

## A SIMULATION APPROACH TO ANALYSE ERROR IN BUFFER SPATIAL ANALYSIS

**Chui Kwan CHEUNG, Wenzhong SHI**  
Department of Land Surveying & Geo-Informatics  
The Hong Kong Polytechnic University, China  
[tracy.cheung@hongkong.com](mailto:tracy.cheung@hongkong.com); [ls wzshi@polyu.edu.hk](mailto:ls wzshi@polyu.edu.hk)

**KEY WORDS:** Error Propagation, GIS, Modelling, Simulation, Uncertainty, Vector.

### ABSTRACT

This paper proposes a simulation model for analysing error propagation in buffer spatial analysis in a vector-based geographical information system. Since spatial data stored in GIS is not error-free, errors propagate and accumulate over GIS operations. Therefore, there is necessary to study error propagation in the process. The proposed model in this study measures the propagated error in buffer spatial analysis. The errors covered in this study include error of commission and error of omission of a buffer analysis. These errors are due to the difference between the measured location and the expected location of a buffer. In order to study the error propagation, a simulation approach is proposed and implemented to estimate these errors.

### 1 INTRODUCTION

In a geographical information system (GIS), we normally store selected and processed geographical features of the real world in a discrete manner. The difference of a map representation of geographical features to the reality cannot be prevented. It is partially dependent on a map scale. In the past, people always ignore this difference but serious problems will be led if they analyse geographical data and make decisions with GIS (Goodchild, 1991). The American Congress of Surveying and Mapping (ASCM) provided standards on spatial data quality in 1982. Other organizations also defined standards on the spatial data quality with different quality parameters. Aalders (1999) summarized these standards in details. Generally, errors are most likely to be classified into five categories: lineage error, positional error, attribute error, logical inconsistency, and completeness. Moreover, the errors of spatial data may be transferred to the newly generated objects via GIS operations and so the derived spatial data accumulates more errors. This paper intends to investigate the positional error propagation through a GIS buffer operation.

In GIS, there exist different operations and analysis methods to analysis data and retrieve information you require. It is impossible to create one mathematical expression to model propagated errors for all operations, although we know that the propagated errors are functions of the source errors. Now, existing research form error propagation models for each operation only. Some researches have been conducted on the error propagation in the overlay operation (Brinkmann and Hinrichs, 1998; Chrisman, 1987), while some considered the error propagation in vector to raster conversion (Carver and Brunson, 1994; Veregin, 1989). For another important GIS operation buffering, Veregin (1994, 1996) simulated its error propagation in a field-based GIS but only few attempts such as the absolute and the relative accuracy of a buffer operation proposed by Zhang et al (1998) have so far been made at modelling the error propagation for the buffer operation in an object-based GIS.

In the previous study on the error propagation for buffer operation, an absolute accuracy of the buffer operation was considered. It was concluded that the absolute accuracy of the buffer operation was inverse proportional to the width of the buffer for the spatial feature. The weakness of this model for the spatial feature is that in the first step, we have to determine which error-band model should be used for describing error of the source spatial feature but error-band models include epsilon band (Chrisman, 1982), E-band (Caspary and Scheuring, 1992), S-band (Shi and Ehlers, 1993) and so forth. Based on the error-band model you prefer, the error propagation model for the spatial feature can be obtained. Up to now, no standard methods can be implemented for determining the most feasible error-band model for the specified feature.

This paper aims at developing an error propagation model for the buffer spatial analysis in an object-based GIS by simulation. This idea of the model is same as the analytical model on the error propagation for the buffer spatial analysis proposed by Shi and Cheung (2000a). The analytical model does not involve the error-band model. It measures the propagated error with the discrepant area of the buffer. An expected location and a measured location of the buffer

zone bound the discrepancy. In order to sum up most possible locations of the measured location of the buffer zone, the discrepant area was expressed in terms of integrals. Since no exact solution of this integral, a numerical integration was implemented for calculating its approximation. It implied that knowledge of numerical analysis must be required in the analytical model. In this paper, we will propose another method to model the error propagation for buffer spatial analysis – a simulation one, which is easier to understand.

## 2 SOURCE ERROR IN BUFFER SPATIAL ANALYSIS

Buffer operation is a basic GIS function in which a buffer of some specified width is delineated around a spatial feature. Figure 1 illustrates buffers around spatial features: point, linear and areal features. When a desired buffer distance is given, GIS builds the buffer outward from the selected features. Buffering allows potential users to retrieve features that lie within the desired distance of the features.

