

空间数据多尺度表达的 不确定性分析模型

Uncertainty Analysis Models for Multi-scale Representation of Spatial Data

徐丰 牛继强 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

国家自然科学基金项目“空间数据多尺度表达不确定性及其可视化研
国家自然科学基金项目(41201387)
信阳师范学院第六批一级重点学科——地理学
河南省高等学校青年骨干教师资助计划(2012-GGJS-127)

4411244

空间数据多尺度表达的 不确定性分析模型

Uncertainty Analysis Models for Multi-scale Representation of Spatial Data

徐丰 牛继强 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

空间数据多尺度表达的不确定性分析模型/徐丰,牛继强著. —武汉:
武汉大学出版社,2014.4

ISBN 978-7-307-12800-2

I. 空… II. ①徐… ②牛… III. 空间信息系统—不确定系统—系
统模型—研究 IV. P208

中国版本图书馆 CIP 数据核字(2014)第 020475 号



责任编辑:鲍 玲

责任校对:鄢春梅

版式设计:韩闻锦

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件: cbs22@whu.edu.cn 网址: www.wdp.com.cn)

印刷:武汉中远印务有限公司

开本: 720 × 1000 1/16 印张: 17.25 字数: 243 千字 插页: 1

印数: 1 - 1050 册

版次: 2014 年 4 月第 1 版 2014 年 4 月第 1 次印刷

ISBN 978-7-307-12800-2 定价: 39.00 元

版权所有, 不得翻印; 凡购我社的图书, 如有质量问题, 请与当地图书销售部门联系调换。

前 言

随着 GIS 应用领域的不断扩展和需求层次的日益提高,人们越来越多地需要对多尺度的空间数据进行分析、处理和表达。如今,空间数据的多尺度表达已成为地理信息科学研究的前沿问题之一。不确定性问题存在于空间数据的生命周期全过程,只要有空间数据就有不确定性问题存在。因此,空间数据的多尺度表达必然存在不确定性。

空间数据的不确定性问题的早已引起国内外 GIS 界的重视,是当前 GIS 理论研究的热点和难点之一。国外对于空间数据的不确定性问题的非常重视。美国地理信息科学大学研究中心(UGGIS-CI)、美国国家地理信息和分析中心(NCGIA)、美国国家经济与社会研究委员会(ESRS)、荷兰的空间数据分析专家中心(NEXPRI)都把地理信息的不确定性作为重要课题进行研究。在国际空间数据处理会议(ISDH)、欧洲地理信息系统会议(EGIS)、美国地理信息系统年会(AGIS)、自然资源数据库空间数据不确定性等国际会议中,都设立了关于空间数据不确定性专题讨论组。以空间数据不确定性为主题的国际空间数据质量大会已经召开了 8 届;两年一届的自然资源与环境科学中的空间精度评价国际会议已经召开了 10 届。国内外对空间数据不确定性的研究取得了长足进步,主要从“GIS 不确定性的研究内容与体系、数据不确定性的来源、数据不确定性处理方法、空间分析过程中的不确定性传播机理”等方面展开。目前国内已有较为系统的研究和出版的专著,如史文中(2005)的《空间数据与空间分析不确定性原理》,承继成、郭华东、史文中等人(2004)的《遥感数据的不确定性问题》,邬伦、高振记等人(2010)的《地理信息系统中的不确定性问题》、胡圣武(2006)的《GIS 质量

评价与可靠性分析》等。

关于空间数据多尺度表达不确定性的问题,早已引起学者的关注,Lam 等人在 1992 年将尺度和数据不确定性之间的关系归纳为四个方面:不同的量测单位其结果不一样;不同比例尺具有不同的不确定性;大尺度的研究覆盖较大的研究区域,代表性强,不确定性小;运行尺度越小把握性越大,不确定性越小。随着空间数据多尺度表达问题的升温,空间数据多尺度表达的不确定性问题也越来越引起学者们的重视,第八届(2013)国际空间数据质量大会上为数篇空间数据多尺度表达不确定性相关的论文参与交流,在自动综合算法的质量评价中也都包含了对综合结果不确定性的评估,而且许多空间数据不确定性的研究与空间数据多尺度表达不确定性的研究相关,其成果可为空间数据多尺度表达不确定性研究利用和借鉴。

