

点格局分析中边缘校正的一种新算法及其应用

时培建^{1,2}, 戈 峰^{1,*}, 杨清培³, 王建国²

(1. 中国科学院动物研究所, 北京 100101; 2. 江西农业大学农学院, 南昌 330045; 3. 江西农业大学林学院, 南昌 330045)

摘要:在点格局分析中,通常选取一个矩形作为研究区域,而 $K(d)$ 函数估计值的方差倾向于随着距离尺度的增加而增加。作为一种粗略的指导,距离尺度的最大值一般为矩形最小边长度的一半。在这种情况下,边缘校正的权重最小值为 0.25。通过在校正圆上等弧长取点,用校正圆上落在研究区域之内的点数除以整个校正圆上的点数,作为边缘校正权重的近似值。点数越多,这种近似算法越接近传统的精确算法。这种近似算法不仅适用于计算研究区域为矩形的边缘校正权重,而且适用于计算研究区域为任意多边形的边缘校正权重。此外,当矩形研究区域中点事件的信息不足时,这种算法可以允许计算接近到距离尺度的上限(即矩形对角线长度的一半)对应的 $K(d)$ 函数。

关键词:点格局; 距离尺度; 边缘校正; $K(d)$ 函数

文章编号:1000-0933(2009)02-0804-06 中图分类号:Q141, Q147 文献标识码:A

A new algorithm of the edge correction in the point pattern analysis and its application

SHI Pei-Jian^{1,2}, GE Feng^{1,*}, YANG Qing-Pei³, WANG Jian-Guo²

1 Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

2 College of Agronomy, Jiangxi Agricultural University, Nanchang 330045, China

3 College of Forestry, Jiangxi Agricultural University, Nanchang 330045, China

Acta Ecologica Sinica, 2009, 29(2): 0804 ~ 0809.

Abstract: In the point pattern analysis, the study region is generally chosen as a rectangle. Because the variance of the estimation of the $K(d)$ function tends to increase with the distance scale, its maximum is usually less than one-half the length of the shorter side of the rectangle at estimating the $K(d)$ function. In this case, the minimum of edge-corrected weight is proved to be 0.25. Then a new algorithm of edge-corrected weight is proposed in this paper. A number of points are drawn every an identical segment of the edged-corrected circle. The proportion of the number of points in the study region to the number of points in the whole edge-corrected circle, is approximately equal to the edge-corrected weight. Obviously, the larger the number of points is, the more accurate $K(d)$ function calculated with the algorithm is. With respect to the advantage of the algorithm, it can be applicable to estimating the $K(d)$ function when the study region is a rectangle or an arbitrary polygon. Furthermore, as the information of the point events in a rectangle is not enough, the algorithm can permit us to estimate the $K(d)$ function corresponding to the upper limit of the distance scale (i. e. one-half the length of the diagonal of the rectangle).

Key Words: point pattern; distance scale; edge correction; $K(d)$ function

基金项目:国家重点基础研究发展计划资助(973 计划)(2009CB119200);国家科技支撑计划资助项目(2006BAD08A07-3-2)

收稿日期:2007-09-28; 修订日期:2008-07-10

致谢:写作过程中得到张金屯教授和 Peter J. Diggle 教授的指导和帮助,江西农业大学植物学点 06 级研究生郭世权、杨青、李赞参与了野外调查,特此致谢。

* 通讯作者 Corresponding author. E-mail: gef@ioz.ac.cn

点格局分析是一种较新的植物种群空间格局分析方法。它通过划定一定面积的研究区域,以植物种群个体在二维空间上的坐标为基础数据,构成点图,再以点图为基础进行格局分析^[1,2]。点格局分析不仅对于研究植物的空间分布规律和传播扩散路径具有十分重要的意义,对于一些具有空间分布特征的动物,比如具有固定栖息地的昆虫、具有相对固定停歇地的鸟类,亦有重要的研究价值。本论文提出了对边缘校正的一种新算法,并进行了实际应用验证。

