

doi:10.3969/j.issn.1000-6362.2023.03.006

司丽丽,赵亮,魏铁鑫,等.多源强降雨灾情可信度智能判别方法[J].中国农业气象,2023,44(3):228-237

## 多源强降雨灾情可信度智能判别方法\*

司丽丽<sup>1,2,3</sup>, 赵亮<sup>1,2,3</sup>, 魏铁鑫<sup>1,2,3</sup>, 霍治国<sup>4\*\*</sup>, 李姣<sup>1,2,3</sup>

(1. 河北省气象与生态环境重点实验室, 石家庄 050021; 2. 中国气象局雄安大气边界层重点开放实验室, 雄安新区 071800;  
3. 河北省气象灾害防御和环境气象中心, 石家庄 050021; 4. 中国气象科学研究院, 北京 100081)

**摘要:** 真实的灾情信息是有效防范和减轻强降雨灾害损失的重要参考。本研究以过程降雨强度(R)为指标, 构建1984–2020年河北省县级多源气象灾情与致灾过程相匹配的强降雨灾害事件库。经过人工质控, 获得真实灾情信息2305组, 伪灾情信息263组。采用相关分析法确定与灾害发生程度(灾度)显著相关的降雨关键特征因子, 基于单类支持向量机和十折交叉检验法, 随机抽取10次样本, 建立强降雨灾情气象因子致灾判别模型, 并进行检验优化, 以探索智能化、易用性的多源强降雨灾情可信度智能判别方法。结果表明:(1)与灾度显著相关的降雨关键特征因子共计11个, 分别为最大降雨量、最小降雨量、过程平均降雨量、日均降雨量、平均小时雨强、1h最大雨量、3h最大雨量、6h最大雨量、12h最大雨量、24h最大雨量及前10日降雨总量, 均通过了0.01水平显著性检验。(2)采用11个因子建立10个致灾判别模型(M1–M10), 依据真实灾情判别准确率确定最优模型为M9, 其证真率为96.4%, 证伪率为67.6%, 表明该模型对灾情真伪判定较为片面, 应进一步优化。(3)通过自相关检验, 以最大降雨量、平均小时雨强、1h最大雨量及前10日降雨总量4个因子作为输入因子, 重新构建强降雨灾情气象因子致灾判别模型(M11–M20), 最优模型为M20, 其证真率和证伪率分别达到96.2%和82.9%。综合分析认为, 由4个因子构建的气象因子致灾判别模型评估强降雨灾情可信度比11个因子建立的模型更可靠。

**关键词:** 强降雨灾害; 支持向量机; 相关分析; 致灾因子; 可信度判别

## An Intelligent Method for Discriminating the Reliability of Multi-Source Heavy Rainfall Disaster Information

SI Li-li<sup>1,2,3</sup>, ZHAO Liang<sup>1,2,3</sup>, WEI Tie-xin<sup>1,2,3</sup>, HUO Zhi-guo<sup>4</sup>, LI Jiao<sup>1,2,3</sup>

(1. Key Laboratory of Meteorology and Ecological Environment of Hebei Province, Shijiazhuang 050021, China; 2. China Meteorological Administration Xiong'an Atmospheric Boundary Layer Key Laboratory, Xiong'an New Area 071800; 3. Hebei Meteorological Disaster Prevention and Environment Meteorology Center, Shijiazhuang 050021; 4. Chinese Academy of Meteorological Sciences, Beijing 100081)

**Abstract:** The real disaster information is an important reference for preventing and reducing heavy rainfall disaster losses effectively. Taking process rainfall intensity as the index, this paper constructs a heavy rainfall disaster event database matching the multi-source meteorological disasters and disaster-causing processes in Hebei Province during 1984–2020. After manual quality control, 2305 groups of real disaster information and 263 groups of false information are obtained. In this study, correlation analysis was used to determine and select the rainfall eigenfactors that are significantly related to the disaster degree. Based on the One-class support vector machine (OCSVM) and 10

\* 收稿日期: 2022-01-16

基金项目: 河北省气象局科研开发项目(19ky06); 2022年山洪地质灾害防治气象保障工程建设项目

\*\* 通讯作者: 霍治国, 二级研究员, 主要从事农业气象灾害预测与评估研究, E-mail: huozg@cma.gov.cn  
第一作者联系方式: 司丽丽, E-mail: sll\_0312@163.com

folds cross-validation method, 10 samples were randomly selected to establish the meteorological factor disaster discriminant model and test and optimize, so as to explore the intelligent and easy-to-use intelligent discriminant method of multi-source heavy rainfall disaster credibility. The results showed that: (1) there are eleven rainfall eigenfactors related to disaster degree at 0.01 significant level, which are maximum rainfall, minimum rainfall, average process rainfall, average daily rainfall, average hourly rainfall intensity, hourly maximum rainfall, 3-hour maximum rainfall, 6-hour maximum rainfall, 12-hour maximum rainfall, 24-hour maximum rainfall and rainfall in the first 10 days. (2) Ten models (M1-M10) were established using 11 rainfall eigenfactors. According to the identification accuracy of the real disaster, the optimal model was determined to be M9, with the authenticity rate of 96.4% and the falsification rate of 67.6%, which indicated that the model is one-sided in determining the authenticity of the disaster situation and should be further optimized. (3) Through the autocorrelation test, maximum rainfall, average hourly rainfall intensity, hourly maximum rainfall and rainfall in the first 10 days were taken as input factors. Ten models (M11-M20) were reconstructed, and the optimal model is M20 with 96.2% proof rate and 82.9% false rate. Based on comprehensive analysis, the model established by 4 factors is more reliable than the model established by 11 factors.

