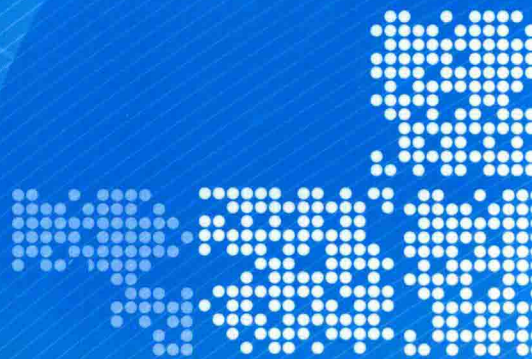




地统计学（现代空间统计学）

郑新奇 吕利娜 主编



科学出版社

地统计学（现代空间统计学）

郑新奇 吕利娜 主编

科学出版社

北京

内 容 简 介

地统计学是以具有空间分布特点的区域化变量理论为基础,研究自然现象或人文现象的空间变异与空间结构的一门学科。它属于研究土地利用变化、土壤理化性状、矿产分布、资源禀赋、生物群落、地貌类型、气象气候监测及社会经济领域有着特定地域分布特征现象的空间统计学范畴。本教材主要介绍地统计学在研究空间分布数据的结构性和随机性、空间相关性和依赖性、空间格局与变异,以及对空间数据进行最优无偏内插、模拟空间数据的离散性及波动性中的应用。读者通过学习本教材,可以掌握基本的地统计学理论、方法和技术,并借助地统计软件工具解决实际问题。

本教材可作为地质学、测绘科学与技术、地理学、环境科学、气象气候学专业的本科生、研究生教材,还可供社会经济领域需要进行空间统计分析的科技工作者阅读参考。

图书在版编目(CIP)数据

地统计学(现代空间统计学)/郑新奇,吕利娜主编. —北京:科学出版社, 2018.7

ISBN 978-7-03-056141-1

I. ①地… II. ①郑… ②吕… III. ①地质统计学 IV. ①P628

中国版本图书馆 CIP 数据核字(2017)第 317797 号

责任编辑:杨红 程雷星/责任校对:樊雅琼

责任印制:吴兆东/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>

北京厚诚则铭印刷科技有限公司 印刷

科学出版社发行 各地新华书店经销

*

2018年7月第一版 开本:787×1092 1/16

2018年7月第一次印刷 印张:11 3/4

字数:300 000

定价:49.00元

(如有印装质量问题,我社负责调换)



前 言

美籍瑞士地理学家、制图学家、加州大学圣塔芭芭拉分校 Waldo Tobler 教授提出了地理学第一定律,即相邻的事物相似,远离的事物相异。正是由于地学现象这一规律,才成就了地统计学。地统计学的基本思想自 20 世纪 50 年代初开始提出,经过广大地质工作者、数学地质研究者、地质统计学学者、矿山地质和采矿设计专家及其他地质统计学应用者和爱好者的不断努力,现在已经形成了一套独立的理论体系,成为在自然科学和社会科学中研究活跃的一个领域。

地统计学,又称地质统计学、空间统计学。它假设空间分布的属性表现出某种程度的连续性,旨在描述地学空间中某一属性的分布。地统计学区别于经典统计学的最大特点即是地统计学既考虑样本值的大小,又重视样本空间位置及样本间的距离,弥补了经典统计学忽略空间方位的缺陷。地质统计学研究的内容主要包括:区域化变量空间变异结构分析、变异函数理论、克里金空间估计及随机模拟,其中变异函数理论和区域化变量的空间变异结构分析等重要内容是进行空间克里金估值的基础和前提。

作者从 2008 年开始在中国地质大学(北京)面向研究生开设“地统计学”课程,中间虽有间断,也历经了近 10 年。“地统计学”从最初的小众研究到现在的自然科学和社会科学通用,越来越走向大众化;地统计学从最初在专业人员和研究生中的高级应用,随着计算机技术的迅速发展,越来越走向普及化。无论是自然科学还是社会科学,应用地统计学的理论和方法解决空间变异和空间插值问题的研究者越来越多,尤其在本科毕业论文写作中日益得到广泛应用。对于理工科学生而言,对空间信息的量化处理方法掌握是一个必需的基本技能。新兴的、发展迅速的地统计学也在多专业得到了广泛应用。目前,“地统计学”也从 GIS 及其相关专业的部分章节内容,尤其是时空插值方法和超级计算的支持,越来越扩展到独立设课。一些理工科行业性高校的本科中独立开设“地统计学”的也越来越多。

全书内容分为三大模块:第 1~3 章为第一模块,是通识基础知识,在写法上不追求面面俱到,而是“点到”为止,与本书内容密切的部分则适当进行了扩展。读者可以借助现在互联网和泛在学习环境,对有关的知识进行拓展学习。第 4~6 章为第二模块,是本书的重点理论和原理部分。通过较为详细的内容阐述,告诉读者实际问题时的思路、方法选择和注意事项,避免可能由对方法原理等理解不全面而导致对结果的误判。第 7 和第 8 章为第三模块,是实践部分。读者通过介绍的主要软件特点及常用和相对流行的软件(ArcGIS、Surfer 等)作为主要工具进行上机学习。读者通过软件操作可以基本掌握软件的使用。

