

空间统计学 林业应用案例

闫秀婧 汪浩然 著



清华大学出版社

空间统计学

林业应用案例

闫秀婧 汪浩然 著

清华大学出版社

北 京

内 容 简 介

本书第一部分较为系统地介绍了空间统计学的区域化变量理论、变异函数及其结构分析、克里格插值、缓冲区分析、叠加分析、网络分析等方面的原理和应用方法；第二部分介绍了空间统计学在林业中的应用情况，结合作者的研究成果，综合性地从城市绿地植被配置、土壤异质性、经济林种植适宜区区划等方面给出空间统计学在林业中的具体应用。这些案例均来自作者对近年来研究工作的总结，具有示范作用，实用性和可操作性强。通过应用案例，读者可以更好地掌握空间统计学理论知识，更好地把握其在实际中的应用。

本书同时为广大林业研究者和林业基层工作者提供了理论基础和技术指导，也可供从事林业、空间分析相关工作的人员作为参考书使用。

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

图书在版编目(CIP)数据

空间统计学林业应用案例 / 闫秀婧，汪浩然 著. —北京：清华大学出版社，2014

ISBN 978-7-302-38142-6

I. ①空… II. ①闫… ②汪… III. ①空间科学—统计学—应用—林业 IV. ①S711

中国版本图书馆 CIP 数据核字(2014)第 227848 号

责任编辑：施 猛 易银荣

封面设计：常雪影

版式设计：方加青

责任校对：曹 阳

责任印制：沈 露

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>, <http://www.wqbook.com>

地 址：北京清华大学学研大厦 A 座 邮 编：100084

社 总 机：010-62770175

邮 购：010-62786544

投稿与读者服务：010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈：010-62772015, zhiliang@tup.tsinghua.edu.cn

课 件 下 载：<http://www.tup.com.cn>, 010-62796865

印 装 者：北京密云胶印厂

经 销：全国新华书店

开 本：185mm×260mm

印 张：13.75 字 数：277 千字

版 次：2014 年 10 月第 1 版

印 次：2014 年 10 月第 1 次印刷

印 数：1~1000

定 价：48.00 元

前 言

林业是生态建设的主体，是经济社会可持续发展的重点所在。进入21世纪，林业经营模式发生了根本性的变化，从过去单一、粗放的经营管理模式转向应用现代高新技术对林业生产、经营管理、战略决策过程中的自然、经济和社会信息进行收集、存贮、传递、处理、分析和利用的精准化、系统化模式。然而，森林因为受物理环境、自然干扰、人为干扰、树种特性、树种对干扰的反应等因素及其相互作用的影响，使得数据具有更强的分散性、复杂性和综合性，传统方法和技术已经不能满足空间自相关性、空间变异性、空间分布等方面的研究需要，因此，必须通过空间统计学去解决这些复杂的问题。

空间统计学(地统计学)是应用数理统计学的一个分支。与传统的统计学相比，能够对研究对象的空间格局进行检验、模拟和估计。林业研究中的生物或非生物因子的时空依赖性普遍存在的，尤其是现代生态学所揭示的不同空间尺度上生物和非生物因子分布的异质性、土壤养分和土壤环境异质性等充分证实了这一规律。但是，在生态学和土壤因子空间过程研究中，针对如何定量地测度空间尺度和更精确地绘制这些自然因子的空间分布图，以往的方法存在着许多的缺陷，尤其是没有一个有关可信程度和误差的评价，而地统计学则能够解决这一问题。由此可推知，空间统计学将成为林业研究领域的一种新工具。但空间统计学在国内林业中的应用才刚刚起步，绝大多数林业工作者或科研人员对空间统计缺乏基本了解，基于这一情况，我们编写了本书。《空间统计学林业应用案例》介绍了空间统计学的基本理论、原理和方法，列举了空间统计学在生态环境分析、植被配置模式选择、经济林土壤变异分析、经济林种植适宜区评价等方面的综合应用，具有较强的实用性和可操作性，旨在为广大林业研究者和基层工作者提供理论基础和技术指导。

本书第一部分较为系统地介绍了空间统计学的区域化变量理论、变异函数及其结构分析、克里格插值、缓冲区分析、叠加分析、网络分析等方面的原理和应用方法；第二部分介绍了空间统计学在林业中的应用情况，结合作者的研究成果，综合性地从城市绿地植被配置、土壤异质性、经济林种植适宜区区划等方面给出空间统计学在林业中的具

