

矢量GIS中属性数据的不确定性分析*

刘文宝^{1,2}, 邓敏¹, 夏宗国^{2,3}

(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, 波士顿, 02125;

3. 武汉测绘科技大学, 湖北 武汉, 430079)

Analysis of Uncertainties of Attributes in Vector GIS

LIU Wen-bao^{1,2}, DENG Min¹, XIA Zong-guo^{2,3}

(1. Department of Geosciences, Shandong University of Science and Technology, Taian, Shandong, 271019; 2. Department of Earth & Geographic Sciences, University of Massachusetts at Boston, Boston, MA 02125; 3. Wuhan Technical University of Surveying and Mapping, Wuhan, Hubei, 430079)

Abstract: This paper discusses the measurement of uncertainties of attribute data in vector GIS. The uncertainties considered include uncertainties in thematic classification, positional errors of boundaries and sampling errors of quantitative attributes. A quantitative index is proposed for describing the uncertainties related to thematic classification, which includes the effects of complexity of boundaries and areas of polygons. Correlation functions between sample values in attribute data are used to measure data sampling errors quantitatively. The second part of this paper focuses on modeling the propagation of attribute uncertainties in logical and arithmetical operations in GIS analysis using fuzzy set and statistical techniques. Two examples are provided to show how the models can be utilized in specific GIS applications.

Keywords: attribute data; uncertainty; error propagation

摘要: 本文从属性区域分类不确定性、边界定位误差和区域内部定量属性数据的抽样误差出发, 综合进行属性数据不确定性的度量 and 传播分析。针对由边界定位误差引起的属性分类不确定性, 在顾及区域面积与边界线数等因素下, 提出了一种定量度量指标。对于属性区域内部的定量属性抽样值, 则建立了抽样值的相关函数表达式。针对GIS中常见的两大类操作, 即逻辑操作和算术操作, 分别根据模糊集合论和概率统计理论建立了相应的属性不确定性传播模型, 并通过算例说明了这些模型在GIS中的实际应用。

关键词: 属性数据; 不确定性; 误差传播

1 引言

近年来GIS数据质量的研究集中在位置数据的误差上, 而对属性数据不确定性的研究较少^[1]。在以属性数据分析为主的许多GIS应用领

域中, 例如土壤化学、环境科学和农林科学, 属性数据质量的重要性甚至远大于位置数据^[2]。

属性数据的不确定性主要来自数据源的不确定性、数据建模的不确定性和分析过程中引入的不确定性等。其中数据源的不确定性又来源于数

* 收稿日期: 1998-12-17, 截稿日期: 1999-08-31。刘文宝, 男, 35岁, 博士后, 教授。主要研究领域为遥感与GIS中的数据质量、误差与不确定性及信息可视化等方面。
国家自然科学基金(编号: 49801016; 49671063)资助项目。

据采集过程中的测量、人为判断和假设等。例如，在 GIS 中将现实空间分为具有渐变特征的连续空间区域和具有跳跃特征的离散空间区域^[3]；对于后者，通常假定区域内部属性是均匀且同质的，其重要特征的变化发生在边界上，像土壤类型分布图。由于区域边界是经过人为判断和边界定位确定的，因此，区域多边形中属性数据的不确定性将主要取决于属性分类不确定性、边界定位误差和属性测量误差。这里，误差是指观测值与其真值间的差异^[3]，是一种具有统计意义的概念；而不确定性则是指被测量对象知识缺乏的程度，它表现为随机性和模糊性^[4]，而且受尺度、分辨率、抽样等许多因素影响^[5]。

在矢量专题图的 GIS 应用分析中，属性数据的不确定性问题主要是属性不确定性的度量和在 GIS 分析过程中的传播。为此，本文探讨这类问题。

2 属性不确定性的度量

2.1 区域边界的属性不确定性

如上所述，由于属性区域边界存在定位误差，进而导致边界附近区域属性的分类不确定性。在矢量专题图上，多边形是由若干边界线要素围成的封闭区域，其边界位置误差可由误差环表达^[6]。下面在假定属性区域边界线段为“e-带”的情况下^[4]，探讨由属性区域边界定位误差引起的属性不确定性。

