The Method for Regional Lightning Frequency Statistics Based on the Inpolygon*

Yang Zhongjiang[1,2] ,Jiang Ke[1,2]

1. Key Laboratory for Aerosol-Cloud-Precipitation of China Meteorological Administration，Nanjing University of Information Sci & Tec, Nanjing 210044, P. R. China;

2. School of Atmospheric Physics，Nanjing University of Information Sci & Tec, Nanjing 210044, China;

ABSTRACT: The integrity and accuracy of lightning data is the basis of lightning disaster risk assessment, they 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 we explore the time distribution of lightning , 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. We first apply inpolygon function method 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, we compared lightning frequency counted by the traditional method 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.

KEYWORDS: Lightning data; Shaperead function; Inpolygon function; Geographical map; Error

INTRODUCTION

We often encountered this kind of circumstance in meteorology, geography subject and so on: the data have a value in a given area, and the data out of area have not to consider. If we encounter this situation, first of all, we need to make sure that it is inside area data or outside area data, then eliminate outside area data. It is relatively easy to judge the relationship of position between point and plane when the given area boundary can be described by simple,

---

* The work is supported by The National Natural Science Foundation of China (Grant No. 41175003)

A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD)

Contact information: Yang Zhongjiang, Institution, Nanjing University of Information Sci & Tec, Nanjing 210044, China, Email: mashimaro2974@126.com

sure and regular mathematical model. But it is very hard to handle the data need in meteorology, geography subject and so on, because they often occur in some irregular administrative area whose boundary cannot be described by simple mathematical model. Artificial process is a feasible method when only a small amount of data, but when there is a huge amount of data (such as lightning data), artificial process becomes more difficult, usually with the aid of computer.

Lightning disaster damage increases year by year, lightning detection is more and more attention by people, lightning monitoring and location plays an increasingly important role in the meteorological observation [1-4].Lightning location system can be used to get to the parameters such as geographical position, the intensity and polarity of lightning flashing, lightning frequency and so on, which is applied to the analysis of the time and spatial distribution of lightning activities [5-10]. Orville, et al. [11] analyzed cloud-to-ground lightning in North America from 1998 to 2000, counted that a total of 88.7 million lightning strikes in three years. There were 31 million times, 29 million times and 28 million times in 1998, 1999 and 2000 respectively; Zhang Tengfei, et al. [12], Cheng Lin, et al. [13] used the lightning data in Yunnan province and Jiangsu province respectively analyzed regional lightning activity time distribution, cloud-to-ground lightning frequency time distribution and so on; In analysis of lightning risk about karst landform scenic spot which toke Shilin who is belong to Yunnan for example, Liu Pingying, et al. [14] used nearly 35 years thunderstorm statistical data of Shilin County, lightning location system observation data and thunder disaster reporting data of Yunnan province from 2006 to 2010, combined with the special geographical location of Shilin scenic spot, karst topography and geological structure features, analyzed thunderstorms characteristic parameters in the scenic spot, including the intensity, density and frequency of lightning and so on. But there are bigger problems in the accuracy of lightning data used in the above research. One reason is that lightning location system itself exists error [15-18]. Another is the lightning statistical data is not completely in a particular area, also means the selection of lightning data exist error. Take Jiangsu province for example, comparing the border surrounded by latitude and longitude of a rectangular area of Jiangsu province with the actual boundary around area of Jiangsu province, the difference has almost doubled. So there will be obviously a large error when take the lightning frequency of former area to represent actual lightning frequency in Jiangsu province. The statistical error is changed by different boundary shape of the area, the longer and narrower the borders is, the greater the lightning frequency statistical error is.

Lightning location system itself needs to continuously eliminate error through improved device hardware and positioning algorithms. So it is an important way to improve the accuracy of the data by eliminating lightning data outside the area. Some software (such as ArcGIS, Surfer, et al.) can eliminate outside data from the produced figure, but cannot export them in a text. This study divides the raw lightning data by applying the inpolygon function of Matlab into two parts: the data inside the area (including the border) and the data outside area, almost can completely eliminate the lightning data outside the boundaries of area, completely eliminate the data selection error causes by the second reason given above, to provide a reliable basis for lightning disaster risk assessment and a series of scientific research which depends on the lightning data, also reduce the workload of post-processing and statistics.

