

王瑛, 祁京, 王姝苓, 等. 基于聚类与统计分级的河北省气象灾害综合区划研究[J]. 灾害学, 2023, 38(2): 85–88, 113. [WANG Ying, QI Jing, WANG Shuling, et al. Study on Regionalization of Meteorological Disasters in Hebei Province—Computer Realization Based on Cluster and Statistics Classification[J]. Journal of Catastrophology, 2023, 38(2): 85–88, 113. doi: 10.3969/j.issn.1000-811X.2023.02.013.]

基于聚类与统计分级的河北省气象灾害 综合区划研究^{*}

王 瑛^{1,2}, 祁 京^{1,2}, 王姝苓^{1,2}, 李 璨^{3,4}, 俞海洋^{3,4}

(1. 北京师范大学 环境演变与自然灾害教育部重点实验室, 北京 100875;
2. 应急管理部教育部减灾与应急管理研究院, 北京 100875; 3. 河北雄安新区气象局,
河北 雄安 071700; 4. 中国气象局雄安大气边界层重点开放实验室, 河北 雄安 071100)

摘 要: 传统的气象灾害区划方法过于依赖专家经验。因此, 该文探索基于聚类与统计分级方法, 由计算机独立完成区划的技术。以河北省暴雨洪涝、冰雹、大风等 9 种气象灾害为例, 通过 K 均值聚类的方法实现了全自动的区域划分, 将河北省气象灾害综合区划分为 6 个区, 并对比了不同单灾种危险性分级对最终区划的影响, 基于分位数法得到的区划方案最优, 其各个区域的分布与历史灾情分布较为一致。

关键词: 气象灾害; 灾害区划; 聚类; 分级; 河北省

中图分类号: P429; X43; X915.5 **文献标志码:** A **文章编号:** 1000-811X(2023)02-0085-05

doi: 10.3969/j.issn.1000-811X.2023.02.013

我国暴雨洪涝、冰雹、台风等气象灾害种类较多, 多种气象灾害在不同地区的组合特征不同, 对地区造成的影响也不同^[1-3], 因此, 对各地进行气象灾害危险性区划, 识别高危险区域有十分重要的现实意义。

自然灾害区划已有诸多研究成果, 冯丽文等^[4]根据十种气象灾害, 通过不同季节主导灾种的频率分布与构成地域组合类型, 依据区划原则, 制成不同季节的综合灾害区划; 陈楠^[5]通过专家打分法、层次分析法、气象灾害风险指数法、加权综合评价等方法对致灾因子的危险性、孕灾环境和承灾体易损性以及防灾减灾能力这四个方面对菏泽市的气象灾害进行风险评估与区划。石菊松等^[6]通过建立相应的滑坡灾害危险性评价体系, 基于 GIS 技术的信息量模型和敏感性评价方法, 实现了巴东县新城区滑坡灾害危险性区划。

自然灾害综合区划的基础数据通常是各类单灾种危险性分级数据, 不同分级方法对区划结果有较大的影响。朱奇峰等^[7]探讨分级方法对坡度制图的影响, 采用四种分级方法进行分级, 得出自然断点分级方法效果在坡度制图中较好; 解明礼等^[8]以四川省汶川县为例, 采用 9 种分级方法对地质灾害易发性指数进行分级, 得出“历史地质灾害累计比例分段法”结果较好; 张丹红等^[9]采用地图编制中常用的自然断点法、几何间隔法、累计比例法、等差法、分位数法 5 种分级方法, 得出累计比例法在 6 个县区综合制图效果适用性最好。

总的来看, 上述自然灾害区划较多地依赖于

专家经验。随着统计学的发展, 聚类方法有了极为广泛的应用, 它可以较好的自动区分出特征相近的类别, 因此, 也可以将其应用于灾害区划研究中。另一方面, 自然灾害危险性区划的数据基础, 大多是单灾种危险性分级数据, 分级多采用 ArcGIS 软件的自然断点法^[10-16]。分级方法不同对最终区划方案有不同影响, 而如何准确分级是目前计算机自动区划流程中需要解决的关键问题。