2.1.1 由边界定位误差引起的属性不确定性

如图 1，除了传统的属性分类误差本身外^[7]，由于属性区域边界存在定位误差，也将导致难以判断属性区域单元 P 的属性值是否为 A。若将 P 视为一个不含定位误差的点元，根据模糊

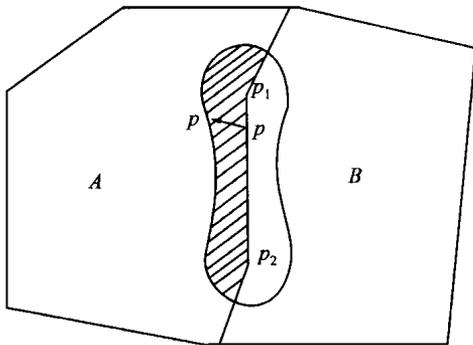


图 1 不确定的属性关系判定

Fig 1 The determination of uncertain attribute relations

数学原理，经分析可选用下述隶属函数^[8]

$$\mu(d) = \begin{cases} \frac{1}{2} e^{-\frac{d^2}{2\sigma^2}} & d < 0 \\ 1 - \frac{1}{2} e^{-\frac{d^2}{2\sigma^2}} & d \geq 0 \end{cases} \quad (1)$$

表示单元 P 的属性值为 A 的隶属度。其中，d 为 P 到属性区域边界线段 P₁P₂ 的垂直距离；d = 0 表示 P 在属性区域内部或边界线上，d < 0 表示 P 在属性区域外部；P_i 为 P 向边界线 P₁P₂ 作垂线的交点，σ 为点 P_i 的点位标准差。

根据式(1)可对任意位置的属性区域点元计算其属性值为 A 的隶属度。例如 μ(d) = 0.90 表明点元 P 属性值为 A 的可能性为 0.90。若将隶属度相同的点连接起来，则可得一系列等隶属度区域，如图 2。其中，A(α) (0 < α < 1) 表示任意位置的点元属性值为 A 的可能性大于等于 α 的一个区域。

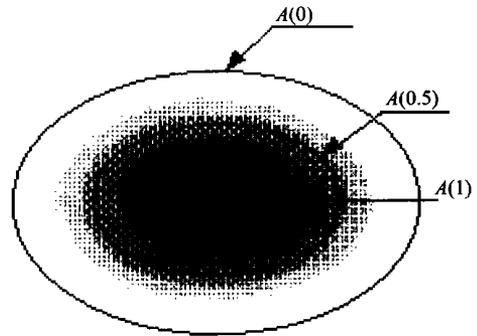


图 2 三种不同隶属度区域示意图

Fig 2 Graphic illustration of three regions with different memberships

2.1.2 区域边界属性的不确定性指标

由边界定位误差引起的多边形属性不确定性，不仅与边界线的定位精度有关，而且与多边形的形状和面积及相邻多边形中属性特征值的变化程度有关^[1-3]。故可利用下列数值指标衡量由边界定位误差引起的多边形属性不确定性：

$$T_i = \sum_{j=1}^{n_i} w_{ij} S_{ij} / S_i \quad (2)$$

式中：T_i 为第 i 个多边形的边界属性不确定性值；n_i、S_i 分别为第 i 个多边形的边界线条数和面积；S_{ij} 为第 i 个多边形中第 j 条边界线的“e-带”面积；w_{ij} 反映第 j 条边界线相邻多边形的属性差异程度。下面利用模糊集合论探讨权 w_{ij} 的确定方法。

