

生物地理统计学

——生物种群时空分析的方法及其应用

周国法 徐汝梅 著

内 容 简 介

本书提出了生物地理统计学的概念、方法和理论框架。从种群内部以及种群与环境间的时空相关性原理出发,指出了生物统计学方法的缺陷,提出了研究种群时空动态的方差图和相关图方法及时空预测的模型。重点研究了种群空间格局的时空相关形成过程和空间格局的时空相关分析方法、种群空间分布的估计和预测的各种时空模型、扩散的时空检验方法和非线性扩散模型、传统生物统计学的改进和空间最优抽样设计方法。

本书可供农业、林业、环境、景观、海洋以及相关的生物学、生态学研究人员及教师参考,也可作为大学相关专业的本科生、研究生教材。

图书在版编目(CIP)数据

生物地理统计学——生物种群时空分析的方法及其应用/周国法,徐汝梅著.
-北京:科学出版社,1997

ISBN 7-03-006239-6

I. 生… II. ①周… ②徐… III. 生物统计;生物地理学 IV. Q·332

中国版本图书馆 CIP 数据核字(97)第 19858 号

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1998年6月第一版 开本:787×1092 1/16

1998年6月第一次印刷 印张:10 3/4

印数:1—2 000 字数:243 000

定价:24.00元

前 言

生态学的主要目的是研究生物学现象的生物学特征及其时序变化。生态学现象一般是发生在一定区域的与生物生存环境密切相关的时空过程。显然,生态学空间现象的不同部分间、生物与生存环境间都存在相互影响,因此要研究空间发生的生态学或生物学现象,就不可避免的要研究空间相互作用;而同样显然的是这种空间相互作用与组分间的空间相对位置有关,不同方向、不同距离的组分间的相互作用是不同的;正是这种空间相互作用的结果才形成了各种不同的空间格局。同时,空间相互作用随时间的变化构成了生态学和生物学现象的动态。

研究与位置有关的空间相互作用,显然不能假设不同位置的样本间是相互独立的,因此传统的生物统计学方法在分析空间相互作用时就行不通了。实际上,生物统计学根本不考虑空间相关性。研究空间相互作用的地质统计学(geostatistics)是60年代发展起来的方法,主要用于研究地质学现象的空间结构和进行空间估值。80年代后期,许多生态学研究人员试图用地质统计学方法研究生物学和生态学空间现象,并取得了一些研究结果。遗憾的是,生态学家们只是简单地套用地质统计学的方法,没有仔细研究那些方法是否适合生态学现象的研究。事实上,生态学现象远比地质地矿学现象复杂,多数地质学现象在较短甚至在很长的时间内都是不发生明显变化的,而生态学现象即使在短时间内也会发生较大的变化(例如动物种群的空间动态),而地质统计学方法根本不涉及时间动态只研究空间静态现象。

与时间相关的是时间序列方法,它是生态学家们常用的预测种群数量随时间变化的方法。时间序列的基本假设是认为种群数量在时间上是相关的,用数学表达为: $X_t = f(X_{t-1}, X_{t-2}, \dots, X_{t-p})$,即 t 时的种群数量 X_t 依赖于 t 之前的 p 个时期 $t-1, \dots, t-p$ 的数量。显然对于空间小区域分布的种群,使用时间序列方法预测其数量变化是合理的。而更多的情况是要预测大范围内种群数量的变化,如一个县甚至一个省,如果把这种大范围看作整体,预测其平均数量或总量显然太粗糙了。可以考虑把大范围分为若干小区域,分别预测每个小区域内的种群数量。由于相邻小区域间存在相互影响即空间相关性,因此统计学家研究了多维时间序列方法,但是多维时间序列方法至少存在以下问题:①多维时间序列考虑的相关只是数量问题,它采用的是调查数据的协方差矩阵,而没有考虑与空间距离有关的相关性;②实践中,预测区域总是分解为数千甚至数万的子区域,虽然理论上维数并不影响参数估计的理论结果,但是数万维的时间序列在协方差推断和参数估计乃至最后的预测方面在计算上却是不可取的,甚至是不可能的;③多维时间序列的参数估计的不唯一性是它的另一个致命弱点。因此必须考虑其他的时空预测途径。

为了克服生物统计学、时间序列分析和地质统计学的缺陷,或者说在吸取生物统计学、时间序列分析和地质统计学优点的基础上,徐汝梅和周国法(1994)提出了生物地理统计学的概念和理论框架。

生物地理统计学是在时空相关假设基础上,以空间信息和地理信息为背景,研究生物学和生态学现象的空间相互作用及其时序变化的统计学方法。它的基本假设是认为任何空间发生的生物学或生态学现象的组分(component)间都存在空间上的自相关性,不同的空间现象间存在空间上的互相关性,①这种空间相关性随距离的增大而减弱,同时②不同方向上的相关强度及相关特征可以是不同的,而且③空间相关性随时间的变化而变化。空间相关性假设从根本上改变了传统生物统计学及生态学方法以样本都是随机选取为基础的理论体系,可以说是生物统计学的突破性发展。生物地理统计学主要研究传统生物统计学、时间序列分析方法无法研究的空间问题和地质统计学无法研究的空间现象的时间动态问题,研究的对象可以是个体、个体群、种群和群落水平的,甚至也可以是景观、区域和全球水平的问题。

概念提出了,但不等于问题解决了。实际上,生物地理统计学只是刚刚开始。首先是生物地理统计学的理论远没有形成更谈不上完善,它的方法还很有限,应用的领域亟待开发,能解决什么样的问题或适应的范围也有待研究。

