空间数据分析

苏世亮　李　霖　翁　敏　编著

科学出版社

空间数据分析

苏世亮　李　霖　翁　敏　编著

科学出版社

北京

内 容 简 介

空间数据分析是分析空间数据、挖掘空间信息、统计空间规律、解决空间问题所涉及的基本理论、方法与技术的总称。作者结合多年的教学和科研体会，在重视与大学数学基础课程、地理信息科学其他专业课程衔接的基础上，遵循从理论基础到实际应用的主线，强调不同方法之间相互关联的逻辑关系，以全新视角重新构建了空间数据分析的知识体系。指导思想是力求深入浅出地为读者提供空间数据分析的思路、方法和应用途径。

本书既可作为地理信息科学专业及相关专业本科生、研究生教材，也可供科研工作者参考。

图书在版编目（CIP）数据

空间数据分析 / 苏世亮，李霖，翁敏编著. —北京：科学出版社，2019.6
ISBN 978-7-03-059515-7

Ⅰ. ①空… Ⅱ. ①苏… ②李… ③翁… Ⅲ. ①空间信息系统-数据处理-研究 Ⅳ. ①P208

中国版本图书馆 CIP 数据核字（2018）第 256589 号

责任编辑：杨 红 郑欣虹 / 责任校对：樊雅琼
责任印制：张 伟 / 封面设计：陈 敬

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

北京中石油彩色印刷有限责任公司 印刷
科学出版社发行 各地新华书店经销

*

2019 年 6 月第 一 版 开本：787 × 1092 1/16
2019 年 6 月第一次印刷 印张：16 1/4
字数：398 000

定价：59.00 元

（如有印装质量问题，我社负责调换）

前　言

空间数据分析(spatial data analysis，SDA)是分析空间数据、挖掘空间信息、统计空间规律、解决空间问题所涉及的基本理论、方法与技术的总称。SDA 是地理信息系统(geographic information systems，GIS)的核心和灵魂，是 GIS 区别于一般的信息系统、计算机辅助设计或者电子地图系统的主要标志之一。SDA 已广泛应用于地理学、地质学、环境学、生态学、社会学、管理学、气象学，以及公共卫生等领域。因此，系统阐述 SDA 的基础理论与方法，为对它进行深入学习和促进其发展奠定良好基础是本教材的基本出发点。

武汉大学地理信息类专业在国内较早地开设了 SDA 的相关课程，强调理论、前沿、实践并重，重视对学生学习兴趣的引导和动手能力的培养。在武汉大学“双一流”学科建设的推动下，作者结合多年的教学和科研体会，在重视与大学数学基础课程、GIS 其他专业课程衔接的基础上，遵循从理论基础到实际应用的主线，强调不同方法之间相互关联的逻辑关系，以全新视角重新构建了 SDA 的知识体系。本教材的指导思想是力求深入浅出地为地理信息科学专业及相关专业本科生、研究生和科研工作者提供 SDA 的思路、方法和应用途径。因此，本教材对于每一种方法都特别注重阐释原理、适用条件、计算过程、结果解释，尽可能淡化数学推导过程。为了使学生能够应用所学的方法和技术解决实际问题，作者特别精心挑选了 16 个典型案例，形成了本教材的姊妹版《空间数据分析案例式实验教程》，为读者创造性地应用所学知识提供范例。此外，作者编写了本教材的理论扩展版《空间智能计算》和应用扩展版《社会地理计算》《健康地理学》，紧密结合 SDA 的新理论、新方法和新技术，以期为学生了解和掌握更多地理信息科学前沿知识以解决实际社会问题提供参考(图 1)。

