Article ID: 1009 5020(2005) 03 183 06 **Document code:** A

Relationship of Uncertainty Between Polygon Segment and Line Segment for Spatial Data in GIS

LIU Chun TONG Xiaohua

ABSTRACT The mathematic theory for uncertainty model of line segment are summed up to achieve a general conception, and the line error band model of ε_σ is a basic uncertainty model that can depict the line accuracy and quality efficiently while the model of ε_m and error entropy can be regarded as the supplement of it. The error band model will reflect and describe the influence of line uncertainty on polygon uncertainty. Therefore, the statistical characteristic of the line error is studied deeply by analyzing the probability that the line error falls into a certain range. Moreover, the theory accordance is achieved in the selecting the error buffer for line feature and the error indicator. The relationship of the accuracy of area for a polygon with the error loop for a polygon boundary is deduced and computed.

KEY WORDS spatial data; ε error band; polygon segment; uncertainty
CLC NUMBER P208

Introduction

The uncertainty of spatial data in GIS is regarded as one kind of generalized error that includes measurable and immeasurable error or numerical and concept error. The random of spatial data, complexity, illegibility and inconsistence of the data are recognized as the main contents of the uncertainty issue, so that the uncertainty of spatial data has a rather large research area. However, the measurement indicator on the uncertainty, presentation format and transmit rules are commonly considered as the key issues in the theory discussion on the uncertainty of spatial data, so such issues are paid more emphasis on during the application process in GIS. The point, line and polygon are defined as the fundamental elements for spatial data in a vector GIS, so the recearch of the spatial data in GIS is mainly put on the uncertainty of point, the uncertainty of line and the uncertainty of polygon.

The accuracy / the uncertainty of a plane point is usually described by root mean squared error or position error of the point. Coordinate root mean squared error, point root mean squared error and error ellipse are used to give a full description of the error distribution in each direction for a two dimensional point[1].

Uncertainty presentation for error band model has become a research focus on the uncertainty of the plane line and curved line. The aim of the discussion on line error band model is to realize the error distribution of the plane line error and to give the description method for it[2].

The location of a polygon can be determined by its boundary, so the uncertainty of a polygon should be determined by its boundary line error band and the complicated nonlinear relationship between them may imcrease some difficulties. So the polygon segment uncertainty depicted by line segment will be discussed as a main content in this paper. Further, the relationship of spatial uncertainty among point, line and polygon are achieved during consideration of the polygon segment uncertainty model. The research of po-

Received on April 20, 2005.
Project supported by the National Natural Science Foundation of China (No. 40301043).
LIU Chun, Ph.D, post doctoral fellow, associate professor; Department of Surveying and Geo Informatics, Tongji University, 1239 Siping Road, Shanghai 200092, China; Key Laboratory of Advanced Engineering Survey of SBSM, Tongji University, 1239 Siping Road, Shanghai 200092, China.

sitional uncertainty on spatial data in GIS will be consummated and meaningful application with vector data uncertainty is also found.

1 Error band model for plane line segment

Plenty of research on line positional uncertainty model is propagated by the ε band indicator (the epsilon band), which was put forward by Perkal[3]. Chrisman[4] and Caspary[5] firstly began depict line segment uncertainty in vector GIS with Perkal's concept of 'ε band'; Dutton[6] later depicted the distribution of line segment uncertainty with simulation method; Shi[7,8] put forward the confidence region and probability distribution of the line segment.

As the enlargement of the ε band model, variance and covariance of an arbitrary point on the line is deduced on the basis of considering the error of two end points for the line, which has a hypothesis of the coordinate error irrelevance or coordinate error relativity, and the root mean squared error (σ in Fig. 1) of an arbitrary point on the direction that perpendicular to the line is expressed as the error band width. The ε band is entitled as ε_σ band or ε_σ model[2]. On the other hand, deep research indicates that the point on the line has a maximum root mean squared error direction (m in Fig. 1), so the ε_m band or ε_m model is achieved when the line segment error band is depicted with 'maximum direction root mean squared error'.

Separately, Fig. 2 gives the visualization presentation of the line error band of ε_σ model and ε_m model.

