

“十三五”国家重点出版物出版规划项目

经济科学译丛

空间数据分析

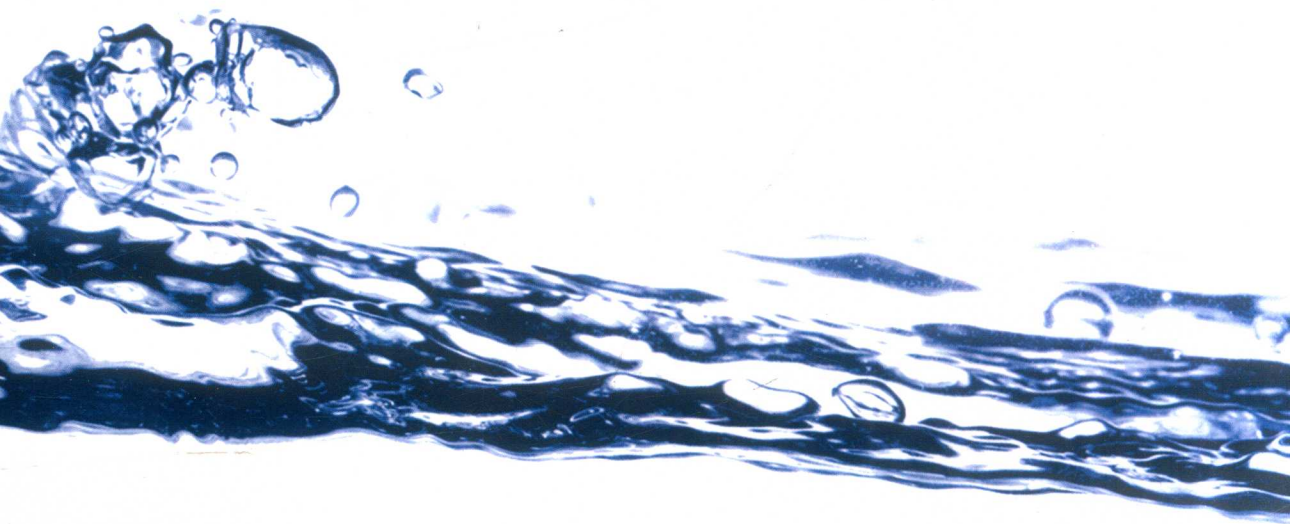
模型、方法与技术

曼弗雷德·M. 费希尔 (Manfred M. Fischer)

著

王劲峰 (Jinfeng Wang)

Spatial Data Analysis
Models, Methods and Techniques



“十三五”国家重点出版物出版规划项目

经济科学译丛

空间数据分析

模型、方法与技术

曼弗雷德·M. 费希尔 (Manfred M. Fischer) 著

王劲峰 (Jinfeng Wang)

张璐 肖光恩 吕博才 译

肖光恩 总校译

Spatial Data Analysis
Models, Methods and Techniques



中国人民大学出版社
· 北京 ·

图书在版编目 (CIP) 数据

空间数据分析：模型、方法与技术 / () 曼弗雷德·M. 弗希尔，王劲峰著；张璐，肖光恩，吕博才译. —北京：中国人民大学出版社，2018. 10
(经济科学译丛)
ISBN 978-7-300-25304-6

I. ①空… · II. ①曼… ②王…③张…④肖…⑤吕… 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 (2017) 第 311960 号

“十三五”国家重点出版物出版规划项目
经济科学译丛

空间数据分析：模型、方法与技术

曼弗雷德·M. 费希尔 王劲峰 著
张璐 肖光恩 吕博才 译
肖光恩 总校译
Kongjian Shuju Fenxi: Moxing Fangfa yu Jishu

出版发行	中国人民大学出版社		
社 址	北京中关村大街 31 号	邮政编码	100080
电 话	010-62511242 (总编室)		010-62511770 (质管部)
	010-82501766 (邮购部)		010-62514148 (门市部)
	010-62515195 (发行公司)		010-62515275 (盗版举报)
网 址	http://www.crup.com.cn		
	http://www.ttrnet.com(人大教研网)		
经 销	新华书店		
印 刷	北京东君印刷有限公司		
规 格	185 mm×260 mm 16 开本	版 次	2018 年 10 月第 1 版
印 张	6.25 插页 2	印 次	2018 年 10 月第 1 次印刷
字 数	79 000	定 价	36.00 元