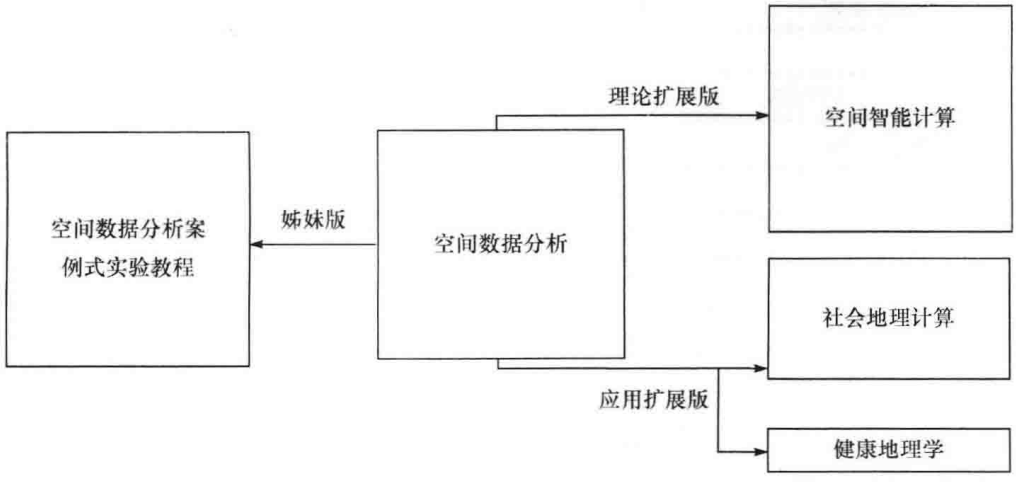

图 1　空间数据分析系列丛书

本教材共分为 7 章。第 1 章为空间数据分析概论，在阐明空间数据性质的基础上，总结了空间数据分析内涵和发展；第 2 章为经典统计学基础，介绍了经典统计学的理论基础和常用方法(相关、聚类、判别、主成分)；第 3 章为统计关系分析，详细阐述了常用的回归模型、

结构方程模型和时间序列分析模型；第 4 章为空间依赖与空间异质性，在阐述空间权重、空间自相关统计量的基础上，重点总结了空间回归及地理加权回归的原理、方法和应用；第 5 章为空间可达性，主要涉及空间距离关系、邻近度分析、叠加分析、网络分析和空间可达性分析；第 6 章为空间格局，阐述了空间格局测度的方法，着重介绍了点格局、空间句法及景观格局的分析方法；第 7 章为空间插值，在介绍方法分类体系的基础上，重点阐述了地统计学的理论和方法模型。

在教材编写过程中，参考了大量国内外优秀教材、文献资料和科研成果；硕士研究生徐梦雅、李泽堃、谭冰清、张倩雯、万琛、皮建华等承担了大量的资料搜集和数据整理工作，在此向被引用资料的作者及这些学生表示特别感谢。作者特别开设了微信公众号(wurg2016)，方便与广大读者交流 SDA 的理论和实践，敬请关注。

由于作者自身知识和水平的局限，教材中难免有疏漏之处，敬请读者批评指正。

作　者

2018 年 12 月

目　录

第 1 章　空间数据分析概论

1.1　空 间 数 据

空间数据(spatial data)是对现实世界中空间事物和现象时空特征及过程的抽象表达和定量描述。这个表达过程首先需要对现实世界进行高度的抽象，建立概念模型来描述对空间实体(spatial entity)的认知、抽象与概括；然后建立适用于计算机存储与表达的数据模型，空间实体被数字化为空间对象(spatial object)，用以记录和描述空间实体的位置、形状、大小及其分布特征等，空间对象构成了空间数据分析中可操作和分析的基本数据单元，具有定位、时间和空间关系等特性。其中，定位特性是指任何空间对象在已确定的坐标参考系统中都只有唯一的空间位置；时间特性是指空间对象的获取是在确定的时间中获取的，同时它的属性随着时间的变化而变化；空间关系特性通常用拓扑关系来表示，指空间对象之间的拓扑特性。空间数据具有的特性导致了空间数据分析和传统的统计分析有很大区别，理解和认识空间数据的特性和性质是进行空间数据分析的重要基础。本章主要从空间数据分析的角度对空间数据的性质与特点进行介绍，如图 1-1 所示。

