

基于法平面的 GIS 平面随机线元概率算法

汤仲安^{1,2}, 朱建军¹

(1. 中南大学 信息物理工程学院, 湖南 长沙 410083; 2. 湖南省 国土资源信息中心, 湖南 长沙 410011)

Probability Algorithm for 2-Dimension Random Line Segment Based on the Normal Plane

TANG Zhong-an^{1,2}, ZHU Jian-jun¹

(1. School of Inforphysics and Geomatics Engineering, Central South University, Changsha 410083, China; 2. Info Center for National Land & Resources of Hunan Province, Changsha 410011, China)

Abstract: Probability algorithm regarding a random line segment falls into its corresponding error model with a certain probability has been put forward based on normal plane of the random line segment in GIS and Probability Analysis and Numerical Value Integral, and some important results are obtained by comparing with probability algorithm based on modeling theory of error model as well as by case study and visualization analysis——both of the probability algorithms are of equivalence for modeling of error model to random line segment, furthermore, the former is current for modeling of error model to generic curves in GIS.

Key words: GIS; random line segment; normal plane; error model; probability analysis; algorithm

摘 要: 基于概率论与数值分析, 研究基于线元法平面的 GIS 中平面随机线元落入其相应误差模型内的概率算法, 通过实例计算与可视化分析, 比较基于误差模型建模机理的概率算法, 并得出如下结论: 基于线元法平面的概率算法与基于误差模型建模机理的概率算法对 GIS 中随机线元误差模型的建模是等价的, 而且对于 GIS 中一般曲线的误差模型建模而言, 前者具有通用性。

关键词: 地理信息系统; 随机线元; 法平面; 误差模型; 概率论; 算法

1 引 言

在 GIS 位置不确定性理论研究中, 线状实体误差模型规模的确定是制约误差模型建模及其应用的技术瓶颈。基于概率论, 线状实体落入其相应误差模型内的概率算法研究十分重要, 这不仅直接影响到概率计算精度, 而且更关乎线状实体误差模型规模的最终确定。

文献[1][2]提出了“标准误差椭圆缩放系数”、误差模型“形状因子 $[m(x_0, t)]$ ”、“尺度因子 (x_0) ”的概念, 研究了“等概率密度误差模型”(The equivalent probability density error model, “Tepdem”)建模机理。基于“Tepdem”, 文献[2][3]探讨了一般曲线误差模型包络线, 同样, 随机线元“Tepdem”包络线是线元任意点误差椭圆包络点(线元误差模型平面内与线元平行且与线元

上任意点误差椭圆相切的两个切点, 它们分别位于线元的两侧, 关于线元上该任意点中心对称, 确定方法详见文献[2]的集合。基于“Tepdem”及其包络点算法, 文献[2]完成了基于“Tepdem”误差模型建模机理的 GIS 随机线元概率算法(下文简称为“概率算法 1”)的研究。

对于随机线元, 线元上任意点误差椭圆两包络点连线必经过该误差椭圆中心(见文献[2]式(4.8.2))。但是, 当线状实体为非线元时, 线状实体上任意点误差椭圆两包络点连线不会经过该椭圆中心, 此时, “概率算法 1”不再适用。然而, 矢量 GIS 中, 通常多见的是 GIS 中一般线状实体(如缓和曲线、拟合曲线等), 因此, 为了线状实体误差模型的建模及其应用, 有必要展开基于线元法平面的随机线元概率算法(下文简称为“概率算法 2”)的研究。

收稿日期: 2004-12-10; 修回日期: 2005-06-16

基金项目: 国土资源部百名优秀青年科技人才计划资助项目(2005)

作者简介: 汤仲安(1971-), 男, 湖南浏阳人, 博士后, 高级工程师, 主要从事 GIS 与遥感信息的不确定性研究、现代测量数据处理理论与应用研究。E-mail: whtza@163.com

