

文章编号: 1001-1595(2001)02-0140-08

中图分类号: P208

文献标识码: A

GIS 结点捕捉的广义算法及误差传播模型

刘文宝^{1,2}, 夏宗国², 崔先国¹

(1. 山东科技大学 地球科学系, 山东 泰安 271019; 2. 美国麻省大学 地球地理科学系, 麻省 波士顿 02125)

A New Generalized Algorithm of Node Snapping and a Universal Model of Error Propagating

LU Wen-bao^{1,2}, XIA Zong-guo², CUI Xian-guo¹

(1. Department of Geosciences, Shandong University of Science and Technology, Taian 271019, China;

2. Department of Earth and Geographic Sciences, University of Massachusetts, Boston, MA 02125, U. S. A.)

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 such as Boolean operations. However, there is a lack of models for dealing with errors in node snapping which is one of the basic operations in most GIS software packages such as ARCED IT in ARC/INFO. Node snapping involves changing the position of points of a spatial feature so that points within a fuzzy tolerance of each other are connected automatically. 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.

In this paper, we propose 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. The impact of statistical dependency of the existing points on the coordinates of the point obtained after snapping and the accuracy of the snapped point are examined through theoretical analysis and numerical simulation. Our results indicate that the level of influence depends on the coefficients of correlation between points involved in the snapping operation. We have shown theoretically that the snapping algorithm widely used in the past merely represents a special case of the generalized algorithm. Several examples are also provided to demonstrate the potential applications of the generalized algorithm and the related error model.

Key words: node snapping; fuzzy tolerance; generalized algorithm; error propagation

摘要: 根据最小二乘原理, 本文提出了结点捕捉的一种广义算法, 并建立了伴随的误差传播模型。针对位于模糊公差范围内待捕捉点组中各点坐标误差统计特性的各种可能的特殊情况, 进一步导出了相应的简化算法及其误差传播模型, 并从理论和数值模拟两方面系统地分析了

收稿日期: 2000-01-04; 修回日期: 2000-09-29

基金项目: 国家自然科学基金资助项目(49801016)

作者简介: 刘文宝(1965-), 男, 山东沂水人, 山东科技大学教授兼美国麻省大学研究员, 博士, 主要从事遥感、地图制图学与地理信息系统中的数据质量和信息可视化研究。

结点坐标的相关性对捕捉结果的影响。理论推导表明, 现有的 GIS 结点捕捉算法属广义算法的一种特殊情况。最后通过算例说明了广义算法和误差传播模型的实际应用前景。

关键词: 结点捕捉; 模糊公差; 广义算法; 误差传播

1 引言

数据质量是近年来国际 GIS 界的重要研究专题之一^[1]。从应用角度看, 数据质量的研究重点集中在误差与不确定性的控制、削弱和消除上, 这方面的工作在当前的 GIS 软件中是通过数据编辑实现的^[2]。在许多现有 GIS 软件的数据编辑中, 捕捉属基本功能之一^[3]。若以 ARC/INFO 为例, 则捕捉类型有: 结点捕捉、弧捕捉和一般捕捉^[4]。其中, 结点捕捉常需要将两个或多个位于某一给定的模糊公差范围内的一组弧结点捕捉为同一坐标点, 从而消除原本为同一结点, 但因数据来源不同所存在的数据质量间的矛盾, 建立正确的多边形或弧/结点拓扑关系, 即拓扑关系清理 (CLEAN)^[5,6]。ARC/INFO 结点捕捉的常用命令有 ADD、MOVE、ROTATE 和 ADJUST 等^[4,5]。例如, 若通过判断可知这组点中存在精度较高或较可靠的一点, 则通过 MOVE 或 ROTATE 命令将精度相对较低的点移位或旋转到精度最高或确认为正确的点上。显然, 这种捕捉后所得点的精度将简单地继承了它原有的精度, 而忽略了精度较低的那些点中所含有的位置信息。当数据源为小比例尺地图时, 这种忽略会引起很大的误差。如果经判断需要采用 ADJUST 命令进行结点捕捉, 其算法则为简单地取点组坐标的算术平均值。由于缺少其中的误差传播模型, 至今无法评估捕捉后所得新点坐标的精度。当位于模糊公差范围内点组坐标的精度间存在较大差异时, 现有 ADJUST 命令所依据的简单算术平均值算法是不合理的。为此, 本文提出结点捕捉的一种广义算法, 并重点讨论与其相应的误差传播模型, 以便为改造现有 GIS 中结点捕捉的 ADJUST 算法提供理论依据。

2 算法与误差传播模型

2.1 广义算法与误差模型

令 $z_i = (x_i, y_i)^T, (i = 1, 2, \dots, n)$ 为模糊公差范围内点组 (以下简称“捕捉点组”) 中的某点, 其

$$\text{协方差阵为 } \Gamma_{ii} = \begin{bmatrix} \sigma_{x_i}^2 & \sigma_{x_i y_i} \\ \sigma_{y_i x_i} & \sigma_{y_i}^2 \end{bmatrix}。$$

