

用 PFU 法研究微型生物群集过程中 数据的处理

王继忠 袁育才* 沈韞芬

(中国科学院水生生物研究所, 武汉)

提 要

根据 MacArthur-Wilson 的岛屿区系平衡模型 $S_t = S_{eq}(1 - e^{-GT})$, 可以从野外生态效应试验和室内毒性试验中, 提出 3 个功能参数 (S_{eq} 、 G 、 $t_{90\%}$) 进行比较。本文提出两种计算方法: 复合梯形法和最小二乘法, 后者已在计算机上实现了 BASIC 计算程序。从数学理论上论证, 最小二乘法误差较小, 但如果实验布局合理, 两种计算方法能得到十分一致的结果。实验模型是否符合理论模型, 可以用统计学上的拟合差异度检验法来检验。

水生态系中各种生物和非生物因子处在十分复杂的、相互作用的网络内。如果能通过正确的方法处理实验中获得的各种数据, 就能正确掌握水生态系的结构与功能。微型生物群落是水生态系中的一个亚级单位, 是目前生态学家用以研究群落级结构与功能的对象之一。Cairns 等 (1969、1979)^[5,6] 用聚氨酯泡沫塑料块 (Polyurethane Foam Unit, 简称 PFU) 方法研究原生动物的群集过程, 认为符合 MacArthur-Wilson 1963^[8] 的岛屿区系平衡模型。沈韞芬等 (1985)^[1] 已把 PFU 法应用到生物监测中, 并在 1983、1984 年举办两期“微型生物群落在生物监测中的应用”培训班。来自全国 21 个省、市的环境生物工作者们就实验数据的处理问题进行了讨论。作者之一曾与日本京都大学寺本英教授进行了探讨。本文集生物、数学、计算机三方面的专家和专业人员的知识, 提出两种计算方法 (复合梯形法和最小二乘法) 供读者在应用 PFU 法时使用。

计 算 方 法

(一) MacArthur-Wilson 的岛屿区系平衡模型

MacArthur 和 Wilson (1963)^[8] 在岛屿生物地理分布的理论中提出种类的多样性是迁入过程和消失过程之间平衡的结果。Wilson 和 Simberloff (1969、1970)^[9,10] 用烟薰的办法把 4 个岛上的鸟赶走, 隔 1—2 年后分别记录 4 个岛上鸟的迁入种类和消失种类, 迁入速度随时间而下降, 消失速度随时间而上升, 两条曲线交叉时, 就是群落达到了平衡。Cairns 等认为如果把 PFU 看作为一个荒芜小岛, 原生动物的群集过程也符合这

* 在中国科学院武汉分院工作。
1984年12月30日收到。

一模型^[6]。

设 I 为迁入种数; E 为消失种数; S_t 、 S_{eq} 、 S^* 分别为 t 时、平衡时和种库的种数; k_1 、 k_{-1} 分别为迁入速度和消失速度的常数, $k_1 + k_{-1} = G$ 。设 PFU 内微生物的入侵速度与 $(S^* - S_t)$ 成正比, 消失速度与 S_t 成正比, 则

$$\frac{dI_t}{dt} = k_1(S^* - S_t) \quad (1)$$

$$\frac{dE_t}{dt} = k_{-1}S_t \quad (2)$$

而群集速度为

$$\frac{dS_t}{dt} = \frac{dI_t}{dt} - \frac{dE_t}{dt} \quad (3)$$

代(1)(2)式于(3)式得

$$\frac{dS_t}{dt} = k_1S^* - (k_{-1} + k_1)S_t = k_1S^* - GS_t \quad (4)$$

当群集过程达到动态平衡时, 种数对时间的变化速度应为零, 即 $\frac{dS_t}{dt} = 0$, 此时平衡种数 S_{eq} 根据公式(4)写成:

$$S_{eq} = \frac{k_1S^*}{G} \text{ 或 } k_1S^* = S_{eq}G \quad (5)$$

将(5)式代入(4)式得:

$$\frac{dS_t}{dt} = GS_{eq} - GS_t = G(S_{eq} - S_t) \quad (6)$$

(6)式也可写成:

$$\frac{dS_t}{S_{eq} - S_t} = G dt$$

$$S_t = C e^{-Gt} + S_{eq} \quad (C \text{ 为积分常数})$$

当 $t = 0$ 时, $S_t = 0$, $C = -S_{eq}$, 微分方程(6)式满足初值的特解为:

$$S_t = S_{eq}(1 - e^{-Gt}) \quad (7)$$

此(7)式便是 MacArthur-Wilson 的平衡模型公式。

(二) 群集过程中参数的估计

Cairns 等^[5]首次将 MacArthur-Wilson 的平衡模型公式应用于原生动物的 PFU 上的群集过程中, 并提出群落级的 3 个功能参数—— S_{eq} 、 G 、 $t_{90\%}$ 。 $t_{90\%}$ 是指达到 90% 平衡种数所需要的时间, 可由公式(7)中求得。

$$(1 - e^{-Gt}) = 90\%, \quad e^{-Gt} = 0.10, \quad -Gt = \ln 0.10$$

$$t_{90\%} = \ln 0.1 \times (-G^{-1}) = \frac{2.303}{G} \quad (8)$$

在 PFU 微生物群集过程的实验中, 一般 PFU 的曝露天数为 1, 3, 6, 11, 15, 30 天, 依生境而异。每次收集的 PFU 挤出液中, 总有首次遇到的种类, 称为新见种 (N)。

新见种有 3 种趋向——每次都出现;断断续续地出现;从此消失。我们把断续出现的种类叫复见种 (R)。1983 年的培训班学员们经过讨论认为复见种必须符合 3 个条件:(1)上次采样中未遇见,(2)以前(不包括上次)遇见过,(3)本次复见。因此迁入种数 $I = N + R$, 消失种数 $E = (I + S_{n-1}) - S_n$, 分别除以曝露天数, 即得迁入速度和消失速度。种库 S^* 是每次新见种之和。

如何从上述这些实验数据中求 k_{-1} 和 k_1 , Gilroy (1975)^[7] 提出 4 种理论上的计算方法。现讨论这 4 种方法在实际应用中的问题。第一种方法先直接求 G 值。由 (7) 式可得:

$$\ln \frac{S_{eq}}{S_{eq} - S_t} = Gt, \quad G = \frac{1}{t} \ln \frac{S_{eq}}{S_{eq} - S_t} \quad (9)$$

以 t 作为自变量, $\ln \frac{S_{eq}}{S_{eq} - S_t}$ 作为因变量作相关分析。这方法的缺点是:(1)微型生物群集过程中平衡时间不长时, S_{eq} 值可直接测知。但平衡时间较长时, S_{eq} 的值只能估计, 误差大;(2)由于原生动植物群落必须进行即席活体观察以作正确的分类鉴定, 时间上受到限度, 因此采样次数不可能很多。获得的 S_t 值次数不多。因此用 (9) 式来估算 G 值的置信度不高。

第二种方法是 k_1 的极限求法。当 $t \rightarrow 0$ 时, $S \rightarrow 0$; 当 $t \rightarrow \infty$ 时, $S \rightarrow S_{eq}$ 。此时 (1) 式分别为:

$$\left(\frac{dI}{dt}\right)_{t \rightarrow 0} = k_1 S^*, \quad \left(\frac{dI}{dt}\right)_{t \rightarrow \infty} = k_1 (S^* - S_{eq}),$$

则

$$k_1 = \frac{\left(\frac{dI}{dt}\right)_{t \rightarrow 0} - \left(\frac{dI}{dt}\right)_{t \rightarrow \infty}}{S_{eq}} \quad (10)$$

$\left(\frac{dI}{dt}\right)_{t \rightarrow 0}$ 和 $\left(\frac{dI}{dt}\right)_{t \rightarrow \infty}$ 在实验中都不可能实现, 只能用 S_t 的最初值和最终值来代替。加之 S_{eq} 仍然是估计值, 置信度降低。此法不宜采用。

第三种方法可以避免对 S_{eq} 值的估计法, 即用两次邻近的 S_t 值, 然后用二次方程式解 G 值。如

$$\begin{aligned} S_{t_1} &= S_{eq}(1 - e^{-Gt_1}) \\ S_{t_2} &= S_{eq}(1 - e^{-Gt_2}) \end{aligned}$$