2 概率算法原理

如图 1 所示, 线元 $Z_0 Z_1$ 的空间概率密度分布曲面被平行于 $x-O-y$ 的平面 $F(x, y) = (0)$ 所截, 其截面封闭边沿线即为文献 [1][2] 所定义的随机线元“Tepdem”。显然, “Tepdem”可划分为图 1 所示的 B_1, B_2, B_3 3 个相互独立的区域。其中, 端点 Z_0, Z_1 处线元空间概率密度分布曲面与过两端点处线元法平面之间所笼罩下的空间误差区域被平面 $F(x, y) = (0)$ 所截而得的截面区域分别为“Tepdem”的一部分, 设为 B_1, B_3 , 并设其对应的空间误差区域分别为 Ω_1, Ω_3 ; 而将线元上任意离散点处线元法平面与线元空间概率密度分布曲面之间的截面边沿线——1 维正态概率密度分布曲线 $f_t, t \in [0, 1]$ 叠加所笼罩下的空间误差区域被平面 $F(x, y) = (0)$ 所截而得的截面区域定义为 B_2 , 其空间误差区域设为 Ω_2 。

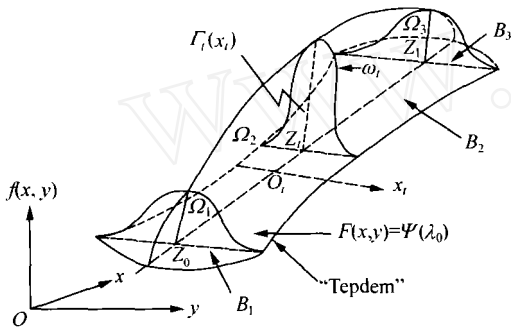


图 1 由随机线元法向 1 维正态概率密度分布曲线 $f_t, t \in [0, 1]$ 叠加所构成的概率密度分布曲面
 Fig. 1 The spatial probability density distribution curve surface of the random line segment $Z_0 Z_1$ structured by normal probability density distribution curves $f_t, t \in [0, 1]$

根据概率论与数理统计, 空间误差区域 $\Omega_1, \Omega_2, \Omega_3$ 对应的体积 V_1, V_2, V_3 即为随机线元在“Tepdem”各误差区域 B_1, B_2, B_3 内的概率, 记为 $P(B_1) = V_1, P(B_2) = V_2, P(B_3) = V_3$, 因此, 随机线元落于其相应“Tepdem”内的概率为 $P_0 = P(B_1 \cup B_2 \cup B_3)$, 即

$$P_0 = P(B_1) + P(B_2) + P(B_3) - P(B_1 \cap B_2) - P(B_1 \cap B_3) - P(B_2 \cap B_3) + P(B_1 \cap B_2 \cap B_3) \quad (1)$$

由于 B_1, B_2, B_3 各区域内随机线元上各点的随机变化事件相互独立, 因而式 (1) 中任意两个事件同时发生的概率等于其单独发生的概率的积。

例如: $P(B_1 \cap B_2) = P(B_1) P(B_2)$, 于是, 根据文献 [4][5], 式 (1) 又可以如下写出

$$P_0 = P(B_1) + P(B_2) + P(B_3) - P(B_1) P(B_2) - P(B_1) P(B_3) - P(B_2) P(B_3) + P(B_1) P(B_2) P(B_3) \quad (2)$$

这样, 求出 $P(B_1), P(B_2), P(B_3)$, 我们即可确定随机线元落入其“Tepdem”内的概率值。

3 公式推导

首先确定随机线元端点处对应概率 $P(B_1), P(B_3)$ 。

设线元精度较低的端点 Z_0 所对应“标准误差椭圆缩放系数”(见文献 [1][2]) 为 m_0 。根据概率论与数理统计, 经过某 2 维点误差椭圆中心且垂直于该误差椭圆所处的坐标平面的任意平面必定平分该 2 维点对应的概率密度分布曲面。因此, 有:

1. 在 Z_0 处, $P_{B_1} = V_1 = \frac{1}{2} (1 - \exp(-\frac{m_0^2}{2}))$ 。
2. 根据文献 [1][2], Z_1 处“标准误差椭圆缩放系数”为 m_1 。