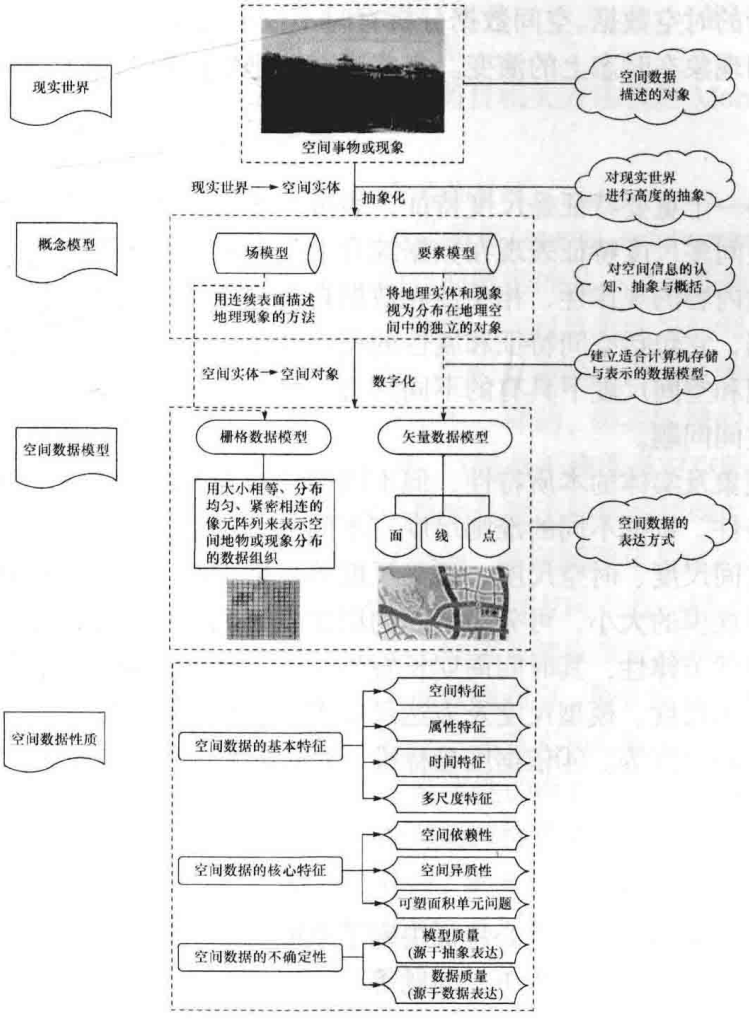

图 1-1　空间数据的认知、表达与性质

1.1.1 基本特征

1. 空间特征

空间特征是空间数据最基本的特征，空间数据记录了空间实体的空间分布位置、几何形状及与其他空间实体的空间关系等空间信息。空间实体的位置通常采用坐标系统进行描述；空间实体的几何特征表示了空间实体的大小、形状及空间维度；空间实体可以依据其维度分为点、线、面、体四种基本类型；空间关系描述了空间实体间的相互关系，在空间分析中起着关键作用，主要包括拓扑关系、方位关系、距离关系等。其中，空间拓扑关系主要描述空间实体间的相交、包含、邻接、相离等；空间方位关系描述空间实体间的绝对方位和相对方位；空间距离关系主要用来度量空间实体间的远近程度。

2. 属性特征

空间数据的属性特征描述了空间数据的内涵和性质，属性数据随着空间实体的不同而变化。属性特征可以从不同的角度进行定义，通常分为定性与定量两种；从空间数据分析度量的角度，属性数据可以归纳为名义属性、次序属性、间距属性、比率属性及周期属性等。

3. 时间特征

空间数据的时间特征描述了空间实体随着时间变化而变化的特征。这种特征是指空间数据对空间实体的描述总是与其采集的某一特定的时间或时间段对应，顾及时间特征的空间数据构成了更为复杂的时空数据。空间数据分析有时不仅要考虑空间实体在空间上的分布特征，同时也要考虑空间现象在时态上的演变。有些研究也把空间数据分为空间特征数据与时间特征数据两类。

4. 多尺度特征

空间数据的另一个重要特征是尺度特征，表现在不同观察层次上的信息被表达、分析的详细程度不同。空间多尺度特征表现在数据综合上。数据综合类似于数据抽象或制图概括，是指根据数据表达内容的规律性、相关性和数据自身规则，可以由相同的数据源形成再现不同尺度规律的数据，它包括空间特征和属性的相应变化。多尺度的空间数据反映了空间现象及实体在不同时间和空间尺度下具有的不同形态、结构和细节层次，应用于解决宏观、中观和微观各层次的空间问题。

尺度是空间现象及实体的本质特性，但不同的学科对尺度的定义不同，其定义取决于尺度使用的环境和条件。依据不同的准则尺度有不同的分类形式：①依据兴趣领域，尺度可以分为空间尺度、时间尺度、时空尺度、语义尺度等。数据的空间多尺度是指空间范围大小或地球系统中各部分规模的大小，可分为不同的层次；时间多尺度指的是空间过程及空间实体的特征有一定的自然节律性，其时间周期长短不一。②依据研究过程，尺度可以分为现实尺度、数据尺度、采用尺度、模型尺度及表达尺度等。③依据研究范围，尺度可以分为宏观尺度、中观尺度及微观尺度等。④依据度量标准，尺度可以分为命名尺度、次序尺度、间隔尺度及比率尺度等。

1.1.2 核心特征

在地学问题中，空间数据的核心特征决定了空间数据分析的特殊性，这些核心特征包括空间依赖、空间异质性、可塑面积单元问题等。

1. 空间依赖

空间依赖(spatial dependence)被认为是空间数据最基本的特质。空间依赖的存在决定了地理学空间的有序性与决定性。Tobler(1970)指出：空间上距离相近的实体之间的相似性比距离远的实体间的相似性大[①]，可以视为对空间依赖的一种定性描述。空间依赖的含义是空间上某一位置的空间现象与其自身及近邻位置上的同一空间现象有关。空间依赖产生的原因是十分复杂的，一般认为是空间实体的相互作用、空间现象的集聚、扩散及各种测量误差等造成的。一般而言，观测数据的采集通常是和空间单元相关联的，如人口普查单元、行政区域等，这将导致测度上的误差。当采集数据的边界不能精确地反映产生样本数据的基础特征时，将会表现出空间依赖(图 1-2)。