本书的主要目的是回答下面的问题:①如何量化时空相关性的特征即距离、方向变化规律?②如何应用空间相关分析方法研究传统的生物统计学和生态学问题?③如何利用时空分析的结果进行种群空间分布的估计与预测?④在空间相关条件下如何使用传统统计分析方法?⑤如何进行空间抽样设计?这些都是生态学家们关心而用传统的方法不能解决的问题。

实际上,我们在10年前已经开始研究空间分析方法了。在对昆虫种群(主要是马尾松毛虫和温室白粉虱)的空间格局及其动态的时空分析中,我们逐渐发现了传统生物统计学、地质统计学、空间统计学和时间序列分析的许多不足之处,尤其是在研究大范围如几个县甚至一个省的种群动态问题时,传统方法存在无法克服的缺陷,几乎无法完成大范围的预测问题。为此,我们和合作者们共同研究,提出和发展了许多利用地理信息和空间相关分析研究种群空间格局和种群动态的新方法,这些研究直接导致了生物地理统计学概念和内容框架的提出。

作者特别感谢美国农业部林务局(USDA Forest Service)的 Andrew M. Liebhold 博士,他给作者之一提供了进修的机会,并提供了许多美国农业部林务局的有关调查数据,书中的许多方法的研究是和 Liebhold 博士的帮助分不开的;其次应该感谢中国林业科学研究院的李天生教授,作者之一在时空分析方面和马尾松毛虫的种群动态研究是在李教授的指导下完成的,李教授无私地提供了她自己多年积累的调查数据。最后应该感谢蒋南青博士,她对本书的许多方法提出了宝贵的意见和建议;徐岩、盖宇、沈静三位博士生和黄英硕士生的许多建议对书中方法的完整性大有裨益,在此一并致谢。

当然,生物地理统计学还在初创阶段,它的理论有待完善,方法有待充实提高,所有这些都需生态学研究人员的进一步努力,本书只起抛砖引玉的作用,可以相信生物地理统计学不久就会成为生态学家关注的焦点。

目 录

前言

第一章 生物统计学与空间相关	1
1.1 统计模型与空间相关	1
1.2 传统生物统计学方法的局限性	3
1.3 本书的基本内容	10
第二章 空间相关性与空间格局	12
2.1 空间现象与空间数据	12
2.2 空间数据的一般性分析	13
2.3 空间相关分析的方法	17
2.4 空间自相关显著性的检验	27
2.5 空间分布格局的形成	29
2.6 空间格局的检验	34
2.7 时空相关性分析	40
第三章 空间密度分布的估计	42
3.1 密度估计的一般概念	42
3.2 距离方法及最优化	44
3.3 一般克立格方法	47
3.4 方差图的模型与建模	50
3.5 实例研究	53
3.6 方差函数模型对 OK 结果的影响	57
3.7 空间自回归估计	61
3.8 趋势面分析	63
3.9 估计优劣的交叉检验	65
第四章 时空预测模型	66
4.1 时间序列与时空预测	66
4.2 空间转移概率方法	66
4.3 空间-时间序列预测模型	76
4.4 三维自回归趋势面	78
4.5 预测克立格方法	80
4.6 距离预测方法	81
4.7 广义时空回归模型	84
第五章 空间分析中的方法	89
5.1 空间相关分析的计算方法	89
5.2 空间相关分析的理论模型	91
5.3 空间相关中的方向性	93
5.4 空间估值的随机性模型	94

5.5	空间估值的非随机性方法	108
5.6	关于数据变换	110
第六章	时空回归与自回归模型	111
6.1	时间序列模型	111
6.2	趋势面分析	115
6.3	空间回归-自回归模型	119
6.4	扩散的自相关分析及模型	125
6.5	扩散过程的模拟	130
第七章	传统生物统计学方法的改进	134
7.1	样本均值的估计	134
7.2	两样本均值的比较	135
7.3	线性回归的残差分析	136
7.4	改进形式的回归分析	137
7.5	实例研究	139
第八章	空间抽样设计	146
8.1	空间抽样的方法	146
8.2	在 N 个位置上选取 n 个样本的实施	149
8.3	样方面积大小的影响	151
8.4	样方面积效应的纠正	154
8.5	样方面积改变时方差函数的估计	156
结束语:	进一步的设想	158
参考文献	159

第一章 生物统计学与空间相关

生物统计学是生物学试验设计及数据分析的基本方法(刘来福 1988, 马育华 1985)。生物统计学方法的基本假设是:所有样本都是随机样本,换句话说样本间都是相互独立、具有相同的概率分布;以此假设为基础,进一步得到了生物统计学关于统计估值、假设检验以及回归、判别、聚类等一系列结果。随着空间自相关问题中样本间空间相关假设的提出(Geary 1954, Cliff et al. 1973, 1981, Sokal 1978a, b),传统生物统计学方法的缺陷逐渐暴露出来。由于空间相关在生物学、生态学的各种现象中经常出现(Cliff et al. 1981, Sokal et al. 1987, Oden et al. 1986, Liebhold et al. 1989, 1991, Guo 1994, Kareiva 1994, Rhodes 1994, 李天生等 1991),因此如何解决空间相关样本的分析问题已成为许多生物统计学和统计生态学研究人员注目的研究方向。地质统计学(geostatistics)方法的大量引入就是其中的方向之一(Journel 1978, David 1977, Schotzko et al. 1989, Liebhold et al. 1991, Hohn et al. 1993, Beliaeff et al. 1995, 周国法等 1991a)。