不妨设两相邻多边形 i 和 k 属性值的模糊集合为

$$\mu_i = (\mu_{i1}, \mu_{i2}, \Lambda, \mu_{in}) \quad (3)$$

$$\mu_k = (\mu_{k1}, \mu_{k2}, \Lambda, \mu_{kn}) \quad (4)$$

其中 n 为专题图上属性的类型数,且满足归一化要求,即有 $\sum_{q=1}^n \mu_{iq} = 1$ 和 $\sum_{q=1}^n \mu_{kq} = 1$ 。则权的计算公式为

$$w_{ij} = \frac{1}{2} \sum_{q=1}^n |\mu_{iq} - \mu_{kq}| \quad (5)$$

式中: $0 < w_{ij} < 1$ 。如果不考虑相邻多边形属性特征值的变异程度,可令 $w_{ij} = 1$,则式(2)简化为

$$T_i = \sum_{j=1}^{n_i} S_{ij} / S_i \quad (6)$$

2.2 区域内的属性误差

属性区域内的属性值,例如环境科学中土壤化学成份的含量等,都是通过抽样测量得到的。不妨将某一特定位置处的空间属性表示为

$$A_i(x_t) = b_i(x_t) + \Delta_i(x_t) \quad (7)$$

其中: $A_i(x_t)$ 为 x_t 处第 i 个空间属性的观测值,带有随机误差 $\Delta_i(x_t)$; $b_i(x_t)$ 为其真值。对于不同的 x 值序列,函数 $A_i(x)$ 和 $\Delta_i(x)$ 分别为一元随机过程。而 $A(x_t) = g(A_1(x_t), A_2(x_t), \Lambda, A_m(x_t))$ 和 $\Delta(x_t) = h(\Delta_1(x_t), \Delta_2(x_t), \Lambda, \Delta_m(x_t))$ 则为 m 元随机场。

根据随机场理论^[9],空间属性观测量的统计不确定性可由下列相关函数表达

$$\sigma_{ij}(x_1, x_2) = \rho_{ij}(x_1, x_2) \sigma_i(x_1) \sigma_j(x_2) \quad (8)$$

其中: $\rho_{ij}(x_1, x_2)$ 为第 i, j 个数据层中 x_1, x_2 处属性观测量的相关系数,而 $\sigma_i(x_1), \sigma_j(x_2)$ 为其方差。

3 属性不确定性的传播

在 GIS 应用过程中,常常需要利用空间数据库中的一些基础属性数据派生一部分新属性数据。例如,根据某一区域的土壤类型、坡度以及湿度属性数据产生一幅关于在这区域种植某种农作物的适宜性地图^[1]。这种 GIS 空间操作函数可表达为

$$Z(x) = f(A_1(x), A_2(x), \Lambda, A_m(x)) \quad (9)$$

式中: $Z(x)$ 为对 m 个空间属性 $A_i(x)$ ($1 \leq i \leq m$) 进行空间操作 f 得到的新属性域。实用中,模型(9)常表现为逻辑模型和算术模型^[10]。

3.1 逻辑模型

GIS 中传统的逻辑模型是布尔模型,只能处理确定性的属性数据分析。为了处理不确定性的属性数据,下面对传统的布尔逻辑模型进行扩

展,建立模糊逻辑不确定性传播模型。

以任意位置处的属性值满足属性值条件的程度作为隶属度,记

$$\mu_{A_i}(T = v_i(x)) = f(v_i, \alpha_{ln}, \alpha_{lr}, \beta_{lm}, \beta_{lr}) \quad (10)$$

其中: $v_i(x)$ 为位置 x 处第 i 个属性的属性特征值; $\mu_{A_i}(T = v_i(x))$ 为位置 x 处属性值为 A_i 的隶属度,简记为 μ_{A_i} ; 参数 α_{ln}, α_{lr} 分别为属性值条件的左、右临界值; β_{lm}, β_{lr} 分别为左、右临界值缓冲宽度。