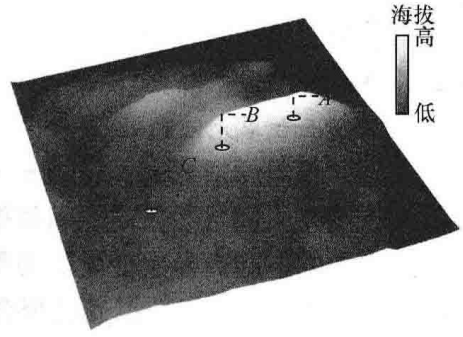

图 1-2　空间依赖的表现

对于真实地形中 A、B、C 三点的高程采样数据，因为 A、B 间直线距离比 A、C 间距离近，所以采样得到的 A、B 高程值之差更小，即“空间上距离相近的实体之间的相似性比距离远的实体间的相似性大”。对于大量的高程采样数据而言，这一现象可能表现为空间依赖

空间自相关可以视为空间依赖的定量描述，空间依赖程度是通过空间自相关量测的，这是两个直接关联的概念。空间自相关的指标体系多样，大体可以分为两种类型：全局观测和局部尺度。全局方法是指对研究区域的整体给出一个参数或指数，而局部方法提供和数据观测点等量的参数或指标。常用的度量空间自相关方法包括 Moran's I、Geary's C、Ripley's K、Join-Count 指数、半变异函数等。

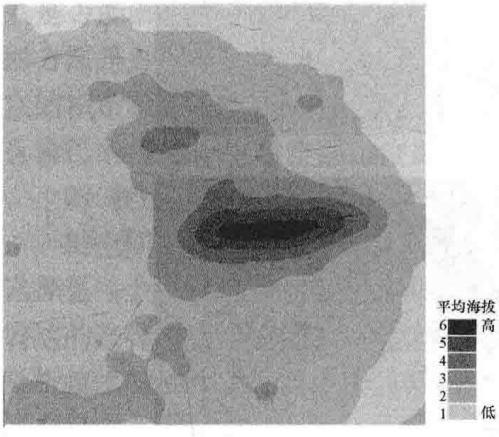

图 1-3　空间异质性的表现

根据高程采样点绘制等高线图，在 1～6 级的不同海拔带上会有不同的植被分布，而相同的高度带内往往分布同类型植被。这张图突出表现了高程属性在不同高度带的空间异质性和同一高度带内的局部自相关性

2. 空间异质性

空间异质性(spatial heterogeneity)是与空间依赖相对应的空间数据的另一个重要性质，表现为空间实体间的差异性。空间异质性与空间上行为关系缺乏稳定性有关，意味着描述空间关系的参数在研究区域的不同地方是不一样的，但在区域的局部其变化是一致的(图 1-3)。空间非平稳性是对空间异质性的描述，表示空间异质性是源于局部的特殊性，空间现象在空间上是非平稳的。各向同性是与此密切相关的一个概念，即假设模式在所有方向上是一样的。空间异质性在根本上是源于地球系统在演化过程中的分异结果，对于大部分空间数据而言，假设空间过程的非平稳性和各向异性可以更为真实地反映空间现象的实质。

空间异质性的存在导致在空间数据分析过程中，需要强调对局部性质的识别和分析，否则很难保证结果的可靠性，甚至得到错误的结论。而一般的全局模型和全局统计量对空间过程之间的复杂相互作用进行了平均，从而得到的是单一结果，这样

① 美国地理学家 Tobler(1970)提出：“Everything is related to everything else, but near things are more related than distant things”，也称为 Tobler 第一定律。

可能导致空间数据误差和不确定性有空间集聚的倾向，即分析结果中某些空间区域出现较大的误差和不确定性。空间异质性的定量分析方法主要包括局部 Moran's I、局部 Geary's C、空间非平稳性指数等。另外，在实际分析中，一般是利用对空间中有限的观测数据，根据问题的性质、有关的理论及研究人员的经验归纳分析出空间数据存在的关系，并没有实现充分利用样本信息对每一个样本进行估计，在分析时这种基于经验的归纳可能会遗漏没有考虑或发现的关系或性质。

3. 可塑面积单元问题

可塑面积单元问题(modifiable areal unit problem，MAUP)是对空间数据分析结果产生不确定影响的主要原因之一，是空间数据的另一个核心特征，表现为空间数据分析的结果随着面积单元定义的不同而发生变化。面积单元对分析结果的影响主要体现在两方面：①尺度效应(scale effect)，即当空间数据通过聚合而改变其粒度大小时，空间数据分析的结果也随之变化。②划区效应(zoning effect)，即在同一粒度或聚合水平上，不同的分区方法将导致不同的分区结果(图 1-4)。