考虑已有的地统计学(地质统计学)教材偏理论而轻实践的情况,本教材在兼顾和保留尽可能多的理论知识的基础上,重点强化了利用地统计学进行实践的内容,让读者在了解基本理论和原理的前提下,可以快速解决实际问题。总之,本书全面梳理了国内外研究的新进展,对地统计学的学科进行了重新定位。在继承已有研究成果的基础上,将新的理论、方法、技术、应用写入教材,以反映本领域的最新成果,其中包括作者的研究和教学成果。同时兼顾了自然科学和社科科学研究领域;理论和实践密切结合,以应用引导理论学习,提高读者的动手能力。整本书构成了了解进展与发展特点—理解基本知识和基本理论—掌握基本方法及限制条件—

提出解决问题的思路和途径—通过技术手段解决问题—决策服务的体系。通过系统的培训，使一般工作者达到会用，为高级工作者提供开放探索空间。

本书是集体智慧的结晶：一是按照作者多年从事教学这个课程的体会和思考，拟定了编写提纲；二是编写过程中，让参与学习和培训的部分研究生参与了资料的整理和编写工作；三是主要内容让不同专业的同学进行阅读试用，最后形成目前这个版本。参与本书编写的人员有：高晓晨、罗广芳、李国勇、李红涛、吕慧敏、董芳玢、范丽娜、王洪丹、施园园、秦倩、李胜男、王平安、段晓燕、姚思瑶、陈宇琛、陈学砧、徐阳、王林、张月恒等。郑新奇、吕利娜参与了全部书稿的编写，最后由郑新奇、吕利娜统稿、定稿。

本书成稿过程中，获得教育部本科教学改革与教学质量工程试点项目“地理信息科学专业综合改革”、教育部首批新工科项目“地学特色新工科计算机通识课程体系优化设计与实践探索”资助，尤其是得到王家耀院士等专家的指导，同时受到中国地质大学(北京)教务处、信息工程学院 GIS 教研室老师们在讨论过程中的启发。在此表示衷心感谢！

本书写作过程中参阅了大量文献，除了书中已经标注的文献外，可能还有一些没有标注出来。在这里对已标注的和给我启发帮助的文献的作者一并表示感谢。

限于时间和作者水平，书中还存在这样或那样的问题，敬请读者批评指正。

郑新奇

2017年12月

目 录

前言

第 1 章 绪论	1
1.1 地统计学概念及发展	1
1.2 地统计学学科定位	3
1.3 本书目的与内容安排	7
1.4 地统计学学习要点	7
第 2 章 数理统计基础	10
2.1 经典统计学方法	10
2.2 空间统计学方法	17
第 3 章 地统计学理论基础	26
3.1 地统计学的基本假设	26
3.2 区域化变量的概念及性质	28
3.3 变异函数与结构分析	30
3.4 协方差函数	40
第 4 章 克里金法及应用条件	44
4.1 概述	44
4.2 简单克里金法(SK)	46
4.3 普通克里金法(OK)	49
4.4 泛克里金法(UK)	52
4.5 对数正态克里金法(LK)	56
4.6 指示克里金法(IK)	57
4.7 析取克里金法(DK)	58
4.8 协同克里金法(CK)	60
第 5 章 地统计学建模新方法	67
5.1 概述	67
5.2 时空多元空间信息统计分析	70
5.3 分形空间信息统计分析	81
5.4 模式识别空间信息统计分析	87
第 6 章 估值的不确定性和采样策略	94
6.1 空间变异性及尺度效应	94
6.2 估值的不确定性	99
6.3 采样设计策略	100
6.4 样本形状的影响和最佳样块大小	104

第 7 章 地统计学计算工具及结果可视化	110
7.1 地统计学计算工具发展历程	110
7.2 IDRISI 操作与应用	110
7.3 GeoDa 操作与应用	120
7.4 ArcGIS 操作与应用	129
7.5 Surfer 操作与应用	138
第 8 章 上机练习	145
8.1 ArcGIS 上机练习	145
8.2 Surfer 上机练习	167
8.3 IDRISI 上机练习	172
主要参考文献	178