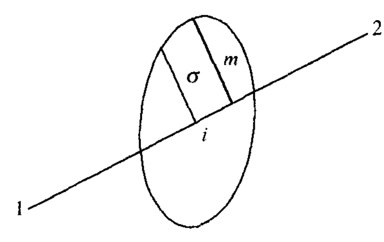

Fig. 1 Error ellipse for an arbitrary point on a line

Fig. 2 ε band model for line segment

From the visualization of error band, we can see that the accuracy may be low when the two ends of the line have a large position error whereas the accuracy will be high when the middle point on the line has a small position error. Clearly, the error band has a trend of inclining to one side, so the line error band may has a certain direction when using the error ellipse as the uncertainty model to establish the plane line error band. Under the instances for the different accuracy of two end points on the line, the shape of the line error band may change accordingly with the error of two end points. Besides, the line uncertainty credibility band[9] and information entropy error band[10] are all used to discuss the related issue surrounding line segment uncertainty model in GIS.

2 Probability analysis of plane line error band

It is useful to analyze the plane line error distribution status by discussing the probability of plane line segment error falling inside different range and to realize its influence of line segment uncertainty on GIS and its application.

Making a further discussion, according to the characteristic of distribution function of multi-dimension random variable, each point error on the line segment has a probability falling into the range of $\Omega(-\sigma_{xr}<x_r<\sigma_{xr};\ -\sigma_{yr}<y_r<\sigma_{yr})$ and can be represented as:

$$P_r\int_\Omega f(X)\,\mathrm{d}X=\int_\Omega\frac{1}{(2\pi)^{n/2}\mid D_x\mid^{\frac{1}{2}}}\cdot$$

$$\exp\left\{-\frac{1}{2}(X-\mu_x)^{\mathrm{T}}D_x^{-1}(X-\mu_x)\right\}\mathrm{d}X \quad (1)$$

where X is a vector composed by the coordinates of n points; D_x, μ_x are the corresponding mean and variance.

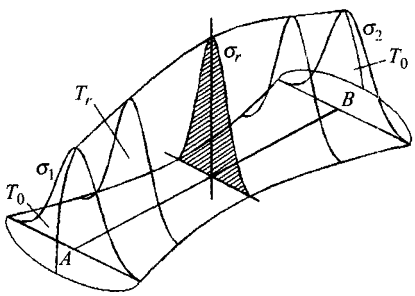

Fig. 3 Probability representation for the plane line error band

When the points on the line segment are adjacent to each other infinitely, all the points on the line will fall into the range Ω simultaneously, and will have the probability as a ratio of the volumes that showed in Fig. 3 and represented by the following equation:

$$P_r = \frac{T_r + T_0}{T_l + 1} \quad (2)$$

where T_r is the volume composed by the error curved surface with all points except the two ends perpendicular to the line segment inside the range of $\Omega_r(-\varepsilon < x_r < \varepsilon; 0 < y_r < l)$; T_l is the volume composed by the error curved surface inside the range of $\Omega_0(-\infty < x_r < \infty; 0 < y_r < l)$; T_0 is the volume composed by the point error curve of the two ends of the line segment. Thus we have

$$T_r = 2\int_0^{\sigma_r}\int_0^l f(x_r)\mathrm{d}x_r\mathrm{d}y_r = 2\int_0^1\int_0^1 \frac{l}{\sqrt{2\pi}}\cdot \exp\left\{-\frac{x_r^2}{2}\right\}\mathrm{d}x_r\mathrm{d}r$$

$$T_l = 2\int_0^{\infty}\int_0^1 \frac{l}{\sqrt{2\pi}\sigma_{xr}}\exp\left\{-\frac{x_r^2}{2\sigma_{xr}^2}\right\}\mathrm{d}x_r\mathrm{d}r = l \quad (3)$$

During the computation of the probability of the line segment falling into the different range, the line is taken as the range boundary and denoted by:

$$C = k\{(1-r)\sigma_{x_1} + r\sigma_{x_2}\} \quad (4)$$

If $k=1$, the probability that the plane line segment falling into the range is about 72%-78%; when $k=1.67$, the probability is about 92%-95%; $k=2$, it is about 96%-98%; $k=3$, it is more than 99%.

