

327-331

12035(17)

空间自相关与分布型指数研究

李天生

(中国林业科学研究院林业研究所, 北京, 100091)

周国法

(北京大学分校数学系)

Q141

A

摘要 本文介绍了地学统计学(geostatistics)中关于自相关函数、方差图的基本概念。提出了扩散性指数 I_z , 当 I_z 从 0→1 变化时, 种群的空间分布由均匀→随机→聚集。 I_z' 可反映不同方向上种群分布情况。以松毛虫种群调查资料为例, 介绍了这些新概念的应用。结果表明, 空间分布型与空间位置、方向有关。当研究昆虫种群空间分布型时, 不考虑空间位置是不全面的。

关键词 地学统计学, 方差图, 变程, 扩散性指数, 空间自相关, 空间分布型。

关于空间分布型的研究, 从早期的频次分布法到以后的分布型指数, 以及 Taylor 幂法则, Iwao 回归及各种改进形式^[1-5], 都依赖于扩散系数 ($C_v = S^2/\bar{x}$) 的定义及相应的分布型判定准则, 它们只能判定简单的空间分布型, 这些方法都受样方大小和抽样空间大小的影响^[6,7]。本文继续探讨这个问题, 并将空间位置引入研究中, 用地学统计学的理论和方法研究昆虫种群空间分布型问题, 提出相应的扩散性指数为判定空间分布型的指标。

1 自相关函数及其性质

空间自相关是指昆虫个体都占据一定的生存空间, 不同位置上的个体间存在着相互影响, 而研究个体间的相互关系时, 必须考虑空间位置关系。描述这种现象仅靠分布型指数远远不够, 这里介绍地学统计学中的方差图及影响范围的概念及其方法^[8-10]。地学统计学与经典生物统计学的本质区别就在于前者考虑样方的空间位置及样方间的距离。

设 $Z(x)$ 是某种昆虫在研究域中 x 处(二维或三维空间点的坐标)的个体数量, 通常以 x 为中心的样方内的个体总数。对于未知的空间分布 $Z(x)$ 是随机变量, 与空间位置有关, 并且假设 $Z(x)$ 满足下面条件:

- (1) $E[Z(x)] = \mu$ 平均值不依赖分布位置,
- (2) $E[Z(x+h) - Z(x)]^2$ 只依赖于距离 h 。

在上述假设下, 定义:

$c(h) = \text{cov}[Z(x+h), Z(x)]$ 称协方差函数,

$r(h) = \frac{1}{2} \text{var}[Z(x+h) - Z(x)]$ 称为方差函数。

根据假设(1)、(2)容易证明: $r(h) \geq 0$ 且 $r(0) = 0, r(-h) = r(h), r(+\infty) = c(0), c(h) + r(h) = c(0), h-r(h)$ 图象称为方差图。

方差图的形状反映了昆虫种群空间分布的结构或空间相关的类型, 同时给出了这种空间

收稿日期: 1992 08 15, 修改稿收到日期: 1993 05 20。

相关的范围。如果存在空间自相关, $r(h)$ 会随 h 增大而增大, 见图 1, 典型的情况是当 $r(h)$ 增大到一个平衡状态才停止, 这时的 $r(h)$ 值称为基台。达到基台值后, $r(h)$ 就稳定在基台值附近, 这时的 h 称为相关程或变程 (range), 即存在空间相关的最大距离。理论上 $r(0)=0$, 但由于种种原因, 如取样误差、小尺度变化性等, 使得样本值在小于 $h=a$ 的距离分离时具有更大的不相似性, 于是 $r(h)$ 外插到 $h=0$ 时并不取 0 值, 这个值称为金块效应, 它表示了由于短距离变化较大和抽样误差引起的方差。某些情况下, $r(h)$ 会呈直线状, 这表明数据的完全随机性或完全均匀性, 说明在抽样尺度下没有空间相关性。水平直线状态称纯金块效应。

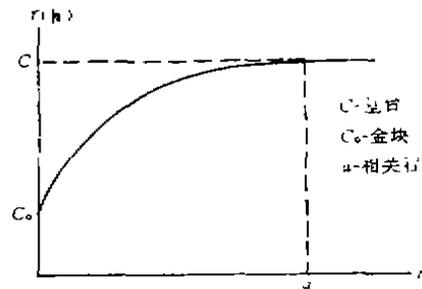


图 1 一维情况下 $r(h)$ 的典型图象
Fig. 1 Typical graph of $r(h)$ in one dimension c —sill, c_0 —nugget, a —range