第 1 章 绪 论

1.1 地统计学概念及发展

现实中常遇到这样的场景：农作物施肥时，需要采用测土配方施肥技术有针对性地按需按量补充作物所需的营养元素，实现各种养分平衡供应，满足作物的需要。常规做法是采集土壤样本进行土壤养分检测，大致了解每块土地的养分情况。但是这样得到的结果实际上只代表土壤样本的情况，若样本相距较远，则了解样本和样本之间的情况也很重要。例如，环保工作者希望了解 $PM_{2.5}$ 浓度在城市中的空间连续分布情况；地貌学家需要知道地形在空间的连续起伏变化情况。也就是说，人们关心的不仅是空间采样点的情况，同时关心自然现象在空间上的分布状况。因此，此类问题可归结为“如何将离散的空间采样点转化为连续表面”的问题。那么如何做到这一点呢？以获取地形分布数据为例，一种解决方法是增加高程点采集密度，然而由于人力、物力、财力等客观因素的限制，采集数量不可能无限增多，事实上也不可能做到在无限多的点上进行数据采集；另一种方法是通过已有的高程值来估计其他未取点位上的高程值，从而得到地形在空间的连续分布情况，即空间插值。

空间插值常用于将离散点的数据转换为连续的数据曲面，其方法很多，如反距离加权插值法、全局多项式插值法、径向基函数插值法等，这类方法往往直接通过周围观测点的值内插或者通过特定的数学公式来内插，而较少考虑观测点的整体空间分布情况。相较于前述方法，地统计插值法则是基于空间自相关分析，依据数据的空间变异规律进行插值得到最优无偏估计量及相应插值结果的精度。相比于经典概率论和数理统计学，地统计学在空间预测和不确定性分析方面具有明显的优势，其应用领域从最初的地质、采矿领域，已逐步拓展到农业、环境、生态、社会科学等多个领域。

1.1.1 地统计学的起源

地统计学(geostatistics)，又称地质统计学，形成于 20 世纪 50 年代，由南非矿产地质工程师克里金(Kriging)和西舍尔(Sichel)等在估计南非金矿储量时提出，是一种根据样品的空间位置、相关程度及样品品位权重，进行滑动加权平均，从而估计未知样点上样品平均值的方法，即克里金法。该方法克服了经典统计学将地质变量看成纯随机变量而忽略其空间相关性的不足，能够降低估计误差。随后，法国著名统计学家 G.Matheron 教授在克里金和西舍尔两人工作的基础上，于 1962 年提出“地统计学”概念，并在出版的《应用地统计学论》(*Traité de Géostatistique Appliquée*)专著中阐明了地统计学原理，采用随机函数来描述地质变量的结构性和随机性，将地质统计学与传统的统计学分开，提出了区域化变量、简单克里金、普通克里金、泛克里金的概念，将早期的零散科研成

果理论化和系统化,奠定了地统计学的理论基础。从此,地统计学作为一门新兴的边缘学科诞生了。随后地统计学理论与方法得到了进一步的完善和改进,特别是在实践应用中获得了快速的发展,形成了两种类型的理论体系:一类是有参数的克里金方法,另一类是没有参数的克里金方法,有参数的克里金方法是指所研究的数据必须符合正态分布,如析取克里金;而没有参数的克里金方法对所研究的变量的分布没有特殊要求,如指示克里金和概率克里金。20世纪90年代三维和时空地统计学得以发展,并出现了大量的相关软件,地统计学进入成熟阶段。2000年至今,地统计学进入创新性的二次开发阶段,在原有基础上开发出不确定性的和新型的地统计方法与模型,使其应用领域得到进一步扩展。

地统计学在我国的发展起源于20世纪70年代后期。随着1977年美国福祿尔采矿金属有限公司(Flour Mining & Meta Incorporation)的H.M.Parker博士的来华访问,我国的数学地质研究者及勘探、矿山设计人员系统地学习了地统计学的基本概念和内容,将其深化到了我国地质矿业的应用领域,并编译了一系列著作与教材,如侯景儒编著的《矿业地质统计学》《实用地质统计学》,王仁铎和胡光道编著的《线性地质统计学》,孙洪泉编著的《地质统计学及其应用》,张仁铎的《空间变异理论及应用》等,为我国的地质统计学理论与应用研究打下了坚实的基础。

1.1.2 地统计学的概念

地统计学是20世纪60年代由法国著名统计学家G.Matheron创立的一门新的统计学分支,并应用与发展于采矿学、地质学等地学领域。G.Matheron(1962)首先采用了“地统计学”一词,并将其定义为:“地统计学即以随机函数的形式在勘查与估计自然现象中的应用”。之后,随着地统计学的发展,他又将地统计学定义修改为:“地统计学是区域化变量理论在评估矿床上的应用(包括采用的各种方法和技术)”。然而,地统计学发展至今,不仅在地质学,而且在土壤、农业、气象、海洋、生态、环境等各学科领域都得到应用和发展。因此,一些地统计学工作者将这一概念修订为:“地统计学是以区域化变量理论为基础,以变异函数为主要工具,研究在空间分布上既有随机性又有结构性,或空间相关和依赖性的自然现象的科学”。