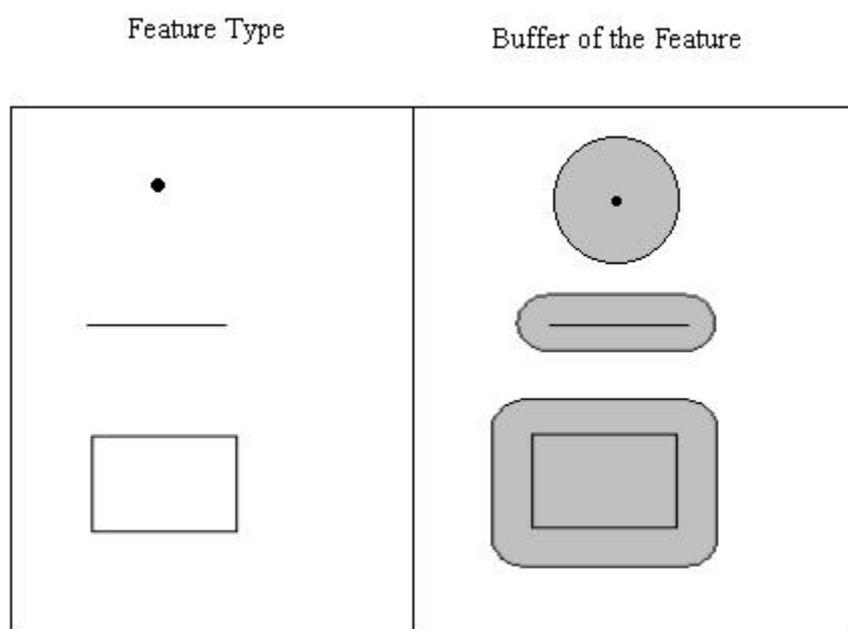


Figure 1. Buffer around features.

Since a spatial feature in GIS is not error-free, errors in this feature will be transferred to the newly generated feature during a buffering process. As a result, the derived spatial feature may accumulate more errors and have different error characteristics from the source feature. And errors in this derived spatial feature should be depending on errors in source spatial feature (source errors). Let  $S\_Error(x)$  denote a function used for the source error measurement of a spatial feature  $x$ ,  $B\_Error$  denote an input error function for buffer size, and  $g$  denote an error propagation function in buffer spatial analysis. Then, the propagated error  $P\_Error(x)$  can be expressed in term of the error propagation function and the source error as Equation (1).

$$P\_Error(x) = g(S\_Error(x), B\_Error) \quad (1)$$

This source error can be caused by age of data, map scale, density and pattern of observation and storage data format (Shi, 1994). Also, positional error and attribute error exist. During buffer operation, some errors may introduce because we may input an imprecise buffer size. In this study, we concentrate on positional error of a spatial feature as the source error in buffer spatial analysis under an assumption that there are no input errors for buffer size.

Models on positional errors of a spatial feature were proposed (Ali, 1998; Chapman et al, 1997; Dutton, 1992; Easa, 1995; Shi, 1998). Some of them developed their models based on the assumption of error distribution of a point (or

node) on a spatial feature. In this study, this assumption is elementary, mainly due to the fact that each spatial feature is represented by nodes in a GIS database.

### 3 ASSUMPTION OF NODAL ERROR DISTRIBUTION

In this study, it is assumed that nodal errors are bivariate normally distributed. Mean is referred to be the expected location of a node. How can the covariance matrix in the bivariate normal distribution be determined? In GIS, coordinates of nodes are usually captured from existing map using digitiser or scanner. Let  $(\hat{x}, \hat{y})$ ,  $(\bar{x}, \bar{y})$ ,  $(\Delta s_x, \Delta s_y)$ , and  $(\Delta m_x, \Delta m_y)$ ,  $(\Delta d_x, \Delta d_y)$  denote the expected location of a point, its measured location, errors due to surveying, errors due to map, and errors due to digitising. Then in a 1:M scale map, the expected location of a point is listed in Equation (2).

