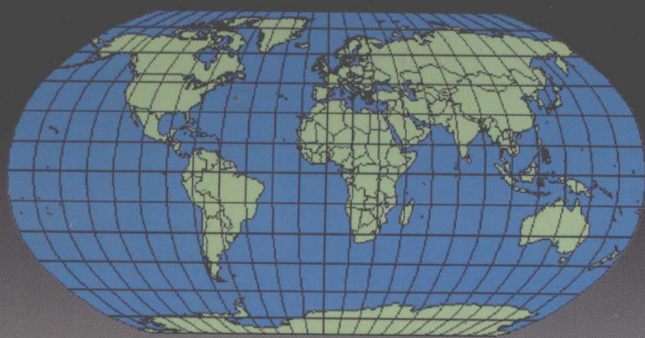


Accuracy Analysis and Quality Control of Spatial Data in GIS

刘大杰 史文中 童小华 孙红春 著

GIS 空间数据的 精度分析与质量控制

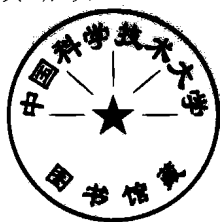


上海科学技术文献出版社

GIS 空间数据的精度分析与质量控制

刘大杰 史文中 著
童小华 孙红春

国家自然科学基金资助项目



上海科学技术文献出版社

责任编辑：祝静怡

封面设计：林 翌

GIS空间数据的精度分析与质量控制

刘大杰 史文中 著
童小华 孙红春

国家自然科学基金资助项目

*

上海科学技术文献出版社出版发行

(上海市武康路2号 邮政编码200031)

全国新华书店经销

上海教育学院印刷厂印刷

*

开本 850×1168 1/32 印张 7.5 字数 215 000

1999年10月第1版 1999年10月第1次印刷

印数：1-1100

ISBN 7-5439-1499-9/T·572

定 价：20.00 元

内 容 摘 要

GIS 数据的精度分析和质量控制问题是 GIS 的基础理论研究课题之一, 已成为国际 GIS 界十分关注的问题。GIS 空间数据的精度分析和质量控制所涉及的理论和实际问题很多, 其内容十分广泛。基于对 GIS 数据的精度分析和质量控制问题的重要性, 本书汇集了国家自然科学基金项目(49671065)和香港研究基金项目(HKP65/95)的主要成果, 重点是论述 GIS 空间数据的精度分析、不确定性模型和质量控制方面的某些问题。主要内容是:

1. 分析空间数据随机误差的统计性质和分布, 通过分布检验, 认为:

数字化数据的误差分布不一定是正态分布, 而可能是 p -范分布($1 < p < 2$), 并讨论地图曲线数字化数据的精度分析方法。

2. 阐述地图数字化数据的平差模型和分级平差方法, 并讨论空间数据的最小二乘估计和 p -范估计问题。

3. 论述 GIS 中的平面直线, 圆曲线, 缓和曲线, 一般拟合曲线和空间直线的不确定性模型, 以及其可视化问题。

4. 结合 GIS 数据的质量特征和 GIS 产品的特点, 讨论 GIS 产品质量的过程控制和抽样检验问题。

本书可供从事 GIS 数据质量控制方面的研究人员和有关高等学校的师生阅读, 也可供从事 GIS 系统建设和应用开发的科技人员参考。

Main contents

Accuracy analysis and quality control of data in GIS is one of the basic theories. They have been paid great attentions in GIS area, and many theories and actual issues have been involved in them. Based on the importance of accuracy analysis and quality control of data in GIS, this book summarized the research outcomes of two research projects which were supported by grants from China National Natural Sciences Foundation (Project No. 49670165) and Research Grants Council of the Hong Kong SAR (Project No. HKP65/95E). The book emphasized on issues such as accuracy analysis of spatial data, uncertainty model and quality control in GIS. The main contents are as follows:

1. The statistic property and distribution of the random error of spatial data was analyzed. It is considered that the error distribution of digital data mustn't be normal distribution and may be p-norm distribution ($1 < p < 2$). And accuracy analysis method of digital data in map curve was discussed.