**Key words:** Heavy rainfall disaster; OCSVM; Correlation analysis; Disaster-causing factor; Reliability discriminant

真实准确的灾情数据有利于政府部门掌握灾害发生发展动态趋势,是有效开展灾害风险管理和减灾救灾的基础依据<sup>[1]</sup>,这就要求做到快速、精准地甄别强降雨灾情的真实性,提供真实可靠、广覆盖的灾情信息<sup>[2]</sup>。以往的灾情获取模式主要为人工调查逐级上报,较为单一。但随着3S、图像拍摄、互联网等技术的发展,灾情数据的来源愈加丰富<sup>[3-4]</sup>,气象灾情监测采集模式由传统的单一来源向多种来源转变,河北省气象灾害防御和环境气象中心建立了基于“文献史料+部门共享+互联网+社会化”的多源气象灾情采集业务,使得灾情数据向海量化发展,这为全方位灾情监测、预警评估、风险区划提供了数据基础,但同时因其繁杂性、多样性,呈现出愈加明显的质量低、可信度低等特点<sup>[5]</sup>。伪灾情的出现将影响对灾害发展趋势的研判,从而降低抗灾救灾决策的科学性,制约气象灾害防御进展。因此,如何快速、有效、准确地开展灾情数据可靠性判别就显得至关重要。

目前国内气象灾情信息鉴定工作主要通过人力资源开展<sup>[6]</sup>,以咨询相关部门、现场调查、查阅相关资料等方式进行主观鉴定,存在成本高、效率低等问题。因此,针对灾害风险管理和灾情信息鉴定的迫切需求,很有必要提供更为高效、精准的判别方法。当前有关气象灾情可信度判别的相关研究仅在雷电灾害方面有少量研究报道,何文革等对一般雷电灾害的调查鉴定方法进行了探讨<sup>[7]</sup>,吴孟恒等以雷

电灾害调查与鉴定实践经验为基础,论述了雷电灾害调查与鉴定的基本原理、具体程序等<sup>[8]</sup>,这类方法以灾害现场人工核查及资料分析为主。近年来,对数据可信度的研究主要侧重在数据可信度影响因素方面的相关实证分析<sup>[5,9-12]</sup>,李吉顺等利用经验方法推算出部门上报灾情信息中所包含的灾害损失数据的可信度<sup>[5]</sup>,徐绪堪等基于D-S理论对突发事件多源数据可信度进行了评估<sup>[12]</sup>,这些方法适用于对含有直接经济损失等数值型信息进行可信度判别,但当前随着互联网、社会化等多源灾情信息的出现,非数值型灾情信息占比更大,因此,开展不依赖于信息中所含数值的灾情信息可靠性判别研究非常必要。

强降雨灾害是中国发生频率最高、累积影响最重、社会影响面最广的气象灾害<sup>[13]</sup>,仅河北省“16.7”降雨过程即造成直接经济损失约574.57亿元<sup>[14]</sup>。而农作物往往是该类灾害中脆弱性较强、暴露度较高的承灾体,相较于其他承灾体,极端强降水引起的暴雨洪涝灾害对农业生产和产量更是产生直接危害,造成巨大损失<sup>[15]</sup>。鉴于此,本研究以强降雨灾害为研究对象,基于灾害学原理,利用多源气象灾情数据以及同期降雨特征因子,通过相关分析确定建模因子,采用机器学习算法,利用十折交叉验证方式,构建气象因子致灾判别模型,以期探究既能精准判别灾情可靠性,又简便易用的方法,解决依赖人工开展灾情可靠性判别的成本高、效率低等问题,为开展灾害监测与评估提供服务。

# 1 资料与方法

## 1.1 数据来源与处理

### 1.1.1 多源强降雨灾情数据

1984–2020 年河北省县（区）域多源强降雨灾情数据来源于应急管理部门、灾情直报、灾害普查、《中国气象灾害大典》（河北卷）、网络媒体及志愿者（表 1），数据字段包括灾害发生的起始时间、结束时间、影响区域、灾害描述、直接经济损失、受灾人口、死亡人口和信息来源 8 个字段。多种来源的灾情数据互为补充，解决了单一来源灾情覆盖面不足的问题，但也存在灾情数据重复问题。本研究通过分析其重复特征，制定了多源强降雨灾情数据的去重与质控规范和流程，筛选同一灾害过程、不同数据来源的灾情信息，最终保留 1 条作为入库灾情数据。通过整理得到 1984–2020 年强降雨灾情数据 2761 条，再经过人工核对质控，确认真实灾情信息 2498 条，伪灾情信息 263 条，以此构建河北省多源强降雨灾情数据库。

### 1.1.2 强降雨特征数据

尽管形成暴雨洪涝的灾害系统异常复杂，但其致灾因子主要是过强或过于集中的降水<sup>[16–17]</sup>。因此，为了对灾情数据的可靠性精准判别，需要收集同期的降雨数据，统计甄别降雨过程并将降雨过程与灾情数据进行匹配，找到导致灾害发生的强降雨过程，并提取计算可能致灾的关键因子，包括降雨过程的起止时间、降雨量、降雨时长、降雨强度、降雨范围、前期降雨情况等降雨特征数据（表 2）。上述数据来源于河北省气象局 142 个国家地面气象观测站点和 3047 个地面气象自动监测站点，其中，国家地面气象观测站点降雨数据的时间尺度为 1984–2020 年共 37a，由于地面气象自动监测站点建站较晚，故其时间尺度略短，为 2006–2020 年共 15a。

## 1.2 灾情可信度智能判别模型建模过程及方法

### 1.2.1 建模过程

第一步，筛选与灾情相匹配的致灾降雨过程。提出致灾过程匹配方案，并依据方案提取致灾过程，建立供模型构建与检验的灾害事件库。第二步，通过相关分析筛选触发灾害的关键气象因子。采用 Pearson 相关分析法，以县域为单位，匹配降雨特征因子与灾害强度（以下简称灾度）进行相关性分析，初步筛选与灾害发生关系密切的特征因子，作为气象致灾关键因子。第三步，采用单类支持向量机算法（One-Class SVM, OCSVM）及十折交叉验证方式<sup>[18]</sup>，构建并检验气象因子致灾判别模型，最终获得最优判别模型。