本章主要介绍什么是生物种群的空间自相关,为什么要研究生物地理统计学,本书的主要内容和相关的参考资料。

1.1 统计模型与空间相关

自然界中,生物种群或者是均匀分布的,或者是随机分布的,或者是呈斑块状聚集的,或构成梯度变化的或其他类型的空间结构——包括方向和范围的分布格局。生态学家们利用经典生物统计学和统计生态学提供的手段给出了多种判定这种空间结构的方法(徐汝梅 1987, 丁岩钦 1994)。在研究多种种群间相互作用以及一个种群内个体间、种群与其生存环境间的相互作用时,则考虑了空间异质性(Odum 1971, Dempster et al. 1986, 祝宁等 1996)的问题;这种空间异质性更广义的可称为空间相关性,它在生态学理论中起着关键作用。多种生态学理论及模型都假设生态系统中的单元间在空间上或时间上如果“距离”更接近,则其相互作用就强,“距离”越远相互作用就越弱,如传染病及其他灾变模型,种群竞争理论,群落的稳定性、演替、进化和同化,多样性的维持,种群的扩散、遗传、寄生,捕食者与猎物间的相互作用等;所有这些理论及模型的假设或结论都表明,空间相互作用(即空间相关性:种群内个体间的、种群间的、种群与环境间的)构成了生态学理论的基本成分。

事实上,空间相关存在于一切由相互影响的不同部分组成的、空间发生的生态学现象中。这些空间现象可以划分为两大类:空间相互作用和扩散现象,以及在一个充满某个空间区域内发生的现象。所谓空间相互作用是指一个地方发生的现象会影响其他与之相关的位置(地方)的结果,而且一般来说这种影响与位置间的空间关系即距离及方向有关。扩散是个体间或种群间或种群与生存环境间空间相互作用的最好例子。一般来说,扩散总是

从扩散源开始逐步向外发展的,离扩散源近的地方会先受到扩散的影响;成虫迁飞是另一类空间相关现象,因为它的发生可能是在距离较远的位置上进行的。所谓充满某个空间区域的现象是指许多相邻的位置受同一现象的影响,如森林害虫的暴发在相邻地区是相关的,因为一般的暴发过程会同时影响这些地区。正是由于不同位置上现象的相互影响才导致某种特定的格局或空间结构,生物地理统计学的目的就是分析并预测这些空间相互影响、相互影响的时序变化及其可能导致的结果,从而揭示生态学现象的规律和实质。

生物地理统计学的主要内容包括空间数据的选择、分析,通过分析使空间现象减少为明显的和有用的部分;然后把这些有用的成分与理论的期望相比较,从而得到关于空间分布的结论;同时用这些有用的成分来分析空间分布的未来发展趋势。而理论分布的假设主要有以下几种可能。

①相互独立同分布(iid)模型:这是经典生物统计学的基本假设(刘来福 1988);它的假设是“对于生态学现象的观测都是在同样条件下(同质性假设)进行的并且是相互独立(没有相互影响假设)完成的,不同空间位置的现象具有相同的概率分布(同分布假设)”。于是从现象中观测的数据是一组随机样本,这样就可以用标准的统计程序建立统计模型、估计模型参数、检验统计假设。例如,对某个生物种群的几次抽样结果是 6、5、12、3、8、3、1,假设这些数据来自于随机抽样也许合理;但若另外的调查结果是 9、5、6、4、2、11、1,它也可以看作随机样本;但是由于它们是对同一种群的调查,因此两组数据应该组合在一起处理,这样由于样本量的增加,得到的结论应该是更合理的。那么组合起来的数据是否是随机样本呢?答案是不确定的。

②非齐次(异质)性模型:当数据具有异质性特性时,通常是在统计模型中加上非常数均值假设,如组合 Poisson 分布(Pielou 1977, 周国法 1996, 索思伍德 1985)就是假设不同位置上分布密度的均值不同。但这只是考虑了大尺度上(平均)的变化性,仍需要考虑小尺度上变化的异质性。舍弃独立性假设是得到这种现象统计模型的最途径。注意强调空间特性是研究种群动态的基本条件。

③相关性模型:独立性是一个十分方便的假设,它使得生物统计学的理论得以形成和完善,但实际情况则可能是更多地存在空间的统计依赖性(Cliff et al. 1973, Sokal et al. 1987, Oden et al. 1986, Ord 1975)。对于空间数据,这种空间依赖性存在于任何方向,且随距离变化和时间变化而不同。一般来讲,个体位置越分散,个体间的这种相关性越弱;对于个体群和种间相互作用结论是相同的。

我们生活的空间是四维的,空间的方向是无限的,时间的方向是唯一的。对于那些距离接近的数据(空间上或时间上接近,种群内个体间的或种群间的或种群与环境间的距离接近)认为是相关的(即非统计独立)是自然的,并且已被生态学家成功地应用于构造生态学现象的模型。纯时间模型或时间序列模型(Legendre et al. 1983, 陈兆国 1988)就属于这种模型,它通常是同一分布的观测,但数据在时间上是依赖的。

真的空间模型只是在最近 10 年才加入生物统计学课程的。追溯到地质统计学中的空间相关则是 30 年前建立的适合地质矿产估值的一系列理论和方法。地质学、土壤科学、流行病学、农业及林业科学、生态学、大气科学、环境科学、海洋科学,或其它任何工作,只要它的数据是在空间不同位置上收集的,都需要发展一种模型(可以不是统计的)以表征不同空间位置上测量的依赖性。显然这种模型远比时间序列复杂,因为空间的方向不是唯一

