

杨仲江, 江珂, 曹洪亮, 等. 基于 inpolygon 的区域外闪电剔除方法研究 [J]. 灾害学, 2014, 29(3): 72 - 76. [Yang Zhongjiang, Jiang Ke, Cao Hongliang, et al. Inpolygon-based Elimination Method of Regional Lightning [J]. Journal of Catastrophology, 2014, 29(3): 72 - 76.]

基于 inpolygon 的区域外闪电剔除方法研究*

杨仲江^{1,2}, 江珂^{1,2}, 曹洪亮^{1,2}, 楼嘉懿^{1,2}

(1. 南京信息工程大学 中国气象局气溶胶与云降水重点开放实验室, 江苏 南京 210044;
2. 南京信息工程大学 大气物理学院, 江苏 南京 210044)

摘 要: 闪电资料的完整性、准确性是雷电灾害风险评估的基础, 将影响雷电灾害风险评估结果的准确性, 并影响一系列依赖于闪电资料的科学研究。探究雷电地域分布时, 区域外的数据不影响探究准确性, 但探究雷电时间分布时, 只有区域内的数据是有效的。针对用传统方法统计经纬度跨度形成的矩形区域内的闪电频次来代表某地区闪电频次出现的严重误差, 首创以省市县地理底图为边界, 应用 inpolygon 函数方法, 简便快速地区分统计区域内外闪电数据, 大幅提高闪电原始资料准确性。以江苏省为例, 对比用传统疆界边缘经纬度所围矩形区域和用新方法统计江苏省边界所围区域的闪电频数, 研究表明: 用新方法统计的江苏省 2012 年的省内闪电数据占用传统矩形区域法统计的 50.46%。也即, 传统方法统计的 2012 年江苏省闪电频数误差高达 98.18%。通过剔除省外闪电, 得到省内真实有效的闪电频数数据, 大大提高了雷电灾害风险评估和雷暴分布规律分析的准确性和可靠性。

关键词: 闪电资料; shaperead 函数; inpolygon 函数; 地理底图; 误差

中图分类号: X4; P4 **文献标志码:** A **文章编号:** 1000 - 811X(2014)03 - 0072 - 05

doi: 10.3969/j.issn.1000-811X.2014.03.014

在气象、地理等学科中经常会遇到这种情况, 在给定区域内的数据有利用价值, 而区域外数据不必考虑。遇到该情况, 首先需要确定区域内和区域外数据, 其后剔除区域外数据。如果给定区域边界能用比较简单、确定、规则的数学模型来描述, 则判断点与平面位置关系相对容易。但气象、地理等学科需要处理的资料常常发生在某个不规则行政区域, 无法用简单的数学模型来描述边界, 处理会非常困难。对于少量数据, 人工处理是一种可行的方法, 但当数据量非常庞大时, 如闪电数据, 人工处理变得较难, 一般借助计算机处理。

闪电灾害带来的损失逐年增加, 闪电探测越来越受人们关注, 其监测定位在气象观测中扮演的角色也越来越重要^[1-4]。利用闪电定位系统可以获得地闪发生的地理位置、闪电强度的幅值、极性、闪电次数等参数, 用于分析雷电活动的时空分布规律等^[5-10]。Orville 等^[11]分析 1998 - 2000 年北美地区云地闪电, 3 年共计发生 8 870 万次闪电, 1998 年、1999 年和 2000 年分别为 3 100 万次,

2 900 万次和 2 800 万次; 张腾飞等^[12]、程琳等^[13]利用闪电资料分别对云南和江苏区域性雷电活动时间分布、地闪频次时间分布等方面进行了分析; 刘平等^[14]在以云南石林为例的喀斯特地貌景区雷击风险分析中, 利用石林县近 35 年的雷暴数据统计资料、2006 - 2010 年云南省闪电定位系统观测及雷灾上报资料, 结合石林景区特殊的地理位置、喀斯特地形地貌和地质构造特征, 分析景区所在区域的闪电强度、闪电密度、闪电频数等雷暴特征参数。但以上研究所使用的闪电数据的准确性有较大问题。原因之一是闪电定位系统自身存在误差^[15-18]。原因之二是统计的闪电数据并非完全在特定区域内, 即选取的闪电资料存在一定的误差。例如以江苏省边界经纬度所围的矩形区域与江苏省实际边界所围区域面积来比, 相差几乎有一倍, 因而以前者区域内的闪电频数来代表江苏省实际发生的闪电频数显然会出现很大误差。不同边界形状的区域统计出来的误差各不一样, 疆界越是狭长倾斜, 统计出来的闪电频数误差越大。