$$\begin{cases} \hat{x} = \bar{x} + \Delta s_x + M \Delta m_x + M \Delta d_x \\ \hat{y} = \bar{y} + \Delta s_y + M \Delta m_y + M \Delta d_y \end{cases} \quad (2)$$

Casparly and Scheuring (1993) used standard derivation to describe the total error of a spatial feature in GIS. Its variances in the x-direction  $\sigma_{T_x}^2$  and in the y-direction  $\sigma_{T_y}^2$  should be

$$\begin{cases} \sigma_{T_x}^2 = \sigma_{\text{surveying in the x-direction}}^2 + M^2 \sigma_{\text{map in the x-direction}}^2 + M^2 \sigma_{\text{digitizing in the x-direction}}^2 \\ \sigma_{T_y}^2 = \sigma_{\text{surveying in the y-direction}}^2 + M^2 \sigma_{\text{map in the y-direction}}^2 + M^2 \sigma_{\text{digitizing in the y-direction}}^2 \end{cases} \quad (3)$$

Since there is a correlation within a node on a spatial feature, their correlation coefficient should be estimated but we are not going to derive it in this study. The error distribution of a node can be defined given the mean and the variance of the distribution. However, this distribution is bell shaped and tail off. We will construct an error ellipse for each node to limit the possible positions of nodes inside the error ellipse with a predefined confident coefficient for simplification.

For  $j = 1$  or  $2$ , let  $(x_j, y_j)$  denote a node of a spatial feature;  $\rho_{x_j y_j}$  denote a correlation coefficient of  $x_j$ 's error and  $y_j$ 's error; and  $\sigma_{x_j}$ ,  $\sigma_{y_j}$  denote sample derivations of  $x_j$ 's error and  $y_j$ 's error respectively. Then, the mathematical expression of an error ellipse for the node is given by

$$\left( \frac{x_j - \mu_{x_j}}{\sigma_{x_j} / \sqrt{-2 \log(\alpha)}} \right)^2 + \left( \frac{y_j - \mu_{y_j}}{\sigma_{y_j} / \sqrt{-2 \log(\alpha)}} \right)^2 - 2\rho_{x_j y_j} \left( \frac{x_j - \mu_{x_j}}{\sigma_{x_j}} \right) \left( \frac{y_j - \mu_{y_j}}{\sigma_{y_j}} \right) = 1, \text{ for } i = 1, 2 \quad (4)$$

where  $\alpha$  is the confidence coefficient.

### 4 ERROR PROPAGATION MODEL

Error propagation model for a buffer around a point feature and a linear feature using simulation has been proposed by Shi and Cheung (2000b). Thus, this study will focus on a buffer around an areal feature. Also in the previous part, the relationship between errors of a spatial feature and propagated errors in buffer spatial analysis has been shown (see Equation (1)). This study assumes that potential users will key-in the value of buffer size and so Equation (1) should be further modified to Equation (5).

$$P\_Error(x) = g(S\_Error(x)) \quad (5)$$

Function  $g$  in Equation (5) will be studied in the following. First, we will define how propagated error in buffer spatial analysis is modelled.

A discrepancy of a buffer around an areal feature is bounded by the measured location and the expected location of the buffer, and their tangents. Figure 2 shows an example of an areal feature with four nodes. For  $j = 1, 2, 3$  or  $4$ ,  $(\mu_{x_j}, \mu_{y_j})$  and  $(x_j, y_j)$  represent the expected node and the measured node of the areal feature respectively. The boundary of the expected areal feature is the thick solid polyline and that of the measured areal feature is the thick dotted polyline. The region around the solid polyline is the buffer around the expected areal feature while the region around the dotted polyline is the buffer around the measured areal feature. The shaded region in light grey is due to the error of the commission while that in dark grey is due to the error of omission.