1.1.3 地统计学的内容

从定义来看,地统计学主要包含三方面内容。

1. 理论基础——区域化变量(regionalized variable)理论

自然科学研究中的许多变量都具有空间分布的特点,它们通常随所在空间位置的不同表现出不同的数量特征,如降水量、气温、海拔、土壤有机元素含量等,这些变量称为区域化变量,其所描述的现象称为区域化现象。区域化变量也称为区域化随机变量,是根据其在一个区域内的位置不同而取值,即它是与位置有关的随机函数。因此,区域化变量具有两个最显著也最重要的特征,即随机性和结构性。一方面,区域化变量是随机函数,它具有局部的、随机的、异常的特征;另一方面,区域化变量具有结构性,即在空间位置上相邻的两个点具有某种程度的自相关性。

2. 主要工具——协方差函数和变异函数

区域化变量的结构性和随机性需要一种合适的函数和模型来表述,使其两者均能兼顾,这就是协方差函数和变异函数。协方差函数和变异函数是地统计学中以区域化变量理论为基础建立的两个最基本函数,是描述区域化变量的主要工具。

3. 主要内容——克里金插值

克里金插值法,又称空间局部估计法或空间局部插值法,是地统计学的主要内容之一。克里金法建立在变异函数理论及结构分析的基础之上,其实质是利用区域化变量的原始数据及其相应变异函数的结构特征,对未知点的区域化变量取值进行线性无偏最优估计。该方法最初由南非矿产工程师克里金应用于寻找金矿,因此 G.Matheron 就以“克里金”的名字命名了该方法。

1.1.4 地统计学的发展

当前,地统计学的理论体系不断完善和发展,与其他学科的相互渗透也促进了地统计学理论体系与应用水平的不断提高。地统计学的发展趋势主要表现在以下四个方面:①注重学科交叉,发展旧理论与探索新方法并行。例如,探寻替代估计变异函数的模型方法,发展不确定性地统计学;研究机理模型与地统计学之间的耦合,发展基于地统计学的不确定性决策;加强地统计学与专家系统、地理信息系统、神经网络及人工智能等的结合与应用。②加强时空地统计学的研究,真正实现在时间-空间域上的动态估值。③完善地统计学软件功能,着重加强软件的可视化研究及图形输出质量。④注重实际应用,拓宽应用领域。进一步加强与拓宽地统计学与其他学科的相互渗透,研发新的空间数据估值应用领域,如健康与公共卫生、社会科学等。

1.2 地统计学学科定位

1.2.1 地统计学的学科分支

基于概率统计视角,地统计学可以说是应用统计学的一个新分支;基于地质应用视角,地统计学是数学地质的一个独立分支;基于空间信息统计分析的视角,地统计学又是地理信息科学的一个重要分支。因此,地统计学是统计学、地质学、地理学等的一个交叉学科。

1.2.2 地统计学的研究内容

1. 空间估值

空间估值是根据在空间上分布的离散采样点值估求出未知点值,或将离散的空间数据点转化为连续的空间数据曲面。地统计学领域中将这种估值方法统称为克里金法。它是一种以得到无偏最优估计量为目标的广义的最小二乘回归算法,即估值误差的数学期望值为 0,方差达到最小。

2. 局部不确定性预测

克里金无偏最优估计量存在两个假设条件:①假设估计误差的频率分布是对称的;

②克里金误差只与数据构型相关,而与具体数值无关。但是实际情况中,数据往往会存在低值高估、高值低估的问题,或者出现待估点被一个大值和小值所包围的估值误差往往大于被两个同等规模小值包围时的误差的情况。因此,在进行未知点估值时还应考虑待估点周围样本点的影响,利用条件概率模型来推断局部不确定性。局部不确定性预测法有参数法(如众高斯方法)和非参数法(如指示克里金法)两种。

3. 随机模拟

根据随机变量的定义,每个变量可以有多个实现(realization)。也就是说,在总体趋势是正确的前提下,每个未知点上的变量估值可以有多种情况,这种方法称为随机模拟。随机模拟可以利用各种类型数据(如“硬”的采样点测量数据,“软”的各种类型的间接测量数据)生成众多的实现,每一个实现采用不同的表现方式展现同一种空间格局。随机模拟方法有高斯序列模拟、LU分解模拟、高斯指示模拟、P-field模拟、模拟退火算法等。但是,克里金法所获得的估值是唯一的,它虽然完成了对空间格局的认知,但没能使其再现。

4. 多点地统计学

传统地统计学利用变异函数来量化空间格局,但是变异函数只能度量空间上两个点之间的关联,即在二阶平稳或内蕴假设下空间上任意两点之间的相关性,而难以表征复杂的空间结构和再现复杂目标的几何形态。例如,不同弯曲河道的变异函数在同一方向上可能是十分相似的,因而不能通过变异函数加以区分。对于关联性很强的情况或者研究对象有较为明显的曲线特征时,要想量化其空间格局就需要包含多个空间点。多点地统计学则可以通过多个点的训练图像来取代变异函数从而有效反映目标的空间分布结构。该方法产生并主要应用于石油领域。

1.2.3 地统计学的应用领域

涉及空间分布数据的结构性和随机性、空间相关性与依赖性 & 空间格局与变异的研究,以及对这些数据进行无偏最优内插估计或数据离散性、波动性的模拟,均可考虑采用地统计学的理论与方法。