本文以河北省为研究区, 选取造成损失较大的暴雨洪涝、冰雹、大风、大雾、低温霜冻、干旱、高温热浪、台风、雪灾 9 种气象灾害, 运用聚类算法得到将 9 种气象灾害综合在一起的区划方案, 并采用多种分级方法得到区划方案, 通过与历史灾情的对比, 分析区划方案的优缺点, 以期最大可能的实现计算机自动灾害区划。

1 研究区概况

河北省位于华北平原, 处于内陆与沿海的交会地带。其地势呈西北高、东南低, 地貌类型复杂多样, 兼有平原、高原、山地、湖泊、丘陵和海滨。在复杂孕灾条件影响下, 河北省内暴雨洪涝、冰雹、干旱、大风等气象灾害频发。河北省处于我国京津经济圈, 作为我国重要的经济总量大省, 频发的气象灾害每年都造成河北省巨大损失, 例如, “1996.8 暴雨”“2009.11 暴雪”“2012.7.21 暴雨”等。

^{*} 收稿日期: 2022-10-09

修回日期: 2023-01-31

基金项目: 河北雄安新区气象局“雄安新区重大气象灾害风险评估及对策研究”项目

第一作者简介: 王瑛(1974-), 女, 汉族, 云南曲靖人, 教授, 主要从事灾害风险评估和区划研究。E-mail: wy@bnu.edu.cn

通信作者: 俞海洋(1983-), 男, 汉族, 江苏南通人, 高级工程师, 主要从事气象灾害风险管理研究。E-mail: yuhaiyang@cma.cn

河北省各类气象灾害的区域分布差异明显,北部张家口、承德地区多沙尘暴、冰雹、低温霜冻等气象灾害;西部太行山地区多干旱、暴雨洪涝等气象灾害;东部沿海地区易受到台风灾害的影响;沧州、衡水、邢台、保定部分地区是河北省的干旱重灾区域。

2 数据与方法

2.1 数据来源

本文选用的暴雨洪涝、冰雹、大风、大雾、低温霜冻、干旱、高温热浪、台风、雪灾 9 种气象灾害数据均来自《河北省气象灾害风险地图集》^[17],冰雹由于缺少强度数据,采用冰雹灾害年均频次(1984—2015 年),其他灾种均采用 50 a 一遇致灾危险性数据,包括暴雨洪涝(年最大小时降水量)、大风灾害(最大风强)、大雾灾害(年大雾日数)、低温霜冻灾害(日最低气温)、干旱综合强度指数、高温热浪(持续时间)、台风(日最大风速)、雪灾(降雪 1 d 积雪深度)。历史灾情数据主要来自于气象灾害大典、新闻报道和民政灾情。

2.2 研究方法

2.2.1 K 均值聚类算法

K 均值聚类是通过迭代求解将一个集合分成类似对象组成的多个类的分析算法。本文以 K 均值聚类对各个单灾种进行分类,将欧氏距离作为相似度测度,采用均值作为聚类准则函数,求解对应某一形心的最优分类^[18],误差平方和达到最小时聚类终止。

在样本数据集中随机选取 k 个样本点作为初始形心 $c_i (i=1, 2, \dots, k)$ 。在第 j 次迭代中,对任意一个样本点 $p_i (i=1, 2, \dots, n)$ 求其到 k 个形心的欧氏距离:

$$E_d(t, i) = \sqrt{(p_i - c_i)^2} \quad (1)$$

将该样本归类到与其距离最小的形心所在的簇。采用均值更新各簇的形心值。重复步骤,直到各簇的形心值不再变化。

$$c_i = \frac{1}{n_i} \sum_{t=1}^{n_i} p_{it} (i=1, 2, \dots, k) \quad (2)$$

2.2.2 区域命名

聚类后,需要对各个区域进行命名,经过多次试验,确定以下命名规则:

