观地展示了不同类别之间的差异。说明运用 K -means聚类方法效果良好,并为后续开展具有针对性的降雨时空相似度分析奠定了基础。

2.3 主成分分析融合权重

经过上述距离计算得出三个距离矩阵分别是 M_R , M_C , M_E ,为了消除不同属性间的单位差异,确保每个属性对距离计算的贡献相等。将 M_R , M_E 使用最小-最大规范化方法将矩阵内所有数据缩放致0~1范围内,对于给定的原始数据 D ,将其进行最小-最大规范化处理后得到的新数据 D' ,计算方法为

$$D' = \frac{D - D_{\min}}{D_{\max} - D_{\min}} \tag{19}$$

式中, D_{\max} 和 D_{\min} 分别是原始数据 D 的最大值和最小值。这样处理后的数据 D' 取值范围为[0,1]。

更新后的距离矩阵为 X_R , X_C , X_E ,使用2.3节提及的方法对三个距离矩阵进行处理,得出各主成分对应的方差和权重系数,并将权重系数归一化,结果如表4所列。

设定 $\omega_1 = 0.14$, $\omega_2 = 0.35$, $\omega_3 = 0.51$,则任意两场降雨之间的综合距离 $D_{\text{总}}$ 计算方法为

$$D_{\text{总}} = \omega_1 D_{i,j}^R + \omega_2 D_{i,j}^C + \omega_3 D_{i,j}^E \tag{20}$$

式中, $D_{i,j}^*$ 表示3个距离矩阵中,第 i 场降雨与第 j 场

表3 聚类中心
Table 3 Clustering center

类 别	数 目	降雨历时 /h	强降雨历时 /h	降雨量平均值 /mm	降雨量最大值 /mm	小时最大雨强 /mm·h ⁻¹	降雨中心最大雨强 /mm·h ⁻¹
1	126	11	1	9.16	16.82	3.84	9.38
2	60	15	3	24.71	43.63	9.22	21.22
3	31	25	5	51.51	77.69	11.85	26.85
4	6	38	9	87.84	132.59	19.87	40.29
5	1	19	13	206.77	265.49	39.98	56.41
6	1	59	19	264.12	316.76	25.53	31.43