1. 地统计学在地质学中的应用

地统计学传统应用于地质、采矿领域,属于地质学领域中数学地质的一个分支,在应用中积累了较多的资料和经验,主要集中在以下三个方面。

(1) 矿产资源储量计算及平均品位估计。地统计学从地质、采矿的实际出发,根据矿床地质变量的特点,最大限度地利用勘探工程所提供的各种信息,进行储量的整体估计和局部估计,并能够在开采前定量地给出储量的估计精度。同时,该方法能够与计算机相结合,实现储量计算的自动化。相比于传统储量计算方法,地统计学对于矿产资源储量计算及平均品位估计等问题具有无可比拟的优势。目前,应用地质统计学的企业公司也相继增多,世界上已经有 200 多个矿山成功地应用地统计学计算储量,涉及的矿种有铁、铜、金、银、钨、锡、铅、锌、铀、铬等金属矿产及煤气、煤田、磷矿等非金属矿床。

(2) 矿产资源预测及找矿勘探。地统计学可以利用泛克里金法、对数正态克里金法等进行地质数据处理;利用估计方差进行储量分级、矿床勘探类型的定量分析;利用变

异函数及其参数研究地质变量空间分布规律;采用泛克里金法进行成矿预测发现异常远景区;利用变异函数和估计方差进行勘探网度的优化与评价。

(3)石油勘探开发。随着地统计学在石油勘探开发中的应用日益广泛与深入,逐渐发展形成了一门新兴学科——石油地质统计学,主要包括地质建模、地震数据反演及生产数据整合三方面。①地质建模:主要利用多点地统计学进行储层规模模拟,估计地层的埋深、层厚、孔隙度、渗透率和含油饱和度等地质和地球物理参数的空间分布,绘制各种地质图件;利用变异函数研究储层的非均质性及各向异性;实现储层预测。②地震数据反演:主要是基于井数据及变差函数模拟波阻抗值,利用估值转换为合成地震记录以便与真实地震对比预测。③生产数据整合:整合地震、测井、钻井和露头等各种信息并进行建模。

2. 地统计学在土壤学中的应用

许多自然现象在空间上是连续变化的,空间上相近的点相较于距离较远的点在理化性质等方面具有更大的相似性,即统计学意义上的相互依赖性,这与区域化变量和地统计学应用的前提是一致的。目前,土壤特性的空间变异研究是土壤科学热点之一,地统计学在其中的应用主要集中在以下四个方面。

(1)土壤物理性质空间变异。主要是利用地统计学在土壤物理参数、状态参数等方面进行空间变异研究探索,如土壤颜色、颗粒组成、土壤水分、土壤水力传导度、团聚体大小、饱和水压、孔径等土壤物理性质方面的空间变异研究。

(2)土壤化学性质空间变异。主要是利用地统计学对土壤化学性质进行空间变异性研究,如针对氮、磷、钾、钙、镁、土壤pH等土壤养分的空间相关性研究。利用变异函数及其模型和克里金插值方法,从不同空间尺度、不同时间尺度、不同景观尺度及不同环境因素条件对土壤养分等化学性质的空间变异性进行大量探索,描述和归类土壤化学性状的空间属性,从而为土壤养分管理、土壤环境背景制图等提供必要的数据和他方法。

(3)土壤学试验设计和采样方法。地统计学方法能够分析土壤特性的空间变异规律,指导土壤采样数目、采样密度、样点分布及采样方法的确定,从而有效解决野外采集数据不具备统计意义的问题。

(4)土壤质量管理。土壤质量是全球生物系统赖以持续发展的关键因素之一。利用地统计学方法可以通过已知取样点的数据去估测未知采样点土壤特性指标,判读其是否超过某一阈限,为制定管理规范、确定经营策略等土壤质量管理工作提供重要参考指标。近年来,地统计学在土壤质量管理上的应用主要集中在土壤养分管理和土壤污染研究方面。

3. 地统计学在生态学中的应用

生态学是研究生物之间及生物与非生物环境之间相互关系的学科,不同物种在不同时间、不同地点的分布是该学科重要的研究内容之一。在生物体分布和环境因素之间的空间变化分析中,空间依赖性分析尤其重要。但是,传统的统计分析方法认为样本间是相互独立的,因而忽略了这个问题。鉴于地统计学的特点及在土壤空间变异性 and 格局研究中的优越性,生态学者将其引入生态学领域,并广泛应用到昆虫生态学、水生态学、景观生态学和植被生态学等多个专业方向。地统计学为生态学家提供了一个非常有效的分析和解释空间数据的方法,具体如下。

(1)生态学变量空间变异性的定量描述和解释。通过分析变量空间格局的尺度、几何形状、变异方向等信息,将空间格局与生态学过程联系起来,有助于更好地理解研究对象的空间格局特征,给出合理的生态学解释。例如,通过对气候、地形、土壤、植被等蝗虫生境因子的研究,以及解释蝗虫空间分布特征的成因,进一步分析生境因子和蝗虫之间的关系,为蝗虫防治和蝗灾预测模型的研究提供科学依据。