The selection of line error indicator and error band buffer should be determined according to the probability of line segment. Generally, the convenience shall be considered and the position of the plane line segment shall be ensured not to be influenced by other factors in GIS, so the parameter k shall be determined as 2, and thus we take the line

$$x = 2\{(\sigma_{x_2} - \sigma_{x_1})y/l + \sigma_{x_1}\} = 2\{(1-r)\sigma_{x_1} + r\sigma_{x_2}\} \quad (5)$$

as the upper boundary of the error buffer.

3 Polygon area accuracy and polygon error band model

3.1 Polygon area accuracy

If polygon has n margins and its vertexes is denoted as $p_i(x_i, y_i)(i=1, 2, \cdots, n)$, the variance of their coordinate is expressed as:

$$\boldsymbol{D}_i = \begin{bmatrix} \sigma_{x_i}^2 & \sigma_{x_iy_i} \\ \sigma_{x_iy_i} & \sigma_{y_i}^2 \end{bmatrix} \quad (6)$$

Assuming that each vertex's coordinates are independent, then the area of the polygon can be calculated by:

$$S = \frac{1}{2}\sum_1^n x_i(y_{i+1} - y_{i-1}) = \frac{1}{2}\sum_1^n x_i\Delta y_{i-1,i+1} \quad (7)$$

where $y_0 = y_n$, $y_1 = y_{n+1}$.

The differential coefficient equation is gained as follows.

$$\mathrm{d}S = \frac{1}{2}\sum_1^n(\Delta y_{i-1,i+1}\mathrm{d}x_i + \Delta x_{i-1,i+1}\mathrm{d}y_i) \quad (8)$$

The variance of area for the polygon is

$$\sigma_{s_1}^2 = \frac{1}{4}\sum_1^n(\Delta y_{i-1,i+1}^2\sigma_{x_i}^2 + \Delta x_{i-1,i+1}^2\sigma_{y_i}^2 + 2\Delta x_{i-1,i+1}\Delta y_{i-1,i+1}\sigma_{x_iy_i}) \quad (9)$$

Supposing that

$$\sigma_{x_i}^2 = \sigma_{y_i}^2 = \sigma_0^2, \sigma_{x_iy_i} = 0$$

then

$$\sigma_{s_1}^2 = \frac{1}{4}\sum_1^n(\Delta y_{i-1,i+1}^2 + \Delta x_{i-1,i+1}^2)\sigma_0^2 = \frac{1}{4}\sum_1^n l_{i-1,i+1}^2\sigma_0^2 \quad (10)$$

where $l_{i-1,i+1}$ is the distance between two adjacent points of the polygon.

For a regular n polygon whose edge length is l, it can be written as:

$$\sigma_{s_1} = \sqrt{n}\sin\left(\frac{\pi}{2} - \frac{\pi}{n}\right) \cdot l\sigma_0 \qquad (11)$$

3.2 The influence of boundary line segment error on polygon area

Assuming the boundary line of the polygon has a error band, and the area of the error band is Δ_i, so the total influence on the area of the polygon exerted by the error band is:

$$\Delta_s = \Delta_1 + \Delta_2 + \cdots + \Delta_i + \cdots + \Delta_n, \; i = 1, 2, \cdots, n \qquad (12)$$

If each error band is independent, then we have:

$$\sigma_{s_2}^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_i^2 + \cdots + \sigma_n^2$$
$$\sigma_{s_2} = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_i^2 + \cdots + \sigma_n^2} \qquad (13)$$

where σ_i is the standard error of the error band, and is also the half area of the line segment error band ε_σ, thus we have $\sigma_i = \Delta_i / 2$. For a regular n polygon whose edge length is l, the area error equation is

$$\sigma_{s_2} = \sqrt{\sum_{i=1}^{n} \sigma_i^2} = \sqrt{n} \cdot \frac{\Delta_i}{2n} = \frac{\Delta_i}{2\sqrt{n}} \qquad (14)$$

3.3 Area of the ε_σ error band of the polygon

The polygon is composed of the plane line segments, so the plane line segment ε_σ error band can exert the influence on the polygon segment which inflicted by the line segments. For a rectangle in Fig. 4, the area of the error band loop is

$$S = 4(A_i - \pi\sigma_{x_i}^2 - A_\Delta) \qquad (15)$$

where $A_i = 2l\int_0^1 \sigma_r \mathrm{d}r + \pi\sigma_{x_i}^2$;

$$\sigma_r = \sqrt{(1-r)^2\sigma_{x_i}^2 + r^2\sigma_{x_{i+1}}^2 + 2r(1-r)\sigma_{x_i x_{i+1}}};$$

and A_Δ is the area enclosed by the inner polygon boundary and the error ellipse.

