

粒子群优化算法拟合 Logistic 方程的研究

袁利国, 聂笃宪

(华南农业大学 理学院, 广东 广州 510642)

摘要:提出了一种新的拟合 Logistic 方程与估计其参数的方法. 首先利用三段和法对参数进行初步估计, 得到参数的粗略值, 然后利用粒子群优化算法进行寻找最优参数, 并对所得结果进行统计检验. 结果表明三段和法与粒子群优化算法相结合具有很强的拟合 Logistic 方程的能力. 此方法可推广应用到对其他非线性生长曲线方程的拟合.

关键词:粒子群优化算法; Logistic 方程; 参数估计; 三段和法

中图分类号: TN911.26

文献标志码: A

文章编号: 1001-411X(2011)03-0114-05

Study on Fitting Logistic Equations with Particle Swarm Optimization Algorithms

YUAN Li-guo, NIE Du-xian

(College of Sciences, South China Agricultural University, Guangzhou 510642, China)

Abstract: A new method was developed for fitting of Logistic equations and estimating the parameters. Rough estimation of the parameters was conducted first with three-sum methods, and then the optimum parameters were found by particle swarm optimization algorithm (PSO). Statistical test was applied to the data. The results showed that the combined method of three-sum method and particle swarm optimization algorithm can fit the Logistic equations well. This method can be generalized to fit other nonlinear growth curve equations.

Key words: particle swarm optimization; Logistic equation; estimating the parameters; three-sum method

连续的 Logistic 方程是种群增长规律的一个经典数学模型, 在生态学、生物医学、人口统计学、经济学等方面都有广泛应用, 其微分方程形式可记为

$$\frac{dy(t)}{dt} = by(t) \left(1 - \frac{y(t)}{K} \right), \quad (1)$$

它描述了有限空间内种群增长规律, 其中 K 为环境容纳量, $y(t)$ 为 t 时刻的种群数, b 为种群内禀增长率. 参数 K, b 是生物进化对策论中的重要概念. 方程

(1) 的积分形式为 $\ln \left| \frac{K - y(t)}{y(t)} \right| = a - bt$, 或

$$y(t) = \frac{K}{1 + e^{a-bt}}, \quad (2)$$

记 $e^a = A$, 则方程(2)可写成

$$y(t) = \frac{K}{1 + Ae^{-bt}}, \quad (3)$$

其中 K, A (或 a)、 b 为待估计参数^[1-8].

对 Logistic 方程的拟合一直引起生物与数学工作者的广泛兴趣, 被研究并应用于处理拟合试验数据, 但是由于是非线性函数, 且不能线性化, 因此对其参数的估计非常困难. 到现在为止, 已有一些经典的方法^[1-8], 本文在这些方法的基础上, 提出一种新的拟合方法——三段和法与粒子群优化算法 (PSO) 相结合的方法, 这样可解决参数初值范围较难确定的问题, 并运用到拟合草履虫 (双小核草履虫 *Paramecium aurelia* 和大草履虫 *P. caudatum*) 试验数据 (表 1)^[9]. 所得结果表明, 此方法具有很高的拟合精度, 且收敛速度快, 易实现. 经统计检验, 实测数据与所得预测数据相关系数高, 差异不显著, 此方法与编制的程序可推广到对其他非线性生长模型的

收稿日期: 2010-01-26

作者简介: 袁利国 (1979—), 男, 讲师, 博士研究生, E-mail: liguoy@scau.edu.cn

基金项目: 国家青年基金 (60702030); 华南农业大学校长基金 (2008K011, 2009K023)

拟合.

1 方程参数的初步估计(三段和法)

Logistic 方程 3 种常规的参数估计方法:三段和法、三点法、差分法. 本文先采用三段和法^[4]对方程(3)中的参数 K, A, b 的值做一个粗略的估计,以它作为初始值,再利用粒子群优化算法对方程(2)中的参数 K, a, b 进行更进一步的寻优.

