

CAMBRIDGE

# Spatial Data Analysis Theory and Practice

Robert Haining

## 空间数据分析 理论与实践

[英] 罗伯特·海宁 著 李建松 秦昆 译



WUHAN UNIVERSITY PRESS

武汉大学出版社

P208  
H029

CAMBRIDGE

Spatial Data  
Analysis  
Theory  
and Practice

Robert Haining

空间数据分析  
理论与实践

[英] 罗伯特·海宁 著 李建松 秦昆 译



WUHAN UNIVERSITY PRESS

武汉大学出版社

P208  
H029

## 图书在版编目(CIP)数据

空间数据分析理论与实践/(英)罗伯特·海宁著;李建松,秦昆译.  
—武汉:武汉大学出版社,2009.11  
ISBN 978-7-307-07395-1

I. 空… II. ①海… ②李… ③春… III. 地理信息系统  
IV. P208

中国版本图书馆CIP数据核字(2009)第193768号  
著作权合同登记号:图字17-2009-116号

*Spatial Data Analysis: Theory and Practice*, 1st edition ( ISBN 978 -0 -521 -77319 -5 ) by Robert Haining first published by Cambridge University Press 2003

All rights reserved.

This simplified Chinese edition for the People's Republic of China is published by arrangement with the Press Syndicate of the University of Cambridge, Cambridge, United Kingdom.

© Cambridge University Press & Wuhan University Press 2009

This book is in copyright. No reproduction of any part may take place without the written permission of Cambridge University Press and Wuhan University Press.

This edition is for sale in the People's Republic of China (excluding Hong Kong SAR, Macau SAR and Taiwan Province) only.

此版本仅限在中华人民共和国境内(不包括香港、澳门特别行政区及台湾地区)销售。

责任编辑:胡艳 责任校对:黄添生 版式设计:马佳

---

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:湖北恒泰印务有限公司

开本: 720 × 1000 1/16 印张: 23.25 字数: 461千字 插页: 1

版次: 2009年11月第1版 2009年11月第1次印刷

ISBN 978-7-307-07395-1/P · 164 定价: 48.00元

---

版权所有,不得翻印;凡购我社的图书,如有缺页、倒页、脱页等质量问题,请与当地图书销售部门联系调换。

献给我的夫人雷切尔（Rachel），以及我们的孩子  
西莉亚（Celia）、莎拉（Sarah）和马克（Mark）

## 原书出版序

《空间数据分析：理论与实践》(Spatial Data Analysis: Theory and Practice)对空间数据分析领域进行了广泛的论述。该书综述了空间数据分析和位置(地点、背景和空间)在科学以及决策相关研究方面的重要性。涵盖了人们所关注的最新的探索性空间数据分析和空间建模方法,以及如何表达在地理空间中的属性的基本问题。引领读者进入支撑空间数据分析的关键领域,为读者了解和领会这个领域的诸多关键问题提供了一个平台。此书部分内容适合大学本科生和硕士研究生,还包含了让从事空间分析领域研究的地理、社会、经济、环境和统计等学科的科学工作者感兴趣的、具有足够挑战性的内容。

罗伯特·海宁(Robert Haining)是剑桥大学人文地理学教授。他在空间数据分析领域著述广泛,尤其是在经济地理学、医学地理学和犯罪地理学领域的应用方面。他之前的著作《社会和环境科学中的空间数据分析》(Spatial Data Analysis in the Social and Environmental Science)(剑桥大学出版社,1993)得到了好评,并广为引用。

## 译者序

在数字化空间数据大量积累的今天，关于空间数据分析的理论、技术方法及其应用引起了人们极大的兴趣。分析空间数据可以为人们回答科学和决策领域的诸多空间问题。在地理信息科学和应用空间统计学方面，对空间数据分析具有重要影响的一系列理论和实践的发展、在地理信息系统中直接获取数据能力的增长以及大量支持空间数据分析软件的出现，都为提高空间数据分析的能力及应用水平提供了必要的动力。

罗伯特·海宁（Robert Haining）教授于2008年10月应邀到武汉大学进行了为期一周的讲学。他关于空间数据分析方面的精彩讲学内容，深深地感动了每一位听众。他的著作《空间数据分析：理论与实践》涉及内容广泛，理论和实践性强，是引领读者进入空间数据分析研究应用领域的重要读物。结合我国地理信息系统高等教育和空间数据分析研究应用的需要，我们历经近一年的时间翻译了这部著作。希望对我国在该领域的教学、科研和应用起到促进和推动作用。

