非线性模型的误差传播及其在 GIS 中的应用*

胡圣武　陶本藻

(武汉测绘科技大学地学测量工程学院,武汉市珞喻路 39 号,430070)

摘　要　本文给出非线性模型的误差传播公式,并应用于 GIS 空间数据误差分析和处理。

关键词　非线性模型;误差传播;GIS 空间数据分析

分类号　P 207;P 208

在测量的数学模型中,大多函数关系属非线性型。由于常规测量遇到的一般为微弱非线性型,所以仍可通过线性化处理来满足其精度要求。随着 GPS、RS、GIS 技术的应用和发展,仅用线性化来处理非线性模型有时就不能满足要求了,研究和发展非线性模型的理论和方法作为迫切需要而提到日程上来了。

GIS 原始数据本身带有误差,在空间数据库中进行各种操作、转换和处理也将引入误差。与一般测量数据类似,根据误差对结果影响的性质,GIS 数据误差亦可分为随机误差、系统误差和粗差,此外还有自然界随时间的变化或随空间的变迁而引起的变差。变差实际上是动态模型误差。所以,GIS 的误差理论和误差分析方法与一般测量数据的处理是一致的,只是 GIS 的数学模型种类繁多而且有的比较特殊,应用已有理论不是那么直接和方便,相应地还要发展结合 GIS 情况的误差理论。GIS 空间数据误差分析和处理中出现的许多数学模型,也大多是非线性的,其参数估计和误差传播亦应从非线性模型出发研究。

本文仅讨论非线性模型的误差传播并举例说明其在 GIS 中的应用。

1　非线性模型误差传播

设有观测向量 $L=(L_1\ \ L_2\ \ \cdots\ \ L_n)^{\mathrm{T}}$,其协方差为 $D_L=E[(L-\tilde{L})(L-\tilde{L})^{\mathrm{T}}]$,式中 $\tilde{L}=E(L)$,对应的误差向量为:

$$\Delta = L - E(L) = L - \tilde{L}$$

且有 $E(\Delta)=0, D(\Delta)=D_L$。若 u 维向量 $X=(X_1\ \ X_2\ \ \cdots\ \ X_u)^{\mathrm{T}}$ 是观测值 L 的函数:

$$X = \varphi(L_1\ \ L_2\ \ \cdots\ \ L_n) = \varphi(L) \qquad (1)$$

其中,

$$\varphi(L) = [\varphi_1(L)\ \ \varphi_2(L)\ \ \cdots\ \ \varphi_u(L)]^{\mathrm{T}}$$

$\varphi(L)$ 为非线性函数,则由 D_L 求 D_X,即为非线性模型 (1) 的误差传播问题。

非线性模型误差传播的研究始于 80 年代初,Wolf 提出了顾及函数二次项的误差传播公式[1],陶本藻对此公式作了推证 (1985)。徐培亮给出了非线性函数的协方差传播公式[2],完善了这一问题的理论研究。Teunissen、Knickmeyer 给出了非线性最小二乘法估计量的方差计算式[3],等等。由于当时在测量中所处理的大都是微弱非线性模型,此类成果并未引起测绘界的注意和应用。

非线性模型的误差传播在现代测量数据处理中有着广泛应用。为此,本文先简略介绍已有成果,在此基础上,推荐一个实用公式,以便应用。

由 (1) 式出发,根据方差定义:

$$D_X = E[(X - EX)(X - EX)^{\mathrm{T}}]$$

可得[2]:

$$D_X = \Phi^{\mathrm{T}} D_L \Phi - [\mathrm{tr}(\Psi D_L)]^2/4 + E[(\Delta^{\mathrm{T}}\Psi\Delta)^2]/4 + E(\Phi^{\mathrm{T}}\Delta\Delta^{\mathrm{T}}\Psi\Delta) + R \qquad (2)$$