对 N 幅单一属性专题图进行逻辑操作,得到一幅具有 N 种属性的新专题图,其并、交、补模糊逻辑不确定性传播模型分别为

$$\mu_{(A_1, A_2, \Lambda, A_N)}(T = v(x)) = \max(\mu_{A_1}, \mu_{A_2}, \Lambda, \mu_{A_N}) \quad (11)$$

$$\mu_{(A_1, A_2, \Lambda, A_N)}(T = v(x)) = \min(\mu_{A_1}, \mu_{A_2}, \Lambda, \mu_{A_N}) \quad (12)$$

$$\mu_{A_i}^-(T = v_i(x)) = 1 - \mu_{A_i}(T = v_i(x)) \quad (13)$$

3.2 算术模型

在 GIS 属性数据分析中,算术模型是派生新属性数据另一类常用模型,如水土流失量 FAO 计算模型^[1]。设对 N 幅单一属性数据层进行算术运算,得到新专题图的属性数据为

$$Z(x) = f(A_1(x), A_2(x), \Lambda, A_N(x)) \quad (14)$$

上式线性化后,应用协方差传播定律得^[11]

$$D_{ZZ} = e D_{AA} e^T \quad (15)$$

其中, $e = (e_1, e_2, \Lambda, e_N) = \left(\frac{\partial f}{\partial A_1} \Big|_0, \frac{\partial f}{\partial A_2} \Big|_0, \Lambda, \frac{\partial f}{\partial A_N} \Big|_0 \right)$, $D_{AA} =$

$$D_{AA} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \Lambda & \sigma_{1N} \\ \sigma_{21} & \sigma_2^2 & \Lambda & \sigma_{2N} \\ \Lambda & \Lambda & \Lambda & \Lambda \\ \sigma_{N1} & \sigma_{N2} & \Lambda & \sigma_N^2 \end{pmatrix}, \text{ 顾及}$$

$\sigma_{ij} = \rho_{ij} \sigma_i \sigma_j$, 展开得

$$\hat{\sigma}_z^2 = \sum_{i=1}^N \sum_{j=1}^N \rho_{ij} \sigma_i \sigma_j e_i e_j \quad (16)$$

若 $N = 1$, 则为对一幅专题图进行 GIS 算术操作,则式(16)简化为

$$\hat{\sigma}_z^2 = e^2 \sigma^2 \quad (17)$$

4 算例分析

由边界定位误差引起的属性不确定性指标的计算比较简单。例如对于一个属性多边形,通过式(2)直接计算其位置误差带加权面积与多边形本身面积的比值得出。对于一幅专题图,可以对所有的多边形属性不确定性指标值进行统计,得到总体的属性不确定性指标值。因此,下面重点讨论属性不确定性的传播。

4 1 逻辑模型

下面是 GIS 中根据布尔逻辑模型进行空间信息查询的一个例子。

“如果区域的坡度 10%；粘土含量 41 mg/kg；3 5 PH 8 5；那么该区域为所选的植被生长适宜土地。”

基于普通集合论思想，则上述知识规则可表示为

$$\mu_{(A_1, A_2, A_3)}(v(x)) = \begin{cases} 1 & \text{坡度 } 10\% ; \text{粘土含量 } 41 \text{ mg/kg;} \\ & \text{并且 } 3.5 \leq \text{PH} \leq 8.5 \\ 0 & \text{其他} \end{cases}$$