体应用，这些案例均是作者近年来研究工作的总结，具有示范性。这些应用案例可以帮助读者更好地掌握空间统计学理论知识，也可以帮助读者把握其在实际生活中的应用。

本书由闫秀婧教授负责总体设计、组织、审稿和定稿工作。第一篇及第二篇第一章内容由闫秀婧教授撰写；第二篇第二章、第三章内容由汪浩然副教授撰写。此书在编写过程中，得到了甘肃林业职业技术学院姚德生教授、武戌良副教授、吴霞副教授、陈西仓副教授、李树金副教授、刘岚实验师、王瑞君实验师，甘肃省林业科学技术推广总站辛永清副教授、王集桂工程师、孙力工程师，甘肃省天水市麦积区林业区史永国高级工程师、杨军平工程师，甘肃省小陇山林业实验局张宋智正高级工程师，东芬兰大学侯正阳博士、徐晴博士，甘肃省林业规划设计研究院王顺彦高级工程师等多位专家的大力支持和帮助，在此表示衷心感谢！

本书由甘肃省自然基金项目(基于生态效应城市绿地植被配置模式挖掘研究，项目编号：1208RJZE259)、甘肃省财政项目(城市绿化生态环境质量优化评价研究，项目编号：2060203)、甘肃省林业科学技术推广总站项目(麦积区经济林种植土壤适宜性空间数据平台建立与应用研究，项目编号：20110423)资助出版。

因作者水平有限，书中难免存在不妥和疏漏之处，敬请读者批评指正。反馈邮箱：wkservice@vip.163.com。

编 者
2014年7月于天水

目 录

第一篇 空间统计分析基础理论

第一章 空间半变异分析	3
第一节 空间信息特征和表示	3
一、林业空间信息特征	3
二、空间信息表示	3
第二节 地统计基本原理	6
一、前提假设	6
二、区域化变量	7
第三节 变异分析	7
一、常用变异函数	7
二、变异模型确立	13
第二章 空间插值	17
第一节 空间插值概述	17
第二节 精确性插值	18
一、反距离加权插值方法(IDW)	18
二、径向基插值方法	20
第三节 克里格插值	21
一、普通克里格模型	21
二、简单克里格插值模型	22
三、泛克里格模型	22
四、指示克里格模型	22
五、析取克里格模型	22
六、协同克里格模型	22
第四节 插值交叉验证	22

一、常见的交叉验证形式	23
二、空间插值交叉验证	23
第五节 空间预测面的建立	24
一、数据检验	24
二、空间插值	25
三、交叉验证	25
第六节 重分类	26
第三章 缓冲区分析	28
第一节 缓冲区类型	28
一、单重缓冲区	28
二、多重缓冲区	28
第二节 缓冲区分析基本算法	29
一、算法的基本定义	29
二、算法的实现	30
第三节 缓冲区建立	32
一、点缓冲区建立	32
二、线、面缓冲区建立	33
第四章 叠加分析	35
第一节 叠加分析分类	35
一、视觉信息叠加	35
二、点与多边形叠加	35
三、线与多边形叠加	36
四、多边形叠加	36
五、栅格图层叠加	36
第二节 叠加分析算法	36
一、布尔逻辑运算	36
二、叠加分析算子	37
第三节 叠加分析实现	44
一、裁剪	44
二、合并	45
三、擦除	46
四、求交	46
五、同一	47
六、对称差	47

七、更新	48
第五章 网络分析	49
第一节 网络的组成	49
一、线状要素——链	49
二、点状要素	49
第二节 网络分析的基本功能和原理	50
一、基本功能	50
二、基本原理	51
第三节 网络分析实现	52
一、流向分析实现	53
二、追踪分析实现	53

第二篇 空间统计在林业中的应用实例

第一章 空间统计在林业中的应用综述	56
第一节 空间统计分析研究进展	56
第二节 空间统计分析在林业中的应用	59
一、应用的必要性	59
二、空间统计分析在林业应用中的研究	59
第二章 空间统计分析在城市绿地植被配置模式上的应用	63
第一节 研究的目的和意义	63
第二节 研究的现状	64
一、城市绿色空间的内涵	64
二、城市绿色空间的分类体系	65
三、空间统计学在生态分析中的应用	66
四、小尺度植物群落结构及其动态研究	66
五、城市绿色空间生态环境效应研究	66
六、绿色空间在调控城市生态系统平衡中的重要性	67
七、城市绿地植被配置模式挖掘研究	68
第三节 应用实例	69
一、生态服务和生态环境质量数据获取与处理	70
二、遥感数据处理	72
三、绿地生态服务和生态环境质量空间优化组合研究	74
四、绿地生态服务和生态环境质量优化组合分析	109