若 $\sigma_{x_i y_i} = \sigma_{y_i x_i} = 0, (i = 1, 2, \dots, n)$, 则称 z_i 为自相关向量。记 $z^* = (x_1, y_1, x_2, y_2, \dots, x_n, y_n)^T$, 则捕捉点组中的 n 个点的协方差阵为

$$\Gamma_{z^* z^*} = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1 y_1} & \sigma_{x_1 x_2} & \sigma_{x_1 y_2} & \dots & \sigma_{x_1 x_n} & \sigma_{x_1 y_n} \\ \sigma_{y_1 x_1} & \sigma_{y_1}^2 & \sigma_{y_1 x_2} & \sigma_{y_1 y_2} & \dots & \sigma_{y_1 x_n} & \sigma_{y_1 y_n} \\ \sigma_{x_2 x_1} & \sigma_{x_2 y_1} & \sigma_{x_2}^2 & \sigma_{x_2 y_2} & \dots & \sigma_{x_2 x_n} & \sigma_{x_2 y_n} \\ \sigma_{y_2 x_1} & \sigma_{y_2 y_1} & \sigma_{y_2 x_2} & \sigma_{y_2}^2 & \dots & \sigma_{y_2 x_n} & \sigma_{y_2 y_n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{x_n x_1} & \sigma_{x_n y_1} & \sigma_{x_n x_2} & \sigma_{x_n y_2} & \dots & \sigma_{x_n}^2 & \sigma_{x_n y_n} \\ \sigma_{y_n x_1} & \sigma_{y_n y_1} & \sigma_{y_n x_2} & \sigma_{y_n y_2} & \dots & \sigma_{y_n x_n} & \sigma_{y_n}^2 \end{bmatrix} = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} & \dots & \Gamma_{1n} \\ \Gamma_{21} & \Gamma_{22} & \dots & \Gamma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{n1} & \Gamma_{n2} & \dots & \Gamma_{nn} \end{bmatrix} \quad (1)$$

其中

$$\Gamma_{ij} = \begin{bmatrix} \sigma_{x_i y_j} & \sigma_{x_i x_j} \\ \sigma_{y_i x_j} & \sigma_{y_i y_j} \end{bmatrix}$$

若 $\Gamma_{ij} = 0 (i \neq j)$, 则称 z_i 与 $z_j (i \neq j)$ 为互相关向量; 若对部分或全部的 $i, j = 1, 2, \dots, n (i \neq j)$, 有 $\Gamma_{ij} = 0 (i \neq j)$ 成立, 则称 z^* 为部分或全互相关向量。记 Ψ 为 $\Gamma_{z^* z^*}$ 的逆矩阵, 即 $\Gamma_{z^* z^*}^{-1} = \Psi_{2n \times 2n}$ 。令 $\Phi_{ij} (i, j = 1, 2, \dots, n)$ 为 $2n \times 2n$ 阶方阵 $\Psi_{2n \times 2n}$ 中与逆矩阵 $\Gamma_{z^* z^*}$ 中的 Γ_{ij} 相对应的 2×2 阶分块子矩阵; $z = (x, y)^T$ 为经结点捕捉后所得新点坐标的最优估值向量, 其协方差阵为 $\Gamma = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix}。$

由于模糊公差范围内点组的结点捕捉算法实质上属测量平差中对一组直接观测量 (即点组坐标) 进行平差, 故根据最小二乘原理^[7], 可导出结点捕捉坐标估值的广义算法公式

$$z = \left(\sum_{i=1}^n \sum_{j=1}^n \Phi_{ij} \right)^{-1} \sum_{i=1}^n \sum_{j=1}^n (\Phi_{ij} z_j) \quad (2)$$

在结点捕捉过程中, 由捕捉点组向量 z^* 中的点坐标误差到捕捉后所得新点坐标 z 的误差传播模型为

$$\Gamma = \left(\sum_{i=1}^n \sum_{j=1}^n \Phi_{ij} \right)^{-1} \quad (3)$$

显然, 结点捕捉广义算法式 (2) 和误差传播模型式 (3) 适用于捕捉点组为多分辨率、多源数据的任意

情况。

2.2 纯自相关算法与误差模型

在GIS的实际应用中,常遇到捕捉点组中各点间相互独立的情况,例如当这 n 个点分别来自 n 层相互独立的数据层时。此时,对所有的 $i, j = 1, 2, \dots, n$, 均有 $\Gamma_{ij} = 0 (i \neq j)$, 故对部分或全部 $\sigma_{x_i y_i} = \sigma_{y_i x_i} = 0, (i = 1, 2, \dots, n)$ 的情况,可称 z^* 为纯自相关向量。因此,式(1)变为分块对角矩阵,即 $\Gamma_{z^* z^*} = \text{diag}\{\Gamma_{11} \Gamma_{22} \dots \Gamma_{nn}\}$, 而 $\Psi = \Gamma^{-1} = \text{diag}\{\Gamma_{11}^{-1} \Gamma_{22}^{-1} \dots \Gamma_{nn}^{-1}\}$ 。于是,式(2)简化为下列纯自相关算法

$$z = \left(\prod_{i=1}^n \Gamma_{ii}^{-1} \right)^{-1} \left(\prod_{i=1}^n \Gamma_{ii}^{-1} z_i \right) \quad (4)$$