若根据上述隶属函数进行区域选择，则易导致土壤属性值偏高或偏低于属性临界值而使选择失误。为此，下面根据式(12)进行属性不确定性传播分析。

不妨假定满足上述属性值条件的隶属函数分别为线性函数，即

$$\mu_{A_1}(v_1(x)) = \begin{cases} 1 & v_1 \leq 9\% \\ \frac{11 - v_1}{2} & 9\% < v_1 < 11\% \\ 0 & v_1 \geq 11\% \end{cases}$$

$$\mu_{A_2}(v_2(x)) = \begin{cases} 0 & v_2 \leq 38.5 \\ \frac{43.5 - v_2}{5} & 38.5 < v_2 < 43.5 \\ 1 & v_2 \geq 43.5 \end{cases}$$

$$\mu_{A_3}(v_3(x)) = \begin{cases} 0 & v_3 \leq 3 \\ \frac{v_3 - 3}{4} & 3 < v_3 < 7 \\ 1 & 7 \leq v_3 \leq 8 \\ \frac{9 - v_3}{9} & 8 < v_3 < 9 \\ 0 & v_3 \geq 9 \end{cases}$$

根据上述 3 个隶属函数，首先计算某区域 Ω 中任意位置处分别属于属性值 A_1, A_2 和 A_3 的隶属度，列于表 1、2、3；然后利用式(12)计算 $\mu(T = v(x))$ ，结果列于表 4。其中行、列数分别表示区域任意位置的 x, y 局部区域标识坐标^[1]。

表 1 数据层 1 上区域 Ω 中任意位置属性值的模糊表示
Tab 1 Fuzzy representation of attribute values in the region Ω of the first data layer

0	1	2	3	4	5	6	7	8
1	0	0.11	0.23	0.35	0.51	0.34	0.28	0.09
2	0.06	0.20	0.38	0.56	0.70	0.57	0.35	0.16
3	0.13	0.30	0.54	0.71	0.85	0.73	0.54	0.25
4	0.35	0.57	0.80	0.92	1.00	1.00	0.76	0.43
5	0.48	0.72	1.00	1.00	1.00	0.96	0.78	0.57
6	0.64	0.88	1.00	0.90	0.86	0.79	0.62	0.43
7	0.41	0.68	0.76	0.70	0.64	0.44	0.26	0.09
8	0.22	0.40	0.49	0.46	0.30	0.27	0.20	0

表 2 数据层 2 上区域 Ω 中任意位置属性值的模糊表示

Tab 2 Fuzzy representation of attribute values in the region Ω of the second data layer

0	1	2	3	4	5	6	7	8
1	0.14	0.35	0.70	0.82	0.84	0.71	0.65	0.40
2	0.36	0.62	0.80	0.92	1.00	0.88	0.76	0.48
3	0.48	0.70	0.93	1.00	0.96	0.72	0.56	0.44
4	0.74	0.89	1.00	1.00	0.88	0.68	0.48	0.40
5	0.94	1.00	1.00	0.86	0.72	0.56	0.44	0.40
6	0.82	1.00	0.94	0.80	0.60	0.40	0.24	0.20
7	0.68	0.86	0.72	0.54	0.42	0.23	0.11	0
8	0.42	0.54	0.48	0.36	0.27	0.10	0	0

表 3 数据层 3 上区域 Ω 中任意位置属性值的模糊表示

Tab 3 Fuzzy representation of attribute values in the region Ω of the third map layer

0	1	2	3	4	5	6	7	8
1	0.43	0.46	0.54	0.44	0.38	0.34	0.34	0.24
2	0.60	0.68	0.70	0.62	0.52	0.40	0.42	0.32
3	0.62	0.72	0.78	0.70	0.64	0.50	0.54	0.40
4	0.70	0.90	1.00	0.88	0.76	0.56	0.60	0.48
5	0.90	1.00	1.00	1.00	1.00	0.76	0.54	0.32
6	1.00	1.00	0.80	0.74	0.72	0.60	0.44	0.32
7	0.86	0.80	0.66	0.54	0.54	0.36	0.22	0
8	0.62	0.60	0.40	0.35	0.36	0.11	0.06	0

表 4 叠置层上区域 Ω 中任意位置属性值的模糊表示

Tab 4 Fuzzy representation of attribute values in the region Ω of the overlaid map layer

