MODEL OF ERROR PROPAGATING IN TOPOLOGY CLEAN OPERATION

Min DENG[1, 2], Wenbao LIU [3], Xiaoyong CHEN[1]

[1]Geoinformatics, School of Advanced Technology, Asian Institute of Technology PO BO X 4, Klong Luang, Pathumthani 12120, THAILAND
(st028028, xychen)@ait.ac.th
[2]Department of Geomatics, Institute of Remote Sensing and Information Engineering, Wuhan University, 229 Road Luoyu, Wuchang, Wuhan, 430079, P.R.China
[3]Department of Earth & Geographic Sciences, University of Massachusetts at Boston, MA 02125,USA
Wenbao_liu@yahoo.com

Commission II, WG Ⅱ/6

KEY WORDS: Data Cleaning, Topological Construction, Node Snapping, ARC Snapping, General Snapping

ABSTRACT:

Data quality in GIS has been widely recognized as one of the most critical issues in GIS applications in recent years. There are numerous effective error models for managing error propagation in the process of data manipulation and analysis. But there is still a lack of error analysis and its propagation in topological operations. In practical applications, data sources, which have different accuracy of spatial data, will be a necessity of causing inconsistency of topological relations. So this paper analyses topological inconsistency of inaccurate data, and compares their differences of different processing methods. Then, an error propagation model is built, as is easy for users to analyze and understand the effect of error in GIS operations. Finally, a detailed example illustrates its application of the model built in this paper.

1. INTRODUCTION

Spatial databases are one of the most expensive parts of GIS, while data quality in GIS has been widely recognized as one of the most critical issues in GIS applications in recent years (Veregin, 1999). There are numerous effective error models for managing error propagation in the process of data manipulation and analysis such as Boolean operations (Chrisman, 1987; Shi and Liu, 2000). However, there is a lack of models for dealing with errors in topology clean operation which is an indispensable process in order to ensure the correctness of topological relations between spatial features, further, it will have a serious effect on spatial analysis, spatial query, computation and graphical display.

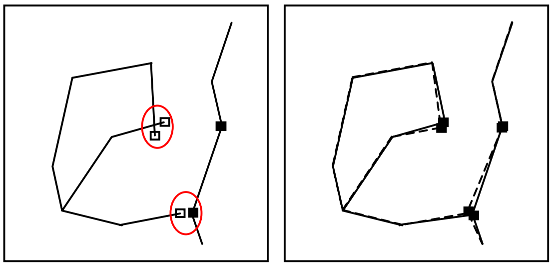
In most GIS software, errors in topological relations are corrected by means of data editing tool such as 'snap' functionality. Snap, which is one of the basic operations in most GIS software Packages such as ARCEDIT in ARC/INFO, is usually used to handle this kind of error, and is classified as Node snapping, Arc snapping and General snapping (Jones, 1997). However, above snapping operations involve changing the position of points of a spatial feature so that points within a fuzzy tolerance of each other are connected automatically. There are some basic commands for adjustment of inconsistent points, such as 'MOVE' and 'ADJUST'. At Present, there is a lack of methods of tracking uncertainty for these operations and some metric indicators. In addition, although this involves an adjustment of feature coordinates, in which the coordinates of the snapped node are simply the means of coordinates of two or more points, there are no quantitative techniques for assessing the spatial accuracy of the final node created by automatic snapping. Subsequently, adjustment of coordinates of point, including node and vertex, will directly lead to changes of Arc/line and polygon. Hence, it is very difficult to obtain their accuracy information without that of point coordinates. This paper aims to address these issues systematically.

In this paper, we firstly look at various cases for topological inconsistency between point and point, point and line, point and polygon, Arc/line and Arc/line, Arc/line and polygon, polygon and polygon, summarizing some existing methods in section 2. Section 3 presents a generalized algorithm of node snapping using the least square method and develop a universal model for handling error propagation related to node snapping. Simplified algorithms and models are derived for some special cases with different statistical characteristics of point errors within a fuzzy tolerance. And the expression formula of accuracy estimation of line and polygon and its spatial attribute such as length of line and area of polygon are also given. In section 4, a detailed example is provided to demonstrate the potential applications of the generalized algorithm and the related error model, and some comparatives under different handling operations. Finally, this paper ends with some conclusions in section 5.