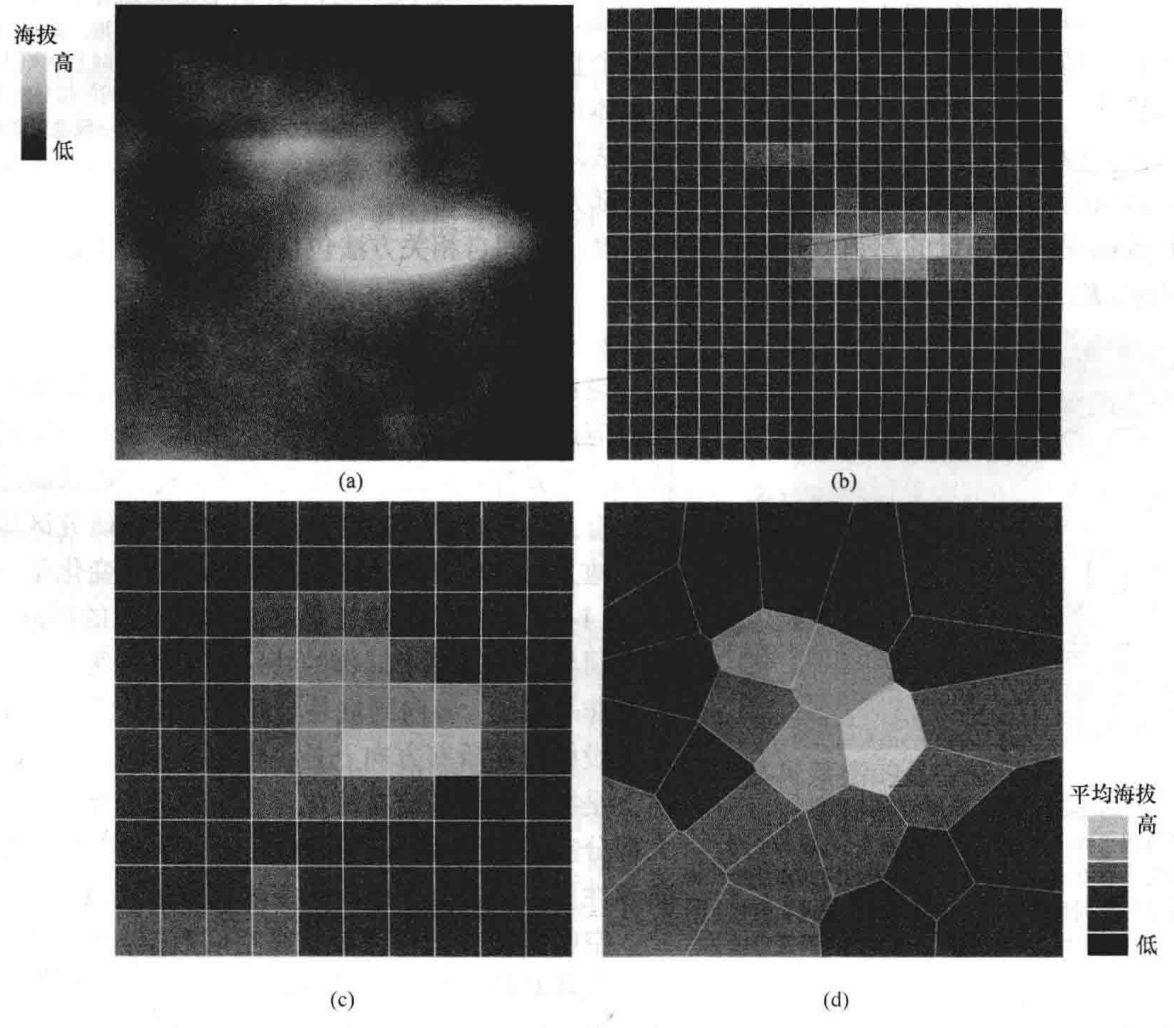

图 1-4　尺度效应和划区效应

对比图(a)中的真实 DEM 数据，图(b)和图(c)反映了尺度效应对于求取平均海拔的影响，图(c)和图(d)反映了划区效应对于求取平均海拔的影响，这说明在某一特定尺度的分析结果很难推广到其他尺度

随着空间单元的聚合和划区方案的不同，空间分析的结果也产生相应的变化。因此必须注意分析结果仅对于所采用的面积单元有效，在其他尺度上则不尽然。将某一尺度上的结果推广到其他精细的尺度上将导致“生态谬误”(ecological fallacy)，这是生态学中关于可塑面积单元问题的术语。空间现象及实体是多尺度的，分析时注意空间聚合的尺度和划区效应对于正确的空间数据分析过程及其结果的获取是非常重要的。