0	1	2	3	4	5	6	7	8
1	0	0.11	0.23	0.35	0.38	0.34	0.28	0.09
2	0.06	0.20	0.38	0.56	0.52	0.40	0.35	0.16
3	0.13	0.30	0.54	0.70	0.64	0.50	0.54	0.25
4	0.35	0.57	0.80	0.88	0.76	0.56	0.48	0.40
5	0.48	0.72	1.00	0.86	0.72	0.56	0.44	0.32
6	0.64	0.88	0.80	0.74	0.60	0.40	0.24	0.20
7	0.41	0.68	0.66	0.54	0.42	0.23	0.11	0
8	0.22	0.40	0.40	0.35	0.27	0.10	0	0

分析表 4 知，经模糊逻辑运算得到的新专题图，其属性值不确定性与原始数据层属性值不确定性密切相关。下面以满足属性值程度为 0.7 作为选择各数据层上区域范围的截值，分别得区域 Ω_1, Ω_2 和 Ω_3 ，如图 3(a)、(b) 和 (c) 所示，将其进行逻辑叠置操作运算，结果如图 3(d)。

4 2 算术模型

在某环境评价专题图上存在一多边形区域，其中铅、钙两种元素含量属性数据的抽样值及其精度列于表 5。在 GIS 中利用下列模型^[3]

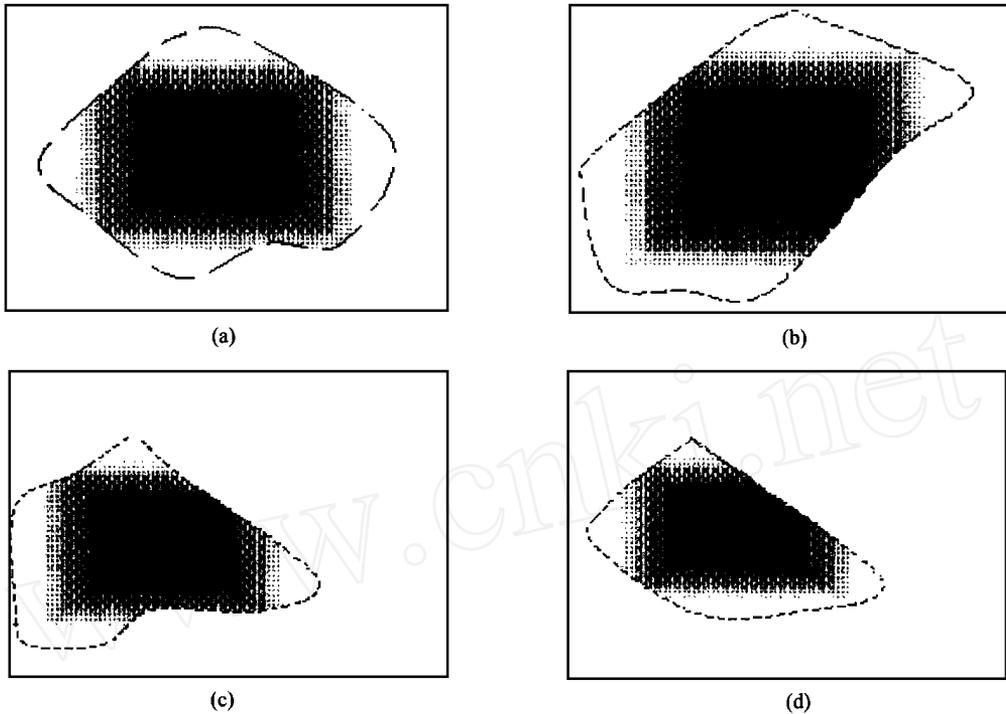


图 3 逻辑操作下的属性不确定性传播

Fig 3 Propagation of attribute uncertainty in logical operation

$$R = Pb + 13Ca \quad (18)$$

来预测土壤中铅的浓度 (Pb) 和钙的浓度 (Ca) 对环境的危害系数 R 。利用公式 (17) 计算的危害系数的标准差列于表 5 中最后一列。

表 5 不同误差状态下计算危害系数的起始数据和计算数据 (单位: 毫克/千克)

