

21世纪高等院校教材

空间数据分析方法

◎ 王远飞 何洪林 编著



科学出版社

www.sciencep.com

21 世纪高等院校教材

空间数据分析方法

王远飞 何洪林 编著

科学出版社

北京

内 容 简 介

本书较全面地介绍了空间数据的分析方法。在系统介绍空间分析的概念、内容、结构体系和研究进展的基础上,对空间数据的性质进行了探讨;然后以空间数据的类型为线索,分别研究了空间点数据、面数据、连续数据的分析方法;并讨论了空间回归(包括地理加权回归)方法;最后介绍了 GIS 软件系统中的地图代数语言及其建模技术。本书是一本面向 GIS 专业空间分析课程的教材,通过大量的应用实例,展示了空间分析方法的应用价值,较系统地反映了空间分析方法和技术的最新进展与成果。

本书既可用作高等院校地理、生态、环境、卫生、经济等专业的本科生教材,同时也可供相关领域的研究者和技术人员参考。

图书在版编目(CIP)数据

空间数据分析方法/王远飞,何洪林编著. —北京:科学出版社,2007
21 世纪高等院校教材
ISBN 978-7-03-018966-0

I. 空… II. ①王…②何… III. 地理信息系统-高等学校-教材 IV. P208
中国版本图书馆 CIP 数据核字(2007)第 067399 号

责任编辑:郭 森 杨 红 李久进 / 责任校对:陈丽珠
责任印制:张克忠 / 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>

双 青 印 刷 厂 印 刷

科学出版社发行 各地新华书店经销

*

2007 年 6 月 第 一 版 开本:B5(720×1000)

2007 年 6 月 第一次印刷 印张:15 3/4

印数:1—3 000 字数:298 000

定价:24.00 元

(如有印装质量问题, 我社负责调换〈双青〉)

前 言

20 世纪科学技术的飞速发展促进了地理学研究的飞跃。对地观测技术、计算机技术、网络通信技术的重大突破是推动地理学大发展的动力。在地理学近代发展中, 计量革命和地理信息技术这两大激动人心的事件, 使得地理学研究的理论方法和技术手段焕然一新。20 世纪 60 年代的计量革命推动了地理学研究从定性走向定量; 产生于同一时代的地理信息技术虽然滞后于计量革命, 但是其影响却更为深远。地理信息技术的出现不仅为地理学研究提供了新的技术支持, 而且形成了以地理信息科学为核心的新的研究领域、理论和方法。在地理信息系统技术的支持下, 地理学中的理论、模型在应用中焕发光彩。

空间分析是地理信息科学的核心领域, 它是地理学分析方法中区别于其他定量分析方法的推演分析框架和体系。较之先前的计量革命中注重移植统计分析方法, 空间分析则围绕地理学问题的“空间”本质建立模型方法。计量地理学的主要技术方法是直接套用数理统计方法建立地理学模型, 解释地理问题, 虽然有一些空间分析的模型方法, 但不是主流; 由于统计方法关注的是全局性的建模问题, 缺乏空间作用关系的描述, 丢失了地理预见性, 致使模型在解释地理问题时遇到困难。这也是计量地理运动遭受诟病的原因。事实上, 计量地理学家已经注意到统计方法忽略空间特征对地学分析带来的影响, 在统计分析方法的基础上引进描述空间关系和作用的特征量, 发展适合于描述地理问题的模型方法是空间分析技术的主要任务。面向地理学问题的本质, 学界提出了地理学第一定律, 建立了空间自相关、空间集聚性、MAUP 问题等一系列地理概念和理论模式, 发展了空间统计、空间点模式、空间回归、地统计、空间相互作用等描述地理问题的分析方法和技术, 至今已经形成了较为完整的体系。这些内容和方法正是空间分析的精髓, 也是本书关注的主要内容。

随着 GIS 技术的广泛使用, 这些空间分析方法被各个领域所熟悉, 地理分析方法和技术日益表现出解决空间相关问题的强大能力, 原来局限于理论分析的模型和方法正在演变成各个领域日常工作的普通工具。目前, 关于空间分析的教材已经丰富起来, Internet 以其强大的知识扩散能力成为人们获取知识的平台, 本书所涉及的很多内容从数据到实例都可轻易地从 Internet 上获取。特别需要提