在国家自然科学基金项目“空间数据多尺度表达不确定性及其可视化研究”(No. 41001219)的资助下,本书作者对空间数据多尺度表达的不确定性及其评价指标、量化模型,空间数据多尺度表达空间关系不确定性及其评价指标、量化模型,空间数据多尺度表达不确定性可视化方法等进行了系统深入的研究,取得了一些阶段性成果。本书除了对这些成果进行全面的总结,还借鉴了国内外相关领域的最新研究进展,力图对空间数据多尺度表达不确定性的理论和方法进行系统的阐述,初步建立空间数据多尺度表达不确定性研究的理论和方法体系。

本书共分为 7 章。第 1 章(绪论)阐明空间数据多尺度表达不确定性的相关概念及相关研究进展;第 2 章(空间数据多尺度表达不确定性的理论基础)介绍相关数学基础以及空间数据不确定性模型;第 3 章(空间数据多尺度表达不确定性及其形式化描述)研究空间数据多尺度表达不确定性的来源和分类,并归纳不同尺度空间数据不确定性的统一的形式化描述方法,作为开展相关研究的基础;第 4 章(空间数据多尺度表达的不确定性建模)深入研究各地理事要素多尺度表达的不确定性评价指标和量化模型;第 5 章(空间数据多尺度表达的空间关系不确定性建模)深入研究拓扑关系和方

向关系多尺度表达的不确定性评价指标和量化模型；第6章(空间数据多尺度表达的不确定性可视化)系统研究空间数据多尺度表达不确定性的可视化方法；第7章(结语)对本书进行总结。

本书由信阳师范学院城市与环境科学学院的徐丰、牛继强共同拟定提纲；徐丰撰写了本书的前言、第1章、第2章的第1节、第3章、第4章的第1节和第2节、第5章的第1节和第2节等几个部分，牛继强撰写了本书的第2章的第2节和第3节、第4章的第3节至第6节、第5章的第3节至第5节、第6章、第7章等几个部分。李卓凡、蔡珺同学在整理资料、绘制图表和校对方面做了大量的工作，王帅帅、王腾、付素华等同学也参与了资料整理与校对工作。最终，由徐丰、牛继强统稿。在本书的研究和撰写过程中，得到了信阳师范学院城市与环境科学学院领导的关怀和同事的支持，在此表示诚挚的谢意。

本专著是在国家自然科学基金项目“空间数据多尺度表达不确定性及其可视化研究”(No. 41001219)的资助下取得的研究成果之一，并得到国家自然科学基金(No. 41201387)、信阳师范学院第六批一级重点学科——地理学一级重点学科和河南省高等学校青年骨干教师资助计划(No. 2012-GGJS-127)的联合资助。

由于研究问题的复杂性和研究时间的限制，本专著的研究仍是初步的，在一些方面尚不成熟，错误和疏漏在所难免，敬请读者不吝指正。

徐 丰 牛继强

2013年12月于信阳师范学院

目 录

第 1 章 绪论	1
1.1 基本概念	1
1.2 空间数据多尺度表达的不确定性问题	9
1.3 空间数据多尺度表达不确定性的研究意义	14
1.4 相关研究进展	15
1.5 本章小结	29
第 2 章 空间数据多尺度表达不确定性的理论基础	30
2.1 数学基础	30
2.2 空间数据的不确定性模型	45
2.3 本章小结	65
第 3 章 空间数据多尺度表达不确定性及其形式化描述	66
3.1 空间数据多尺度表达的实现方案	66
3.2 空间数据多尺度表达的不确定性来源	72
3.3 空间数据多尺度表达的不确定性类型	85
3.4 空间数据多尺度表达的不确定性概念模型	90
3.5 空间数据多尺度表达不确定性的形式化描述	94
3.6 本章小结	96
第 4 章 空间数据多尺度表达的不确定性建模	98
4.1 点元多尺度表达的位置不确定性	98
4.2 线状要素多尺度表达的不确定性	100
4.3 面状要素多尺度表达的不确定性	123