1 点格局分析的基本原理

研究区域内,位于二维空间的点构成点事件。用 $K(d)$ 函数来反映点事件空间布局的依赖程度^[3],其定义如下:

$$K(d) = \lambda^{-1} E(\#(d_{ij} \leq d)) \quad (1)$$

式中, λ 表示研究区域内每单位面积点事件的密度, E 表示在一定距离尺度下点事件数量的期望, $\#$ 表示数量, d_{ij} 表示研究区域内任意一点到其它点的距离, d 表示距离尺度。如果在一定距离尺度下点事件的分布是随机的,则有:

$$E(\#(d_{ij} \leq d)) = \lambda \pi d^2 \quad (2)$$

此时, $K(d) = \pi d^2$ 。当点事件为聚集分布时,如果距离尺度较小, $K(d)$ 的观测值将大于 πd^2 。公式(1)的估计式为:

$$\hat{K}(d) = \frac{A}{n^2} \sum_i^n \sum_j^n \frac{I_{ij}(d)}{w_{ij}(d)} \quad (3)$$

式中, A 表示研究区域面积, $\hat{\lambda} = n/A$, $\hat{E} = n^{-1} \sum_i^n \sum_j^n \frac{I_{ij}(d)}{w_{ij}(d)}$, n 是研究区域内点事件的数量, i, j 是样方内任意两个具有相同特征的点事件, $I_{ij}(d) = 1$ (当 $d_{ij} \leq d$ 时成立, d_{ij} 是 i 到 j 之间的距离), $w_{ij}(d)$ 是边缘校正的权重,等于以点事件 i 为圆心以 d_{ij} 为半径的圆落在研究区域内的弧长和整个圆周长的比值。

通常用 $L(d)$ 函数的估计值来判断在此尺度下点事件的分布类型。在距离尺度 d 下,如果 $\hat{L}(d) > 0$,为聚集分布;如果 $\hat{L}(d) = 0$,为随机分布;如果 $\hat{L}(d) < 0$,为均匀分布。

$$\hat{L}(d) = \sqrt{\hat{K}(d)/\pi} - d \quad (4)$$

一般利用 Monte Carlo 方法^[3]来求解随机分布的上下包迹线,如果研究对象实际的值落在上下包迹线的区间内,则在此距离尺度下植物种群是随机分布的。具体方法是选取 m 组 n 个随机数,每组都利用公式(4)求解在不同距离尺度下的 $L(d)$ 值,然后把对应的每个 d 下的 m 个 $L(d)$ 由小到大进行排序,分别保留最大值和最小值作为上下包迹线的取值范围,建议使用 99 组随机模拟值来确定包迹线的区间^[4]。

2 边缘校正的传统算法

对于研究区域是任意多边形的情况,因为边缘校正权重很难求出具体的数学表达式,所以一般把研究区域简化为矩形或圆形^[3]。在实际的研究中多是选取一定面积的矩形(本文称之为样方,包括正方形和长方形)作为研究区域。样地面积 $A = a \times b$, a 表示矩形较小边的长度, b 表示矩形较大边的长度。汤孟平等^[5]在 Hanus 等人^[6]研究的基础上,提出边缘校正的 4 种情况及其精确算法,即:

(i) 当圆完全包含在样方内时,即 $d_{ij} \leq d_{ib}$, 权重

$$w_{ij}(d) = 1 \quad (5)$$

(ii) 当圆与样方的一条边相交,即 $d_{ij} > d_{ib}$, 不等式 $d_{ix} < d_{ij}$ 与 $d_{iy} < d_{ij}$ 中有且只有一个成立, 权重

$$w_{ij}(d) = 1 - \left[\arccos\left(\frac{d_{ib}}{d_{ij}}\right) \right] \cdot \frac{1}{\pi} \quad (6)$$

(iii) 当圆与样方的两条相邻的边相交,即 $d_{ij} > d_{ib}$, 不等式 $d_{ix} < d_{ij}$ 与 $d_{iy} < d_{ij}$ 同时成立, 且 $\arccos\left(\frac{d_{ix}}{d_{ij}}\right) +$