而式(3)简化为下述纯自相关误差传播模型

$$\Gamma = \left(\prod_{i=1}^n \Gamma_{ii}^{-1} \right)^{-1} \quad (5)$$

可以说,即使捕捉点组中各点间是互相关的,但却无法获取这种互相关信息时,也只能近似地视为上述纯自相关情况。

2.3 等自协方差阵算法与误差模型

在上述 n 个待捕捉点间无互相关性,或因缺少互相关信息而忽略不计的前提下,如果捕捉点组又是来自等精度的数据源,则有 $\Gamma_{ii} = \Gamma_{00} =$

$$\begin{bmatrix} \sigma_{x_0}^2 & \sigma_{x_0 y_0} \\ \sigma_{y_0 x_0} & \sigma_{y_0}^2 \end{bmatrix}, (i = 1, 2, \dots, n)。$$

例如,当对同一幅地图的数字化数据进行拓扑关系清理时,就会遇到这种情况。此时,式(4)简化为下列等自协方差阵算法

$$\left. \begin{aligned} x &= \frac{1}{n} \sum_{i=1}^n x_i \\ y &= \frac{1}{n} \sum_{i=1}^n y_i \end{aligned} \right\} \quad (6)$$

相应地,式(5)简化为下述等自协方差阵误差传播模型

$$\left. \begin{aligned} \sigma_x &= \sigma_{x_0} / \sqrt{n} \\ \sigma_y &= \sigma_{y_0} / \sqrt{n} \\ \sigma_{xy} &= \sigma_{y_x} = \sigma_{x_0 y_0} / n \end{aligned} \right\} \quad (7)$$

分析式(6)、式(7)可见,尽管捕捉点组中的各点是纯自相关的,即每一坐标对 $x_i, y_i (i = 1, 2, \dots, n)$ 间都存在相关性,但在捕捉点组为等精度的情况下,捕捉后所得新点坐标对 x, y 的估值及其方差 σ_x^2, σ_y^2 , 却均与这种相关性无关,故其值可分别单独计算。此时,只有新点坐标对 x, y 间的自协方差 σ_{xy} 与捕捉点组中坐标对 $x_i, y_i (i = 1, 2, \dots, n)$ 间

的自协方差 $\sigma_{x_0 y_0}$ 有关。

2.4 实用算法与误差模型

实际工作中有时遇到的情况常常比较简单,一般是捕捉点组中无互相关性,且各点坐标分量 x_i 和 $y_i (i = 1, 2, \dots, n)$ 的精度相等,即 $\Gamma_{ij} = 0 (i \neq j)$ 且 $\sigma_{x_i}^2 = \sigma_{y_i}^2 = \sigma_i^2, (i = 1, 2, \dots, n)$ 。

2.4.1 基本算法与模型

首先考虑捕捉点组为非等精度的基本情况,即 $\Gamma_{ij} = \Gamma_{ji}, (i, j = 1, 2, \dots, n; i \neq j)$ 。顾及相关系数的定义^[7], 即 $\rho_i = \sigma_{x_i y_i} / \sigma_{x_i} \sigma_{y_i}$, 则有

$$\Gamma_{ii} = \begin{bmatrix} \sigma_{x_i}^2 \sigma_{y_i}^2 & \\ & \sigma_{y_i}^2 \sigma_{x_i}^2 \end{bmatrix} = \sigma_i^2 \begin{bmatrix} 1 & \rho_i \\ \rho_i & 1 \end{bmatrix} \quad (8)$$

于是,式(4)变为下列基本算法

$$z = \left[\prod_{i=1}^n \frac{1}{\sigma_i^2 (1 - \rho_i^2)} \begin{bmatrix} 1 & -\rho_i \\ -\rho_i & 1 \end{bmatrix} \right]^{-1} \times \left[\prod_{i=1}^n \frac{1}{\sigma_i^2 (1 - \rho_i^2)} \begin{bmatrix} x_i - \rho_i y_i \\ y_i - \rho_i x_i \end{bmatrix} \right] \quad (9)$$

而式(5)简化为下列误差传播基本模型

$$\Gamma = \left[\prod_{i=1}^n \frac{1}{\sigma_i^2 (1 - \rho_i^2)} \begin{bmatrix} 1 & -\rho_i \\ -\rho_i & 1 \end{bmatrix} \right]^{-1} \quad (10)$$

从式(9)、式(10)可知,在结点捕捉点组为基本模型的前提下,捕捉后所得新点坐标对 x, y 的估值及其方差 σ_x^2, σ_y^2 均与捕捉点组中各点的自相关系数 ρ_i 有关,故理论上必须联合估计。

2.4.2 等自相关系数算法和误差模型

尽管捕捉点组中各点的精度不同,但当遇到影响各点自相关性因素相同的情况时,则会出现捕捉点组自相关系数全等的情况,即 $\rho_i = \rho_0, (i = 1, 2, \dots, n)$ 。此时,式(9)简化为下列等自相关系数算法

$$\left. \begin{aligned} x &= \frac{1}{n} \sum_{i=1}^n (\sigma_i^2 x_i) / \frac{1}{n} \sum_{i=1}^n \sigma_i^2 \\ y &= \frac{1}{n} \sum_{i=1}^n (\sigma_i^2 y_i) / \frac{1}{n} \sum_{i=1}^n \sigma_i^2 \end{aligned} \right\} \quad (11)$$