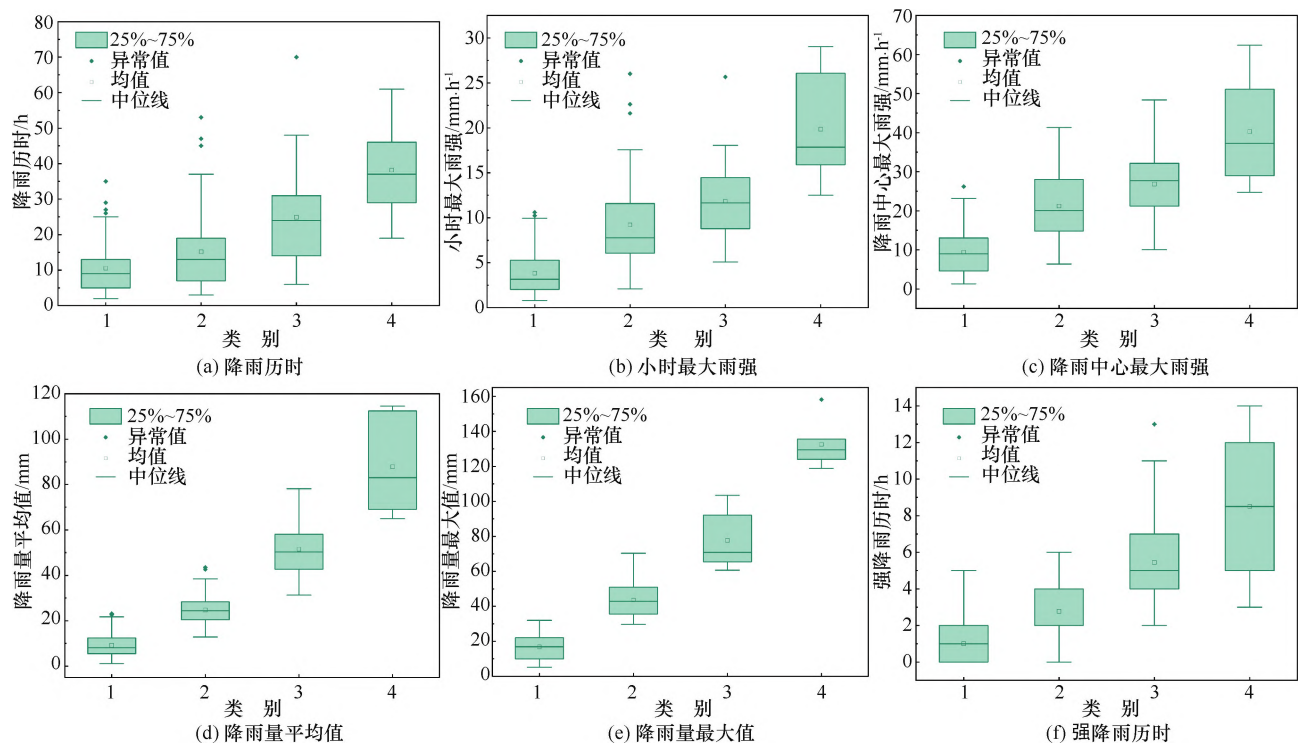


图 6 降雨聚类变量箱型图

Fig. 6 Box plot of rainfall clustering variables

表 4 三个特征距离的权重系数

Table 4 Weight coefficients of the three feature distances

距离特征	降雨总量距离	降雨中心距离	降雨要素距离
权重系数	0.14	0.35	0.51

降雨对应的距离。

2.4 相似性度量及评级

根据权重系数,也可以得到综合距离矩阵

$$\boldsymbol{M}_{\text{总}} = \omega_1 \boldsymbol{X}_R + \omega_2 \boldsymbol{X}_C + \omega_3 \boldsymbol{X}_E \quad (21)$$

在综合距离矩阵中,每一行或每一列都表示该场降雨与其他所有降雨的综合距离,将其从小到大排列,距离越短即相似度越高,用 1 减去综合距离矩阵中的每一个数值构成一个新矩阵,为综合相似矩阵 \boldsymbol{M}_S 。

经过对 225 场降雨进行距离计算后,得出的综合相似度,如表 5 所列。

表 5 场次降雨综合相似度

Table 5 Comprehensive similarity of rainfall events

降雨场次序号	1	2	3	...	223	224	225
1	1	0.412	0.486		0.478	0.469	0.565
2	0.412	1	0.506		0.588	0.551	0.586
3	0.486	0.506	1		0.497	0.544	0.575
⋮	⋮	⋮	⋮	...	⋮	⋮	⋮
223	0.478	0.588	0.497		1	0.628	0.584
224	0.469	0.551	0.544		0.628	1	0.613
225	0.565	0.586	0.575		0.584	0.613	1

将相似度从小到大进行排列,统计所有相似度的分布情况(见图 7),并提取相应位置的相似度大小用于评级,具体划分如表 6 所列。划分为 4 个等级,即相似度大小位于 $[0.78, 1)$ 之间评级为 A,意味着极为相似;相似度大小位于 $[0.7, 0.78)$ 之间评级为 B,意味着较相似;相似度大小位于 $[0.55, 0.7)$ 之间评级为 C,意味着一般相似;相似度大小位于 $[0, 0.55)$ 之间评级为 D,意味着不相似。

表 6 相似度评级

Table 6 The grading results of similarity

累计百分比/%	相似度	评 级
25	0.46	D
50	0.55	C
95	0.70	B
99	0.78	A

3 结果与讨论

3.1 实验结果

以 33 号“20090730”降雨为例,依据本文所提出的多距离融合的方法进行相似性比较。提取降雨特征并计算与其他降雨的 3 个距离特征,找出其相似降雨。表 7 展示了与 33 号降雨的综合相似度排名前五名及后五名的场次降雨,根据结果显示,33 号降雨与 120 号“20140729”降雨最为相似,综合相似度