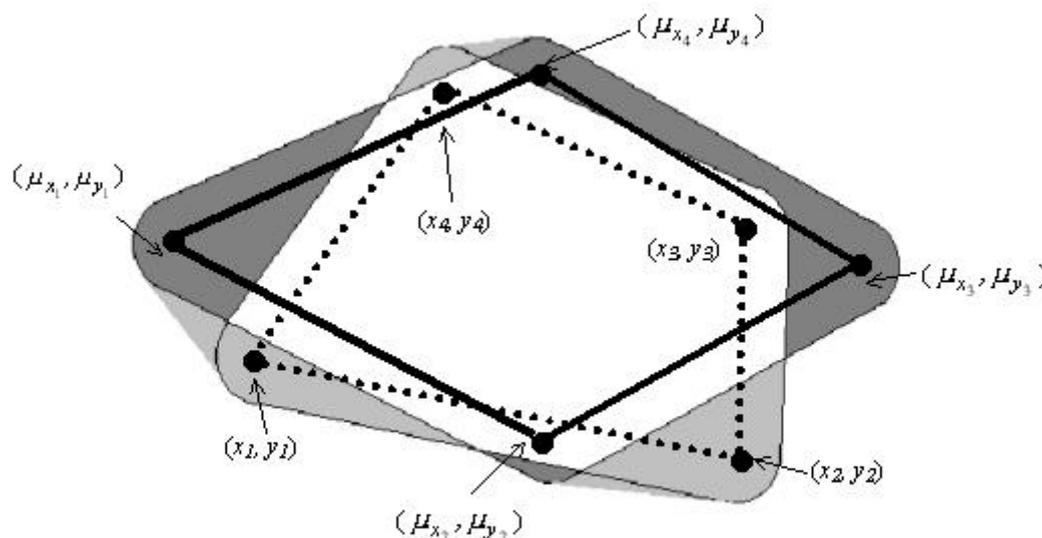


Figure 2. The error of commission is shaded in light grey while the error of omission is shaded in dark grey in the buffer spatial analysis for an areal feature with four nodes.

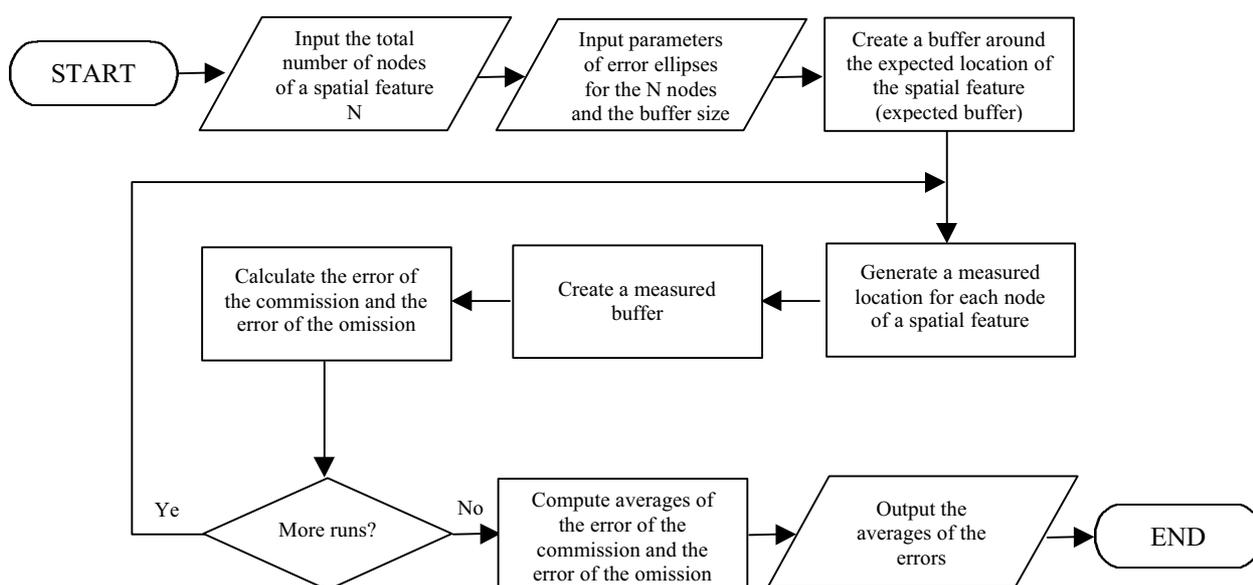


Figure 3. Steps in the simulation study are charted.

In order to estimate the error of the commission and the error of the omission in the buffer spatial analysis, simulation is implemented based on our assumption of the nodal error distribution. Under our assumption, thousands of possible measured locations for each node on a spatial feature will be generated and so there are thousands of measured areal features. For each measured areal feature, their corresponding buffers are created after the buffer size is inputted. These measured buffers will then be compared with the expected buffer. After the comparison, averages of the error of the commission and of the error of the omission can be determined. A flowchart in Figure 3 illustrates the overall concept of this simulation model.