本书的前言、致谢、引言、第1章至第5章、第8章至第11章以及附录部分由李建松翻译；秦昆翻译了第6章和第7章。李建松对全书翻译稿进行了统稿。

感谢武汉大学出版社王金龙编辑对该书出版所做的工作，感谢胡艳编辑对本书所做的编辑工作。

由于水平有限，在翻译过程中难免出现错误和不当的地方，敬请读者指正。

译者

2009年6月于珞珈山

# 前 言

科研界对分析空间数据越来越感兴趣。这反映了一个良好设定的问题或假说的成立，即空间位置、高质量的空间数据和适用的统计学扮演了一个重要角色。

在本书的写作过程中，为了阐明在分析空间数据方面实际的和潜在的兴趣范围，我涉猎了一系列与科学和决策相关的领域。为了提供这个领域的概览，我把以下两个研究领域放在突出位置：地理信息科学（GISc）和应用空间统计学。

作为理解空间数据分析结果过程的一部分，定义地理实体与如何在一个数据库中以包含属性数据和位置数据的数据矩阵形式获取这个地理实体之间的关系是重要的。对这些数据矩阵操作的有效性——修改或改善初始的表达（如空间平滑）、假设检验（如这幅地图模式中含有事件的空间聚类吗？）或拟合模型（如根据社会经济协变量解释犯罪模式或健康分析结果），将取决于要表达的地理实体在数据矩阵中获取的好坏程度。认识这种联系是重要的，关于这个问题更深层的理解可以参阅GIS学科的相关文献。

我汲取了空间统计学的发展成就，这些成就被用于选定的来自连续表面和来自将区域划分为子区域的数据（如一个城市被分成选区或统计区域）。为了涵盖这方面的资料，我企图描述这些重要的思想并同时用来引导读者涉猎相关专业资源和原创性论文。本书并没有无遗漏地包含空间统计的所有方面（如没有覆盖点处理），也没有涵盖空间分析的所有方面（如没有包含制图建模）。

如果支持分析的软件环境许可，编写一些空间数据分析的程序是非常有帮助的。现在地理信息系统软件被广泛应用于处理空间数据。在大量增加的软件中，一部分实现的空间统计分析方法与地理信息系统软件有关。本书附录向读者介绍了一些相关软件。

## 读者群体

我希望本书介绍空间数据分析技术和模型的方式在适合研究团体的同时，也能适合更为广泛的阅读群体。

本书的部分内容已经作为剑桥大学地理学专业本科二年级学生的课程内容，用8个学时给他们介绍地理信息科学特定的领域和空间分析方法。这些内容包括第1章、第2章和第3章的3.1、3.2.1、3.2.3、3.2.4（a），第4章的部分选定内容

(如探测误差、异常值 (outliers) 和面插值问题), 第 7 章的部分选定内容 (7.1.3 地图平滑) 以及利用正态线性回归模型 (normal linear regression model) 的建模和制图输出例子, 选择了与实际相联系的用于热点探测的简单方法 (7.3.1 (a) 的前半部分) 和逻辑回归建模 (logistic regression for modeling) (11.2.2 (a))。

本书的部分内容已经作为剑桥大学地理信息系统和遥感专业一年制硕士生的课程内容。前 16 学时的课程介绍空间分析的一般方法, 特别是针对连续表面的数据。作为对第 1 章到第 4 章内容的补充, 在 4.4.2 中特别介绍了高斯 (Gauss) 数据的克里格 (Kriging) 方法 (包括第 10 章以及关于半方差函数 (semi-variogram) 建模和估计的引用文献)。后 16 学时通过对犯罪和健康数据的分析, 介绍探索性空间数据分析 (exploratory spatial data analysis) 和空间建模的内容, 这主要是针对区域数据。第 7 章的内容包括了在第 5 章中描述的概念框架提供的介绍。建模方面的部分课程内容选自第 9 章, 实例集中在第 9 章和第 11 章。



## 致 谢

本书是最近两年在剑桥大学完成的,但是本书源于多年前我在谢菲尔德大学一直从事的教学和研究工作。这本书要追溯到 20 世纪 70 年代初在西北大学 Michael Dacey 关于空间自相关的一次学术报告。通过这次报告,我首次接触到空间数据分析的问题。20 世纪 80 年代末,Michael Goodchild 邀请我在圣巴巴拉的 NCGIA 度过的时光,也被证明是具有启蒙意义的。