从而, $P_{B_3} = V_3 = \frac{1}{2} (1 - \exp(-\frac{m_1^2}{2}))$, 其中, m_1 为线元上任意点在以该点对应的标准误差椭圆两轴为坐标轴线的相对坐标系 u_t-v_t 中的坐标方差, 且满足 $\frac{m_1^2}{u_t^2 v_t^2} = \frac{m_0^2}{x_t^2 y_t^2} (1 - \frac{r_{x_t y_t}^2}{x_t^2 y_t^2})$ (其中 $r_{x_t y_t}$ 为坐标 x_t, y_t 的相关系数。下同)。文献 [1][2] 已有详细的证明与分析, 此处从略。

其次求算 $P(B_2)$ 。

由图 1 知, 线元上任意点 Z_t 处线元法平面与其空间概率密度曲面的截面线为一元正态分布函数图像, 形状关于该点处 2 维概率密度分布曲面的中轴线对称。若以此中轴线与平面 $x-O-y$ 的交点 O_t 为原点、以线元法平面与平面 $x-O-y$ 的交线为 x_t 轴, 以经过原点 O_t 、平行于 $f(x, y)$ 轴且垂直于平面 $x-O-y$ 的中轴线为 $f_t(x_t)$ 轴, 则相对坐标系 $x_t - O_t - f_t(x_t)$ 中该截面线的概率密度分布函数为 $f_t(x_t) = \frac{1}{\sqrt{2\pi} \sigma_t} \exp[-\frac{(x_t)^2}{2\sigma_t^2}]$ 。

根据文献 [1][2] 知: 参数 t 处线元上对应点 Z_t 处“标准误差椭圆缩放系数”为 $m(t)$, 其误差椭圆参数为 $E(t) = m(t) u_t$ 及 $F(t) = m(t) v_t$ 。

由于 Z_t 处一元正态分布概率密度分布函数振幅与对应点 2 维正态分布概率密度分布函数振幅相等,故

$$\frac{1}{2\sqrt{t}} = \frac{1}{2u_t v_t} \Rightarrow t = \sqrt{2} u_t v_t \quad (3)$$

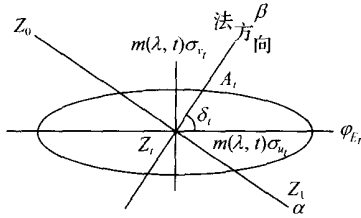


图 2 随机线元任意点 Z_t 处对应的 2 维概率密度分布曲面与 $F(x, y) = c_0$ 的截面图

Fig.2 The section figure built by plane $F(x, y) = c_0$ and 2D probability density distribution curve surface of arbitrary point Z_t on random line segment

如图 2 所示,随机线元 $Z_0 Z_1$ 的方位角为可以如下计算:

$$\text{令 } k_0 = \frac{y_1 - y_0}{x_1 - x_0}, \text{ 若 } k_0 > 0, \theta = \arctan(k_0);$$

反之, $\theta = +\arctan(k_0)$ 。

设随机线元法向与此误差椭圆主轴的夹角为 t , 则

$$t = \frac{\pi}{2} - |\theta - \alpha|$$

由图 2 知: $|Z_t A_t|$ 即为 Z_t 处线元法向一元正态概率密度分布曲线在截平面 $F(x, y) = c_0$ 上的截距。根据解析几何,容易求得

$$|Z_t A_t| = \frac{m(\lambda, t) \cdot u_t v_t}{\sqrt{\cos^2 t \cdot v_t^2 + \sin^2 t \cdot u_t^2}}$$

结合式(3),令

$$|Z_t A_t| = c_0 \Rightarrow c_0 = \frac{|Z_t A_t|}{t}$$

$$P(B_2) = \lim_{m \rightarrow \infty} \sum_{k=0}^{m-1} S(c_0, t(k)) \cdot t = \lim_{m \rightarrow \infty} \left\{ \frac{1}{3m} \left[\sum_{k=0}^{m-1} (S(c_0, t(k)) + S(c_0, t(k+1)) + \sqrt{S(c_0, t(k)) S(c_0, t(k+1))}) \right] \right\}$$