### 1.2.2 灾度计算和分级

（1）灾度计算。灾度是对灾害进行评价的最优指标<sup>[19]</sup>，利用灾度数据筛选触发强降雨灾害的关键因子。引入前人研究气象灾害等级所用的灾度公式<sup>[20]</sup>，即

$$G = \lg(50DHE) - 5 \tag{1}$$

式中，G 为灾度，D 为死亡人数，H 为受灾人数，E 为 2020 年可比经济损失（万元），其计算方法为

$$E = SC \tag{2}$$

式中，S 为当年经济损失（万元），C 为物价指数，资料均来自河北省当年度对应县（区）的统计年鉴。式（1）中，如果 D 或 H 值为 0，对数公式则无意义，因此将出现 0 值时的 D 和 H 均赋值为 1 再行计算。

（2）灾度分级。采用最优分割法对灾害强度进行特强、强、中等和低 4 个等级划分。最优分割法是 Fisher 在 1958 年最先提出的，该方法能够不破坏原有样本，通过计算分级后，组内离差平方和最小，但不同组间离差平方和最大，是对有序样本或可变为有序样本进行科学分级的一种聚类分级方法<sup>[21]</sup>。

表 1 多源强降雨灾情数据来源和数量（条）

Table 1 Source and quantity of multi-source heavy rainfall disaster information (piece)

数量 Quantity	应急管理部门灾 情报告 Reports of disaster from the emergency department	气象部门灾情直 报 Reports of disaster from the meteorological department	气象部门灾害普 查 Disaster survey from meteorological department	《中国气象灾害大典》 （河北卷）China Meteorological Disaster Ceremony(Hebei Volume)	网络媒体 Networking medium	志愿者 Volunteer
	317	360	1341	178	524	41

表 2 与致灾降雨过程有关的特征因子统计

Table 2 Eigenfactor statistics related to one-time rainfall process

类别 Categories	序号 Order	降雨特征因子 Eigenfactor of rainfall	说明 Caption
降雨时间 Rainfall time	1	开始时间(年-月-日 时)Start time(YYYY-mm-dd hh)	一次降雨过程中最早出现降雨的时间 The earliest time of rainfall in one rainfall process
	2	结束时间(年-月-日 时)End time(YYYY-mm-dd hh)	一次降雨过程中最晚结束降雨的时间 The latest time to end rainfall in one rainfall process
过程降雨量 Rainfall	3	过程平均降雨量 Average rainfall(mm)	所有站点过程降雨之和除以降雨站数 Total rainfall of all stations divided by rainfall station
	4	最大降雨量 Max rainfall(mm)	一次降雨过程中降雨量最大值 The max. rainfall in all stations in one rainfall process
	5	最小降雨量 Min rainfall(mm)	一次降雨过程中降雨量最小值 The min. rainfall in all stations in one rainfall process
降雨时长 Duration of rainfall	6	日均降雨量 Daily average rainfall(mm·d <sup>-1</sup> )	一次降雨过程中每日降雨的平均值 Daily average rainfall in one rainfall process
	7	持续雨日 Duration(d)	一次降雨过程中降雨持续日数 Duration days in one rainfall process
	8	持续雨时 Duration(h)	一次降雨过程中降雨持续小时数 Duration hours in one rainfall process
降雨强度 Rainfall intensity	9	累计雨时 Total time of rainfall(h)	持续雨时与降雨间歇时段的差值 The D-value between duration and rainfall interval
	10	过程降雨强度 Rainfall intensity (mm)	见式(4)Formula (4)
	11	平均小时雨强 Average hourly rainfall intensity (mm)	一次降雨过程中小时降雨的平均值 1-hour average rainfall in one rainfall process
	12	1h最大雨量 1-hour max. rainfall(mm)	逐小时降雨最大值 1-hour max rainfall
	13	3h最大雨量 3-hour max. rainfall(mm)	滑动累计 3h 降雨最大值 Sliding max. 3-hour total rainfall
	14	6h最大雨量 6-hour max. rainfall(mm)	滑动累计 6h 降雨最大值 Sliding max. 6-hour total rainfall
	15	12h 最大雨量 12-hour max. rainfall (mm)	滑动累计 12h 降雨最大值 Sliding max. 12-hour total rainfall
降雨范围 Rainfall range	16	24h 最大雨量 24-hour max. rainfall (mm)	滑动累计 24h 降雨最大值 Sliding max. 24-hour total rainfall
	17	降雨站数 Number of rainfall stations	一次降雨过程中出现降雨的所有气象站数量 Total stations with rainfall in one rainfall process
	18	覆盖范围 Coverage(%)	降雨站数与总站数的百分比 Percentage of rainfall stations to total stations
前期降雨情况 Prophase rainfall	19	前 10 日降雨次数 Rainfall times in the first 10 days	致灾降雨过程开始前 10 日内降雨总次数 The first 10 days rainfall frequency since the beginning of the disaster process
	20	前 10 日降雨累计时长 Rainfall duration in the first 10 days (h)	致灾降雨过程开始前 10 日内降雨时间的累计值 The first 10 days rainfall duration since the beginning of the disaster process
	21	前 10 日降雨总量 Rainfall in the first 10 days (mm)	致灾降雨过程前 10 日内每次降雨过程平均降雨量之和 The first 10 days rainfall since the beginning of the disaster process

注：表中数据均以县为单位，并综合统计县域内所有国家站和区域站。

Note: All the data in the table are based on the county, and all of the national meteorological stations and regional automatic meteorological stations in a county are counted comprehensively.