* 收稿日期: 2013 - 11 - 25 修回日期: 2014 - 01 - 05

基金项目: 国家自然科学基金项目(41175003); 江苏高校优势学科建设工程资助项目(PAPD)

作者简介: 杨仲江(1961 -), 男, 江苏射阳人, 副教授, 高级工程师, 主要从事雷电研究。E-mail: mashimaro2974@126.com

闪电定位系统自身的误差需要通过改进设备硬件及定位算法来不断消除。而剔除区域外闪电数据成为提高资料准确性的重要途径。很多软件(如 ArcGIS、Surfer 等)可以实现在已制作的图上剔除区域外数据,但无法将区域内数据导出。本研究应用 Matlab 中的 inpolygon 函数将原始闪电数据分为区域内(包括边界上)和区域外两部分,几乎可以完全剔除区域边界外的闪电数据,完全消除上述第二种原因造成的资料选取误差,为雷电灾害风险评估开展和一系列依赖于闪电资料的科学研究提供可靠依据,也为后期处理和统计数据减少工作量。

1 多边形的点包容性

多边形的点包容性在气象、地理等领域应用越来越广泛,该判定问题是进行气象、地理等资料处理的基础。1985 年 Preparata 和 Shamos 提出多边形的点包容性检验是计算几何中最基本问题之一,它在计算机图形学、地理信息系统(GIS)等领域有广泛应用。该问题描述为:给定多边形 P 和任意点 Q ,判定点 Q 是否位于多边形 P 内^[19]。

关于多边形的点包容性判断的经典方法有:①由 Foley、van Dam 等学者和 Haines 使用的经典射线交叉法(rar-crossing)^[20-21],过测试点向任意角度发一条射线,若与多边形各条边交点个数之和为偶数,则测试点在多边形之外,否则在多边形之内。若有交点为多边形顶点则要另选一条射线重算。该方法简单、可靠,可处理任意多边形,通常被用作与其他算法对比的基准算法。但却难以处理边界点在射线上、边界与射线共线等奇异情况。②由 Preparata、Hormonn、钟晓平等学者使用的累计角度法^[19,22-23],相对于闭合曲线围成的区域,若测试点与曲线上点的连线相对于 X 轴所形成的角度增量累计值为 360° ,则该点必位于该区域内部;若该累计值为 0° ,则该点必位于该区域外部。尽管目前已有大量多边形的点包容性检测算法,但寻找适合领域要求的解决方案仍然是重要的研究内容。

上述解决方法并不适用于判断闪电资料中落雷点与边界的关系,原因有以下几点:①闪电资料并不是只有落雷点在直角坐标系中的位置信息,

还包括闪电发生的日期、时刻、强度、陡度、误差等一系列参数;②统计在省、市或县内的闪电数据,是以该省地理底图作为边界,其经纬度构成的坐标点多达数千个;③一年闪电资料包含了几十万甚至几百万条数据,数据量非常大,研究某个地区闪电时间、空间分布时,常常要统计多年数据才能准确反应其规律。上述解决方法虽然原理简单,但实施过程相当繁琐,所用时间比较长。因此,本文应用 Matlab 中 inpolygon 等函数方法,简单、快速处理闪电资料,区分区域内外数据。

2 区域闪电频数统计方法基础及研究思路

基于 Matlab 强大的数据处理和制图功能,运用 inpolygon 函数和 shaperead 函数将需要处理的数据在边界限制下,区分为区域内(包括边界上)和区域外数据,而后对区域内(包括边界上)数据进行统计。

2.1 预处理

预处理包括在闪电资料中提取落雷点在直角坐标系中的位置信息和创建边界文件。表 1 所示为江苏省 2012 年部分闪电原始资料,反映其基本格式。随着闪电定位仪的升级,有些闪电原始资料已经定位到县级,但目前仍有许多闪电定位系统只能给出落雷点的经纬度,无法判断落雷点所在的省、市、县。

将闪电资料的每一列转换成字符串格式,应用 textread 函数读取闪电资料,提取需要的信息,再将其转换回数字格式,最终输出。主要程序如下:

```
wj = strcat(num2str(kk),'.txt');
[xuhao, riqi, time, n1, n2, n3, b1, b2, b3] = textread(wj,'%s %s %s %s %s %s %s %s %s %s');
n = size(n1, 1); % n 为 n1 的行数
for i = 1 : n
s1 = n1{i, 1}; l1 = length(s1); % 纬度
s2 = n2{i, 1}; l2 = length(s2); % 经度
s3 = n3{i, 1}; l3 = length(s3); % 强度
.....
```

