

国家自然科学基金资助项目

编号 40101022

矢量GIS中的不确定性 理论及应用

戴洪磊 著

地震出版社

国家自然科学基金资助项目（编号 40101022）

矢量 GIS 中的不确定性理论及应用

戴洪磊 著

地震出版社

图书在版编目(CIP数据)

矢量 GIS 中的不确定性理论及应用/戴洪磊著. —北京:地震出版社, 2004.9

ISBN 7-5028-2561-4

I. 矢… II. 戴… III. 地理信息系统 IV. P208

中国版本图书馆 CIP 数据核字 (2004) 第 086810 号

地震版 XT200400259

内容简介 本书首先介绍了 GIS 所能解决的基本问题以及空间数据质量问题、不确定性问题和相关研究背景;其次论述了从事矢量空间数据不确定性研究所依据的统计数学基础;然后,详细探讨了矢量 GIS 中点、线、面目标的位置数据不确定性的度量问题;此外,通过衡量位置数据不确定性的综合模型和位置数据不确定性在典型 GIS 空间分析中的传播说明了位置不确定性理论的应用问题。最后,总结了本书中的主要研究成果,并给出了该领域中需要进一步研究的问题和方向。本书可供从事地理信息系统理论研究和应用开发的科研和技术人员、管理人员,以及大专院校师生参考。

矢量 GIS 中的不确定性理论及应用

戴洪磊 著

责任编辑:陈晏群

责任校对:庞娅萍

出版发行:地震出版社

北京民族学院南路 9 号

邮编:100081

发行部:68423031 68467993

传真:88421706

门市部:68467991

传真:68467991

总编室:68462709 68423029

传真:68467972

E-mail: seis@ht.rol.cn.net

经销:全国各地新华书店

印刷:北京地大彩印厂

版(印)次:2004 年 9 月第一版 2004 年 9 月第一次印刷

开本:787×1092 1/16

字数:218 千字

印张:8.5

印数:001~500

书号:ISBN 7-5028-2561-4/P·1204 (3187)

定价:25.00 元

版权所有 翻印必究

(图书出现印装问题,本社负责调换)

前 言

“数字地球”概念的提出推动了 GIS 的迅速发展。作为空间信息应用技术, GIS 已被广泛应用于社会、经济、环境和工程等多个领域的建模和辅助决策中, 例如城市管理、区划、环境和交通整治等。作为空间信息应用学科, GIS 的许多相关理论问题也引起广泛重视和深入研究, 例如空间数据误差与不确定性、空间认知和空间关系理论等。本书主要探讨空间数据的不确定性理论及实践问题。

本书共分八章。第一章介绍了 GIS 所能解决的基本问题和空间数据质量问题, 以及不确定性问题和相关研究背景。第二章介绍了从事矢量空间数据不确定性研究所依据的统计数学基础。第三至五章分别探讨了矢量 GIS 中点、线、面目标的位置数据不确定性的度量问题。第六章在介绍了衡量位置数据不确定性的综合模型和影响因素理论后, 探讨了位置不确定性理论的应用问题。第七章讨论位置数据不确定性在典型 GIS 空间分析中的传播问题。第八章简单总结了本书中的主要研究成果, 并给出了该领域中需要进一步研究的问题和方向。

本书系国家自然科学基金资助项目“遥感与 GIS 集成中空间实体随机表达的理论和方法”(编号 40101022)的部分研究成果。同时, 也吸收了国家自然科学基金资助项目“矿山三维动态 GIS 几何数据的误差理论”(编号 49801016)和“GIS 图形数据位置不确定性理论问题的研究”(编号 49671063), 以及煤炭院校优秀青年基金项目“矿山空间数据质量的可视化理论”(编号 97-039)等课题组的部分研究成果。另外, 本书中所研究的问题和内容与作者的硕士和博士两篇学位论文密切相关。借此机会, 向我的指导教师山东科技大学卢秀山教授、徐泮林教授、刘文宝教授, 武汉测绘科技大学杜道生教授致以衷心的感谢! 此外, 非常感谢鼓励、关心和支持我学业发展的陶华学教授、陶本藻教授、黄杏元教授、欧吉坤教授和陈兰森先生等。