2. The adjustment model and the multi-level adjustment method of map digital data were expounded, and the least square estimation and p-norm estimation of spatial data were discussed.

3. The uncertainty models and the visual issues of straight line, circular curve, transition curve and fitting curve in GIS were expounded.

4. Combining the quality property of GIS data and the property of GIS products, the process control and sampling inspection for the quality of GIS products were discussed.

The book is suited for these researchers who engaged in the quality control of GIS data and correlative teachers and students, and it can also act as the reference to these technological personnel who engaged in system construction and application exploitation of GIS.

前 言

地理信息科学是本世纪发展起来的一门新兴科学。它已成为现代科学的一个重要分支。地理信息系统(GIS)是地理信息科学发展的基础技术。地理信息系统的发展已经带动若干令人瞩目的应用,如信息高速公路、数字地球和空间数据基础设施等。

空间数据是 GIS 的一个重要组成部分。而其中空间数据的质量直接影响到数据的适用性及 GIS 应用的成败。因此,国际上 GIS 的研究机构(如 NCGIA 和 UCGIS)均把 GIS 中的精度分析与质量控制列为现代 GIS 研究的战略重点。

GIS 数据质量研究中涉及到位置精度、属性精度、时域精度、完整性和逻辑一致性等多方面的理论和实际问题,其内容十分广泛,许多问题还有待进一步研究。本书主要讨论空间数据和属性数据的精度分析和质量控制方面的问题。全书以概率论与数理统计为理论基础,介绍作者近年来的研究成果,包括空间数据的误差分布、GIS 空间目标的精度分析、空间数据处理方法、不确定性模型、GIS 产品的抽样检验与质量控制。这些方面构成了空间数据精度分析与质量控制的理论基础。笔者期望借此对 GIS 的质量研究理论的发展起到一点抛砖引玉的作用。

本书第一章简述了 GIS 空间数据精度分析与质量控制研究的重要意义及研究现状;考虑到 GIS 空间数据的精度分析与质量控制的研究是以概率统计为理论基础进行的,因此,本书第二章简要地介绍了有关的统计分布与分布检验方法;第三章分析空间数据随机误差的性质和分布,并对手工数字化和扫描数字化误差进行分布检验,认为其误差的分布不一定是正态分布,而可能是 p -范分布($1 < p < 2$),还探讨了地图数字化数据的精度;第四章阐述地图数字化观测数据的平差模型,包括几种基本模型、道路曲线的

平差模型和分级平差方法,并阐述应用最小二乘估计和 p-范估计方法进行空间数据处理和精度估计等方面的理论问题和实际计算问题;第五章论述 GIS 中主要要素的不确定性模型,包括平面直线、圆曲线、缓和曲线、一般拟合曲线及空间直线的不确定性模型,讨论以直线和曲线的法线方向的中误差 σ 为带宽的 ϵ_0 模型,及以最大方向误差为带宽的 ϵ_m 模型;第六章在分析质量管理和质量控制对 GIS 空间数据的重要性的基础上,简要地介绍质量管理的基本概念,并阐述过程控制和抽样检验的基本原理,以及在 GIS 空间数据的质量控制中的有关问题;最后对 GIS 数据质量控制问题的理论研究和实用的发展作一些介绍。

参加本书写作的有刘大杰教授、史文中教授、童小华博士、孙红春高级工程师;此外,孟晓林博士、姚连璧博士,以及博士研究生刘春、余晓红和华慧、许谷声等也参加了部分工作。

本书得到了国家自然科学基金(项目号:49671065)和香港研究基金委员会(项目号:HKP65/95E)的资助。

作 者

1999 年 1 月

Preface

Geographic information science has been developed since this decade. It has been one of the most active branches of modern science and technology. Geographic information systems (GIS) is one of the most essential technologies, which contributes to the development of geographic information science. GIS forms the basis for several new application areas, such as information highway, digital earth and spatial data infrastructure.