2. TOPOLOGICAL CLEANING

2.1 Processing Inconsistency for Points

In vector GIS, point is a basic unit of representing graphic features. For example, a line segment consists of its start- and end points, a polygon consisting of a series of ordered points. So we may regard spatial database in GIS as point field, and accuracy of spatial database will be measured by positional error of points. Topologically, uncertainty will possibly cause a conflict among spatial features from different data layers, as is more difficult to make further spatial operations, queries etc, and some unexpected results will possibly appear. In general, topological cleaning operations are necessary to make spatial data useful. But such operations involve changes of spatial position, and have different results, as are illustrated in figure 1. In addition, in figure (b), solid line denotes the result from the command 'MOVE'; vanish line the result from the command 'ADJUST'.



(a) Case before node snapping; (b) Case after node snapping

Figure 1. Comparison of different results from node snapping

2.2 Processing Inconsistency for Lines

Topologically, inconsistency of line features is equivalent to the damage of 'equal' relations, which may be summarized as three kinds, overlay, meet and disjoint, see figure 2(a), (b) and (c). Apparently, this inconsistency will further make topological relations between polygon objects incorrect. In addition, another case is to the damage of connection. That is to say, L_1 and L_2 are two adjacent line features in practice, but they are not adjacent since the effect of error and uncertainty such as digitalization error, illustrated in figure 2(d).

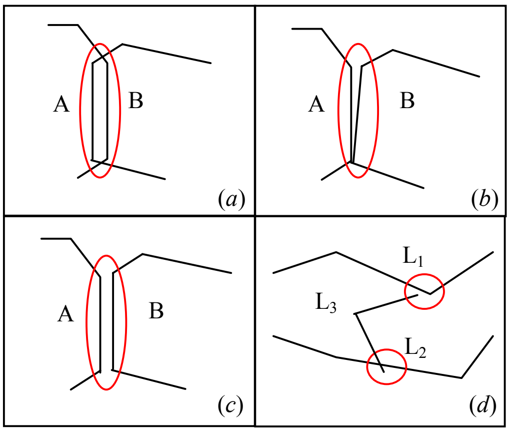


Figure 2. Inconsistent lines by integrating input data from different sources

For above inconsistency of lines, there are two existing methods; one is to replace all inconsistent lines within fuzzy tolerances with a line having the higher accuracy, another to make a simple adjustment for these inconsistent lines, generating a new line replacing them. But the results by these two methods are distinct, as are shown in figure 3.

2.3 Processing Inconsistency for Polygons

Many silvers will generate with inaccurate data when spatial overlay operation is made, which increase greatly the storage volume of computer, leading to the difficulty of spatial analysis and queries. Early in 1970's, many researchers focused on how to process large amounts of silvers, which are also named as meaningless polygon in some literatures (MacDougall, 1975; Chrisman, 1987; 1989; Harvey, 1994). Two major processing approaches are presented in these researches, that is, preventing-focus approach and correcting-focus approach. So-called preventing-focus approach is to snap all points that their distances are within the range of fuzzy tolerances as one point, therefore, inconsistent topological relationships such as Node and Node, Node and Arc, Node and Polygon, Arc and Arc, Arc and Polygon, and Polygon and Polygon do not happen. While correcting-focus approach is to process the existed silvers after overlay operation. The most common used methods is deleting some inconsistent boundary lines, merging all silvers to their neighbor polygons. Although this process causes some attribute match errors, data volumes will be decreased greatly. In most of current commercial GIS softwares, such as ARC/INFO, include some tools used for process of inconsistent data.