五、绿地植被配置模式挖掘	111
六、结论	117
第三章 空间统计分析在经济林土壤异质性研究上的应用	118
第一节 研究的目的是和意义	118
第二节 研究的现状	119
一、空间统计分析在土壤物理特性空间变异上的应用	119
二、空间统计分析在土壤化学性质空间变异上的应用	120
三、土壤盐分空间变异研究	121
四、土壤重金属空间变异的研究	122
第三节 应用实例	123
一、经济林种植现状数据获取与处理	123
二、经济林土壤养分与土壤环境空间变异分析及预测	169
三、经济林种植适宜性区划	179
四、结论	182

附 录

附录A 绿地生态因子实测数据一览表	185
附录B 经济林土壤养分和土壤环境实测数据一览表	194
参考文献	207

空间统计分析基础理论

空间统计分析,即地统计学(Geostatistics),于20世纪50年代初开始形成,是20世纪60年代在法国著名统计学家Matheron的大量理论研究工作基础上形成的一门新的统计学分支。地统计学是以区域化变量理论(Theory of Regionalized Variable)为基础,以变异函数(Variogram)为基本工具来研究分布于空间并呈现出一定的随机性和结构性的自然现象的科学(冯益明等,2004)。由于空间现象之间存在不同方向、不同距离成分等的相互作用,因此,需要通过空间统计方法进行分析。

空间分析是基于地理对象位置和形态特征的空间数据分析技术,其目的在于提取和传输空间信息。空间信息可以分为空间位置、空间分布、空间形态、空间距离、空间方位。空间位置是借助空间坐标系来传递的空间物体的个体定位信息;空间分布反映了同类空间事物的群体定位信息;空间形态是空间物体的几何特征,一些特征易被视觉感受,如走向、连通性等,另一些特征则必须用数值来描述,如面积、周长、坡度;空间距离反映的是空间物体之间几何上的接近程度;空间方位是用方位描述两个物体之间的位置关系。空间统计分析由空间变异与结构的半变异函数和空间局部估计的克里格插值法组成,是GIS空间分析的一个重要技术手段。

空间分析的发展及在各行业中的应用始于20世纪60年代地理和区域科学的计量革命。开始是应用定量分析手段分析点、线、面的空间分布模式,后来则更多地强调地理空间本身的特征、空间决策过程和复杂空间系统的时空演化过程。空间分析的模型和方法没有形成统一的体系结构,对空间分析的基本内容的界定没有得到广泛认同。空间分析主要包括以下几方面:①基于图的分析,主要有缓冲区分析、叠置分析、复合分析、邻近分析及空间联结,这些方法以图形操作为主,在现有的GIS软件中已比较成熟。②基于数据的分析,这部分的理论基础是空间统计学。空间统计学基于两个本质假设:任意距离的两点间的差值的数学期望为零;任意距离的两点间的差值的方差最小。这里的任意距离指的是空间物体间的距离,由于空间物体又可分为点、线、面、体4类,而点与点之间的距离是最本质、最简单的距离形式,因此,空间统计学在这两个本质假设基础上,通过对样本点得出的经验变率函数的拟合得到真实的变率函数,以便作进一步分析。③基于事件机理的分析,有些事件有其相应的机理提供公式构架,由环境信息提供初始边界条件,这样可直接利用前人总结的成果来描述事件。空间分析的目的是提取空间信息,而空间分析的对象是空间数据。空间分析是在空间数据的基础上进行的,空间数据是描述地球表层(有一

定厚度)一定范围内的地理事物及其关系的数据, 可以是关于空间物体及其空间变量和属性的数据。以空间物体为定义域, 随空间物体的延展而变化的地理现象是空间变量; 相反, 不随空间物体的延展而变化的地理现象是空间物体的属性。一般认为, 空间数据包括三大要素: 几何、属性和时间, 从空间分析的角度出发, 空间变量也是空间数据的要素之一, 空间分析中的很多内容是针对空间变量的。