## 5 CASE STUDIES AND DISCUSSIONS

In the following examples, meter (m) is used as length unit. It is considered that an areal feature with three nodes: (0, 0), (1000, 1000) and (2000, -2000). It is supposed that  $\sigma_{x_j}$  and  $\sigma_{y_j}$  (parameters in Equation (4)) are 7.75m for  $j=1, 2$  or  $3$ ; the confidence coefficient is 0.5; and there are no correlations between and within nodes of the areal feature. The average error of commission is 528246.7 m<sup>2</sup> and the average error of omission is 550852.8 m<sup>2</sup> when the buffer size is 500m.

Next, the effect of the buffer size is studied. When the buffer size changes from 500m to 2500m, the errors of the commission and of the omission are demonstrated in Figure 4. The solid line represents a plot of the error of the omission against the buffer size while the dotted line represents a plot of the error of the commission against the buffer size.

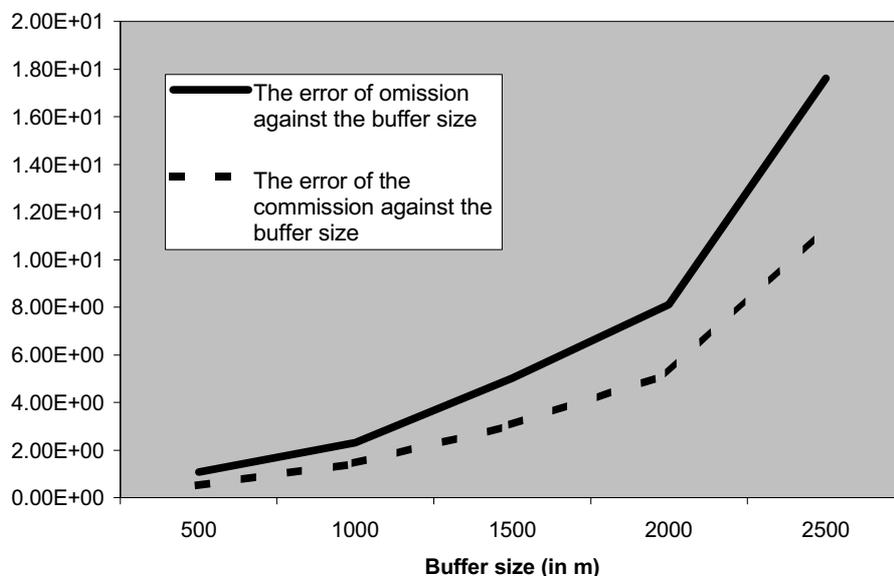


Figure 4. The error of the commission and the error of the omission against the buffer size are plotted.

First, it is noticed that in both the solid line and the dotted line, there is an increasing trend on the error of the omission and the error of the commission respectively when the buffer size increases. Moreover, their slopes are changing steadily. We can divide each of them into three parts. When the buffer size is in the range of 500m to 1000m, the slope in the error of the commission case is  $1.85 \times 10^3 \text{m}$  and that in the error of the omission case is  $6.04 \times 10^2 \text{m}$ . In the case of the buffer size ranging from 1000m to 2000m, the slopes of the solid line and the dotted line are  $3.78 \times 10^3 \text{m}$  and  $2.04 \times 10^3 \text{m}$  respectively; while in the case of the buffer size ranging from 2000m to 2500m, these slopes are  $1.23 \times 10^4 \text{m}$  and  $6.65 \times 10^3 \text{m}$ . Hence, it implies that these errors are increasing steadily over a certain range of the buffer size.

## 6 SUMMARY AND DISCUSSIONS

This study proposed an approach to model error propagation of a buffer spatial analysis in vector-based GIS. As a further development of previous research on error models of GIS features by simulation technique, this study focuses on error propagation in GIS spatial analysis by the simulation technique. In this study, we found that the buffer size affected the error propagation to a certain degree in a buffer spatial analysis.

In this study, the error of the commission and the error of the omission of a buffer spatial analysis are used to measure the error propagation based on the assumption of nodal error distribution. These errors vary for different scale of maps, due to the fact that these errors are computed based on the assumption of nodal error distribution and the variance are affected by scale.