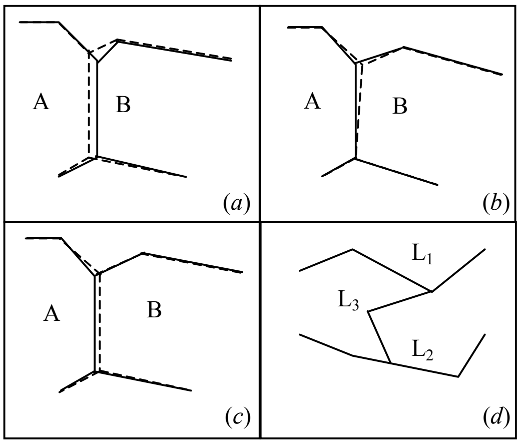


Figure 3. Comparatives of results of inconsistency processing

3. MODEL OF ERROR PROPAGATING

3.1 Generalized Algorithm and Error Model of Node Snapping

Let $z_i = (x_i, y_i)^{\mathrm{T}}, (i = 1,2,\cdots,n)$ be the i -th node, from a group of points within a specified fuzzy tolerance, with the covariance matrix of $\boldsymbol{\Gamma}_{ii} = \begin{bmatrix} \sigma_{x_i}^2 & \sigma_{x_i y_i} \\ \sigma_{y_i x_i} & \sigma_{y_i}^2 \end{bmatrix}$. Furthermore, if there exists $\sigma_{x_i y_i} = \sigma_{y_i x_i} \neq 0, \quad (i = 1,2,\cdots,n)$, then z_i is an auto-correlation vector.

Let $z^* = (x_1, y_1, x_2, y_2, \cdots, x_n, y_n)^{\mathrm{T}}$, then the covariance matrix of the n -point snapping group is as follows

$$\boldsymbol{\Gamma}_{z^*z^*}=\begin{bmatrix}\sigma_{x_1}^2 & \sigma_{x_1y_1} & \sigma_{x_1x_2} & \sigma_{x_1y_2} & \cdots & \sigma_{x_1x_n} & \sigma_{x_1y_n}\\ \sigma_{y_1x_1} & \sigma_{y_1}^2 & \sigma_{y_1x_2} & \sigma_{y_1y_2} & \cdots & \sigma_{y_1x_n} & \sigma_{y_1y_n}\\ \sigma_{x_2x_1} & \sigma_{x_2y_1} & \sigma_{x_2}^2 & \sigma_{x_2y_2} & \cdots & \sigma_{x_2x_n} & \sigma_{x_2y_n}\\ \sigma_{y_2x_1} & \sigma_{y_2y_1} & \sigma_{y_2x_2} & \sigma_{y_2}^2 & \cdots & \sigma_{y_2x_n} & \sigma_{y_2y_n}\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ \sigma_{x_nx_1} & \sigma_{x_ny_1} & \sigma_{x_nx_2} & \sigma_{x_ny_2} & \cdots & \sigma_{x_n}^2 & \sigma_{x_ny_n}\\ \sigma_{y_nx_1} & \sigma_{y_ny_1} & \sigma_{y_nx_2} & \sigma_{y_ny_2} & \cdots & \sigma_{y_nx_n} & \sigma_{y_n}^2\end{bmatrix}$$

$$=\begin{bmatrix}\boldsymbol{\Gamma}_{11} & \boldsymbol{\Gamma}_{12} & \cdots & \boldsymbol{\Gamma}_{1n}\\ \boldsymbol{\Gamma}_{21} & \boldsymbol{\Gamma}_{22} & \cdots & \boldsymbol{\Gamma}_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ \boldsymbol{\Gamma}_{n1} & \boldsymbol{\Gamma}_{n2} & \cdots & \boldsymbol{\Gamma}_{nn}\end{bmatrix} \tag{1}$$