Tab 5 Data for computation of risk factor and results under different error statements (unit: mg/kg)

序号	Pb	σ_{Pb}	Ca	α_{Ca}	$\rho_{Pb, Ca}$	R	σ_R
1	210.0	3.5	4.7	0.12	0	271.1	3.83
2	210.0	3.5	4.7	0.12	0.6	271.1	4.24
3	215.0	3.9	5.0	0.14	0	280.0	4.30
4	215.0	3.9	5.0	0.14	0.8	280.0	4.92
5	220.0	3.7	4.8	0.13	0	282.4	3.99
6	220.0	3.7	4.8	0.13	-0.4	282.4	3.66

分析表 5 可知, 危害系数 R 的误差取决于铅和钙浓度的误差及其相关系数, 其中钙的浓度抽样误差对危害系数 R 计算值精度的影响程度大于铅的。

5 结论与建议

(1) 矢量专题图上定量属性数据的不确定性主要取决于属性分类不确定性(或边界判断不确定性)、边界定位误差和属性数据本身的抽样误

差。

(2) 模糊逻辑不确定性传播模型适于模拟 GIS 逻辑操作中由属性分类产生的模糊属性数据的不确定性; 算术模型用于模拟属性区域内部的随机定量属性数据中的统计误差。

(3) 属性数据的不确定性研究是一个比位置数据误差更为复杂的课题, 相对而言, 后者则比较单一、简单。而属性数据的表现形式多种多样, 例如定量的、定性的、数值的、文本的等, 其不确定性也与多种因素有关, 故需分别进行系统研究。

References:

- [1] BURROUGH P A, MCDONNELL R A. Principles of Geographical Information Systems [M]. New York: Oxford University Press, 1998. 231-240
- [2] MOWRER H T, CZAPLEW SKIR L, HAMRE R H. Spatial Accuracy Assessment in Natural Resources and Environmental Sciences [C]. Fort Collins Colorado: United States Department of Agriculture Forest Service, Rocky Mountain Forest and Range Experiment Station, 1996
- [3] HEUVELINK G B M. Error Propagation in Quantitative Spatial Modeling: Application in GIS [D]. Netherlands: ITC Publication, 1993
- [4] GOODCHILD M F. Modeling Error in Objects and

- Fields[A]. GOODCHILD M F, GOPAL S. Accuracy of Spatial Databases[C]. London: Taylor & Francis, 1989. 107-113
- [5] LU Wen-bao. A Theory of Uncertainty in Spatial Data within GIS[D]. Wuhan: Wuhan Technical University of Surveying and Mapping, 1995. 1-93 (in Chinese).
- [6] LU Wen-bao, DA IHong-lei, XU Pan-lin, GUO Jin-yun. Models of Positional Error Donut for Planar Polygon in GIS[J]. Acta Geodaetica et Cartographica Sinica, 1998, 27 (4): 338-344 (in Chinese).
- [7] CONGALTON R G, KASS G. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices[M]. Boca Raton: Lewis Publications, 1999.
- [8] YANG Lun-biao, GAO Ying-yi. Principles and Applications of Fuzzy Mathematics[M]. Guangzhou: Hua'nan University of Science and Technology Press, 1993 (in Chinese).
- [9] PITERBARG V I. Asymptotic Methods in the Theory of Gaussian Processes and Fields[M]. New York: The American Mathematical Society Press, 1996.
- [10] LU Wen-bao, Li Zong-hua. Error Propagation in Models of Knowledge Influence[J]. Bulletin of Science and Technology of Wuhan Technical University of Surveying and Mapping, 1996, 1: 5-8 (in Chinese).
- [11] MIKHAEL E M. Observations and Least Squares[M]. New York: IEP, 1976.