本书可供从事地理信息系统理论研究和应用开发的科研和技术人员、管理人员，以及大专院校师生参考。

作者受水平和时间所限，书中错误难免，望读者批评指正。

作者

2004年8月1日

目 录

第一章 绪论	(1)
§ 1.1 地理信息系统 (GIS) 简介	(1)
1.1.1 地理信息系统的内容与分类	(1)
1.1.2 地理信息系统的应用范围	(1)
§ 1.2 空间数据模型	(2)
1.2.1 栅格数据结构	(2)
1.2.2 矢量数据结构	(3)
§ 1.3 空间数据	(4)
1.3.1 位置坐标	(4)
1.3.2 属性数据	(4)
1.3.3 空间关系	(4)
§ 1.4 空间数据的不确定性	(5)
1.4.1 不确定性概念	(5)
1.4.2 不确定性分类	(6)
1.4.3 不确定性来源	(6)
1.4.4 国内外研究现状	(10)
1.4.5 位置数据可靠性	(12)
1.4.6 位置数据不确定性在叠置操作中的传播	(12)
§ 1.5 本书的范围和结构	(12)
1.5.1 本书研究的特点	(12)
1.5.2 本书各章研究的内容简介	(13)
第二章 数学基础	(14)
§ 2.1 随机变(向)量	(14)
2.1.1 基本概念	(14)
2.1.2 分布函数与分布密度	(14)
2.1.3 数字特征	(15)
2.1.4 随机向量	(15)
§ 2.2 随机过程	(17)
2.2.1 概率结构	(17)
2.2.2 数字特征	(17)

§ 2.3 随机场	(18)
2.3.1 基本概念	(18)
2.3.2 数字特征	(19)
§ 2.4 本章小结	(20)
第三章 随机点位不确定性的度量	(21)
§ 3.1 随机点的分布函数与分布密度	(21)
§ 3.2 点位协方差阵	(22)
§ 3.3 点位误差数值指标	(23)
3.3.1 点位方差	(23)
3.3.2 任意方向 φ 的位差	(23)
§ 3.4 平面及三维随机点的误差曲线和误差椭圆及误差椭球	(24)
3.4.1 二维点位误差曲线	(24)
3.4.2 点位误差椭圆	(24)
3.4.3 点位误差椭圆与点位误差曲线的比较	(24)
§ 3.5 点位不确定性的可视化度量指标	(26)
3.5.1 一维随机点的可视化度量指标	(26)
3.5.2 二维随机点的可视化度量指标	(27)
3.5.3 三维随机点的可视化度量指标	(29)
§ 3.6 本章小结	(31)
第四章 随机线位不确定性的度量	(33)
§ 4.1 随机线元的分布函数与分布密度	(33)
4.1.1 三维随机线元的分布函数和分布密度	(33)
4.1.2 二维随机线元的分布函数和分布密度	(35)
§ 4.2 线位相关函数阵	(36)
§ 4.3 平面随机线位误差带的解析及可视化模型	(36)
4.3.1 线元上任意点的位差公式	(36)
4.3.2 误差椭圆参数	(38)
4.3.3 误差椭圆方程	(39)
4.3.4 广义误差带解析方程	(40)
4.3.5 “g—带”形状特征讨论	(45)
4.3.6 算例分析	(46)
§ 4.4 平面随机线位置信带与可靠度带	(47)
4.4.1 随机线位置信带“C—带”(Confience Band)	(47)
4.4.2 随机线位可靠度带“R—带”(Reliability Band)	(48)

§ 4.5 平面随机线元的误差曲线带模型	(49)
4.5.1 线位误差曲线带	(50)
4.5.2 算例分析	(52)
4.5.3 结论与建议	(52)
§ 4.6 三维随机线元误差椭球域的解析及可视化模型	(53)
§ 4.7 本章小结	(55)
第五章 随机面位不确定性的度量	(57)
§ 5.1 平面随机面元的分布函数和概率密度	(57)
§ 5.2 平面随机面位误差环的解析及可视化模型	(59)
5.2.1 面位误差环定义	(59)
5.2.2 随机面位误差环的构成机理	(59)
5.2.3 面位误差环“g—环”的解析表达式	(60)
5.2.4 算例	(62)
§ 5.3 随机面位置信环和可靠度环模型	(65)
5.3.1 面位置信环“C—环”(Confience Donut)模型	(65)
5.3.2 面位可靠度环“R—环”(Reliability Band)模型	(65)
§ 5.4 随机面位误差环的可视化度量指标	(65)
5.4.1 前提假设	(65)
5.4.2 精度最高点落入它本身误差椭圆的概率	(65)
5.4.3 面位误差环度量指标	(66)
§ 5.5 本章小结	(67)
第六章 位置数据不确定性度量的综合模型及其应用	(68)
§ 6.1 平面及三维随机曲线的分布函数与分布密度	(69)
6.1.1 随机曲线的分布函数和概率密度	(69)
6.1.2 平面二维曲线的分布函数和概率密度	(70)
§ 6.2 平面随机曲线线位误差带模型	(71)
6.2.1 随机曲线的统计模型	(71)
6.2.2 随机曲线的误差模型	(72)
6.2.3 随机曲线长度方差估计	(74)
6.2.4 算例分析	(74)
6.2.5 结论	(75)
§ 6.3 平面随机折线线位误差带模型	(77)
6.3.1 随机折线误差带概念	(77)
6.3.2 随机折线误差带边界线方程推证	(77)
6.3.3 算例分析和讨论	(79)