而式(10)简化为

$$\left. \begin{aligned} \sigma_x &= \sigma_y = \pm 1 / \sqrt{\frac{1}{n} \sum_{i=1}^n \sigma_i^2} \\ \sigma_{xy} &= \sigma_{yx} = \rho_0 / \frac{1}{n} \sum_{i=1}^n \sigma_i^2 \end{aligned} \right\} \quad (12)$$

由式(11)、式(12)可知,捕捉后所得新点坐标对 x, y 的估值及其方差 σ_x^2, σ_y^2 均与捕捉点组中各点的自相关系数 ρ_0 无关,但新点坐标对 x, y 间的自协方差 σ_{xy} 与 ρ_0 有关。



2.4.3 纯独立算法和误差模型

进一步, 如果捕捉点组中的所有点均是非自相关的, 即自相关系数 $\rho_i = 0, (i= 1, 2, \dots, n)$, 那么式(11)保持不变, 而式(12)简化为

$$\left. \begin{aligned} \sigma_x = \sigma_y = \pm 1 / \sqrt{\sum_{i=1}^n \sigma_i^2} \\ \sigma_{xy} = \sigma_{yx} = 0 \end{aligned} \right\} \quad (13)$$

2.4.4 纯独立且等方差模型

如果捕捉点组中所有点均是非自相关的, 且均是等精度的, 即自相关系数 $\rho_i = 0$ 且 $\sigma_i^2 = \sigma_0^2 (i= 1, 2, \dots, n)$, 那么式(11)退化为式(6), 而式(13)简化为

$$\left. \begin{aligned} \sigma_x = \sigma_y = \sigma_0 / \sqrt{n} \\ \sigma_{xy} = \sigma_{yx} = 0 \end{aligned} \right\} \quad (14)$$

上式也可直接从式(7)经相同的简化假设后导出。

2.5 算法与误差模型理论分析

综合分析式(7)、式(10)、式(12)、式(13)和式(14)可见, 若捕捉点组中各点均是自相关的, 则捕捉后的新点坐标对 x, y 间也是自相关的。反之, 亦成立。同时, 由式(6)、式(7)、式(11)和式(12)可知, 尽管捕捉点组中各点是自相关的, 但若满足该两模型的前提假设条件, 则捕捉后点的坐标估值及其方差将不受这种相关性影响, 故可不考虑这种相关性, 而分开单独估计。

算法式(6)和模型式(7)适用于来自同一数据层的点组捕捉情况, 因为来自同一数据层的数据常常是等精度的。在现行ARC/NFO 的拓扑关系清理时, 正是以式(6)为结点捕捉中ADJUST 命令的算法基础, 只是缺少式(7)的误差传播算法。故捕捉后的结果只提供坐标值, 而缺少精度信息。

2.6 捕捉点组的相关性影响数值分析

文献[8, 9]研究表明, 尽管同一层上的流方式手工数字化数据间存在着强相关性, 但很难估计这种相关值。因此, 在实际应用上述算法和误差模型时, 一个十分重要的问题是: 若捕捉点组中事实上存在自相关性或互相关性, 但因缺少这种相关性数值信息而不加考虑时, 则会对捕捉后的结果产生怎样的影响? 为此, 下面通过设计的一组数值模拟试验进行分析。

2.6.1 试验设计要点

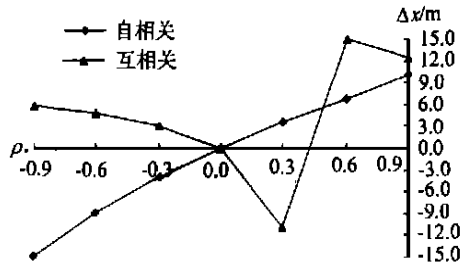
由于实际捕捉点组中的点数一般为 2~ 4 个, 故选取 3 个点进行试验。假设同一点的 x, y 坐标精度相等(实际情况常如此), 记为 σ^2 , 令自相关系数为 $\rho_i (i= 1, 2, 3)$, 而互相关系数为 ρ_{12}, ρ_{13} 和 ρ_{23} 。为了分离自相关性和互相关性对分析结果的影响, 在考虑其中一个时假定另一个为 0。对每一种相关性, 又分单相关和多相关分别讨论。将每一种情况下的计算结果与相关系数均为 0 时的结果相比较, 即可得到当存在相关性但忽略不计时, 对结点捕捉结果的影响。