将容量为 n 的样本 $\{(x_i, y_i), i = 0, 1, \dots, n\}$ 按时间顺序排列后,再等分成 3 组. 令 $r_1 = \text{int}((n + 1)/3)$ (其中 $\text{int}()$ 表示取整函数),如果 $n = 3r_1$,则舍掉样本的第 1 个数据. 如果 $n = 3r_1 + 1$,则舍掉样本的最前面 2 个数据. 这样,剩余样本刚好能三等分,且每组数据点数为 r_1 ,对方程(3)两边取倒数,得

$$\frac{1}{y(t)} = \frac{1}{K} + \frac{A}{K}(e^{-b})^t, \text{记 } S_i (i = 1, 2, 3) \text{ 分别如下}$$

$$S_1 = \sum_{t=0}^{r_1-1} \frac{1}{y(t)} = \frac{r_1}{K} + \frac{A}{K} \left(\frac{1 - e^{-br_1}}{1 - e^{-b}} \right), \quad (4)$$

$$S_2 = \sum_{t=r_1}^{2r_1-1} \frac{1}{y(t)} = \frac{r_1}{K} + \frac{A}{K} e^{-br_1} \left(\frac{1 - e^{-br_1}}{1 - e^{-b}} \right), \quad (5)$$

$$S_3 = \sum_{t=2r_1}^{3r_1-1} \frac{1}{y(t)} = \frac{r_1}{K} + \frac{A}{K} e^{-2br_1} \left(\frac{1 - e^{-br_1}}{1 - e^{-b}} \right), \quad (6)$$

记 $D_1 = S_1 - S_2, D_2 = S_2 - S_3$, 则

$$\begin{cases} D_1 = \frac{A(1 - e^{-br_1})^2}{K(1 - e^{-b})}, \\ D_2 = \frac{Ae^{-br_1}(1 - e^{-br_1})^2}{K(1 - e^{-b})}, \end{cases} \quad (7)$$

由式(7)得

$$b = -\frac{1}{r_1} \ln \frac{D_2}{D_1}, \quad (8)$$

由式(5)、(6)、(8)得

$$K = r_1 \left(S_1 - \frac{D_1^2}{D_1 - D_2} \right)^{-1}, \quad (9)$$

由式(7)得

$$A = \frac{KD_1^2}{C(D_1 - D_2)}, \quad (10)$$

其中 $C = \frac{1 - e^{-r_1 b}}{1 - e^{-b}}$.

式(8)、(9)、(10)统称为 Logistic 方程(3)的参数估计三段和法公式. 全文均采用草履虫试验数据(表 1)^[9],利用三段和法估计 Logistic 方程(3)参数,当观测数据为一环时,去掉第 0 天与第 1 天 2 个数据,得 $K = 452.3451, A = 14.3834, b = 0.8077$,故得到方程(3)的三段和法参数估计的 Logistic 预测方程

$$\hat{y}(t) = \frac{452.3451}{1 + 14.3834e^{-0.8077t}}, \quad (11)$$

又因为 $e^a = A$, 所以 $a = 2.6661$, 方程(11)可等价写成方程(2)的形式

$$\hat{y}(t) = \frac{452.3451}{1 + e^{2.6661 - 0.8077t}}, \quad (12)$$

其他试验数据可类似处理,这种三段和方法得到的参数估计较粗略. 利用优化算法估计参数,可得到比较精确的值,但若事先不知参数大致落在什么范围,也很难对估计参数初始化,也就很难利用优化算法估计参数. 因此,本文提出先在三段和法的基础上粗略估计出方程参数,将其作为初始值,构造参数的初始化范围,然后在此基础上,以残差平方和最小为准则,建立相应的适应度函数,利用粒子群优化算法对 Logistic 方程的参数进一步估计,从而获得更优的拟合预测模型.