上面指出距离是方差图的重要特征量, 另一重要特征量是方向。当昆虫种群分布区域的环境条件不同时, 就会出现空间分布上的差异。经典生物统计学总是把所有样本点一视同仁, 然而对于某些林业害虫, 在阳坡和阴坡分布就有显著差别。地学统计学讨论了关于方向问题。即各向同性和各向异性。异向性是指某种昆虫在某个方向上比其它方向有更好的连续性。各向同性、异性可以通过不同方向上计算 $r(h)$ 来检验。

关于相关程, 亦称变程 (记 a), 它反映了数据的空间相关距离。一个分布, 如果 a 存在, 总能找到它; 当 $a > 0$ 时, 如果两点间的间隔 $|h| \leq a$, 两点处的个体是相互影响的, 相反两处则互不影响。不同方向的 a 值反映不同方向的影响范围。且 a 值不受样方大小的影响, 如果相对尺度发生变化, a 的绝对大小是固定的。如果 a 的固有大小是 4 (单位距离), 当样方大小取 1 (单位距离) 时, $a=4$, 当样方大小取 2 时, 则 $a=2$ 。可见 a 的固有大小是不变的。另外 c_0 值的大小反映了密度变化的幅度, $c_0 \rightarrow 0$ 反映连续变化, c_0 越大变化幅度越大, 指标 $c_0/(c_0+c)$ 反映这种变化程度。

对于观测数据 $Z(x_i), i=1, 2, \dots, n$, 要得到 $r(h)$ 的表达式, 首先要计算样本方差函数值:

$$\hat{r}(h) = \frac{1}{2N(h)} \sum_{x_i - x_j = h} (Z(x_i) - Z(x_j))^2$$

其中 $N(h)$ 是相距 h 的点对 (x_i, x_j) 的个数; 可用 $\hat{r}(h)$ 的值拟合 $r(h)$ 模型。常用的 $r(h)$ 模型有:

(1) 线性模型 $r(h) = A + B \cdot h, h > 0$

(2) 指数模型 $r(h) = c_0 - e^{-3h/a}, h > 0$

(3) 球状模型 $r(h) = \begin{cases} c_0 + c \left[\frac{3}{2} \frac{h}{a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right], & 0 < h \leq a \\ c_0 + c; & h > a \end{cases}$

(4) 高斯模型 $r(h) = \begin{cases} 0, & h = 0 \\ c_0 + c(1 - e^{-3h^2/a^2}), & h > 0 \end{cases}$

模型中的 a 值正好是相关程。

关于方差模型的选择及参数估计方法, 可用最小二乘法。但存在着许多问题。一是关于计算 $r(h)$ 时取值的间隔, 它对模型选择有很大影响, 如果间隔取大, 会损失许多有用的信息; 间

隔太小时 $N(h)$ 对于不同的 h 值, N 值会有差别, $N(h)$ 值太小相应的 $r(h)$ 值可差别很大。如果样本数为 n , 则 $N(h)$ 值不应少于 $\frac{1}{2}n$ 。建议除去 $N(h)$ 较小的 $r(h)$ 值后, 用最小二乘法拟合模型。

2 扩散性指数 I_z 的定义及意义

在经典生物统计学中, 样本方差 S^2 反映了生物种群个体分布的不均匀程度, 根据潘松分布的特点知道扩散系数 $C_v = \frac{S^2}{\bar{X}}$ 是一个很好的反映分布特征的指标, 由于它依赖于样方大小, 因此必须给它加上一个距离尺度, 使其不依赖于样方的选择。一个简单办法是以样方大小作权计算 \bar{X} 和 S^2 , 结果得到 $C'_v = C_v/A$, A 是样方大小, C'_v 的单位正好是单位面积虫数, 它的缺点是仍然只反映总体平均, 没有反映空间特征。由于 a 反映了空间影响范围, 为消除尺度影响, 可用 L/a 反映空间影响的变化情况 (L 是抽样区的半径), 为消除量纲影响, 用 S^2/\bar{X}^2 代替 C_v , 最后可用 $J = VL/m^2 a$ 来反映空间分布的结构, 其中 V, m 分别是 $Z(x)$ 的方差和。

由于 J 在 $(0, +\infty)$ 变化, 不便于应用, 故用 I_z 代替 J :

$$I_z = \sin(\operatorname{tg}^{-1} J) = \frac{V^2 L^2}{a^2 m^2 + V^2 L^2}$$

$I_z \in [0, 1]$, I_z 与 J 的性质相似。 I_z 称为扩散性指数, 可以看作是分布型指数的一种推广形式, 但它有空间上的意义。 I_z 的定义也可在不同方向讨论, 从而得到各个方向上的分布变化情况。

