

文章编号:1671-6833(2004)02-0094-03

GIS 中道路曲线的整体误差模型与试验研究

郭同德<sup>1</sup>, 杜云海<sup>2</sup>, 卢 艳<sup>1</sup>

( 1. 郑州大学环境与水利学院, 河南 郑州 450002; 2. 郑州大学工程力学系, 河南 郑州 450002)

摘 要: 以随机过程的均方微积分为工具描述了曲线要素的整体误差, 用过程的数字特征定义了曲线整体误差的具体指标, 分析了误差指标的几何意义. 根据过程模型设计了曲线整体误差的扫描数字化试验, 用三次参数样条高密度地插值出一条道路曲线, 对道路中心线进行反复的采集, 经转换获取了随机过程的 100 条样本曲线, 在此基础上估计出了误差指标值, 讨论了过程的平稳性, 分析了扫描数字化中系统误差的成因以及相应的削减系统误差的方法. 试验数据大体上代表了道路中心线扫描数字化的精度.

关键词: 曲线; 随机过程; 均方积分; 误差函数; 总误差

中图分类号: P 207 文献标识码: A

0 引言

曲线是GIS 中最基本的几何要素, 其位置不确定性对GIS 的各种分析与输出结果起着至关重要的影响, 是合理评价输出结果可靠性的依据. 与点要素相比, 线要素, 尤其是曲线要素的不确定性的理论与实验研究较为薄弱. 研究曲线位置不确定性时, 常将曲线分为类型-I 线与类型-II 线. 类型-I 线是指现实世界中有明确定义的线, 如道路中心线、地籍边界等; 而类型-II 线是指无明确定义的线, 如森林边界等. 本文拟研究的是类型-I 线, 将分别对以下几个问题进行研究: ①曲线误差的定义; ②曲线误差的数学模型; ③道路中心线的扫描数字化误差试验.

1 曲线误差的模型

曲线误差的定义是一个正在讨论的问题, 文献 [1] 对折线段定义了误差指标, 文献 [2] 从线段的概率分布出发, 研究了直线段的误差模型. 对一般光滑曲线的误差, 除定性的描述外, 尚未见到明确的误差定义. 下面首先建立测量线与真实线上点的对应关系, 并以对应点的误差对弧长的积分作为整条曲线的误差. 考虑到测量线节点个数以及节点坐标的随机性, 最终以一维随机过程作为描述曲线误差的数学模型.

在矢量GIS 中, 曲线是用折线近似表达的, 折线上的节点一般不会恰好在曲线上, 产生误差是

必然的, 距离是误差的最直接的体现. 下面先将点的距离推广到曲线上.

如图 1 所示, 虚折线  $l$  为测量线, 实曲线  $L$  为真实线. 过真实线上点  $P$  作曲线的法线, 法线与测量线的交点为  $N$ , 若曲线是封闭的, 则  $N$  与  $P$  一一对应; 若曲线不封闭, 在曲线端点附近可能会出现无法对应的情况, 这时只需对折线进行适当的延长或截取, 仍可以建立对应关系. 一条折线通常由相当多节点构成, 端点的延长或截取对整个曲线误差影响是微弱的.

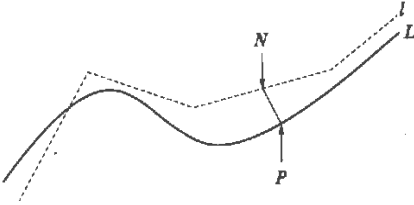


图 1 点的对应关系

Fig. 1 Corresponding points between two curves

对应点之间误差的定义可以仿照有向曲线的定义方法, 定义曲线的侧, 并分别以正负号表示  $N$  点位于曲线的右侧和左侧.

定义 1 设线段  $PN$  的长为  $d$ , 称 
$$Z = \begin{cases} -d, & P \text{ 在 } l \text{ 的左邻域} \\ d, & P \text{ 在 } l \text{ 的右邻域} \end{cases}$$

为点  $N$  与点  $P$  的误差. 若  $P$  点的坐标为  $P(x(s), y(s))$ , 对应点  $N$  的坐标为  $N(a(s), b(s))$ ,  $N$  与

$P$  的误差为  $Z(s)$ , 其中参数  $s$  为真实线的弧长. 则真实线与测量线的整体误差可以定义为

**定义 2** 称  $Z(s)$  对弧长的积分  $\int_0^L Z(s)ds$  为曲线  $L$  与  $\bar{L}$  的平均误差, 称  $\int_0^L |Z(s)|ds$  为  $L$  与  $\bar{L}$  的总误差.