的,而且空间同一位置上发生的生态学现象还随时间变化(徐汝梅 1990,周国法 1992,林一中等 1988,Liebhold et al. 1989, Gage et al. 1990,Epperson 1993, 1994),这样的模型在传统的生物统计学中是不研究的,而地质统计学只建立静态模型(Journel 1978, Armstrong 1989, Isaaks et al. 1989, Cressie 1991)无法用来研究生态学现象的时间变化性。本书的目的就是介绍有关空间生态学现象的分析方法以及空间生态学现象随时间变化的分析方法。

在介绍生物地理统计学的内容之前,先看一看为什么必须研究这种新的统计学分支。

1.2 传统生物统计学方法的局限性

前面已经指出,空间相关性在生态学现象中随处可见,而这种空间依赖性在经典生物统计学中被忽略了,从文字上还不足以说明经典的生物统计学方法存在什么问题。下面从几个例子说明忽略空间相关性的经典方法存在的问题。

1.2.1 分布型指数方法

空间分布格局是研究一个种群的空间行为的基础。任何种群都是在空间不同位置分布的,但由于种群内个体间的相互作用及种群对环境的适应,使得不同种群、同一种群在不同环境条件下会呈现不同的空间分布格局,这种格局显然也随时间而变化。传统生态学中研究空间分布型有多种方法,但基本上都与样本数据的均值和方差或样本数据的概率分布有关(皮洛 1991,徐汝梅 1990,丁岩钦 1994)。图 1-1 是一特定区域的马尾松毛虫卵

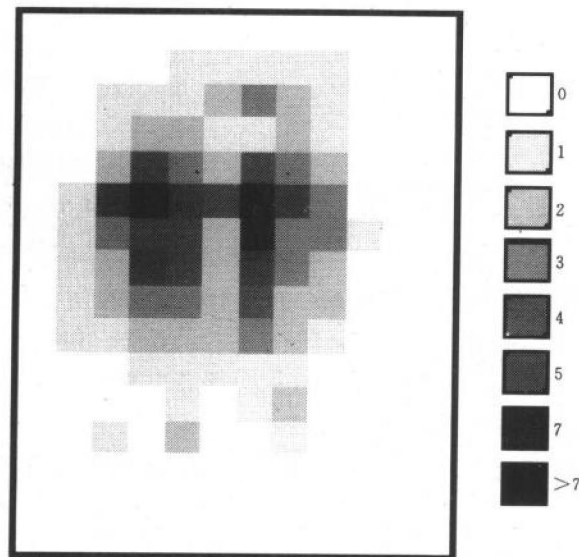


图 1-1 马尾松毛虫卵块的分布密度

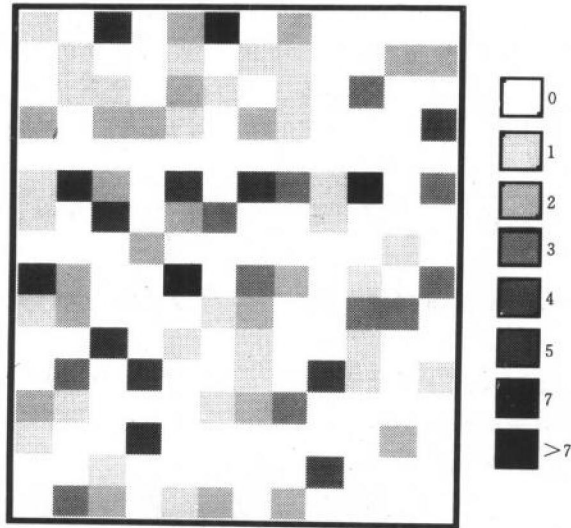


图 1-2 图 1-1 密度的空间随机化

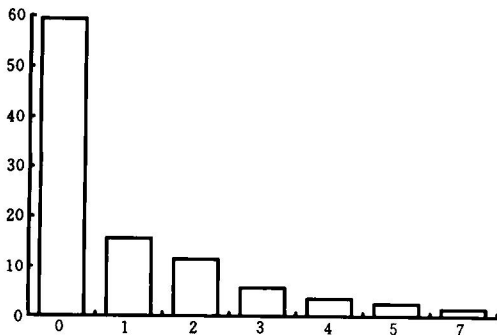


图 1-3 图 1-1 中密度分布的直方图

块(周国法等 1987)的空间分布图,从图中可以看出卵块分布呈明显聚集状;其分布特点是密度从区域中心向四周减小,且距离相近(尤其是相邻)的格间卵块数差异不大(也称为空间相似),但距离较远的格间卵块数目差别较大。

图 1-2 和图 1-1 是同一调查数据,但对图 1-1 的分布位置作了随机处理,即从图 1-1 中选一数据然后用随机方法给出它的位置从而得到图 1-2 的分布格局。从图形上看,

二者具有明显不同的空间分布特征。原始分布具有明显的空间相关趋势,而随机化后的分布更像是随机的或均匀分布,没有任何明显的空间趋势。但二者的数据完全相同,因此二者的均值、方差、变化系数、峰值系数、直方图都完全相同(图 1-3,表 1-1),所以如果用经典统计学方法拟合的分布型也相同,无法找出空间分布上的任何差异,但二者的空间格局显然不同。因此需要有更进一步的方法来区分这两个空间分布格局。