空间统计学以区域化变量理论为基础, 借助变异函数, 可进行空间数据的结构性和随机性空间相关性和依赖性空间分布与变异有关的研究, 并对这些数据进行最优无偏内插估计。与传统统计学相比, 地统计学更注重随机变量的空间过程, 通过研究对象在空间上不同间隔的抽样点上的差异, 定量描述其空间变化规律。空间统计学主要通过半方差图描述研究对象的空间相关性的方法; 通过空间上抽样点的调查数据, 应用空间局部内插理论, 对空间上未测点进行估计 (Robertson, 1987)。

空间分析除在找矿勘探、矿体圈定、储量计算、采矿设计、矿山生产及地学科研等方面具有明显的优越性外, 在石油地质、生物学、生态学、岩石学、地球化学、地震地质、海洋地质、农业、水文、古气候、古地理、气象学、遥感地质、环境、林业、医学等许多方面都有成功应用的实例(肖斌等, 2000)。因此, 地统计学在不到50年的研究和实践中, 已发展成为研究自然界具有随机性和规律性双重特征变量的具有普遍性的科学方法。

本部分将介绍空间统计分析中半变异分析、空间插值、缓冲区分析、叠加分析、网络分析的基本原理和方法, 力求使读者对空间统计分析方法有一个全面的认识和了解。

第一章 空间半变异分析

空间数据是地理空间抽象的数字描述和离散表达。它具有客观存在性与抽象性、概括性与多态性、可存储性与可传输性、可度量性与近似性、可转换性与可扩充性、商品性与共享性等数据的一般特征(刘湘南等, 2005)。空间信息作为地理信息的载体, 具有时空特征、多维结构、多尺度性、不确定性和海量性的特征。它是多源、多维、多尺度、多分辨率和多时态的, 它反映了地球复杂巨大系统和地球上各种自然和社会的非线性特征(王新洲, 2003)。空间信息的来源有航测、实测、遥感影像提取和空间插值预测等多种形式, 将这些数据加载到ARCGIS和GS(Geostatistics for the Environmental Sciences)中, 通过半变异分析可以确立空间因子的最大影响矢距和半变异函数, 进行空间现象的预测和统计, 从而为行业提供空间决策依据。

第一节 空间信息特征和表示

一、林业空间信息特征

地统计学以区域化变量理论为基础, 它的应用已被扩展到分析各种自然现象的空间异质性和空间格局。而森林由于受物理环境、自然干扰、人为干扰、树种特性、树种对干扰的反应、土壤等因素及其相互作用的影响, 使得它本身具有高度空间异质性和时间异质性, 又因这些因子的测量方法不同、来源不同、载体也不同, 从而使得测量结果具有多维性。传统的数理统计方法已无法解决, 必须通过空间统计分析方法来实现。

二、空间信息表示

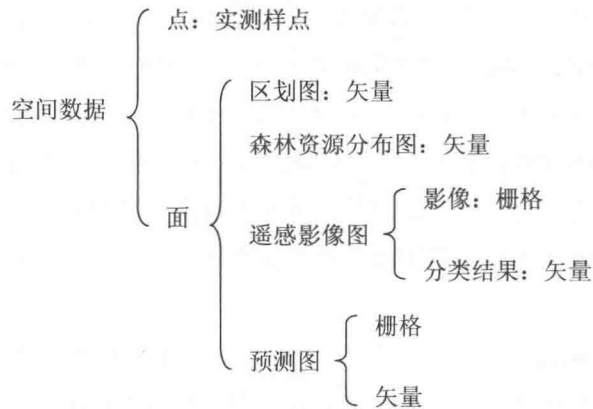


图1-1-1 林业调查中空间数据分类

空间分析是基于地理对象的位置和形态特征的空间数据分析技术。空间分析方法必须受到空间数据表示形式的影响和制约。因此,在进行空间分析时,必须考虑空间数据的表示方法和数据模型。

空间数据表示的基本任务是将以图形模拟的空间物体表示成计算机能够接受的数字形式,如林业上的实测数据、遥感影像数据、遥感影像提取的分类数据等,其表达有栅格形式、矢量形式或属性形式。按照空间维数又划分成两种类型:点数据和面数据,详细分类如图1-1-1所示。实测数据主要记录采样点的相关属性,在ARCGIS中根据地理坐标加载到系统中,再将该属性值导出为点的矢量数据,生成一个点层。面层主要有遥感影像图、区划图、通过遥感影像提取的森林资源分布图、预测图层,其中通过遥感影像提取的森林资源分布图有矢量和栅格两种格式:区划图为矢量形式,预测的图层有栅格和矢量两种。栅格数据模型和矢量数据模型是描述地理现象最常见、最通用的数据类型。