考虑到参数 $t(k)$ 的离散化取值,结合满足 GIS 精度要求的线元离散点个数 m , 有

$$P(B_2) = \frac{1}{3m} \sum_{k=0}^{m-1} (S(c_0, t(k)) + S(c_0, t(k+1)) + \sqrt{S(c_0, t(k)) S(c_0, t(k+1))})$$

$$\frac{m(c_0, t)}{\sqrt{2(\cos^2 t \cdot v_t^2 + \sin^2 t \cdot u_t^2)}} \quad (4)$$

由于 Z_t 处线元法平面在线元空间概率密度分布曲面内被平面 $F(x, y) = c_0$ 所截而得的截面面积即为一元正态曲线在区间 $[t_0, t_1]$ 内的概率,而截面面积

$$S(c_0, t) = \int_{t_0}^{t_1} \frac{1}{\sqrt{2t}} \exp\left(-\frac{(x_t)^2}{2t}\right) dx_t = \int_{t_0}^{t_1} \frac{1}{\sqrt{2t}} \exp\left(-\frac{t^2}{2}\right) dt = 2 \int_{t_0}^{t_1} \dots$$

因此,由式(4)计算出 $S(c_0, t)$ 后, $P(B_2)$ 可按数值积分(复合辛普生公式)计算。由此可见: $S(c_0, t)$ 是参数 t 的函数。将截面面积 $S(c_0, t)$ 对线长 $L_{Z_0 Z_1}$ 积分,则有

$$V_2 = \int_0^{L_{Z_0 Z_1}} S(c_0, t) dl = P(B_2)$$

由于 $S(c_0, t)$ 是参数 t 的函数,又 $l:0 \rightarrow L_{Z_0 Z_1} \Leftrightarrow t:0 \rightarrow 1$, 故

$$P(B_2) = \int_0^{L_{Z_0 Z_1}} S(c_0, t) dl = \int_0^1 S(c_0, t) dt \quad (5)$$

为便于计算,我们将随机线元离散化为有限个点,利用数值分析来求算 $P(B_2)$ 。根据 GIS 分析与精度要求,离散点数目是一个具有确定范围的数,其确定方法在文献[1][2]中已有详细论述。设 m 为满足 GIS 精度要求的线元离散点个数的下限,且对于任意的非负整数 $k \in [0, m]$, 令 $t(k) = k/m$, 则 $t(k) \in [0, 1]$ 。于是,由数值分析,结合式(5),有

$$P(B_2) = \int_0^1 S(c_0, t) dt = \lim_{m \rightarrow \infty} \sum_{k=0}^{m-1} S(c_0, t(k)) \cdot t$$

为了概率计算的精确性,随机线元两相邻法向上一元正态曲线、平面 $F(x, y) = c_0$ 以及线元概率密度分布曲面之间围成的不规则空间体积可按棱台体积公式计算,即有

为了保证计算精度,在离散点 m 的基础上,不断加大 m 的值,使得计算精度达 10^{-6} 。

最后,分别将求得的 $P(B_1), P(B_2), P(B_3)$ 代入式(2),即可求算出 Z_0 处任意 c_0 所对应的随机线元落入其相应的“Tepdem”内的概率值。

4 实例计算与可视化分析

已知线元 $Z_0 Z_1$ 的端点坐标分别为 (845, 765), (593, 705), 单位:m。其坐标间方差-协方差矩阵为

$$\begin{pmatrix} 2.534 & 3 & -0.028 & 7 & -0.064 & 5 & 0.026 & 5 \\ & & 2.143 & 2 & 0.047 & 3 & -0.086 & 7 \\ \text{对} & & & & 1.145 & 3 & 0.023 & 5 \\ & & & & & & 0.732 & 3 \\ \text{称} & & & & & & & \\ (\text{cm}^2) & & & & & & & \end{pmatrix} \quad (6)$$

4.1 误差模型及其包络线的比较

当 θ_0 固定时,根据文献[1][2]中“Tepdem”建模机理及随机线元端点坐标精度,我们可以确定符合“概率算法 1”的该线元“Tepdem”及其相应的边界包络线。