则

$$\frac{S_{t_1}}{S_{t_2}} = \frac{1 - e^{-Gt_1}}{1 - e^{-Gt_2}} \quad (11)$$

当 $\frac{t_2}{t_1}$ 不为整数或较大 (如 $\frac{t_2}{t_1} \geq 5$) 时, 就很难从 (11) 式中解出 G 值。如果在实验中碰到例外情况, 如 $S_{t_2} < S_{t_1}$ 时, 就不符合数学模型, 缺乏自动校正的手段, 求出的 G 值不可靠。

第四种方法是复合梯形法, 比较适用。此法的优点在于它有一定的自动补偿作用, 置信度较高。对 (1) (2) 式进行积分, (1) 式为:

$$\sum_n I_n = k_1 S^* t_n - k_1 \int_0^t S dt \quad (12)$$

(2) 式为:

$$\sum_n E_n = k_{-1} \int_0^t S dt \quad (13)$$

采用梯形法得到积分计算的近似公式

$$\int_0^t S dt = \sum_n \frac{(t_n - t_{n-1})(S_n + S_{n-1})}{2}$$

代入 (12) (13) 式得:

$$k_1 = \frac{\sum_n I_n}{S^* t_n - \sum_n \frac{(t_n - t_{n-1})(S_n + S_{n-1})}{2}} \quad (14)$$

$$k_{-1} = \frac{\sum_n E_n}{\sum_n \frac{(t_n - t_{n-1})(S_n + S_{n-1})}{2}} \quad (15)$$

从 PFU 实验原始数据中获得的 S_n 、 S_{n-1} 、 I 、 E 等值代入 (14) (15) 式就获得 k_1 和 k_{-1} 值。再代入 MacArthur-Wilson 的平衡模型公式 (7) 和公式 (8) 中, 就可求得微型生物群集过程中的 3 个功能参数—— G 、 S_{eq} 、 $t_{90\%}$ 。由于采用了复合梯形法计算积分 $\int_0^t S_t dt$, 其近似误差为 $\alpha = \frac{-t\dot{h}^2}{12} S''^{[2]}$ 。但因为 $S'_t = -S_{eq} G^2 e^{-Gt} < 0$, 所以 $\alpha > 0$, 即用复合梯形法计算上凸函数 S_t 的积分值比真值偏小, 而凹函数 S_t 的积分值比真值偏大。如果群集曲线愈接近线性, 则近似程度愈好。当然, 在实际中是不可能的。于是我们认为有必要建立用最小二乘法来估计这 3 个参数, 把误差降低到最小的限度。

设 SS_n 为实验进行中 t_n 时的总种数, 则总误差

$$SR \triangleq \sum_n (SS_n - S_n)^2$$

代入 (7) 式,

$$SR \triangleq \sum_n [SS_n - S_{eq}(1 - e^{-Gt_n})]^2 \quad (16)$$

SR 称为最小二乘目标函数。实际上 SR 是计算值 S_n 和实验值 SS_n 的总误差平方和。 SR 是参数 (G 、 S_{eq}) 的二元函数, 不能够用初等方法求其极小值。本文选用非线性稳态模型的参数估计法之一——Marquardt 法^[3]来估计参数 G 和 S_{eq} 。当然, 也可用其它方法, 如单纯形搜索法、变尺度法、Gauss 法^[3]...等。我们在 Sharp PC-1500 袖珍计算机和 TRS-80 微型计算机上实现了 BASIC 计算程序^[3]。利用此参数估计程序可得到最小二乘指标下的参数估计值 \hat{G} 和 \hat{S}_{eq} 。代入 (7) 式, 便可得到一试验模型, 至于其模型是否适宜 MacArthur-Wilson 公式, 可用统计学上的拟合差异度检验法 (Lack of Fit Test 简写 LOF) 来检验^[4]。

LOF 法——设有几组试验数据的集合 A ,

$$A = \{(t_i S_i) / i = 1, 2, \dots, n\}.$$

现将 A 分割成 k 个子集合

$$A_j (j = 1, 2, \dots, k),$$

记

$$A_j = \{(t_{ij} S_{ij}) / t_{1j} = t_{2j} \dots = t_{n_j j}\},$$

并且:

$$\sum_{j=1}^k n_j = n$$

$$\sum_{j=1}^k A_j = A$$

$$k < n \quad A_j \neq \phi$$

令

$$\bar{S}_{\cdot j} \triangleq \frac{1}{n_j} \sum_{i=1}^{n_j} S_{ij}$$

$$SE_j \triangleq \sum_{i=1}^{n_j} (S_{ij} - \bar{S}_{\cdot j})^2 = \sum_{i=1}^{n_j} S_{ij}^2 - n_j \bar{S}_{\cdot j}^2$$