POINT IN POLYGON

The application of polygon point is found more widely in the field of meteorology, geography and so on, it is the basic method of handling the meteorological, geographic data. In 1985, Preparata and Shamos put forward that algorithm for point in polygon is one of the most fundamental problems in computational geometry, it can be a range application in the field of computer graphics, geographic information system .The question can be described as: A given polygon P and arbitrary point Q, and then judge whether the point Q is inside the polygon P or not [19].

The classic methods of the judgment about point in polygon are: one is by scholars, such as Foley, Van Dam and Haines [20, 21] using the classical ray-crossing method, across a test points ends a ray to any angle, if the sum of the node number with the each side of the polygon is even, then the test point is outside the polygon, otherwise the point is within the polygon. If you have any intersection of polygon across the vertices, then you should choose a ray for a recalculation. This method is simple, reliable, and it can handle any arbitrary polygon, it is often used as a benchmark algorithm compared with other algorithms. But it is difficult to handle the boundary points in ray, boundary line with ray. Another is the cumulative angle method used by Preparata, Hormonn and Zhong Xiaoping[19,22-23], relative to the closed curve of area, if the test point and the attachment point of the curve relative to the angle formed by X axis increment accumulative total value is 360°, the point will be located within the region; If the total value is 0°, the point will be located outside the region. Although there are a large number of detection algorithm of point in polygon, looking for suitable for the field requirements solution is an important research content.

The solution above does not apply to judge the relationship between flash point and boundary in the lightning data, the reasons are the following: ①the lightning data is not only include the location information of the ground lightning flashpoint in the rectangular coordinate system, but also including the parameters like time , intensity, slope, error and so on; ②we use geographic base map of the province as the boundary to add up the lightning data in the province, city or county, but it contains thousands of coordinates points which are reproduction of its latitude and longitude;③It contains hundreds , thousands or even millions lightning data in every year, it is a large amount of data. To study the regular pattern of the time and spatial distribution of lightning, years of the statistical data can correctly reflect the laws. Although the principle of the solution is simple, the implementation process is very complicated, the time is longer. So, this article apply inpolygon function of Matlab ,which can handle lightning data simply and rapidly, and distinguish the data into inside and outside of the area.

STATISTICAL BASIS AND RESEARCH METHODS

Based on powerful data processing and mapping function of Matlab, using inpolygon function and shaperead function divide the data who need to handle into inside (including the border) and outside data under the boundary limit, then counting the data who are inside of area (including the border).

1) Preprocessing

Preprocessing include extract position information about ground lightning flash point in the rectangular coordinate system and create boundaries file. Table 1 reflect the basic format what the part of raw lightning data of Jiangsu province in 2012.

Table 1 Part of raw lightning data of Jiangsu province in 2012

No.	Date	Time	North latitude	East longitude	Intensity /kA	Slope (kA/μs)	Error/ %
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

Turn each column of the lightning data into a string format, apply textread function reads the lightning data, then extract the needed information, and convert them back into digital format, at last output them. The main program is as follows:

```
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');
n=size(n1,1);                %n is row of n1
for i=1:n
s1=n1{i,1}; l1=length(s1);     % latitude
s2=n2{i,1}; l2=length(s2);     % longitude
s3=n3{i,1}; l3=length(s3);     % intensity
 ……
w=str2num(s1(4:l1));         % w is latitude data
j=str2num(s2(4:l1));          % j is longitude data
q=str2num(s2(4:l1));         % q is intensity data
y1=riqi{i,1};                 %y1 is month
t1=time{i,1};                %t1 is time
wd(i,1)=w;
jd(i,1)=j;
qd(i,1)=q;
y=str2num(y1(6:7));
t=str2num(t1(1:2));
end
sd=[yf,rq,sk,jd,wd,qd];
```

Data in the Table 2 is the lightning data need to count which Tab.1 after the above process.