1.1.3　不确定性

空间数据的质量受到从地理世界到抽象概念模型及从映射到数据模型两个阶段的影响，一个体现在地理空间的抽象表达方式，包括选取的属性、量测的方式等；另一个体现在确定表达模型后空间数据量测的确定性，包括其地理坐标、属性值等。

1. 模型(抽象表达)质量

模型质量是评测空间数据不确定性的第一阶段，是建立映射复杂现实世界所使用的模型的表达质量，即根据选择的有限空间对象、空间关系或变量来离散化表达现实世界的程度。模型的质量可以根据表达的精度、清晰性、完整性和一致性(对象如何被表达)进行评估。模型质量评估的内容包括空间表达方式的有效性及具体表达方式的合理性，如像素分辨率和聚合程度等。具体包括以下两个方面。

1) 指标有效性

在模型质量中，属性表达的不确定性反映在属性的选取上。空间数据分析中，体现在确定研究主题后，建立映射指标体系时对变量或要素的选择，以及映射的算法和逻辑操作上。应当考虑选取的空间数据是否能够有效地表达对应的地理现象，以及选取的空间数据指标体系是否能够完整表达。另外还应当考虑的是，对于不同的研究区域而言，同一指标体系及映射方法是否同样适用，例如，基于过度拥挤、失业率、住房拥有率、轿车拥有量四个维度建立的城市社会不均等性尺度分析不一定适用于郊区，因为存在郊区的人口虽然倾向于不能拥有自己的房屋，生活空间过度拥挤，但却普遍拥有机动车的情况。

2) 表达合理性

对空间现象及实体表达方式的选择，受到特定研究及分析的空间尺度等影响。表达方法的选择是一个十分复杂的选择，例如，一个空间现象可以抽象为点、线、面的形式来表示，这种表达的选择不仅对用于获取这个空间现象的数字化对象类型具有影响，也会进一步对空间数据分析技术的选择及后续使用的可视化形式产生影响。表达方式的选择受到相关理论、研究主题、已有经验、技术、概念认知等多方面的影响，没有绝对的正确性。例如，对于一个面向对象的表达(如城市)，若在分析中不需要考虑其内部结构及性质，则其按点状对象的表达可以视为是合适的；而个体层面的数据往往倾向于给定点的位置信息来表达，精细获取的地理编码数据提供了在位置上的精确信息，但可能位置数据集的精度对于研究不一定合适；在一些研究中，过于精细化的数据可能产生伦理问题，如基于家庭统计的收入信息可能侵犯个人隐私等。

针对栅格数据表达的不确定性，主要体现在栅格数据结构将空间划分为等面积的像元，空间对象基于单元格的分类构成，每个单元格只有一个值。然而现实中可能存在混合像元的情况，即某一个单元格不完全由一个属性构成，而栅格单元格的值表示的是混合像素的优势值或中心值。单一结果的表达会导致空间信息的丢失，因而基于栅格数据表达将一定程度上扭曲空间对象的形状，这种表达的不确定性与像元的分辨率有关；针对矢量数据表达的不确定性，主要体现在分析空间尺度的影响上。例如，点数据是表示某些类型的社会经济数据较

好的方式，如个体信息统计等。但过于详细的信息可能导致侵犯隐私、数据集规模过大等问题；在一些研究中，出于多种原因，经常先精确定位个体，然后获取近似的数据或将数据按区域水平聚集表示，经常会导致数据空间表示的不合理性，不能充分展示数据的分布等特征问题。这种不确定性由聚合程度所决定，产生的原因是地理空间不能自然地形成分析单元，离散化方式不能很好地表现现实世界相关性或某种现象的不均匀性。另外，聚合程度还将影响数据集的“噪声”。某些区域的平均值可能基于很多数据点，而另一些则基于少量的数据点，在进行统计检验时，会降低区域间比较的可靠性。如果过程是随机分布的，则一般样本的总体越小，误差的方差就越大。

2. 数据质量

从数据使用者的角度，高质量的数据是能充分满足用户需求的数据。数据质量评估是针对给定的数据集，选择合适的评估指标，从而描述数据定义或模型满足相应评价规则的程度。数据质量评估在依赖二次数据源(即间接获取的数据)进行空间数据分析的领域十分重要。数据质量的评估包括空间实体位置和属性数据的质量两个方面，二者又是相互依赖的，因为空间数据还具有时间特征，所以这三者都将对数据质量产生影响。

数据质量可能是空间不均匀的，误差结构在整个地图上可能变化，即数据质量可能是有空间异质性的。误差的异质性可能来自于测量过程和局部地表特征之间的相互作用。例如，遥感影像上的误差随传感器类型不同和地形性质的不同而不同。