$\arccos\left(\frac{d_{iy}}{d_{ij}}\right) \geq \frac{\pi}{2}$, 权重

$$w_{ij}(d) = 1 - \left[\arccos\left(\frac{d_{ix}}{d_{ij}}\right) + \arccos\left(\frac{d_{iy}}{d_{ij}}\right) + \frac{\pi}{2} \right] \cdot \frac{1}{2\pi} \tag{7}$$

(iv) 当圆与样方的两条相邻的边相交, 即 $d_{ij} > d_{ib}$, 不等式 $d_{ix} < d_{ij}$ 与 $d_{iy} < d_{ij}$ 同时成立, 且 $\arccos\left(\frac{d_{ix}}{d_{ij}}\right) +$

$\arccos\left(\frac{d_{iy}}{d_{ij}}\right) < \frac{\pi}{2}$, 权重

$$w_{ij}(d) = 1 - \left[\arccos\left(\frac{d_{ix}}{d_{ij}}\right) + \arccos\left(\frac{d_{iy}}{d_{ij}}\right) \right] \cdot \frac{1}{\pi} \tag{8}$$

式中, d_{ix} 是点事件 i 到样方长边的最近距离, d_{iy} 是点事件 i 到样方宽边的最近距离, d_{ib} 是样方内点事件 i 到最近邻边的距离, $d_{ib} = \min(d_{ix}, d_{iy})$, 其余符号与上文相同。

其实这种算法是在默认了距离尺度 $d \leq a/2$ (即距离尺度最大只能取到样方最小边长的一半), 但当距离尺度超过样方最小边长的一半时, 这种算法就不再适用。例如 $w_{ij}(d) \in (0, 0.25)$ 当时, 传统算法无法计算边缘校正的权重(图1)。

如果 $0 \leq d \leq a/2$, 则有 $0.25 \leq w_{ij}(d) \leq 1$ 。证明如下:

(i) $w_{ij}(d) = 1 \in [0.25, 1]$

(ii) $\because w_{ij}(d) = 1 - \left[\arccos\left(\frac{d_{ib}}{d_{ij}}\right) \right] \cdot \frac{1}{\pi}$

$\therefore \arccos\left(\frac{d_{ib}}{d_{ij}}\right)$ 越大, $w_{ij}(d)$ 越小

$\because 0 \leq d \leq a/2$, 不等式 $d_{ix} < d_{ij}$ 与 $d_{iy} < d_{ij}$ 中有且只有

一个成立

$\therefore 0 < d_{ij} \leq a/2$

又

$\because d_{ib} = \min(d_{ix}, d_{iy})$, 且 $d_{ib} < d_{ij}$

$\therefore 0 \leq d_{ib} < a/2$ (当第 i 点恰好落在样方一条边上时等号成立)

当 d_{ij} 一定时, $\arccos\left(\frac{d_{ib}}{d_{ij}}\right) \leq \arccos\left(\frac{0}{d_{ij}}\right) = \frac{\pi}{2}$, 于是

$w_{ij}(d) \in [0.5, 1] \subset [0.25, 1]$

(iii) $\because d_{ij} > d_{ib}, d_{ix} < d_{ij}, d_{iy} < d_{ij}$, 且 $\arccos\left(\frac{d_{ix}}{d_{ij}}\right) + \arccos\left(\frac{d_{iy}}{d_{ij}}\right) \geq \frac{\pi}{2}$

\therefore 当且仅当 $d_{ix} = d_{iy} = 0$ 时(第 i 点恰好落在样方的一个顶点上), $w_{ij}(d)$ 最小

$\min\{w_{ij}(d)\} = 1 - \left[\frac{\pi}{2} + \frac{\pi}{2} + \frac{\pi}{2} \right] \cdot \frac{1}{2\pi} = 0.25$

此时, $\arccos\left(\frac{d_{ix}}{d_{ij}}\right) + \arccos\left(\frac{d_{iy}}{d_{ij}}\right) = \pi$