Table 2 The lightning data need to count

Month	Day	Hour	North latitude	East longitude	Intensity /kA
2	5	15	113.9688	29.8927	239.2
2	5	15	112.9953	28.0758	-162.2
2	5	18	115.1317	30.2823	11.9
2	5	22	117.6521	29.7236	138.0

Create boundary files can apply MeteoInfo or ArcGIS software select the needed map from the benchmark in the map, and output for shape file. Let shaperead function read the master file as a boundary shape file.

2) Shaperead function

Shaperead function is aimed at reading the vector features and properties from shape file. Its usage is: S=shaperead ('a.shp') reads in a shapefile, filename is a, and returns an N-by-1 geographic data structure array in projected map coordinates (a map struct). The geographic data structure combines geometric and feature attribute information.

Shaperead supports the ordinary 2-D shape types: 'Point', 'Multipoint', 'PolyLine', and 'Polygon'.

3) Inpolygon function

Inpolygon function[24] is aimed at judging whether a given point in a polygon or not. Its usage is:[IN ON] = inpolygon(X, Y, xv, yv). The input values of X and Y, respectively represents abscissa and ordinate values of test point, whereas xv and yv respectively represents abscissa and ordinate values of polygon fixed-point. If the given point is in the polygon, IN return to 1, otherwise to 0, that if a point on the boundary of the polygon, ON return to 1, otherwise to 0.

Inpolygon function requires the points who definite polygon must be continuous along the polygon, also means the closed polygon consists of a series of points. As shown in fig.1, the base map is a pentagon, "+"and"●"are original data points that can reflect clearly that "●" are the inside (including the border) data," + "are the outside data, divide these two parts data into two text files and export them.

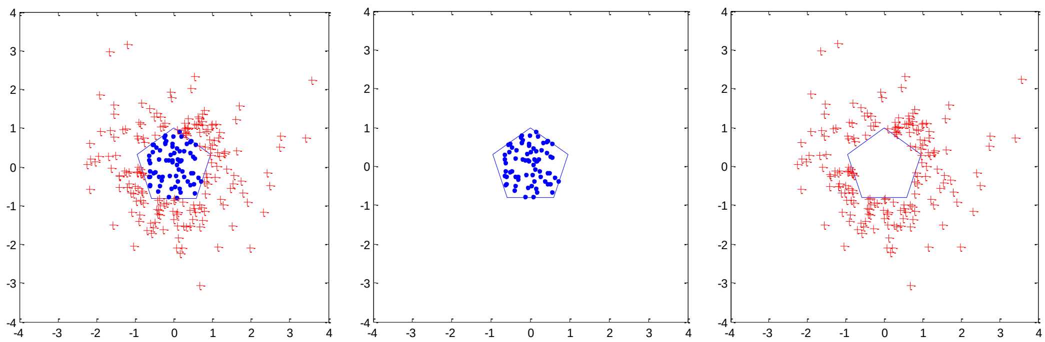
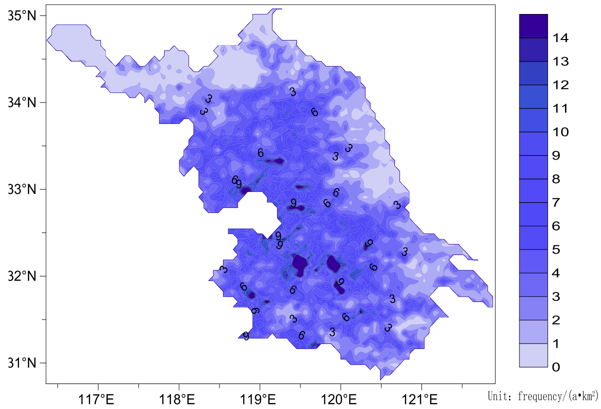


Fig.1 Distinguish the area figure is internal or external data

APPLICATION OF PROCESSING LIGHTNING DATA

The most common problems about describing isolines in the meteorology often conducted only in an administrative area, rather than a rule area. As shown in fig.2, take Jiangsu province for example, it is feasible that take Jiangsu province borders for blanking base map to improve accuracy when make the figure about lightning frequency distribution of Jiangsu province. But we cannot quantitatively analyze total lightning frequency, positive lightning frequency and negative lightning frequency.