where $\boldsymbol{\Gamma}_{ij}=\begin{bmatrix}\sigma_{x_ix_j} & \sigma_{x_iy_j}\\ \sigma_{y_ix_j} & \sigma_{y_iy_j}\end{bmatrix}$. If $\boldsymbol{\Gamma}_{ij}\neq 0 \quad (i\neq j)$, $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ $(i\neq j)$ are cross-correlation vectors. If for all or some of $i,j=1,2,\cdots,n$ $(i\neq j)$, we have $\boldsymbol{\Gamma}_{ij}\neq 0\,(i\neq j)$, then $\boldsymbol{z}^*$ is a partial or full cross-correlation vector. Let $\boldsymbol{\Psi}$ be the inverse matrix of $\boldsymbol{\Gamma}_{z^*z^*}$, i.e., $\boldsymbol{\Gamma}_{z^*z^*}^{-1}=\boldsymbol{\Psi}_{2n\times 2n}$, $\boldsymbol{\Phi}_{ij}$ $(i,j=1,2,\cdots,n)$ be a 2×2 partitioned matrices taken from a $2n\times 2n$ square matrix $\boldsymbol{\Psi}_{2n\times 2n}$, related to the same order of $\boldsymbol{\Gamma}_{ij}$ in the inverse matrix $\boldsymbol{\Gamma}_{z^*z^*}$, and $\boldsymbol{z}=(x,\,y)^{\mathrm{T}}$ be the optimally estimated vector, of the coordinates of the new point created by snapping, with covariance matrix of $\boldsymbol{\Gamma}=\begin{bmatrix}\sigma_x^2 & \sigma_{xy}\\ \sigma_{yx} & \sigma_y^2\end{bmatrix}$.

Because the algorithm of node snapping of a group of points within fuzzy tolerance is essentially an adjustment algorithm of direct observations on the basis of the least square principle in geodesy (Mikhail, 1976), we can derive the following generalized formula for estimating the coordinates of the snapped point

$$\boldsymbol{z}=(\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\Phi}_{ij})^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{\Phi}_{ij}\boldsymbol{z}_j) \tag{2}$$

At the same time, we have the universal error propagation model of coordinates through the snapping operation from $\boldsymbol{z}^*$ to $\boldsymbol{z}$

$$\boldsymbol{\Gamma}=(\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\Phi}_{ij})^{-1} \tag{3}$$

Apparently, the generalized algorithm of node snapping (2) and the universal error propagation model (3) can be used to multiple data sources with varying resolutions.

3.2 Error Propagating of Line Feature with Snapping Processing

3.2.1 Positional Error Estimation of Any Point of Line

As we know, the snapped line consisted of a set of points. Let $L^*=(x_1^*,y_1^*,x_2^*,y_2^*,\cdots,x_n^*,y_n^*)^T$, and let $z_i^*=(x_i^*,y_i^*)^T$ be the i-*th* snapped point, its coordinates can be computed by equation (1). Mathematically, one line segment is determined by its start- and end points, one poly-line by a series of line segments, i.e., an ordered set of points. So we may compute positional error of its any point as a partial measure, and total error band consisted of positional error ellipse of all points. Let any point in line feature L^* be $z_{it}^*(x_{it}^*,y_{it}^*)$, which is located in the segment $z_i^*z_{i+1}^*$, it can be expressed by [7]

$$\begin{cases}x_{it}^*=(1-t)x_i^*+tx_{i+1}^*\\ y_{it}^*=(1-t)y_i^*+ty_{i+1}^*\end{cases} \tag{4}$$

Furthermore, equation (3) can be rewritten as

$$z_{it}^*=\begin{pmatrix}x_{it}^*\\ y_{it}^*\end{pmatrix}=K_{2\times 2n}L^* \tag{5}$$

where $K_{2\times 2n}=\begin{pmatrix}0,0,\cdots,1-t,0,t,0,\cdots,0,0\\ 0,0,\cdots,0,1-t,0,t,\cdots,0,0\end{pmatrix}$.