1.2.3 致灾过程提取

(1) 基本思路。首先界定降雨过程起止时间，再以灾情开始时间及发生县域为初始条件，筛选可

能致灾的降雨过程，通过计算提取降雨强度最强过程作为致灾降雨过程。

(2) 降雨过程界定。对于强降雨灾害，尤其是

城市内涝、洪水、滑坡等灾害事件而言，往往是一场雨造成的，“场雨”对于自然灾害研究更具意义。因此，本研究引入目前国际上常用的最小降雨间歇 (Minimum Inter-event Time, MIT)法划分降雨过程，Melillo 等根据统计情况，在干旱区将春夏时期发生的前后 6h 未检测到雨情的降雨过程，划分成一场降雨事件<sup>[22]</sup>。中国较常用的 MIT 值为 2h 和 6h。李雨欣等研究认为将 6h 作为 MIT 值进行河北省降雨划分，符合河北省降雨特征，是准确可行的<sup>[23]</sup>。本研究借鉴其研究结果，选取 6h 作为两次降雨过程的时间间隔，对河北省各个站点降雨数据进行计算分析，进而统计每次降雨过程的降雨特征因子 (表 2)。上述因子充分描述了区域内一次降雨过程的时空变化特征，为避免各因子间的量级差异<sup>[24]</sup>，对因子进行了归一化处理，即

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3)$$

式中， $x_{\min}$  为各因子序列最小值， $x_{\max}$  为各因子序列最大值。

(3) 致灾过程确定。考虑到强降雨灾害相对降雨过程存在一定的滞后性，因此，在确定致灾降雨过程时，将自灾害开始当日及其前 2 日作为致灾过程的选择时段，对所有站点降雨数据进行筛选。依据降雨过程界定方法查找该时段内的降雨过程，并计算相应的过程降雨强度 (R)，该指标可综合体现降雨致灾能力，因此，选择降雨强度最大的降雨过程作为致灾降雨过程，计算该过程的降雨特征因子值作为建模待用数据。

其中，过程降雨强度 (R) 选用了气象行业标准《降雨过程强度等级》<sup>[25]</sup>推荐的公式进行计算。由于河北省本地强降雨致灾特点为短时强降雨略多于过程强降雨，因此，将原有小时最大雨强权重由 0.5 调至 0.65，为使该强度计算更加精细化，将公式中日降雨量值替换为小时降雨量。调整后的过程降雨强度计算式为

$$R = \frac{\omega \sum_{i=1}^n (r_{\max})_i + (1 - \omega) \sum_{i=1}^n \left( \frac{\sum_{d=1}^{T_0} r_d}{T_0} \right)_i}{n} \quad (4)$$

式中， $\omega$  代表权重，取 0.65； $n$  代表区域内出现降雨的测站个数， $i$  取值范围 $[1, n]$ ； $r_{\max}$  代表区域内测站小时降雨量最大值 (mm)； $T_0$  代表县域降雨过

程持续时间 (h)， $d$  取值范围 $[1, T_0]$ ； $r_d$  代表县域内测站的过程平均降雨量 (mm)。

灾情数据与致灾过程的匹配结果表明，193 条数据为无对应致灾过程的空过程灾情，无法形成一组完整的灾害案例，将其剔除，最终用于后续研究的真实强降雨灾害案例数据为 2305 条。

### 1.2.4 单类支持向量机

单类支持向量机 (OCSVM) 是一种无监督学习算法，由 Scholkopf 在传统支持向量机算法的基础上发展而来<sup>[26]</sup>。其原理是将所有输入样本作为目标样本，通过核函数将输入样本映射到高维空间中，使得存在一个最优超球体，超球体尽可能将目标样本包含其中，而将非目标样本排除在超球体外，同时最小化该超球体的体积。由于 OCSVM 算法仅需单类样本，即可通过对其训练得到二分类模型，因此适用于解决样本数量分布不均匀时训练检测分类器遇到的问题<sup>[27]</sup>。本研究将该方法应用于构建强降雨灾情致灾因子判别模型。

## 2 结果与分析

### 2.1 强降雨灾情可信度判别的流程设计

对于一条强降雨灾情可信度判别，首先提取该条灾情关键信息，包括灾害发生时间 (精确至小时)、灾害发生地点 (精确至县区)，作为匹配致灾过程的触发因子；然后，按照降雨过程界定规则筛选与该条灾情匹配的降雨过程 ( $R_c$ , 1 个或多个)，再计算每个过程降雨强度 (R)，判定导致该次灾害发生的致灾过程，并提取关键降雨特征因子 ( $R_k$ )；最后将致灾过程的  $R_k$  代入强降雨灾情气象因子致灾判别模型，得到灾情可信度的判别结果 (图 1)。

### 2.2 触发灾害的关键降雨特征因子筛选

#### 2.2.1 灾度分级

将灾害损失数据统一为综合指数即灾度，与降雨特征因子进行相关分析，以筛选触发灾害的关键降雨特征因子。

由于不同来源的灾情调查规范不同，部分灾情缺乏社会经济损失数据，无法进行灾度计算，因此，将含有直接经济损失和受灾人口等数据的 725 条灾情信息进行灾度计算及分级。表 3 给出了各等级对应的灾度值区间及灾害发生频次 (县·次数)。总体来看，河北省以中、低强度的强降雨灾害为主，占灾害总频次的 64.7%，这与历史上河北省强降雨灾害实际发生规律基本相符，表明灾度分级较为适宜。

