DOI: 10.13203/j.whugis1996.04.009

第 21卷第 4期　　武汉测绘科技大学学报　　Vol. 21 No. 4
1996年 12月　　Journal of Wuhan Technical University of Surveying and Mapping　　Dec. 1996

GIS叠置前后同名点元的方差估计*

刘文宝　史文中
(山东矿业学院地球科学系,泰安市岱宗大街 223号, 271019)

摘　要　定义了同名坐标观测值矩阵概念,建立了多层叠置同名观测值的多元分析模型,导出了同名点元叠置前后位置方差的一般估计式。

关键词　叠置操作;同名点元;方差估计
分类号　P 207

为了对矢量 GIS多层叠置输入、输出图的位置精度进行评定,需要针对点、线、面三个基本单元研究叠置前后的方差估计方法。这是 GIS界迄今尚未解决的一簇理论问题[1]。本文只限于讨论一种特殊情况,即叠置前后同名点元的方差估计。

1　同名坐标观测值矩阵

设点坐标观测量序列 $\{(X_j, Y_j), j=1, 2, \cdots, n\}$在 GIS空间数据库中 m个叠置层上的 n个坐标观测值序列为 $\{(x_{ij}, y_{ij}), i=1, 2, \cdots, m; j=1, 2, \cdots, n\}$,写成矩阵形式为:

$$L=\begin{bmatrix} x_{11}y_{11} & x_{12}y_{12} & \cdots & x_{1n}y_{1n} \\ x_{21}y_{21} & x_{22}y_{22} & \cdots & x_{2n}y_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1}y_{m1} & x_{m2}y_{m2} & \cdots & x_{mn}y_{mn} \end{bmatrix}=(l_1\ l_2, \cdots, l_{2n}) \tag{1}$$

其中当 $j=1, 3, 5, \cdots, 2n-1$时,$l_j=(x_{1k}, x_{2k}, \cdots, x_{mk})^{\mathrm{T}}$;当 $j=2, 4, \cdots, 2n$时,$l_j=(y_{1k}, y_{2k}, \cdots, y_{mk})^{\mathrm{T}}$。显然,矩阵 L中的每一列为不同层上同名点元坐标的观测值。定义 L为同名坐标观测值矩阵,简称观测值矩阵。

2　两层叠置前后同名点元的方差估计

在 (1)式中,当 $m=2$时,即为两层叠置时的观测值矩阵:

$$L=\begin{bmatrix} x_{11}y_{11} & x_{12}y_{12} & \cdots & x_{1n}y_{1n} \\ x_{21}y_{21} & x_{22}y_{22} & \cdots & x_{2n}y_{2n} \end{bmatrix} \tag{2}$$

假定 (2)式中所有的坐标观测值间均是独立的,但精度不同。不妨设各坐标观测值的权阵为:

$$P=\begin{bmatrix} p_{x_{11}}p_{y_{11}} & p_{x_{12}}p_{y_{12}} & \cdots & p_{x_{1n}}p_{y_{1n}} \\ p_{x_{21}}p_{y_{21}} & p_{x_{22}}p_{y_{22}} & \cdots & p_{x_{2n}}p_{y_{2n}} \end{bmatrix} \tag{3}$$

由 (2)式知,各同名点元坐标观测值之差分别为:

收稿日期: 1996-06-07.刘文宝,男, 32岁,博士,现从事 GIS数据质量研究。

$$\Delta x_j = x_{2j} - x_{1j},\quad \Delta y_j = y_{2j} - y_{1j}\quad (j = 1, 2, \cdots, n) \tag{4}$$

显然,Δx_j和Δy_j就是各同名点元坐标观测值之差数的真误差。因为

$$\begin{cases} \Delta x_j = x_{2j} - x_{1j} = (x_{2j} - x_{1j}) - (X_{2j} - X_{1j}) \\ \Delta y_j = y_{2j} - y_{1j} = (y_{2j} - y_{1j}) - (Y_{2j} - Y_{1j}) \end{cases} \tag{5}$$