表 1 江苏省 2012 年部分闪电原始资料

序号	日期	时间	北纬	东经	强度/kA	陡度/(kA/ μ s)	误差/%
0	2012-02-05	15:39:19.7711777	29.8927	113.9688	239.2	33.6	73.6
1	2012-02-05	15:46:30.6500521	28.0758	112.9953	-162.2	-25.0	0.0
2	2012-02-05	18:26:54.3438596	30.2823	115.1317	11.9	26.5	0.0
3	2012-02-05	22:09:42.0556855	29.7236	117.6521	138.0	19.2	73.7

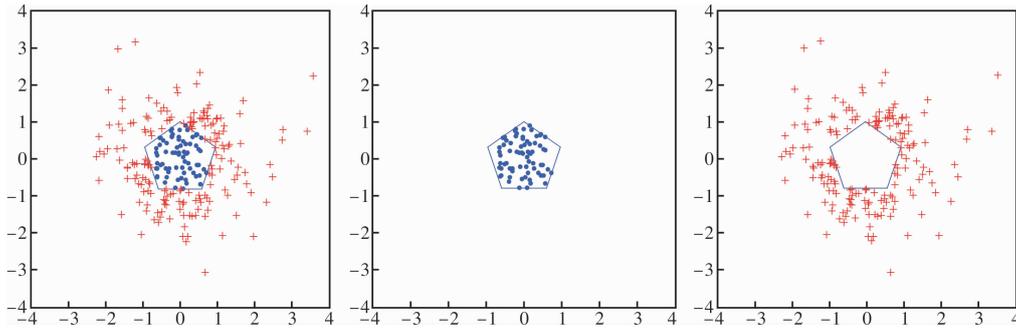


图1 区域内外数据区分图

```

w = str2num(s1(4: 11));      % w 为纬度数据
j = str2num(s2(4: 11));      % j 为经度数据
q = str2num(s2(4: 11));      % q 为经度数据
.....
wd(i, 1) = w;
jd(i, 1) = j;
qd(i, 1) = q;
.....
end
sd = [yf, rq, sk, jd, wd, qd];

```

表2所示为表1中数据经过上述程序处理后的待统计闪电资料。

月份/月	日期/日	时刻/时	东经/(°)	北纬/(°)	强度/kA
2	5	15	113.968 8	29.892 7	239.2
2	5	15	112.995 3	28.075 8	-162.2
2	5	18	115.131 7	30.282 3	11.9
2	5	22	117.652 1	29.723 6	138.0

边界文件的创建,应用MeteoInfo或ArcGIS等软件将所需地图从基准地图中选取,输出为shape文件。由shaperead函数读取shape文件的主文件作为边界文件。

2.2 shaperead 函数

shaperead函数的作用是从shape文件中读取矢量特性和属性。它的用法是:

$S = \text{shaperead}('a. \text{shp}')$ 。(1)

读入一个shape文件,主文件的后缀是.shp,文件名为a,返回n+1个地理数据结构数组在地图上的投影坐标(一个地图结构)。地理数据结构结合了几何和特征属性信息。shaperead函数支持普通二维形状类型:“点”、“多点”、“多线”和“多边形”。

2.3 inpolygon 函数

inpolygon函数^[24]的作用是判断一个点是否存在于一个多边形内。它的用法是:

$[IN \ ON] = \text{inpolygon}(X, Y, xv, yv)$ 。(2)

其输入值X和Y分别表示测试点的横坐标值和纵坐标值,而xv和yv则分别表示多边形定点的横坐

标值和纵坐标值。若某一点在多边形内,则IN返回值为1,否则IN为0,若某一点在多边形的边界上则ON为1,若不在边界上则ON为0。

inpolygon函数要求定义多边形的点必须沿着多边形连续,即由一系列点围成一个封闭图形。如图1所示,底图为正五边形,“+”和“·”为原始数据点,可以清晰的反映出“·”为区域内(包括边界上)数据,“+”为区域外数据,将这两部分数据分为两个文本文件导出。

3 在闪电资料处理中的应用

气象上最常见的等值线描绘问题常常仅在一个行政区域内进行,而不是一个规则区域。图2中以江苏省为例,制作江苏省闪电频数分布图时,以江苏省边界为底图白化提高准确性,这是可行的。但不能定量分析总闪、负闪和正闪的频次。