The proposed model can be implemented to simulate error propagation over buffer operations in a GIS. This model provides a measure tool to identify the errors of the commission and the omission in buffer spatial analysis. This measurement can further be used to analyse reliability of spatial features and analysis in GIS.

## ACKNOWLEDGMENTS

The work described in this paper was substantially supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region (Project No. G-V642 and G-S407).

## REFERENCES

- Aalders H.J.G.L., 1999. The Registration of Quality in a GIS, Proceedings of the International Symposium on Spatial Data Quality'99, p. 23-32.
- Ali A.A., 1998. Modelling and Managing Uncertainty in Object-Based Geospatial Information System, Ph.D. Dissertation, University of Calgary, Alberta, Canada.
- Brinkmann A., Hinrichs K., 1998. Implementing Exact Line Segment Intersection in Map Overlay, Proceedings of 8<sup>th</sup> International Symposium on Spatial Data Handling, p. 569-579.
- Carver S.J., Brunson C.F., 1994. Vector to Raster Conversion Error and Feature Complexity: an Empirical Study using Simulated Data, IJGIS, Vol. 8, No. 3, p. 261-270.
- Caspary W., Scheuring R., 1992. Error-band as Measures of Geographic Accuracy, Proceedings of EGIS'92, p. 226-233.
- Caspary W., Scheuring R., 1993. Positional Accuracy in Spatial Databases, Computers Environment and Urban Systems, Vol. 17, p. 103-110.
- Chapman M.A., Alesheikh A., Karimi, H., 1997. Error Modelling and Management for Data in Geospatial Information Systems, Proceedings of Coast GIS'97, p. 1-15.
- Chrisman N.R., 1992. A Theory of Cartographic Error and its Measurement Digital Databases, Proceedings of Auto-Carto 5, p. 159-168.
- Chrisman N.R., 1987. The Accuracy of Map Overlays: a Reassessment, Landscape and Urban Planning, Vol. 14, p. 427-439.
- Dutton G., 1992. Handling Positional Uncertainty in Spatial Databases, Proceedings of the 5<sup>th</sup> International Symposium on Spatial Data Handling, p. 460-469.
- Easa, S.M., 1995. Estimating Line Segment Reliability using Monte Carlo Simulation, Surveying and Land Information Systems, Vol. 55, No. 3, p. 136-141.
- Goodchild M.F., 1991. Keynote Address: Symposium on Spatial Database Accuracy, Proceedings of Symposium on Spatial Database Accuracy, p. 1-16.

Goodchild M.F., 1999. Measurement-based GIS, Proceedings of the International Symposium on Spatial Data Quality'99, p. 1-9.

Shi W.Z., 1994. Modelling Positional and Thematic Uncertainty in Integration of GIS and Remote Sensing, Ph.D. Dissertation, University of Osnabruck-Vechta, Enschede, NL: ITC publication 22.

Shi, W.Z., 1998. A Generic Statistical Approach for Modelling Error of Geometric Features in GIS, International Journal of Geographical Information Science, Vol. 12, No. 2, p. 131-43.

Shi W.Z., Cheung C.K., 2000a. Measure error propagation for buffer spatial analysis in GIS using Gaussian Quadrature numerical method, Proceedings of Accuracy2000 (accepted).

Shi W.Z., Cheung C.K., 2000b. Error Propagation in Buffer Spatial Analysis using Simulation, 9th International Symposium on Spatial Data Handling (submitted).

Shi W.Z., Ehlers M., 1993. "S-band", a Model to Describe Uncertainty of an Object in an Integrated GIS/Remote Sensing Environment, Proceedings of IGARSS'93.

Veregin, H., 1989. A Review of Error Models for Vector to Raster Conversion, Operational Geographer, Vol. 71, No. 1, p. 11-15.

Veregin H., 1994. Integration of Simulation Modelling and Error Propagation for the Buffer Operation in GIS, PE&RS, Vol. 60, No. 4, p. 427-435.

Veregin H., 1996. Error Propagation through the Buffer Operation for Probability Surfaces, PE&RS, Vol. 62, No. 4, p. 419-428.

Zhang B., Zhu L., Zhu G., 1998. The Uncertainty Propagation Model of Vector Data on "Buffer" Operation in GIS, ACTA Geodaetica et Cartographica Sinica, Vol. 27, No. 3, p. 259-266 (in Chinese).