6.3.4	结论	(81)
§6.4	几何图形数据的可靠性评价	(82)
6.4.1	制图要素位置特征的数学描述	(82)
6.4.2	图形数据的可靠性	(83)
6.4.3	应用实例	(86)
6.4.4	结论与建议	(88)
§6.5	误差带模型在区域边界线确定中的应用	(88)
6.5.1	目标之间关系的判断	(89)
6.5.2	点与线段目标之间关系的判断	(90)
6.5.3	点与面目标之间关系的判断	(91)
6.5.4	不确定性在区域边界线确定中的应用	(93)
§6.6	本章小结	(94)
第七章	GIS 空间叠置中位置不确定性的传播	(95)
§ 7.1	多边形叠加	(95)
§ 7.2	GIS 叠置前后同名点元的方差估计	(96)
7.2.1	同名坐标观测值矩阵和观测权阵	(96)
7.2.2	两层叠置前后同名点元的方差估计	(97)
7.2.3	多层叠置前后同名点元的方差估计	(99)
§ 7.3	GIS 叠置后同名线元的方差估计	(101)
7.3.1	多层叠置后同名线元端点的方差估计	(101)
7.3.2	多层叠置后同名线元的精度估计	(103)
7.3.3	多层叠置后同名线要素长度的方差估计	(104)
§ 7.4	GIS 叠置后同名面元的方差估计	(106)
7.4.1	同名面元物理特征的方差估计	(106)
7.4.2	同名面元几何特征的方差估计	(108)
§ 7.5	多边形叠置中位置误差的传播模型	(110)
7.5.1	叠置后的交点坐标及精度	(110)
7.5.2	交点多边形的误差状态	(113)
§ 7.6	本章小结	(116)
第八章	总结与展望	(118)
8.1	回顾总结	(118)
8.2	研究展望	(119)
参考文献		(121)

第一章 绪 论

§ 1.1 地理信息系统 (GIS) 简介

1.1.1 地理信息系统的内容与分类

地理信息系统 (Geographical Information System, GIS) 是决策支持系统之一。作为一门新兴的交叉学科, 它主要是研究描述、存储、分析和输出空间信息的理论和方法; 作为一个技术支持系统, 它主要用于从地理空间数据库中提取和传输多元空间信息, 并采用相关空间分析模型, 完成辅助空间决策任务。地理信息系统将现实世界中的地理要素和地理现象简化为一系列由空间位置参考信息和非位置信息两部分组成。

按其内容和应用范围不同, 地理信息系统可以分为三大类: 专题地理信息系统, 区域信息系统和通用地理信息系统工具。而一般所指的地理信息系统常常是指通用的地理信息系统工具或软件平台。一个完整的 GIS 主要由计算机硬件系统、软件系统、空间数据、系统管理与操作人员五部分构成。GIS 的核心部分是计算机软、硬件系统, 而空间数据则是地理内容在 GIS 中的反映, 用户和管理人员则决定系统的工作方式和维护系统的正常运行。

GIS 的应用虽然千差万别, 但其所解决的主要问题可归结为与空间位置相关的五种类型:

- (1) 位置问题: 在某个特定的位置有什么?
- (2) 条件问题: 在何处有满足某些条件的东西?
- (3) 趋势问题: 在某处存在何种地理现象及该现象呈现何种空间、时间变化趋势?
- (4) 模式问题: 已经发生或正在发生的事件与哪些因素有关?
- (5) 模型问题: 采用什么模型来实现最优决策过程?

上述任务在 GIS 中是通过不同的功能来实现的。虽然各种商用 GIS 软件包的侧重点不同, 并且它们在实现这些功能时所采用的技术也不一样, 但大多数商用 GIS 软件包都具备下列功能: 数据的获取系统、数据的初步处理系统、数据的存储及检索系统、数据的查询与分析系统和图形的显示与交互。

1.1.2 地理信息系统的应用范围

显然, 地理信息系统从节省人力、物力、财力和缩短系统建立周期等方面的考虑, 决定了对通用地理信息系统工具平台的强烈需求。当前, 有许多平台式 GIS 软件可供选择。例如, 美国 ESRI 公司 (Environmental Systems Research Institute Inc.) 的系列产品, 包括运行于 UNIX/Windows NT 平台上的 Workstation ArcInfo 和 Desktop ArcInfo (ArcGIS), 以及运行于 Windows 平台上的 ArcView。Workstation ArcInfo 具有最基本和必要的 GIS 功能, 包括数据