(2)生物特征的估计。基于采样点数据反映的空间结构特征,估计未知点的密度、数量等生物特征,以获取研究对象的空间分布格局。

(3)生态学研究对象的时空变化规律分析、不同相关研究对象的时空动态及耦合关系分析。分析研究某一物种在不同时期的空间变异性和空间分布特征,发现该物种发展过程中的时空变化规律。分析研究不同物种的时空分布,探究其样本之间的时空相关性及依赖性,说明其时空动态及耦合关系。

同时可将地统计学与地理信息系统、遥感、全球卫星导航定位系统结合起来以快速准确地获取生物环境信息,并对信息进行有效的管理和分析使用。

4. 地统计学在环境科学中的应用

环境是一个时空连续体,其样本具有空间相关性的特征,难以利用经典的统计分析方法全面描述和探究环境变量自身的空间变异特征及不同环境因子之间的空间相关关系。地统计学则能够有效地解决这一问题,其在环境科学研究中的应用主要表现在以下三个方面。

(1)土壤环境研究。土壤重金属污染是目前土壤环境治理的重要内容与研究热点之一。通过利用地统计学中的变异函数和克里金插值技术进行重金属空间结构分析、模拟和估值,从而描述和模拟污染物的空间分布特征及估算未采样点的取值,揭示污染物在空间上的分布迁移趋势。此外,结合现有的污染背景资料,还可以识别出各种可能的污染源。

(2)水环境研究。地统计学在水环境研究中的应用主要包括:利用半方差函数分析和克里金插值分析等进行地下水位埋深异质性分析及预测;利用地统计学优化环境监测网点的位置和数目,分析预测水环境污染物浓度、研究水质参数的估算及变异性,从而考察植物生长、土地利用等对水质的影响。

(3)其他相关领域研究。地统计学在优化大气质量评估中采样位置、大气污染物分布、声环境评价研究中也取得了很好的效果。

5. 地统计学在气象学中的应用

数值天气预报和日常气象分析经常需要将不规则的站点资料插值到规则的网格进行分析,常采用以距离函数为权重的插值方法,如反距离加权插值法和逐步订正法等。随着地统计学方法的兴起,克里金法已经逐步应用于气象学领域,对气温、降水、光合有效辐射量、蒸发量等气象数据,通过构建变异函数进行了空间、时空插值分析,得到了较高的精度。

目前,地统计学除在传统的地质、采矿领域及土壤学、生态学、环境科学、气象学领域应用外,在其他空间相关领域中也崭露头角。例如,将地统计学引进园艺和农业领域,研究作物长势;将地统计学应用于临床医学,利用对偶克里金法重建了脊椎侧突的三维影像;将地统计学引入遥感领域,结合时间和空间信息,用克里金法处理时间序列

为 30 年的地面覆盖物遥感图像, 提高地面覆盖物变化预测准确率。此外, 地统计学在公共卫生、社会科学等领域也有所涉足, 显示出越来越强大的生命力。

1.3 本书目的与内容安排

地统计学是以具有空间分布特点的区域化变量理论为基础, 研究自然现象或人文现象的空间变异与空间结构的一门学科。它是针对像土地、土壤、矿产、资源、生物群落、地貌, 以及人类活动等有着特定的地域分布特征的空间统计方法。本书主要目的是在介绍经典统计学方法与空间统计学方法的基础上, 重点阐述地统计学区域化变量、变异函数、协方差函数等基础理论与概念, 并对常用的克里金插值方法与应用条件进行描述。同时, 本书也对地统计学建模的新方法进行了论述, 并为学习者介绍了目前主流地统计软件及工具的使用流程。读者通过阅读学习可以掌握基本的地统计学方法和技术, 以及借助地统计软件工具解决实际问题的基本技能。

考虑已有的地统计学(地质统计学)教材偏理论而轻实践的情况, 本教材则在兼顾和保留尽可能多的理论的基础上, 重点强化了利用地统计学进行实践的内容, 让读者在了解基本理论和原理的前提下, 可以快速解决实际问题。

1.4 地统计学学习要点

1.4.1 学习地统计学课程总体要求

1. 打好数理统计基础

地统计学的理论基础是区域化变量理论, 主要研究那些分布于空间并显示出一定结构性和随机性的自然现象; 而协方差函数和变异函数是以区域化变量理论为基础建立起来的, 是地统计学的两个最基本的函数; 地统计学的主要方法之一的克里金法, 就是建立在变异函数理论和结构分析基础之上的, 因此, 打好数理统计基础是掌握地统计学知识的前提。

2. 融会贯通地统计分析基本理论

地统计学是一门理论性很强的学科, 具有一套较完整的理论体系, 具有完善的分析方法、实用的程序系统和数据库。因此, 在学习中一定对地统计学的每个要点都要逐步融会贯通, 如对空间变异与结构分析不够理解, 那么对空间局部估计的分析也就不会正确, 甚至空间插值出来的结果是完全错误的。