而 $(X_{2j} - X_{1j})(Y_{2j} - Y_{1j})$ 恒为零。对 (4) 式施加权倒数传播律[2],可得Δx_j和Δy_j的权倒数分别为:

$$1/p_{\Delta x_j} = 1/p_{x_{2j}} + 1/p_{x_{1j}},\quad 1/p_{\Delta y_j} = 1/p_{y_{2j}} + 1/p_{y_{1j}} \tag{6}$$

从而有:

$$p_{\Delta x_j} = p_{x_{1j}} p_{x_{2j}} / (p_{x_{1j}} + p_{x_{2j}}),\quad p_{\Delta y_j} = p_{y_{1j}} p_{y_{2j}} / (p_{y_{1j}} + p_{y_{2j}}) \tag{7}$$

于是,根据方差的定义,则由真误差计算的方差因子估值为:

$$\hat{\sigma}_0^2 = \sum_{j=1}^{n} (p_{\Delta x_j} \Delta x_j^2 + p_{\Delta y_j} \Delta y_j^2) / 2n \tag{8}$$

利用 $\hat{\sigma}_0^2$可求得不同叠置层上各同名点元坐标观测值 (叠置操作输入值) 的方差:

$$\hat{\sigma}_{x_{ij}}^2 = \hat{\sigma}_0^2 p_{x_{ij}}^{-1},\quad \hat{\sigma}_{y_{ij}}^2 = \hat{\sigma}_0^2 p_{y_{ij}}^{-1} \tag{9}$$

而叠置输出图上第 j 个点的坐标值及其方差分别为:

$$\hat{x}_j = \Big(\sum_{i=1}^{2} p_{x_{ij}} x_{ij}\Big) \Big/ \sum_{i=1}^{2} p_{x_{ij}},\quad \hat{y}_j = \Big(\sum_{i=1}^{2} p_{y_{ij}} y_{ij}\Big) \Big/ \sum_{i=1}^{2} p_{y_{ij}} \tag{10}$$

$$\hat{\sigma}_{x_j}^2 = \hat{\sigma}_0^2 \Big/ \sum_{i=1}^{2} p_{x_{ij}},\quad \hat{\sigma}_{y_j}^2 = \hat{\sigma}_0^2 \Big/ \sum_{i=1}^{2} p_{y_{ij}} \tag{11}$$

如果各叠置层上的坐标观测值精度相同,不妨令第一层的权为 p_1,第二层的权为 p_2,则 (7)、(8) 两式变为:

$$p_{\Delta} = p_{\Delta x_j} = p_{\Delta y_j} = p_1 p_2 / (p_1 + p_2) \tag{12}$$

$$\hat{\sigma}_0^2 = [p_{\Delta} / (2n)] \sum_{j=1}^{n} d_j^2 \tag{13}$$

其中 $d_j^2 = \Delta x_j^2 + \Delta y_j^2$为同名坐标观测值点间的欧氏距离。于是,(9)、(10) 和 (11) 式分别变为:

$$\hat{\sigma}_{x_{ij}}^2 = \hat{\sigma}_{y_{ij}}^2 = [p_{\Delta} / (2n p_i)] \sum_{j=1}^{n} d_j^2 \tag{14}$$

$$\hat{x}_j = \Big(\sum_{i=1}^{2} p_i x_{ij}\Big) \Big/ \sum_{i=1}^{2} p_i;\quad \hat{y}_j = \Big(\sum_{i=1}^{2} p_i y_{ij}\Big) \Big/ \sum_{i=1}^{2} p_i \tag{15}$$

$$\hat{\sigma}_{x_j}^2 = \hat{\sigma}_{y_j}^2 = p_{\Delta} \sum_{j=1}^{n} d_j^2 \Big/ \Big(2n \sum_{i=1}^{2} p_i\Big) \tag{16}$$

如果两叠置层上的所有坐标观测值精度相同,令其权为 1,则 (13)、(14)、(15) 和 (16) 式分别变为:

$$\hat{\sigma}_0^2 = (1/(4n)) \sum_{j=1}^{n} d_j^2 \tag{17}$$

$$\hat{\sigma}_{x_{ij}}^2 = \hat{\sigma}_{y_{ij}}^2 = (1/(4n)) \sum_{j=1}^{n} d_j^2 \tag{18}$$

$$\hat{x}_j = (1/2) \sum_{i=1}^{2} x_{ij},\quad \hat{y}_j = (1/2) \sum_{i=1}^{2} y_{ij} \tag{19}$$

$$\hat{\sigma}_{x_j}^2 = \hat{\sigma}_{y_j}^2 = (1/(8n)) \sum_{j=1}^{n} d_j^2 \tag{20}$$