我深深感谢这些年和我一起工作的朋友和同事。在 20 世纪 90 年代中期,谢菲尔德大学卓有远见地投资创立了研究中心——谢菲尔德地理信息和空间分析中心(SCGISA)。这给了我解决一些不同理论和应用问题的机会,并能够与校内和地方机构展开广泛合作。在此,我特别感谢 Max Craglia、Ian Masser 和 Steve Wise,他们与我共同创立了 SCGISA,完成了很多项目,并开展了诸多有趣的讨论。

在同很多优秀的研究者一起工作的时候,我受益颇多。特别要感谢 Judith Bush、Paul Brindley、Vania Ceccato、Sue Collims、Andrew Costello、Young-Hoon Kim、Jingsheng Ma、Xiaoming Ning、Paola Signoretta 和 Dawn Thompson。在剑桥,我和 Jane Law 在 WinBUGS 项目中,学习应用贝叶斯方法处理犯罪和健康数据。本书第 10 章和第 11 章的例子,很大程度上得益于她的辛勤劳动。我和 Jane 也一起在剑桥地区向当地机构推广这些方法。

本书关于误差传播、数据缺失评估和空间采样方面内容的写作得益于同 Giuseppe Arbia、Bob Bennett、Dan Griffith、Luis Flores、Jinfeng Wang 的研究合作。一些可视化资料和探索性数据分析得益于同 Steve Wise 合作的两个 ESRC 项目。我和 Max Craglia 完成了几个具有较强决策维的项目,特别是最近的两个英国内务部的项目。关于空间建模的一些资料得益于同 Eric Sheppard、Paul Plummer 合作的关于在相互依存的市场中价格变化的建模研究。我对空间分析方法在健康领域应用的研究兴趣,受到同 Marcus Blake、Judith Bush、Dawn Thompson 合作项目的启发,也得益于在谢菲尔德与 David Hall、Ravi Maheswaran 的合作,以及最近和 Andy Cliff 合作的关于美国麻疹数据的研究工作。Martin Kulldorff 热心地给予了我他关于筛查检验的建议。我对于空间数据分析方法在犯罪领域应用的研究兴趣,除了受相关报告的影响外,很大程度上得益于 Tony Bottoms、Andrew Costello、Paul Wiles 的建议。同时也感谢很多人激发了我关于很多相关领域的兴趣,在此,请原谅我不能一一列举他们的姓名。

本书的部分内容已通过剑桥大学本科和研究生教学测试,在此感谢他们为我做出的贡献。同时感谢三个匿名读者读了我著作的前半部分后,提出的良好建议。

感谢剑桥大学地图学实验室的 Phil Stickler 为我绘制插图。感谢剑桥大学出版社的 Tracey Sanderson、Carol Miller、Anne Rix 对本书编辑和出版的指导。

感谢以下同行们允许我在例子中使用他们的数据: Dawn Thompson 和谢菲尔德卫生机构关于乳腺癌的筛查数据;南约克郡警察局的关于犯罪数据集,包括盗窃和欺诈犯罪数据;谢菲尔德儿童服务机构的被开除学生的数据; James Reid 提供的关于剑桥郡的最新行政境界数据;癌症情报部门的 Sara Godward 提供的剑桥郡的肺癌数据, Andy Cliff 提供的美国麻疹数据。

最后感谢我的父母这么多年给予我的鼓励。我的妻子,也是最好的朋友 Rachel 对本书也做出了特别的贡献。正是她的支持,至少由于她尝试新鲜事物的毅力和热情鼓舞了我。

## 版权说明

本书中一些表示境界线的资料的图形版权属于 ED-LINE 机构的 Crown。普通测绘数据是由 EDINA 数字地图(JISC 支持服务)提供的。本书中的一些图形在原出版者的授权下经过了再处理。列举如下,并在附录中有详细说明。

图 3.5:克鲁沃(Kluwer)学术出版社。取自《Geo-ENV II 环境应用地理统计》,由 J. Gomez-Hernandez、A. Soares、R. Frodevaux(1998) Savelieva 等编辑。《切尔诺贝利市(乌克兰)辐射性尘埃的条件随机协同模拟》,图 14,第 463 页。

图 4.3:介子(Pion)有限公司,伦敦。取自 M. Tranmer and D. G. Steel(1998)的《生态谬误(生态学谬误)调查》,《环境和规划》,A(30)图 1,第 827 页和图 2,第 830 页。

图 6.2:泰勒和弗朗西斯(Taylor and Francis)有限公司,伦敦。取自 M. Craglia、R. Haining 和 P. Wiles(2000)的《城市犯罪模式分析方法的比较评估》,《城市研究》(37),图 5,第 725 页。