表 1 草履虫试验数据¹⁾

Tab. 1 Experimental data of paramecium

$t_{\text{培养}}/d$	双小核草履虫		大草履虫	
	一环	半环	一环	半环
0	2	2	2	2
1	6	3	6	5
2	24	29	31	22
3	75	92	46	16
4	182	173	76	39
5	264	210	115	52
6	318	210	118	54
7	373	240	140	47
8	396		125	50
9	443		137	76
10	454	249	162	69
11	420	219	124	51
12	438	255	135	57
13	492	252	133	70
14	468	270	110	53
15	400	240	113	59
16	472	249	127	57

1)表中数据为 0.5 mL 培养液中的个体数,数据来源于文献[9].

2 粒子群优化算法

粒子群优化 (PSO) 算法由 Kennedy 等^[10]发明,系统随机初始化一组维粒子群(其中包括位置初始值与速度初始值),记第 i 个粒子的速度与位置分别为 $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ 与 $v_i = (v_{i1}, v_{i2}, \dots, v_{in})$,在每一次迭代过程中,粒子通过跟踪粒子本身所找到的最优解(个体极值的第 i 个分量,记: $pbest_i$)与整个粒子群的最优解(全局极值,记: $gbest$)来更新自己,找到以上两极值后,依据以下公式更新速度与位

置^[10-15]

$$v_i^{k+1} = wv_i^k + c_1r_1(pbest_i - x_i^k) + c_2r_2(gbest - x_i^k), \quad (13)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1}, \quad (14)$$

其中,式(13)与(14)中的参数及经验设置通常如下: k 是迭代数, v_i^{k+1} 是 v_i 第 $k+1$ 次迭代, x_i^{k+1} 是粒子 x_i 第 $k+1$ 代; c_1 、 c_2 称为学习因子,经验取值为 $c_1 = c_2 = 2$,调节学习最大步长; w 是惯性加权因子,调节对解空间的搜索范围,取值 $0.1 \sim 0.9$; r_1 、 r_2 是 $(0,1)$ 之间的随机数.每一维粒子的速度都会被限制在最大速度(v_{max})范围内,如果某一维更新后的速度超过 v_{max} ,则这一维的速度限定为 v_{max} ,通常设为粒子的范围宽度.同其他智能优化算法比较,PSO 是根据自己的速度来搜索,且有记忆功能,采用实数编码简单、容易实现,且没有许多参数需要调整,粒子群优化算法有极其广泛的应用^[10-15].

2.1 适应度函数与粒子群初始化

记待拟合 Logistic 方程(2)为 $y = f(K, a, b, t)$, 其中 K, a, b 为待估计参数,由前面的三段和法已得到其初值.对表 1 中实测样本数据 $\{(t_i, y_i) : i = 1, 2, \dots, n\}$, Logistic 方程(2)式 $y = f(K, a, b, t)$ 的拟合问题可转化为优化模型参数向量 (K, a, b) , 使得

$$\min \sum_{i=1}^n [y_i - f(K, a, b, t_i)]^2, \quad (15)$$

式(15)即以实测值与估计值之间的残差平方和最小作为粒子群优化算法中的适应度函数.这样把一个统计中的拟合与估计预测问题转化为一个最优化问题.再利用粒子群优化算法,对问题(15)进行寻优.

根据前面三段和法,得到表 1 中的一环数据时, Logistic 方程(2)式参数的粗略值为 $K = 452.3451$, $a = 2.6661$, $b = 0.8077$, 则在利用 PSO 算法过程中,对参数粒子的初始化取值范围可适当放大作如下处理: $410 \leq K \leq 510$, $1.5 \leq a \leq 5.0$, $0.01 \leq b \leq 1.80$. 对其他 3 组实测值的拟合均做类似处理,这是三段和法与粒子群优化算法相结合的关键之处.