录入和编辑、投影变换、制图输出、查询分析及其分析功能（缓冲区分析、叠加复合分析等等），栅格数据的分析功能，支持栅格矢量一体化查询和叠加显示，并提供了二次开发语言 AML 以及开放开发环境 ODE，供用户通过扩展通用平台来建立自己的 GIS 应用系统。此外，对于一些具有专业化特色，并广泛应用的专门功能，也提供了相应的扩展模块，例如 TIN（基于不规则三角网的地表模型生成、显示和分析模块，可以根据等高线、高程点、地形线生成 DEM，并进行通视、剖面、填挖方计算等），GRID（栅格分析处理模块，可以对栅格数据进行输入、编辑、显示、分析、输出，其分析模型包括基于栅格的市场分析、走廊分析、扩散模型等）和 NETWORK（网络分析模块，提供了最短路径选择、资源分配、辖区规划、网络流量等功能，可以应用于交通、市政、电力等领域的管理和规划）等。ArcGIS 包括 Arc Catalog、Arc Toolbox 和 Arc Map 三部分。其中 Arc Catalog 用于元数据的定位、浏览和管理空间数据；Arc Toolbox 是由常用数据分析处理功能组成的工具箱；Arc Map 用于地图数据的显示、查询和分析。ArcView GIS 是 ESRI 的桌面 GIS 系统，能够对地图数据、结构化的属性数据、统计图、地图图面配置、开发语言等多种文档进行管理，拥有二次开发语言 Avenue 和一些常用的空间分析“插件”扩展模块，例如 Spatial Analyst（栅格数据的建模分析）；Network Analyst（网络分析）；和 3D Analyst（利用 DEM 实现三维透视图的生成）等。

§ 1.2 空间数据模型

自然界中的空间实体可抽象为点、线、面和曲面等多种类型。要使计算机能识别和处理它们，必须对这些特征实体进行数据表达，即建立一定的数据模型。在表达特征实体时，应注意到点是构成空间实体的基本元素。若用一个没有大小的点坐标来表达基本点元素时，称矢量表示法，即点、线和面，其中点不占有面积，线和面由一系列内部相关联的坐标形成，一定的面或线则能与一定的属性连接。若采用一个有固定大小的像元来表达基本点元素时，称为栅格法，即一系列 x, y 坐标定位的像元，每个像元独立编码，并载有属性值。相应地，表达空间实体的数据模型可分为矢量数据模型和栅格数据模型。

1.2.1 栅格数据结构

栅格数据结构实际就是每个像元位置由行列确定的像元阵列。点实体由聚集在一起的相邻像元结合表示。栅格结构按一定的规则排列决定了所表示的实体位置很容易隐含在网络文件的存储结构中。根据每个存储单元在文件中的记录位置可方便地得到其行列位置，这种行列坐标可以很容易地实现坐标转换。网络文件中的每个代码本身明确地代表了实体的属性或属性的编码。

栅格数据是量化和近似离散的数据，它所表示的地表是不连续的。因此，像元所表达的地表一定面积内的空间数据是近似的，是按某种规则在像元内提取的值，例如平均值、主成分值等。另外，栅格大小与地表相应单元大小之比视作栅格数据的比例尺。这种比例尺的大小对长度、面积等的度量有较大影响。此外，计算长度、面积的方法也对长度和面积的度量等有较大影响。一般像元越大，这种误差越增加。

在决定代码时要尽可能地保证最大的信息容量，以便保持表达地表的真实性。为了保证

栅格数据的精度, 当在同一栅格下的原图对应了几种不同的属性值时, 由于每一个单元只能取一个值, 故需要设计一定的取值方法赋值。常用的取值方法有中心点法、面积占优法、长度最占优法和重要性法。中心点法也叫做“网格交点归属法”, 是根据栅格元中心处的地物类型或现象特性决定栅格代码的, 此法适合于具有连续分布特性的地理要素, 如降雨量分布、人口密度图等。面积占优法是以占栅格面积最大的地物类型或现象特征决定栅格单元的代码, 这种方法最适合分类较细、地物类别斑块较小的情况。长度最占优法是以覆盖的格网过中心部位时, 横线占据该格中的大部分长度的属性值定为该栅格单元的代码。重要性法是根据栅格内不同地物的重要性, 选取最重要的地物类型决定相应的栅格单元代码。这种方法常用于具有特殊意义而面积较小的地理要素, 例如城镇、交通枢纽、河流水系等, 以及适合保留特别重要的地理实体, 例如稀有金属矿产区域等。

此外, 提高栅格数据精度的另一个途径是缩小单个栅格单元的面积, 即增加栅格单元总数。随着行列数的增加, 每个栅格单元可代表更为精细的地面矩形单元。于是, 混合单元减少, 量算精度可以大大提高, 从而能表现更细小的地物类型, 逼近真实形态。所带来的问题是大量数据冗余。为此, 又发展了一系列栅格数据压缩编码方法, 如键码、游程长度编码、块码和四分树编码等。

