

陈洁, 苏炳彦. 基于聚类法的陕西省雷电灾害风险区划 [J]. 陕西气象, 2015 (S1): 4-6.

文章编号: 1006-4354 (2015) S1-0004-03

# 基于聚类法的陕西省雷电灾害风险区划

陈 洁, 苏炳彦

(渭南市雷电预警防护中心, 陕西渭南 714000)

**摘 要:** 选取 2002—2007 年陕西省雷电灾害易损性分析指标的统计研究结果作为变量, 采用 Ward' s 法对陕西 10 个地市进行层次聚类分析, 并用迭代聚类分析法检验聚类解。结果表明西安属高风险区, 渭南、咸阳属次高风险区, 宝鸡属中风险区, 其它地市属低风险区。

**关键词:** 聚类分析; 雷电灾害易损性; 风险区划

**中图分类号:** P429

**文献标识码:** A

区域雷电灾害风险评估是相对于单体建筑物评估模式提出的, 考虑评估所在地的雷电活动时空分布特征、地域环境特征, 以综合多种功能的区域为评估对象, 以灾害等级或事故后果为研究内容, 对风险做出评价与估算的过程。

目前, 区域雷电灾害风险评估尚处探索阶段, 没有规范的评估方法或成熟经验技术, 现有的区域雷电灾害风险研究主要是易损度区划。由于环境差异性的存在, 使得不同区域雷电灾害易损度差别也很大, 学者们利用气象台站观测资料、闪电定位资料或其它资料结合评估区域特征参数, 通过引入多种数理模型开展了大量区域雷电灾害风险研究, 对各地的雷灾易损度区划分等级<sup>[1-4]</sup>, 李彩莲等<sup>[5]</sup>根据 2002—2007 年相关资料研究了陕西省雷电灾害易损性。

由于研究中各指标变量等级存在阈值赋予的主观性, 有必要对区划结果进行检验。本文根据陕西省雷电灾害易损性分析指标的统计研究结果, 采用 SPSS (Statistical Package for the Social Sciences) 软件中一种基本凝聚层次方法, 即 Ward' s 法对 10 个地市的特征参量聚类分析, 并结合检验效果得出雷电灾害风险区划等级。

## 1 资料和方法

### 1.1 资料来源

根据文献 [5] 的研究结果, 陕西省 10 个地

市雷电灾害易损性分析的研究指标见表 1, 雷暴日数选取 2002—2007 年各地气象站观测资料, 雷电灾害资料取自全国和陕西省雷电灾害报告,

**表 1 2002—2007 年陕西省各地市雷电灾害易损性分析指标**

| 地市 | 雷暴日 | 雷电灾害频度 | 经济损失模数 | 生命损失模数 |
|----|-----|--------|--------|--------|
| 安康 | 24  | 3.7    | 0.006  | 0.011  |
| 宝鸡 | 15  | 1.3    | 0.021  | 0.021  |
| 汉中 | 24  | 1.8    | 0.008  | 0.013  |
| 商洛 | 23  | 2.2    | 0.008  | 0.018  |
| 铜川 | 22  | 2.3    | 0.017  | 0.021  |
| 渭南 | 17  | 5.2    | 0.022  | 0.042  |
| 西安 | 11  | 8.0    | 0.122  | 0.080  |
| 咸阳 | 15  | 4.8    | 0.039  | 0.049  |
| 延安 | 28  | 2.3    | 0.009  | 0.006  |
| 榆林 | 30  | 1.7    | 0.007  | 0.008  |

雷暴日数—平均每年发生雷暴的天数, 单位 d/a;  
雷电灾害频度—平均每年发生雷电灾害的次数, 单位 次/a;  
经济损失模数—区域单位面积上的潜在受灾经济量, 单位万元/km<sup>2</sup>;  
生命损失模数—区域单位面积上的潜在受灾人数, 单位万人/km<sup>2</sup>。