2.2 参数估计的 PSO 算法步骤

基于 Matlab 编程^[16], 粒子群优化算法拟合 Logistic 方程的算法流程如下,程序结束条件为最大迭代次数:

1) 定义全局变量;

2) 构建初始化函数:包括预设最大代数、最大速度控制量、全局最优值、学习因子、粒子的取值范围、初始化粒子群规模 $n = 100$ 与粒子群 A . 粒子群初始化,我们以一个 $n \times 11$ 维矩阵 A 来储存,其中第 1~3 列元素分别为 K, a, b 的初始值(即粒子的初始位置);第 4~6 列元素分别为 K, a, b 的粒子速度值;第

7~9 列元素分别为 K, a, b 的粒子个体最优值;第 10 列元素为粒子个体历史最优适应值;第 11 列元素为粒子当前个体适应值;

3) 按方程(15)构建计算适应度函数,计算每个粒子的适应度值储存在 $A(:, 11)$ 中,并与对应的 $A(:, 10)$ 比较更新,找出当前全局最优适应值 $\min(A(:, 10))$ 与最优参数;

4) 依据公式(13)与(14)更新粒子位置与速度;

5) 达到预设最大迭代数,PSO 算法结束循环;

6) 构建输出函数,输出适应度函数值(残差平方和)及参数值的收敛过程示意图(表 1 中的一环数据的收敛过程示意图见图 1~4,其他情形的收敛示意图略去).

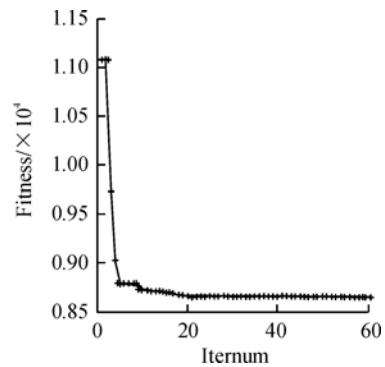


图 1 残差平方和的收敛过程

Fig. 1 Convergence process of residual sum of squares

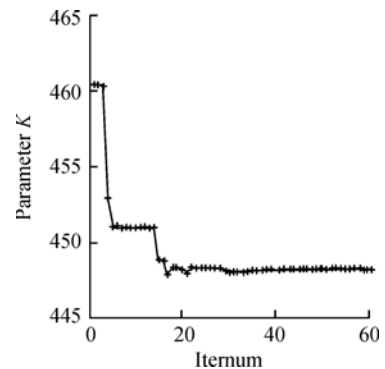


图 2 参数 K 的收敛过程

Fig. 2 Convergence process of parameter K

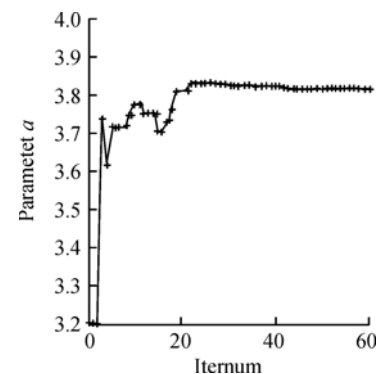


图 3 参数 a 的收敛过程

Fig. 3 Convergence process of parameter a

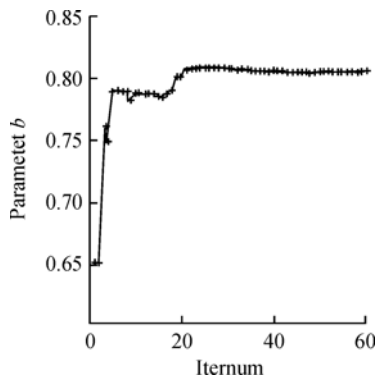


图 4 参数 b 的收敛过程

Fig. 4 Convergence process of parameter b

3 拟合结果与统计检验

表 2 是本文利用三段和法与粒子群优化算法结合对表 1 中草履虫试验数据拟合计算的结果,并与其他经典的方法所得结果做了比较.从表 2 可知,本文方法所得结果具有很好的拟合效果,与以往最好结果基本一致,且算法思想易实现,程序运行时间短.现对实测值与预测值的适合性进行统计检验,这里我们采用相关系数法与 χ^2 检验法^[17].