当 I_z 从 $0 \rightarrow 1$ 变化时, 种群的空间分布从均匀分布 \rightarrow 随机分布 \rightarrow 聚集分布, I_z 越接近 1, 聚集程度越高, 扩散性越强。

对于完全的随机分布, 当样本量较大时, I_z 的均值 $E(I_z) = \frac{1}{1 + \left(\frac{a}{L}\right)^2 m^2}$ ($a \approx L, m = V$) 因

此理论上, 对完全的均匀分布 ($a = \infty$) $I_z = 0$, 完全随机分布 $I_z = \frac{1}{1 + \left(\frac{a}{L}\right)^2 m^2}$, 完全聚集分布

($a = 0$) $I_z = 1$ 。

与扩散系数 C_v 相比, I_z 至少有以下几方面优点: (1) I_z 可反映不同方向上种群分布的情况; (2) I_z 是无量纲指标, 且只在 $[0, 1]$ 变化; (3) I_z 不受特异值的影响, 具有较大稳定性; (4) I_z 可能为抽样及建模提供信息^[6]。 I_z 的缺点是计算比较复杂。

3 实例研究

对于任何生物种群, 一旦它的分布协方差函数 $c(h)$ 或方差函数 $r(h)$ 已知, 关于种群空间分布的结构、密度的估计等都迎刃而解^[6, 10]。下面以马尾松毛虫调查资料为例^[7], 用实验方差函数值拟合理论模型 $r(h)$, 以求得 c_0, c 和 a 。

图 2, 图 3 均可看出, 对于该地松毛虫各向同性的条件不成立。图 3a 为平面等方差线, 说明各方向变化不一致。b, c, d 分别为 3 个方向方差图的示意图 (单位相同), 可比较其差别。各向异性与研究区的地形有关。研究区是由相邻的 3 个山坡组成。南北为山坡走向, 东西方向是高度相近的山坡相连。从图 3 也可见东西和南北方向的相关程有较大差异。结果表明水平方向的相互影响比垂直方向的相互影响范围要大。可见空间位置对昆虫分布的影响是不可忽视的。

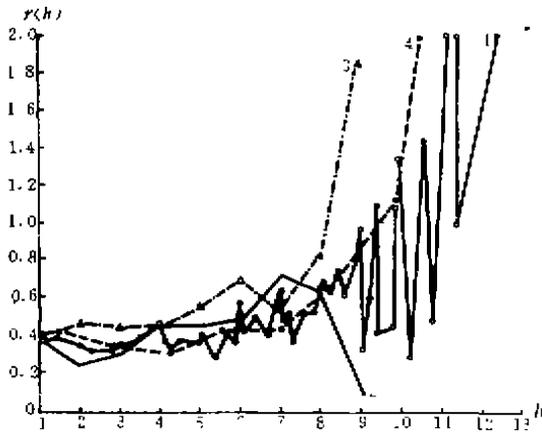


图2 实验方差图

Fig. 2 Experimental variogram

- 1. 全方位 All directional
- 2. 南北方向 N-S direction
- 3. 东西方向 E-W direction
- 4. 东南-西北方向 ES-WN direction

表1为用球状模型拟合经验方差函数及相应 I_z 值。均值 $m=0.3$, 方差 $S^2=0.51$ 。从表1看到, 虽然使用同样的调查数据, 但按不同方向计算得到的 I_z 值有很大不同, 再一次说明了方向在研究空间分布型时有重要意义。

从表1的 a 值看, 南北方向的影响范围较小, 大约只有东西方向的一半。与上面结论一致。但从整体上看, a 值较大。可能存在着大范围的空间相关。这有待于进一步研究。

表1 方差函数的球状模型参数及扩散性指数 I_z 值
Table 1 Parameters in spherical model and values of I_z

参数 Parameter	c_0	c	a	L/a	I_z
全方位 All direction	0.376	0.135	39.781	0.320	0.461
南北方向 S-N direction	0.391	0.110	7.070	1.414	0.549
东西方向 E-W direction	0.505	0.007	14.140	0.707	0.233
东南-西北 ES-WN direction	0.376	0.135	30.150	0.420	0.340

关于分布型问题, 从 I_z 值看, 东西方向、南北方向间存在较大差别, 但两者都不能看作极端情况 0 或 1, 因此从整体上看, 研究区内松毛虫的分布可以说是随机的, 但东西、南北两方向的分布存在较大差异。南北方向有聚集分布的趋势。用传统的方法研究空间分布问题, 不考虑方向是不全面的。这一结果对研究种群扩散有一定的指导意义。