收稿日期: 2014-12-18

作者简介: 陈洁 (1982—), 女, 汉族, 陕西西安人, 硕士, 助理工程师, 从事雷电灾害风险评估研究。

国民生产总值采用统计年鉴和政府工作报告中的数据。

### 1.2 聚类分析法

多元统计学中常用的指标分类有判别分析法和聚类分析法，判别分析是事先已知事物的类别和分类数，聚类分析是事先不知道事物的类别和分类数。对 10 个地市雷电灾害风险区划的结果是缺乏先验知识的直接样本分类，用聚类法把性质相近的风险区域归为同一类统计分析。

聚类分析是多元统计学中的多变量统计技术，根据分类对象的不同，聚类分析法可分为 R 型聚类和 Q 型聚类。前者是分类筛选变量，了解变量间关系的亲疏程度；后者是分类处理样本，综合利用多个变量的信息对样本进行谱系图分类。

聚类分析多采用欧氏距离（公式 1）和 Pearson 相关系数（公式 2）计算相似性测度，欧氏距离表征多维空间聚类点的间距，距离越近表示两者关系越密切，划分为同一类的可能性就越大，相反，距离越远则表示两者可能分属不同种类。

欧氏距离

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \quad \text{。} \quad (1)$$

Pearson 相关系数

$$r_{ij} = \frac{\sum_{k=1}^m (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^m (x_{ik} - \bar{x}_i)^2 \sum_{k=1}^m (x_{jk} - \bar{x}_j)^2}} \quad \text{。} \quad (2)$$

R 型聚类时， $m$  为样本数， $x_i, x_j$  为任意两个变量；Q 型聚类时， $m$  为变量数， $x_i, x_j$  为任意两个样本。

为了强调同类事物的内部差异小，不同类事物差异大，选取聚类分析法中 Ward’s 法风险区划，方法是将每个对象作为一个簇的开始，通过计算其它对象到初始簇中心的离差平方和，逐步合并距离最小的两个簇，直至所有对象都合并

在一个簇中或满足预先设定的终结条件时停止循环。

### 2 雷电灾害风险区划

将陕西省 10 个地市作为样本，与雷电灾害易损性相关的 4 个影响因子（即雷暴日数、雷电灾害频度、经济损失模数和生命损失模数）作为 4 个指标变量，由于各变量的量纲不一致，利用 ward’s 法先对 10 个样本的 4 个指标所包含数据标准化处理后进行系统聚类分析，二叉树样本聚类过程见图 1。

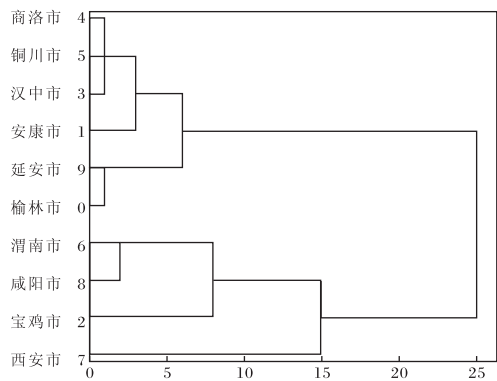


图 1 二叉树样本聚类过程

从图 1 看出，聚类过程进行了 9 步，每一步均对应一个距离测算值和一个聚合系数，映射值越大表明聚类过程的置信度越高。根据类间距欧氏距离映射值 5 来分析，将样本分为 4 类时较为理想，得到结果：西安，渭南、咸阳，宝鸡，安康、汉中、商洛、铜川、榆林、延安。

为验证以上聚类解的稳定性，以 4 类分类结果为迭代聚类分析的初始聚类中心（表 2）。在满足收敛准则 0.01，两次迭代间聚类中心变化

表 2 初始聚类中心值

| 初始<br>类别 | 地市      | 雷暴<br>日 | 雷电灾<br>害频度 | 经济损<br>失模数 | 生命损<br>失模数 |
|----------|---------|---------|------------|------------|------------|
| 1        | 西安市     | 11      | 8.0        | 0.122      | 0.080      |
| 2        | 渭南市、咸阳市 | 16      | 5.0        | 0.030      | 0.046      |
| 3        | 宝鸡市     | 15      | 1.3        | 0.021      | 0.021      |
| 4        | 其他      | 25      | 2.3        | 0.009      | 0.013      |