到的是书中介绍的很多内容有些已经是主流 GIS 软件系统的基本分析模块，有些以扩展模块的形式可随时加载，例如，ArcGIS 中的空间数据分析工具箱，MapInfo 的 Vertical Map 等；部分特殊领域使用的专业性强的分析方法也以各种软件包（包括自由软件）的形式出现，例如，开放源代码的 R 语言统计分析系统中集成了包括从一般的数理统计方法到空间统计分析方法的各种技术，这些软件包大多能够与 IS 软件包自由地交换数据。

规划、管理、决策以及地学试验研究中对地理信息技术的巨大需求，促使大学中很多学科领域纷纷设置地理信息科学专业，空间分析逐渐成为大学的地理学及其相关学科领域的基本课程，而相关教材却跟不上教学发展的需要，系统介绍这些内容的书籍较少。因此，编写一本空间分析领域的教材，对于地理学及相关专业的学生掌握必备的技能、提高分析问题和解决问题的能力是十分有益的。作者根据大量的国内外著作与文献尝试编写了这本教材，在写作过程中不仅参考了列于主要参考文献之中的文献，而且从中引用了许多内容，此外还参考了很多其他的文献和大量的网络资料以及国外著名学者的电子版教学讲义等。对这些文献的阅读、学习、翻译、整理构成了本书的基础。

空间数据分析包括的内容十分广泛，作为一本教材，既要考虑内容的全面性，又要有所取舍。因此，在参照国外教材体系及文献内容的前提下，考虑国内教学和课程体系的需要，本书的内容主要包括空间数据的性质、探索性空间数据分析与可视化、空间点模式、空间面模式、空间回归、空间连续数据建模、地图代数与栅格数据分析、空间建模等。第 1 章绪论部分对空间分析的概念、内容、框架体系和研究进展作了系统介绍；第 2 章探讨了空间数据的性质及其有关问题；第 3 章研究了近年来受到重视的探索性数据分析和可视化方法；然后以空间数据的类型为线索，在第 4、5、7 章分别研究了空间点数据、面状数据、连续数据的分析方法；在第 6 章中对空间回归（包括地理加权回归）方法进行了深入探讨；第 8、9 章介绍了 GIS 空间分析实现的内容，包括 GIS 软件系统中的地图代数语言及其建模技术。全书较系统地反映了空间数据分析方法和技术的最新进展与成果。其中，王远飞撰写了第 2~6 章，并与何洪林一起编写了第 1 章及第 7~9 章。

在本书的编写过程中得到了很多人的帮助，首先感谢香港中文大学梁怡教授，是他为我们创造了在香港中文大学访问学习、全面接触和了解这一领域国际发展的机会，而且梁怡教授本人又是该领域内造诣深厚的著名专家，其学识使我们受益匪浅。这期间对大量国外文献的阅读，激发了我们的浓厚兴趣，是我们

编写这本教材的原动力；其次感谢张超教授，是他的不断鼓励和鞭策才使我们最终完成了本书的编写。此外，感谢王铮教授在本书编写过程中的一贯鼓励和支持。研究生朱海燕、孙健鹤、丁鹏飞、罗向欣、刘黎明、武占云、丛晓男等在文献资料整理和文稿校核方面付出了辛勤劳动，对他们表示诚挚的谢意。

由于时间、资料、知识、精力等多方面的局限性，书中疏漏难免。望对这一领域有兴趣的读者、专家不吝赐教，以使作者完善本书。

作 者

2007年4月

目 录

前言

第 1 章 绪论	1
1.1 空间分析的概念与研究进展	1
1.2 空间分析的研究内容	5
1.3 空间分析与地理信息科学.....	11
1.4 小结.....	12
第 2 章 空间数据的性质	13
2.1 地理世界的概念模型与数据模型.....	13
2.2 空间数据的性质.....	20
2.3 空间数据的不确定性.....	26
第 3 章 探索性空间数据分析与可视化	31
3.1 关于 EDA、ESDA 与可视化	31
3.2 EDA 与可视化的基本方法	33
3.3 ESDA 与空间数据可视化.....	47
3.4 交互技术与 ESDA	52
第 4 章 空间点模式方法	57
4.1 空间点模式的概念与空间分析技术.....	57
4.2 基于密度的方法——样方计数法与核函数法.....	60
4.3 基于距离的方法.....	71
4.4 G 函数与 F 函数	76
4.5 K 函数与 L 函数	83
4.6 K 函数方法的扩展——二元模式与空间-时间模式	89
第 5 章 面状数据空间模式分析方法	94
5.1 空间接近性与空间权重矩阵.....	94
5.2 面状数据中趋势分析	100
5.3 空间自相关的概念	104
5.4 名义变量的空间自相关测度——连接计数法	107
5.5 空间自相关统计量——Moran's I 和 Geary's C	110
5.6 广义 G 统计量	113