3. 用一种编程语言实现地统计分析

地理信息科学专业的学生, 必须掌握一门编程语言。C#在继承 C 和 C++强大功能的同时, 又去掉了它们的一些复杂特性, 具有简单的可视化操作、高效的运行效率、强大的操作能力、优雅的语法风格、创新的语言特性及便捷的面向组件编程支持等优势, 成为 .NET 开发的首选语言。同时, 可以很好地结合地理信息系统二次开发平台(ArcGIS Engine), 实现地统计的空间变异与结构分析、空间局部估计两大功能, 不但可以加深对地统计分析的理解与认识, 还可以激发学生对编程的热情, 使他们从中得到乐趣。

1.4.2 明确地统计学的落脚点

1. 揭示空间数据的相关规律

空间统计分析方法假设研究区中所有的值都是非独立的,在空间或时间范畴内其相互之间存在一定的相关性,即自相关性。根据空间数据之间自相关性,可以利用已知样点值对任意未知点进行预测。但事实上,空间数据之间的相关性在预测之前是未知的,因此揭示空间数据的相关规律是空间统计分析的重要任务之一,即在考虑样本点位置、方向和彼此之间距离的基础上,直接测定空间数据结构的相关性和依赖性,研究具有一定随机性和结构性的各种变量的空间分布及变异规律。

2. 利用相关规律进行未知点预测

空间统计下的另一重要任务就是利用空间数据相关规律对任意未知点进行预测。地统计学中的空间预测方法为克里金法,是建立在变异函数理论及结构分析基础上,在有限区域内对区域化变量的取值进行无偏最优估计的一种方法;在估计未知点数值时,它不仅估计样点与邻近已知样点的空间位置,还考虑了各邻近样点彼此之间的位置关系,同时利用已知观测值空间分布的结构特征,从而使这种估计比其他传统的估计方法更精确,更符合实际。

1.4.3 学习地统计学课程的捷径

1. 理论学习与软件应用相结合

地统计学是一门涉及统计、地理空间分析理论的课程,学生需要在系统学习地统计学理论的基础上,结合地统计学专用软件进行大量的上机操作练习,这样既能强化、巩固地统计理论知识,又能够使学生灵活地运用地统计分析方法,为解决实际问题奠定基础。

2. 课程学习与实践应用相结合

地统计学既是一门方法论学科,又是一门实践性很强的学科,要想真正地掌握地统计分析方法,必须在实践中运用。通过实际案例,将抽象、枯燥的理论与生动的生产实践活动相结合,以利于对地统计理论知识的理解和升华,起到举一反三、触类旁通的作用。

1.4.4 能熟练使用软件工具

随着计算机技术的发展,目前国内外已开发出一些公共领域地统计软件及模块,如GS+、ArcGIS地统计模块、GeoDa、GSLIB、Surfer、Geo-EAS、SGeMS等。

GS+(Geostatistics for the Environmental Sciences)软件由美国Gamma Design软件公司制作,是目前业界常用的专业地质统计分析软件。GS+软件涉及几乎所有的地统计分析功能,包括三维条件下数据的基本统计分析、分形分析、协方差分析、变异函数分析及普通克里金法、协同克里金法、条件模拟等地统计学常用的空间分析与估值方法。相较于其他地统计学软件,GS+的亮点之处在于能够根据输入的数据自动拟合包括高斯、椭圆和指数模型在内的实验变异函数。同时,GS+能够导入导出Surfer、ArcGIS Grid等常用的网格文件,并具有强大的绘图、输出功能,可以将计算结果直接绘图输出。

ArcGIS地统计分析模块是嵌入于美国环境系统研究所(Environmental Systems

Research Institute, ESRI)研发的 ArcGIS 软件中的一个扩展模块。此模块的优势是能够充分利用 GIS 强大的空间分析功能,实现探索性空间数据分析、空间插值分析及结果检验、高斯地统计模拟等功能,以及地统计学与 GIS 的融合。

GeoDa 是一个基于栅格数据探索性空间数据分析(ESDA)的软件工具集合体。该软件向用户提供一个友好的和图示的界面用以描述空间数据分析,如自相关性统计和异常值指示等。

GSLIB(Geostatistical Software Library)是由美国斯坦福大学应用地球科学系 C.V.Deutsch 和 A.G.Journel 开发的地统计程序工具箱。该工具箱能够完成大部分地统计学功能计算,包括变异函数拟合、克里金估计、随机模拟及相应的空间制图。