2.6.2 自相关性影响分析

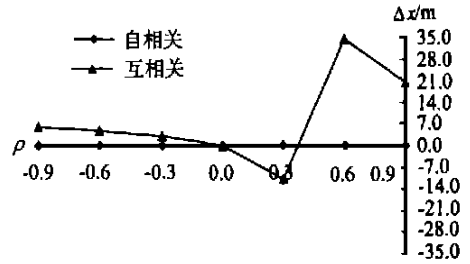
考虑单自相关和多自相关两种情况。对于单自相关, 不妨设仅第一个点中存在自相关性, 即 $\rho_1 \neq 0, \rho_2 = \rho_3 = 0$; 对于多自相关, 又分为 3 个点中的自相关性呈相同、相近和相异较大 3 种情况。对单自相关, 令 ρ_1 以步长 0.1 从 ± 0.1 到 ± 0.9 依次变化取值, 捕捉后所得新点的坐标 x, y 根据广义算法式(2)计算, 方差 σ_x^2, σ_y^2 和协方差 σ_{xy} 利用误差模型式(3)估计, 而新点坐标分量 x, y 间相关系数 ρ 由 $\rho = \sigma_{xy} / \sigma_x \sigma_y$ 确定。比较 ρ_i 取不同值时的有关计算结果与 $\rho_i = 0$ 时的相应值, 即可得到忽略自相关性所引起的影响值, 具体计算公式为

$$\left. \begin{aligned} \rho &= (\rho_1, \rho_2, \rho_3)^T; \Delta \rho = \rho(\rho) - \rho(\rho=0) \\ \Delta x &= x(\rho) - x(\rho=0); \Delta \sigma_x = \sigma_x(\rho) - \sigma_x(\rho=0) \\ \Delta y &= y(\rho) - y(\rho=0); \Delta \sigma_y = \sigma_y(\rho) - \sigma_y(\rho=0) \end{aligned} \right\} \quad (15)$$

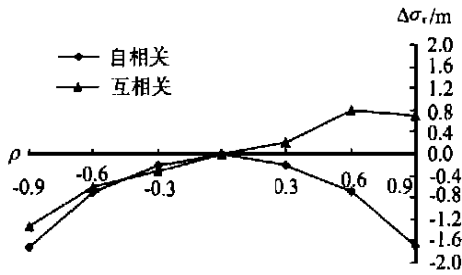
图 1 分别以 $\Delta x, \Delta \sigma_x$ 和 $\Delta \rho$ 为例绘出了它们与自相关系数 ρ 间的关系曲线。其中, 图 1(a) 为自相关和互相关系数对 x 坐标的影响图, 单位为米; 图 1(b) 为自相关和互相关系数对 σ_x 的影响图, 自相关情况的单位为米, 互相关情况的单位为 $\Delta \sigma_x \times 10$ 米; 图 1(c) 为自相关和互相关系数对 ρ 的影响图, 互相关情况的为 $\Delta \rho / 1000$ 。类似地, 对多自相关也进行了数值模拟分析。其中, 多自相关性呈相同的关系曲线如图 2 所示, 而多自相关性呈相近和相异较大的结果分别列于表 1、2。图 2 中, 图 2(a) 为自相关和互相关系数对 x 坐标的影响图, 单位为米; 图 2(b) 为自相关和互相关系数对 σ_x 的影响图; 图 2(c) 为自相关和互相关系数对 ρ 的影响图, 互相关情况的为 $\Delta \rho / 1000$ 。



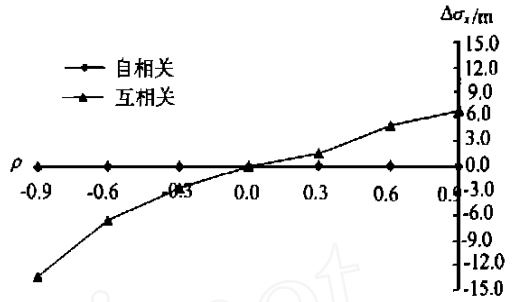
(a)



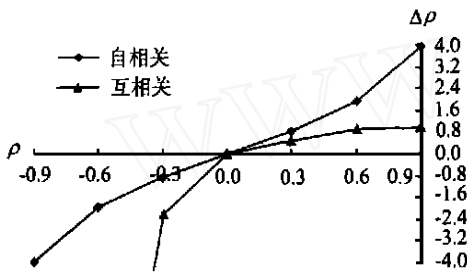
(a)



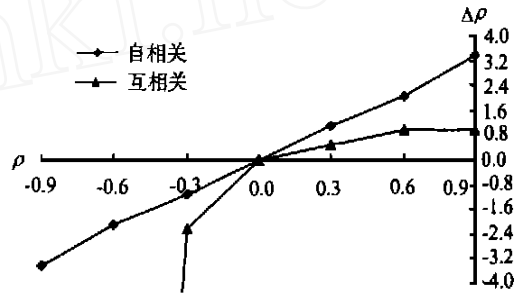
(b)



(b)



(c)



(c)

图1 单自相关和单互相关影响图

Fig 1 Effect of correlation coefficients of simple auto-correlation and cross-correlation of existing data points on the coordinates, standard deviation and coefficient of auto-correlation of the snapped point

图2 等相关系数的多自相关和多互相关影响图

Fig 2 Effect of equal correlation coefficients of multiple auto-correlation and cross-correlation of existing data points on the coordinates, standard deviation and coefficient of auto-correlation of the snapped point