5.7	局部空间自相关统计量	115
第 6 章	空间回归分析	120
6.1	回归分析方法	120
6.2	空间自回归模型	121
6.3	空间回归模型的实例	129
6.4	地理加权回归模型	132
6.5	GWR 应用的实例研究	141
第 7 章	空间连续数据分析方法	147
7.1	探索性分析方法	147
7.2	趋势面分析	151
7.3	连续数据的空间依赖性测度——协方差图和半方差图	152
7.4	克里格方法	160
7.5	克里格建模方法实例研究	171
第 8 章	地图代数与基本地理计算	176
8.1	理解栅格数据	176
8.2	地图代数中的基本运算	181
8.3	地图代数中的函数与类型	188
8.4	局部函数	190
8.5	邻域函数	196
8.6	类区函数	200
8.7	块函数	202
8.8	全局函数	203
第 9 章	GIS 空间建模	209
9.1	空间建模	209
9.2	距离函数的应用——缓冲区、区位配置和最短路径	213
9.3	地图的比较分析——交叉表和 KIA 指数	218
9.4	函数的综合应用	222
9.5	栅格数据集的滤波处理	227
9.6	空间决策与实例分析	232
	主要参考文献	240

第 1 章 绪 论

地理信息技术的出现,促进了地理分析方法的发展和应用,开创了地理学研究的崭新时代。与人类活动密切相关的各种信息大多是空间信息或地理信息,但是,长期以来由于缺乏有效的数据采集手段和分析技术的支持,空间数据的分析处理无法满足科学决策和管理的需要,导致地理学研究在决策管理中的重要价值始终不能得到很好的体现。20世纪60年代诞生的地理信息系统(GIS)技术是空间信息采集、管理、分析和表示的工具或技术,地理学理论、知识、分析方法等随着GIS技术的广泛应用而得到重视和发展。目前GIS技术的应用已经覆盖了城市管理、生态环境管理和规划、交通规划、森林管理、应急服务、自然风险分析、市场分析、物流管理、公安司法等诸多领域。在GIS的应用中,各个领域面临的共同问题是在日益丰富的数据环境中,提供怎样的分析技术来挖掘数据中的规律、提取地理知识促进管理创新和科学决策,是空间分析应当解决和回答的问题。目前,各种GIS原理和应用方面的书籍都强调空间分析是GIS的核心功能,但是一般的GIS软件或商业GIS软件中的空间分析都有局限性,远远满足不了实际应用领域对空间分析的需要。发展空间数据分析方法无论是对于地理学的发展还是GIS技术的深入应用都具有重要的意义。

1.1 空间分析的概念与研究进展

1.1.1 空间分析的概念

虽然地理学家应用空间分析(spatial analysis)方法研究地理问题的历史由来已久,但是空间分析作为一个独立的概念而使用是伴随着GIS技术而出现的。文献中关于空间分析、空间数据分析的提法多种多样,Unwin(2003)在其著作*Geographic Information Analysis*中进行了分析总结,他认为,不同领域的文献中至少存在4种相互联系的空间分析概念,分别是空间数据操作(spatial data manipulation)、空间数据分析(spatial data analysis)、空间统计分析(spatial statistical analysis)、空间建模(spatial modeling)。

1) 空间数据操作

主要出现在GIS中,但通常被称为空间分析(spatial analysis)。一般包括缓冲区分析,包含分析,相交分析,叠加分析,距离、面积、路径计算,以及基

于空间关系的空间查询等简单的数据分析功能。GIS的空间分析操作主要基于地理对象的几何特征，对于属性描述则主要表现在可视化的制图方面。

2) 空间数据分析

一般指对空间数据的描述性和探索性分析技术和方法，是所有空间分析过程中的一个重要步骤，特别是对于规模庞大的数据集，通过将数据图形化或地图化的探索性分析技术，研究数据中潜在的模式、异常等，为后续的分析做准备。