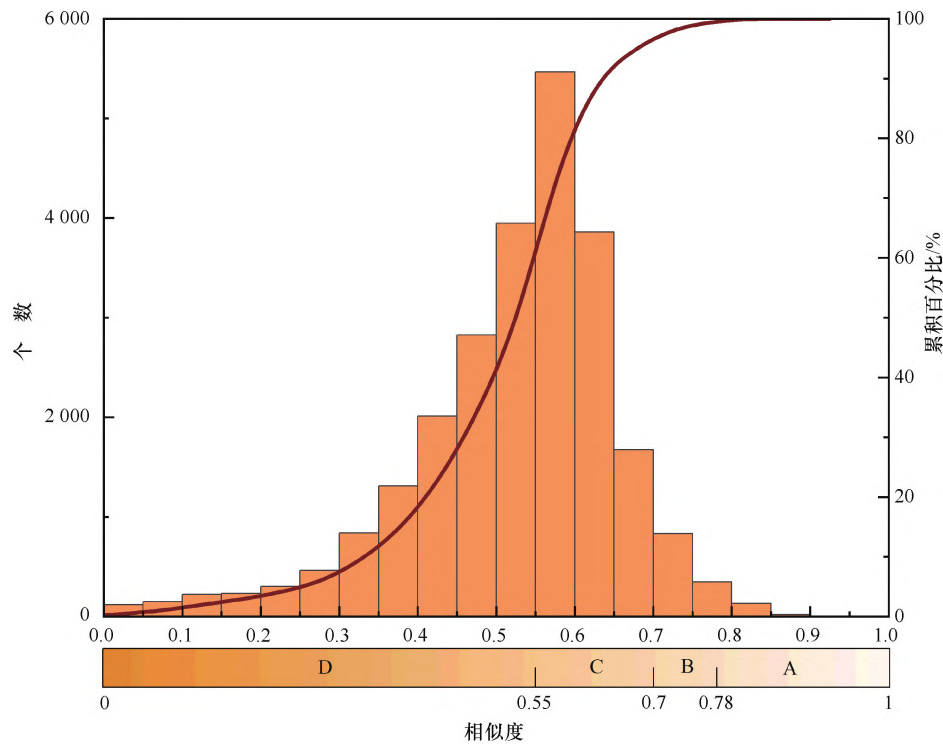


图 7 相似度分布

Fig. 7 The distribution of similarity

表 7 距离计算及相似度排名情况

Table 7 Calculation of distances and ranking of similarity

降雨场次序号	降雨总量距离	降雨中心距离	降雨要素距离	综合距离	综合相似度/%	相似度排名	相似度评级
120	0.084	0.351	0.158	0.215	78.46	1	A
79	0.267	0.327	0.231	0.270	73.04	2	B
66	0.184	0.682	0.088	0.309	69.08	3	C
28	0.169	0.578	0.201	0.329	67.14	4	C
131	0.251	0.427	0.295	0.335	66.51	5	C
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
64	0.298	0.987	0.612	0.699	30.11	220	D
215	0.720	0.933	0.603	0.735	26.48	221	D
178	0.766	0.929	0.633	0.755	24.50	222	D
83	0.991	0.976	0.645	0.809	19.11	223	D
158	0.995	0.962	0.804	0.886	11.37	224	D

为 78.46%；与 158 号降雨相似度最差，仅为 11.37%。相似度评级结果为 A 级 1 个，为 120 号降雨；B 级一个，为 79 号降雨；C 级 49 个；D 级 173 个。

选取最为相似的 120 号降雨与其进行对比分析。如图 8 所示，展示了两场降雨的降雨分布图，从图中可看出两场降雨分布趋势大致相同，其降雨总量距离为 0.084。图 9 展示了两场降雨的降雨中心范围及重叠情况，降雨主要分布在研究区中心处，两场降雨中心距离 0.351，表明其中心重叠度达到 64.9%。表 8 所列为综合相似度排名前五的场次降雨要素特征值，

可以直观对比降雨历时、降雨量、降雨强度等降雨要素。可以看出 120 号降雨与综合相似度最高的 33 号降雨的各降雨要素特征相差不大，其降雨量平均值仅差 1.39 mm。

为了进一步验证该方法的准确性，随机挑选几场相似度评级在 B 级以上的降雨，并用降雨分布图展示其相似性对比情况(见图 10)。

3.2 讨论

3.2.1 不同类别相似性对比分析

经过聚类分析将场次降雨划分为 4 个类别，结合相似矩阵，可以评估不同类别相似性的分布情况。为

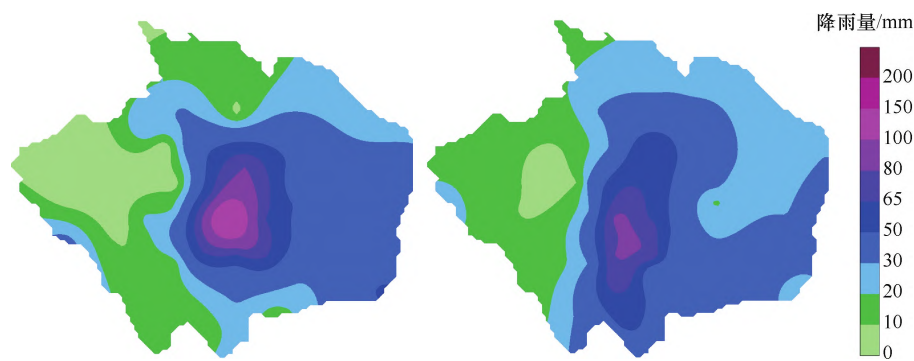


图8 33号(左)与120号(右)降雨分布
Fig. 8 Rainfall distribution of No. 33 (left) and No. 120 (right)