表1 自相关系数近似相等的多自相关

Tab 1 Multiple auto-correlation with approximately equal coefficients of correlation

($\rho_1 \rho_2 \rho_3, \rho_{12} = \rho_{13} = \rho_{23} = 0$)

ρ_1	ρ_2	ρ_3	$\Delta x/m$	$\Delta y/m$	$\Delta \sigma_r/m$	$\Delta \sigma_r/m$	$\Delta \rho$
- 0.7	- 0.8	- 0.9	4 556.7	3 937.7	- 0.0917	- 0.0917	- 0.0029
- 0.5	- 0.7	- 0.8	5 885.6	4 954.4	- 0.1626	- 0.1626	- 0.0023
- 0.2	- 0.3	- 0.5	2 048.6	0 791.2	- 0.0561	- 0.0561	- 0.0010
- 0.1	- 0.2	- 0.3	1 687.0	0 777.3	- 0.0295	- 0.0295	- 0.0005
0	0	0	0	0	0	0	0
0.1	0.2	0.3	- 1 696.2	- 0 077.5	- 0.0295	- 0.0295	0.0005
0.2	0.3	0.5	- 2 620.6	0 963.2	- 0.0561	- 0.0561	0.0010
0.5	0.7	0.8	- 4 754.7	2 239.6	- 0.1626	- 0.1626	0.0023
0.7	0.8	0.9	- 5 683.4	4 513.0	- 0.0917	- 0.0917	0.0029

表 2 自相关系数相异较大的多自相关

Tab 2 Multiple auto-correlation with significantly different coefficients of correlation ($\rho_{12} \rho_{13} \rho_{23} = 0$)

ρ_1	ρ_2	ρ_3	$\Delta x/m$	$\Delta y/m$	$\Delta \sigma_x/m$	$\Delta \sigma_y/m$	$\Delta \rho$
- 0.9	- 0.5	0	- 13 840.7	- 8 820.0	- 0.716 9	- 0.716 9	- 0.003 5
- 0.5	- 0.2	0	- 5.755 5	- 3.110 9	- 0.232 9	- 0.232 9	- 0.001 6
0	- 0.5	0.5	3.778 3	11.846 9	- 0.699 7	- 0.699 7	- 0.000 6
0	- 0.5	0.9	- 15.046 0	30.728 6	- 2.111 6	- 2.111 6	0.000 2
0	0	0	0	0	0	0	0
0	- 0.9	0.9	1.310 2	48.715 8	- 4.614 1	- 4.614 1	- 0.003 4
0.5	0.2	0	3.864 3	0.829 0	- 0.232 9	- 0.232 9	0.001 6
0.9	0.5	0	5.581 7	- 0.588 4	- 0.716 9	- 0.716 9	0.003 5
0.9	- 0.9	0.9	27.348 7	30.333 8	- 8.574 7	- 8.574 7	0.006 5
0.5	- 0.5	0.5	11.697 0	13.584 5	- 1.853 1	- 1.853 1	0.001 0

从图 1 可见, 在单自相关情况下, 随着自相关系数的增大, 略其不计对捕捉后所得新点坐标 x, y , 标准差 σ_x, σ_y 和相关系数 ρ 的影响也增大。其中对坐标 x, y 的影响呈近似线性特征, 如图 1(a); 对标准差 σ_x, σ_y 的影响呈近似抛物线特征, 如图 1(b); 对相关系数 ρ 的影响呈近似对称的一般曲线, 如图 1(c)。从图 2 可见, 在自相关性相同的多自相关情况下, 忽略相关性将不影响新点坐标 x, y 及其标准差 σ_x, σ_y 的估值, 如图 2(a)、2(b); 但对自相关系数 ρ 的估计存在影响, 影响曲线呈近似线性特征, 如图 2(c)。这与直接从式 (11)、式(12)的理论分析中所得出的结论一致。从

表 1、2 可知, 在多自相关系数存在差异的情况下, 忽略相关系数对捕捉结果的影响总是存在的。

2.6.3 互相关性影响分析

类似地, 考虑了 4 种情况, 且计算仍采用式 (15), 但其中的相关系数为互相关系数。单互相关情况下的结果如图 1, 多互相关且相关系数相等的结果如图 2, 相关系数相近和差异较大情况下的结果分别列于表 3、4。从图 1、2 中可见, 对单互相关和互相关系数相等的多互相关, 忽略相关性对捕捉后所得新点坐标 x, y 及其标准差 σ_x, σ_y 和相关系数 ρ 的影响曲线的趋势类似。

表 3 互相关系数近似相等的多互相关

Tab 3 Multiple cross-correlation with approximately equal coefficients of correlation ($\rho_1 = \rho_2 = \rho_3 = 0, \rho_{12} \rho_{13} \rho_{23}$)