3 多层叠置前后同名点元的方差估计

为了利用同名坐标观测值矩阵推求同名点元叠置前后的方差,首先构成 GIS 多层叠置同

名坐标观测值的多元分析模型。式 (1) L 中第 j 列的 Gauss-Markov 模型为:

$$E\{l_j\} = AZ_j,\ D\{l_j\} = \sigma_j^2 P^{-1} \tag{21}$$

其中 Z_j为同名点坐标未知参数向量。当 $j = 1, 3, \cdots, 2n-1$ 时, $Z_j = X_k (k = 1, 2, \cdots, n)$;当 $j = 2, 4, \cdots, 2n$ 时, $Z_j = Y_k (k = 1, 2, \cdots, n)$。$A_{m\times 1} = (1\ 1 \cdots\ 1)^{\mathrm{T}}$, P 为 l_j的观测权阵。顾及各个同名点元坐标观测值间的相关性,设协方差阵为:

$$\mathrm{cov}\{l_i, l_j\} = \sigma_{ij}^2 P^{-1} \quad (i, j = 1, 2, \cdots, 2n) \tag{22}$$

令 Z 为 $1\times 2n$ 维的未知参数行向量,即 $Z = (Z_1, Z_2, \cdots, Z_{2n})$, Γ_0为广义单位权协方差阵:

$$\Gamma_0 = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{12n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{22n} \\ \cdots\cdots & & & \\ \sigma_{2n1} & \sigma_{2n2} & \cdots & \sigma_{2n}^2 \end{bmatrix} \tag{23}$$

则观测矩阵 L 的多元线性 Gauss-Markov 模型 (21) 可扩展为:

$$E\{L\} = AZ; \quad D\{L\} = \Gamma_0 \otimes P^{-1} \tag{24}$$

其中 $\Gamma_0 \otimes P^{-1}$应理解为向量 $\mathrm{vec}\{L\}$ (矩阵 L 的拉直变换) 的协方差阵,$\otimes$ 为矩阵 Kronecker 乘积。显然,式 (24) 就是 GIS 多层叠置同名观测值的多元分析模型。于是,问题变为: 已知 (24) 式中 A 和 P,求 Z 及 Z 和 L 的协方差阵。为此,首先导出 Z 和 Γ_0的估计公式,基本思想是利用矩阵拉直运算 vec 将多元分析模型 (24) 化为广义线性模型。

记 $l = \mathrm{vec}\{L\}$ 为 $2mn\times 1$维向量, $z = Z^{\mathrm{T}}$为 $2n\times 1$维向量,则 (24) 式可化为下列广义线性模型[3]:

$$E\{l\} = (I_{2n} \otimes A)z, \quad D\{l\} = \Gamma_0 \otimes P^{-1} \tag{25}$$

对模型 (25) 施加最小二乘准则,可得同名点坐标在叠置输出图上的最优无偏估计为:

$$\hat{z} = [I_{2n} \otimes (A^{\mathrm{T}}PA)^{-1}A^{\mathrm{T}}P]l \tag{26}$$

写成矩阵形式为:

$$\hat{Z} = (A^{\mathrm{T}}PA)^{-1}A^{\mathrm{T}}PA \tag{27}$$

从 (27) 式可见, Z 的最优无偏估计与 Γ_0无关。换言之,即使不知道各个同名坐标观测值的方差因子 σ_i^2和协方差因子 σ_{ij},也能估计输出图上的坐标值,并且结果不受任何影响。这说明,各个同名点元的坐标观测值可以单独平差,而不必顾及它们间的相关性。因此,第 j 个同名点在输出图上的坐标估值 $\hat{Z}_j$及其协因数分别为:

$$\hat{Z}_j = \begin{cases} (\sum_{i=1}^{m} p_i)^{-1} \sum_{i=1}^{m} (p_i x_{ik}) & \text{当 } j = 1, 3, \cdots, 2n-1 \text{ 时} \\ (\sum_{i=1}^{m} p_i)^{-1} \sum_{i=1}^{m} (p_i y_{ik}) & \text{当 } j = 2, 4, \cdots, 2n \text{ 时} \end{cases} \tag{28}$$

$$Q_{\hat{Z}_j\hat{Z}_j} = (\sum_{i=1}^{m} p_i)^{-1} \tag{29}$$

而 $\hat{Z}$ 的协方差阵为:

$$D(\hat{Z}) = \hat{\Gamma}_0 \otimes (A^{\mathrm{T}}PA)^{-1} \tag{30}$$

从 (24) 和 (30) 式可见,欲求 $D(\hat{Z})$ 和 $D(L)$,应先求出 $\hat{\Gamma}_0$,这与 $\hat{Z}$ 的估计是不同的。为此,令 $V = L - A\hat{Z}$ 及 $\hat{L} = L + V$。易证[4]:

$$\hat{\Gamma}_0 = V^{\mathrm{T}}PV/(m-1) = (L - A\hat{Z})^{\mathrm{T}}P(L - A\hat{Z})/(m-1) \tag{31}$$

是 Γ_0的无偏估计。写成分量形式,则单位权方差和协方差分别为:

$$\hat{\sigma}_j^2 = (1/(m-1))\sum_{i=1}^{m} p_i(\hat{Z}_j - l_{ij})^2 \tag{32}$$

$$\hat{\sigma}_{jk} = (1/(m-1))\sum_{i=1}^{m} p_i(\hat{Z}_j - l_{ij})(\hat{Z}_k - l_{ik}) \tag{33}$$

于是,第 j 个同名点在输出图上的坐标 $\hat{Z}_j$的方差为:

$$\hat{\sigma}_{Z_j}^2 = \hat{\sigma}_j^2\left(\sum_{i=1}^{m} p_i\right)^{-1} \tag{34}$$

而同名点坐标观测值的方差阵为:

$$D\{l\} = \hat{\sigma}_j^2 P^{-1} \quad (j = 1, 2, \cdots, 2n) \tag{35}$$

写成分量形式,则得同名点坐标观测值的方差。其中,当 $j = 1, 3, \cdots, 2n-1$时,有:

$$\hat{\sigma}_{x_{ij}}^2 = \hat{\sigma}_j^2 p_i^{-1} \quad (i = 1, 2, \cdots, m) \tag{36}$$

而当 $j = 2, 4, \cdots, 2n$ 时,有:

$$\hat{\sigma}_{y_{ij}}^2 = \hat{\sigma}_j^2 p_i^{-1} \quad (i = 1, 2, \cdots, m) \tag{37}$$

各同名点元坐标观测值间的协方差阵为:

$$\mathrm{cov}\{l_i, l_j\} = \hat{\sigma}_{ij} P^{-1} \quad (i, j = 1, 2, \cdots, 2n) \tag{38}$$

特别地,当各叠置层上的所有坐标观测值精度相同时,式(28)、(32)、(33)、(34)、(36)、(37)和(38)分别简化为:

$$\hat{Z}_j = \begin{cases} (1/m)\sum_{i=1}^{m} x_{ik} & \text{当 } j = 1, 3, \cdots, 2n-1 \text{时} \\ (1/m)\sum_{i=1}^{m} y_{ik} & \text{当 } j = 2, 4, \cdots, 2n \text{时} \end{cases} \tag{39}$$

$$\hat{\sigma}_j^2 = \begin{cases} (1/(m-1))\sum_{i=1}^{m}(\hat{Z}_j - x_{ik})^2 & \text{当 } j = 1, 3, \cdots, 2n-1 \text{时} \\ (1/(m-1))\sum_{i=1}^{m}(\hat{Z}_j - y_{ik})^2 & \text{当 } j = 2, 4, \cdots, 2n \text{时} \end{cases} \tag{40}$$

$$\hat{\sigma}_{jk} = (1/(m-1))\sum_{i=1}^{m}(\hat{Z}_j - l_{ij})(\hat{Z}_k - l_{ik}) \tag{41}$$

$$\hat{\sigma}_{Z_j}^2 = \hat{\sigma}_j^2/m, \quad \hat{\sigma}_{x_{ij}}^2 = \hat{\sigma}_{y_{ij}}^2 = \hat{\sigma}_j^2, \quad \mathrm{cov}\{l_i, l_j\} = \hat{\sigma}_{ij} \tag{42}$$