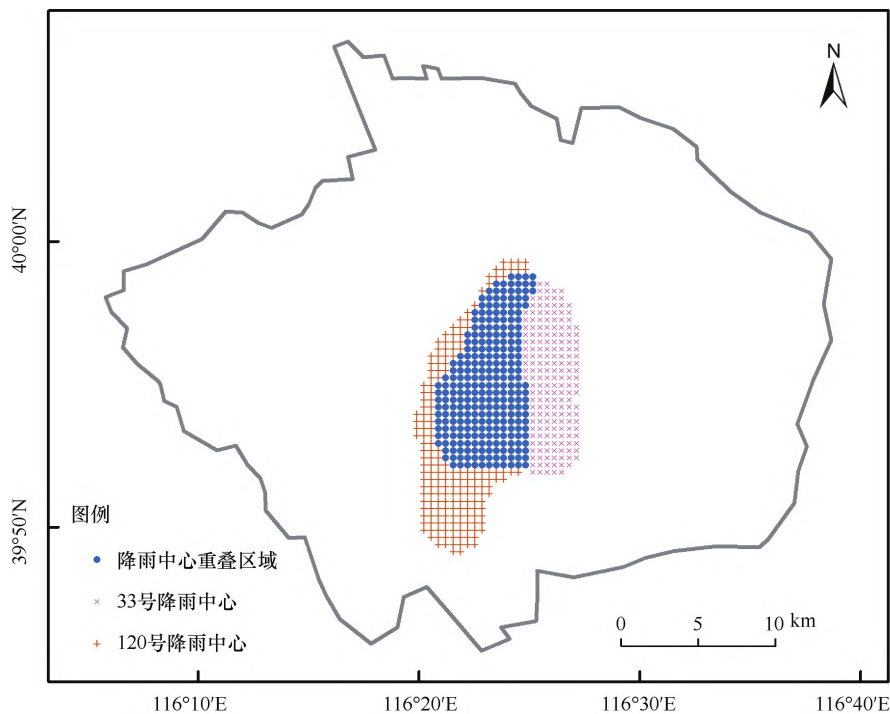


图9 降雨中心范围及重叠区域
Fig. 9 Rainfall centers and overlapping area

表8 排名前五的场次降雨要素特征值

Table 8 Feature values of the top five ranked rainfall events

降雨场次序号	起始时间	结束时间	T	T_t	R_d	R_m	I	I_c	综合相似度/%
33	2009-07-30T14:15	2009-07-31T03:55	14	4	33.73	67.61	11.96	33.20	—
120	2014-07-29T21:20	2014-07-30T08:40	12	5	32.34	57.58	8.77	22.51	78.46
79	2012-06-23T20:05	2012-06-25T04:50	33	6	62.07	79.61	12.85	24.24	73.04
66	2011-07-26T21:35	2011-07-27T02:45	6	4	36.54	67.15	15.71	32.69	69.08
28	2009-07-05T16:45	2009-07-06T10:05	19	1	20.02	31.31	12.77	21.81	67.14
131	2015-06-17T14:25	2015-06-17T20:30	7	1	8.52	17.15	6.37	15.82	66.51

了更直观地反映不同类别的整体相似度水平及差异情况,本文绘制了不同类别相似度分布的直方图(见图11)和箱型图(见图12)。研究发现,随着类别的增加,样本相似度呈现整体减小的趋势,雨量较小的类