则实验的纯误差

$$SE = \sum_{j=1}^k SE_j$$

又因为 SE_j 的自由度为 $n_j - 1$, 故 SE 的自由度为:

$$\sum_{j=1}^k (n_j - 1) = n - k$$

记实验平均误差

$$ME = \frac{SE}{n - k}$$

因为总误差 $SR = \sum_{i=1}^n (SS_i - \hat{S}_i)^2$ 的自由度为 $n - 2$, 所以拟合误差 $SL = SR - SE$ 的自由度为:

$$n - 2 - (n - k) = k - 2$$

记平均拟合误差

$$ML = \frac{SL}{k - 2}$$

由数理统计知:

$$F \triangleq \frac{SL / (k - 2)}{SE / (n - k)} = \frac{ML}{ME} \quad (17)$$

则 F 服从自由度为 $(k - 2, n - k)$ 的 F -分布。因为 SE 表示数据本身的误差大小, 如果 SR 增大, 即表明拟合误差增大, 此时 F 值增大。于是取显著性水平 α , 如果 $F > F_\alpha(k -$

2, $n - k$), 则称数据 (t_i, S_i) 对模型的拟合有显著性差异; 也就是说数据不符合选定参数所对应的模型。但因为参数选择的最优性, 一般便可以认为数据不符合此模型。如果 $F < F_{\alpha}(k - 2, n - k)$, 则称数据对模型有较好的拟合, 即符合此模型。

(三) 举例示范

根据对鸭儿湖 IV 号氧化塘室内生物测试所得原生动动物群集过程的下列数据:

PFU 曝露天数	1		3		6		9		13	
	A	B	A	B	A	B	A	B	A	B
总种数 S	12	10	16	19	24	24	24	26	39	22
新见种数 N	12	10	12	6	16	14	12	8	19	8
复见种数 R					1	1	3	4	6	3

分别用复合梯形法和最小二乘法求 S_{eq} 、 G 、 $t_{90\%}$ 3 个功能参数值计算, 并用 LOF 法检验, 获得结果如下:

	复合梯形法	最小二乘法
S_{eq}	28.21	28.4
G	0.34	0.348
SR	191.287	190.406
$t_{90\%}$	6.77	6.61
F	0.3228	0.315
$F_{\alpha}(k - 2, n - k)$	16.7	16.7
LOF 检验	符合模型	符合模型

讨 论

1. 用最小二乘法估计微型生物群集过程的 3 个功能参数 (S_{eq} 、 G 、 $t_{90\%}$), 求得的结果较复合梯形法正确。因为最小二乘法的立足点是取在最小二乘意义下的最优参数 (\hat{G} 、 \hat{S}_{eq}), 它所对应的误差函数 SR 是最小的, 所以有理由认为最小二乘法要比复合梯形法更为精确。

2. 复合梯形法中要援引新见种数、复见种数、迁入种数和消失种数。这就要求种类鉴定必须正确无误。最小二乘法只援引总种数, 不要求迁入种数和消失种数。这不仅减少工作量, 而且可以减少误差。复合梯形法可能由于较早地引入了误差(因为在区别新见种类还是复见种类时, 由于工作人员掌握分类知识的程度不同而造成人为的误差), 并随着误差的传递, 最后求出的 3 个参数值就可能有较大的误差。而最小二乘法是一次引入误差, 它避免了误差在传递过程中增大的谬误, 而且将计算误差在最小二乘意义下降到最小限度。

3. 本文举的例子显示出最小二乘法和复合梯形法的结果十分接近。但由于在 LOF 检验中, 复合梯形法对应下的 F 值会大于最小二乘法对应下的 F 值。这时就可能出现数据符合最小二乘法计算的参数所对应的模型, 而不符合复合梯形法计算的参数所对应的模

型。可是这种例外是允许而又正常的。因为 LOF 检验是对一组特定参数所对应的模型进行检验。可以设想, 如果改进复合梯形法, 求得更接近真实的模型参数, 数据就有可能符合这组改进参数所对应的模型。自然, 它所对应的 SR 不会小于最小二乘法对应的 SR 。