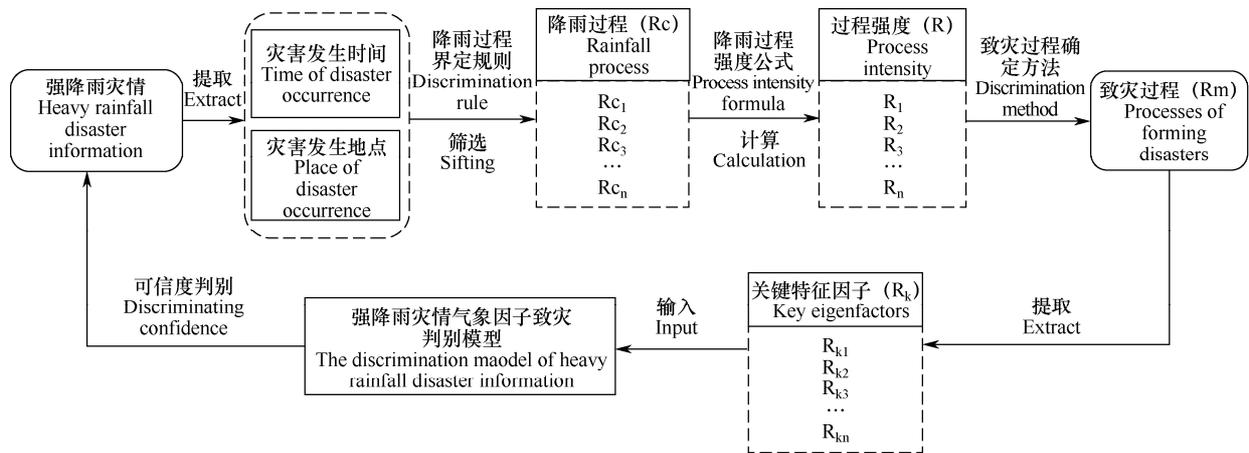


图 1 强降雨灾情可信度判别流程

Fig. 1 Procedure for discriminating the reliability of heavy rainfall disaster information

表 3 强降雨灾度分级阈值及发生次数统计

Table 3 Grading threshold of disaster degree and number of occurrences at each grade

等级 Grade	灾度 Disaster degree	次数(县·次) Times(county·times)	占比 Ratio(%)
特强 Extra strong	(3.82, 7.40]	70	9.7
较强 Stronger	(2.30, 3.82]	186	25.7
中等 Medium	(1.09, 2.30]	247	34.1
低等 Slight	[0, 1.09]	222	30.6

2.2.2 降雨特征因子的筛选

将强降雨灾害案例中过程降雨量、降雨时长、雨强、前期降雨情况 4 类特征因子与相应灾度等级进行相关性分析，最终筛选出相关系数通过了 0.01 水平显著性检验的因子共计 11 个，结果见表 4。由表可见，最大降雨量、最小降雨量、过程平均降雨量、日均降雨量、平均小时雨强、1h 最大雨量、3h 最大雨量、6h 最大雨量、12h 最大雨量、24h 最大雨

量以及前 10 日降雨总量均与灾度级别呈极显著相关，表明过程降雨量、雨强以及前期降雨量 3 类降雨因子对于触发强降雨灾害起到了关键作用，三者配合触发强降雨灾害的发生发展，而代表降雨时长的各特征因子与灾度均无显著相关关系，说明其不是强降雨灾害发生的限定因素。

2.3 气象因子致灾判别模型构建与检验优化

2.3.1 模型构建

采用 OCSVM 及交叉验证的方式构建并检验模型。即在总的样本数据中按比例随机分成两份，其中 1 份用于模型训练，另 1 份用于模型检验，多次按比例随机抽取训练集和检验集，并分别构建模型，通过比较这些模型的准确率，从中选择出最优模型。数据多次抽取相当于进行了多次交叉建模，进行的次数越多，相当于进行交叉建模次数就越多，结果就越具有代表性。对 2305 组真实灾情案例随机抽取

表 4 降雨特征因子与灾度的相关系数

Table 4 Correlation coefficient(R) between the disaster degree and eigenfactors of rainfall

降雨特征因子 Eigenfactors of rainfall	相关系数 Correlation coefficient	降雨特征因子 Eigenfactors of rainfall	相关系数 Correlation coefficient
最大降雨量 Max. rainfall	0.365**	3h 最大雨量 3-hour max. rainfall	0.266**
最小降雨量 Min. rainfall	0.328**	6h 最大雨量 6-hour max. rainfall	0.321**
过程平均降雨量 Average rainfall	0.415**	12h 最大雨量 12-hour max. rainfall	0.366**
日均降雨量 Daily mean rainfall	0.348**	24h 最大雨量 24-hour max. rainfall	0.388**
平均小时雨强 Average hourly rainfall intensity	0.137**	前 10 日降雨总量 Rainfall in the first 10 days	0.291**
1h 最大雨量 1-hour max. rainfall	0.162**		

注：\*、\*\*分别表示相关系数通过 0.05、0.01 水平的显著性检验。下同。

Note: \* is P<0.05, and \*\* is P<0.01. The same as below.

