

21

普通高等教育“十二五”规划教材

21 世纪全国高校应用人才培养信息技术类规划教材



空间数据分析

毕硕本 编 著



北京大学出版社
PEKING UNIVERSITY PRESS

普通高等教育“十二五”规划教材

21 世纪全国高校应用人才培养信息技术类规划教材

空间数据分析

毕硕本 编著



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

空间数据分析/毕硕本编著. —北京: 北京大学出版社, 2015. 8

(21世纪全国高校应用人才培养信息技术类规划教材)

ISBN 978-7-301-25553-7

I. ①空… II. ①毕… III. ①地理信息系统—高等学校—教材 IV. ①P208

中国版本图书馆CIP数据核字(2015)第033864号



书 名	空间数据分析
著作责任者	毕硕本 编著
责任编辑	温丹丹
标准书号	ISBN 978-7-301-25553-7
出版发行	北京大学出版社
地 址	北京市海淀区成府路205号 100871
网 址	http://www.pup.cn 新浪微博: @北京大学出版社
电子信箱	zyjy@pup.cn
电 话	邮购部 62752015 发行部 62750672 编辑部 62765126
印 刷 者	三河市北燕印装有限公司
经 销 者	新华书店
	787毫米×1092毫米 16开本 18.5印张 460千字
	2015年8月第1版 2015年8月第1次印刷
定 价	38.00元

未经许可, 不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有, 侵权必究

举报电话: 010-62752024 电子信箱: fd@pup.pku.edu.cn

图书如有印装质量问题, 请与出版部联系, 电话: 010-62756370

前 言

本书根据高等院校培养综合型交叉学科本科人才的发展目标编写，介绍空间数据分析的原理与方法。全书共分 17 章，主要内容包括绪论、空间数据的性质、空间数据的完备化、空间数据的标准化、探索性空间数据分析、空间点模式分析、地统计数据插值、格数据统计、格数据回归分析、空间回归分析、面状数据空间模式分析、空间连续数据分析方法、非参数统计、空间抽样方法、空间度量算法、空间分析算法和空间统计分析算法。

本书内容丰富、结构合理、针对性强，理论叙述严谨、能力培养目标明确。读者学完本门课程后，能够掌握空间数据分析的基本知识、基本原理与方法。

本书可作为地理信息系统、计算机科学与技术、遥感科学与技术、测绘工程、地理科学等相关专业学生的教科书，同时也适合于从事空间数据分析、地理信息系统应用的人员参考。

本书在编写过程中得到了单位领导和同人的热情帮助和支持，在此表示衷心的感谢！本书的编写参考了部分同行专家的著作和成果，在此对他们表示衷心的感谢！

由于时间仓促，加之作者的水平有限，书中难免有遗漏和不足之处，恳请同行专家和广大读者批评指正。

编 者

2015 年 7 月

本教材配有教学课件，读者如有需要，请加 QQ 群 (279806670) 或发电子邮件至 zyjj@pup.cn 索取，也可致电北京大学出版社 (010-62765126)。

目 录

第1章 绪论	1
导读	1
1.1 空间分析的概念与研究发展	1
1.2 空间数据分析的研究内容	4
1.3 小结	8
思考及练习题	8
参考文献	9
第2章 空间数据的性质	10
导读	10
2.1 地理世界的概念模型与数据模型	10
2.2 空间数据的性质	16
2.3 空间数据的不确定性	19
2.4 小结	22
思考及练习题	23
第3章 空间数据的完备化	24
导读	24
3.1 空间数据的缺值处理	24
3.2 空间数据的插值	28
3.3 算例	39
3.4 小结	40
思考及练习题	40
参考文献	41
第4章 空间数据的标准化	42
导读	42
4.1 问题的提出	42
4.2 基本原理	42
4.3 方法	45
4.4 算例	50
4.5 小结	52
思考及练习题	53
参考文献	53