若某区域内,第 i 灾种($i=1, 2, \dots, 9$)强度最高级和次高级的面积和 S_i ,在各类灾种中最大;第 j 灾种($j=1, 2, \dots, 9$)强度最高级和次高级的面积和 S_j ,在各类灾种中第二大;第 k 灾种($k=1, 2, \dots, 9$)强度最高级和次高级的面积和 S_k ,在各类灾种中第三大;且 $S_i/S_t, S_j/S_t, S_k/S_t \geq 60\% (\pm 5\%)$ (S_t 为该区域的总面积),则 i, j, k 灾种被认为是该区域的主导灾种。若 $S_i/S_t \geq 60\% (\pm 5\%)$, $S_j/S_t, S_k/S_t < 60\% (\pm 5\%)$ (S_t 为该区的面积),则仅 i 灾种被认为是该区域的主导灾种。

如果上述条件不满足,第 m 灾种($m=1, 2, \dots, 6$)强度最高级、次高级、中间级面积为 SS_m, S_m, A_m ,若 $SS_m + S_m + A_m \geq 80\% (\pm 5\%)$,则 m 灾种被认为是该区域的主导灾种。

2.2.3 区划方案的评价方法

区划方案的优劣判断,始终是区划研究中的难题。对于自然灾害而言,区划方案是否可以全面反映历史灾情,可以作为判断方案的参考标准之一,因此,本文采用空间重合度的方法来判断

区划方案与历史灾情分布的吻合情况,进而确定区划方案的优劣。

空间重合度 Ω 的计算方法(图 1)如下所示:

$$B_{evaluated} = \sum_{i=1}^{nX} X_i + \sum_{j=1}^{nZ} Z_j \quad (3)$$

$$B_{historical} = \sum_{i=1}^{nX} X_i + \sum_{k=1}^{nY} Y_k \quad (4)$$

$$\alpha = \frac{B_X}{B_{evaluated}} \quad (5)$$

$$\gamma = \frac{B_Z}{B_{evaluated}} \quad (6)$$

$$\Omega = \alpha - \beta - \gamma \quad (7)$$

由上式可知, Ω 的数值位于 $-2 \sim 1$ 之间, Ω 越接近于 1,重合度越高; Ω 越接近于 -2 ,重合度越低。

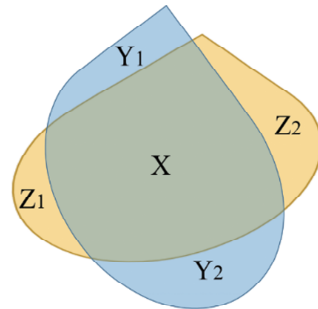


图 1 空间重合度计算示意图

2.2.4 单灾种危险性分级方法

由于各个单灾种的危险性指标完全不同,不具有可比性。因此需要将各个灾害的危险性数据进行分级,再基于分级数据进行聚类。目前 GIS 软件中,常用的分级方法有以下 4 种。

分位数分级法,是指将一个随机变量的概率分布范围分为几个等分的数值点。本文采用 5 分位法,把所有数值由小到大排列并分成五等分,处于 4 个间隔点(20%, 40%, 60%, 80%)位置的数值就是 5 分位数。

自然断点法,通过不断地迭代计算分级,能根据数值统计的分布规律进行自动分级和分类,并且能够使结果类别间的差异最大化^[19]。这个算法的功能是减少同一级中的差异、增加级间的差异。

几何间隔法,按照数值几何间距进行分级,是用于显示要素属性值与平均值之间的差异,将使用与标准差成比例的等值范围创建分类间隔。

均值—标准差方法,是利用单灾种危险性均值和不同的标准差倍数的组合来划分各个单灾种的危险性等级。标准差反映了危险性相对于平均危险性的偏离程度,用均值和标准差能够反映不同区域危险性的变异。基于上述理论,将研究区的危险性影响以 μ (均值)、 $\pm 0.5 \text{ std}$ 、 $\pm 0.1 \text{ std}$ 为分割点划分为 5 级。