3.4 Comparison between standard error of the polygon area and area of error band of polygon

By the above two methods, standard error of the area and the area of error band of several regular polygons or irregular polygons are calculated and the results are listed in Table 1, in which the standard error of the vertex of the polygon is 1/100 of the length of margin. The results gained by Eqs. (9) and (13) are presented with Fig. 5. Conclusions can be achieved from Table 1 and Fig. 5 as follows:

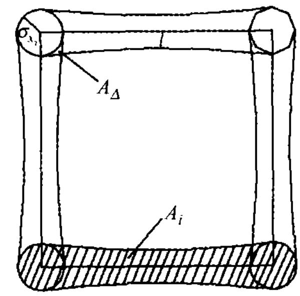

Fig. 4 Polygon error loop for a square

1) The difference is small between standard errors of the area for a polygon which are calculated separately by Eqs. (9) and (13), and the value may smaller than the ratio of 1/5 of the standard error of the area.

2) Calculation results with Eqs. (9) and (13) are very close to each other, which indicates that the error of the area for a polygon is affected by the line segment. The analysis on line segment error for each line of a polygon can effectively depict the error distribution and statistical characteristic of probability of the polygon area error.

3) The sum of ε_σ error band area for each line segment of a polygon segment can be regarded as a measurement indicator of polygon segment error. However, the mean and variance are two random variable measurement indicators, and the sum of line segment ε_σ error band area has a big difference comparing to the standard error of polygon area.

4) The two polygon area accuracy determination methods are also related with the area of the polygon. The difference value is calculated for the different cases for the area of the polygon, and is visualized in Fig. 6. The difference will be larger when the edge of the polygon is manifold.

4 Conclusions

Spatial analysis such as overlay analysis should depend on the fundamental elements such as point, line and polygon in GIS. So the uncertainty of spatial data may affect the reliability of the application development and decision making in GIS. The issue of uncertainty has been regarded as one of main basic theoretical research contents in GIS. The problem, to realize the uncertainty of the line segment and polygon segment, is resolved after a series of typical line segment uncertainty models are first summarized in this paper. The ε_σ error band model for a line segment is regarded as a basic uncertainty model used to reflect the line segment accuracy and uncertainty preferably here. The ε_m model, error entropy model and some other models are also discussed as a further development of it. ε_σ error band model for line segment can depict the error distribution of a line segment composed of two end points felicitously, and the characteristic with the smaller error at middle while the larger at ends. The achieved characteristic can reflect and depict the influence of the line segment error on the polygon segment. Furthermore, by analyzing the probability of a line segment error falling into a certain range, the probabilistic statistical characteristic is gained, which is helpful to determine the error indicator and error buffer for a line segment based on rigorous theory in GIS. Finally, a certain relationship has been found between the standard error of polygon area and error band area of the polygon segment by dedu-

Table 1 Computational results of different stand error of polygon area and the area of polygon error loop

polygon	RMS of area σ_{s_1}	Area of error loop Δ_s	Influence of error band σ_{s_2}	Area of polygon
Regular quadrangle	141.42	586.6	142.15	10 000
Regular pentagon	212.66	979.35	218.99	23 777
Regular hexagon	212.13	982.30	200.51	25 981
Regular heptagon	206.85	987.55	186.63	27 365
Regular octagon	200.00	995.77	176.03	28 285
Regular enneagon	192.84	1 000.70	166.78	28 926
Regular decagon	185.87	1 007.28	159.27	29 389
Rectangle	223.60	975.10	243.78	20 000
Echelon	187.08	814.01	203.50	15 000
Pentagon	183.71	849.88	190.04	15 000
Heptagon	264.57	1462.77	276.44	35 000
Octagon	129.90	811.69	143.49	8 750
Decagon	244.95	1624.62	256.88	32 500