测量线与真实线所夹无意义多边形的面积亦能较好地反映两曲线的整体贴近程度. 容易证明, 定义 2 给出的总误差, 与无意义多边形的面积相当接近, 其差的绝对值不超过  $\frac{1}{2}kZ^2$ , 这里  $k$  为曲线的切线转过的角度,  $Z$  为  $Z(s)$  的最大值. 由于测量线节点个数以及节点坐标是随机的, 因此  $Z(s)$  实际上为随机变量.

**定义 3** 称随机过程  $\{Z(s), 0 \leq s \leq l\}$  为真实曲线的误差过程, 称  $m(s) = EZ(s)$  为曲线的平均误差函数; 称  $M(s) = E|Z(s)|$  为曲线的中误差函数; 称误差过程的均方积分的数学期望 (见文献 [3])  $\sigma_T = E \int_0^L |Z(s)|ds$  为曲线的总误差. 总误差的几何意义为: 对曲线测量或数字化时, 会产生无意义多边形, 总误差在数值上近似等于无意义多边形面积的理论平均值.

## 2 道路曲线误差过程的数字化试验

### 2.1 试验条件

取一图幅为  $44.5\text{ cm} \times 38.6\text{ cm}$ 、比例尺为  $1:50\,000$  的地形图进行扫描, 用 **Mpinfo** 采集图中一段道路曲线, 以采集点作为样条插值的型值点对道路曲线进行高密度的插值 (原密度的 20 倍). 插值后的道路曲线几乎是光滑的, 如图 2 所示, 将各插值点看作曲线的理论坐标. 利用 **Photoshop** 擦除扫描图像中相应的公路, 以插值所得公路曲线代之, 重新设定图廓点并用绘图仪输出图形, 以此图作为数字化试验用图. 扫描试验用图并保存为 **gif** 格式.



图 2 插值道路曲线

Fig. 2 The road given by cubic spline

依照上述过程设定的试验用图具有以下优点: ①尽可能地剔除了原图本身的误差, 试验所得误差主要由数字过程引起; ②道路曲线上各点的坐标是已知的, 便于分析; ③插入公路曲线的色彩、宽度等与原图基本一致, 使得模拟条件与数字

化操作的实际情况较为接近, 能较客观地考察数字化误差的大小和分布规律.

按照上述过程进行的实验, 误差的主要部分来自于数字化的随机误差, 包括控制点的定位误差和采集过程中的随机误差. 由于只对一张图进行一次扫描, 没有反复, 因此, 由图纸变形引起的误差以及扫描误差、绘图误差等将以系统误差的形式作用于各采集点.

### 2.2 试验过程

用 **Mpinfo** 建立一表并导入上述 **gif** 格式文件, 对插值的公路进行采集并保存为 **Mf** 格式. 反复这一过程, 得到 100 个 **Mf** 文件.

### 2.3 样本曲线的提取

公路曲线是用三次参数样条函数插值得的, 线上各点的坐标、切线斜率等都可以求出, **Mf** 文件记录了数字化折线上各节点的坐标, 因此, 根据一个 **Mf** 文件不难计算出一条样本曲线  $Z(s)$  来. 图 3 为一条样本曲线的一部分 (靠近曲线上端 40 个节点), 为清晰起见, 图中纵坐标  $Z(s)$  放大了 20 倍. 样本曲线位于横轴上方部分所对应的数字化线在真实线的右边, 同样, 横轴下方曲线对应的数字化线位于真实线的左边.



图 3 部分样本曲线

Fig. 3 A piece of sample curve

## 3 误差指标分析

按下面公式, 不难对各误差指标作出估计. 平均误差函数:

$$\hat{m}(s_j) = \frac{\sum_{i=1}^{100} Z_i(s_j)}{100} \quad (1)$$

中误差函数:

$$M(s_j) = \frac{\sum_{i=1}^{100} |Z_i(s_j)|}{100} \quad (2)$$

总误差:

$$\hat{\sigma}_T = \frac{1}{100} \sum_{i=1}^{100} \sum_{j=1}^{520} |Z_i(s_j)| (s_j - s_{j-1}) \quad (3)$$

式中:  $s_j$  为曲线弧上第  $j$  段的弧长, 521 个点组成 520 个弧段. 经计算, 总误差为  $0.74\text{ cm}^2$ .