3) 空间统计分析

用统计的方法描述和解释空间数据的性质以及数据对于统计模型是否典型或是否如所期望。这里的统计方法是与传统的统计模型完全不同的空间统计方法。地理学家很早就注意到了空间数据所描述的地理现象是空间相关的，这一特征违背了传统统计理论关于独立性的假设，因此需要不同的统计分析方法来测度空间相关性并对数据进行统计分析，发展了专门用于空间数据分析的空间统计方法。

4) 空间建模

空间建模主要包括构建模型预测空间过程及结果。在地理学研究中，根据某些理论和假设，建立模型描述地理现象的分布模式和动态过程是相当普遍的研究方法，例如人文地理学领域中描述人和货物流动的空间相互作用模型，根据人口分布的服务设施区位分析和选址模型，环境过程中的污染扩散模型等。空间建模技术是空间分析的深入发展。

实际应用中，这些术语所指代的方法相互之间密切联系，很难给出一个严格的界限进行区分。在地理研究及相关研究中，基本上都包括了空间分析这四个方面的：数据在GIS环境中存储并可视化；描述和探索性的数据分析技术提出问题并建议相应的分析理论或模型；在此基础上通过空间统计方法建立统计模型；或在理论指导下对特定的问题进行空间建模分析和预测。

综上所述，我们认为空间分析是能够揭示出比数据本身更多的信息和知识的一组分析技术或方法。地理学家用空间分析技术或方法研究地理对象或现象的分布模式、动态演化过程和空间相互作用规律，发展和检验地理模型，增进对地理信息的理解，创新地理知识。

1.1.2 空间分析研究的进展

地理学家使用空间分析方法研究地理学及其相关的问题有着悠久的历史，但是空间分析技术被广泛使用，成为解决地理相关问题的重要分析方法，和计算机的出现、GIS的发展密不可分。20世纪60年代的计量革命和目前仍然在发展的计量地理学方法是空间分析的重要内容。地理学的计量革命改变了地理学以记述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。根据国内外大量的计量地理学的教材，我们不难发现在计量地理学中从数理统计领

域移植过来的统计分析方法所占有的主导地位,其主要内容包括相关分析、回归分析、聚类分析、因子分析等多元统计分析的内容;而空间模式、空间过程、空间相互作用等理论和方法在计量地理学中并没有作为重要的内容进行介绍,文献中广泛出现的计量地理分析方法主要是多元统计分析的内容。因此 Fortheringham 于 2001 年认为“线性回归”是计量革命的核心技术。由于计量地理方法专注于统计分析技术的应用,而忽略了地理问题空间本质,遭受到了学术界的批评和质疑。因为在大多数情况下,描述地理对象或现象的空间数据不再满足传统统计分析方法对数据的基本假设,如对数据的正态分布假设等。传统的统计分析方法是非空间的方法,用于地理建模是不充分的。

1970~1980 年是计量地理学方法或现代空间分析方法发展过程中非常重要的时代,在这一时期围绕地理现象的空间本质或地理数据的空间性质,建立起了地理学的空间分析方法或体系。

Tobler 于 1969 年提出了描述地理现象空间作用关系的“地理学第一定律”,Tobler 指出,“任何事物都是空间相关的,距离近的事物的空间相关性大”,这一定律的提出使得地理现象的空间相关性和异质性特征在研究中得到重视。

Clifford 在 1973 年出版的专著中揭示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性程度,他们清晰地表达了由自回归问题引起的建模错误,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,他们揭示了空间加权矩阵的本质,并提示了两个主要的自回归统计量 Moran' I 和 Gearcy' C 应用于统计检验的详细步骤。通过这样的工作,使得建模过程能够分析寻找更多的合适变量,以避免由于数据的空间自相关性引起的建模和结果的谬误。这些方法被描述为“空间回归模型”。

统计学家 Ripley 于 1981 年对空间点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 等对空间数据中的可塑面积单元问题(简称为 MAUP 问题,又称为生态谬误问题)进行了深入的探讨。这一问题在地理学研究中很早就被提出,其本质是空间尺度变化对于变量统计结果以及变量之间相关性产生的影响。问题分为两类,一类是聚集效应,另一类是划区效应。这些问题的提出对于正确地使用空间数据以及解释空间结果非常有价值,特别是 MAUP 问题对于某些区划问题有重要的实用价值。