图 6.3:牛津大学(Oxford University)出版社。取自《公共健康医学学报》(1994),R. Haining、S. Wise 和 M. Blake 的《构造小面积健康分析的区域》,图 3,第 433 页。

图 6.4:克鲁沃学术出版社。取自《数学地质学》,20(1989),M. Oliver 和 R. Webster 的《多元分类中的空间加权地理统计基础》,图 3,第 15—35 页。

图 6.7:克鲁沃学术出版社。取自 G. Verly 等(1984)的《自然资源特性化的地理统计》,N. Cressie 的《面向约束的地理统计》,图 8,第 33 页。

图 6.8—图 6.12:施普林格(Springer-Verlag)出版社。取自《地理系统学报》2(2002),R. Haining 等的《为空间数据分析提供科学可视化》,第 121—140 页。

图 7.1:约翰威立父子(John Wiley and Sons)有限公司。取自《医学统计》(1999),K. Kafadar 的《美国郡县模拟平滑和调整的死亡率》,图 2(a)—(d)。也感谢 Kafadar 博士提供的这些图形的数字版本。

图 7.3 和图 7.7:劳特利奇(Routledge)出版社。取自《GIS 和健康》,由 A. Gatrell 和 M. Loytonen 编辑, M. Kulldorff 的《空间流行病学统计方法》,图 4.1 和图 4.2。

图 7.8:约翰威立父子有限公司。取自《医学统计》(1988),R. Stone 的《假定源附近的过度环境风险调查》,图 3。

图 8.1 和图 8.2: 俄亥俄州立大学 (Ohio State University) 出版社, 哥伦布市, 俄亥俄。取自《地理分析》(1991), R. Haining 的《空间数据二元相关》, 图 2、图 3 和图 5。

图 8.3: 国际生物统计学会 (International Biometric Society)。取自《生物统计》(1997), A. Cerioli 的《空间数据  $2 \times 2$  表格中的独立性调整型检验》, 图 2, 第 619—628 页。

图 10.1: 俄亥俄州立大学出版社, 哥伦布市, 俄亥俄。取自《地理分析》(1994), D. Griffith 等的《空间数据集中属性采样误差的异质性》, 图 1, 第 31a 页。

图 11.2: 泰勒和弗朗西斯有限公司, 伦敦。取自《城市研究》(2001), M. Craglia 等的《英国城市高密度犯罪区域建模》, 第 1931 页。

# 目 录

前言	1
致谢	1
版权说明	1
引言	1
0.1 关于本书	1
0.2 什么是空间数据分析?	3
0.3 本书的写作动机	4
0.4 本书的组织结构	6
0.5 空间数据矩阵	8

## 第一部分 空间数据分析的背景

第1章 空间数据分析:科学和决策背景	11
1.1 科学领域的空间数据分析	11
1.1.1 科学解释领域中地点、背景和空间的一般性问题	12
1.1.2 空间过程	15
1.2 科学解释特定领域中的地点和空间	16
1.2.1 空间分支科学的定义	16
1.2.2 举例:选择的研究领域	17
1.2.3 问题求解中的空间数据分析	24
1.3 决策领域的空间数据分析	26
1.4 空间数据分析中产生的问题举例	28
1.4.1 描述和地图解释	29
1.4.2 信息冗余	29
1.4.3 建模	29
1.5 小结	29

<b>第2章 空间数据的性质</b> .....	31
2.1 空间数据矩阵:概念化和表达问题 .....	31
2.1.1 地理空间:对象、空间域和地理几何表达 .....	31
2.1.2 地理空间:属性值的空间相关性 .....	33
2.1.3 变量 .....	34
2.1.4 样本还是样本总体? .....	38
2.2 空间数据矩阵:它的形式 .....	39
2.3 空间数据矩阵:它的质量 .....	42
2.3.1 模型质量 .....	43
2.3.2 数据质量 .....	45
2.4 空间相关性的量化 .....	54
2.5 小结 .....	63

## 第二部分 空间数据:数据获取及质量问题

<b>第3章 通过采样获取空间数据</b> .....	67
3.1 空间数据源 .....	67
3.2 空间采样 .....	68
3.2.1 空间采样的目的和方法 .....	68
3.2.2 基于模型和设计的空间采样方法 .....	71
3.2.3 采样方案 .....	73
3.2.4 选择采样方法的问题 .....	76
3.3 模拟制图 .....	83
<b>第4章 数据质量:对空间数据分析的影响</b> .....	86
4.1 数据和空间数据分析中的误差 .....	86
4.1.1 测量误差模型 .....	86
4.1.2 粗差 .....	88
4.1.3 误差传播 .....	92
4.2 数据分辨率与空间数据分析 .....	94
4.2.1 变量的精度和显著性检验 .....	95
4.2.2 支持改变问题 .....	96
4.2.3 使用聚合数据分析关系 .....	102
4.3 数据一致性和空间数据分析 .....	112
4.4 数据完整性与空间数据分析 .....	113