$w_{ij}(d) \in [0.25, 1] \subset [0.25, 1]$

(iv) $\because \arccos\left(\frac{d_{ix}}{d_{ij}}\right) + \arccos\left(\frac{d_{iy}}{d_{ij}}\right) < \frac{\pi}{2}$

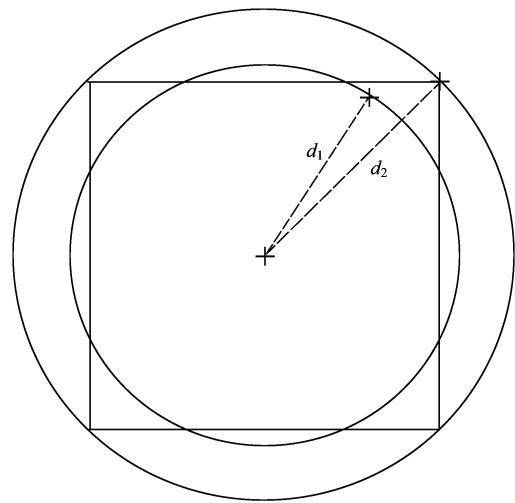


图1 边缘校正的传统算法所无法计算的情况

Fig.1 Case of the disability of the traditional algorithm to calculate the edge-corrected weight

$$\therefore w_{ij}(d) > 1 - \frac{\pi}{2} \cdot \frac{1}{\pi} = 0.5$$

$$w_{ij}(d) \in (0.5, 1) \subset [0.25, 1]$$

3 边缘校正的一种新算法

当距离尺度的取值范围为 $[0, a/2]$, 而边缘校正的权重的范围为 $[0.25, 1]$, 对于研究区域内任意两点的坐标是已知的, 以任意一点为圆心的圆方程可以计算得到。当在校正圆上等弧长取点时, 假设点数足够多, 落在样方内的点数除以整个校正圆上的点数的结果, 可以作为边缘校正权重的近似值。这种算法还可以计算研究区域为任意多边形的边缘校正权重。

3.1 新算法的原理

以研究区域内任意一点为圆心 i , 以它与其他任意一点 j 的距离为半径形成校正圆, 将整个圆的圆心角等分为 s 个等份, 圆上对应的点是均匀排列的。不过有些点落在研究区域内而有些点则落在研究区域外, 用落在研究区域内的点数除以整个圆上的点数, 其结果近似看作是边缘校正的权重 (图 2)。当等分的角度越小, 这种算法越精确:

$$w_{ij} = \frac{n_1}{n_1 + n_2} \quad (9)$$

式中, n_1 表示落在研究区域内的点数, n_2 表示落在研究区域外的点数。

3.2 新算法的语言表达

3.2.1 当研究区域为矩形时新算法的语言表达

```

counter = 0;
for  $\theta = 0: \frac{2\pi}{s}: (2\pi - \frac{2\pi}{s})$ 
if  $0 \leq x_i + d_{ij} \cos\theta \leq b$  and  $0 \leq y_i + d_{ij} \sin\theta \leq a$ 
    counter = counter + 1;
end
end
 $w_{ij} = \frac{\text{counter}}{s}$ 
    
```

算法中, counter 是计数器, 用来记录圆上落在样地内的点数, θ 代表圆上各点与起点构成的圆心角, s 代表在圆上选取点的总数, $2\pi/s$ 表示步长, (x_i, y_i) 代表圆心的坐标。

3.2.2 当研究区域为任意多边形时新算法的语言表达

对于研究区域是任意多边形的情况 (图 3), 首先需要利用各条边上的点拟合出每条边的函数表达, 然后利用下述算法求出边缘校正:

```

counter = 0;
for  $\theta = 0: \frac{2\pi}{s}: (2\pi - \frac{2\pi}{s})$ 
if  $z_k \leq x_i + d_{ij} \cos\theta < z_{k+1}$  and  $f_{k1}(\cdot) \leq y_i + d_{ij} \sin\theta \leq f_{k2}(\cdot)$ 
    counter = counter + 1;
end
end
 $w_{ij} = \frac{\text{counter}}{s}$ 
    
```