90%样本作为训练集，10%样本作为检验集，共随机抽取 10 次建立 10 个 OCSVM 模型 (MCDDM, 以下简称 M)。建模流程见图 2。

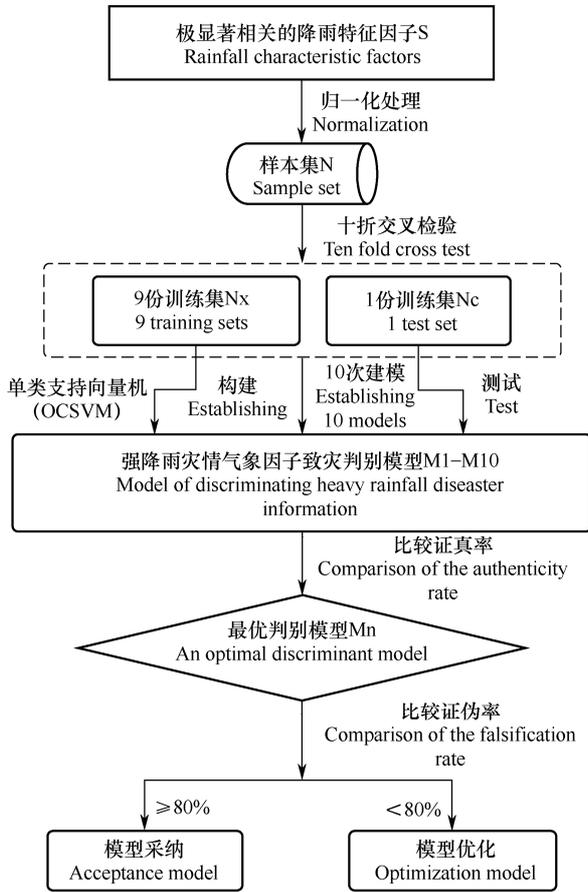


图 2 致灾判别模型构建流程图

Fig. 2 Procedure for establishing the model of discriminating disaster

2.3.2 模型检验

将随机抽取的 10%测试样本分别输入 10 个 OCSVM 模型中进行检验，选取最优模型，为保证伪灾情不被误识为真实灾情，将 263 条伪灾情作为独立测试集对最优模型进行检验，以准确率 80%为采纳标准，低于 80%则需对模型优化，结果见图 3。由图可见，M1-M10 的判别准确率均在 90%以上，其中，M9 的准确率最高，为 96.4%。由于该模型的验证集仅为单一的真实灾情（正样本）数据，缺乏对伪灾情（负样本）数据进行验证，因此，将 263 组经人工质控的伪灾情数据进行二次验证，准确率为 67.6%，综合准确率可达 82%。由此可见，尽管该模型对于真实灾情的判别较好，但对于伪灾情判别能力较低，因此，模型仍需进一步优化完善。

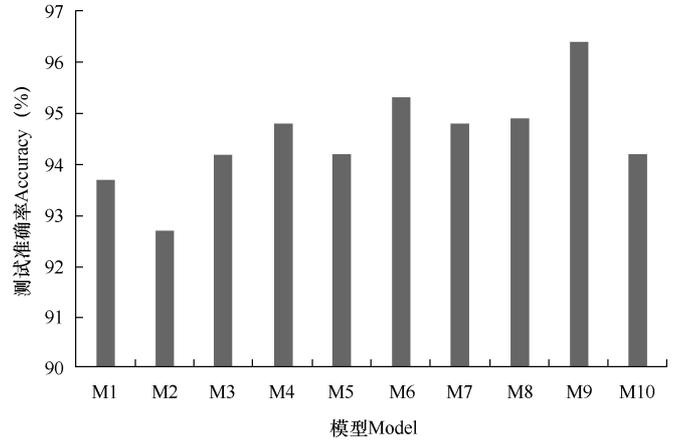


图 3 1984-2020 年真实灾情信息十折交叉检验

Fig. 3 10-fold cross-validation of the real disaster information in 1984-2020

2.3.3 模型优化

分析自变量数据发现，与灾度相关显著的因子中，最大降雨量、最小降雨量、日均降雨量、过程平均降雨量 4 种降雨特征因子分别反映了完整过程的降雨情况，均属于过程降雨因子，自相关检验存在线性相关关系，这也是影响模型的稳定性和效力的关键<sup>[28]</sup>。因此，通过比较相关系数，结合专家经验，最终选择最大降雨量作为建模因子之一。其他因子的选择参照过程降雨相关因子的筛选模式，采用两因子相关检验，结合“系数最大原则”及河北本地气候特点，确定了平均小时雨强、1h 最大雨量及前 10 日降雨总量作为其它建模因子。其中，平均小时雨强与 1h 最大雨量的相关系数相对较小，且代表含义不同，因此，亦将其作为建模因子；各时段最大雨强中尽管 24h 最大雨强与灾度的相关性最强，但考虑河北本地强降雨灾情特点，致灾降雨过程中仍有大量短时强降雨，是致灾的重要因子，因此，选择 1h 最大雨量作为雨强要素的代表因子。最终确定最大降雨量、平均小时雨强、1h 最大雨量及前 10 日降雨总量 4 个特征要素作为模型构建的输入因子。

将优化筛选的降雨特征因子与灾度重新构建模型 M11-M20，检验可见，真实灾情判别准确率最高的为 M20（表 5），准确率为 96.3%，较 M9 的准确率低 0.2 个百分点，但伪灾情判别准确率达 82.9%，较 M9 提升 15.3 个百分点，综合准确率可达 89.6%，提升 7.6 个百分点。对比发现，应用剔除自相关显著的因子后，模型判别能力明显提升，表明该模型更适于评估强降雨灾情可信度。

表5 1984–2020年真实灾情判别准确率

Table 5 The results of discriminating real disaster information in 1984–2020

模型 Models	判别正确(条) Accuracy(piece)	判别错误(条) Error(piece)	判别准确率 Accuracy rate(%)	最优模型 Optimal model(√)
M11	216	15	93.4	
M12	218	13	94.2	
M13	216	15	93.6	
M14	219	12	94.8	
M15	220	11	95.1	
M16	220	11	95.1	
M17	220	11	95.4	
M18	216	15	93.6	
M19	217	14	94.1	
M20	222	9	96.3	√

### 3 结论与讨论

#### 3.1 结论

利用最小降雨间歇(MIT)法和相关分析法筛选致灾过程并提取与灾度显著相关的降雨量、降雨强度及前期降雨3类11个关键特征因子,相关系数在0.137~0.415,均通过0.01水平显著性检验,具有较好的灾情信息可信度判别潜力。