Guptill 和 Morrison(1995)提出了空间数据质量评估的通用标准，定义了空间数据质量包括的七个维度：数据源及其描述、定位精度、属性精度、完整性、逻辑一致性、时序规范、语义准确性(包括特征、关系、属性编码的准确性，或给定的表达规则描述的准确性)。这些不确定性是相互联系的。空间数据分析强调数据以下四种类型的不确定性。

1) 数据的准确性

数据的准确性表现在数据误差大小，即观测值和真值之间的差别，因为测量过程不可避免的不精确性，以及对对象定义的主观性，任何测量都会产生一定的误差；改善实验过程及测量的方式，如提高仪器质量和测量的技术，可以减少这种类型的不确定性，尽管不能完全消除它。

2) 分辨率或尺度

分辨率对数据质量影响最重要的方面是数据的空间尺度，另外还包括数据的时空分辨率及变量分辨率。

空间分辨率对地理现象的表达具有影响。高分辨率的数据，空间单元较小，可能包含高水平的噪声，对识别结构产生影响。例如，小区域的疾病率常表现出高水平的变化，即统计意义中的极值率较大；但在描述具体信息时，越小的空间单元对空间对象的描述就越精确。另外，当数据集处于不同的空间框架或数据集具有不同的分辨率尺度时，空间分辨率将对属性值的精度产生影响。

时空分辨率是指数据集时间推移周期不同，可能产生类似于空间分辨率导致的问题。例如，根据时间周期长度进行聚合计数可能掩盖小尺度上的变化，但会增加样本的计数，计算的比率更加稳定。

变量分辨率是指属性或其他测量结果的精度。在空间数据可视化中，它是指分类的详细程度，包括分区标准及数量的选择等。

3) 一致性

一致性被定义为数据在数据库中没有矛盾，它是指关于空间数据结构的逻辑规则及属性

规则及其用于描述数据集和其他数据的数据兼容性。可建立来自数学理论的规则和公式检查，以检查数据集内部和不同数据层之间的数据一致性，如空间对象的拓扑规则检查、属性值之间的矛盾性检查等。

4) 完整性

数据的完整性反映了数据集缺失数值、不足计数和过量计数等问题，与数据误差可能存在重叠。空间数据的不完整性可能会对比较分析、空间变化分析和描述带来影响。另外，还存在随着时间的推移，研究对象发生变化的情况，这需要在数据库中及时更新相关属性信息或者按新的对象对数据库进行补充、删除、更新等操作。

1.2　空间数据分析

1.2.1　内涵

空间数据分析从地理空间的视角描述和分析问题，是地理学日益受关注的研究手段与方法。空间数据分析本质上是一种思维方式和工具，且具有明显的多学科交叉特征，其显著特点是思想多源、方法多样、技术复杂。因此，不同领域对空间数据分析的方法内涵和外延不同，关于空间数据分析的称谓也有所不同，如空间分析(spatial analysis)、空间数据分析(spatial data analysis)、空间统计(spatial statistics)、地统计(geostatistics)等。空间统计和空间数据分析主要包括点状分布现象的空间格局，着重研究建立各种空间结构回归模型及其稳健解法，以及当今研究中引入的智能计算方法如神经网络、遗传算法等。地统计学的主要内容是利用统计学中的矩方法、变异函数和最小二乘法进行空间线性推测的克里金(Kriging)方法。鉴于空间数据分析内涵的丰富性，学者们从不同的角度对空间数据分析的定义进行了分析总结。

Unwin 是较早对空间数据分析概念进行论述的学者。他将地理数据分为点、线、面和空间连续性数据四类，并进行参数描述和图形分析(Unwin, 1981)。O'Sullivan 和 Unwin(2003)认为在不同领域中至少存在四种相互联系的空间数据分析概念，分别是空间数据操作(spatial data manipulation)、空间数据分析、空间统计分析(spatial statistical analysis)及空间建模(spatial modeling)。Ripley(1981)则从空间统计学的角度对空间数据进行了运算与分析。Goodchild(1987)将空间分析定义为对数据的空间信息、属性信息及二者共同信息的统计描述或说明，并首次对空间分析的框架做了较为系统的研究。他将空间分析分为两大类：一类是“产生式(product mode)分析”，通过这些分析可以获取新的信息，尤其是综合信息，实质是提取显式存储空间信息；另一类是“咨询式(query mode)分析”，旨在回答用户的一些问题，其实质是提取隐式存储空间信息。Haining(1980)认为空间数据分析是基于空间对象及空间布局的地理数据分析，对实体的空间分布进行了定量研究与描述。Bailey 和 Gatrell(1995)认为空间数据分析是指应用逻辑或数学模型分析空间数据或空间观测值，为制定规划和决策服务的一项技术。李德仁(1993)认为空间查询和空间数据分析是指从 GIS 目标之间的空间关系中获取派生的信息和新的知识。郭仁忠(2001)认为空间数据分析是基于地理对象位置和形态特征的数据分析技术，其目的在于提取和传输空间信息。