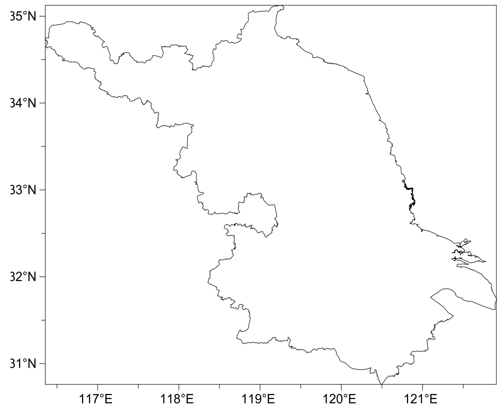


Fig.2 Lightning frequency distribution of Jiangsu province in 2012

Fig.3 Geographic base map of Jiangsu province

The statistical data by traditional method exists in the area where administrative area of latitude and longitude span formed the rectangular. The latitude and longitude of Jiangsu province (as shown in fig. 3)is: east longitude is $116°18'\sim121°57'$, north latitude is $30°45'\sim35°20'$, the formation of the rectangular area is about 319,100 square kilometers, but the area of Jiangsu province is about 102,600 square kilometers, accounts for about 32.15% of the rectangle. Suppose lightning data distribute average in the province, the statistical lightning data of traditional method nearly 67.85% out of the province in theory, the error reach to 211.04%. Considering the detection efficiency of lightning location systems, lightning concentrated in the province, the real error is smaller than theoretical error.

There are 819,594 lightning strikes in 313 days in Jiangsu province in 2012 after the program run and count. There are 413,566 lightning strikes in the province (including the border) after eliminating data outside of province, the inside data account for about 50.46% of the total, the error reach to 98.18%.

Table 3 is the number and ratio of the lightning number in the province and the original about Jiangsu province in every month in 2012.The largest proportion about lightning data in the province occupied all lightning data is July, accounts for about 58.25%, followed by August, accounts for about 51.30%; The smallest is January, the measured data are all outside of the province, its error is also the largest, followed by February, error is 1182.05%; The smallest error is July, nearly 71.67%. Among them, the error in seven months (January, February, march, may, September and November and December) is greater than the theoretic error statistics by traditional method. Possible reason is that the lightning data relatively few and random, so that the statistical result is not universal.

Table 3 The ratio of the lightning number in the province and the original about Jiangsu province in every month 2012

month	sdin	sd	ratio	error/%
1	0	20	0	--
2	55	705	7.80%	1182.05
3	2066	7358	28.08%	256. 13
4	2576	6719	38.34%	160.82
5	534	3108	17.18%	482.07
6	15433	39774	38.80%	157.73

7	206624	354710	58.25%	71.67
8	157201	306436	51.30%	94.93%
9	24985	88264	28.31%	253.23%
10	3595	10487	34.28%	191.72%
11	318	1125	28.27%	253.73%
12	179	888	20.16%	396.03%
total	413566	819594	50.46%	98.18%

Note: sdin is the province lightning number; sd is the original(inside of the rectangle) number of lightning.

Fig.4 is the ground lightning point in Jiangsu province in September 2012. According to the statistics, the data in the province accounted for only about 28.31% of all lightning data in September. In Fig.4 (a) the red"●"are the data outside the province, the blue"●" are the data in the province within the boundaries, it is clearly see from the picture that most of the ground lightning point are outside region. Fig.4(b) is Lightning point of Jiangsu province in September 2012，ground lightning points are within the blue border, that means all in the province. Export the lightning data in figure 4 (b) as the original data for lightning disaster risk assessment and a series of scientific research which depends on them.