第5章 探索性空间分析	55
导读	55
5.1 线性相关性分析	55
5.2 回归分析	55
5.3 主成分分析	56
5.4 层次分析	56
5.5 地理探测器	58
5.6 空间聚集探测检验与地理探测器的比较	60
5.7 小结	60
思考及练习题	60
参考文献	61
第6章 空间点模式分析	62
导读	62
6.1 空间点模式的概念与空间分析技术	62
6.2 基于密度的方法——样方计数法与核函数法	64
6.3 最近邻距离法	73
6.4 G 函数与 F 函数	77
6.5 K 函数与 L 函数	84
6.6 K 函数方法的扩展——二元模式与空间-时间模式	91
6.7 小结	94
思考及练习题	95
参考文献	96
第7章 地统计数据插值	97
导读	97
7.1 空间数据插值	97
7.2 空间数据插值原理	97
7.3 计算公式	100
7.4 趋势面方法	108
7.5 反距离加权法 (IDW)	109
7.6 核心估计函数法	109
7.7 小结	110
思考及练习题	111
参考文献	111
第8章 格数据统计	112
导读	112
8.1 概述	112
8.2 格数据统计原理	113
8.3 格数据的空间相关性计算方法	115

8.4 格数据相关性分析实例	120
8.5 可变面元问题	123
8.6 空间热点探测方法	124
8.7 小结	126
思考及练习题	126
参考文献	127
第9章 格数据回归分析	129
导读	129
9.1 概述	129
9.2 地理加权回归	129
9.3 二阶空间回归模型	131
9.4 空间回归方程的通用模型	132
9.5 二阶空间回归模型似然函数求解	133
9.6 格数据回归计算实例	134
9.7 小结	135
思考及练习题	135
参考文献	136
第10章 空间回归分析	137
导读	137
10.1 引言	137
10.2 回归分析方法	137
10.3 空间自回归模型	138
10.4 空间回归模型实例	145
10.5 地理加权回归模型	148
10.6 小结	155
思考及练习题	156
第11章 面状数据空间模式分析	157
导读	157
11.1 引言	157
11.2 空间接近性与空间权重矩阵	157
11.3 面状数据的趋势分析	163
11.4 空间自相关的概念	167
11.5 名义变量的空间自相关测度——连接计数法	168
11.6 局部空间自相关统计量	172
11.7 小结	178
思考及练习题	178
参考文献	179

第 12 章 空间连续数据分析方法	180
导读	180
12.1 引言	180
12.2 探索性分析方法	180
12.3 趋势面分析	184
12.4 连续数据的空间依赖性测度——协方差图和半方差图	185
12.5 小结	192
思考及练习题	193
第 13 章 非参数统计	194
导读	194
13.1 非参数统计的概念	194
13.2 非参数统计的基本统计量	195
13.3 不同个数样本的情形及其检验	197
13.4 非参数统计实例：洪水频率分析	205
13.5 小结	207
思考及练习题	207
参考文献	207
第 14 章 空间抽样方法	209
导读	209
14.1 问题的提出	209
14.2 经典抽样方法	209
14.3 空间抽样方法	215
14.4 应用实例	224
14.5 小结	229
思考及练习题	229
参考文献	230
第 15 章 空间度量算法	231
导读	231
15.1 空间距离与方向度量算法	231
15.2 面积度量算法	235
15.3 体积度量算法	239
15.4 坡度、坡向度量算法	240
15.5 小结	242
思考及练习题	243
第 16 章 空间分析算法	244
导读	244
16.1 引言	244
16.2 路径分析算法	244

16.3 资源分配分析算法	251
16.4 缓冲区分析算法	254
16.5 叠置分析算法	264
16.6 小结	270
思考及练习题	271
参考文献	272
第17章 空间统计分析算法	274
导读	274
17.1 引言	274
17.2 多变量统计分析算法	274
17.3 空间分类统计算法	279
17.4 层次分析算法	282
17.5 小结	285
思考及练习题	286
参考文献	286

第 1 章 绪 论

导 读

本章通过引入空间分析概念的几种定义,介绍空间数据分析的概念,并简述空间分析与空间数据分析的异同.通过简述空间分析的发展情况,可以了解到空间数据分析的发展.空间数据分析的内容很多,本章仅概述了空间数据模型及其表示、空间数据的性质、探索性空间数据分析、空间数据的点模式分析、面数据的空间分析、空间连续数据的分析六部分内容.常用的空间数据分析方法将在后续的章节中展开介绍.本章涉及的基本概念较多,需要仔细研读、认真体会.