1.2.2 矢量数据结构

矢量数据结构是通过记录坐标的方式表达点、线、面地理实体。与栅格数据结构那样需要进行量化处理不同, 将矢量坐标空间假定为连续空间。矢量数据结构的优点是能更精确地定义位置、长度和大小。其数据存储是以隐式关系以最小的存储空间存储复杂的数据。

1.2.2.1 点实体

点实体的位置由单独的 x, y 坐标对来表达, 属性数据由文本、数字和制图符号等描述。点是基本的地理特征要素, 是空间上不可再分的地理实体。点既可以是抽象的, 也可以是具体的, 例如地物点、线段网络的结点或文本位置点等。若点是一个与其他信息无关的符号, 则应记录符号类型、大小、方向等有关信息。若点是文本实体, 则应记录字符大小、字体、排列方式、比例、方向以及其他非图形属性的联系方式等信息。

1.2.2.2 线实体

线实体是由两对以上的 x, y 坐标定义的直线元素组成的各种线性要素。最简单的线实体只需要存储它的起止点坐标、属性和显示符等。而在输出时既可以用实线, 也可以用虚线等符号信息描绘。利用弧和链等 n 个坐标对的集合, 可以表达任何连续而复杂的曲线。显然, 选取线元越短, x, y 坐标数量越多, 越能精确地表达复杂曲线, 但这样增加了数据存储量和数据处理工作量。线实体主要用于表达线状地物, 例如公路、水系、山脊线、符号线和多边形边界等。线实体有时也称为“弧”、“链”、或“串”等, 其矢量编码包括惟一标识码、线标识码、起始点、终止点、坐标对序列、显示信息和非几何属性。

空间分析中涉及的与线实体直接相关的一大类是网络实体, 例如供排水网、道路网和水系。简单线或链中携带的彼此互相连接的空间信息是各种网络分析中的必要信息。所以, 在数据结构中要建立以结点为基础的指针系统, 以便让计算机在复杂的线网结构中逐线跟踪每一条线。比如在建立水系网络时, 这种指针系统可以表达每条支流之间的连接关系。指针系统包括了结点指向线的指针, 并使每条从结点出发的线汇于结点处的角度, 以便完整地定义

线网络的拓扑关系。

1.2.2.3 面实体

面实体是指区域空间实体。具有名称属性和分类属性的区域实体由多边形数据描述,例如行政区、土地类型、植被分布等;而具有标量属性的区域空间实体也用等值线描述,例如地形、降雨量等。

多边形数据模型比较复杂,因为除了要表达位置和属性外,还要表达区域的拓扑关系特征,例如形状、邻域和层次结构等,以便进行空间分析、专题图资料的分析和操作。为了对多边形数据进行编码,多边形网需要满足一定要求:①区域多边形要有惟一的形状、周长和面积;②空间分析要求的数据结构应像水系网中记录连接关系一样能够记录每个多边形的邻域关系;③专题图上的多边形可能是多边形内嵌套小的次一级的多边形,并不都是同一等级的多边形,例如在土地利用图上湖泊的水域线是个岛状多边形,而湖中的岛屿则为“岛中之岛”。

§ 1.3 空间数据

空间数据模型是通过空间数据来表达空间实体的。空间数据是 GIS 用户或操作人员利用 GIS 的软、硬件系统生产图表等应用产品的原料。GIS 所涉及的空间数据主要是指以地球表面空间位置为参照的自然、人文、社会、经济和景观等信息,这种信息可以是数字、图像、图形、表格和文字等。空间数据的获取方式分为两大类:①数据直接来源于野外测量、地面调查、相片和遥感影像等;②数据来源于现有的文件,如地图、图表等。第二类空间数据则是由系统建立者通过扫描仪、数字化仪、键盘、磁带机或其他系统通讯录入 GIS 的。概括地讲, GIS 的空间数据主要包括:位置坐标、属性数据和空间关系。

1.3.1 位置坐标

用于表达地理目标在自然界或包含某个区域的地图中的空间位置,可以是高斯平面直角坐标、极坐标或经纬度等。采用屏幕或数字化仪输入时通常采用数字化仪直角坐标或屏幕直角坐标。

1.3.2 属性数据

属性是与地理实体相联系的地理变量或地理意义,可分为定性和定量两种,一般经过抽象的概念,通过分类、命名、量算、统计得到。定性属性包括名称、类型、特性等,例如土壤种类、行政区划等;定量属性包括数量和等级,例如面积、长度、土地等级和人口数量等。任何地理实体至少有一个属性,而地理信息系统的分析、检索和表达主要是通过属性的操作运算实现的。