随着对地理数据空间特殊性的重视和地理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。Anslin 等提出描述局部相关性的测度方法和统计量 LISA,成为研究某些现象分布模式的局部热点区域(hot spot)的重要方法。在这一时期考虑空间相关性的空间回归模型(spatial regression model)或空间自回归模型(spatial auto-regression model)被提出并在计量经济学中得到重要的应用,导致了空间计量经

济学的出现。

这些对地理现象或地理数据空间特征的研究以及地理空间统计方法和模型的提出成为现代空间数据分析的转折点，也奠定了现代空间数据分析的理论基础。

几乎同一时期，计量地理学中占有重要地位的空间相互作用模型在 20 世纪 70 年代也遭到了批评。空间相互作用模型的本质是牛顿引力模型的类比模型，这一模型企图通过应用引力的概念描述人口迁移、出行、交通流量等复杂的人文过程，模型本身对于城市交通规划、服务设施的布局等领域作用巨大。但是，由于这类模型缺乏行为背景的描述机制而限制了其应用价值的发挥。于是在 20 世纪 70 年代，开始注重对不同场合中个体选择行为的研究，如工作出行、购物出行、迁移等。与此平行发展的是离散的空间选择模型，提出了新的离散选择模型的变体，这种空间选择模型建立了微观水平上人口流动中的个体决策和其他可观测的宏观变量之间的联系。客观上讲，空间相互作用模型对于个体选择行为的研究需要大量的数据支持，而在当时的发展水平上，数据获取手段的缺乏成为限制这一模型发展的重要原因。随着空间数据获取手段的巨大变化，在处理人口、货物和资源在区域间流动的广泛变化的空间相互作用模型开始活跃，空间上的区位-配置问题成为这些模型重要的应用方向（如零售中心的定位和就业区位），并且越来越多的工作都基于 GIS 环境。因为，一方面 GIS 为这些模型的试验分析提供数据来源，另一方面 GIS 为模型运行的结果提供可视化。总之，GIS 是这些模型运行试验的平台。

进入 20 世纪 90 年代后，空间分析的发展和 GIS 的发展密切结合在一起。随着个人计算机的普及，空间数据获取不再成为分析地理问题的瓶颈，特别是 GIS 技术发展成熟为地理现象和过程的分析提供了新的平台，空间特征的研究受到了前所未有的关注，空间数据分析领域的研究十分活跃。在最近十多年的发展中，空间分析的关键技术发生了重要变化，地理信息系统和遥感等新的技术保证了空间数据的丰富环境，新的处理空间问题的分析模型和方法不断提出。由于分析过程受到不断增长的大量空间数据的驱动，从数据出发的探索性空间分析技术、可视化技术、空间数据挖掘技术、基于人工智能的空间分析技术等面向海量空间数据的分析方法受到重视，并且在最近几年中得到深入的发展。这些方法和技术对于大规模空间分析问题中的不精确性和不确定性有着较高的容许能力。20 世纪 90 年代是以 GIS 为计算环境的空间分析大发展的时期，其推动来自于以下四个方面（M. M. Fisher, A. Getis, 1997）：

- 1) GIS 数据革命极大地促进了空间分析在众多领域的应用。

- 2) 地理学家面临的数据环境发生了巨大的改变，大量的空间分析基于海量空间数据环境，迫切需要新一代的以数据为驱动的地理探索和建模工具，使得分析处理过程中数据丰富环境下的多维复杂性不被忽略。

3) 高性能计算机的出现,为需要复杂的空间数据处理和地理知识表示的空间分析活动提供了可行的环境。

4) 神经网络、遗传算法等可应用的实用智能计算工具提供了空间分析的新范例。基于计算智能的空间分析为改善空间分析技术和模型以满足大规模数据处理需求提供了基础。

1.2 空间分析的研究内容

当前,空间分析呈现出丰富的应用领域,和城市与区域问题、市场、交通和自然资源问题等构成很多联系。GIS和RS技术极大地增强了空间分析的需求。根据发展的趋势,本书主要的研究内容包括:数据模型与地理表达,空间数据的性质及其测度方法,空间分布模式,探索性空间数据分析与可视化,空间统计与地统计方法,空间回归模型,地图代数与空间建模技术等可集成在GIS环境中的空间分析方法和技术。