表 1-1 由图 1-1 或图 1-2 得到的统计量

统计量	均值	标准差	峰值系数	变化系数
样本值	1.000	1.808	3.387	1.808

同样的问题在研究种间相互作用时也存在,可用类似的手段发现研究种间关联度的经典方法也存在很多缺陷。后面将讨论这个问题。

虽然皮洛(1991)、Greig-Smith(1952)等提出了连续带和可变网格方法,但他们采用的仍然是对调查的数据建立概率分布模型进行比较,或对得到的均值和方差进行比较,这些方法仍然没有空间特征的描述,因此无法从根本上解决上述问题,也不能看作真正空间分析方法。

1.2.2 空间估值

考虑下面的简单模型,在所有生物统计学课程开始就可以找到。假设 $Z(1), Z(2), \dots, Z(n)$ (简记为 Z_1, Z_2, \dots, Z_n) 是来自于 iid (相互独立且具有相同概率分布) 随机变量的样本,相同的分布为正态分布 $N(m, \sigma^2)$, 其中均值 m 未知,方差 σ^2 已知。于是 m 的最小方差无偏估计(刘来福 1988)为:

$$m^* = Z_E = \sum_i Z_i/n$$

其中 Z_E 为样本平均值。而且可以直接推断出 m 的分布为 $N(m, \sigma^2/n)$, 于是 m 的 95% 置信区间为 $(m^* - 1.96\sigma/\sqrt{n}, m^* + 1.96\sigma/\sqrt{n})$ 。

现在假设数据是正相关的,相关性随分离距离下降,即:

$$\text{Cov}(Z(i), Z(j)) = \sigma^2 \rho^{|i-j|}$$

其中 $0 < \rho < 1$, 这种形式的相关是从一阶自回归模型(陈兆国 1988)得到的:

$$Z(i) = \rho Z(i-1) + \epsilon(i), i \geq 1$$

其中 $\epsilon(i)$ 是 iid 正态 $N(0, \sigma^2(1-\rho^2))$ 分布且与 $Z(i-1)$ 相互独立的。现在有

$$\text{Var}(m^*) = \frac{1}{n^2} \sum_i \sum_j \text{Cov}(Z(i), Z(j)) = \frac{\sigma^2}{n} \left(1 + \frac{2\rho}{1-\rho} \left(1 - \frac{1}{n} \right) - 2 \left(\frac{\rho}{1-\rho} \right)^2 \frac{1-\rho^{n-1}}{n} \right)$$

当 $n=10, \rho=0.26$ 时, $\text{Var}(m^*) = 1.608\sigma^2/10$, 于是 m 的 95% 的置信区间为 $(m^* \pm 2.46\sigma/\sqrt{10})$ 。因此当数据间存在正相关时,置信区间的经典估计太小了,本例中的真正覆盖概率是 87.8% 而不是 95%。显然如果 $-1 < \rho < 0$, 则经典估计又太宽了。

空间相关对估计的影响可从下面的表达式理解:

$$\text{Var}(m^*) = \sigma^2/n'$$

其中

$$n' = n / \left(1 + \frac{2\rho}{1-\rho} \left(1 - \frac{1}{n} \right) - 2 \left(\frac{\rho}{1-\rho} \right)^2 \frac{1-\rho^{n-1}}{n} \right)$$

n' 可解释为独立观测的等价样本数,例如 $n=10, \rho=0.26$ 时 $n'=6.2$, 即 10 个相关观测值的结果只相当于 6 个独立样本的近似值。对于充分大的 n 值有 $n' \approx n(1-\rho)/(1+\rho)$, 因此即使 n 充分大也同样存在自相关对统计估值的较大影响,因为估计结果是依赖空间相关系数的。

正是由于存在自相关时,均值及方差的估计都受影响,因此由抽样得到的分布型指数

也受到影响,从而导致分布型判断上的错误。

1.2.3 预测问题

同样在数据间空间相关存在时,预测结果也受到影响。继续前例假设要预测 $Z(n+1)$ 的值,同样假设 $Z(1), Z(2), \dots, Z(n)$ 是 iid 正态分布的,均值 m 未知,方差 σ^2 已知。假设 $Z^*(n+1)$ 是预测值,预测满足无偏的条件是:

$$E(Z^*(n+1)) = m$$

最小平方平均误差条件是:

$$E(Z(n+1) - Z^*(n+1))^2 = \min$$

上述条件下的预测值是 $Z^*(n+1) = m^*$ 即样本均值,最小平方误差为 $MS = \sigma^2(1+1/n)$ 。

现在加入相关性假设 $\text{Cov}(Z(i), Z(j)) = \sigma^2 \rho^{|i-j|}$, 此时最小平方误差预测结果为:

$$Z^{**}(n+1) = \rho Z(n) + (1-\rho)\{Z(1) + (1-\rho) \sum_{1 \leq i \leq n-1} Z(i) + Z(n)\} / ((1-\rho)n + 2\rho)$$

容易得到在空间相关假设条件下用样本平均值估计的平均平方误差为:

$$E(Z(n+1) - m^*)^2 = \sigma^2 \{1 + 2\rho(n\rho^n - 1) / n(1-\rho) - 2\rho^2(1-\rho^{n-1}) / n(1-\rho)^2\}$$

当 $\rho=0.26, n=10$ 时上式为 $(1.0905\sigma^2)$, 而 $MS=1.1\sigma^2$, 二者相差不大。进一步,当空间相关随距离呈几何级数递减时,经典预测区间基本上都是正确的。但是如果采用空间相关假设,且采用估计量 $Z^{**}(n+1)$, 则有最小平方平均误差:

$$MS_\rho = E(Z(n+1) - Z^{**}(n+1))^2 = \sigma^2 \{1 - \rho^2 + (1+\rho)(1-\rho^2) / ((1-\rho)n + 2\rho)\}$$