Spatial data is one of the fundamental parts of GIS. The quality of spatial data directly determines the fitness-for-use of GIS and affects the result of GIS applications. Therefore, accuracy analysis and quality control of spatial data in GIS is regarded as one of the fundamental theoretical research issues internationally, such as NCGIA and UCGIS.

The research framework of spatial data quality involves such issues as position accuracy, attribute accuracy, temporal accuracy, completeness and topological consistency. In this book, we focus on accuracy analysis and quality control for positional and attribute data. The theoretical basis of this book is statistics and probability theory. The book introduces research results of our group developed in these years. These cover the error distribution of spatial data, accuracy analysis of features in GIS, spatial data process methods, uncertainty model quality test and control for GIS products. These form the basis of the theory for spatial data analysis and quality control. The development of quality studying of GIS needs a long time, and this book is just a start of

the development.

The first chapter of this book analyses the significance and research status of accuracy analysis and quality control of spatial data in GIS; The second chapter introduces the theories on the statistical distributions and tests that are the theoretical bases for the research; the random error characteristics and distributions of spatial data are analyzed in the third chapter, error distribution tests for manual and scanning digitization are analyzed and it appears that the errors are not always normal but likely to be p-norm distribution ($1 < p < 2$), and accuracy of map digitization is discussed; the fourth chapter develops adjustment models for map digitized data, including basic adjustment models, road curve digitization adjustment models and multi-level adjustment method for complex features, and compares the p-norm estimation method with least square adjustment method; uncertainty models for spatial features are developed in the fifth chapter, including straight line, circular curve, transition curve, ordinary curve and three-dimensional line curve features, and the ϵ_s and ϵ_m models for describing spatial features uncertainties are presented and discussed; the sixth chapter studies the theories and methods of quality management and quality control for GIS data, and explains principles of the process control and sample test for handling data quality in GIS; in the end, the further research on the developments in theory and implementation for the accuracy analysis and quality control of spatial data in GIS is concluded.

This book was jointly written by Prof. Dajie Liu, Prof. Wenzhong Shi, Dr. Xiaohua Tong and Ms Hongchun Sun (Senior Engineer). Dr. Xiaolin Meng, Lianbi Yao and several PhD candidates, Chun Liu, Xiaohong Yu, Hui Hua and Gusheng Xu, also contributed to the book preparation. The work described in

this book was supported by grants from China National Natural Science Foundation (Project No. 49670165) and Research Grants Council of the Hong Kong SAR (Project No. HKP65/95E).

Authors

January 1999

目 录

第一章 绪论	(1)
§ 1.1 空间数据精度分析与质量控制研究的意义	(1)
§ 1.2 空间数据的精度分析与质量控制研究概况	(6)
第二章 统计分布与检验	(10)
§ 2.1 几种离散分布	(10)
§ 2.2 正态分布与 p-范分布	(13)
§ 2.3 几种常用统计量的分布	(19)
§ 2.4 顺序统计量的分布与极差分布	(30)
§ 2.5 分布拟合检验	(36)
第三章 数字化数据的精度分析	(43)
§ 3.1 地图数字化数据的误差	(43)
§ 3.2 随机误差的统计分析	(45)
§ 3.3 手工数字化数据的精度和分布检验	(53)
§ 3.4 地图曲线数字化数据的精度分析	(60)
§ 3.5 扫描数字化数据的精度与分布检验	(68)
第四章 GIS 空间数据的平差模型与精度分析	(72)
§ 4.1 数字化数据的基本平差模型	(73)
§ 4.2 最小二乘条件平差的迭代解法	(76)
§ 4.3 数字化数据的分级平差	(81)
§ 4.4 道路曲线数字化的平差模型	(87)
§ 4.5 空间数据的 p-范估计	(96)
§ 4.6 地图数字化软件系统及其应用	(101)
第五章 空间数据的不确定性模型	(109)
§ 5.1 直线要素的不确定性模型	(110)
§ 5.2 面要素的不确定性模型	(116)