4 算例分析

4.1 两层叠置情况

叠置层Ⅰ、Ⅱ上同名多边形 A 的顶点坐标值分别来自 1∶1 000和 1∶5 000的两幅地图的手工数字化,数据列于表 1。现求多边形顶点坐标:① 在两个叠置层上的数字化数据方差;② 在叠置输出图上的顶点坐标方差。

模拟数字化数据的理论方差(实地) $\sigma_{\mathrm{I}}^2 = 0.25\ \mathrm{m}^2$, $\sigma_{\mathrm{II}}^2 = 1.25\ \mathrm{m}^2$,定权公式为:

$$p_i = \sigma_0^2/M_i \quad (i = \mathrm{I}, \mathrm{II}) \tag{43}$$

其中 M_i为数字化原图比例尺分母,取 $\sigma_0^2 = M_{\mathrm{I}}$,则 $p_{\mathrm{I}} = 1$, $p_{\mathrm{II}} = 1/5$,于是,由(12)式计算得 $p_{\Delta} = 1/6$,而 $\sum_{j=1}^{6} d_j^2 = 18.074$。由(13)式得 $\hat{\sigma}_0^2 = 0.251\ \mathrm{m}^2$,由(14)式分别得 $\hat{\sigma}_{x_{\mathrm{I}j}}^2 = \hat{\sigma}_{y_{\mathrm{I}j}}^2 = 0.251\ \mathrm{m}^2$, $\hat{\sigma}_{x_{\mathrm{II}j}}^2 = \hat{\sigma}_{y_{\mathrm{II}j}}^2 = 1.255\ \mathrm{m}^2$。可见,由(13)和(14)式所估计的方差值与理论值一致,估计结果

是可靠的。

表 1　多边形顶点坐标数据 (/m)

点号	理论值		I		II	
	x	y	x	y	x	y
1	1 000$\sqrt{3}$	500	1 732. 555	500. 500	1 730. 937	498. 873
2	1 500$\sqrt{3}$	1 000	2 597. 579	999. 485	2 596. 958	1 001. 101
3	1 500$\sqrt{3}$	2 000	2 598. 587	2 000. 512	2 599. 224	2 001. 154
4	1 000$\sqrt{3}$	2 500	1 731. 558	2 499. 513	1 733. 178	2 498. 886
5	500$\sqrt{3}$	2 000	866. 512	2 000. 507	865. 893	1 998. 855
6	500$\sqrt{3}$	1 000	865. 515	999. 508	867. 130	998. 885

由 (16) 式得 $\hat{\sigma}^2_{x_j} = \hat{\sigma}^2_{y_j} = 0.209\ \mathrm{m}^2$。这提醒我们,当两层叠置图上存在同名点时,可以利用 (16) 式来部分地解决 Goodchild& Gopal提出的难题[1],即“当 1∶24 000和 1∶100 000的两幅地形图数字化后,进行两层数据的叠置操作,则输出图的比例尺和精度各是多少?”。对输出图的比例尺的大小,人们的选择可以不同,但由 (16) 式估计的实地精度却是确定不变的。

4. 2　多层叠置情况

在 1∶500 1∶1 000 1∶2 000和 1∶5 000四幅地形图上的一条同名边界线的折点为同名点,数字化坐标数据列于表 2 仍采用 (43) 式定权,取 $\hat{\sigma}^2_0 = M_{II}$,则有 $p_{I} = 2, p_{II} = 1, p_{III} = 1/2, p_{IV} = 1/5$ 根据 (32), (36) 和 (37) 式估计的各层数据方差值列于表 2的最后一行。

表 2　同名边界线折点坐标的数字化数据 (/m)