上述 4 种分级方法得到的单灾种空间格局之间存在差异,对最终的区划方法也有影响。

3 不同的区划方案与评价

9 个单灾种危险性都采用同样的分级方法,对分级数据采用 K 均值聚类,类别取 6,最终得到的区划方案如图 2 所示。

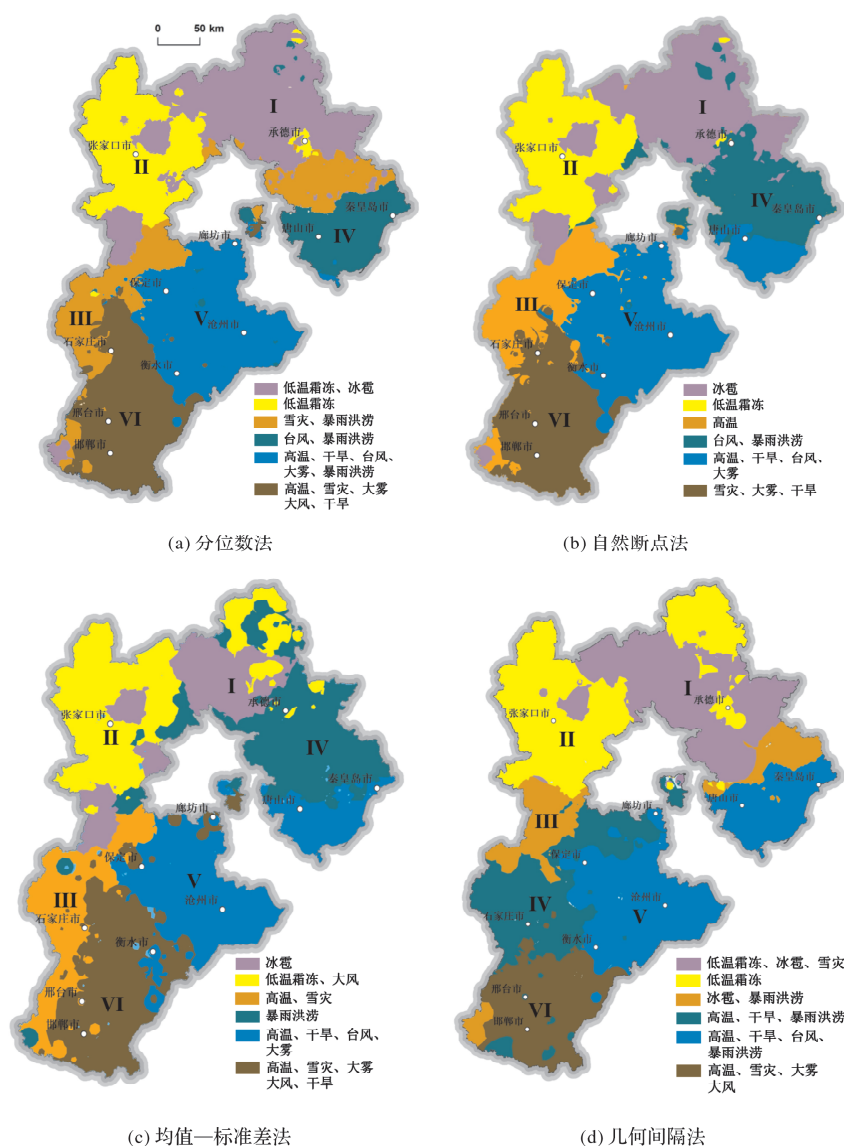


图2 基于不同分级方法的河北省气象灾害综合区划方案(审图号:冀S(2018)41号)