以11个关键特征因子为输入项,结合单类支持向量机法和十折交叉检验法,10次随机挑选建模样本,构建10个强降雨灾情气象因子致灾判别模型。以真实灾情判别准确率最高原则选择得出最优模型(M9),真实灾情判别准确率为96.4%,但伪灾情判别准确率为67.6%,说明该模型评估伪灾情潜力不足。

剔除自相关显著的降雨特征因子,以最大降雨量、1h最大雨量、前10日降雨总量及平均小时雨强4个因子作为输入因子,重新构建强降雨灾害气象因子致灾判别模型,检验可见,最优模型(M20)证真准确率为96.3%,证伪准确率为82.9%,较M9提升15.3个百分点,综合准确率可达89.6%。表明M20可以较好地表征强降雨灾情可信度,且大幅减少输入因子,更有利于在业务实践中计算应用。

#### 3.2 讨论

本研究建立的气象因子致灾判别模型不依赖于灾情数据所含有的具体数值,对数值型和非数值型的强降雨灾情信息可信度判别均适用,与传统的灾情可信度判别方法<sup>[5,9-12]</sup>相比,更具普适性。

对降雨特征因子与灾度进行相关分析发现,代表降雨时长的各特征因子与灾度均无显著相关关

系,说明单独的降雨时长因子不是触发强降雨灾害发生的关键要素,而降雨量以及融合降雨量与降雨时长的降雨强度因子是导致灾害发生的重要条件。

由于农业是强降雨灾害中最为主要的承灾体之一,因此,本研究所建立的模型适用于强降雨所致的农业气象灾情的可信度判别。但依据该类致灾因子对于农业生产影响的机理不同,可划分为更加精细化的灾害种类,如农业洪灾、涝灾和湿害等,如需对此类灾情进行真伪甄别,尚需更加细化的分灾种研究。

本研究将数理统计和人工智能的方法相结合,利用多源数据建立了针对强降雨灾情信息真伪识别效果较好的气象因子致灾判别模型,但仍存在一定局限性。从应用尺度来看,本研究以县(区)级行政区域为基本研究单元,研究方法可为地市级、省级或划定区域内的面灾情信息可信度判别提供参考,而对于有具体位置的点灾情信息判别,致灾因子精度尚需提高。从应用数据来看,灾害的发生是致灾因子、承灾体、孕灾环境多要素综合作用的结果,本研究所构建的气象因子致灾判别模型,主要基于降雨特征因子(即致灾因子)的危险性视角,而缺少承灾体脆弱性及孕灾环境适宜性等方面的信息支撑。因此,下一步将融入承灾体、孕灾环境等信息,并增加对点灾情信息判别的研究,以期提升灾情可信度判别模型的应用精度及准确度。

#### 参考文献 References

- [1] 赵竹飞.基于互联网新闻的灾情信息采集系统构建[D].上

- 海:上海师范大学,2018.
- Zhao Z F.Construction of disaster information collection system based on internet news[D].Shanghai:Shanghai Normal University,2018.(in Chinese)
- [2] 孔锋,薛澜,乔枫雪,等.新时代我国综合气象防灾减灾的综述与展望[J].首都师范大学学报(自然科学版),2019,40(4):67-72.
- Kong F,Xue L,Qiao F X,et al.Summary and prospect of comprehensive meteorological disaster prevention and reduction in China in the new era[J].Journal of Capital Normal University(Natural Science Edition),2019,40(4):67-72.(in Chinese)
- [3] 帅向华,胡素平,刘钦,等.地震灾情网络媒体获取与处理模型[J].自然灾害学报,2013,22(3):178-184.
- Shuai X H,Hu S P,Liu Q,et al.Internet media-based acquisition and processing model of earthquake disaster situation[J].Journal of Natural Disasters,2013,22(3):178-184.(in Chinese)
- [4] 白华,林勋国.基于中文短文本分类的社交媒体灾害事件检测系统研究[J].灾害学,2016,31(2):19-23.
- Bai H,Lin X G.Sina weibo disaster information detection based on Chinese short text classification[J].Journal of Catastrophology,2016,31(2):19-23.(in Chinese)
- [5] 李吉顺,马振骅,佟旭.上报灾情可信度判别研究[J].中国减灾,2004(10):20-21.
- Li J S,Ma Z H,Tong X.Study on reliability judgment of reporting disaster situations[J].Disaster Reduction in China,2004(10):20-21.(in Chinese)
- [6] 刘邕,刘洋,周燕川,等.DB12/T 776-2018 气象灾害调查工作规范[S].天津:天津市气象灾害防御技术中心,2018.
- Liu Y,Liu Y,Zhou Y C,et al.DB12/T 776-2018 Working regulation for meteorological calamity investigation[S].Tianjin:Tianjin Meteorological Disaster Prevention Technology Center,2018.(in Chinese)
- [7] 何文革,张骥,戴栋凌.“一般雷电灾害”调查鉴定方法探讨[J].现代信息科技,2019,3(2):25-29.
- He W G,Zhang J,Dai D L.Investigation and appraisal method of “general lightning disaster” [J].Modern Information Technology,2019,3(2):25-29.(in Chinese)
- [8] 吴孟恒,张彦勇.雷电灾害调查与鉴定技术[J].气象科技,2010,38(6):786-789.
- Wu M H,Zhang Y Y.Investigation and appraisal methods of disasters caused by lightning[J].Meteorological Science and Technology,2010,38(6):786-789.(in Chinese)
- [9] Alhakad N,Serranoalvarado P,Busnel Y,et al.Trust evaluation of a system for an activity with subjective Logic[C]//11th International Conference(Trust, Privacy, and Security in Digital Business. 2014),Proceedings: LNCS 8647,2014: 48-59.
- [10] 张仕斌,许春香.基于云模型的信任评估方法研究[J].计算机学报,2013,36(2):422-431.
- Zhang S B,Xu C X.Study on the trust evaluation approach based on cloud model[J].Chinese Journal of Computers, 2013,36(2):422-431.(in Chinese)
- [11] 李玲玲,朱芬芬,姚致清,等.基于可信度的可靠性度量云模型[J].电力系统保护与控制,2012,40(8):90-94.
- Li L L, Zhu F F, Yao Z Q,et al.Reliability measure cloud model based on credibility[J].Power System Protection and Control,2012,40(8):90-94.(in Chinese)
- [12] 徐绪堪,楼昱清,于成成.基于 D-S 理论的突发事件多源数据可信度评估研究[J].情报理论与实践,2019,42(8):67-72.
- Xu X K,Lou Y Q,Yu C C.Research on the requirements of multi-source data fusion emergency decision-making based on Dempster-Shafer theory[J].Information Studies:Theory & Application,2019,42(8):67-72.(in Chinese)
- [13] 周月华,彭涛,史瑞琴.我国暴雨洪涝灾害风险评估研究进展[J].暴雨灾害,2019,38(5):494-501.
- Zhou Y H,Peng T,Shi R Q.Research progress on risk assessment of heavy rainfall and flood disasters in China [J].Torrential Rain and Disasters,2019,38(5):494-501.(in Chinese)
- [14] 陈小雷.河北省暴雨洪涝灾害评估:2016年7月19-21日特大暴雨过程[M].北京:气象出版社,2018.
- Chen X L.Assessment about storm and flood disaster in Hebei province:the torrential rain process from July 19 to 21 in 2016[M].Beijing:China Meteorological Press,2018.(in Chinese)
- [15] 潘敖大,王珂清,曾燕,等.长江三角洲近46a气温和降水的变化趋势.大气科学学报[J],2011,34(2):180-188.
- Pan A D,Wang K Q,Zeng Y,et al.Trends of temperature and precipitation variation in the Yangtze River Delta from 1961 to 2006[J].Transactions of Atmospheric Sciences, 2011,34(2):180-188.(in Chinese)
- [16] 秦鹏程,刘敏,李兰.有效降水指数在暴雨洪涝监测和评估中的应用[J].中国农业气象,2016,37(1):84-90.
- Qin P C, Liu M, Li L.Application of effective precipitation index in rainstorm flood disaster monitoring and assessment