According to the propagation law of error, we will get the following equation,

$$\Gamma_{z_{it}^*z_{it}^*}=K\Gamma_{L^*L^*}K^T \tag{6}$$

where $\Gamma_{L^*L^*}$ is a covariance matrix, and having

$$\Gamma_{L^*L^*} = \begin{bmatrix} \sigma^2_{x_1^*} & \sigma_{x_1^*y_1^*} & \sigma_{x_1^*x_2^*} & \sigma_{x_1^*y_2^*} & \cdots\sigma_{x_1^*x_n^*} & \sigma_{x_1^*y_n^*} \\ \sigma_{y_1^*x_1^*} & \sigma^2_{y_1^*} & \sigma_{y_1^*x_2^*} & \sigma_{y_1^*y_2^*} & \cdots\sigma_{y_1^*x_n^*} & \sigma_{y_1^*y_n^*} \\ \sigma_{x_2^*x_1^*} & \sigma_{x_2^*y_1^*} & \sigma^2_{x_2^*} & \sigma_{x_2^*y_2^*} & \cdots\sigma_{x_2^*x_n^*} & \sigma_{x_2^*y_n^*} \\ \sigma_{y_2^*x_1^*} & \sigma_{y_2^*y_1^*} & \sigma_{y_2^*x_2^*} & \sigma^2_{y_2^*} & \cdots\sigma_{y_1^*x_n^*} & \sigma_{y_1^*y_n^*} \\ \vdots & & & & & \vdots \\ \sigma_{x_n^*x_1^*} & \sigma_{x_n^*y_1^*} & \sigma_{x_n^*x_2^*} & \sigma_{x_n^*y_2^*} & \cdots\sigma^2_{x_n^*} & \sigma_{x_2^*y_n^*} \\ \sigma_{y_n^*x_1^*} & \sigma_{y_n^*y_1^*} & \sigma_{y_n^*x_2^*} & \sigma_{y_n^*y_2^*} & \cdots\sigma_{y_n^*x_n^*} & \sigma^2_{y_n^*} \end{bmatrix}$$

So equation (6) will be developed as

$$\Gamma_{z_{it}^*z_{it}^*} = \begin{bmatrix} \sigma^2_{x_{it}^*} & \sigma_{x_{it}^*y_{it}^*} \\ \sigma_{y_{it}^*x_{it}^*} & \sigma^2_{y_{it}^*} \end{bmatrix}$$

Where,

$$\sigma^2_{x_{it}^*} = (1-t)^2\sigma^2_{x_i^*} + 2t(1-t)\sigma_{x_i^*x_{i+1}^*} + t^2\sigma^2_{x_{i+1}^*}$$

$$\sigma_{x_{it}^*y_{it}^*} = (1-t)^2\sigma_{x_i^*y_i^*} + t(1-t)(\sigma_{x_i^*y_{i+1}^*} + \sigma_{x_{i+1}^*y_i^*}) + t^2\sigma^2_{x_{i+1}^*y_{i+1}^*}$$

$$\sigma^2_{y_{it}^*} = (1-t)^2\sigma^2_{y_i^*} + 2t(1-t)\sigma_{y_i^*y_{i+1}^*} + t^2\sigma^2_{y_{i+1}^*}$$

Apparently, variance and covariance estimation are only the accuracy of its adjacent two points. It is still difficult to express equation (6) in one equation, because there is cross-correlation and auto-correlation between all the snapped points. We may complete this computation by above two steps in order.

3.2.2 Variance Estimation of Length of Line

Length is an important geometric property of line feature. Its computation equation is:

$$l^* = \sum_{i=1}^{n-1} z_i^* z_{i+1}^* = \sum_{i=1}^{n-1}[(x_{i+1}^* - x_i^*)^2 + (y_{i+1}^* - y_i^*)^2]^{1/2} \qquad (7)$$

Thus we obtain its variance computation, expressed in general form as follows:

$$\sigma^2_{l^*} = Q\Gamma_{L^*L^*}Q^T \qquad (8)$$

Here, $Q = (\alpha_1, \beta_1, \alpha_2, \beta_2, \cdots, \alpha_n, \beta_n)$, and

$$\alpha_1 = -(x_2^* - x_1^*)/z_1^*z_2^*$$
$$\beta_1 = -(y_2^* - y_1^*)/z_1^*z_2^*$$
$$\alpha_2 = (x_2^* - x_1^*)|/z_1^*z_2^* - (x_3^* - x_2^*)/z_2^*z_3^*$$
$$\beta_2 = (y_2^* - y_1^*)/z_1^*z_2^* - (y_3^* - y_2^*)/z_2^*z_3^*$$