版权所有 侵权必究 印装差错 负责调换

《经济科学译丛》

编辑委员会

学术顾问

高鸿业 王传纶 胡代光 范家骧 朱绍文 吴易风

主 编

陈岱孙

副主编

梁 晶 海 闻

编 委 (按姓氏笔画排序)

王一江	王利民	王逸舟	贝多广	平新乔	白重恩
刘 伟	朱 玲	许成钢	张宇燕	张维迎	李 扬
李晓西	李稻葵	杨小凯	汪丁丁	易 纲	林毅夫
金 碚	姚开建	徐 宽	钱颖一	高培勇	梁小民
盛 洪	樊 纲				



《经济科学译丛》总序

中国是一个文明古国，有着几千年的辉煌历史。近百年来，中国由盛而衰，一度成为世界上最贫穷、落后的国家之一。1949年中国共产党领导的革命，把中国从饥饿、贫困、被欺侮、被奴役的境地中解放出来。1978年以来的改革开放，使中国真正走上了通向繁荣富强的道路。

中国改革开放的目标是建立一个有效的社会主义市场经济体制，加速发展经济，提高人民生活水平。但是，要完成这一历史使命绝非易事，我们不仅需要从自己的实践中总结教训，也要从别人的实践中获取经验，还要用理论来指导我们的改革。市场经济虽然对我们这个共和国来说是全新的，但市场经济的运行在发达国家已有几百年的历史，市场经济的理论亦在不断发展完善，并形成了一个现代经济学理论体系。虽然许多经济学名著出自西方学者之手，研究的是西方国家的经济问题，但他们归纳出来的许多经济学理论反映的是人类社会的普遍行为，这些理论是全人类的共同财富。要想迅速稳定地改革和发

展我国的经济，我们必须学习和借鉴世界各国包括西方国家在内的先进经济学的理论与知识。

本着这一目的，我们组织翻译了这套经济学教科书系列。这套译丛的特点是：第一，全面系统。除了经济学、宏观经济学、微观经济学等基本原理之外，这套译丛还包括了产业组织理论、国际经济学、发展经济学、货币金融学、财政学、劳动经济学、计量经济学等重要领域。第二，简明通俗。与经济学的经典名著不同，这套丛书都是国外大学通用的经济学教科书，大部分都已发行了几版或十几版。作者尽可能地用简明通俗的语言来阐述深奥的经济学原理，并附有案例与习题，对于初学者来说，更容易理解与掌握。

经济学是一门社会科学，许多基本原理的应用受各种不同的社会、政治或经济体制的影响，许多经济学理论是建立在一定的假设条件上的，假设条件不同，结论也就不一定成立。因此，正确理解掌握经济分析的方法而不是生搬硬套某些不同条件下产生的结论，才是我们学习当代经济学的正确方法。

本套译丛于 1995 年春由中国人民大学出版社发起筹备并成立了由许多经济学专家学者组织的编辑委员会。中国留美经济学会的许多学者参与了原著的推荐工作。中国人民大学出版社向所有原著的出版社购买了翻译版权。北京大学、中国人民大学、复旦大学以及中国社会科学院的许多专家教授参与了翻译工作。前任策划编辑梁晶女士为本套译丛的出版做出了重要贡献，在此表示衷心的感谢。在中国经济体制转轨的历史时期，我们把这套译丛献给读者，希望为中国经济的深入改革与发展做出贡献。

《经济科学译丛》编辑委员会



译者序言

这是一本介绍空间数据分析方法的初级读物，它对普及读者的空间计量分析知识具有重要的指导作用。

空间计量经济学分析源于对地理区位作用的可视化分析或量化研究，更深化于对空间互动效应或空间溢出效应的识别、测度与建模技术的不断探索。已有的空间计量分析专著重在研究空间数据的建模及其估计方法，对空间交互数据建模技术的探讨仍不全面和深入。本书的重要贡献主要有：一是作者从地理信息分析系统的视角介绍了空间数据的可视化、空间自相关的测度方法以及区域数据的建模方法；二是作者重点研究了空间交互数据的建模与估计方法，主要是空间交互数据模型的不同类别及其估计策略，特别是对泊松空间交互模型的一般化以及独立空间交互模型相关问题的研究。因此，对读者深入地学习和领悟空间交互数据的建模方法和估计策略具有重要的指导意义和画龙点睛的作用。