4.4	点群多尺度表达的不确定性	126
4.5	城镇街区多尺度表达的不确定性	136
4.6	本章小结	144
第5章	空间数据多尺度表达的空间关系不确定性建模	146
5.1	多尺度空间面目标的粗集表达	146
5.2	多尺度空间拓扑关系不确定性建模	147
5.3	多尺度空间方向关系不确定性建模	185
5.4	多尺度空间关系不确定性的统一建模	211
5.5	本章小结	212
第6章	空间数据多尺度表达的不确定性可视化	214
6.1	概述	214
6.2	空间数据多尺度表达的不确定性可视化设计与 流程设计	219
6.3	二维可视化	221
6.4	三维可视化	231
6.5	动态可视化	237
6.6	多媒体可视化	240
6.7	本章小结	242
第7章	结语	243
	主要参考文献	245

第 1 章 绪 论

空间数据的多尺度表达与 GIS 数据的精度分析和质量控制问题是当今地理信息科学和国际 GIS 界十分关注的两大前沿课题。数据质量的关键是对数据不确定性的评估。许多著名的国际学术机构和组织都把空间数据的不确定性问题作为 21 世纪的重大研究课题。目前, GIS 不确定性的研究集中在 GIS 基本实体(点、线、面等)的单一数据集的不确定性的研究上, 很少涉及多尺度表达(数据转换)的空间数据的不确定性(可靠性)评价的模型、度量指标和评价体系。因此, 为了适应地理信息数据多尺度表达的发展需要, 必须研究空间数据多尺度表达不确定性评价模型、度量指标和评价体系, 一方面为正确、合理、有效地建立 GIS 多尺度表达实用系统提供依据和标准, 另一方面为客观、科学地理解、分析和使用多尺度表达的空间数据提供基本保障, 降低空间决策的风险。

1.1 基本概念

1.1.1 空间数据

“空间”的概念在不同的学科有着不同的解释(王家耀, 2001)。地理学中的地理空间是指物质、能量、信息的存在形式在形态、结构过程、功能关系上的分布方式和格局及其在时间上的延续。地理空间涵盖上至大气电离层、下至地幔莫霍面的范围, 是地球上大气圈、水圈、生物圈、岩石圈和土壤圈交互作用的区域。依附地理空间存在的各种事物或现象, 它们可能是物质的, 也可能是非物质

的,如土地类型、资源分布、行政区划、水系、人口分布、工农业布局、城市规划、道路网结构等。这些事物和现象的一个典型特征,就是与一定的地理空间位置相关,都具有一定的几何形态,因此称为地理空间实体。在地理空间中,实体不仅反映了事物和现象的地理本质内涵,而且反映了它们在地理空间中的位置、分布状况以及它们之间的相互关系。

地理信息科学中的地理空间被定义为绝对空间和相对空间两种形式,我们称之为地理信息空间。绝对空间是具有属性描述的空间位置的集合,它是由一系列不同位置的空间坐标值组成;相对空间是具有空间属性特征的实体的集合,它是由不同实体之间的空间关系构成。可以简单地将地理信息空间理解为一个空间目标(对象)组合排列集,其每个目标都具有位置、属性和时间信息,及与其他对象的拓扑关系、语义关系等。

从地理信息空间的定义可知,它是信息世界的地理空间,是地理学地理空间的抽象表达,可以通过对地理空间进行抽象或空间采样来描述,其结果就是地理空间数据。因此,地理空间数据是地理空间信息的数字描述,是地理空间的离散表达,由空间对象的几何信息、属性信息、时间信息,以及与其他对象之间所具有的拓扑关系和语义关系等数据构成。我们可以把这些数据分为三类:属性数据、空间数据以及时域数据,其中空间数据是指含有空间或位置信息的数据,分为关系数据和几何数据两种。

关系数据是描述各个不同空间实体之间关系(如邻接、关联、包含、连通、接近度)的信息,也就是拓扑关系、方向关系、距离关系。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。图1.1为空间数据的分类。