1.3.3 空间关系

地理实体间的空间关系包括:拓扑关系、度量关系和方位关系等。拓扑关系是 GIS 分析中最基本的关系,它定义了地物之间连通、邻接等关系,包括网络结点与网络线之间的枢纽

关系。度量关系定义了两个空间实体间的距离远近。方位关系定义了两个地物之间的相对方位。拓扑关系比较复杂,不仅包括了实体与实体间的拓扑关系,有时还存在面实体与岛或内部点的包含关系等。空间拓扑关系是地理信息系统特色之一,对于地理空间数据的编码、格式转换、存储管理、查询检索和模型分析都有重要意义。

空间数据的来源和数据类型繁多,但主要的类型有:①地图数据,包括普通地图和专题地图;②影像数据,包括卫星遥感、航空遥感和雷达遥感影像等数据;③地形数据,主要是地形等高线图的数字化和已建立的数字高程模型与其他实测的地形数据等;④属性数据,包括各类调查报告、实测数据、文献资料和解译信息等;⑤数据说明,包括由各类纯数据通过调查、推理、分析和总结得到的有关数据的数据,例如数据来源、数据权属、数据产生的时间、数据精度、数据分辨率、源数据比例尺、数据转换方法等。

§ 1.4 空间数据的不确定性

在 GIS 中,数据是最基本和最重要的组成部分,其投入费用总额占到系统建立和维护的 70% 左右。GIS 数据质量的优劣直接影响着系统应用分析结果的可靠程度和系统应用目标是否能真正实现。由于 GIS 是基于应用而形成的一门地球信息学科,以往的研究重点集中在系统的建立、功能的改善和应用方面,在 GIS 初步形成之后,才提出了建立 GIS 基础理论的课题,其中数据质量中的空间数据不确定性理论被列为 20 世纪 90 年代重点研究的课题之一。GIS 空间数据不确定性理论的研究,是 GIS 中极为重要的基础理论课题之一,它对确定 GIS 录用数据质量标准,评价和控制 GIS 产品质量,优化空间数据分布结构,改善 GIS 算法,减少 GIS 设计与开发的盲目性,以及 GIS 的其他研究领域都有深远影响。

地理信息是对现实世界的抽象和表达。由于现实世界的无限复杂性和模糊性,以及人类认识和表达能力的局限性,这种抽象和表达总是不可能完全达到真实值,而只能在一定程度上接近真值。而且,真值往往是不可知的或不可预测的,因此误差总是存在着。

支持 GIS 各种应用的是经分析和处理后提供给用户的文档、图和表等不同类型的 GIS 产品。由于生产 GIS 产品的“原料”——GIS 原始录用数据本身包含着不可避免的误差,描述数据的模型也只能是客观实体的一种近似,并且在 GIS 产品的“生产”过程中,各种空间操作和处理也会引入新的误差和不确定性。因此,由 GIS 生成的各种精美图件与其内在质量不相符合。

GIS 空间数据误差处理和分析正是针对上述背景而提出的研究课题。所研究的主要对象是 GIS 数据中的固有误差和操作处理中产生的误差;内容为这些误差的性质、度量和传播,以便建立一套误差分析和处理理论体系,使未来的 GIS 在向用户提供产品的同时,附带提供该产品的质量指标。

1.4.1 不确定性概念

不确定性可以看作一种广义的误差,既包含了可度量的和不可度量的误差,又涵盖了数值和概念上的误差。一般而言,不确定性是指被测量对象知识缺乏的程度,通常表现为随机性(Randomness)和模糊性(Fuzziness)。从信息论的角度看,不确定性具有多方面的含义,

数据的误差、数据和概念的模糊性及不完整性都可视作不确定性的内容。例如在测量中,当强调不确定性的统计内涵时,测量工作者常习惯于将不确定性称为观测误差(Observation Error),而在强调不确定性的抽象特征时,地理工作者直接称为不确定性(Uncertainty)。近年来人们普遍认为,“不确定性”是一个比“观测误差”含义更广的概念,并越来越多地倾向于使用前者。可认为它是真值的不能被肯定的程度,是自然界各种实体或现象本身固有的属性,在形式上它一般是包含了真值的一个范围,这个范围越大,数据的不确定性就越大。

1.4.2 不确定性分类

1.4.2.1 空间定位不确定性

有时也称空间位置不确定性,是指 GIS 中被描述物体的坐标数据与其地面上真实位置间的接近程度。这种不确定性通常以空间三维坐标数据精度来表示,包括数学基础精度、平面精度、高程精度、接边精度、形状再现精度(形状保真度大小)和像元定位精度(分辨率)等。平面精度和高程精度又可分为相对精度和绝对精度。例如,若平面点 $Z_0(x_0, y_0)$ 的真值为 $\tilde{x}_0, \tilde{y}_0$, 或然值为 $\hat{x}_0, \hat{y}_0$, 则 $\Delta x_0 = \hat{x}_0 - \tilde{x}_0$, $\Delta y_0 = \hat{y}_0 - \tilde{y}_0$ 则为平面点 $Z_0(x_0, y_0)$ 的位置不确定性,通常以二维或三维坐标值的方差来度量。