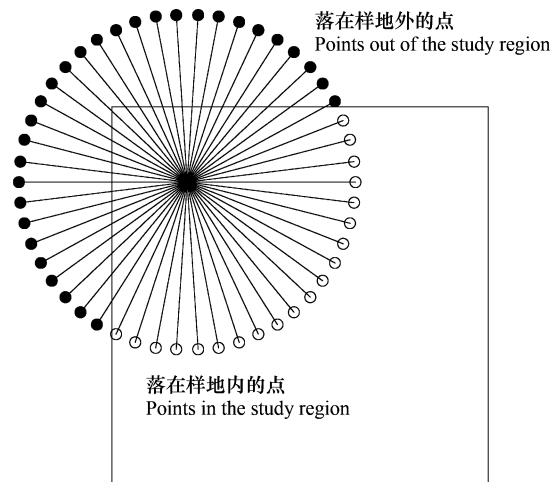


图 2 边缘校正新算法的原理图

Fig. 2 Principle of the new algorithm of the edge correction

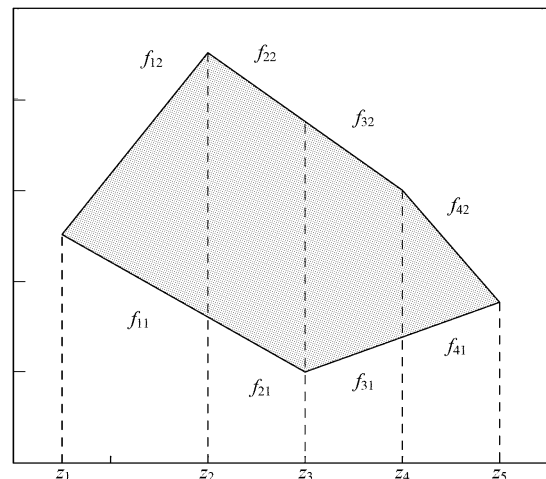


图 3 研究区域为任意多边形的示例

Fig. 3 Illustration that the study region is an arbitrary polygon

```

    counter = counter + 1;
end
end

$$w_{ij} = \frac{\text{counter}}{s}$$


```

算法中, $[z_k, z_{k+1})$ 表示任意多边形交点相邻横坐标构成的区间。当 k 为多边形交点横坐标的次最大值时, 区间为完全闭区间。 $f_{k1}(\cdot)$ 和 $f_{k2}(\cdot)$ 分别为区间 $[z_k, z_{k+1})$ 对应的较小的函数和较大的函数, “ \cdot ”表示 $x_i + d_{ij} \cos\theta$ 。对于任意多边形的 Monte Carlo 的模拟具有一定的难度, 此处给出了一种计算机的算法:

```

X = range × rand(1, n);
for q = 1:n
if  $z_k \leq X(q) < z_{k+1}$ 
     $Y(q) = f_{k1}(\cdot) + [f_{k2}(\cdot) - f_{k1}(\cdot)] \times \text{rand}(1, 1)$ 
end
end
end

```

算法中, 假设任意多边形中共有 n 个点事件, X 表示随机生成点的横坐标数组, range 表示研究区域横坐标的最大值, rand(1, n) 表示生成一行 n 列介于 0 到 1 之间的随机数, $X(q)$ 表示第 q 个随机点的横坐标, $Y(q)$ 表示第 q 个随机点的纵坐标, “ \cdot ”表示 $X(q)$, rand(1, 1) 表示生成一个介于 0 到 1 之间的随机数, (X, Y) 表示生成的 n 个随机点的数组。

4 新算法的应用实例

2007 年 6 月初, 在江西省新建县上罗村(北纬 $28^{\circ}45.378'$, 东经 $115^{\circ}49.625'$), 对一块 $5 \times 5\text{m}^2$ 样方中一年蓬 (*Erigeron annuus*) 的空间分布 ($n = 204$) 进行了调查; 同时, 还对一块 $5 \times 5\text{m}^2$ 样方中豚草 (*Ambrosia artemisiifolia*) 的空间分布 ($n = 276$) 进行了调查。对调查的数据使用 MATLAB7.0 编程计算 $L(d)$ 。