无限的地理实体离散化会产生有限的表示,因为计算机模型和

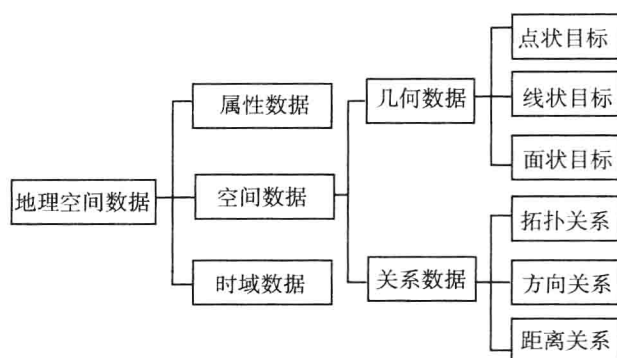


图 1.1 空间数据的分类

数据获取方法是有限的，这个离散化是通过两个步骤来完成的：分辨率控制和特征抽象。按照所表达地理实体的知识，利用分辨率可控制其详细程度(Bruegger, 1994)。依据其详细程度，空间数据有三个水平上层次结构：要素层、目标层、几何细节层。要素是具有相同语义特征的目标集；目标是具有独立地理意义的表达实体，是构成要素的基本单位；几何细节是几何表达上划分的结构体，是构成目标的基本单位，如构成河流目标的“弯曲”特征，构成面状目标的三角形剖分单元等。

1.1.2 空间数据多尺度表达

尺度(scale)是易混淆和易误解的概念，通常对于不同的学科和应用背景有不同的含义(Goodchild & Quattrochi, 1997)。尺度是指在研究某一物体或现象时所采用的空间或时间单位，同时又可指某一现象或过程在空间和时间上所涉及的范围和发生的频率。前者是从研究者的角度来定义尺度，而后者是根据所研究的过程或现象的特征来定义尺度。

尺度也是与地理信息相关的最基本的和难以理解的概念之一。在地理信息科学中，尺度既用来指研究范围的大小(如地理范围)，也用于指详细程度(如地理分辨率的层次、大小)，还用于表明时

间的长短以及频率(即时间尺度)。

在研究的范围大小上(即空间尺度),往往以粒度(grain)和幅度(extent)来表达。空间粒度是指地理空间中最小可辨识单元所代表的特征长度、面积或体积(如样方、像元)。对于空间数据如影像资料而言,其粒度对应于像元大小,与分辨率有直接关系。幅度是指研究对象在空间或时间上的持续范围或长度。具体地说,所研究区域的总面积决定该研究的空间幅度。

在时间尺度上,时间粒度是指某一现象或事件发生的(或取样的)频率或时间间隔。例如,一些测量仪器的取样时间间隔或某一干扰时间发生的频率。一般根据研究项目持续的时间长短来确定其时间幅度。

关于详细程度,可以用人们所熟悉的地图比例尺概念来理解。“比例尺”定义为表达空间(地图图面)中的距离与实际地理空间距离的比率。大比例尺的地图覆盖了较小的地理区域并一般包含了较为详细的信息;相反,小比例尺的地图覆盖了较大的地理区域,但通常所包含的有关该区域的信息较少。不同的比例尺所表达的信息密度有很大的差异,通常以不同分辨率或抽象层次来表达同一地理要素不同详细程度的几何、拓扑结构和属性方面的信息。王艳慧、陈军等人(2003)对此问题作了非常形象直观的解释,在几何层面上,同一地理要素在不同的尺度下表现出不同抽象程度的几何形状,反映在数据库中则可能被抽象为不同的几何类型。如图 1.2 所示,道路网中同样几条路段,在较大比例尺下,各路段表达为双线目标,道路节点用面状几何要素表示;但在小比例尺下,同样的路段则用其道路中心线表示,立交桥、路口等节点则用简单的点状目标来表示。在要素层面上,几个要素可在不同的抽象层次下,基于不同的几何、时态或语义准则聚合成新的复合要素。这样的一个复合要素可能在不同尺度下存在不同的表达,且各种表达相互独立,在不同比例尺转换时会发生出现/消失或聚集/分解的情况。这种情况下,复合要素和底层的对应要素间具有层次性关系,高层要素由低层要素组合而成。如图 1.3 所示,河南省行政区划要素由 17 个市(地区)级行政区划要素组成,后者又由若干个县(市、区)级行

政区划要素组成,彼此之间存在行政上的隶属关系和空间上的聚集/分解关系。在属性层面上,同一地理要素在不同抽象层次下的表达属性也不相同。表 1.1 中,点、线、面在不同尺度背景(村级、乡(镇)级、县(市、区)级、市(地区)级、省级、国家级)下反映的要素的属性具有不同的含义。低层抽象对象的属性值比高层对象的属性值更准确。与前面对象层次性不同的是,要素层次上各尺度下的抽象对象是客观存在的,而属性值层次中的对象实际上并不存在,且属性值与尺度(分辨率)有关。