3.1 区划方案的比较

以分位数分级法得到的区划方案,见图2a所示。全省共分为6个区,大致呈南北格局。自北向南,分别为:I区东北部以低温霜冻、冰雹为主区;II区西北部以低温霜冻为主区;III区中部雪灾、暴雨洪涝为主区;IV区东部以台风、暴雨洪涝为主区;V区东南部以干旱、高温、大雾、暴雨洪涝、台风、大风为主区;VI区南部以大雾、雪灾、高温、干旱、大风灾害为主区。

基于自然断点分级法的区划方案,如图2b所示。除III、IV、V区外,其他区的空间格局与图2a大体相似。III区的主导灾种从雪灾、暴雨洪涝变为高温,且范围变小,只分布在河北西部,东部没有分布;东部减少的部分变成了IV区台风、暴雨洪涝灾害。原IV区唐山部分,变成V区一部分,V区的主导灾种减少为干旱、大雾、高温、台风。VI区范围略缩小,主导灾种减少为雪灾、大雾、干旱。

采用标准差法分级的区划方案(图2c)与断点法的区划方案(图2b)较类似。主要区别是I区的范围缩小,承德北部变为低温霜冻、暴雨洪涝为

主区。III区的范围扩大,向VI区西部带状延伸,灾种由高温变为高温、雪灾为主区。V区的北缘略向南缩。

几何间隔法分级的区划方案(图2d)与分位数分级的方案有较大差异。图2d的I区向图2a的III区的东部扩展,II区向图2a的I区东北部扩展,两种方案的III、IV区分布差异最大,灾种也完全不同,III区成为冰雹、暴雨洪涝为主区,范围变小,IV区成为高温、干旱、暴雨洪涝为主区,分布在石家庄、保定一带。V区分布范围变化也较大,将唐山、秦皇岛部分也纳入了本区,灾种变为高温、干旱、台风、高温、暴雨洪涝区。VI区的范围缩小,灾种也有所变化。

3.2 区划方案与历史灾情验证

上述区划方案差别较大,为了判断各个方案的优劣,采用空间相似度的方法,计算历史受灾县域与区划的重合情况,重合度高的方案就认为是最优方案。将1980—2015年河北省历史暴雨受灾人口县、冰雹、大风、大雾、干旱、高温、低温霜冻、台风、雪灾等受灾县次数占到总数60%以上的地区作为历史受灾县域。

表 1 四种区划方案与历史受灾县域的空间相似度

	暴雨洪涝	冰雹	大风	大雾	低温霜冻	干旱	高温	台风	雪灾
分位数法	-0.20	-0.15	-0.29	-0.87	-0.85	-0.59	-0.89	-0.87	-0.77
自然断点法	-0.45	-0.25	无	-0.89	-0.92	-0.61	-0.89	-0.89	-1.13
标准差法	-0.04	-1.25	-0.69	-0.86	-0.91	-0.62	-0.87	-0.82	-0.86
几何间隔法	-0.26	0.478	-1.28	-1.28	-0.91	-0.69	-0.84	-0.89	-0.67