我国学者也根据研究的需要自行开发了其他应用软件,并取得了很好的效果。例如,北京科技大学侯景儒教授领导研制了地统计学方法研究程序集,能够实现二维地统计学分析(普通克里金法、对数正态克里金法、泛克里金法等)、二维非参数地统计学分析(指示克里金法)、二维多元地统计学分析(协同克里金法、因子克里金法、空域中的多元地统计学)及三维地统计学(三维普通克里金法、三维协同克里金法)。中国地质大学(武汉)王仁铎、胡光道教授领导编制的地统计学软件包,包括常用克里金插值和条件模拟。浙江大学唐启义研发的 DPS 数据处理系统,包含地统计模块,可完成变异函数估计、分析,克里金插值, IDW 插值等。中国林业科学研究院资源信息研究所开发的统计之林(ForSat),包含了遥感数据的地统计分析,在区域遥感分析中得到应用。原武警黄金指挥部黄金地质研究所的 CGES(Chinese Geology and Exploration System)系统、李裕伟教授领导研制的 KPX 固体矿产勘查评价自动化系统,均含有地统计学的相关内容。

地统计学理论上的发展、应用要远远超过其软件的发展,使得新方法的推广由于软件的局限性受到了很大的限制。现有软件对地统计分析的理论热点涉及仅限于极少数,且大多数基于 DOS 或者 UNIX 系统开发,通用性较低。此外,作为一种空间数据的分析工具,地统计学软件在高质量图像表达与可视化方面还远远不够。地理信息系统(geographic information system, GIS)是对空间数据进行搜集、存储、检索、转换、显示及分析的一门技术。作为一个强大的数据库系统,它可以存储、编辑、操作具有同样空间范围的多种专题信息。但对空间数据分布格局进行建模,抽取其特征还很欠缺。地统计学近年来在国际上发展迅猛,特别是 GIS 的发展,对空间分析功能提出了一个新的要求,使得地统计学成为多个学科关注的焦点。但目前为止,二者之间的结合还很少,或非常欠缺。例如,大型软件 ArcGIS,从 8 版本以后加入了地统计学扩展模块,但内容仅限于克里金系列方法,而对于模拟方法还是一片空白。所以,未来将两者结合起来将是一种必然的趋势。

第 2 章 数理统计基础

2.1 经典统计学方法

2.1.1 经典统计学的起源

原始社会后期,最原始的统计产生了,这个时期的统计主要是一些简单的计数活动。奴隶社会时期,奴隶主为了控制奴隶而进行一些较复杂的统计是很有必要的,这一时期随着需求的增加统计的地位也日显重要。封建社会时期,因为国家这个社会形态的存在更为普遍,统治者为了更好地维护自己的统治,需要对各方面进行统计,所以统计的地位更显重要。随着资本主义的兴起,统计已经深入社会的方方面面。

虽然统计的起源很早,但是统计学作为一门系统的学科只有 300 多年的历史。统计学的发展分为三个主要的时期,分别是萌芽期、近代期、现代期。

萌芽期(17世纪中叶~18世纪),这一时期主要有国势学派和政治算术学派两个学派。国势学派,以德国的 H.Conring 和 G.Achenwall 为代表,主要是对国家的重要事项进行记录,而忽视了量的分析,因此又称为“记述学派”。但这一学派的研究对象和研究方法都不符合统计学的方法。政治算术学派,以英国的“政治经济学之父”W.Petty 为代表,该学派并不满足于社会经济现象的数量登记、列表、汇总、记述等过程,还要求把这些统计经验运用分类、制表及各种指标来加以全面系统地总结,并从中提炼出某些理论原则,同时它第一次运用度量的方法,依靠数字解释与说明社会经济现象。

近代期(18世纪末~19世纪),这一时期主要有数理统计学派和社会统计学派。法国的 P.S.Laplace 作为数理统计学派的代表人物之一,第一个把概率论引进统计学领域。他对统计学的贡献是阐明了统计学的大数法则和进行了大样本推断的尝试。数理统计学派的另一代表人物比利时的天文学家、数学家、统计学家 A.Qutelete 实现了统计学和概率论的结合,使统计学开始进入更为丰富发展的新时期。他的主要贡献是发现了大量现象的统计规律性和开创性地应用了许多统计方法。他把统计学发展中的三个主要源泉:国势学派、政治算术学派、古典概率学派加以统一、改造并融合成具有近代意义的统计学。可以说, A.Qutelete 是古典统计学的完成者,也是近代统计的先驱者,又是数理统计学派的奠基人,在统计发展史上具有承上启下、继往开来的地位。19世纪后半叶,由德国大学教授 K.G.A.Knies 首创,并以 C.L.E.Engel 等为代表,兴起了社会统计学派。该学派认为,统计学的研究对象是社会现象,研究目的在于明确社会现象内部的联系和相互关系;统计应包括资料的收集、整理及其分析研究。同时,本学派还认为包含人口普查、工业普查、农业普查等的全面调查在社会统计中处于重要地位。

现代期(20世纪初至今)。19世纪末~20世纪初的统计学主要是描述统计学,20世纪30年代统计学逐渐进入现代统计学的范畴,20世纪60年代以后统计学逐步发展成现代意义上的统计学。统计学的发展主要有三个明显的趋势:①随着数学的不断发展和完