1) 相关系数 r 计算方法:

表 2 0.5 mL 草履虫试验数据不同方法计算结果的比较¹⁾

Tab. 2 Comparison of the values of K, a, b from 0.5 mL paramecium data with different methods

方法	双小核草履虫								大草履虫							
	一环				半环				一环				半环			
	K	a	b	RSS	K	a	b	RSS	K	a	b	RSS	K	a	b	RSS
枚举法 ^[1]	443	5.068 3	1.045 1	1 248	9 241	5.202 1	1.528 4	2 692	130	4.010 1	1.141 4	2 245	59	3.208 6	0.937 3	971
麦夸方法 ^[2]	448	3.817 1	0.804 0	8 640	244	4.251 2	1.237 8	2 367	131	3.713	1.059 7	222	60	2.819	0.812	956
单纯形 ^[3]	448	3.814 3	0.804 0	8 640	245	4.240 7	1.234 5	2 364	131	3.697 7	1.054 4	2 221	60	2.859 7	0.820 0	956
遗传算法 ^[5]	448.17	3.816 406	0.804 688	8 639.92	244.5	4.253 906	1.238 281	2 364.50	130.8	3.721 680	1.062 500	2 221.55	59.89	2.813 476	0.819 336	956.27
三次设计 ^[6]	448	3.787 198	0.797 419	8 507	246	4.190 265	1.207 555	2 345	132	3.813 561	1.084 6	2 184	61	2.738 796	0.811 953	919
改进单纯形 ^[7]	448.175	3.816 978	0.804 668	8 639.92	244.5	4.252 751	1.238 261	2 364.50	130.805	3.722 258	1.062 480	2 221.55	59.89	2.813 476	0.819 397	956.27
本文方法	448.183	3.817 134	0.804 755	8 639.92	244.1	4.272 786	1.245 889	2 380.87	130.847	3.712 931	1.059 669	2 221.15	59.89	2.819 087	0.820 953	956.27

1) RSS 为残差平方和(Residual sum of squares).

$$r = \frac{\sum (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{n\sigma_y\sigma_{\hat{y}}}, \quad (16)$$

其中 σ_y 表示实测值 y_i 的标准差, $\sigma_{\hat{y}}$ 表示预测拟合值 \hat{y}_i 的标准差, \bar{y} 与 $\bar{\hat{y}}$ 表示相应的均值.

2) χ^2 检验计算方法:

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i}. \quad (17)$$

为检验其适合性,查 χ^2 表,自由度 $f = n - 1$ 时,可查临界值 $\chi_{0.05}^2$,若 $\chi^2 < \chi_{0.05}^2$,差异不显著,即预测值与实测值相吻合;若 $\chi^2 > \chi_{0.05}^2$,差异显著,即预测值与实测值不相吻合.由于 χ^2 值过大或过小的异常估值很敏感,所以可用该检验方法进行异常数据的显著检验,对表 1 中草履虫试验数据的检验结果见表 3.

表 3 相关系数 r 与 χ^2 检验结果¹⁾

Tab. 3 Results of correlation coefficient r and chi-square test χ^2

方法	双小核草履虫		大草履虫	
	一环	半环	一环	半环
r	0.991 0	0.991 2	0.970 8	0.936 3
χ^2	48.423 4*	17.273 3	21.449 0	22.576 4

1) * 代表差异显著.

4 结论

由于 Logistic 方程参数范围不可知,因此,在进行参数估计寻优时,参数的初始范围的确定始终是个难点.本文利用三段和法先确定 Logistic 方程参数的一组粗略值,参数的最优值应在这组值附近,从而给出这组参数的一个大致取值范围,再结合粒子群优化算法寻得最优参数.与其他方法所得结果进行比较,并对所得结果进行相关系数与 χ^2 统计检验,结果表明,此方法能取得很好的拟合效果.