在一年蓬的分析中, 距离尺度的范围为 0 到 2.5m, 步长为 0.1m; 在豚草的分析中, 距离尺度为 0 到 1.7m, 最大尺度在最长边的 $1/3$ 附近^[7], 步长为 0.1m。从图 4 可以看出, 新算法在计算一年蓬的 $L(d)$ 函数时和传统算法近乎一致。

显然, 如果研究区域为矩形时, 新算法其实是传统精确算法的一种近似算法, 在校正圆上取点越多, 这种近似算法越接近传统精确算法, 尤其是当 A/n^2 很小时, 所产生的误差在计算 $K(d)$ 或 $L(d)$ 时是可以忽略的。

其实对新算法中取点多少反映最敏感的是 $n \times \hat{E}(d)$; 但是新算法可以应用于计算研究区域为任意多边形的边缘校正权重, 这是传统算法无法作到的。

另外, 当研究区域为矩形时, 一般而言扩大距离尺度可以避免边缘校正的计算, 对于许多植物而言, 这种方法是可行的。但是当点事件的信息不充足时, 扩大距离尺度计算的 $K(d)$ 函数可能只是反映了分布区的点事件异质性变化, 而不代表点事件在同质环境下的扩散规律。例如调查的豚草空间分布, 在图形下方的分布稀少主要是因为当地的农民将牛粪堆积于此, 牛粪所在区域基本没有豚草分布, 这实际上反映的是空间的异质性, 而不是豚草的自然扩散的分布格局。此时在现有点事件记录的研究区域内进一步扩大距离尺度 $K(d)$ 变得更加有意义, 新算法可以允许计算距离尺度接近样方对角线长度一半时的边缘校正权重(因为只要边缘校正权重大于零, 对应的距离尺度函数的计算就仍有意义)。虽然这会进一步增加 $K(d)$ 估计值的方差, 但是这对于在尽可能大的距离尺度下分析某些地理分布稀少的植物的空间格局, 具有一定的参考价值。当 $w_{ij} \in (0, 0.25)$ 时, 新算法仍然可以用来计算 $K(d)$ 函数。