4. 在群落生态学的研究中, 群集过程 (Colonization process) 反应出群落的一种功能现象, 它从一个角度窥知群落内繁多有机体的生命活动。群集过程中的迁入速度和消失速度是指整个群落而言。当两种速度相等时, 群落就达到平衡状态。MacArthur-Wilson 的平衡模型公式 $S_t = S_{eq}(1 - e^{-Gt})$ 已被证明符合于鸟类和原生动物的群集过程。如果环境起了变化, 群落的平衡被破坏, 群落内的有机体要进行种种生命调节以适应这种变化, 其中也包括群集过程的调节。从理论上和实践上都已证明, 用 PFU 法进行原生动物的群集过程的方法能很好地对环境污染进行监测和预报。 S_{eq} 、 G 、 $t_{90\%}$ 3 个参数能反映水体从污染源到恢复正常的系列变化。但是, 用统计学上 P 值 95% 时, 有时 LOF 检验表明不符合平衡模型公式, 尤其是在污染的环境中。此外, 在流水中达到平衡所需的时间很短 (1—数天), 理想的 PFU 取样时间为 1, 2, 4, 8, 16, 24, 48, 72 小时。但在实际工作中有许多困难。于是 LOF 检验时, 也就难以符合平衡理论模型。除了在工作中设法把采样间隔时间缩短 (如增加 12 小时的 PFU) 以期减少误差外, 是否可以设想出比 MacArthur 平衡模型更好的数学模型, 还有待于今后的进一步研究。

参 考 文 献

- [1] 沈韞芬等, 1985。用 PFU 原生动物的群集进行生物监测的研究。本刊本期。
- [2] 张德荣等, 1983。计算方法与算法语言。327 页。人民教育出版社。
- [3] 席少霖等, 1983。最优化计算方法。461 页。上海科学技术出版社。
- [4] Bethea, R. M., Duran, B. S. and T. L. Boullion, 1975. Statistical methods for engineers and scientists. Marcel Dekker. Inc., pp. 583. New York.
- [5] Cairns, J. Jr., Dahlberg, M. L., Dickson, K. L., Smith, N. and W. T. Wallar, 1969. The relationship of freshwater protozoan communities to the MacArthur-Wilson equilibrium model. *Amer. Nat.*, **103**: 439—454.
- [6] Cairns, J. Jr., Dickson, K. L. and J. L. Plafkin, 1979. Protozoan colonization of artificial substrates. In Weitzel, R. L. ed. Methods and measurements of attached microcommunities, pp. 39—57.
- [7] Gilroy, O., 1975. The determination of the rate constants of island colonization, *Ecology*, **56** (4): 915—923.
- [8] MacArthur R. H. and E. O. Wilson, 1963. An equilibrium theory of insular zoogeography. *Evolution*, **17**(4): 373—387.
- [9] Simberloff, D. S. and E. O. Wilson, 1970. Experimental zoogeography of islands. A two-year record of colonization. *Ecology*, **51**(5): 934—937.
- [10] Wilson, E. O. and D. S. Simberloff. 1969. Experimental zoogeography of islands. Defaunation and monitoring techniques. *Ecology*, **50**(2): 267—278.

DATA HANDLING IN STUDYING THE PROCESS OF PROTOZOAN COLONIZATION BY MEANS OF PFU METHOD

Wang Jizhong Yuan Yucui* and Shen Yunfen
(*Institute of Hydrobiology, Academia Sinica, Wuhan*)

Abstract

Through colonizing microbial communities by means of the polyurethane foam unit (PFU), we can evaluate water quality at the community level, and predict the ecological effects of waste water. According to the MacArthur-Wilson equilibrium model, three function parameters (S_{eq} , G and $t_{50\%}$) were brought forward in the field ecological effect test and the laboratory toxicity test. In this article, the compound trapezoidal method and the least square method are used for calculating the parameters. A BASIC program of the least square method has been made for the microcomputer TRS-80 and Sharp PC-1500. Either method can be used if an experiment is designed properly. The lack of fit test (LOF) can be applied to verify whether an experimental model is adapted to the theoretical model.

Key words model, parameter calculation, program, LOF test

* Wuhan Branch of Academia Sinica.