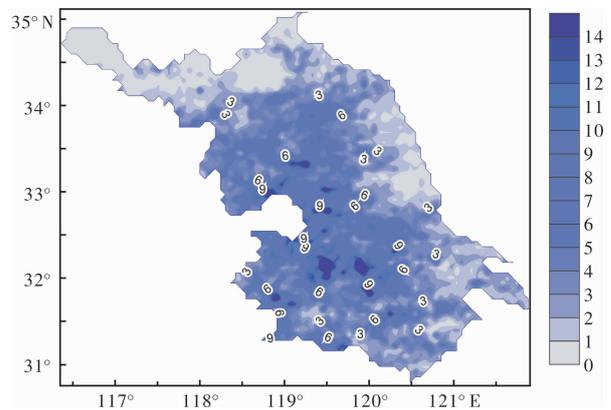


图2 2012年江苏省闪电频数分布图

传统方法统计数据存在于行政区域经纬度跨度形成的矩形区域内。江苏省经纬度为(图3): $116^{\circ}18' \sim 121^{\circ}57'E, 30^{\circ}45' \sim 35^{\circ}20'N$,其形成的矩形面积约为 $31.91 \times 10^4 \text{ km}^2$,但江苏省面积约为 $10.26 \times 10^4 \text{ km}^2$,约占了矩形面积的32.15%。假设闪电数据平均分布在省内,传统方法统计的闪电理论上近67.85%在省外,误差为211.04%。考虑到闪电定位系统的探测效率问题,闪电集中在省内,真实误差比理论误差小。

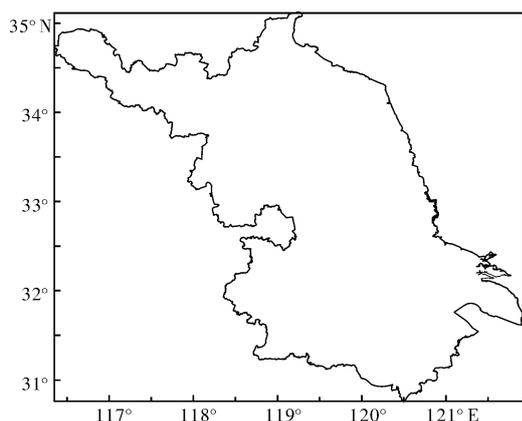


图3 江苏省地理底图

将 2012 年江苏省闪电资料经过上述程序运行后统计得出, 在 313 d 中, 共有 819 594 次闪电, 将省外数据剔除后得到省内(包括边界上)数据, 共有 413 566 次, 约占了总数的 50.46%, 误差达到 98.18%。

表 3 所示为江苏省 2012 年 12 个月省内闪电数量与原始闪电数量, 以及它们之比。闪电资料中省内闪电比重最大的是 7 月, 约占了 58.25%, 其次是 8 月, 约占了 51.30%; 最小的为 1 月, 所测得的数据均为省外数据, 其误差也最大, 其次为 2 月, 达到 1 182.05%; 误差最小的为 7 月, 近 71.67%。其中, 1 月、2 月、3 月、5 月、9 月、11 月和 12 月这 7 个月的误差大于传统方法统计的理论误差。可能原因是这几个月闪电数据相对较少且具有随机性, 使统计结果不具普适性。

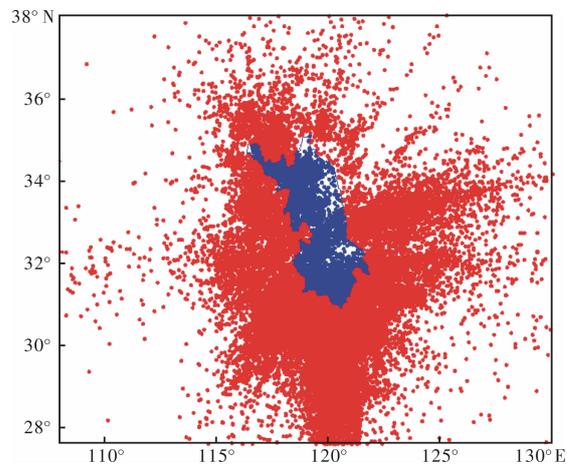
表 3 江苏省 2012 年 12 个月省内闪电数量与原始闪电数量之比

月份/月	sdin	sd	比例/%	误差/%
1	0	20	0	--
2	55	705	7.80	1182.05
3	2 066	7 358	28.08	256.13
4	2 576	6 719	38.34	160.82
5	534	3 108	17.18	482.07
6	15 433	39 774	38.80	157.73
7	206 624	354 710	58.25	71.67
8	157 201	306 436	51.30	94.93
9	24 985	88 264	28.31	253.23
10	3 595	10 487	34.28	191.72
11	318	1 125	28.27	253.73
12	179	888	20.16	396.03
总数	413 566	819 594	50.46	98.18