本书是我带领研究生团队翻译和编写的系列空间计量分析著作或

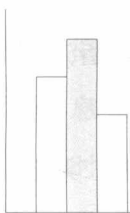
教材的第六本。本书初译由吕博才完成，后由张璐进行全面翻译，最后由肖光恩进行通译和总校译。本书的出版得到了“武汉大学马克思主义团队建设项目”资金和“武汉大学首批来华留学英语授课品牌课程《国际商务》项目”资金的资助。本书的出版也得到了中国人民大学出版社王晗霞编辑的全力支持，在此深表感谢。

由于水平有限，错误与疏漏之处在所难免。敬请读者批评指正，以便再版时修正。欢迎把修改意见发送至邮箱 xiaoguang@whu.edu.cn。

肖光恩

于武汉大学经济与管理学院

2018年8月



前 言

空间和区位一直在地理和区域科学中处于中心地位。但近年来在主流社会科学中，对空间维度的关注持续增加，同时在自然科学（例如生态学）中这种关注也在增加。在实证研究中，越来越多的社会科学家已经运用了新的方法和技术（例如地理信息系统、全球定位系统、遥感、空间统计学和空间计量经济学）。此外，在理论研究中，人们也更加关注区位和空间交互作用。

广义上讲，空间分析就是对位于地理空间中的空间现象的定量分析（Bailey and Gatrell, 1995）。这一研究领域过于广泛，因此想要在一本书中完全涵盖相关内容的愿望是不能实现的 [见 Fischer 和 Getis (2010) 对此领域多样性的论述]。因此，我们决定将主要研究范围限定于空间分析的一个重要子集——空间数据分析。这样做的理由是：我们主要关注地理空间分析过程中出现的数据，同时考虑运用一些模型、方法以及技术来描述和解释这一分析过程中出现的行为，以及这些行为与其他空间现象之间的关系。通过这种方式来定义空间数据分

析，我们将本书的内容限定在空间数据的统计描述与建模中，重点是一系列特定的方法。因此，本书不考虑一些重要的定量分析方法，例如，网络分析法和区位配置分析法，尽管这两种分析法通常是空间分析的重要内容。

无论空间数据分析是否为一个独立的学术领域，事实上，在过去的20年中，出于对空间数据的研究兴趣以及理解空间数据的需要，空间数据分析显得越发重要。空间数据指的是将观测对象与空间参照物联系起来的数据。空间参照物可能是具体的，如邮寄地址或网格参照物；也可能是不具体的，如遥感图像中的像素。

在过去的几十年中，在这个主题上产生了许多优秀的文章（Cliff and Ord, 1981; Upton and Fingleton, 1985; Anselin, 1988b; Griffith, 1988; Ripley, 1988; Cressie, 1993; Haining, 1990, 2003; Bailey and Gatrell, 1995; LeSage and Pace, 2009）。这些大多数是写给研究人员的，而本书主要是从“数据驱动”而非“理论驱动”的角度向研究生介绍空间数据分析。根据这个总体目标，我们并没有试图详尽地阐述整个空间数据分析，而是将讨论的范围限制于对两类主要空间数据的分析：一类是区域数据，即与一组固定的区域或涵盖研究范围的区域相关联的数据；另一类是空间交互数据（或从起源地到目的地的流动数据，亦翻译为“起点—终点流数据”，本书此后不再做区分），即对地理空间中代表点或区域的每一对成对数据或连接数据的测量。

本书仅讨论空间分析模型、方法和技术的一部分内容，它们对分析空间数据是非常有用的，且容易掌握。本书讨论的主题包括：一是非正式或探索性的方法和技术，二是正式的统计建模、参数估计和假设检验。

本书分为两部分，每一部分都尽可能独立。第一部分主要考虑区域数据的分析，这些区域可以是规则方格，如遥感图像，或者是一组

不规则形状的区域，如行政区划。第二部分重点分析空间交互数据，这些数据与成对的点或区域相关，被称为从起源地到目的地的流动数据或空间交互数据，这些数据与运输规划、人口迁移、通勤、购物行为、物流以及信息和知识的传播等方面的研究息息相关。

我们不考虑时空数据，但我们假设数据是纯空间性的，要么随着时间的推移而聚合，要么参考时间上的固定点。空间数据的测量、存储和检索都很重要，但不在本书的研究范围之内。GIS 提供的软件工具可以将空间数据和非空间数据、定性数据和定量数据整合到可在一个系统环境下管理的数据库中 (Longley et al., 2001)。我们假设读者对统计学和数学的背景知识有所了解，以便将本书限定在一个可控的篇幅之内。