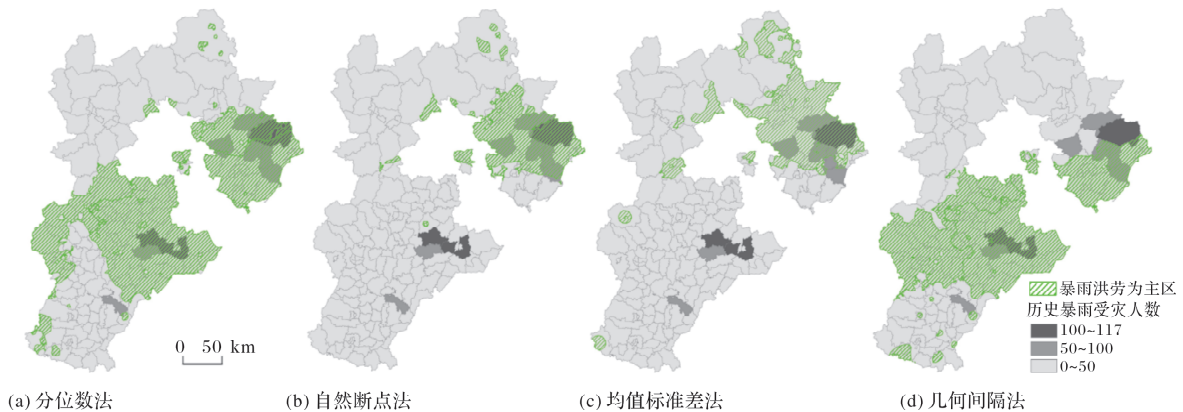


图 3 不同分级方法的暴雨洪涝区划方案与历史灾情的验证

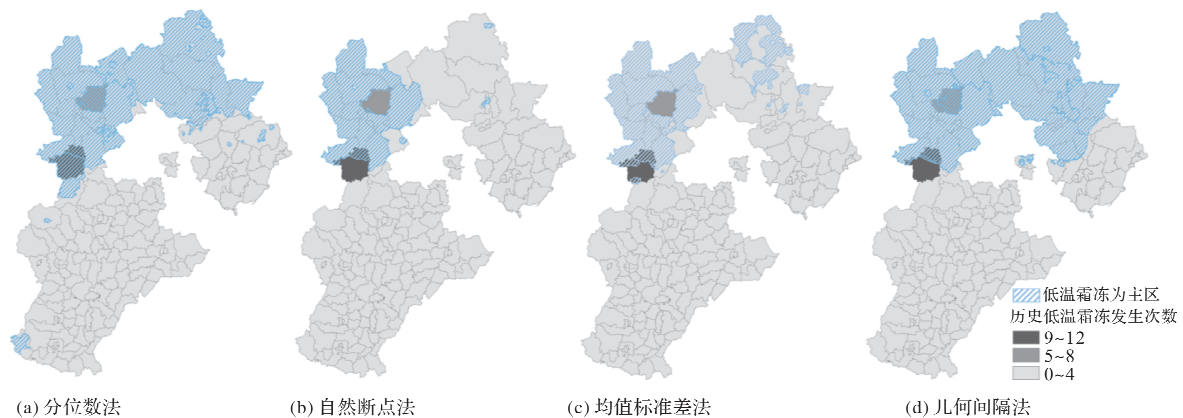


图 4 不同分级方法的低温霜冻区划方案与历史灾情的验证

将 4 种区划方案的各个灾种分区与历史各个受灾县域进行空间相似度计算, 结果见表 1。总体看, 分位数法的主导灾种分布与历史上 9 种灾害的受灾县域空间相似度大都为最高, 仅 1-2 种灾害低于几何间隔法和均值标准差法, 因此, 本文认为根据分位数分级法获得的河北省气象灾害综合区划方案最优。

为了更直观的展示空间相似度, 以暴雨洪涝灾害、低温霜冻为例, 将 4 种区划方案的两种灾害的主要影响区与历史上两种灾害的受灾县域进行空间分布叠加(图 3)。图 3 中, 灰色区域表示历史上受灾县域分布, 斜线网纹为 4 种区划方案所得到的灾害主要影响区分布。可以看到, 分位数法的区划方案中, 两种灾害的分区与历史受灾范围基本重合, 极少量县在分区外。

自然断点法和均值-标准差的方案中, 暴雨洪涝分区(图 3b、图 3c)缺少了沧州等历史上暴雨灾害受灾较严重区域, 低温霜冻分区(图 4b、图 4c)也缺少了承德北部等历史上低温霜冻受灾较严重的区域; 而分位数法和几何间隔法在暴雨洪涝(图 3a、图 3d)和低温霜冻分区(图 4a、图 4d)上则相对吻合历史上暴雨灾害受灾较严重区域, 且分位数法较几何间隔法来说暴雨洪涝和低温霜冻的分区与历史受灾严重区域更加吻合。