…… …

$$\alpha_{\text{n-1}} = (x_{n-1}^* - x_{n-2}^*)/z_{n-2}^*z_{n-1}^* - (x_n^* - x_{n-1}^*)/z_{n-1}^*z_n^*$$
$$\alpha_{\text{n-1}} = (y_{n-1}^* - y_{n-2}^*)/z_{n-2}^*z_{n-1}^* - (y_n^* - y_{n-1}^*)/z_{n-1}^*z_n^*$$
$$\alpha_{\text{n}} = (x_n^* - x_{n-1}^*)/z_{n-1}^*z_n^*$$
$$\beta_{\text{n}} = (y_n^* - y_{n-1}^*)/z_{n-1}^*z_n^*$$

Therefore, substituting above parameters into equation (8), we will have:

$$\sigma^2_{l^*} = \sum_{i=1}^{n}(\alpha_i^2\sigma^2_{x_i^*} + 2\alpha_i\beta_i\sigma_{x_i^*y_i^*} + \beta_i^2\sigma^2_{y_i^*}) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(\alpha_i\alpha_j\sigma_{x_i^*x_j^*} + \alpha_i\beta_j\sigma_{x_i^*y_j^*} + \beta_i\beta_j\sigma_{y_i^*y_j^*}) \qquad (9)$$

In practice, the points in the snapping group are often independent of each other, and all graphical data are from different independent data layers. Therefore, equation (8) is simplified as:

$$\sigma^2_{l^*} = \sum_{i=1}^{n}(\alpha_i^2\sigma^2_{x_i^*} + \beta_i^2\sigma^2_{y_i^*}) \qquad (10)$$

3.3 Variance Estimation of Area of the Snapped Polygon

For a polygon feature, length and area are its two key geometric attributes. Its variance computation of length is like equation (9) and (10). So we only consider its variance estimation of area. Let the snapped polygon consist of points: $(x_1^*, y_1^*), (x_2^*, y_2^*), \cdots, (x_m^*, y_m^*)$, its circumscribing area is expressed by

$$A = \frac{1}{2}\sum_{j=1}^{m}(y_{j+1}^* - y_{j-1}^*)x_j^* \qquad (11)$$

Or

$$A = \frac{1}{2}\sum_{j=1}^{m}(x_{j-1}^* - x_{j+1}^*)y_j^* \qquad (12)$$

Then its derivate factors are computed by:

$$e_j = \frac{\partial A}{\partial x_j^*} = \frac{1}{2}(y_{j+1}^* - y_{j-1}^*)$$
$$f_j = \frac{\partial A}{\partial y_j^*} = \frac{1}{2}(x_{j-1}^* - x_{j+1}^*)$$

Where lower index $j = 1, 2, \cdots, m$; $x^*_{m+1} = x^*_1$; $y^*_{m+1} = y^*_1$. Furthermore, we will its expression of variance as

$$\sigma_A^2 = \sum_{j=1}^{m}(e_j^2\sigma_{x_i^*}^2 + 2e_j f_j \sigma_{x_i^* y_i^*} + f_j^2 \sigma_{y_i^*}^2) + 2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}(e_i e_j \sigma_{x_i^* x_j^*} + e_i f_j \sigma_{x_i^* y_j^*} + f_i f_j \sigma_{y_i^* y_j^*}) \qquad (13)$$

4. EXAMPLES AND DISCUSSIONS

Consider the cases of ARC/INFO application. In the operations of CLEAN or UNION, INTERSECT and CLIP, we often perform the Node snapping so that a group of arc/nodes within fuzzy tolerance can be snapped together (ESRI, 1988). The accuracy of coordinates in the following examples is based on a land use map at a scale of 1:24,000 (Hord, 1976).

As figure 5 shows, Polygon#1 and Polygon#2 are from two independent data layers, and the displayed data are only a part, which is listed in table 1.