1.2.1 空间数据模型与地理世界的表示

空间数据和非空间数据的区别是需要深入探讨的主题,空间数据主要描述地理实体或现象的空间位置。在表达离散和连续现象时分为两类主要的空间数据。分类中的前者是一种实体型的世界观,空间现象被描述为0维的点对象,1维的线对象或2维的面对象。如果空间被描述为连续的现象,如温度、地形、污染物的浓度分布,这是场的世界观。分类中的后者通常根据对离散实体采样的空间位置来获得。

实体观允许空间对象拥有属性描述。典型的空间分析的目的在于观测单元的空间排列,但还可以考虑其他属性信息。如果分析方法仅仅考虑观测空间单元的属性特征而忽视了空间关系则不属于空间数据分析范畴。

鉴于空间数据模型是各类GIS原理和技术教材和著作中探讨的主题,本书主要从空间数据特殊性质的角度探讨空间数据模型,不涉及空间数据存储结构方面的内容。

1.2.2 探索性空间数据分析与可视化

根据前面的分析,空间数据分析的方法可在广义上归纳为探索性数据分析方法,数据可视化方法,空间统计方法,空间建模方法等。但是在很多分析中,这些技术都是结合起来使用的,而采用可视的方式显示数据一般是空间分析的第一步,其次才是探索可能的模式和可能的建模。

任何数据分析的第一步都应当首先对数据进行检查。利用画图 and 地图的方式

进行信息的视觉显示为研究人员提供了建立需要的假设和模型的拟合评价或预测能力的基础。实际上, GIS 出现之前, 地理学家就一直使用地图的方式可视化地表达地理现象的分布模式和空间关系, 地图也是决策分析的重要工具。广泛引用的著名的地图可视化的例子是 1853 年英国伦敦霍乱病暴发时, Snow 的地图分析为决策者提供了重要的依据。随着 GIS 技术的发展可视化的技术也产生了本质的变化, GIS 提供了交互方式进行动态地理空间数据的技术显示, 例如大部分商业 GIS 软件都提供了专题制图和分级分类显示。GIS 的可视化可被用于生产地图并且允许以交互的方式探索空间模式。

在 20 世纪 60 年代, 统计学家 Tuckey 就注意到了从样本出发、基于某些统计理论假设的统计方法在应用中的缺陷, 提出了从数据出发进行探索性数据分析 (EDA) 的方法。目前这一技术在海量数据环境分析中得到了深入的发展和应。数据探索性分析的目的在于期望在不对数据做出满足任何条件的前提下, 通过图形或地图的方法研究数据的特征, 为后续假设或模型建立做准备, 这一阶段也可以应用简单的分析模型。探索性阶段经常和可视化结合在一起, 在计算机支持下通过交互的方式研究数据的特征。

在空间分析中, 地理学家将探索性数据分析推广到空间数据的研究中, 提出了探索性空间数据分析技术 (ESDA), 将地图、统计图表、表格数据等综合在一起使用, 在 GIS 环境中这些数据通过多个窗口表示, 并使用交互刷新技术 (brushing), 数据在一个窗口的变化也相应地反映在其他窗口中, 提高了交互分析的性能。

1.2.3 空间数据的性质

空间数据与一般的属性数据相比, 具有特殊的性质, 例如空间相关性、空间异质性, 以及由尺度变化等引起的 MAUP 效应等。研究空间数据的性质对于空间数据的建模非常重要。空间数据的建模必须整合可能存在的空间依赖性才能更好地表示空间模式和空间关系。空间效应可能是大尺度的趋势也可能是局部效应, 一般前者称为“一阶”效应, 它描述的是某个参数均值的总体变化性; 后者则称为“二阶”效应, 是由空间依赖性所产生的。“二阶”效应表达的是空间上近邻位置上的数值相互趋同的倾向, 可通过其对于均值的偏差计算估计。传统的统计分析方法对于“一阶”效应能够有效地建模, 例如回归技术描述“一阶”效应。而“二阶”效应是对空间相关性局部特征的描述, 显然违反了传统数据分析技术关于数据独立性的假设, 适合的分析技术必须考虑引起这些局部效应的数据的协方差结构。