另一方面,根据线元上任意点坐标 $(x(t), y(t))$,线元的方位角 α ,以及基于误差椭圆求得的线元法向截距 $|Z_t A_t|$,可以得到线元法向上对称于该随机线元的两点 $U(t), D(t)$

$$\left. \begin{aligned} x_{U(t)} &= x(t) + |Z_t A_t| \cos(\alpha - \frac{\pi}{2}) \\ y_{U(t)} &= y(t) + |Z_t A_t| \sin(\alpha - \frac{\pi}{2}) \\ x_{D(t)} &= x(t) + |Z_t A_t| \cos(\alpha + \frac{\pi}{2}) \\ y_{D(t)} &= y(t) + |Z_t A_t| \sin(\alpha + \frac{\pi}{2}) \end{aligned} \right\} \quad (7)$$

式中, $t \in [0, 1]$ 。

这样,图 1 中误差区域 B_2 的边界包络线即为 $U(t), D(t), t \in [0, 1]$ 的集合,而其端点处边界包络线为对应端点误差椭圆的一部分,与“概率算法 1”对应的端点处包络线相同。

为便于分析,可视化操作中离散点数目 $m = 25, \theta_0 = 50$ 。可视化图形的规模在原图的基础上放大了 100 倍,利用计算机编程实现。可视化结果如图 3 所示。

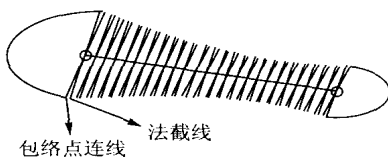


图 3 基于两种算法的误差模型及其包络线叠合图
Fig.3 Superposition for error models and its corresponding boundary envelop discrete points based on both of the probability algorithms

图 3 由两部分组成:一部分为按照“概率算法 1”生成的误差模型及包络点连线;另一部分为按照“概率算法 2”生成的误差模型及截距 $2|Z_t A_t|$ (即 $U(t), D(t)$ 的连线), $t \in [0, 1]$ 。两部分对应线元离散点数目相等、放大比例一致,叠加在同一幅图上便于可视化分析。

由图 3 知,线元上任意点误差椭圆两包络点连线与该点处线元法平面在对应椭圆上的截距之间存在交角,长度上包络点连线大于线元法平面在椭圆上的截距,这与事实吻合;且当 $t \in [0, 1]$ 时, $U(t), D(t)$ 都分布于“Tepdem”边界包络线上,亦即,按照“概率算法 1”与“概率算法 2”绘制的误差模型及其包络线完全一致。因此,线元空间概率密度分布曲面可视为“线元法向 1 维正态概率密度分布曲线叠加所笼罩下的空间误差区域”。相应地,“Tepdem”内误差区域 B_2 可以描述为:线元上任意点处法平面与线元空间概率密度分布曲面之间的截口线——一元正态分布曲线 $f(x, y), t \in (0, 1)$ 的叠加体被平面 $F(x, y) = \theta_0$ 所截而得的面状区域。

4.2 概率计算与分析

为了比较两种概率算法的计算精度,固定线元上离散点个数 m (此处, $m = 2590$, 满足 GIS 精度及其分析要求,确定方法见文献[1][2],变动 θ_0 以求得如表 1 所示的计算结果。

表 1 离散点个数 m 一定而 θ_0 变动时两种算法下的概率计算值

Tab.1 The probability calculation values of two algorithms in the case of certain quantity of discrete points and variation of θ_0

θ_0	1.0	2.0	3.0
概率算法 1	0.738 661	0.891 464	0.944 049
概率算法 2	0.738 745	0.891 579	0.944 236
θ_0	4.0	5.0	6.0
概率算法 1	0.969 478	0.984 129	0.992 010
概率算法 2	0.969 759	0.984 491	0.992 427

由表 1 可得如下结论:

1. 随机线元离散点数目 m 一定时,对于同一 θ_0 ,两种概率算法对应的概率计算值差异体现于万分位上,数值几乎相等,因此,两种概率算法对于决定“尺度因子” θ_0 没有影响;
2. “概率算法 2”的概率计算值比“概率算法 1”的概率计算值略大,因此,当概率置信水平

(95%)一定时,根据“概率算法 1”得到的满足概率置信水平的 ρ_0 (本例为 3.2)也必然使得根据“概率算法 2”计算的概率值满足同样给定的概率置信水平;