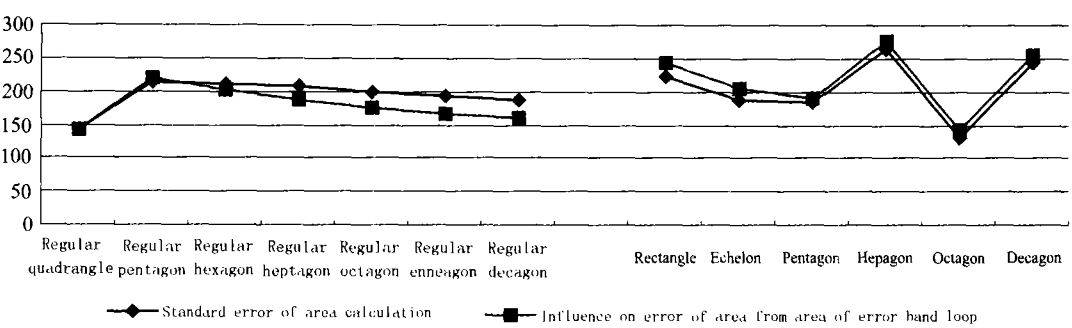

Fig. 5 Comparison of error of mean squares of polygon area with the area of the error loop

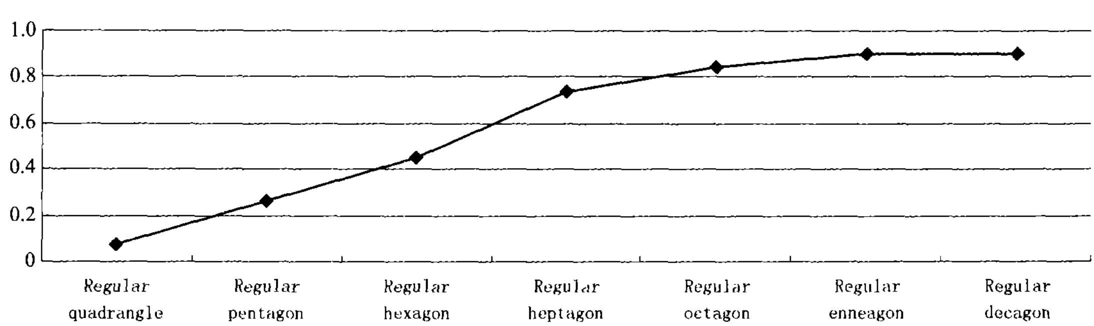

Fig. 6 Difference of the two accuracies of the areas calculated relative to the polygon area

cing and numerical calculation of the influence from error band loop of a polygon on area accuracy. The error of the area of a polygon is inflicted by the error of the line segments which form the polygon. Only an accuracy value is given according to a common calculation of the standard error of polygon area. However, error distribution of polygon area error and its probabilistic statistical characteristic can be achieved that if an error analysis on each line segment of the polygon is conducted. The sum of each line segment ε_σ error band area can be regarded as an indicator for the area error, while it can't be regarded as a standard error of the area due to the great difference between them in some instances.

REFERENCES

1 Yu Z C, Lu L C (1983) Fundamentals of surveying adjustment. Beijing: Press of Surveying and Mapping. (in Chinese)

2 Liu D J, Shi W Z, Tong X H (1999) Accuracy analysis and quality control for spatial data in GIS. Shanghai: Shanghai Press of Science and Technology. (in Chinese)

3 Perkal J (1956) On epsilon length. *Bulletin de l' Academic Polonaise Des Sciences*, (4): 399-403

4 Chrisman N R (1982) A theory of cartographic error and its measurement in digital database. *Auto-Carto*, (5): 159-168

5 Scheruing G W (1993) A position accuracy in spatial database. *Environment and Urban System*, (17): 103-110

6 Dutton G (1992) Handling Positional Uncertainty in Spatial Database. The 5th International Symposium on Spatial Data Handling, Charleston, South Carolina.

7 Shi W Z (1994) Model positional and thematic uncertainties in integration of remote sensing and GIS. Enschede: ITC Publicaiton.

8 Shi W Z (1998) Theory and methods for handing errors in spatial data. Beijing: Science Publication. (in Chinese)

9 Fan A M, Guo D Z (2001) The uncertainty band model of error entropy. *Acta Geodatica et Catrographica Sinica*, 30(1): 48-53