注：4 类分类结果的含义和单位同表 1

距的最小值小于初始类中心距离的 1%时停止迭代，完成分类结果（表 3）。

表 3 迭代聚类结果

| 样本  | 聚类 | 距离    |
|-----|----|-------|
| 安康市 | 4  | 0.216 |
| 宝鸡市 | 3  | 0.000 |
| 汉中市 | 4  | 0.101 |
| 商洛市 | 4  | 0.136 |
| 铜川市 | 4  | 0.211 |
| 渭南市 | 2  | 0.106 |
| 西安市 | 1  | 0.000 |
| 咸阳市 | 2  | 0.106 |
| 延安市 | 4  | 0.175 |
| 榆林市 | 4  | 0.280 |

经迭代聚类后的分类结果与 ward’s 法完全一致，可以将陕西省 10 个地市雷电灾害风险程度分为 4 类的聚类解较稳定，即西安为第 1 类属雷电灾害高风险区，渭南、咸阳为第 2 类属雷电灾害次高风险区，宝鸡为第 3 类属雷电灾害中风险区，安康、汉中、商洛、铜川、榆林、延安为第 4 类属雷电灾害低风险区。

将表 3 的聚类结果与表 1 中 4 个影响因素的数据对比发现，雷电灾害风险等级越高的地区，灾害的发生更频繁，经济和生命的损失情况亦显著，但是与当地雷暴的出现日数不成正比关系。

根据风险等级划分情况，关中较陕南和陕北遭受雷电灾害损失更严重。西安、渭南、咸阳和宝鸡属少雷暴区，然而随着单位面积上的人口基数和电气化密集程度的增加，受灾频数以及造成的经济和人员损失数量也陡增。

3 结论

根据雷电灾害易损性分析指标的统计结果，利用聚类分析法对陕西省 10 个地市的雷电灾害风险等级分类研究得出以下结论。

（1）依据雷电灾害风险的程度划分为 4 类并通过检验，聚类解表明西安属高风险区，渭南、咸阳属次高风险区，宝鸡属中风险区，其他地市属低风险区。雷灾风险区划等级与雷暴日数多少无直接关系，更多的受地区人口数量和电气化设备应用程度的影响。

（2）研究的重点是聚类法在风险区划分析中的可行性，在此基础上得出陕西省雷电灾害风险区划结果，分析数据直接采用已有研究成果，为了使风险区划更具有指导意义，应该采取长时间序列及最新统计数据资料作进一步研究并对本文结果修正。

（3）区划结论是在选取与雷电灾害易损性相关的 4 项特征数据标准化处理后得出的，随着区域雷电灾害风险评估研究的不断深入，将会发现更多适宜评价的影响因子和变量，增加或替换评价指标是否会产生不同区划结果以及区划的合理性有待研究。

参考文献：

[1] 邓春林, 季严飞, 刘刚. 基于雷电定位数据的区域雷击灾害风险评估方法探讨 [J]. 南京信息工程大学学报: 自然科学版, 2010, 2 (3): 230-235.

[2] 严春银, 吴高学, 朱建章. 区域雷灾易损性及其区划的实证分析 [J]. 气象与环境学报, 2007, 23 (1): 17-21.

[3] 曾金全, 张烨方, 王颖波. 基于综合评价算法的雷电灾害易损度区划模型研究 [J]. 气象, 2011, 37 (12): 1595-1600.

[4] 殷娴, 肖稳安, 冯民学, 等. 区域雷灾分布特征及易损度区划 [J]. 气象科技, 2009, 37 (2): 216-220.

[5] 李彩莲, 赵西社, 赵东, 等. 陕西省雷电灾害易损性分析、评估及易损度区划 [J]. 灾害学, 2008, 23 (4): 49-53.