(一) 栅格数据类型

在栅格模型中,地理空间被划分成像元,空间位置由像元的行列号表示。像元的大小反映数据的分辨率,空间物体由若干像元隐含描述。栅格数据的设计思想是将地理空间看成一个连续的整体,在这个空间中处处有定义。

栅格结构以规则阵列表示空间地物或现象分布的数据组织,组织中的每个数据表示地物或现象的非几何属性特征。栅格模型采用面域或空域枚举来直接描述空间目标对象。在栅格模型中,点(点状符号)是由一个或多个像元表示,线是由一串彼此相连的像元构成。在栅格模型中,每一像元的大小是一致的(一般是正方形),而且每一个栅格像元层记录着不同的属性(如植被类型等)。像元的位置由纵横坐标(行列)决定。所以,每个像元的空间坐标不一定要直接记录,因为像元记录的顺序已经隐含了空间坐标。

栅格模型具有如下几个特点:①栅格的空间分辨率指一个像元在地面所代表的实际面积大小(一个正方形的面积)。②对于同一幅图形或图像来说,随着分辨率的增大,存储空间也随之增大。例如,如果每一像元占用一个字节,而且分辨率为100m,那么,一个面积为 $10\text{km}\times 10\text{km}=100\text{km}^2$ 的区域就有 $1000\times 1000=1\,000\,000$ 个像元,所占存储空间为1 000 000个字节;如果分辨率为10m,那么,同样面积的区域就有 $10\,000\times 10\,000=100\,000\,000$ 个像元,所占存储空间近100MB。③表达空间目标、计算空间实体相关参数的精度与分辨率密切相关,分辨率越高,精度越高。④非常适合进行空间分析。例如,同一地区多幅遥感图像的叠加操作等。⑤不适合进行比例尺变化、投影变换等。

总之,栅格数据属性明显,定位隐含,即定位是根据数据在数据集中的位置得到的,所表示实体的位置很容易隐含在网格文件的存储结构中。栅格结构表示的地表是不

连续的,是量化和近似离散的数据。在栅格结构中,地表被分成相互邻接、规则排列的矩形方块,每个方块与一个栅格单元相对应。栅格数据的比例尺就是栅格大小与地表相应单元大小之比。在对许多栅格数据进行处理时,常假设栅格所表示的量化表面是连续的,以便使用某些连续函数。由于栅格结构对地表的量化,在计算面积、长度、距离、形状等空间指标时,若栅格尺寸较大,则会造成较大的误差,同时,由于在一个栅格的地表范围内,可能存在多于一种的地物,而表示在相应的栅格结构中常只能是一个代码。

(二) 矢量数据类型

矢量模型是利用边界或表面来表达空间目标对象的面或体要素,通过记录目标的边界,同时采用标识符(Identifier)表达它的属性来描述空间对象实体。矢量模型能够方便地进行比例尺变换、投影变换以及图形的输入和输出。矢量模型处理的空间图形实体是点(Point)、线(Line)、面(Area)。

点是空间上不可再分的地理实体,可以是具体的,也可以是抽象的,如地物点、文本位置点或线段网络的结点等。如果点是一个与其他信息无关的符号,则记录时应包括符号类型、大小、方向等有关信息;如果点是文本实体,记录的数据应该包括字符大小、字体、排列方式、比例、方向以及与其他图形属性的联系方式等信息。

线实体用其中心轴线上的抽样点坐标串表示其位置和形状;线实体可以定义为直线元素组成的各种线性要素,直线元素由两对以上的 (x, y) 坐标定义。

面实体用范围轮廓线上的抽样点坐标串表示其位置和范围,多边形面数据是描述地理空间信息的最重要的一类数据。在区域实体中,用具有名称属性和分类属性的多类多边形表示,如行政区、植被分布、土地类型;具有标量属性的有时用等值线描述,如地形、降雨量等,是由线形成的闭合多边形。