从空间格局上看, 基于分位数法的河北省气象灾害综合区划(图 2a), 也与河北省气象灾害孕灾环境分异规律较为一致。由于温度的纬向地带性分异规律, 河北省北部 I 区、II 区以低温类气象灾害较多, 高温灾害、干旱灾害主要发生在南部 VI 区。由于临近渤海, IV 区、V 区以台风、暴雨洪涝灾害为主。而由于冀北山地、燕山山地的影响, 使得 I 区冰雹灾害也较多; 由于燕山山地、太行山山地的影响, 使得 III 区以雪灾、暴雨洪涝灾害为主。

5 结论与讨论

本文以河北省为例, 基于暴雨洪涝、冰雹、大风、大雾、低温霜冻、干旱、高温热浪、台风、雪灾 9 种气象灾害的危险性评估等级结果, 进行气象灾害综合区划研究, 得出以下结论:

(1) 基于多个单灾种危险性分级数据, 通过 K 均值聚类的区划方法可以基本实现全自动的区域划分, 减少对专家经验的依赖。

(2) 单灾种危险性评估采用不同的分级方法, 对聚类结果有影响。分位数法、自然断点法、几何间隔法、标准差法 4 种分级方法中分位数法得到

(下转第 113 页)

ban planning. A model that can simulate the temperature difference between the original roof and the green roof in all weather is constructed by projection pursuit modeling. The simulation accuracy of the model is tested through validation experiments, and the simulation of the effect of green roofs on urban heat island effect is carried out in the heat island center area of Nanning city as an example. The results show that the simulation method of the effect of green roof on urban heat island effect based on the projection pursuit model can effectively simulate the mitigation of the heat island effect in any area after the green roof is carried out; the heat island area of the study area can be reduced by 40.1% through the green roof, which can significantly improve the heat island situation in the area.

Keywords: green roofs; heat island effect; projection pursuit model

(上接第 88 页)

的区划方案结果最优, 与历史灾情的重合度最好。

(3) 河北省气象灾害综合区划, 分为 6 个大区: I 区东北部低温霜冻、冰雹为主; II 区西北部低温霜冻为主; III 区中部雪灾、暴雨洪涝为主; IV 区东部台风、暴雨洪涝为主; V 区东南部干旱、高温、大雾、暴雨洪涝、台风、大风为主; VI 区南部大雾、雪灾、高温、干旱、大风灾害为主。

(4) 河北省气象灾害危险性的格局, 有南北差异, 主要由于温度的纬向地带性分异规律, 使得河北北部低温类气象灾害较多, 南部相对较少; 高温灾害主要发生在南部地区。同时, 格局也存在东西差异, 是受到地势、距海远近等因素影响,

本文的分位数法区划方案, 仍需要完善, 几何间隔法对冰雹灾害的分区, 与历史受灾县域重合度更高, 可以考虑不同分级方法的配合应用。此外, 分位数法区划方案中, 太行山区有小部分未划至 V 区中, 在考虑地域共轭的原则下, 这种情况需要进行一定的调整, 仍然需要专家经验的部分参与。

参考文献:

- [1] 史培军, 孙劭, 汪明, 等. 中国气候变化区划(1961—2010 年)[J]. 中国科学: 地球科学, 2014, 44(10): 2294—2306.
- [2] 王静爱, 史培军, 王瑛, 等. 中国城市自然灾害区划编制[J]. 自然灾害学报, 2005, 14(6): 42—46.
- [3] 王平, 史培军. 中国农业自然灾害综合区划方案[J]. 自然灾害学报, 2000, 9(4): 16—23.
- [4] 冯丽文, 郑景云. 我国气象灾害综合区划[J]. 自然灾害学