§ 5.3	圆曲线的不确定性模型	(120)
§ 5.4	缓和曲线的不确定性模型	(126)
§ 5.5	一般拟合曲线的不确定性模型	(134)
§ 5.6	三维空间直线的不确定性模型	(143)
§ 5.7	不确定性模型的可视化	(149)
第六章	GIS 空间数据的质量控制与抽样检验	(159)
§ 6.1	质量管理与质量控制的一般概念	(159)
§ 6.2	统计过程控制与控制图	(163)
§ 6.3	抽样检验的基本概念与抽样检验标准	(171)
§ 6.4	GIS 数据的质量特征	(179)
§ 6.5	数字地图产品的质量检验标准和检验方法	(185)
§ 6.6	实例——某市土地利用现状数字地图的质量 控制	(195)
§ 6.7	GIS 空间数据质量控制的策略	(206)
第七章	GIS 数据质量研究展望	(210)
§ 7.1	现实世界度量的维度空间与 GIS 数据不确定 性的研究	(210)
§ 7.2	关于 GIS 空间数据精度和质量控制的几点结 论	(212)
§ 7.3	GIS 数据质量的研究与展望	(213)

Table of Contents

Chapter 1	Introduction	(1)
§ 1.1	Significance of Accuracy Analysis and Quality Control on Spatial Data	(1)
§ 1.2	A Review of Related Research on accuracy analysis and quality control	(6)
Chapter 2	Statistical Distributions and Tests	(10)
§ 2.1	Several Discrete Distributions	(10)
§ 2.2	Normal Distribution and P-norm Distribution	(13)
§ 2.3	Distributions of Several Statistical Variables	(19)
§ 2.4	Distribution of Sequential Statistical Variables and Maximum Difference Distribution	(30)
§ 2.5	Distribution Fitness Tests	(36)
Chapter 3	Accuracy Analysis of Digitization	(43)
§ 3.1	Error of Map Digitization	(43)
§ 3.2	Statistical Analysis of Random Error	(45)
§ 3.3	Accuracy Analysis and Distribution Test of Manual Digitization Error	(53)
§ 3.4	Accuracy Analysis of Curve-line Digitization Error	(60)
§ 3.5	Accuracy Analysis and Distribution Test of Scanning Error	(68)
Chapter 4	Adjustment Models and Accuracy Analysis	(72)
§ 4.1	Fundamental Adjustment Models for Digitization	

	Data	(73)
§ 4.2	Iterative Method of Least Square Conditional Adjustment	(76)
§ 4.3	Multi-level Adjustment for Digitization Data	(81)
§ 4.4	Adjustment Models for Error of Road Curve Digitization	(87)
§ 4.5	Least p-norm Estimation of Spatial Data	(96)
§ 4.6	Map Digitization Software and Its Application	(101)
Chapter 5	Uncertainty Models for Spatial Features	(109)
§ 5.1	Uncertainty Models for Straight Line Features	(110)
§ 5.2	Uncertainty Models for Area Features	(116)
§ 5.3	Uncertainty Models for Circular Curve Line Features	(120)
§ 5.4	Uncertainty Models for Transition Curve Features	(126)
§ 5.5	Uncertainty Models for ordinary Curve Features	(134)
§ 5.1	Uncertainty Models for Three-Dimension Line Features	(143)
§ 5.1	Visualization of Uncertainty in Spatial Data	(149)
Chapter 6	Quality Control and Sample Test for Spatial Data	(159)
§ 6.1	Overview of Quality Control and Management	(159)
§ 6.2	Statistical Process Control and Control Graphs	(163)