式中,

$$\Phi^{\mathrm{T}} = \left[\left(\frac{\partial\varphi}{\partial L_1}\right)_E\ \ \left(\frac{\partial\varphi}{\partial L_2}\right)_E\ \ \cdots\ \ \left(\frac{\partial\varphi}{\partial L_n}\right)_E\right]$$

$$\Psi = \begin{bmatrix} \left(\frac{\partial^2\varphi}{\partial L_1^2}\right)_E & \left(\frac{\partial^2\varphi}{\partial L_1\partial L_2}\right)_E & \cdots & \left(\frac{\partial^2\varphi}{\partial L_1\partial L_n}\right)_E \\ \left(\frac{\partial^2\varphi}{\partial L_2\partial L_1}\right)_E & \left(\frac{\partial^2\varphi}{\partial L_2^2}\right)_E & \cdots & \left(\frac{\partial^2\varphi}{\partial L_2\partial L_n}\right)_E \\ \cdots\cdots & & & \\ \left(\frac{\partial^2\varphi}{\partial L_n\partial L_1}\right)_E & \left(\frac{\partial^2\varphi}{\partial L_n\partial L_2}\right)_E & \cdots & \left(\frac{\partial^2\varphi}{\partial L_n^2}\right)_E \end{bmatrix}$$

收稿日期:1997-01-09。胡圣武,男,26 岁,硕士生,现从事数据处理研究。

* 国家自然科学基金资助项目,编号 49474204

$$R = E[\alpha^{T}\Delta(R_1 - E(R_1)) + (\Delta^{T}U\Delta - \mathrm{tr}(UD_L))(R_1 - E(R_1))] + E[(R_1 - E(R_1))^2]$$

$$R_1 = (1/3)[(\partial^3 h/\partial L_1^3)_E\Delta_1^3 + (\partial^3 h/(\partial L_1^2\partial L_2))_E\Delta_1^2\Delta_2 + \cdots] + \cdots$$

公式(2)由徐培亮1986年导出,是不考虑 L 分布的、理论上严密的协方差传播公式。

如果函数是线性的,就取(2)式的第一项,它就是线性模型的误差传播式。考虑函数的二次项或更高次项,就要从(2)式出发推导相应项次的公式,因为直接应用(2)式是困难的。

最实用的非线性模型,对误差传播而言,取至二次项一般已足够了。在文献[2]中,由(2)式已得出考虑二次项的实用公式了,该文还指出,Wolf导出的类似公式是近似的[1]。

以下我们从另一途径,经过简单的推导,也可得出顾及二次项的非线性模型误差传播公式。

针对测量数据具体情况,观测值 L 服从正态分布,则有:

$$E(\Delta) = 0,\quad E(\Delta^3) = 0$$
$$E(\Delta^2) = D_L,\quad D(\Delta^4) = 3\sigma^4 \tag{3}$$

再假定各观测值间相互独立,则

$$E(\Delta_i\Delta_j) = 0\quad (i \neq j) \tag{4}$$

非线性函数(1)取至二次项,其误差关系式为:

$$\mathrm{d}X = \alpha_1\Delta_1 + \alpha_2\Delta_2 + \cdots + \alpha_n\Delta_n + (\alpha_{11}\Delta_1^2 + \alpha_{22}\Delta_2^2 + \cdots + \alpha_{nn}\Delta_n^2)/2 + \alpha_{12}\Delta_1\Delta_2 + \alpha_{13}\Delta_1\Delta_3 + \cdots + \alpha_{n-1,n}\Delta_{n-1}\Delta_n \tag{5}$$

式中,

$$\alpha_i = \partial h/\partial L_i,\quad \alpha_{ij} = \partial^2 h/\partial L_i\partial L_j$$

对(5)式取数学期望,并顾及(3)、(4)式,得:

$$E(\mathrm{d}X) = [\alpha_{11}\sigma_1^2 + \alpha_{22}\sigma_2^2 + \cdots + \alpha_{nn}\sigma_n^2]/2 \tag{6}$$

式中,$\sigma_i^2 = E(\Delta_i^2)$,由此即得:

$$D_X = E[\mathrm{d}X - E(\mathrm{d}X)]^2 = E[\alpha_1\Delta_1 + \alpha_2\Delta_2 + \cdots + \alpha_n\Delta_n + \alpha_{11}(\Delta_1^2 - \sigma_1^2)/2 + \cdots + \alpha_{nn}(\Delta_n^2 - \sigma_n^2)/2 + \alpha_{12}\Delta_1\Delta_2 + \cdots + \alpha_{n-1,n}\Delta_{n-1}\Delta_n] \tag{7}$$