在矢量模型中,拓扑关系是进行空间分析的关键。在GIS的拓扑数据模型中,与点、线、面对应的空间图形实体主要有结点(Node)、弧段(Arc)、多边形(Polygon),多边形的边界被分割成一系列的弧和结点,结点、弧、多边形间的空间关系在数据结构或属性表中加以定义。GIS的矢量数据模型具有如下特点:①通过对结点、弧、多边形拓扑关系的描述,相邻弧段的公用结点,相邻多边形的公用弧段在计算机中只需记录一次。②空间图形实体的拓扑关系,如拓扑邻接、拓扑关联、拓扑包含不会随着诸如移动、缩放、旋转等变换而变化,而空间坐标及一些几何属性(如面积、周长、方向等)会受到影响。③一般情况下,通过矢量模型所表达的空间图形实体数据文件占用的存储空间比栅格模型小。④能够精确地表达图形目标,精确地计算空间目标的参数(如周长、面积)。

(三) 栅格数据与矢量的区别

栅格数据与矢量数据最大的区别是,前者用元子空间充填集合表示,后者用点串序

列表达边界形状及分布。因此,栅格数据面向空间的数据结构在布尔运算、整体操作特征计算及空间检索方面有着明显的优势;而矢量数据面向目标的数据结构则很容易实现模型生成、目标显示及几何变换。

第二节 地统计基本原理

地统计学又称空间统计学,是20世纪60年代法国著名统计学家Matheron G.在进行大量理论研究的基础上提出并逐渐形成的一门新的统计学分支。它以区域化变量理论为基础,借助变异函数,可进行与空间数据的结构性和随机性、空间相关性和依赖性、空间分布与变异有关的研究,并对这些数据进行最优无偏内插估计。空间统计分析的目的在于,找出某种属性分布的整体特征和趋势,了解其中的规律,以便科学地对其进行分析和预测。它有两大研究领域:区域化变量的空间相关及变异分析、空间插值(张仁铎, 2005, Bongarcon D Francois, 2004)。

地统计分析理论基础包括前提假设、区域化变量、变异分析和空间插值。

一、前提假设

(一) 随机过程

地统计学与传统统计学的相同点:都是在大量样本的基础上,通过分析样本间的规律,探索其分布规律并进行预测。地统计学认为研究区域中的所有样本都是随机过程的结果,即所有样本值都不是相互独立的,是遵循一定的内在规律的。地统计学就是要揭示这种内在的规律并进行预测。

(二) 正态分布

地统计学和传统统计学一样,假设样本是正态分布的。在获得数据后,首先要对数据进行分析,若不符合正态分布假设,就需对数据进行变换,转为符合正态分布的形式,并尽量选取可逆的变换形式。

(三) 平稳性

对于统计学而言,重复是其理论基础,即从大量重复的观察中可以进行预测和估计,并可以了解估计的变化性和不确定性。对于大部分空间数据而言,平稳性的假设是合理的。平稳性有两类:一类是均值平稳,即假设均值是不变的并且与位置无关;另一类是与协方差函数有关的二阶平稳和与半变异函数有关的内蕴平稳。二阶平稳是假设具有相同的距离和方向的任意两点的协方差是相同的,协方差只与这两点的方差(即变异

函数)是相同的。二阶平稳和内蕴平稳都是为了获得基本重复规律而作的基本假设,通过协方差函数和变异函数可以进行预测并估计预测结果的不确定性。

二、区域化变量

当一个变量呈现一定的空间分布时,称之为区域化变量,它反映了区域内的某种特征或现象。区域化变量与一般的随机变量的不同之处在于,一般的随机变量取值符合一定的概率分布,而区域化变量根据区域内位置的不同而取不同的值。而当区域化变量在区域内确定位置时,表现为一般的随机变量,即它是与位置有关的随机变量。在实际分析中,常采用抽样的方式获得区域化变量在某个区域内的值,即此时区域化变量表现为空间点函数

$$Z(x)=Z(x_u, x_v, x_w) \quad (1-1-1)$$

根据其定义,区域化变量有两个显著特征:随机性和结构性。首先,区域化变量是一个随机变量,它具有局部的、随机的、异常的特征;其次,区域化变量具有一定的结构特点,即变量在点 x 与偏离空间的距离为 h 的点 $x+h$ 处的值 $Z(x)$ 和 $Z(x+h)$ 具有某种程度上的相似性,即自相关性,这种自相关性的程度依赖于两点间的距离 h 及变量特征。除此之外,区域化变量还具有空间局限性、不同程度的连续性和不同程度的各向异性(即各个方向表现出的自相关性有所区别)等特征。