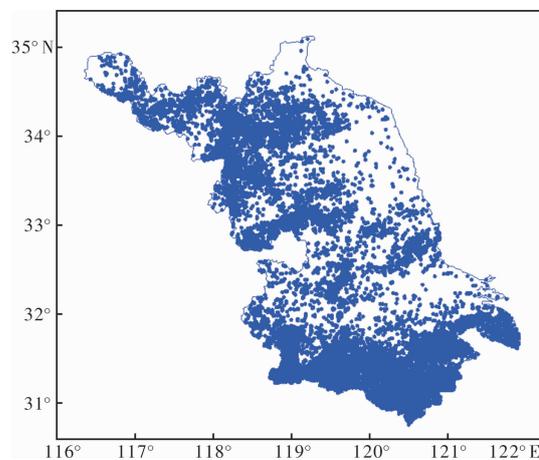
注: sdin 为省内闪电数量; sd 为原始闪电(即矩形内)数量。

图 4 为 2012 年 9 月江苏省内外和江苏省落雷点对比图。据统计, 省内数据占 9 月份所有闪电数据仅约 28.31%。图 4(a) 中红色“+”为省外数据, 在蓝色边界内的“·”为省内数据, 从图中可明显

看到大部分落雷点都处于区域外。图 4(b) 为江苏省内 9 月份落雷点图, 落雷点均在蓝色边界内, 即全处于省内。将图 4(b) 中闪电数据导出作为雷电灾害风险评估和一系列依赖于闪电资料科学研究的原始资料。



(a) 江苏省内外落雷点



(b) 江苏省落雷点

图 4 2012 年 9 月江苏省内外和江苏省落雷点对比图

赵旭寰^[8]统计出 2005 年 9 月 1 日至 2006 年 8 月 17 日, 江苏省气象局闪电定位系统在 291 个闪电日内共监测到闪电 711 866 次, 用本方法统计出其中真正的江苏省内闪电仅为 343 963 次, 约占监测到闪电的 48.32%。

4 结语

针对传统方法统计某区域内的闪电频次的不足, 提出应用于处理气象资料判定点是否位于行政区域内的新方法。创新的将气象、地理等学科经常使用的地理底图直接用作区域边界, 基于 in-polygon 函数, 通过将待测点坐标与边界坐标值作比较, 快速判断点与边界的位置关系, 导出区域内数据。以江苏省为例, 证实传统方法统计的闪电资料存在很大误差, 2012 年江苏省总闪频数误

差近 98.18%，使用本研究所得方法处理后，该类误差得到消除，为雷电灾害风险评估和雷暴分布规律分析提供真实有效的闪电资料，将大大提高雷电灾害风险评估和雷暴分布规律分析的准确性和可靠性。今后只需考虑改善闪电定位系统自身探测效率和减小定位误差问题。

参考文献:

- [1] 范广平, 王建, 扬继安, 等. 雷达与闪电定位资料在高炮人工防雷中的应用[J]. 气象, 2002, 28(4): 40-41.
- [2] 姚叶青, 魏鸣, 王成刚, 等. 一次龙卷过程的多普勒天气雷达和闪电定位资料分析[J]. 南京气象学院学报, 2004, 27(5): 587-594.
- [3] 龚静, 郭三刚, 何生存. 天气雷达和闪电定位仪在青海东部冰雹云监测中的应用[J]. 青海科技, 2005, 12(3): 24-27.
- [4] 胡新华. 人工观测与闪电定位探测雷暴日资料的对比分析[J]. 江西气象科技, 2005, 28(3): 55-56.
- [5] 冯桂力, 陈文选, 刘诗军, 等. 山东地区闪电的特征分析[J]. 应用气象学报, 2002, 13(3): 347-355.
- [6] 郑栋, 孟青, 吕伟涛, 张义军, 等. 北京及其周边地区夏季地闪活动时空特征分析[J]. 应用气象学报, 2005, 16(5): 639-644.
- [7] 李建华, 郭学良, 肖稳安. 北京强雷暴的地闪活动与雷达回波和降水的关系[J]. 南京气象学院学报, 2006, 29(2): 228-234.
- [8] 冯民学, 韦海容, 焦圣明, 等. 南京市闪电定位资料的对比分析[J]. 南京气象学院学报, 2008, 31(2): 151-157.
- [9] 张阳, 张义军, 孟青, 等. 北京地区正地闪时间分布及波形特征[J]. 应用气象学报, 2010, 21(4): 442-449.
- [10] 赵旭寰. 江苏省雷电分布规律及预报研究[D]. 南京: 南京信息工程大学, 2008.
- [11] Richard E Orville, Gary R Huffines, William R Burrows, et al. The North American lightning detection network (NALDN)-first results: 1998-2000[J]. Mon. Wea. Rev., 2002, 130(8), 2098-2109.
- [12] 张腾飞, 尹丽云, 许迎杰, 等. 2007年5-8月云南省雷电活动特点和致灾因子分析[J]. 灾害学, 2009, 24(1): 73-79.
- [13] 程琳, 周俊驰, 许蓓, 等. 2001年度江苏省雷电活动及雷击灾害特征[J]. 南京信息工程大学学报: 自然科学版, 2013, 5(2): 139-142.
- [14] 刘平英, 杨荣建, 李兆华, 等. 喀斯特地貌景区的雷击风险分析——以云南石林为例[J]. 灾害学, 2013, 28(2): 156-160.
- [15] 樊灵孟, 李志峰, 何宏明. 等. 雷电定位系统定位误差分析[J]. 高压技术, 2004, 30(7): 61-63.
- [16] 朱颀, 王振会. 一种确定雷电定位误差的实用方法[J]. 甘肃科技, 2007, 23(10): 80-82.
- [17] 张文娟, 孟青, 吕伟涛, 等. 时间差闪电监测网的误差分析和布局优化[J]. 应用气象学报, 2009, 20(4): 402-410.
- [18] 胡志祥. 雷电定位算法和误差分析理论研究[D]. 武汉: 华中科技大学, 2012.
- [19] Franco P Preparata, Michael I Shamos. Computational Geometry: An introduction [M]. New York: Springer-Verlag, Inc. New York, 1985.
- [20] Foley J D, Van Dam A, Feiner S K, et al. Computer graphics principles and practice[M]. 2nd ed. Addison-Wesley, Reading, MA, 1990.
- [21] Eric Haines. Point in polygon strategies [C]//Pau heckbent. Graphics Gems IV. Waltham: Academic Press, 1994. 24-46.
- [22] Kai Hormann, Alexander Agathos. The point in polygon problem for arbitrary polygons [J]. Computational Geometry: Theory and Applications, 2001, 20(3): 131-144.
- [23] 钟晓平. 点与闭合曲线所围区域关系的客观判定方法[J]. 应用气象学报, 2003, 14(1): 125-128.
- [24] 刘锋涛, 凡友华. 基于 MATLAB 实现二维 delaunay 三角剖分[J]. 工程地质计算机应用, 2008, 51(3): 34-36.

Inpolygon-based Elimination Method of Regional Lightning

Yang Zhongjiang^{1,2}, Jiang Ke^{1,2}, Cao Hongliang^{1,2} and Lou Jiayi^{1,2}

(1. Key Laboratory for Aerosol-Cloud-Precipitation of China Meteorological Administration, Nanjing University of Information Science & Technology, Nanjing 210044, China; 2. School of Atmospheric Physics, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract: The integrity and accuracy of lightning data is the basis of lightning disaster risk assessment, which influence the accuracy of the results, and series of scientific research depended on the lightning data. The data which is outside of the area will not affect the accuracy of inquiry when exploring regional distribution of lightning, but when time distribution of lightning is explored, only the data inside of the area is valid. There are serious errors using the traditional method of lightning frequency statistics in an area, on behalf of a rectangular area bounded by latitude and longitude. Inpolygon function method is firstly applied to distinguish between interior and exterior lightning data of statistical area simply and quickly, using the geographical map of provinces, cities and counties as boundary, thus the accuracy of lightning raw data will be greatly increased. Taking Jiangsu province as an example, lightning frequency counted by the traditional method are compared with the new inpolygon function method. The results show that in 2012, the lightning data of Jiangsu province counted by the new method account for 50.46% of the traditional method, which means the error of the traditional method, is as high as 98.18%. We can get real and effective lightning frequency data in the province by eliminating those outside the province, and greatly improve the accuracy and reliability of the lightning disaster risk assessment and thunderstorm distribution analysis.

Key words: lightning data; shaperead function; inpolygon function; geographical map; error