别相似度集中分布在较高的范围,而雨量较大的类别相似度分布范围较广,平均值较低,相似度评估效果不理想。

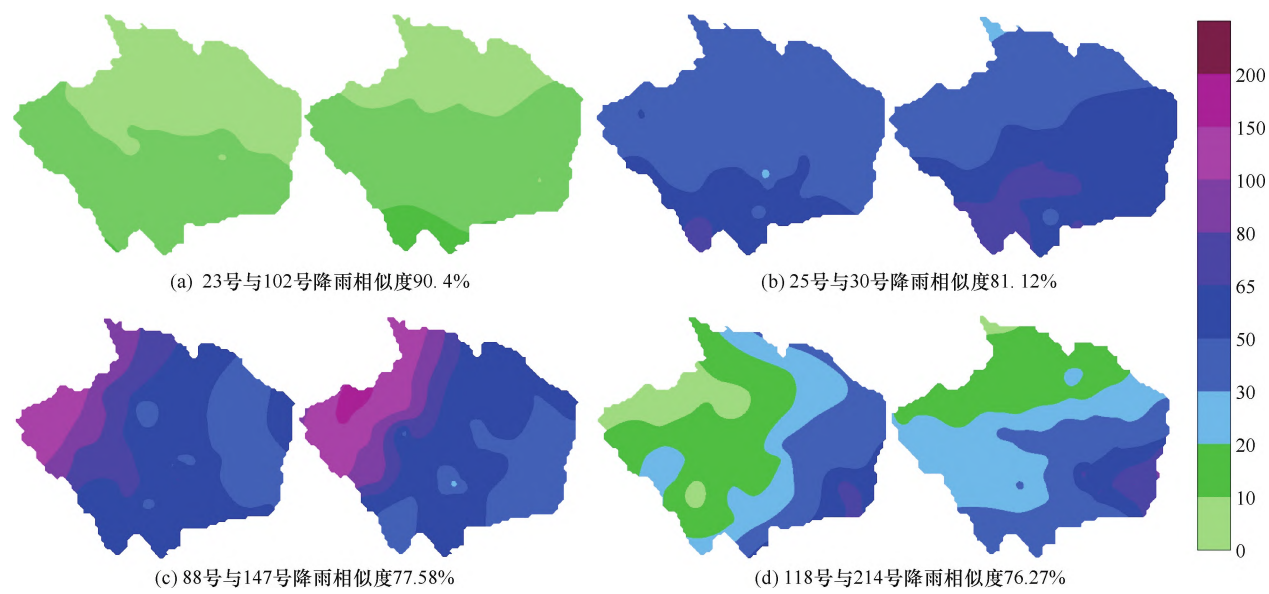


图 10 四场降雨及其相似降雨对比

Fig. 10 Comparison of rainfall distribution of four rainfall events and their similar rainfalls

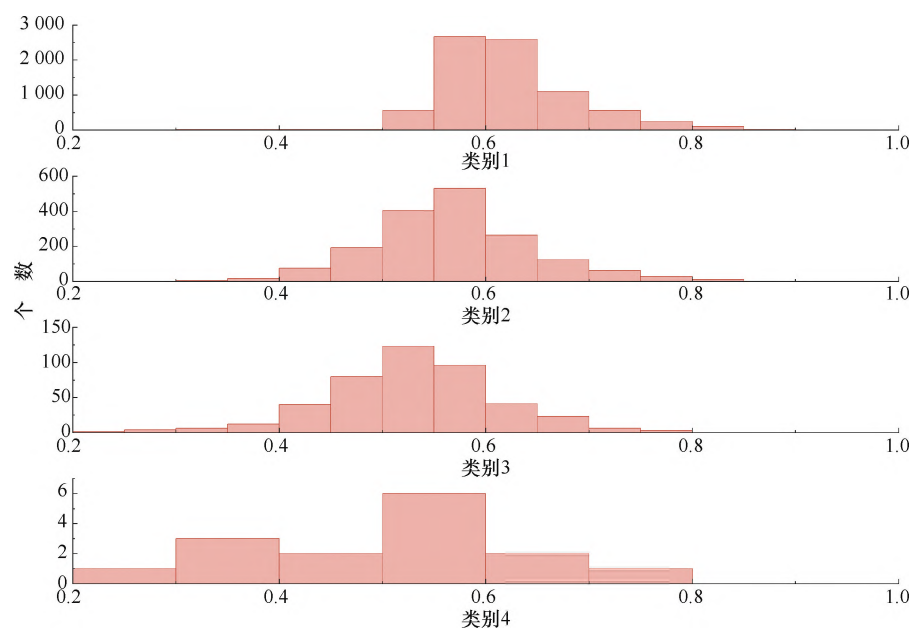


图 11 不同类别降雨相似度分布直方图

Fig. 11 Histogram of similarity distribution for rainfall different categories

3.2.2 强降雨相似性比较的适用性问题

强降雨对城市内涝的影响巨大,然而其相似性分析结果不太理想。原因可能有两个方面。

一是降雨数据库样本量不足。在相似性分析中,样本数量的多少直接影响相似度评估的准确性。由于强降雨和特大暴雨等极端事件的频发概率较低,该类样本数量相对较少。本文类别 3 和类别 4 的样本量仅占样本总数的 16%,导致了这两个类别的相似度较低。在之后的研究中需要扩大降雨库规模,以提高相

似性分析的准确性。XYU 等^[27]在进行北京城区洪涝模型模拟时,设计了多种降雨方案,丰富了北京城区暴雨-内涝情景库。

二是强降雨的降雨过程更加复杂。与小雨中雨相比,强降雨和极端暴雨过程的内在物理机制更为复杂多变,导致同一类别内个体间差异大,相似度评估效果不好。今后可以增加降雨特征要素,如降雨雨型、降雨的时空变化率等,以更全面地描述强降雨事件,提高相似性分析的敏感性和准确性。同时,在今后研