1.4.2.2 属性不确定性

属性不确定性是指实体的属性值与其真实属性相符的程度。属性不确定性通常取决于数据的类型,且常常与位置精度有关,包括要素分类与代码的正确性、要素属性值的正确性及名称的正确性等几个方面。例如,若地面上某一块用地本应是绿化地,但在 GIS 中被描述为建筑用地,则为类别上的属性不确定性。此外,还有描述程度上的不确定性,例如某一地区的绿地覆盖率为 60%,但在 GIS 中被描述为 80%,这是属性量化上的不确定性。

1.4.2.3 时域不确定性

指在描述地理现象时,具有时间含义的描述上的差别。主要为数据的现势性,一般通过数据采集时间和数据更新的时间和频度来表现。例如,某一宗地划归为某一用户的时间是从 1985 年开始到现在,但在 GIS 中误记为 1975 年开始到现在。

1.4.2.4 逻辑一致性

指数据关系上的一致性,包括数据结构、数据内容、空间属性和专题属性,尤其指拓扑关系上的内在不一致性。例如,多边形的不闭合问题,多边形编号丢失或一个多边形对应一个以上的编号等。

1.4.2.5 数据完整性

指地理数据在范围、内容及结构等方面是否达到规范要求,包括数据范围、分层、实体类型、属性数据和名称等方面的完整性。这涉及到,对于给定的目标,GIS 在什么程度上尽可能完整地表达了该实体,这与性质与规则的选择、地图综合与比例尺等诸多因素有关。

1.4.3 不确定性来源

1.4.3.1 图形数据不确定性来源

1. 测量数据的不确定性

测量数据主要是指使用常规大地测量、GPS、城市测量、摄影测量和其他一些测量方法直接观测所得到的被测对象的空间位置信息,其不确定性属于空间数据的定位不确定性。空

间数据的位置通常以坐标对表示,并与其地理坐标,即经纬度来表示之间存在确定的转换关系。在以标准椭球体和标准地球代表地球真实表面和空间时,引入了一定的参考系统误差因素。

测量误差通常考虑的是系统误差、粗差和偶然误差,它包括:①系统误差的产生与一个确定的系统规律有关,由环境因素(如温度、湿度和气压等)、仪器结构与性能以及操作技能等方面因素影响而产生。系统误差不能通过重复观测加以校正或消除,只能用数学模型进行模拟和估计。②粗差是操作人员在使用设备、读数或记录观测时因粗心或观测不当而产生的,需要采用各种方法检查和消除。一般地,粗差可通过简单的几何关系或代数检查验证其一致性,或通过重复观测检查发现和消除。③偶然误差是一种随机性的误差,由一些不可见和不可控制的因素引起的。这种误差具有一定的特征,如正负误差出现频率相同、大误差少、小误差多等。偶然误差可采用随机模型进行估计和处理。

2. 地图数据的不确定性

地图数据是指对现有地图进行数字化产生的数据。地图数据质量问题中,不仅含有地图要素中固有的误差,还包括图纸变形、图形数字化等误差。

(1) 地图固有误差来源。

① 位置误差。地图精度受多方面影响,例如图面精度、制图及地图保管等。对 1:1000 地形图而言,主要误差源有:控制测量误差、碎部测量误差、控制绘图误差、碎部绘图误差、制图综合误差、地图制印误差和清绘误差等。对于不同比例尺地图,其各项误差有所不同,这些可以根据地图测量及制图规范查出。在地图的存放与管理过程中,由于环境温度和湿度的变化,可引起图纸尺寸变化。例如,温度变化从 0℃~25℃可使图纸尺寸变化 1.6%,湿度变化可使其尺寸变化 1.25%~2.5% 的范围。地图要素的总体位置误差是以上各种误差的累积与传递。实际应用中,一般假设只是对以上各真实值的一种近似,上述各误差之间并不一定独立,并且程度不同地影响总误差即并非等权。只是由于这些误差的性质及关系比较复杂而到现在还没有比较完善的方法解求它们,因而用近似的方法解算总误差值。

② 属性误差。地图上的每一目标均有相应的属性描述。例如,在 1:500 地形图上标有电线杆,但是若由于标号错误而误把它标为树,则属于命名误差(或错误)或一种粗差。另外,一种属性误差是量值上的误差。例如,在标记某一地区的建筑密度时误算为 75%,而正确值应为 7.5%。与地图要素的位置误差相比,其属性误差一般讨论较少。事实上,在许多情况下,属性误差的影响是很大的,特别是像命名方面的误差。