新算法理解起来虽然简单, 应用范围较广, 但是当步长间隔过小时, 程序运行不太经济, 占用时间长, 新算法对计算机硬件的运行速度有较高的要求。

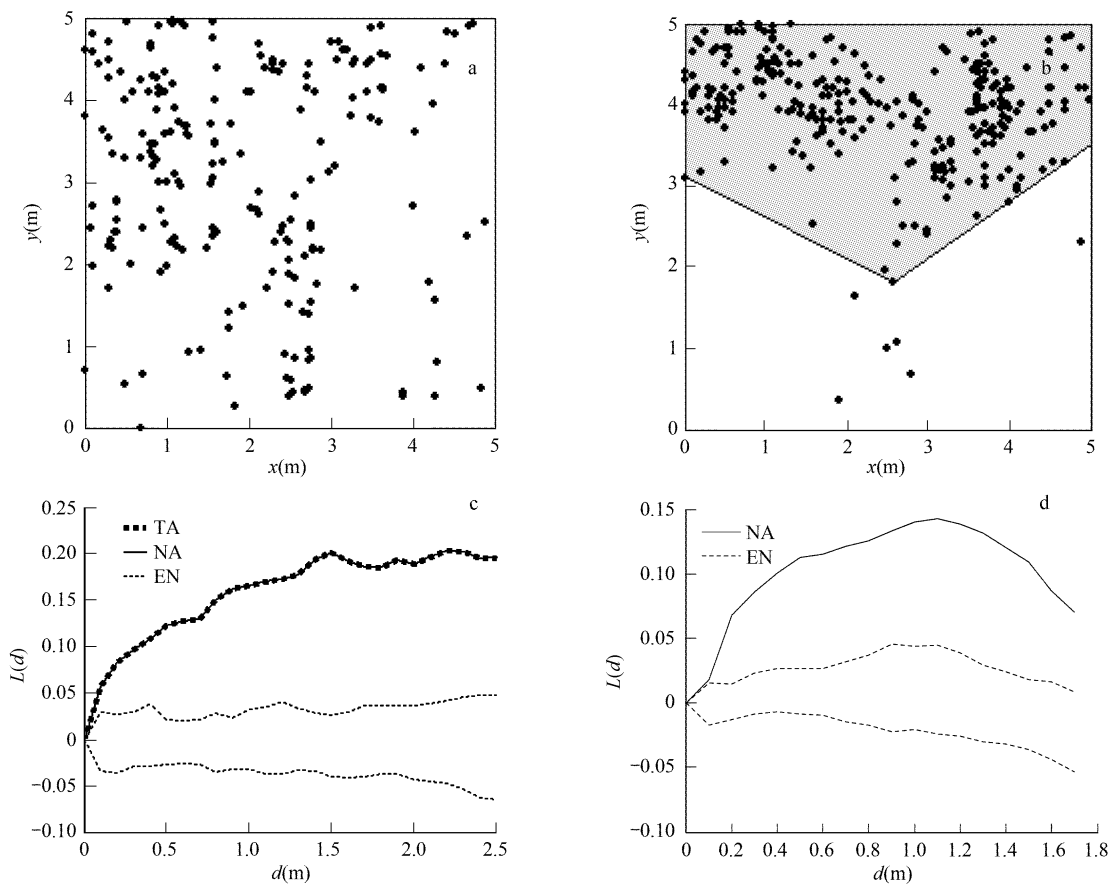


图4 植物的空间分布及函数

Fig.4 Spatial distributions of vegetation and their functions

a. 一年蓬的空间分布; b. 豚草的空间分布; c. 传统算法和新算法在计算一年蓬的函数的比较, d. 用新算法计算任意多边形研究区域内豚草的函数示例, 阴影部分代表研究区域; TA 代表传统算法; NA 代表新算法; EN 代表包迹线

a. Spatial distribution of Annual fleabane; b. Spatial distribution of Ragweed; c. Comparison between the traditional algorithm and the new algorithm on estimation the $L(d)$ functions of Annual fleabane; d. Illustration using the new algorithm to estimate $L(d)$ functions of Ragweed in an arbitrary polygon, and the shaped region represents the study region; TA denotes the traditional algorithm; NA denotes the new algorithm; EN denotes envelopes

References:

- [1] Diggle P J. Statistical analysis of spatial point patterns (second edition). London: Edward Arnold, 2003.
- [2] Zhang J T. Analysis of spatial point pattern for plant species. Acta Phytocologica Sinica, 1998, 22(4): 344—349.
- [3] Manly B F J. Randomization and monte carlo methods in biology. New York: Chapman and Hall, 1991.
- [4] Gatrell A C, Bailey T C, Diggle P J, Rowlingson B S. Spatial point pattern analysis and its application in geographical epidemiology. Transactions of the Institute of British Geographers, New Series, 1996, 21(1): 256—274
- [5] Tang M P, Tang S Z, Lei X D, et al. Edge correction of Ripley's $K(d)$ function on population spatial pattern analysis. Acta Ecologica Sinica, 2003, 23(8): 1533—1538.
- [6] Hanus M L, Hann D W, and Marshall D D. Reconstructing the spatial pattern of trees from routine stand examination measurements. Forest Science, 1998, 44(1): 125—133.
- [7] Rowlingson B S, Diggle P J. Splancs: spatial point pattern analysis code in S-plus. Computers in Geosciences, 1993, 19(5): 627—655.

参考文献:

- [2] 张金屯. 植物种群空间分布的点格局分析. 植物生态学报, 1998, 22(4): 344~349.
- [5] 汤孟平, 唐守正, 雷相东, 等. Ripley's $K(d)$ 函数分析种群空间分布格局的边缘校正. 生态学报, 2003, 23(8): 1533~1538.