我们感谢维也纳经济大学的经济地理与地理信息科学研究所的慷慨支持。还要感谢 Thomas Seyffertitz (经济地理与地理信息科学研究所) 提供的技术援助，他在文字处理系统、格式、索引方面的技术非常专业，对本书的细节也很关注。最后，我们要感谢本系列书的编辑 Henk Folmer 博士，他对本书手稿提出了宝贵的意见。

曼弗雷德·M. 费希尔

于维也纳

王劲峰

于北京

2011 年 5 月



目 录

第一章	引 言	1
1.1	数据和空间数据分析	2
1.2	空间数据类型	4
1.3	空间数据矩阵	5
1.4	空间自相关	8
1.5	空间数据的严格性	11
第一部分	区域数据的分析	13
第二章	区域数据的探索性分析	15
2.1	制图与地理可视化	16
2.2	空间权重矩阵	20
2.3	空间自相关的全局测量和检验	23
2.4	空间自相关的局部测量和检验	28
第三章	区域数据的建模	33
3.1	空间回归模型	35

第二部分

3.2 空间相关性检验	38
3.3 空间德宾模型	40
3.4 空间回归模型的估计	42
3.5 模型参数解释	45
空间交互数据的分析	49
第四章 空间交互数据的模型和方法	51
4.1 空间交互数据的可视化和探索性分析	52
4.2 广义空间交互模型	53
4.3 普通最小二乘回归模型的设定与方法	55
4.4 广义泊松空间交互模型	58
4.5 泊松空间交互模型的最大似然估计	60
4.6 泊松空间交互模型的一般化	62
第五章 空间交互模型和空间相关	65
5.1 独立（对数正态）空间交互模型的矩阵表示	67
5.2 独立空间交互模型的计量经济学扩展	69
5.3 空间交互模型的空间滤波	73
参考文献	76



摘要：在本章中，我们介绍空间数据分析，并且将空间数据分析与其他形式的的分析区分开来。空间数据，我们定义为包含位置和属性信息的数据。我们专注于研究两种类型的空间数据：区域数据和从起源地到目的地的流动数据。区域数据与一种情形有关，即在这种情形中所研究的变量——至少就本书而言——是不连续变化的，但是仅仅在一组固定的区域内或覆盖研究领域的区域中有观测值。这些固定的点可以由规则方格构成（例如遥感图中的像素），或者由不规则区域单元构成（例如人口普查区）。从起源地到目的地的流动数据（也称为空间交互数据，也称为“起点—终点数据”，本书不再做区分）与地理空间中的成对点或成对区域相关。这些数据代表人口、商品、资本、信息或知识从一组起点流动到一组终点，也与运输规划、人口迁移、通勤、购物行为、物流、信息和知识的传播等方面的研究息息相关。我们考虑数据中的空间自相关问题，传统的统计分析不可靠，于是要求我们使用空间分析工具。这个问题是指观测值在空间上是非独立的。我们最后总结一下空间分析研究人员经常面临的一些实际问题。

关键词：空间数据；空间数据的类型；空间数据矩阵；区域数据；从起源地到目的地的流动数据；空间自相关；空间数据的严格性

1.1 数据和空间数据分析

与信息不同，数据是由数字或者在某种意义上中性的符号构成的，与上下文无关。原始的地理事实，如某一特定时间和地点的温度，就是一些数据的例子。根据 Longley et al. (2001, p. 64)，我们可以认为空间数据由原子元素和有关地理世界的事实构成。简单来说，通常将空间数据的一个原子元素（严格来说，数据）和地理位置（地点）相关联，这些元素通常是时间和一些描述性属性或实体属性的数据。例如，“2010年12月24日下午2时，北纬 $48^{\circ}15'$ 、东经 $16^{\circ}21'28s$ 的气温为 $6.7^{\circ}C$ ”。它将位置和时间与大气温度的属性相关联。因此，我们可以说空间（地理）数据将地点（位置）、时间和属性（这里是温度）相关联。

属性有多种形式。有些是物理或环境性质的，而其他则是社会或经济属性。一些属性可以简单地标识一个位置，例如用于记录土地所有权的邮政地址或包裹标识符。其他一些属性是在某个地点的某种测量（例如大气温度和收入），而另外一些属性则用于分类，例如土地利用类别分为农业用地、住宅用地和工业用地。