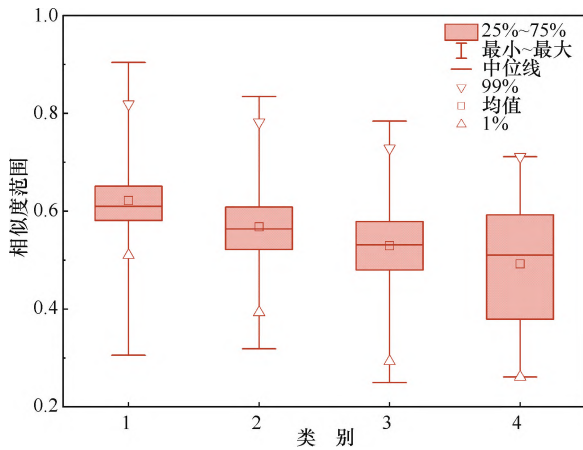


图 12 不同类别降雨相似度分布箱型图

Fig. 12 Boxplot of similarity distribution for different rainfall categories

究中,可以增加特征变量,应用机器学习方法,实现模型的自主学习和优化,提高相似匹配能力,从而提高该方法在城市洪涝预测与防治中的应用价值。刘媛媛等^[28]将机器学习算法引入到降雨时空分布特征的研究中,实现了暴雨时空分布的特征提取。

3.2.3 山区流域中相似性分析方法的适用性问题

在山区流域,降雨站点密度一般较低,导致观测数据相对稀缺。这可能影响插值结果的准确性,进而影响相似性分析的结果。因此,在应用方法前,需要充分考虑数据采集的局限性,通过合理的插值方法来填补数据空缺,以获得更准确的场次降雨数据^[29]。

山区流域具有复杂的地形和地貌,导致降雨分布和特征与城区存在显著差异。山区降雨常呈现出聚集性,高强度等特点,极易引发山洪等灾害。多距离融合相似性分析方法的优势在于能够综合多个降雨特征。因此在山区流域中也可以灵活引入降雨聚集性指标、地形起伏度等指标,以更好地反映山区降雨的变化。

4 结 论

本文通过收集北京市中心城区 2008—2021 年降雨资料,经过降雨场次划分、降雨特征提取、降雨距离计算,提出了一种基于多距离融合的场次降雨相似性比较方法,并对其中两场降雨进行实例分析,得出如下结论。

(1)以往研究多以日降雨数据作为基础,而场次降雨数据更能反映实际的降雨特征,本文将最小时间间隔定为 6 h 进行降雨场次划分,符合实际降雨情况。

(2)城市洪涝的具体情况受雨强、暴雨中心、降雨量、降雨历时等降雨时空特征影响较大。本文通过生成的降雨栅格数据,利用机器学习算法提取计算降雨特征,基于主成分分析确定不同距离的权重,实现了对降雨总量、中心范围和降雨要素的综合考虑,提高了相似性评估的全面性。

(3)经过对所有场次降雨相似度大小统计分析,将相似度划分为 A、B、C、D 四个等级,能够定量判定不同场次降雨的相似程度。其中相似度大小位于 0.78 以上,级别为 A,代表极为相似;相似度大小位于 0.55 以下,级别为 D,代表不相似。

(4)以 33 号 20090730 降雨为例,经过相似性分析得出其与 120 号降雨最为相似,综合相似度为 78.46%,评级为 A,并通过降雨分布图、降雨中心范围图及降雨特征要素等直观对比,验证了结果的准确性及可靠性。

本文为场次降雨相似性分析提供了一种全新的思路和方法,能够准确地评估场次降雨之间的相似度,为城市洪涝预警和应对提供帮助。本文的方法只在城市区域进行了验证,针对山区流域,由于降雨空间变化更为复杂,需要更加谨慎地进行探索和研究。未来的研究方向可聚焦在动态匹配方面,即根据实时降雨预报情况进行动态匹配,提前预知预测城市洪涝情况,提高预警的准确性和时效性。

参考文献 (References):