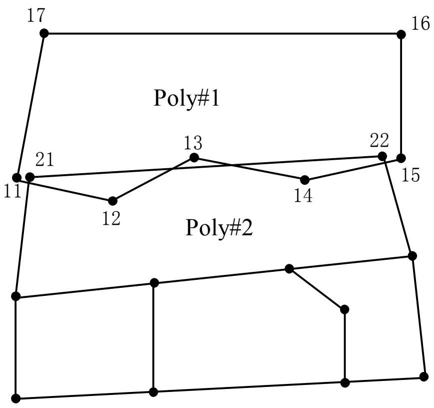


Figure 5. Topological inconsistency of integrating different data sources

ARC#	Point No.	x/m	y/m	σ_x/m	σ_y/m	ρ_{xy}
1	11	768657.60	2935308.40	20.00	20.00	0
	12	768979.20	2935222.00	20.00	20.00	0
	13	769228.80	2935298.80	20.00	20.00	0
	14	769588.80	2935296.60	20.00	20.00	0
	15	769903.00	2935349.20	20.00	20.00	0
	16	769892.40	2935745.80	20.00	20.00	0
	17	768759.80	2935742.60	20.00	20.00	0
2	21	768674.80	2935310.20	10.00	10.00	0
	22	769890.20	2935351.60	10.00	10.00	0

Table 1. Original data

When we make an overlaying operation, many silvers will generate, which is led by different data sources with distinct accuracy. The overlay result is showed in figure 6. Apparently, it needs a processing before making some spatial analyses. Here, for the point pairs 11 and 21, 15 and 22, we take different snapping algorithms to process respectively. Correspondingly, the new generated points are named as 11^* and 22^*. All available methods are MOVE, ADJUST, and General algorithm. For ADJUST algorithm (abbreviated as "AJ"), it can be represented as:

$$x^*_{11} = \frac{1}{2}(x_{11} + x_{21}),\ y^*_{11} = \frac{1}{2}(y_{11} + y_{21}),$$

$$\sigma_{x^*_{11}} = \frac{1}{2}\sqrt{\sigma_{x_{11}}^2 + \sigma_{x_{21}}^2},\ \sigma_{y^*_{11}} = \frac{1}{2}\sqrt{\sigma_{y_{11}}^2 + \sigma_{y_{21}}^2}$$

While for general algorithm (abbreviated as "GA"), it is as follows:

$$x^*_{11} = (\sigma_{x_{11}}^{-2} + \sigma_{x_{21}}^{-2})^{-1}(\sigma_{x_{11}}^{-2} x_{11} + \sigma_{x_{21}}^{-2} x_{21}),$$

$$y^*_{11} = (\sigma_{y_{11}}^{-2} + \sigma_{y_{21}}^{-2})^{-1}(\sigma_{y_{11}}^{-2} y_{11} + \sigma_{y_{21}}^{-2} y_{21}),$$

$$\sigma_{x^*_{11}} = \pm(\sigma_{x_{11}}^{-2} + \sigma_{x_{21}}^{-2})^{-1},\ \sigma_{y^*_{11}} = \pm(\sigma_{y_{11}}^{-2} + \sigma_{y_{21}}^{-2})^{-1}$$

The computation results are listed in Table 2. And data with a rectangle cannot be provided by current GIS software. Further we may compute accuracy of length of line feature and area of polygon feature, listed in table 3.

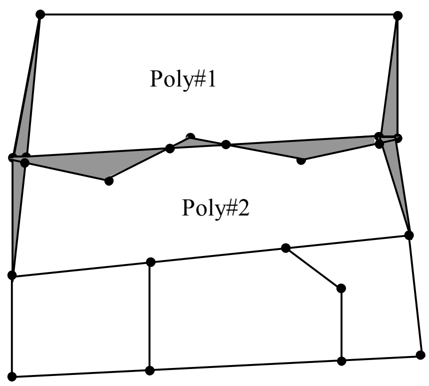


Figure 6. Generations of spurious polygons in topology construction before data cleaning

Methods	IP#.	x/m	y/m	σ_x/m	σ_y/m	ρ_{xy}
MV	11^*	768674.80	2935310.20	10.00	10.00	0
	22^*	769890.20	2935351.60	10.00	10.00	0
AJ	11^*	768666.20	2935309.30	11.18	11.18	0
	22^*	769896.60	2935350.40	11.18	11.18	0
GA	11^*	768671.36	2935309.84	8.94	8.94	0
	22^*	769892.76	2935351.12	8.94	8.94	0