将上式展开,并顾及(3)、(4)式,最后可得:

$$D_X = \sum_{i=1}^{n}\alpha_i^2\sigma_i^2 + (1/2)\sum_{i=1}^{n}\alpha_{ii}^2\sigma_i^4 + \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{ij}^2\sigma_i^2\sigma_j^2\quad (i \neq j) \tag{8}$$

这就是实用计算公式。

2 非线性误差传播在GIS中的应用

由于GIS原始数据存在着误差,导致经变换、处理后生成的产品也存在误差,这是误差传播问题,其函数关系许多是非线性型的。

例如,在利用GIS研究土壤侵蚀情况时,其中由降雨引起的水土流失量FAO(Burrough,1986)公式为:

$$R = 0.011abc + 66 \tag{9}$$

式中,a 为年平均降雨量;b 为日最大降雨量(两年出现一次);c 为两年中的最大暴雨降水量之和。

空间数据库中基本地理信息单元是点线面。线元是由端点的两对坐标值来确定的,长度 l 和方位 α 是线元的两个属性数据,计算式为:

$$l_{12} = [(x_2 - x_1)^2 + (y_2 - y_1)^2]^{1/2} \tag{10}$$

$$\alpha_{12} = \arctan[(y_2 - y_1)/(x_2 - x_1)] \tag{11}$$

多边形的面积也是一个常用的属性指标,面积计算式为:

$$G = (1/2)\left|\sum_{i=1}^{n}(y_{i+1} - y_{i-1})x_i\right| \tag{12}$$

诸如GIS中矢量数据栅格化转换,地图叠置操作中的函数模型,数据因回归分析的建模,内插与推估、坐标转换与统一等等均会出现非线性函数关系。运用上一节给出的误差传播实用公式,可以精确地确定其成果的精度。

下面举两个简单例子,说明公式的应用。

例1 由空间中点 $A(3,5,4)$ 和点 $B(2,3,2)$ 拟合一条直线:

$$l = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2} \tag{13}$$

已知位置方差:

$$D_A = \begin{pmatrix} 1/2 & & \\ & 1/2 & \\ & & 1 \end{pmatrix},\quad D_B = \begin{pmatrix} 1/4 & & \\ & 1/4 & \\ & & 1/2 \end{pmatrix}$$

计算该直线段的线长精度。

解:l 的近似值为:

$$l_0 = \sqrt{(2-3)^2 + (3-5)^2 + (2-4)^2} = 3$$

求导数:

$$\partial l/\partial x_A = -(x_B - x_A)/l_0 = 1/3,$$
$$\partial l/\partial x_B = (x_B - x_A)/l_0 = -1/3,$$
$$\partial l/\partial y_A = 2/3,\quad \partial l/\partial y_B = -2/3,$$
$$\partial l/\partial z_A = 2/3,\quad \partial l/\partial z_B = -2/3;$$
$$\partial^2 l/\partial x_A^2 = 8/27,\quad \partial^2 l/\partial x_B^2 = 8/27,$$
$$\partial^2 l/\partial y_A^2 = 5/27,\quad \partial^2 l/\partial y_B^2 = 5/27,$$
$$\partial^2 l/\partial z_A^2 = 5/27,\quad \partial^2 l/\partial z_B^2 = 5/27;$$
$$\partial^2 l/\partial x_A\partial x_B = -8/27,\quad \partial^2 l/\partial x_A\partial y_A = -2/27,$$
$$\partial^2 l/\partial x_A\partial y_B = 2/27,\quad \partial^2 l/\partial x_A\partial z_A = -2/27,$$

$$\partial^2 l/\partial x_A\partial z_B = 2/27,\quad \partial^2 l/\partial x_B\partial y_A = 2/27,$$
$$\partial^2 l/\partial x_B\partial y_B = -2/27,\quad \partial^2 l/\partial x_B\partial z_A = 2/27,$$
$$\partial^2 l/\partial x_B\partial z_B = -2/27,\quad \partial^2 l/\partial y_A\partial y_B = -5/27,$$
$$\partial^2 l/\partial y_A\partial z_A = -4/27,\quad \partial^2 l/\partial y_A\partial z_B = 4/27,$$
$$\partial^2 l/\partial y_B\partial z_A = 4/27,\quad \partial^2 l/\partial y_B\partial z_B = -4/27,$$
$$\partial^2 l/\partial z_A\partial z_B = -5/27$$