当 $\rho=0.26, n=10$ 时,上式为 $(1.01952\sigma^2)$ 。

对应的 $Z(n+1)$ 的 95% 预测区间为:

$$(Z^{**}(n+1) \pm (1.96\sigma \sqrt{1.0195})) \text{ —— 有相关性假设、采用估计量 } Z^{**}(n+1)$$

$$(m^* \pm 1.96\sigma \sqrt{1.0905}) \text{ —— 无相关性假设、采用估计量 } m^*$$

二者相差 0.03σ , 差别不大。对于充分大的 n , 用 $Z^*(n+1)$ 和 m^* 估计 $Z(n+1)$ 的预测区间长度比为 $(1-\rho^2)$, 当 $\rho=0.26$ 时 $1-\rho^2=0.93$, 二者相差不是太大, 但当 $\rho=0.5$ 时 $1-\rho^2=0.75$, 二者相差就太大了。可见弱的空间相关对预测结果影响不大, 而强的空间相关对预测结果影响很大。

1.2.4 具有空间相关误差的线性模型

假设在空间位置 $\{u_1, u_2, \dots, u_n\}$ 上观测到的空间数据为 $\{Z(u_1), Z(u_2), \dots, Z(u_n)\}$, 其多元线性回归模型为:

$$Z(u) = \sum_i b_i X_i(u) + \delta(u)$$

其中 $X_1(u), X_2(u), \dots, X_q(u)$ 是 q 个可观测的变量(张尧庭等 1983), 它们可能依赖于空

间位置也可能不依赖于空间位置; $\delta(u)$ 是均值为 0 具有有限方差的误差,这是最常见的回归形式的模型。

在假设 $\{Z(u_1), Z(u_2), \dots, Z(u_n)\}$ 相互独立相同分布方差为 1 时即得到经典的多元回归假设,这时参数(系数) $B=(b_1, b_2, \dots, b_n)'$ 的最小二乘估计是

$$B^* = (X'X)^{-1}X'Z$$

其中 X, Z 分别是相应于变量 $X(u), Z(u)$ 的样本矩阵。

下面是作者等(Zhou et al. 1995a)给出的关于美国西弗吉尼亚州中部地区森林对舞毒蛾(Lepidoptera: Lymantriidae)危害敏感性预测的例子。所谓对舞毒蛾危害敏感性指在有舞毒蛾危害时某种树木受害的可能性或多个树种可能受害的平均百分率。在对 2100 个样方调查后,得到了平均敏感性数值,同时每个样方处都测得了海拔、坡度、方位及地形位置。可以假设下面的回归式成立:

$$p = b_0 + b_1(\text{海拔}) + b_2(\text{坡度}) + b_3(\text{方位}) + b_4(\text{山脊}) + b_5(\text{中坡}) + b_6(\text{山谷}) + e$$

用 2040 个样本(随机抽取 60 个样本用于估计精度的交叉检验)构造了回归模型。图 1-4 给出了海拔的分布图,图 1-5 是由回归式得到的研究区域内各点上的敏感率估计值。显然估计结果与海拔图有非常相似的趋势。对 60 个样本作交叉检验的结果表明,估计值和实际值差异显著。那么是这种回归模型不够理想吗?有这种可能。