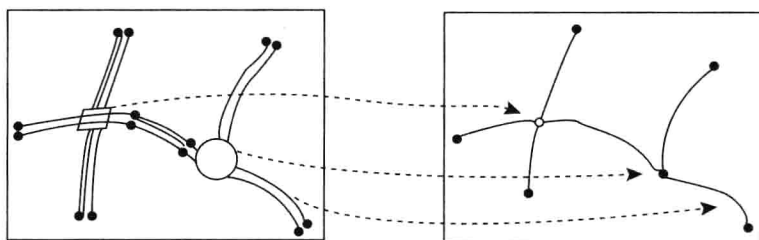


图 1.2 地理要素多尺度集合抽象表达的层次性



图 1.3 地理要素不同尺度抽象层次下的要素聚集特征

在地理信息系统的研究中,一般将尺度简单地理解为比例尺。不同的尺度不仅所表达的信息在密度上有很大的差异,而且还会影响所表达的地理信息是否正确,因为不少地学现象和规律只在一定的尺度出现,只有经过合理的尺度抽象出的空间信息才更具利用价值。因此,从认知科学的观点出发,尺度体现了人们对空间事物、空间现象认知的深度与广度。

表 1.1 同一几何抽象要素在不同尺度下反映不同详细层次的属性信息

要素	点	线	面
村级	位置、点属性	位置、形状、方向	位置、形状、面积
乡(镇)级	位置、点属性	位置、形状、方向	位置、形状、结构、面积
县(市、区)级	位置、点属性	位置、形状、方向	位置、形状、结构、面积
市(地区)级	位置、点属性	位置、走势	位置、形状、结构
省级	位置、高级别点属性	轮廓、走势	位置、结构
国家级	位置、高级别点属性		轮廓

李霖等人(2005)将尺度定义为与空间数据的空间、属性、时间特征同等重要的基本特征。空间数据的多尺度表达就是要解决由于用户对地理数据不同应用的需求,而导致需要对空间数据产生不同表达的问题。王艳慧等人(2006)认为地理要素多尺度表达指的是在数据库中存储、定义、描述、维护地理要素两个或两个以上尺度(比例尺)下的相应表达实例,这些实例以不同分辨率或抽象层次来表达同一要素不同详细程度的几何、拓扑结构和属性方面的信息,使得用户可以根据具体需求选择其中的一种或多种不同分辨率的实例,并能用于多比例尺的层次空间推理和数据更新。由此可见,空间数据的多尺度表达是在不同尺度(比例尺)下以相适应的不同详细程度来表达空间数据,一个特定尺度下的空间数据可以转换到另一个尺度下的空间数据,这种转换可以是向上尺度转换(尺度扩展),也可以是向下尺度转换(尺度收缩)。常用关于尺度转换

(孟斌, 王劲峰, 2005)的表述见表 1.2。

表 1.2 常用关于空间尺度转换的表达

	英文表达	其他中文翻译	具体含义
尺度 扩展	Upscaling Scaling-up Top-down	向上尺度转换 尺度上推	从小尺度(空间)观测中获得较大尺度(空间)上信息的过程,是将信息从精确的尺度(高分辨率)向模糊尺度转换的过程
尺度 收缩	Downscaling Scaling-down Bottom-up	向下尺度转换 尺度下推	将大尺度上(空间)的信息分解到更小尺度(空间)上的过程,是将信息从模糊的尺度向精确的尺度转换的过程

1.1.3 不确定性

“不确定性”是相对于“确定性”而提出来的,“不确定性”是客观世界的固有属性,它遍及万事万物和一切客观现象中。

比较常见的关于不确定性的定义有:不确定性(uncertainty)是指客观世界或实体本身就具有的变异,表现为不精确性、随机性和模糊性;不确定性是指既不能肯定,也不能否定,或可能是对或可能是错,没有准确把握;不确定性表示事物的含糊性、不明确性、不肯定性或指某事物的未决定或不稳定状态;混沌(chaos)理论认为,复杂系统中的不确定性是由于系统中各要素的相互作用一般处于力的不平衡状态,在空间上表现为要素的不均匀性和不稳定性,在时间上则表现为发展的非周期性,它们综合反映在客观实体上就是对象的不确定性(spatial object with uncertainty)。总括前人的研究成果,客观世界的不确定性问题可归为客观世界客观存在的不确定性和人们主观认识上的不确定性两类。客观世界客观存在的不确定性最为典型的是随机性;人们主观认识上的不确定性有:由于定义不清造成的模糊性,灰色系统中的部分已知、部分未知的灰性,以