虽然时间在空间数据分析中是非强制性的，但地理位置是必不可少的，并且地理位置将空间数据分析与其他形式的被认为是无空间或非空间的数据分析区分开来。即使观测单位本身的定义是空间的，假如我们忽略样本位置间的空间关系，单独处理这些属性，我们也不能称这是空间数据分析。即使属性数据是至关重要的，若脱离了它们的空间背景，也会失去价值和意义（Bailey and Gatrell, 1995, p. 20）。无论这些属性是如何测量的，为了进行空间数据分析，我们至少同时需要位置和属性这两个信息。

空间数据分析需要一个基础的空间研究框架，以此定位正在研究的空间现象。Longley et al. (2001) 和其他学者对度量地理的两种基本方式进行了区分：对空间现象是离散或是连续的两种看法。换句话说，区别就在于，前者认为空间充满着“离散对象”，后者则认为空间实质上被“连续曲面”覆盖。前者被认为是对象或实体空间观，后者则被认为是场空间观。

对象空间观认为，被分析的空间现象的种类由它们的维度来识别，占据区域的对象是二维的，通常被称为区域；某些物体更像一维线，包括河流、铁路或道路，并被表示为一维物体，一般被称为线条；其他物体更像是一些零维度的点，如植物、人类、建筑物、地震的震中，等等 (Longley et al., 2001, pp. 67 - 68; Haining, 2003, pp. 44 - 46)。注意，本书不考虑面对象或体对象，即具有长度、宽度和深度三维的物体，它们用于表示自然对象（如流域）或人工现象（如购物中心的人口潜力）。

当然，这样做是否恰当取决于研究的空间尺度（即寻求测量“现实”细节的水平）。如果我们正在全国范围内考察城市居民点的分布，将它们视为点的分布是合理的，但在较小区域的尺度下，它就变得不太敏感了，正如上面提到的把道路看作线的现象。但尺度相关性仍然存在。在较大的比例尺的城市地图上的道路很宽，举例来说，如果你要关注汽车导航问题，这一点就显得尤其重要。线也标记了区域的边界。通过区域我们能更广泛地理解那些在行政上或法律上被定义的实体（例如国家、地区、人口普查区域等），以及“自然区域”（例如地图上的土壤或植被区）。

场空间观认为，重点是空间现象的连续性，地理世界是由有限数量的变量描述的，每个变量都可以在地球表面的任何点测量，并且在表面上数值有所变化 (Haining, 2003, pp. 44 - 45)。如果我们想到自然环境中的现象，例如温度、土壤特征等，那么这些变量可以在地球表面的任何地方观察到 (Longley et al., 2001, pp. 68 - 71)。当然，在实践中这些变量（例如温度）是离散的，在众多位置进行取样并表示为线的集合（所谓的等温线）。土壤特征也可以在一组

离散位置进行采样并表示为连续变化的区域。在所有这些情况下，就是试图用离散抽样来表示潜在的连续性（Bailey and Gatrell, 1995, p. 19）。

1.2 空间数据类型

在描述空间数据的性质时，区分测量变量所处空间的离散性或连续性以及变量（测量值）本身的离散性或连续性是很重要的。如果空间是连续的（场空间观），变量值必须是连续的值，因为连续场不能被保存在离散值的变量中。如果空间是离散的（对象空间观），或者一个连续空间被离散化，则既可以是连续值又可以是离散值（名义值或序数值）（Haining, 2003, p. 57）。

根据两种类型的空间观和测量水平对空间数据进行分类来确定合适的统计方法是用于回答研究问题的必不可少的首要步骤。但仅有分类是不够的，因为相同的空间对象可能代表完全不同的地理空间。例如，点（所谓的质心）也用于表示区域。表 1.1 提供了一个四种空间数据类型的分类。

（i）点式数据：由一些研究区域中的一系列点位置组成的数据集，在这些位置上发生了感兴趣的事件（一般意义上的），例如疾病或某类犯罪的发生。

（ii）场数据（也称为空间连续数据、地理统计数据）：与概念上的连续变量（场空间观）相关，并且观测值是在一组预先确定且固定的位点上抽样得来的。

（iii）区域数据：其观测值与区域单位（区域对象）的固定数量相关，这些区域单位可能形成一个规则的格点图（如遥感图像）或一组不规则的区域或地区（例如县、区、人口普查区域甚至是国家）。

（iv）空间交互数据（也称为起点—终点流动数据或连接数据）：一对点位置或一对区域的测量数据。