关于空间数据分析概念及内容的讨论很多，且在不同领域中应用与论述时各有侧重点。综合现有的研究，通常可以将空间数据分析视为涵盖了空间分析、空间统计与建模等内容的一般意义上的概念，将空间数据分析定义为分析中包含空间对象位置信息的技术与方法(Longley et al., 2005)。一般可以将空间数据分析的内容归纳为：①空间图形分析，包括空间

数据的量算、几何操作，如长度、面积、形状的量测，空间重心的计算，叠加分析，缓冲区分析，相交分析及基于空间关系的查询等。②空间数据统计分析，主要是采用统计方法对空间数据特征、性质进行描述或探测。主要包括两部分内容：一部分是对空间数据的探索性分析，如研究空间数据的基本统计特征，识别异常值，为后续分析做准备等；另一部分是空间统计分析，即发展专门的空间统计分析方法对空间数据的性质进行描述和解释，如空间点格局分析、景观格局分析、地统计学等。③空间回归模型与建模，借助空间回归模型，旨在对空间现象之间的依赖关系或交互作用进行描述，主要包括构建模型、预测空间过程及结果，如社会经济学中的人口迁移模型。④地理模拟与智能计算模型，如元胞自动机、多智能体、神经网络等。基于此，本教材将空间数据分析定义为分析空间数据、挖掘空间信息、统计空间规律、解决空间问题所涉及的基本理论、方法与技术的总称。

1.2.2 发展

地理学家很早就已经采用空间数据分析的方法对各类地理问题进行研究，如样方方法(Gleason, 1920)、应用二元权重对间隔数据的检验(Moran, 1950a; Geary, 1954)、植物群落空间分布模式的研究(Goodall, 1952)、最近邻距离方法(Clark and Evans, 1954)。早期的研究大多从几何学观点出发，侧重于用最近邻方法(nearest neighbor analysis, NNA)及用位置数据描述空间点模式的分析。空间数据分析概念的提出和飞跃式发展很大程度上获益于 20 世纪 20 年代开始的数量地理(quantitative geography)革命。数量地理学是应用数学思想方法和计算机技术进行地理学研究的科学，早期以一般统计方法的应用为主，通过地理事物与现象空间关联的分析，把握人流、物流、信息流等方面的地理信息。20 世纪 60 年代在电子计算机技术推动下的计量革命为地理学带来了新工具与新思维。计量地理学(geographimetrics)的出现改变了地理学以记录和描述地理现象为主要研究手段的传统，促进了地理学定量分析技术的发展，对数学在地理学中的应用起到了普及和推动作用。1970 年，Tobler 提出了描述地理现象空间作用关系的“Tobler 定律”，即“Everything is related to everything else, but near things are more related than distant things. ”，这一定律的提出使得地理现象的空间自相关性在研究中得到重视。Cliff 和 Ord(1969, 1973)提出了空间自相关的概念，使研究者能够从统计上评估数据的空间依赖程度，并展示了在空间随机性条件下如何检验回归分析中的误差，揭示了空间加权矩阵的本质。20 世纪 70 年代中期多元统计方法和随机过程引入地理学研究领域，70 年代末期地理学者引进数据处理技术，并与数据库和信息系统技术相结合，利用数学工具深入研究了区域自然、社会、经济、人口等过程的各类数学模型，阐明了地理现象的空间分布规律和模式，进行了有关地理结构和地理组织的演绎。空间数据分析发展中兼容并蓄了系统论、控制论、决策论、信息论等其他学科的相关内容和方法，极大地丰富了其理论基础。统计学家 Ripley 于 1981 年对空间分布模式进行了卓有成效的研究和总结，提出了测度空间点模式的 K 函数方法等。OpenShaw(1984)等对空间数据中的可塑面积单元问题进行了深入的探讨。Matheron 的《区域化变量理论》为后来的地统计学的迅速发展奠定了基础。随着对地理数据空间特质的重视和空间统计模型的提出，以描述全局特征为主的传统统计分析方法逐渐向以描述局部特征为主的统计分析方法转变，在这一时期，考虑空间自相关的空间回归模型或空间自回归模型被提出并在计量地理学中得到了重要的应用。在这一阶段，围绕地理现象的空间本质和地理数据的空间性质，建立了地理学的空间数据分析理论和方法体系。这一阶段是计量地理学和现代空间数据分析方法发展过程中十分重要的时期，也奠定了现代空间数据分析的理论基础。