4.4.1	缺失数据问题 .....	114
4.4.2	空间插值和空间预测 .....	122
4.4.3	边界、权重矩阵和数据完整性 .....	130
4.5	小结 .....	132

### 第三部分 探索性空间数据分析

<b>第5章</b>	<b>探索性空间数据分析:概念模型 .....</b>	<b>137</b>
5.1	探索性数据分析和探索性空间数据分析 .....	137
5.2	空间变化的概念模型 .....	138
<b>第6章</b>	<b>探索性空间数据分析:可视化方法 .....</b>	<b>142</b>
6.1	数据可视化和探索性数据分析 .....	142
6.1.1	数据可视化:方法和任务 .....	143
6.1.2	数据可视化:通过计算机的发展 .....	145
6.1.3	数据可视化:选择的技术 .....	146
6.2	空间数据可视化 .....	147
6.2.1	聚合数据的数据准备问题:变量值 .....	147
6.2.2	聚合数据的数据准备问题:空间框架 .....	150
6.2.3	空间可视化的特殊问题 .....	155
6.3	数据可视化和探索性空间数据分析 .....	159
6.3.1	空间数据可视化:为单变量数据选择的技术 .....	159
6.3.2	空间数据可视化:为双变量或多变量选择的技术 .....	164
6.3.3	谢菲尔德地区乳腺癌筛查数据的摄取 .....	165
6.4	小结 .....	169
<b>第7章</b>	<b>探索性空间数据分析:数值方法 .....</b>	<b>170</b>
7.1	平滑方法 .....	170
7.1.1	图形绘图的约束平滑 .....	170
7.1.2	空间相关性的约束描述 .....	171
7.1.3	地图平滑 .....	172
7.2	全局地图特性的探索性识别:整体聚类 .....	183
7.2.1	区域数据的聚类 .....	183
7.2.2	标记点模式的聚类 .....	187
7.3	局部地图特性的探索性识别 .....	190

7.3.1	聚类探测 .....	190
7.3.2	焦点检验 .....	201
7.4	地图比较 .....	203

## 第四部分 假设检验和空间自相关

第8章	存在空间相关性的假设检验 .....	209
8.1	空间自相关和空间数据集均值检验 .....	210
8.2	空间自相关和二元关联检验 .....	213
8.2.1	皮尔森积矩相关系数 .....	213
8.2.2	列联表的卡方检验 .....	218

## 第五部分 空间数据建模

第9章	空间数据统计分析模型 .....	223
9.1	描述性模型 .....	225
9.1.1	大尺度空间变化模型 .....	225
9.1.2	小尺度空间变化模型 .....	226
9.1.3	空间变化的多尺度数据模型 .....	236
9.1.4	层次贝叶斯模型 .....	237
9.2	解释性模型 .....	241
9.2.1	连续取值响应变量模型:正态回归模型 .....	241
9.2.2	离散取值区域数据模型:广义线性模型 .....	244
9.2.3	层次模型 .....	247
第10章	空间变化统计建模:描述性建模 .....	251
10.1	表达空间变化的模型 .....	251
10.1.1	连续取值变量模型 .....	251
10.1.2	离散取值变量模型 .....	257
10.2	空间变化建模中若干一般性问题 .....	262
10.3	层次贝叶斯模型 .....	263
第11章	空间变化统计模型:解释性建模 .....	273
11.1	空间数据建模的方法 .....	273
11.1.1	经典方法 .....	273



---

11.1.2	计量经济学方法 .....	275
11.1.3	数据驱动方法 .....	278
11.2	空间数据线性建模的一些应用 .....	280
11.2.1	地区收入收敛检验 .....	280
11.2.2	二元响应模型 .....	282
11.2.3	多级模型建模 .....	285
11.2.4	谢菲尔德入室盗窃的贝叶斯建模 .....	286
11.2.5	被开除学生的贝叶斯建模 .....	293
11.3	小结 .....	295
附录 I	软件 .....	296
附录 II	剑桥郡肺癌数据 .....	298
附录 III	谢菲尔德盗窃数据 .....	302
附录 IV	谢菲尔德被开除的学生数据 .....	308
参考文献	.....	311