§ 6.3	Concepts and Standards of Sample Test	(171)
§ 6.4	Quality Characteristic of GIS Data	(179)
§ 6.5	Standards for Quality Control of Digital Products	(185)
§ 6.6	A Case Study — Quality Control for Digital Land Use Map of Shanghai	(195)
§ 6.7	Strategy for Quality Control of Spatial Data in GIS	(206)
Chapter 7	Conclusions and Further Directions	(210)
§ 7.1	Dimension Space for Measurement of the Real world and Uncertainty Research	(210)
§ 7.2	Major Contents and Outcomes of This Book	(212)
§ 7.3	Further Directions for Quality Study in GIS	(213)

第一章 绪 论

§ 1.1 空间数据精度分析与质量控制研究的意义

地理信息系统(GIS)是一种以采集、存储、管理、分析和描述整个或部分地球表面与空间和地理分布有关的空间信息系统。它通常由数据采集子系统、标准与规范、数据输入子系统、数据库和数据管理系统、数据应用子系统等部分组成。从技术的角度看, GIS 包括软硬件条件、GIS 数据、GIS 基础设施(技术人员、资金支持)等。

在建设一个项目中,硬件条件和 GIS 软件是应用项目的工具和环境,而数据是 GIS 中最基本和最重要的组成部分,也是一个 GIS 项目中的投资重点,一般可占总投资的 50%~70%。近年来, GIS 数据质量问题以及由此产生的 GIS 数据质量控制技术成为了国内外学术界研究的热点问题之一。

GIS 所包含的数据均与空间位置相联系,通常可按其表达形式分为以下两种类型:

1. 空间数据,又称为图形数据。

它是 GIS 的基础数据,是描述空间实体的位置、形状、大小及各个不同实体之间的关系的的信息。空间实体的位置一般以三维或二维坐标串来表示。

空间数据又可分为几何数据和关系数据两种。

几何数据从几何角度把空间目标划分为点状、线状和面状目标三种基本类型。

点状(零维)目标,是在空间有确定的位置(坐标)而没有长度和面积的目标。例如,孤立点、拓扑交点和端点。

线状(一维)目标,是在空间有确定位置和长度而没有面积的目标,其端点由两个点状目标界定。

面状(二维)目标,是在空间有确定的位置,并有长度和面积的目标,它由若干个点状或线状目标界定。

关系数据是描述各个不同空间实体之间的关系(如邻接、关联、包含、连通、接近度)的信息。关系数据的建立有助于各种应用和空间分析。

2. 非空间数据,又称为属性数据。

它包括各个地理单元中的社会、经济或其它专题数据,是对地理单元(实体)专题内容的广泛、深刻的描述。从某种意义上也可以将关系数据归入属性数据。

对 GIS 来说,空间数据是基础;非空间数据是内涵,是地理单元的纵深描述。不难理解,数据质量的优劣将直接影响 GIS 应用的分析结果的可靠程度和应用目标的实现。

一般来说,GIS 的数据质量主要包括以下内容:

1) 位置精度,空间实体的位置通常以三维或二维坐标表示,而位置精度则是表示实体的坐标数据与真实数据的接近程度,因而常以坐标数据的精度来表示。位置精度包括数学基础精度、平面精度、高程精度、接边精度、形状再现精度等。

2) 属性精度,是指实体的属性值与其真实值相符的程度,它通常取决于数据的类型,且常与位置精度有关。属性精度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。

3) 逻辑一致性:指数据之间的关系上的可靠性。包括数据结构、数据内容、空间属性与专题属性,以及拓扑性质上的内在一致性。

4) 数据完整性,包括数据范围、数据的分层、实体类型、属性数据和名称等各方面的数据的完整性。

5) 时间精度,主要是指数据的现势性。一般通过数据采集时间、数据更新时间及更新频度来表现。

6) 数据情况说明,指数据说明的准确性和全面性,要求对数