Table 2. The results from different snapping methods

		ARC#1			Poly#1		
		$l\,(m)$	$\sigma_l\,(m)$	σ_l/l	$A\,(m^2)$	$\sigma_A\,(m^2)$	σ_A/A
Before Cleaning		1265.829	31.173	24/1000	541906.980	22194.848	410/10000
After Cleaning	MV	1216.105	14.142	12/1000	484241.920	19694.486	407/10000
	AJ	1231.086	15.811	13/1000	488564.420	20301.471	416/10000
	GA	1222.097	12.643	10/1000	485968.990	19341.226	398/10000
Differences	MV	49.724	17.031	12/1000	57665.060	2500.362	3/10000
	AJ	34.743	15.362	11/1000	53342.560	1893.377	-6/10000
	GA	43.832	18.531	14/1000	55937.990	2853.622	12/10000

Table 3. Calculation data of length of lines and area of polygons and their standard errors

We may find from table 2 that the results obtained by MV, AJ and GA methods are distinct, and have a large difference. For MV methods, only accuracy of the data source with better quality is considered, ignoring accuracy information of other data sources. AJ methods considers all accuracy information of all data sources only take a simple average value of all accuracies, i.e., equal weights for all accuracies of data sources. GA methods may be regarded as a kind of a weighted average value, which considers not only all accuracies of data sources, but also levels of their accuracies.

5. CONCLUSIONS

Topological inconsistency is almost unavoidable in making some spatial operations involving multiple data sources with distinct accuracies. In this paper, various cases of inconsistency are summarized and some existing methods are also reviewed. The new method can be applied to equal or non-equal accuracy snapping point group with dependency or independency and provides reliable theoretically estimated results of the snapped points and their accuracy information, which is necessary for GIS to estimate quality of its resulting products. On the other hand, the models used for accuracy estimating of length of lines and of area of polygons built in this paper are also suitable for spatial operations like 'intersection', 'union', etc.

ACKNOWLEDGEMENTS

The authors acknowledge funding support received from the National Natural Science Foundation Council of China (grant no. 49801016 and 40101022).

REFERENCES

Burrough PA, McDonnell RA, 1998. *Principles of Geographical Information Systems*. Oxford University Press, Oxford, pp. 220-240.

Chrisman NR, 1987. The Accuracy of Map Overlays: A Reassessment. Landscape and Urban Planning, 14: pp. 427-439.

Chrisman NR, 1989. Modeling Error in Overlaid Categorical Maps. In: *Accuracy of Spatial Databases*, edited by M.F. Goodchild and S. Gopal, Taylor and Francis, London, pp. 21-34.

Clarke KC, 1996. *Analytical and Computer Cartography*. Englewood Cliffs, New Jersey, pp. 198-203.

Guptill SC, Morrison JL, 1995. *Elements of Spatial Data Quality*. Elsevier Science, New York, pp. 1~12.

Harvey F, 1994. Defining Unmovable Nodes/Segments as Part of Vector Overlay, in: *Advances in GIS Research*. Taylor and Francis, London, pp. 159-176.

JONES CB, 1997. *Geographical Information Systems and Computer Cartography*. Longman, Harlow, pp. 87-92,121-131.

LIU Wen-bao, XIA Zong-guo, CUI Xian-guo, 2001. A New Generalized Algorithm of Node Snapping and a Universal Model of Error Propagating. *ACTA GEODAETICA et CARTOGRAPHICA SINICA*, 30(2), pp. 140-147.

MacDougall E B. The Accuracy of Map Overlays. *Landscape Planning*, 1975, 2, pp. 23-30.

SHI Wen-zhong, LIU Wen-bao, 2000. A Stochastic Process-Based Model for Positional Error of Line Segments in GIS. *International Journal of Geographic Information Sciences*, Vol.14 (1), pp. 51-66.

Veregin H, 1999. Data quality parameters. In: Longley P A. et al. (ed.), *Geographical Information System*, 2nd Eds. Vol.1, John Wiley & Sons, New york, pp. 177~189.