3. 线元落入其相应“Tepdem”内的概率值并不随 ρ_0 的变化而快速收敛于 1。

4.3 分析与小结

根据 4.1、4.2 的实例计算与可视化分析,可以得出如下结论:

1. 对于随机线元,“概率算法 1”与“概率算法 2”所对应的误差模型及其边界包络线完全一致,概率计算值近似相等,两种概率算法对随机线元误差模型的建模是等价的。

2. 根据文献[1][2],“Tepdem”由线状实体上任意点“标准误差椭圆缩放系数”、“尺度因子”和标准误差椭圆参数所构成的误差椭圆簇聚合而成;由文献[6]知:“ ρ_0 -带”是线状实体上任意点法向在该点对应的标准误差椭圆上截距的聚合体,而“ m -带”则是线状实体上任意点标准误差椭圆与线状实体距离最大点的集合。由“概率算法 1”所得的误差模型及其包络线相当于基于“Tepdem”建模机理的“ m -带”,而由“概率算法 2”所得的误差模型及其包络线则相当于基于“Tepdem”建模机理的“ ρ_0 -带”。由 4.1 知,此时,两者等价。

3. 由 4.2(2)知,对于同一 ρ_0 ,“概率算法 2”的概率值略大于“概率算法 1”的概率值,因此,当概率置信水平一定时,“概率算法 2”对应的“Tepdem”范围略小,这对给定位置不确定性指标的线状实体定位而言,“概率算法 2”较为严格。由于一般曲线对应的“概率算法 1”无法得出,因此,为了 GIS 中线状实体概率算法的通用性,结合 4.2

结论 1,我们可以选用“概率算法 2”,即基于法平面的概率算法来计算线状实体落入其相应“Tepdem”内的概率值。

参考文献:

- [1] TANG Zhong-an, WANG Xir-zhou, JI Xi-an-hua. Modeling Theory of Error Model to Line Segment in Vector GIS[J]. Geomatics and Information Science of Wuhan University, 2004, (11): 968-972. (汤仲安,王新洲,纪现华. 矢量 GIS 平面随机线元误差模型建模机理[J]. 武汉大学学报信息科学版, 2004, (11): 968-972.)
- [2] TANG Zhong-an. The Equivalent Probability Density Error Model to Line Entity in Vector GIS[D]. Wuhan: Wuhan University, 2004. (汤仲安. 矢量 GIS 线状实体等概率密度误差模型[D]. 武汉: 武汉大学, 2004.)
- [3] TANG Zhong-an, WANG Xir-zhou, CHEN Zhi-hui. Envelop of Error Model of 2D Spatial Curve in GIS[J]. Acta Geodaetica et Cartographica Sinica, 2004, 33(2): 151-155. (汤仲安,王新洲,陈志辉. GIS 中平面一般曲线误差模型包络线[J]. 测绘学报, 2004, 33(2): 151-155.)
- [4] LI Qing-hai, TAO Ben-zao. Probability Theory & Statistics and Its Applications in Surveying[M]. Beijing: The Publishing House of Surveying and Mapping, 1992. (李庆海,陶本藻. 概率统计原理和在测量中的应用[M]. 北京: 测绘出版社, 1992.)
- [5] DAI Hong-lei. Theory and Methods of Positional Uncertainty Measurement and Propagation in Vector GIS[D]. Wuhan: Wuhan Technical University of Surveying and Mapping, 2000. (戴洪磊. 矢量 GIS 中位置不确定性度量与传播的理论和方法[D]. 武汉: 武汉测绘科技大学, 2000.)
- [6] LIU Da-jie, SHI Wen-zhong, TONG Xiao-hua, SUN Hong-chun. Accuracy Analysis and Quality Control of Spatial Data in GIS[M]. Shanghai: Science and Technology Literature Publishing Company of Shanghai, 1999. (刘大杰,史文中,童小华,孙红春. GIS 空间数据的精度分析与质量控制[M]. 上海: 上海科学技术文献出版社, 1999.)

(责任编辑:雷秀丽)