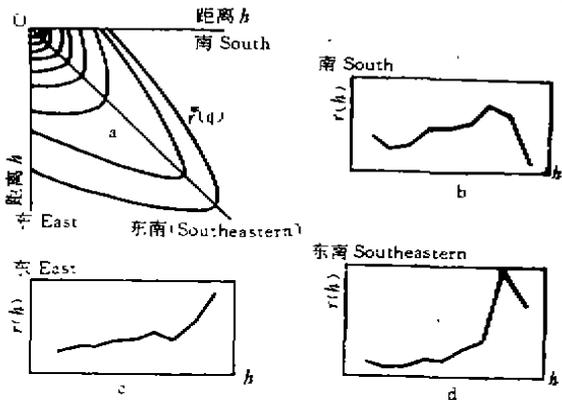


图3 等方差线图及各方向相关图

Fig. 3 Contour line of the vari-function and variogram in each direction

4 讨论

(1) 从大范围研究区总体上看, 松毛虫分布的趋势是随机的, 但在南北方向与东西方向有所差异。沿东西向较为均匀, 南北方向较为聚集, 聚集性不强。这可能是由于南北沿山坡走向, 高度变化大, 东西向高度相近, 松毛虫扩散性更强的原因。

(2) 扩散性指数 I_z 反映了分布的总体特征: 相关的范围有多大(与扩散距离有关), 相关的程度如何(与扩散速度及密度的变化幅度有关)。 I_z 从 0→1 变化时, 种群的空间分布从均匀分布→随机分布→聚集分布。 I_z 越接近 1, 聚集程度越高, 扩散性越强。

(3)文中给出 $r(h)$ 的简单表达式,如果把 $r(h)$ 分解为空间、时间两部分^[11],即可研究昆虫种群时空动态,它包括空间不同位置上种群数量随时间的变化。如果结合遥感监测结果,可以分析大范围虫灾扩散、蔓延和迁移的规律。为大面积虫灾管理提供有效的工具。

(4)国外文献中有用“Jackknife”方法作为选择判断空间内插好坏的标准,它的原理是比较观察值和理论值的误差,而在计算某点观察值的理论值时,把该点的数值删掉计算。

参 考 文 献

- 1 Iwao S. A new regression method for analyzing the aggregation pattern of animal populations. *Res. Popul. Ecol.* 1968, 10, 1—20
- 2 Iwao S. A new method of sequential sampling to classify populations relative to a critical density. *Res. Popul. Ecol.* 1975, 16, 281—288
- 3 徐汝梅,刘来福.改进的 Iwao, m^*-m 模型. *生态学报*, 1984, 4(2), 1—8
- 4 兰星平. m^*-v 模型在检验昆虫空间分布型和抽样调查中的应用. *林业科学*, 1991, 27(5), 511—516
- 5 于秀林,任朝佐.再次改进的 Iwao, m^*-m 模型. *生态学报*, 1986, 6(3), 193—196
- 6 周国法,李天生.应用对分法和样条法判定昆虫种群空间分布型. *林业科学研究*, 1992, 5(1), 26—31
- 7 李天生,周国法.昆虫种群距离聚集度指标的研究. *生态学报*, 1991, 11(4), 345—348
- 8 周国法,李天生.昆虫种群密度估计的空间自相关方法. *林业科学*, 1991, 27(4), 475—478
- 9 Liebhold A M, Elkinton J S. Characterizing spatial patterns of gypsy moth regional defoliation. *For. Sci.* 1986, 35, 557—568
- 10 Isaaks E H, Srivastava R M. *An Introduction to Applied Geostatistics*. Oxford Univ. Press NY 1989, 561
- 11 Bilonick R A. The space-time distribution of sulphate deposition in the northeastern United States. *Atmospheric Environ.* 1985, 19(11), 1829—1845

STUDIES ON THE SPATIAL AUTOCORRELATION AND THE INDEX OF SPATIAL DISTRIBUTION PATTERNS

Li Tiansheng

(Research Institute of Forestry, Chinese Academy of Forestry, Beijing, 100091)

Zhou Guofa

(Department of Mathematics, Branch Campus of Beijing University)

Some concepts of geostatistics, such as variogram and range, were introduced in this paper. A dispersal index I_z was suggested to describe the dispersal of population. When I_z varied from 0—1, population distribution would change from uniform → random → aggregation. The data of pine moth were taken as an example to show the application of these concepts. The results indicated that the spatial distribution was dependent on the spatial location and directions. When the interspecies spatial distribution of insects was studied, their spatial locations had to be taken into account.

Key words: geostatistics, variogram, range, dispersal index.