对于平均误差函数和中误差函数, 限于篇幅, 这里只给出 8 个点的误差指标值, 见表 1.

图 4(a), (b) 分别为中误差函数与平均误差函数的图形, 图中只画出前 40 个节点, 且纵坐标放大了 20 倍.

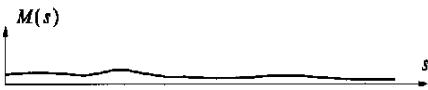
表 1 点的误差指标

Tab. 1 Error characteristics of points

点号	曲率	$m(s)$	$M(s)$
1	0.017 937	-0.01 028	0.012 0
2	0.068 520	-0.00 338	0.006 7
3	0.115 756	-0.01 850	0.020 7
4	0.360 270	-0.012 07	0.014 0
5	0.454 926	0.011 53	0.011 9
6	1.127 055	0.026 71	0.027 1
7	2.332 869	-0.021 09	0.022 3
8	3.731 225	-0.024 69	0.025 1



(a) 平均误差函数



(b) 中误差函数

图 4 误差函数  
Fig. 4 Error functions

将误差函数曲线与原公路线相比较就会发现,数字化线明显偏向真实线凹向的一侧,原因可能是:①公路线条较粗,不易瞄准中心线;②数字化员习惯于以一种“内接”的方式采集曲线,虽然节点已尽可能地接近曲线,但节点之间的部分却有偏向曲线凹向侧的趋势,这一偏差在比例尺较小的情况下分辨不出来;③曲线采集的密度较小. 误差函数与平均误差函数是由 100 条样本曲

线估计出的,误差的正负和大小明显地与真实曲线的形状有关,这一误差已不能视为随机误差,而应当将其看作由数字化员的习惯引起的一种系统误差.数字化员可通过放大比例尺的办法,检查偏差的大小,及时调整采集方式或习惯,尽可能地削弱这一误差. 从误差随机过程的角度看,过程的一阶中心矩及原点矩都不是平稳的.不同点的误差之间有较强的统计相关性. 中误差函数有较明显的几何意义,函数曲线与横坐标轴所围图形的面积近似等于数字化线与真实线所围的无意义多边形的面积.

4 结论

- (1)GIS 中曲线元素沿法线方向的误差可用一维随机过程描述,误差函数与平均误差函数较好地表达了曲线上不同点的误差特性.总误差从整体上度量了数字化线与真实线的差异,有较明显的几何意义.
  - (2)误差过程是一非平稳过程,不同点的误差之间是不独立的,有较强的相关性.
  - (3)在曲线数字化过程中,系统误差的作用较为明显,它与曲线的凹凸方向以及数字化员的作业习惯有关.数字化员应及时调整作业习惯以提高数据采集精度.
- 参考文献:
- [ 1 ] 李德仁,彭美云,张菊清.GIS 中线要素的定位不确定性模型研究[J].武汉测绘科技大学学报,1995,20( 4 ):283~288.
  - [ 2 ] 史文中,刘文宝.GIS 中线元位置不确定性的随机过程模型[J].测绘学报,1998,27( 1 ):37~44.
  - [ 3 ] 唐鸿龄,张元林,陈浩球.应用概率[M].南京:南京工学院出版社,1988.

The Overall Error Model of Road Curve in GIS and Its Digitization Test

GUO Tong -de<sup>1</sup>, DU Yun -hai<sup>2</sup>, LU Yan<sup>1</sup>

( 1.College of Environmental & Hydraulic Engineering ,Zhengzhou University ,Zhengzhou 450002,China ;2.Department of Mechanical Engineering ,Zhengzhou University ,Zhengzhou 450002)

Abstract : Based on the stochastic process theory the error model of curves in GIS is established .Two characteritics concerning the error of points on a curve are proposed in terms of the stochastic process .And the total error of the curve is expressed by a mean square integral of a stochastic process .In order to check the efficiency of the model proposed ,a scan digitizing test is designed .In the test ,a piece of road is digitized for times and 100 sample functions are derived .The result of the test shows that the system error in digitized map isn't negligible .The errors of points are chiefly dependent on the concavity and the curvature of the road curve being tested .

Key words : curve ;stochastic process ; mean square integral ; mean error function ; overall error