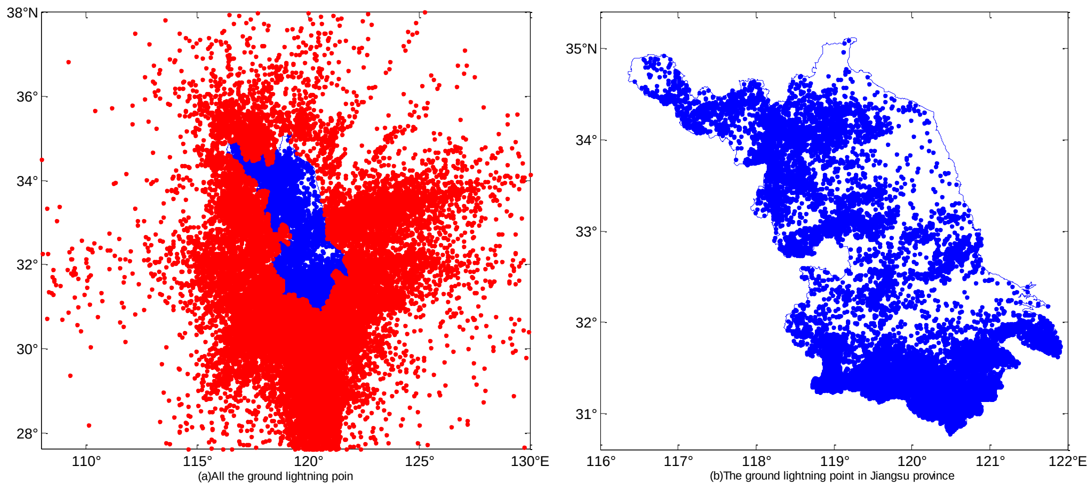


Fig.4 Contrast the ground lightning point in Jiangsu province with Jiangsu province inside and outside in September 2012

Zhao Xuhuan [10] counted 711,866 times lightning in 291 lightning days from September 1, 2005 to August 17, 2006 by lightning location system of Jiangsu province meteorological bureau, however, there were only 343,963 times real lightning in Jiangsu province with the method of this article, about 48.32% of detected lightning.

CONCLUSION

In view of the deficiency about lightning frequency statistics in a certain area by the traditional methods, put forward a new method which apply to judging whether the given point locate in the administrative area or not when handle the meteorological data. Innovatively, directly use the geographical area, such as weather, geography discipline and so on, as the border reproduction, based on inpolygon function, compare the measuring points coordinates with boundary coordinates to judge quickly the relationship about location between the point and boundary, at last export area data. Take Jiangsu province for example, confirmed that there is a big error of lightning data counted by the traditional method, and the total flash frequency error of Jiangsu province in 2012 is nearly 98.18%. This kind of error have be eliminated after processing by method in this study which provide real and effective lightning data for the lightning disaster risk assessment and analysis of the distribution of thunderstorm, it will greatly improve the accuracy and reliability about them. What we need to consider in future is improving the efficiency of lightning detection and decreasing positioning error.

REFERENCE

[1] Fan Guangping; Wang Jian; Yang Ji'an; et al. 2002: The Application of Radar and Lightening Positioner to Hail Suppression by Antiaircraft Gun. *J. Meteorological Monthly*, **28(4)**,40-41.

[2] Yao Yeqing; Wei Ming; Wang Chenggang; et al. 2004: A Case Analysis of Tornado Based on Doppler Radar Data and Lightning Location Data. *J. Journal of Nanjing Institute of Meteorology*, **27(5)**,587-594.

[3] Gong Jing; Sangang; He Shengcun; 2005: The Application of Weather Radar and Lightening Positioner to Hail-cloud Monitoring in the Eastern of Qinghai. *J. Qinghai Science and Technology*, **12(3)**,24-27.

[4] Hu Xinhua; 2005: The Contrast And Analysis of the Thunderstorm Day Data of Artificial Observation and Lightning Orientation. *J. Jiangxi Meteorology Science & Technology*, **28(3)**,55-56.

[5] Feng Guili; Chen Wenxuan; Liu Shijun; et al. 2002: Preliminary Analysis of Lightning Characteristics in Shandong. *J. Quarterly Journal of Applied Meteorology*, **13(3)**,347-355.

[6] Zheng Dong; Meng Qing; Lü Weitao; et al. 2005: Spatial and Temporal Characteristics of Cloud-to-Ground Lightning in Simmer in Beijing and Its Circumjacent regions. *J.Quarterly Journal of Applied Meteorology*, **16(5)**,639-644.