③ 时域误差。时域误差是近年来引入的新概念。一般可认为时域误差有两方面的含义。一方面是,绘制在地图上的物体存在时间上的误差,例如某一建筑物的建筑时间为 1975 年到现在,而在图上错标为 1985 年到现在。由于一般地形图只表达了量测当时的情况,并没有追究地面物体的开始及终结时间,因此时域误差这个问题通常被忽略了。另一个方面是成图时间的早晚,成图时间越接近现在,其表达地面的现势性越强,其时域精度越高。除地形、地质等比较稳定的因素外,地图的可靠性是随其成图日期增长而递减的。

④ 逻辑一致性误差。逻辑一致性误差有两个层次的涵义。其一是物理概念上的,例如一个停车场与临近的道路,步行街与建筑物的连接或水塘、湖、河流和大海之间的水系连接性等。其二是物理概念的逻辑关系,输入到 GIS 后在 GIS 中数据模型上的误差。例如在 GIS 中一条弧的两端点与节点,以及一条弧两侧的左右多边形的相邻关系。逻辑上一致性,尤其

是拓扑关系上的一致性对于相关信息及完整性检查是十分重要的。

⑤ 完整性误差。在地图测图的过程中若有要素丢失,例如某一建筑物漏测,则其测图结果就不具有完整性。检查数据的完整性(或者是新的地物,如新建的一条道路)可能出现在地形动态监测中,例如对 GIS 中的已有区域加入新地物。由于某区域对象的动态性,数据完整性是 GIS 分析的主要问题之一。数据的不完整性应限制在一定的范围内。例如城市规划、管理均对数据现势要求很高。

GIS 中的数据完整性及逻辑一致性检测,通常是一起进行的。例如,美国在 1971 年发展的独立地图编码系统(The Dual independent Map Encoding System-DIME)可用于检测地图数据的完整性。围绕一个街区的道路被表示成许多弧段,每一个弧段有一个起始、终止结点;在每一结点处至少有三条弧段相交;对每一个弧段而言至少有一个左街区和一个右街区;每一个街区中由一组完整的结点及弧段组成。所开发的程序用于检测这些弧、节点、街区是否满足上述关系,实际上就是在检测数据的完整性以及逻辑拓扑一致性。在实际应用中,对于有些数据,例如一孤岛,由于它与周围地物不一定存在必然的逻辑关系,其丢失或逻辑上的错误就不可能用程序完成,只能由人工检查方法确定。

(2) 绘图介质变形产生的误差。

由于图纸的大小随着湿度和温度的变化而变化。在温度不变的情况下,若湿度由 0%增至 25%,则纸的尺寸可能改变 1.6%。纸的膨胀率和收缩率并不完全相同,即使湿度又恢复到原来的大小,图纸也不能恢复到原来的尺寸。一张 90cm 长的图纸因湿度变化而产生的误差可能高达 1.46cm。在印刷过程中,图纸在长、宽方向的净增长约为 1.25 %和 2.5 %,变形误差的范围为 0.24~0.48mm。聚纸薄 MUO 的二底图与纸质地图相比,材料变形产生的误差相对较小。

(3) 图形数字化误差。

大多数 GIS 空间数据库的建立是采用手工或屏幕跟踪数字化实现的。数字化方式主要有手工数字化和扫描数字化两种。

① 数据预处理误差。在进行地图数字化时,必须采用适当的比例因子进行修正。如果从不同的地图上采集信息,应了解地图的投影方式是否一致,比例尺是否匹配,以估计由此可能产生的误差。

② 跟踪数字化误差。线划跟踪技术和扫描数字化所引起的平面误差一般比较小,但在扫描数字化时,要素结合处出现的误差比较大。手工数字化引起的误差与原始资料、操作员、操作方式及数字化仪质量等因素有关。

③ 数字化要素对象引起的不确定性。原始资料的地图负载量,曲线的粗细及复杂程度是影响数字化精度的主要因素。若原图愈复杂且图的负载量大,则愈易出现选择目标遗漏差错(粗差)。粗线的输入精度要低于细线,复杂曲线的输入误差高于简单曲线,密集的要素比稀疏要素更容易引起误差等。

④ 数字化操作员引起的不确定性。数字化操作员的技能与经验不同,对地图数字化精度会有不同的影响。这主要表现在最佳采样点点位的选择、采点的多少、数字化标十字丝与目标对准程度的判断能力等方面。另外,数字化操作人员的疲劳程度和数字化的速度也会影响数字化的质量。

⑤ 数字化仪引起的误差。数字化仪的分辨率和精度对数字化的质量有决定性的影响。通