及未确知性。其中,未确知性是指就客观世界本身来讲它是确定的,但对于决策者来说由于受决策条件限制而导致的认识不清、所掌握的证据不足以确定事物的真实状态和数量关系而带来的纯主观的、认识上的不确定性。

不确定性的含义比较广泛,以下现象均具有不确定性:

①凡是既不能肯定,也不能否定的,为不确定性。

②凡是一个图形、影像、数据或方程具有多解性的,该方程或数据具有不确定性,如遥感数据、地球物理数据等。

③在一个多尺度系统中,物理量是随尺度而变化的。例如,英国或挪威的海岸线不仅随度量的尺度的不同而不同,而且还随比例尺的不同而不同,在1:50万和1:5万地图上测量的长度,除了规定的地图比例差之外,还存在“多尺度系统的不确定性问题”。

④凡是瞬息万变的动态系统,其变化的速度大于仪器可能测量的速度,它是测不准的,是不确定的。

⑤人工模拟产品与客观真实世界之间总存在一定的差异,没有差异是不可能的,也就是具有不确定性。

⑥凡是两类地物之间是渐变的,则两者之间的界线具有不确定性。

⑦凡是概念或语义具有不完备性或不一致性特征的,则具有不确定性。例如,城市的定义在各国是不一致的,所以城市化率的概念也是不确定的;林地农业部、林业部、土地局的定义可能是不一样的,所以是不确定的。

数据是对客观世界性质、特征和状态的描述,但是由于客观世界的复杂性和在数据产生过程中携带了一些和客观世界无关的干扰影响,使得数据产生了与客观世界不一致的状况,这种不一致导致了数据具有不确定性。数据的不确定性大概由三方面的因素所致:

①客观世界本身的不确定性;②人类对客观世界认知的局限性;③量测技术水平及手段的局限性。

客观世界中的绝大部分现象都是不确定的。所谓确定的、规则的现象,只会在一定的前提和条件下发生,只会在局部或者较短的时间内存在。例如,在构成物质的分子内部,其电子围绕原子的运

动轨迹是不确定的；城市与郊区的结合部都明显地呈现边界的不确定性；气候分布、各种林地分布的边界也都呈现出不确定性特征。

人类对客观世界的认知是一个渐进的逐步深入的漫长过程，而这种认知是渐进的、不断深入发展的，在某一时期总是处于一定的水平，并且在认知进程的每一阶段，都无可避免地存在着认知的不确定性。例如，随着航天科技与遥测技术的飞速发展，人们对宇宙的认知进一步深化，其认知的广度是以往任何时期都无与伦比的。

我们对客观目标使用某一量测技术在一定的时期内可能只能达到一定的准确度，而且不同的量测技术可能会导致不同的准确度。例如，使用全球定位系统对地面点坐标进行量测，在具有 SA (selective availability) 政策限制下，定位精度为 100m，取消 SA 限制后定位精度可以达到 20m (实际可能达到约 10m)；而使用差分 GPS 技术，在有足够多的卫星观测值的情况下可以达到 1 ~ 5m 的精度。再进一步，GPS 载波相位定位可达到 1 ~ 2cm，甚至毫米级精度。

“不确定性”存在于自然科学技术的各个领域，如微电子分子生物学、基因工程及医学中，同时还存在于社会经济和人文科学的各个领域。总之，客观世界和我们的生活里充满了不确定性。随着不确定性研究的深入，世界的不确定性特征越来越得到学术界的普遍认可。越来越多的科学家相信，不确定性是这个世界的魅力所在，只有不确定性本身才是确定的。

1.2 空间数据多尺度表达的不确定性问题

1.2.1 空间数据的不确定性

数据是客观实体信息的载体，是科学的基础。空间数据是 GIS 的基础数据。

空间数据通过描述地理空间实体及其实体间各关系来表达和模拟客观地理世界。但是由于主客观的因素，如测量设备的固有精度