[7] Li Jianhu; Guo Xueliang; Xiao Wen; et al. 2006: Relationship of Cloud-to-ground Lightning Activities with Radar Echo and Precipitation in Storms of Beijing. *J. Journal of Nanjing Institute of Meteorology*, **29(2)**,228-234.

[8] Feng Minxue; Wei Hairong; Jiao Shengming; et al. 2008: A Comparative Analysis of the Lightning Strike Positioning Data in Nanjing. *J. Journal of Nanjing Institute of Meteorology*, **31(2)**,151-157.

[9] Zhang yang; Zhang Yijun; Meng Qing; et al. 2010: Temporal Distribution and Waveform Characteristics of Positive Cloud-to-ground Lightning in Beijing Area. *J. Quarterly Journal of Applied Meteorlolgy*, **21(4)**,442-449.

[10] Zhao Xuhuan; 2008: The Research on Lightning Activities and Forecasting over Jiangsu Province. *D. Nanjing University of Information Science & Technology*.

[11] Richard E.Orville; Gary R. Huffines; William R. Burrows; et al. 2002: The North American Lightning Detection Network (NALDN)—First Results: 1998–2000. *J. Mon. Wea. Rev.*, **130(8)**, 2098–2109.

[12] Zhang Tengfei; Yin Liyun; Xu Yingjie; et al. 2009: Analysis of Characteristics and Causes of Lightning Activity in May to August of 2007 in Yunnan. *J. Journal of Catastrophology*, **24(1)**,73-79.

[13] Cheng Lin; Zhou Junchi; Xu Bei; et al. 2013: Analysis on Characteristics of Lightning Activities

and Lightning Disasters in Jiangsu Province in 2011. *J. Journal of Nanjing University of Information Science & Technology (Natural Science Edition)*, **5(2)**,139-142.

[14] Liu pingying; Yang Rongjian; Li Zhaohua; et al. 2013: Analysis on Lightning Risk in Scenic Regions with Karst Landform——Taking the Stone Forest as an Example. *J. Journal of Catastrophology*, **28(2)**,156-160.

[15] Fan LingMeng; Li Zhifeng; He Hongming; et al. 2004: Analysis of Location Error of Guangdong Lightning Location System. *J. High Voltage Engineering*, **30(7)**,61-63.

[16] Zhu Biao; Wang Zhenhui. 2007: A Practical Method to Ensure the Lightning Location Error. *J. Gansu Science and Technology*, **23(10)**,80-82.

[17]Zhang Wenjuan; Meng Qing; Lü Weitao; et al. 2009: Error Analyses and Network Optimization for Time-of-arrival Lightning Locating System. *J. Quarterly Journal of Applied Meteorlolgy*, **20(4)**,402-410.

[18] Hu Zhixiang. 2012: Theoretical Study on Location Algorithm and Error Analysis of Lightning Targets. *D. Huazhong University of Science & Technology*.

[19] Franco P. Preparata; Michael I. Shamos. 1985: Computational Geometry: An Introduction. *Springer-Verlag* New York, Inc. New York.

[20] Foley J D; van Dam A; Feiner S K,etal. 1990: Computer Graphics Principles and Practice. 2nd ed. *Addison-Wesley*, Reading, MA.

[21] Eric Haines. 1994: Point in polygon strategies. *J. Graphics Gems IV*, ed. Paul Heckbert, Academic Press,24-46.

[22] Kai Hormann; Alexander Agathos. 2001: The point in polygon problem for arbitrary polygons. *J. Computational Geometry: Theory and Applications*, **20(3)**,131-144.

[23] Zhong Xiaoping. 2003: The Objective Judgement Method of the Relationship between Point and Inpolgon. *J. Quarterly Journal of Applied Meteorology*, **14(1)**,125-128.

[24] Liu Fengtao; 2008: Fan Youhua. Realize Two-dimensional Delaunay Triangle Subdivision Based on the MATLAB. *J. Engineering Geology Computer Application*, **51(3)**,34-36.