ρ_{12}	ρ_{13}	ρ_{23}	$\Delta x/m$	$\Delta y/m$	$\Delta \sigma_x/m$	$\Delta \sigma_y/m$	$\Delta \rho$
- 0.2	- 0.3	- 0.5	10.170 9	10.170 9	- 8.587 2	- 8.587 2	- 0.061 2
- 0.1	- 0.2	- 0.3	5.204 5	5.204 5	- 3.468 0	- 3.468 0	- 0.003 8
0	0	0	0	0	0	0	0
0.1	0.2	0.3	1.191 7	1.191 7	0.158 9	0.158 9	0.000 1
0.2	0.3	0.5	36.118 3	36.118 3	- 6.494 6	- 6.494 6	- 0.019 0
0.5	0.7	0.8	- 77.806 4	- 77.806 4	- 3.614 0	- 3.614 0	- 0.004 2

表 4 互相关系数相异较大的多互相关

Tab 4 Multiple cross-correlation with significantly different coefficients of correlation ($\rho_1 = \rho_2 = \rho_3 = 0$)

ρ_{12}	ρ_{13}	ρ_{23}	$\Delta x/m$	$\Delta y/m$	$\Delta \sigma_x/m$	$\Delta \sigma_y/m$	$\Delta \rho$
- 0.5	- 0.2	0	3.654 6	3.654 6	- 5.203 9	- 5.203 9	- 0.009 7
0	- 0.5	0.5	21.674 8	21.674 8	1.685 2	1.685 2	0.000 6
0	- 0.5	0.9	0.856 0	0.856 0	0.233 8	0.233 8	0.000 1
0	0	0	0	0	0	0	0
0	- 0.9	0.9	8.292 6	8.292 6	0.684 7	0.684 7	0.000 3
0.5	0.2	0	303.056 1	303.056 1	15.673 4	15.673 4	0.000 7
0.9	0.5	0	15.171 1	15.171 1	7.733 0	7.733 0	0.000 9
0.9	- 0.9	0.9	12.764 3	12.764 3	7.794 1	7.794 1	0.000 9
0.5	- 0.5	0.5	20.339 6	20.339 6	5.743 2	5.743 2	0.000 9

从数值模拟分析中还发现,在所有情况下,忽略捕捉点组的自相关或互相关性对捕捉后所得新点坐标相关系数 ρ 的影响值 $\Delta\rho$ 很小,而对坐标 x, y 及其标准差 σ_x, σ_y 的影响值 $\Delta x, \Delta y, \Delta\sigma_x, \Delta\sigma_y$ 一般小于该点的标准差 σ_x, σ_y 。

3 算例分析

考虑 ARC/NFO 中的实际情况,在数据 CLEAN 操作或 UNION、INTERSECT 及 CLIP 等叠置操作中,常需将位于模糊公差范围内的一组弧节点捕捉为同一点^[4]。下列算例中坐标的精度参照 1:24 000 比例尺的土地利用图设计^[10]。

3.1 等精度结点情况

如图 3,在某数字化要素层单元内存在位于模糊公差范围内的 3 个点 19004、19005 和 19006,其坐标和精度信息列于表 5 中上半部 (Part 1)。为了重建拓扑关系,若在 ARC/NFO 内,则是通过命令 MOVE 或 ADJUST 实现的^[4]。若采用 MOVE 命令,设将 19005 和 19006 两点简单地移位到 19004 点上,则 19004 点即为捕捉后所得的新点,其坐标和精度信息与原 19004 点的相同。若采用 ADJUST 命令,则处理结果列于表 5 下半部中 (Part 2) 第“A”行。若采用本文提出的新方法,即利用广义算法式(2)及其误差传播模型式(3)计算,其结果列于第“N”行。另外,表 5 中的“N-M”为新方法与 MOVE 方法结果间的差值,而“N-A”为新方法与 ADJUST 方法结果间的差值。“--”表示 ADJUST 命令不提供精度信息。

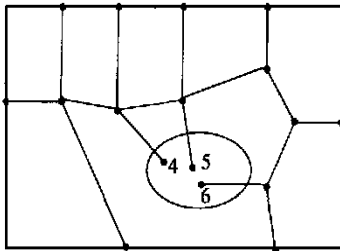


图 3 精度相同的 3 点的捕捉

Fig 3 Node snapping with three nodes of equal accuracy

分析表 5 可见,ARC/NFO 中 ADJUST 命令所依据的是广义算法式(2)的特例,即式(6),由于本例中的 3 个结点为等精度的,故 ADJUST 命令所提供的坐标计算值与广义算法的相同;MOVE 命令可以简单地保留所选择点的坐标和精度信息,尽管实际的 ARC/NFO 中还没有考

虑这种精度信息问题。对于本例,ADJUST 命令的缺点是缺少提供精度信息的功能,而 MOVE 命令忽略了其他两点中所含有的位置信息,后者导致了所提供的坐标值与精度信息与广义算法的结果间存在着几米的差异,见表 5 中末行。

表 5 等精度结点在选用不同捕捉算法下的结果

Tab 5 Different results from different snapping algorithms for data points of equal accuracy

Point No.	x/m	y/m	σ_x/m	σ_y/m	ρ_{xy}
19004	763 289.41	2 959 325.32	10.00	10.00	0
Part 1 19005	763 284.27	2 959 334.56	10.00	10.00	0
19006	763 295.65	2 959 318.27	10.00	10.00	0
A	763 289.44	2 959 326.05	--	--	--
Part 2 N	763 289.44	2 959 326.05	5.77	5.77	0
N-A	0.00	0.00	--	--	--
N-M	-6.21	7.78	-4.23	-4.23	--