第三节 变异分析

一、常用变异函数

(一) 确定曲线类型

1. 变异函数

变异函数或变差函数是空间统计学的基本理论。在一维条件下,当空间点 x 在一维 x 轴上变化时,区域变量 $Z(x)$ 在点 x 和 $x+h$ 处的值 $Z(x)$ 与 $Z(x+h)$ 差的方差的一半定义为区域变量 $Z(x)$ 在 x 轴上的变异函数,记为 $r(x, h)$,即

$$r(x, h) = \frac{1}{2} E[Z(x) - Z(x+h)]^2 - \frac{1}{2} \{E[Z(x)] - E[Z(x+h)]\}^2 \quad (1-1-2)$$

在二阶平衡假设条件下对任意 h 有

$$E[Z(x)] = E[Z(x+h)] \quad (1-1-3)$$

因此1-1-2式可以改写为

$$r(x, h) = \frac{1}{2} E[Z(x) - Z(x+h)]^2 \quad (1-1-4)$$

从1-1-4式可知，变异函数依赖于 x 和 h ，当变异函数仅依赖于 h ，与 x 无关时，变异函数 $r(x, h)$ 可改写成 $r(h)$ ，即

$$r(x, h) = \frac{1}{2} E[Z(x) - Z(x+h)]^2 \tag{1-1-5}$$

2. 内蕴假设

有些自然现象和随机函数，它们具有无限离散性，即无协方差，但有半变异函数，这时区域化变量 $Z(x)$ 的增量 $Z(x) - Z(x+h)$ 满足下列两个条件时，就称该区域化变量满足内蕴假设。

(1) 在整个研究区内，随机函数 $Z(x)$ 的增量 $Z(x) - Z(x+h)$ 的数学期望为0，即

$$E[Z(x)] - E[Z(x+h)] = 0 \quad \forall x, \forall h \tag{1-1-6}$$

(2) 对于所有矢量的增量 $Z(x) - Z(x+h)$ 的方差函数存在且平稳

$$Var[Z(x) - Z(x+h)] = E[Z(x) - Z(x+h)]^2 = 2r(x, h) = 2r(h) \quad \forall x, \forall h \tag{1-1-7}$$

即要求 $Z(x)$ 的半变异函数 $r(h)$ 存在且平稳。

在内蕴假设中，随机函数 $Z(x)$ 的增量 $Z(x) - Z(x+h)$ 只依赖于分隔它们的向量 h (模和方向)，而不依赖于具体位置 x ，这样，被向量 h 分割的每一对数据 $[Z(x), Z(x+h)]$ 可以看成是一对随机变量 $[Z(x), Z(x+h)]$ 的一个不同实现，而半变异函数 $r(h)$ 的估量为

$$r(h) = \frac{1}{2N(h)} \sum_i^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \tag{1-1-8}$$

式中： h 为样本距； $N(h)$ 是分隔间距为 h 的样本对的总个数； $Z(x_i)$ 是样本空间样本 x_i 的属性值； $Z(x_i + h)$ 是距样本 x_i 距离为 h 处的样本的属性值。半变异函数是假设 $Z(x)$ 为区域化变量且满足平衡条件和本征假设的前提下定义的，半变异函数大时，空间相关性减弱。

以 h 为横坐标， $r(h)$ 为纵坐标，绘制半变异函数曲线图(见图1-1-2)，该图可以直观地展示区域化变量 $Z(x)$ 的空间变异性。通过半变异函数计算可得到4个重要的参数，即块金值(Nugget)、基台值(Sill)、偏基台值(Partial Sill)、变程(Range)，其含义如下。

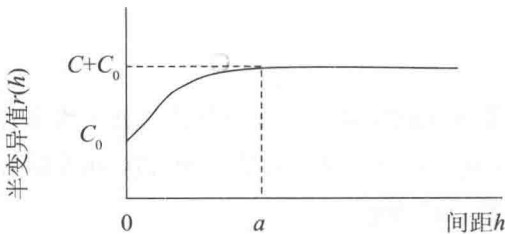


图1-1-2 半变异函数曲线图

块金值 (Nugget)：当采样点的距离为0时，半变异函数值应为0；但由于存在测量误差和空间变异，使得两样点非常接近时，半变异函数值不为0，即存在块金值。测量误差是仪器内误差引起的，空间变异是自然现象在一定空间范围内的变化，它们任意一方或两方共同作用产生了块金值。