点号	I		II		III		IV	
	x	y	x	y	x	y	x	y
1	1 150. 367	1 999. 633	1 150. 515	1 999. 476	1 150. 735	1 999. 265	1 151. 162	1 999. 838
2	1 149. 635	2 000. 358	1 149. 489	2 000. 517	1 149. 265	2 000. 734	1 148. 838	2 001. 162
3	1 200. 358	2 049. 642	1 200. 524	2 049. 485	1 200. 726	2 049. 271	1 201. 157	2 049. 842
4	1 199. 644	2 050. 365	1 199. 487	2 050. 521	1 199. 272	2 050. 728	1 198. 843	2 051. 158
5	1 250. 371	2 099. 864	1 250. 519	2 099. 472	1 250. 728	2 099. 261	1 251. 149	2 098. 835
6	1 249. 628	2 100. 369	1 249. 477	2 100. 513	1 249. 275	2 100. 729	1 249. 851	2 101. 160
方差估值	0. 135		0. 270		0. 540		1. 350	

建议在将 GIS源数据录入空间数据库中时,勿将不同来源、不同尺度、不同分辨率的同名坐标数据简单地附合为某一类型的数据,应把精度指标作为属性一并存入空间数据库中。

参　考　文　献

1　Goodchild M F, Gopal S. The Accuracy of Spatial Databases. New York: Taylor and Francis, 1989.
2　於宗俦,鲁林成.测量平差基础.北京: 测绘出版社, 1983.
3　张尧庭,方开泰.多元统计分析引论.北京: 科学出版社, 1983.
4　刘文宝. GIS空间数据库的不确定性理论: [学位论文].武汉: 武汉测绘科学大学, 1995
5　黄幼才,刘文宝,李宗华,等. GIS空间数据误差分析和处理.武汉: 中国地质大学出版社, 1995.

The Variance Estimation of Corresponding Points in GIS Overlay Operation

Liu Wenbao Shi Wenzhong
(Dept. of Geo-science, Shandong Mining Institute, Daizong Street 223, Tai' an, China, 271019)

Abstract This paper introduces observation matrix of corresponding points coordinates, establishes multivariate analysis model of corresponding points observation in overlay opertions, derives a general formula to estimate the variance of corresponding points before and after overlay operation.

Key words overlay operation; corresponding points; variance estimation

(上接第 416页)

研究方向上有所突破,达到国内先进水平,部分研究方向达到世界先进水平。

计算机科学与工程

本学科的发展目标是: 在“九五”期间建设成省部级重点学科,在空间数据库理论和多媒体数据库技术、图像模式识别与计算机视觉、计算机网络和多媒体技术、智能接口、智能控制系统等方面,创造一批达到国际先进水平的成果;建设空间数据库实验室、图像模式识别实验室。

印刷工程

本学科的发展目标是: 利用现有学科优势,应用新理论、新技术和新方法,两年内使本学科达到省部级重点学科,四年内争取获得印刷工程博士学位授予权。

发展彩色印前系统,实现电子分色机的高端联网,在彩色图像处理与制版方面达到国内领先水平。在印刷数据源、规范化研究方向达到国际先进水平,在电子出版物的设计与制作方面达到国内领先水平。 改造印刷材料和印刷产品测试实验室,使之在国内开放,并成为有权威的印刷产品质量检测和评价中心。

电子信息工程

本学科的发展目标是: 在“九五”期间将进一步把电子信息工程学科建设成为省部级重点学科;将“信号与信息处理实验室”建成为省部级重点实验室,并争取获得“信号与信息处理”专业研究生博士学位授予权。

土地管理

本学科的发展目标是: 依托测绘高技术与 GIS技术,加强土地信息的采集、处理、存储、传递、应用的研究,在信息采集、数据处理自动化等方面进一步形成优势,保持国内领先水平。 申报“土地规划与利用”、“房地产管理”本科专业,争取使土地学学科成为国家级学科并成为省部级重点学科。

城市规划

本学科的发展目标是: 走依托优势、突出重点的道路,进一步完善学科和现有层次教育,“九五”期末把本学科办成省部级重点学科,下世纪初办成博士点。

把“城市信息系统研究室”进一步扩展成为城市规划新技术研究所,完成一批国内领先的科研项目,将“城市综合实验室”建成为省部级重点实验室。