- [J].Chinese Journal of Agrometeorology,2016, 37(1):84-90.(in Chinese)
- [17] 温泉沛,周月华,霍治国,等.湖北暴雨洪涝灾害脆弱性评估的定量研究[J].中国农业气象,2018,39(8):547-557.  
Wen Q P,Zhou Y H,Huo Z G,et al.Quantitative assessment on vulnerability of storm flood disasters in Hubei province[J].Chinese Journal of Agrometeorology,2018,39(8):547-557.(in Chinese)
- [18] 刘学艺,李平,郜传厚.极限学习机的快速留一交叉验证算法[J].上海交通大学学报,2011,45(8):1140-1145.  
Liu X Y,Li P,Gao C H.Fast leave-one-out cross-validation algorithm for extreme learning machine[J].Journal of Shanghai Jiaotong University,2011,45(8):1140-1145.(in Chinese)
- [19] 沈澄,孙燕,尹东屏,等.江苏省暴雨洪涝灾害特征分析[J].自然灾害学报,2015,24(2):203-212.  
Shen C,Sun Y,Yin D P,et al.Characteristic analysis of rainstorm-induced flood disaster in Jiangsu province[J].Journal of Natural Disasters,2015,24(2):203-212.(in Chinese)
- [20] 夏祎萌,何清,李军,等.新疆大风灾害灾度和危险度分析[J].中国沙漠,2012,32(4):1025-1028.  
Xia Y M,He Q,Li J,et al.Disaster degree and risk degree of wind disasters in Xinjiang[J].Journal of Desert Research, 2012,32(4):1025-1028.(in Chinese)
- [21] 武琳琳.基于 Fisher 最优分割法的聚类分析应用[D].郑州:郑州大学,2015.  
Wu L L.Cluster analysis based on the method of Fisher optimal division[D].Zhengzhou:Zhengzhou University,2015.(in Chinese)
- [22] Melillo M,Brunetti M T,Peruccacci S,et al.An algorithm for the objective reconstruction of rainfall events responsible for landslides[J].Landslides,2015,12(2):311-320.
- [23] 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 Y X,Wang Y,Ma Q Y,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)
- [24] 樊高峰,张勇,柳苗,等.基于支持向量机的干旱预测研究[J].中国农业气象,2011,32(3):475-478.  
Fan G F,Zhang Y,Liu M,et al.Study of drought prediction based on support vector machine[J].Chinese Journal of Agrometeorology,2011,32(3):475-478.(in Chinese)
- [25] 王莉萍,王秀荣,王维国,等.QX/T 341-2016 降雨过程强度等级[S].北京:国家气象中心,2016.  
Wang L P,Wang X R,Wang W G,et al. QX/T 341-2016 Rainfall process intensity rating[S].Beijing:National Meteorological Center,2016.(in Chinese)
- [26] 孙霖.人脸识别中的活体检测技术研究[D].杭州:浙江大学,2010.  
Sun L.Research on anti-spoofing in face recognition[D]. Hangzhou:Zhejiang University,2010.(in Chinese).
- [27] 闫龙,胡晓鹏.基于上下文特征与单类支持向量机的人脸活体检测[J].电子技术应用,2020,46(6):32-35.  
Yan L,Hu X P.Face anti-spoofing based on context and OCSVM[J].Electronic Technology Application,2020,46(6): 32-35.(in Chinese)
- [28] 王梅生.线性回归分析中自变量间的相关性问题[J].东北工学院学报,1985(1):94-104.  
Wang M S.On the correlation between variables in linear regression analysis[J].Journal of Northeast Institute of Technology,1985(1):94-104.(in Chinese)