按 (5) 式,函数 (13) 的误差关系式为:

$$\begin{aligned} dl = {} & (1/3)(dx_A - dx_B + 2dy_A - 2dy_B \\ & + 2dz_A - 2dz_B) + (8dx_A^2 + 8dx_B^2 \\ & + 5dy_A^2 + 5dy_B^2 + 5dz_A^2 + 5dz_B^3)/54 \\ & + (-8dx_A dx_B - 2dx_A dy_A \\ & + 2dx_A dy_B - 2dx_A dz_A + 2dx_A dz_B \\ & + 2dx_B dy_A - 2dx_B dy_B + 2dx_B dz_A \\ & - 2dx_B dz_B - 5dy_A dy_B - 4dy_A dz_A \\ & + 4dy_A dz_B + 4dy_B dz_A - 4dy_B dz_B \\ & - 5dz_A dz_B)/27 \end{aligned} \tag{14}$$

代入 (8) 式,即得长度 l 的方差为 $D_l = 1.19$; l 的中误差为 $\sigma_l = D_l^{1/2} = 1.09$ 如果只考虑一次项,则其方差为 1.08

例 2　已知 DTM 上三点高程为 $h_A = 1$, $h_B = 4$, $h_C = 9$,通过内插求 D 点高程精度。

已知点的方差为 $m_A^2 = 1$, $m_B^2 = 1/4$, $m_C^2 = 1/9$ 高程内插公式为:

$$h_D = (\sqrt{h_A} + \sqrt{h_B} + \sqrt{h_C})^2/9 \tag{15}$$

计算内插点高程 $h_D = 4$ 其导数值为:

$$\partial h_D/\partial h_A = 2/3,\quad \partial h_D/\partial h_B = 1/3,$$
$$\partial h_D/\partial h_C = 2/9;\quad \partial^2 h_D/\partial h_A^2 = -5/18,$$
$$\partial^2 h_D/\partial h_B^2 = -1/36,\quad \partial^2 h_D/\partial h_C^2 = 1/162;$$
$$\partial^2 h_D/\partial h_A\partial h_B = 1/36,\quad \partial^2 h_D/\partial h_A\partial h_C = 1/54,$$
$$\partial^2 h_D/\partial h_B\partial h_C = 1/108$$

(15) 式的误差关系式为:

$$\begin{aligned} dh_D = {} & (6dh_A + 3dh_B + 2dh_C)/9 \\ & - (90dh_A^2 + 9dh_B^2 + 2dh_C^2)/648 \\ & + (3dh_A dh_B + 2dh_A dh_C + dh_B dh_C)/108 \end{aligned} \tag{16}$$

算得内插点 D 的高程方差 $D_D = 0.52$,中误差 $\sigma_D = 0.72$ 若按线性误差传播,其方差和中误差分别为 0.47 和 0.68

3　结　果

本文推荐的非线性模型误差传播实用公式,可应用于测量和 GIS 采集的数据分析和处理。顾及二次项对于 GIS 误差传播已经足够了。将现有的误差理论考虑 GIS 数据的具体情况进行扩展和研究必将对 GIS 产品质量和控制起到一定作用。

参　考　文　献

1　Wolf H. Das Fehler for Tpflfanzung Sgestez Mit II Ordnung. ZfV, 1961.

2　徐培亮.非线性函数的协方差传播公式.武汉测绘科技大学学报,1986(2)

3　Teunissen P J G, Knickmeyer E H. 非线性和最小二乘法. CISM Journal ACSGC, 1988, 42(4) 武测译文,1990(1)

4　边馥苓.地理信息系统原理和方法.北京: 测绘出版社,1996.

5　黄幼才,刘文宝,李宗华,肖道纲. GIS空间数据误差分析和处理.武汉: 中国地质大学出版社,1995.

Nonlinear Error Propagation and Its Application in GIS

Hu Shengwu　Tao Benzao

(School of Geo-science and Surveying Engineering, WTUSM, 39 Luoyu Road, Wuhan, China, 430070)

Abstract　The formulas of error propagation for nonliner model are given and improved in order to be simplified and be really used in practice, and its application in error analysis and procession of spatical data of GIS are discussed. This paper studies its importance and usefulness for spatial data of GIS and draws a conclusion that nonlinear error propagation really describles accuracy.

Key words　nonlinear model; error propagation; error analysis of spatial data of GIS