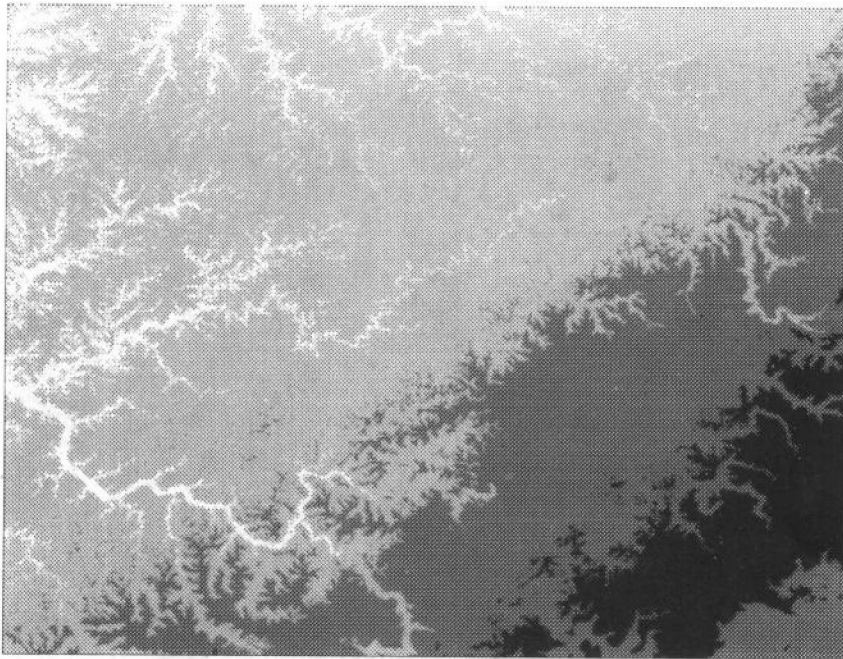


图 1-4 海拔分布图,黑色海拔最高,白色海拔最低

作者的进一步研究是假设估计误差存在空间相关性。进一步用一般克立格(Krige, 后面将详细介绍)估计,得到了研究区内各点上误差的空间相关估计,即对误差作空间相

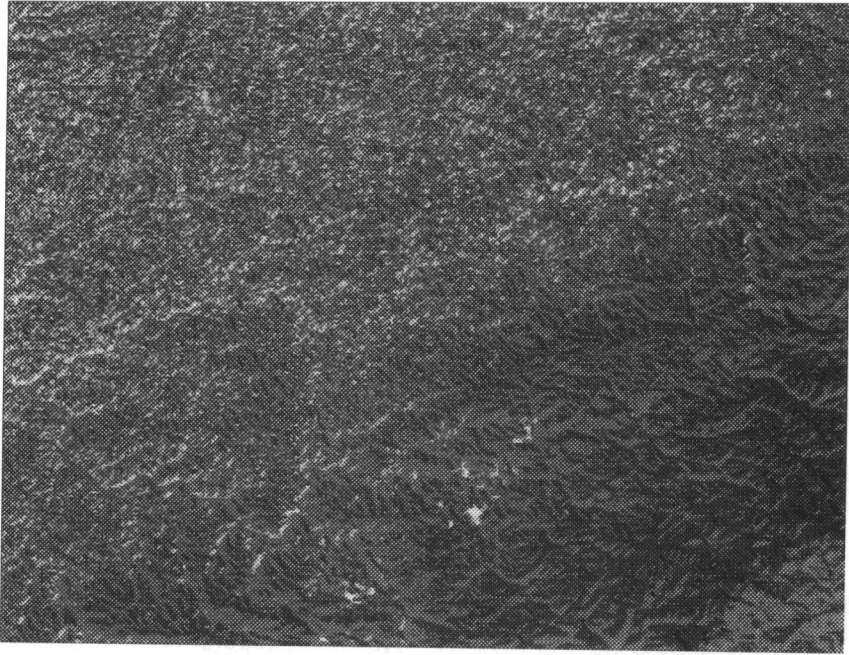


图 1-5 线性回归估计的敏感率分布

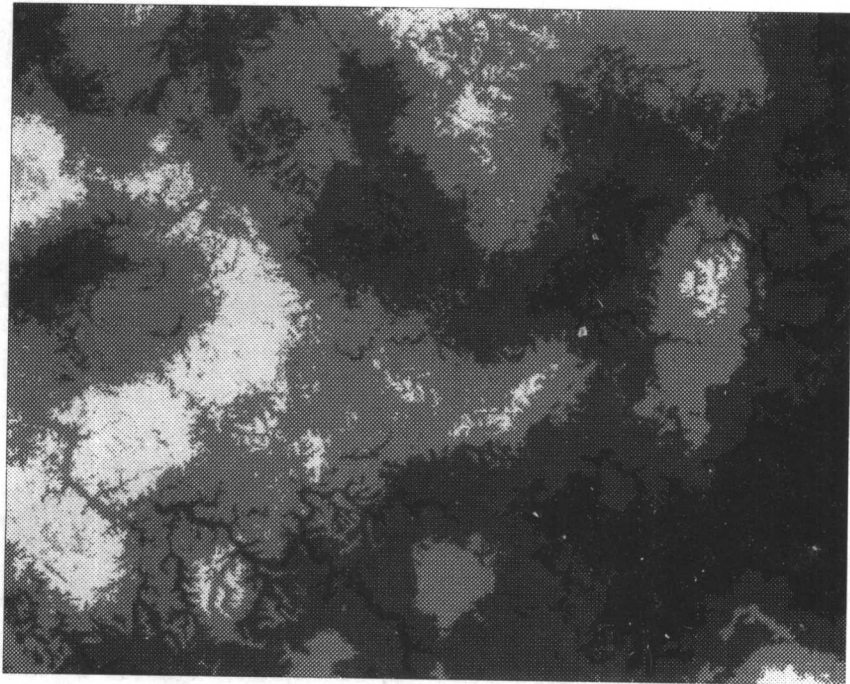


图 1-6 线性回归加上克里格方法估计的敏感性概率分布

关处理,然后用线性估计结果加上误差的相关处理结果作为敏感值的新估计,图 1-6 是这种新估计的结果。显然新的估计至少外表上比线性回归要好,因为这时估计结果不再有与海拔完全一致的趋势了。再用留下的 60 个额外样本作交叉检验,发现这时的估计值与样本值没有显著差异,说明新估计确实比原估计准确多了。

上面的例子说明,当误差间存在空间相关时,经典回归的估计也许是不充分的,必须考虑空间相关的影响。还可以从许多实例说明经典的判别分析、聚类分析等在研究空间相关问题时的局限性。因此必须采用更好的方法才能解决上述问题,本书后面介绍有关的改进方法。

1.2.5 物种间空间相关性

这个问题在研究群落问题时是经常出现的(吴亚等 1980,任立宗等 1988,赵志模等 1990,王献溥 1990,王勇等 1991,杜道林等 1995)。显然,物种的不同组成及其在空间分布的不同格局构成了群落的空间结构,而物种间的不同空间相互作用(种间相互作用必然是空间的)导致了群落的不同功能,种间相互作用的平衡以及环境对种群影响的稳定使得群落得以稳定。下面首先看一下种间关联度能描述多少群落的特征。