- 报, 1994, 3(4): 49—56.
- [5] 陈楠. 菏泽市气象灾害风险评估与区划研究[J]. 山东农业大学学报(自然科学版), 2018, 49(2): 257—261.
- [6] 石菊松, 张永双, 董诚, 等. 基于 GIS 技术的巴东新城区滑坡灾害危险性区划[J]. 地球学报, 2005, 26(3): 275—282.
- [7] 朱奇峰, 杨勤科, 师动, 等. 坡度分级方法对坡度制图的影响[J]. 水土保持通报, 2017, 37(3): 314—320.
- [8] 解明礼, 巨能攀, 赵建军, 等. 区域地质灾害易发性分级方法对比分析研究[J]. 武汉大学学报(信息科学版), 2021, 46(7): 1003—1014.
- [9] 张丹红, 王效科, 张路, 等. 大比例尺土壤保持服务制图分级方法研究[J]. 生态学报, 2021, 41(4): 1391—1401.
- [10] 贺芳芳, 邵步粉. 上海地区低温、雨雪、冰冻灾害的风险区划[J]. 气象科学, 2011, 31(1): 33—39.
- [11] 李楠, 任颖, 顾伟宗, 等. 基于 GIS 的山东省暴雨洪涝灾害风险区划[J]. 中国农学通报, 2010, 26(20): 313—317.
- [12] 张亚杰, 张京红, 陈升学, 等. 海南岛橡胶(*Hevea brasiliensis*)寒害风险区划[J]. 生态学杂志, 2018, 37(9): 2808—2814.
- [13] 邹海平, 陈汇林, 李伟光, 等. 海南岛冬种瓜菜暴雨洪涝灾害风险评估与区划[J]. 中国农业气象, 2015, 36(1): 100—107.
- [14] ZOU H, CHEN H, LI W, et al. Risk assessment and regionalization of rainstorm and Flood disaster of winter melon and vegetable in Hainan Island[J]. Chinese Journal of Agrometeorology, 2015, 36(1): 100—107.
- [15] 轩春怡, 刘勇洪, 杨晓燕, 等. 基于 1km 网格的北京暴雨洪涝灾害风险区划[J]. 气象科技, 2020, 48(4): 579—589.
- [16] 陈施吉, 李家启, 王远谋, 等. 长江重庆航道大雾灾害风险评估与区划研究[J]. 西南大学学报(自然科学版), 2020, 42(9): 165—172.
- [17] 《河北省气象灾害风险地图集》编辑委员会编著. 河北省气象灾害风险地图集[M]. 北京: 科学出版社, 2018.
- [18] 薛薇. 基于 R 的统计分析与数据挖掘[M]. 北京: 中国人民大学出版社, 2014: 272—288.
- [19] JENKS G F, COULSON M R C. Class Intervals for Statistical Maps[M]. Bertelsmann C. 1963.

Study on Regionalization of Meteorological Disasters in Hebei Province Based on Cluster and Statistics Classification

WANG Ying^{1,2}, QI Jing^{1,2}, WANG Shuling^{1,2}, LI Can^{3,4}, YU Haiyang^{3,4}

(1. Key Laboratory of Environmental Change and Natural Disaster of Ministry of Education, Beijing Normal University, Beijing 100875, China; 2. Academy of Disaster Reduction and Emergency Management, Ministry of Emergency Management and Ministry of Education, Beijing 100875, China; 3. Hebei Xiong'an New Area Meteorological Bureau, Hebei 071100, China; 4. China Meteorological Administration Xiong'an Atmospheric Boundary Layer Key Laboratory, Hebei 071100, China)

Abstract: The traditional regionalization depends too much on experience from experts. Therefore, based on clustering and statistical classification methods, This paper explore the technology of zoning independently completed by computer. Taking nine meteorological disasters including rainstorm, flood, hail and gale in Hebei province as examples, This paper realize automatic regional division by k -means clustering method, and divide Hebei province into 6 regions. This paper also compare the influence of different hazardous classification of single disaster on the final zoning, and the results show that the zoning scheme based on quantile method is the best, and the regionalization are basically consistent with the historical disaster distribution.

Keywords: meteorological disasters; disaster regionalization; cluster; classification; Hebei Province