3.2 非等精度结点情况

如图 4,在 ARC/NFO 叠置操作后所产生的新的数据层上,存在位于模糊公差范围内的 3 个结点 97063、97064 和 97065。由于该 3 点来自不同精度的数据层,其坐标的精度是不同的,数据信息列于表 6 上半部 (Part 1)。按类似于 3.1 中的步骤进行计算和分析,结果列于表 6 下半部 (Part 2)。其中,MOVE 命令所选择的相对可靠的点号为 97063。

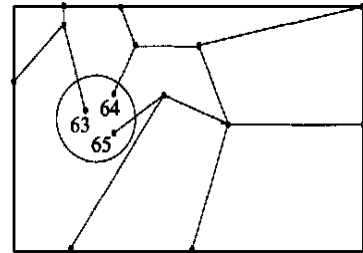


图 4 来自不同精度数据层的结点捕捉

Fig 4 Node snapping with three statistically independent vertices of varying accuracy

表 6 非等精度结点在选用不同捕捉算法下的结果

Tab 6 Different results from different snapping algorithms for points of varying accuracy

Point No.	x/m	y/m	σ_x/m	σ_y/m	ρ_{xy}
97063	768 570.60	2 935 308.40	20.00	20.00	0
Part 1 97064	768 515.80	2 935 367.90	40.00	40.00	0
97065	768 664.50	2 935 230.48	100.00	100.00	0
A	768 583.63	2 935 302.26	--	--	--
Part 2 N	768 562.89	2 935 317.52	17.61	17.61	0
N-A	-20.74	15.26	--	--	--
N-M	-7.71	9.12	-2.39	-2.39	--

从表 6 中可见,对于本例,广义算法与 ARC/NFO 中现有 ADJUST 命令所得坐标估值的结果间存在着较大差异, x 坐标相差近 21 m、 y 坐标相差 15 m;而与 MOVE 命令结果间的差异见表 6 中末行。

4 结论和建议

1. 结点捕捉后所得新点坐标分量间的相关性取决于捕捉点组中各点的自相关性和互相关性。当且仅当捕捉点组中各点均非自相关和非互相关时,捕捉后所得的新点才是非自相关的。捕捉点组的自相关性和互相关性一般将对捕捉后所得新点坐标分量的估值和精度有不同程度的影响,但对自相关系数的影响很小。

2. 现行 ARC/NFO 中结点捕捉的 ADJUST 算法只是本文广义算法的特例,既不能处理非等精度点组的结点捕捉,又不能为等精度捕捉点组的捕捉结果提供精度信息。而广义算法及其广义误差模型可以为等精度或非等精度、独立或相关点组的结点捕捉提供严密的理论结果及其精度信息。

3. 建议对 ARC/NFO 中结点捕捉的现有 ADJUST 算法进行改造,即由简单算术平均值算法改为式(2)所描述的广义算法,并增加式(3)的精度估计功能。后者正是现有 GIS 软件所缺少的,而在数据质量评价中又是必须考虑的一个最基本的问题。

参考文献:

[1] LONGLEY P A. Geographical Information Sys-

tems [M]. New York: John Wiley and Sons, 1999. 175-176

[2] BURROUGHPA, MCDONNELL R A. Principles of Geographical Information Systems [M]. Oxford: Oxford University Press, 1998. 1-220

[3] JONES C B. Geographical Information Systems and Computer Cartography [M]. Harlow: Longman, 1997. 87-92. 121-131.

[4] ESRI Arcedit Users Guide: Interactive Graphics Editor [Z]. California: Environmental Systems Research Institute, 1988

[5] ESRI ARC Command References: Commands A-I, J-Z [Z]. California: Environmental Systems Research Institute, 1992

[6] CLARKE K C. Analytical and Computer Cartography [M]. New Jersey: Englewood Cliffs, 1996. 198-203

[7] MIKHAL E M. Observations and Least Squares [M]. New York: A Dun-Donnelley Publisher, 1976

[8] KEEFER B J, SMITH J L, GREGORE T G. Modeling and Evaluating the Effects of Stream Mode Digitizing Errors on Map Variables [J]. Photogrammetric Engineering and Remote Sensing, 1991, 57(7): 957-963

[9] LU Wen-bao. Models of Data Quality and Dynamic Spatial Relations in GIS-T [R]. Nanjing: Southeast University, 1998. 1-68 (in Chinese).

[10] HORD R. Land-use Map Accuracy Criteria [J]. Photogrammetric Engineering and Remote Sensing, 1976, 42(5): 671-677.

启 事

欢迎订阅《测绘学报》《测绘通报》合订本。《测绘学报》1998~ 2000 年,每年 1 册(精装),定价 40 元。《测绘通报》1997~ 2000 年,每年上下两册(精装),定价 80 元。

需要购买者可通过邮局汇款(另加 20% 邮费)至北京复外三里河路 50 号中国地图出版社期刊室收。邮编: 100045, 联系电话: (010) 83524586 63534931-2184, 2185, 联系人: 金英。