- [1] 湛舟颖,孔锋. 河南郑州“7·20”特大暴雨洪涝灾害应急管理碎片化及综合治理研究[J]. 水利水电技术(中英文), 2022, 53(8): 1-14.
CHEN Zhouying, KONG Feng. Study on fragmentation of emergency management during “7·20” extreme rainstorm flood disaster in Zhengzhou of Henan Province and relevant comprehensive treatment [J]. Water Resources and Hydropower Engineering, 2022, 53(8): 1-14.
- [2] 任汉承,刘舒,谢每超,等. 基于精细化水动力模型的城市内涝特征分析及风险评估[J]. 水利水电技术(中英文), 2022, 53(4): 24-40.
REN Hancheng, LIU Shu, XIE Meichao, et al. Refined hydrodynamic model-based analysis on urban waterlogging features and risk assessment [J]. Water Resources and Hydropower Engineering, 2022, 53(4): 24-40.
- [3] ZHANG Z, JIAN X, CHEN Y, et al. Urban waterlogging prediction and risk analysis based on rainfall time series features: A case study of Shenzhen[J]. Frontiers in Environmental Science, 2023, 11.
- [4] LIU Y, LI L, LIU Y, et al. Dynamic spatial-temporal precipitation

- distribution models for short-duration rainstorms in Shenzhen, China based on machine learning[J]. Atmospheric Research, 2020, 237.
- [5] 袁定波, 艾萍, 熊传圣. 水文时空数据挖掘方法及其应用评述[J]. 水利信息化, 2018(1): 14-17.
YUAN Dingbo, AI Ping, XIONG Chuansheng. A review of hydrological spatio-temporal data mining methods and its applications [J]. Water Resources Informatization, 2018(1): 14-17.
- [6] 汤静静, 王文圣, 万阳, 等. 降雨相似性评价的集对分析法[J]. 成都工业学院学报, 2014, 17(2): 65-68.
TANG Jingjing, WANG Wensheng, WAN Yang, et al. Assessment of rainstorm similarity based on set pair analysis method [J]. Journal of Chengdu Technological University, 2014, 17(2): 65-68.
- [7] 陈建, 李春红, 王峰, 等. 长江上游流域暴雨洪水相似性判别指标研究[J]. 人民长江, 2019, 50(1): 58-63.
CHEN Jian, LI Chunhong, WANG Feng, et al. Study on similarity discriminant index of storm flood in upper reaches of Yangtze River [J]. Yangtze River, 2019, 50(1): 58-63.
- [8] XIAO T, SHEN T, JING F, et al. K-means cluster's classification of grade of regional strong precipitation Events[J]. Proceedings of the Fifth Symposium of Risk Analysis and Risk Management in Western China (WRARM 2017), 2017, 152: 268-274.
- [9] AUNG S S, OHSAWA S, NAGAYAMA I, et al. An attempt to forecast all different rainfall series by dynamic programming approach [J]. Proceedings of the Future Technologies Conference (FTC) 2018, VOL 1, 2019, 880: 530-547.
- [10] OUYANG R, REN L, CHENG W, et al. Similarity search and pattern discovery in hydrological time series data mining[J]. Hydrological Processes, 2010, 24(9): 1198-1210.
- [11] 赵圆元, 张行南, 黄钰凯, 等. 基于动态权重的水文时间序列相似性分析方法[J]. 水电能源科学, 2021, 39(2): 9-12.
ZHAO Yuanyuan, ZHANG Xingnan, HUANG Yukai, et al. Similarity analysis method of hydrological time series based on dynamic weight [J]. Water Resources and Power, 2021, 39(2): 9-12.
- [12] ZHANGLING X, ZHONGMIN L, BINQUAN L, et al. New flood early warning and forecasting method based on similarity theory[J]. Journal of Hydrologic Engineering, 2019, 24(8).
- [13] 钱磊, 邱学兴, 郑淋淋, 等. 基于降水空间分布相似的最优集成降水预报及其检验[J]. 暴雨灾害, 2022, 41(3): 324-335.
QIAN Lei, QIU Xuexing, ZHENG Linlin, et al. Prediction and test of optimal integrated precipitation based on similar spatial distribution of precipitation [J]. Torrential Rain and Disasters, 2022, 41(3): 324-335.
- [14] YU Y, HE X, ZHU Y, et al. Rainfall similarity search based on deep learning by using precipitation images [J]. Applied Sciences-basel, 2023, 13(8).
- [15] 丁凯熙, 张利平, 宋晓猛, 等. 北京地区汛期降雨时空演变特征及城市化影响研究[J]. 地理科学进展, 2019, 38(12): 1917-1932.
- DING Kaixi, ZHANG Liping, SONG Xiaomeng, et al. Temporal and spatial features of precipitation and impact of urbanization on precipitation characteristics in flood season in Beijing [J]. Progress in Geography, 2019, 38(12): 1917-1932.
- [16] 郑鑫源, 王逸辰, 陈驰. 永定河流域水库资源一体化保护与利用研究[J]. 水利发展研究, 2023, 23(5): 57-61.
ZHENG Xinyuan, WANG Yichen, CHEN Chi. Research on Integrated protection and utilization of reservoir resources in the Yongding River Basin [J]. Water Resources Development Research, 2023, 23(5): 57-61.
- [17] 梁世强, 杜龙刚, 刘晨阳, 等. 2020年北京市汛期雨水情分析[J]. 北京水务, 2020(6): 54-57, 62.
LIANG Shiqiang, DU Longgang, LIU Chenyang, et al. Analysis of rainfall in flood season in Beijing in 2020[J]. Beijing Water, 2020(6): 54-57, 62.
- [18] 皋云, 李琼芳, 周正模, 等. 镇江市降雨场次划分与暴雨特性分析[J]. 水资源保护, 2021, 37(3): 43-49.
GAO Yun, LI Qiongfang, ZHOU Zhengmo, et al. Division of rainfall events and analysis of rainstorm characteristics in Zhenjiang City [J]. Water Resources Protection, 2021, 37(3): 43-49.
- [19] 张宇航, 杨默远, 潘兴瑶, 等. 降雨场次划分方法对降雨控制率的影响分析[J]. 中国给水排水, 2019, 35(13): 122-127.
ZHANG Yuhang, YANG Moyuan, PAN Xingyao, et al. Influence of rainfall division method on capture ratio of rainfall [J]. China Water & Wastewater, 2019, 35(13): 122-127.
- [20] 李哲, 周毅, 吴丽珍, 等. 以径流污染为指标的降雨重要性分类[J]. 中国给水排水, 2021, 37(9): 110-115.
LI Zhe, ZHOU Yi, WU Lizhen, et al. Rainfall importance classification based on runoff pollution [J]. China Water & Wastewater, 2021, 37(9): 110-115.
- [21] WANG X, MUEEN A, DING H, et al. Experimental comparison of representation methods and distance measures for time series data [J]. Data Mining and Knowledge Discovery, 2013, 26(2): 275-309.
- [22] 钱堃, 包为民, 李偲松, 等. K均值聚类分析方法在洪水预报中的应用[J]. 水电能源科学, 2012, 30(5): 41-44.
QIAN Kun, BAO Weimin, LI Caisong, et al. Application of K-means cluster analysis in flood forecasting [J]. Water Resources and Power, 2012, 30(5): 41-44.
- [23] 黄冬梅, 郑霞, 赵丹枫, 等. 一种多元台风时间序列的相似性度量方法[J]. 计算机应用研究, 2019, 36(3): 844-850.
HUANG Dongmei, ZHENG Xia, ZHAO Danfeng, et al. Similarity measure method of multiple typhoon time series [J]. Application Research of Computers, 2019, 36(3): 844-850.
- [24] 余宇峰, 贺新固, 万定生, 等. 基于图像特征和深度学习的相似场次降雨模式库构建方法: 中国, 116010795B[P]. 2023-06-