1.1 空间分析的概念与研究发展

1.1.1 空间分析的概念

虽然地理学家用空间分析 (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、空间数据分析、空间统计分析、空间建模.

1. 空间数据操作

主要出现在 GIS 中,但通常被称为空间分析.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4个方面: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,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 空间数据分析与空间分析的异同

空间分析是一个比较广义的概念,GIS中的空间分析是一个狭义的概念,主要指GIS中的空间数据操作,而本书中的空间数据分析是包含Unwin所讲的空间数据分析,以及空间统计分析两部分的内容。空间建模则由于应用领域的不同,具有不同的应用模型,不适合作为通用性、一般性的空间数据分析的内容。

1.1.3 空间分析研究的发展

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,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，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，揭示了空间加权矩阵的本质，提出了两个主要的自回归统计量 Moran' I 和 Geary' 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。Open Shaw 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，以描述全局特征为传统的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。Anselin 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或空间自回归模型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4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 空间数据分析的研究内容

空间数据分析主要研究内容包括:空间数据模型及其表示、空间数据的性质、探索性空间数据分析、空间数据的点模式分析、面数据的空间分析、空间连续数据的分析,以及空间数据的非参数统计、空间抽样、空间分析与空间统计分析等方法与算法.

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.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分析范畴.

1.2.2 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,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，如回归技术描述“一阶”效应。而“二阶”效应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，而当从更高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。

1.2.3 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。利用画图和地图的方式进行信息的视觉显示是研究人员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應用。数据探索性分析的目的是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, 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.

1.2.4 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.

空间点模式是根据事件的空间坐标的分析技术 (如暴发疾病的位置), 事件也可能包含属性信息 (如暴发的时间). 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对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,一般称为多边形;规则的单元称为格子(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。由于数值是对一个区域的表示,因此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非常重要。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。

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,而第二个指标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.

“一阶”方法忽略了局部性和空间相关性等因素, 在很多应用中会带来偏差, 而且其偏差难以估计. 需要“二阶”效应的分析方法, 这一方法的典型是克里格 (Kriging) 方法或地统计方法. 在这类方法中, 用协方差函数或方差图描述采样点上得到的属性值之间的空间依赖性. “二阶”效应的存在将导致短距离间隔上的观测值之间的正协方差和长距离间隔上的低协方差或相关. 方差图描述的是一个协方差随着样本距离的变化而变化的函数, 同样, 相关图也是相关性随着样本点距离的变化而变化的函数. 半方差图是采样点之间距离和方向变化时的方差的图形表达. 对于平稳的空间过程, 这三种方法描述的是相似的信息. 在空间过程中, 半方差图的估计被认为是对于平稳性的更为稳健的估计, 表示为偏离一般趋势. 当“一阶”效应微弱、“二阶”效应强烈存在时更为合适的方法是使用模型拟合方差图. 这些模型被定义为“用眼睛看”的模型, 但是方差图本身不能用于预测值, 可以通过克里格方法获得这一功能. 这是一种加权平均的技术, 利用方差图描述的空间依赖性估计空间分布变量的值.

1.3 小 结

在地理学研究、地理信息科学发展及 GIS 的实际应用中, 空间数据分析都占有至关重要的地位. 本书力图从广度与深度两个方面介绍空间分析与空间数据分析的概念, 以及空间分析的发展、空间数据分析的主要内容. 空间数据分析主要研究内容包括: 空间数据模型及其表示, 空间数据的性质, 探索性空间数据分析, 空间数据的点模式分析, 面数据的空间分析, 空间连续数据的分析; 以及空间数据的非参数统计、空间抽样、空间分析与空间统计分析等方法与算法. 常用的空间数据分析方法将在后续章节中详细介绍.

思考及练习题

1. 什么是空间分析? 空间分析概念有几种类型?
2. 空间数据分析和空间分析有何异同?
3. 简述空间分析研究的发展.
4. 空间数据分析包括哪几部分内容?
5. 名词解释

(1) “一阶”效应; (2) “二阶”效应; (3) 各向同性; (4) 边界效应; (5) 面积单元问题; (6) 平稳过程.

6. 名词解释

(1) 空间点模式; (2) 样方计数法; (3) 核密度估计; (4) 最近邻距离; (5) 同质 Poisson 过程; (6) G 方法.