为了说明问题,下面考虑两个人工的群落(图 1-7)。

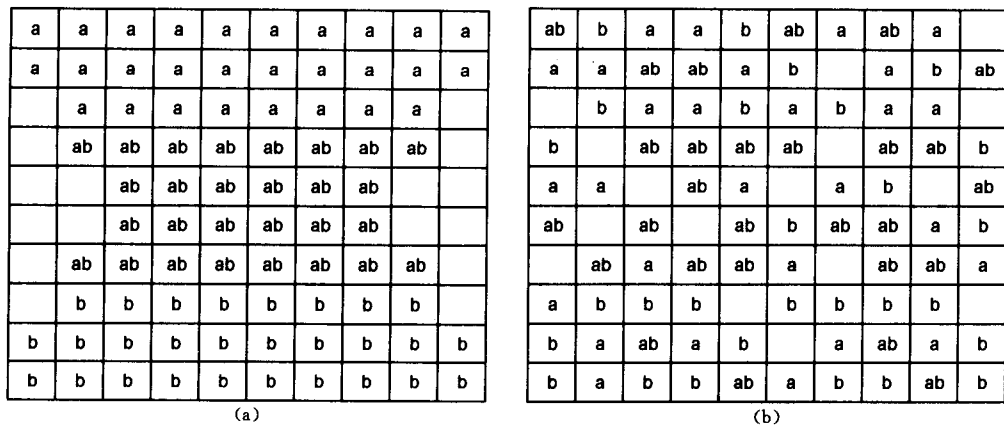


图 1-7 (a) 人工设计 10×10 的群落 AC1 中种群 A、B 的个体分布,a 表示种群 A 的个体, b 表示种群 B 的个体;(b) 群落 AC2 是人工群落 AC1 中个体分布的空间随机化

表 1-2 是图 1-7 的人工群落中物种数量的分布。容易计算出图 1-7(a)和图 1-7(b)中的物种数量分布是相同的。用传统的群落数量特征计算公式可以得到两图的群落物种数、丰富度、多样性、优势度、关联度、种间相关系数等指标值都相同,两个群落的不同在于群落内物种分布格局不同从而导致群落的空间格局(结构)不同。种间的空间关系(用关联度无法度量)不同导致群落的结构不同,从而两个群落的结构和功能有可能完全不同。遗憾的是多数的群落研究(Pielou 1966,Odum 1971,孙儒泳 1992)主要关心的是个体或物种数量的分析结果,对于物种的空间格局、种间的空间相关、群落的空间结构等缺乏数量方

法。例如如何用数量方法而不是定性方法说明图 1-7 的两个群落的物种空间格局和种间空间关系是不同的,过去是没有适当的方法的。

表 1-2 图 1-7 中的物种数量分布

		a	
		0	1
b	0	16	28
	1	28	28

进一步的问题是描述群落内物种多样性、优势度、种间关联性等和描述物种空间格局、群落空间结构等的方法是否可以统一? 或者至少它们之间有什么关系? 这也许是一个复杂的问题,但是它是一个很有实际意义的问题。

1.3 本书的基本内容

生物地理统计学是生态学家、生物统计学家为研究生物种群的时空动态而创立的方法,它的理论涉及许多领域,包括生物统计学、生物地理学、生态学、数学生态学、景观生态学、地理统计学、地质统计学、空间统计学等;它的方法除了涉及一般生物统计学内容外,还包括大量与地理信息有关的内容,如空间过程模型、空间试验设计、空间抽样方法、空间估值与预测;许多方法都是最近几年才出现的。作者在本书中并不准备做面面俱到的介绍,只介绍我们在近几年研究过的领域,主要包括两大部分:种群空间结构的分析和种群密度估计与预测问题。

全书内容可以分两部分:不规则空间数据分析与规则空间数据分析。这种划分不只是根据获取数据的形式(规则与不规则)的不同而划分,更主要的是当一个方法适合于不规则数据格式时,它也适合于规则数据,但是反之则几乎是不可能的。因此在介绍一般方法外,作者还要介绍规则空间数据的一些特有方法,但尽量不重复前面已有的内容。

空间分布格局是生态学家最熟悉的问题之一,前面已经指出了传统方法的不足,因此本书在第二章、第六章介绍散点图、方差图、相关图方法,这些方法不仅可以分析种群的空间分布结构,而且适应于分析种间关系,甚至适应于种群内不同水平之间的比较,如幼虫与成虫空间分布的比较。

空间估值,如某个种群在一定范围内个体密度分布的估计、受害率分布的估计、产量分布的估计等问题,都是生态学家及实际工作者关心的问题,许多问题更是制定决策的依据。本书第三章介绍空间密度估计的一般问题,第五章则介绍了多元方法、非线性方法、百分率估计及空间分布的模拟问题;第四章介绍空间预测问题。

多元分析是经典生物统计学、统计生态学、景观生态学中常用的方法,其主要特点是用相关的可观测的变量去确定某个我们关心的生态学变量的特征。第六章及第七章的部分内容介绍有关多元统计的改进方法,因为许多方法在空间相关假设下已经不成立了。描述变量间的相关也不再使用简单相关系数,而是采用协方差函数和相关图。

对种群动态研究感兴趣的读者可以参考第四章及第六章的有关内容,虽然书中没明确说明种群动态问题,但上述章节涉及空间分布的动态、种群数量分布的预测问题。特别是有关扩散问题,书中多处提及,基本上给出了完整的扩散过程的描述、分析及预测模型。

对传统的生态学和生物统计学抽样问题感兴趣的读者,可以参考本书的第八章,其中介绍了空间抽样的方法和样方大小对抽样结果的影响,以及如何处理不同样方问题的手段。

总体上讲,本书的内容是对近几年生物地理统计学研究的概括,对许多问题的研究都有待于进一步深入,希望本书能起到抛砖引玉的作用。