09.

YU Yufeng, HE Xingu, WAN Dingsheng, et al. Construction method of similar rainfall pattern database based on image feature and deep learning: China, 116010795B[P]. 2023-06-09.

- [25] 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.

- [26] 孙继松, 雷蕾, 于波, 等. 近 10 年北京地区极端暴雨事件的基本特征[J]. 气象学报, 2015(4): 609-623.

SUN Jisong, LEI Lei, YU Bo, et al. The fundamental features of the extreme severe rain events in the recent 10 years in the Beijing area [J]. Acta Meteorologica Sinica, 2015(4): 609-623.

- [27] XYU M, YAO Y, LIU S, et al. Multi-mode surface generalization

supports a detailed urban flooding simulation model [J]. Frontiers in Earth Science, 2021, 9.

- [28] 刘媛媛, 刘洪伟, 霍风霖, 等. 基于机器学习短历时暴雨时空分布规律研究[J]. 水利学报, 2019, 50(6): 773-779.

LIU Yuanyuan, LIU Hongwei, HUO Fenglin, et al. An application of machine learning on examining spatial and temporal distribution of short duration rainstorm [J]. Journal of Hydraulic Engineering, 2019, 50(6): 773-779.

- [29] 郎立晨, 唐诚, 高星, 等. 复杂地形下降降水的高空间分辨率插值方法研究[J]. 地理学报, 2023, 78(1): 101-120.

LANG Lichen, TANG Cheng, GAO Xing, et al. Spatial interpolation of high-resolution daily precipitation over complex terrains [J]. Acta Geographica Sinica, 2023, 78(1): 101-120.

(责任编辑 康 健)