通常空间数据被模式化为平稳过程, 即假定当邻近的观测可能依赖时, 它们独立于观测位置。在平稳过程中, 如果在不同位置上观测数据之间的协方差仅仅

依赖于距离而与方向无关,则这样的空间过程被称为是各向同性的。非平稳过程数据几乎不可能被模拟,因为几乎所有位置上都需要不同的参数集。因此大部分的建模步骤是首先在均值基础上识别趋势,然后按照平稳过程对趋势的偏差建模。

影响空间数据分析的主要因素是数据赖以分析的地理尺度。在局部层次上识别特定的非随机模式是可能的,而当从国家层次上观察时则转为随机变化。另外的问题是很多空间数据集是基于不规则形状单元的,或可能存在方向效应。接近或临近性同样比时间序列分析更加难以清晰地定义。任何类型的空间数据分析都在某种程度上受制于边界效应,即地图边界上的面状单元只是在一个方向上存在邻居。很多数据分析不得不依赖于根据某一特定空间聚集水平来概括数据信息,如果使用的是相同层次的聚集,根据这种分析做出的推断只在此层次上正确。这种情况被称为面积单元问题。

本书将对空间数据的性质等进行探讨,并详细地介绍描述空间相关性的分析方法和统计量的计算,并对产生 MAUP 问题的尺度效应和划区效应进行详细的介绍。

1.2.4 空间数据分析的点模式方法

探讨空间数据分析的方法首先需要对空间数据的类型进行区分。一般分为四种主要类型,分别是:点数据、线数据、面数据、空间连续数据。为了分析的需要,研究者从连续和离散两个方面看待地理现象的分布。而上述数据的分类方法都能够对连续的或离散的世界进行完整地表达。我们首先分析点模式研究的内容。

空间点模式是根据事件的空间坐标的分析技术(例如爆发疾病的位置),事件也可能包含属性信息(例如爆发的时间)。数据的点模式可以是基于所有点事件的完全地图,或者是样本点分布模式。空间点模式研究的重点在于探测点事件的分布是随机的,或是聚集,或是均匀分布的模式。识别关于事件发生的位置相关的随机过程非常重要。空间点模式可根据过程的密度定量地描述,可使用单位面积上平均的事件数量这个一阶性质测度。二阶性质或空间依赖性需要根据点之间或区域之间的关系进行分析。后者通常被解释为聚类分析。

在 GIS 中空间点模式最常用的表示方法是点状地图、密度图等方式。一般根据这样的地图可视地检测模式的随机性是困难的。需要定量的分析方法计算有关的统计量来研究分布模式,一般使用“一阶”或“二阶”效应的空间分析方法。其中点模式的“一阶”效应可通过样方计数和核密度方法计算。样方计数法是首先将区域划分为面积相等的子区域样方,然后根据每一个样方中的事件数量计算概括统计量,这些技术给出的是空间基本过程的密度变化。这一方法的缺点

是将信息聚集到面状数据中，这将引起信息的损失。而核密度估计是这样一种技术，它使用原始的点位置产生光滑的密度图。而点模式的二阶性质是基于点之间距离测度的研究方法，一般使用的是最近邻距离。最近邻距离估计有两种技术，即随机选择的事件及其最近邻之间的距离，或随机选择空间上的位置与最近邻的事件之间的距离。

目前已经有很多空间点模式的分析方法，基于样方计数技术的用离散指数法 (index of dispersion) 检验，而那些使用最近邻距离的用 Clark-Evans 检验法和 K -函数法等。但是各种方法对于分布模式的结论依赖于和完全随机模式 (CSR) 的比较。一个随机空间过程产生的点模式，应当遵守同质 poisson 过程 (homogenous poisson process)。这意味着研究区域中的每一个事件是以等概率发生在区域的任意位置上，并且其发生独立于空间任意位置和其他的事件。因此完全随机过程不存在“一阶”或“二阶”效应。另外可以使用的模型是 A. Getis 和 J. K. Ord (1992) 的 G 方法， G 方法构造的是基于距离的统计量，也可用于评价点模式和面模式的空间自相关问题，还可用于检验用全局方法不能检验的空间依赖的局部热点区域 (hot spot)。

1.2.5 面数据的空间分析方法与空间回归模型

根据 Reiply 的观点，面积单元可构成规则的格子或栅格，也可构成不规则的单元。对于不规则的单元，我们一般称之为多边形；规则的单元称之为格子 (lattice)。无论是规则的还是不规则的面积单元，通常由属性数据描述其特征，一般称其为面数据或格子数据 (lattice data)。由于数值是对一个区域的表示，所以一般不需要对数据进行估计。在面数据中主要强调的是空间模式和趋势的探测与解释，甚至扩展到考虑协变的情况。

在面数据的分析中首先需要研究的是它的表达方法，一般通过设色地图 (choropleth) 进行可视化。正确使用类别间隔和颜色表达设色地图中的数值是非常重要的。统计地图和密度图可用于表达特定区域的重要性，但是必须了解由 MAUP 所产生的问题。同时对比和显示多种数据是可能的，这可通过在设色地图上增加规则的柱状图或符号来完成。

对于面数据的空间分析，需要研究面数据的许多性质，例如，多边形的几何性质和相互之间的接近性 (proximity) 的测度方法，空间自相关性的测度方法等。在建模分析中，一般首先需要研究的是接近性特征，这一信息一般通过产生的接近性矩阵或空间加权矩阵来描述。本书将详细介绍空间加权矩阵的构造方法。

对于面数据同样需要研究其“一阶”效应和“二阶”效应。描述“一阶”效应的方法有空间滑动平均核估计方法等。“二阶效应”主要用于探究属性值对于

均值偏差的空间分布是否存在空间依赖,这就是空间自相关 (spatial auto correlation),它定量描述了不同位置上同一属性值的相关程度。其中最常用的方法是 Moran' I 和 Geary' C 两个指标,第一个指标与连续数据分析中的协变异图密切相关,而第二个方法则与连续数据中的变异图密切相关。Moran' I 等统计量具有强大的分析能力。此外还可以使用相关图以图形化方式实现不同空间间隔 (spatial lag) 上数值之间的相关性。这种空间分析中的相关图对于模式的描述相似于时间序列分析中的相关谱。但是上述方法并没有提供可用于辨识热点区域 (hot-spot) 空间联系的局部化指标。Anselin 于 1994 年描述的 Moran 散点图和空间间隔饼图能以可视方式刻画变异的局部模式,对其定量估计可以使用 Getis 和 Ord 的 G 统计量,或 Anselin (1995) 的空间联系局部指标 (LISA)。与 G 统计量相似,后者是可用于非平稳性的局部 hot-spot 指示器,并且可用来评估全局统计中个别数据的影响,并辨识离群点 (outlier)。

在面数据空间相关性分析的基础上,可深入分析空间背景上变量之间的回归关系,一般通过两种途径建模,一种是空间回归模型,另外一种是地理加权回归模型。这两种模型在数学形式上不同,但是都考虑了空间位置要素,也就是空间相关性对回归模型建立的重要影响,而且模型揭示的信息量比传统的统计分析技术更为广泛。

1.2.6 空间连续数据的分析技术

在空间连续数据以及面状数据的分析中,分析的重点转移到使用属性信息来描述空间模式上。空间连续数据还经常被称为地统计数据。数据通常是对空间上固定点的采样数据。这一分析的主要目标是使用样本采样点上收集的数据描述属性值的空间变化。空间变化可被模拟为一阶和二阶的空间过程。

空间连续数据“一阶”效应的分析方法描述的是要素分布的全局趋势,主要有空间滑动平均、镶嵌方法、核估计方法和趋势面分析。空间滑动平均方法是在给定的近邻样采样点之间内插数值,例如倒数距离加权法 (IDW),通过引入距离加权机制来说明采样点之间的距离的变化对于插值点数值的贡献。而另外可以使用基本样本点的镶嵌方法,这类方法中最常用的是 Delauney 三角形,又被称为不规则三角网 (TIN)。根据 Delauney 三角形可以得到 Dirichlet 镶嵌,或 Voronoi 多边形。这样的 TIN 可用于构造等高线地图或数值地形模型 (DTM)。类似于点模式,同样能用核估计技术将采样点的属性信息转换成表面。对于“一阶”过程可使用趋势面分析,建立普通的多项式最小二乘回归。对结果的处理必须格外注意,因为标准的回归所假设的独立随机误差和异方差性似乎被违反了。大部分的趋势面分析模型主要描述的是整体的趋势,对于局部预测是无价值的。