拟合 Logistic 方程的研究具有理论价值与很强的实用性,因此,寻找易实现且效果好的拟合 Logistic 方程的方法一直是研究者努力的方向.也可尝试采用改进的粒子群优化算法(如:混沌粒子群优化算法,GA-PSO 混合算法等)来拟合非线性生长曲线方程.

参考文献:

[1] 万昌秀,梁中宇.逻辑斯谛曲线的一种拟合方法[J].生态学报,1983,3(3):288-296.
 [2] 王莽莽,李典谟.用麦夸方法最优拟合逻辑斯谛曲线

- [J]. 生态学报, 1986, 6(2): 142-147.
- [3] 马占山. 单纯形加速法拟合生态学中的非线性模型[J]. 生物数学学报, 1992, 7(2): 160-167.
- [4] 杨运清, 杨玉林. 动物生长模型参数的常规估计法[J]. 东北农业大学学报, 1994, 25(2): 132-137.
- [5] 蔡煜东, 陈德辉. 运用遗传算法拟合 Logistic 曲线的研究[J]. 生物数学学报, 1995, 10(1): 59-63.
- [6] 王红春, 陈平留. 三次设计结合模矢法拟合 Logistic 曲线的研究[J]. 生物数学学报, 1999, 14(4): 453-458.
- [7] 吴承祯, 洪伟. 运用改进单纯形法拟合 Logistic 曲线的研究[J]. 生物数学学报, 1999, 14(1): 117-121.
- [8] 王福林, 王吉权. 生长曲线参数估计的一种新方法 - 优化回归组合法[J]. 生物数学学报, 2007, 22(3): 533-538.
- [9] GAUSE G F. The Struggle for Existence [M]. USA: Williams and Wilkins, Baltimore, 1934.
- [10] KENNEDY J, EBERHART R C. Particle swarm optimization [C] // IEEE. Proceedings of IEEE international conference on neural networks, Peth, Australia: IEEE Press, 1995: 1942-1948.
- [11] 陈杰, 潘峰, 王光辉. 粒子群优化方法在动态优化中的研究现状[J]. 智能系统学报, 2009(3): 189-198.
- [12] 蒋建国, 吴琼, 夏娜. 自适应粒子群算法求解 Agent 联盟[J]. 智能系统学报, 2007, 2(2): 69-73.
- [13] 王兆伟, 肖扬, 刘湘黔. 基于粒子群算法的 MIMO CDMA 平坦衰落信道均衡器[J]. 智能系统学报, 2008(1): 38-42.
- [14] SHI Y, EBERHART R C. Empirical study of particle swarm optimization [C] // IEEE. Proceedings of the IEEE congress on evolutionary computation, Washington: IEEE Press, 1999: 1945-1950.
- [15] EBERHART R C, SHI Y. Particle swarm optimization: Developments, applications and resources [C] // IEEE. Proceedings of the 2001 congress on evolutionary computation, Seoul, South Korea: IEEE Press, 2001: 81-86.
- [16] 薛定宇, 陈阳泉著. 高等应用数学问题的 MATLAB 求解 [M]. 北京: 清华大学出版社, 2008.
- [17] 张国权. 应用概率统计 [M]. 北京: 科学出版社, 2003.

【责任编辑 李晓卉】

(上接第 113 页)

参考文献:

- [1] TAMMY H, BETH O. Cymothoid isopod parasites in aquaculture: A review and case study of a Turkish sea bass (*Dicentrarchus labrax*) and sea bream (*Sparus auratus*) farm [J]. Diseases of Aquatic Organisms, 2001, 46: 181-188.
- [2] 于海燕, 丁雪娟, 张剑英. 中国海湾水虱属(等足目, 纺锤水虱科)一新纪录种 [J]. 动物分类学报, 2007, 32(4): 1000-1001.
- [3] 陆小菘, 邹为民, 彭华林, 等. 珠江水系二种鱼虾寄生性等足类新种的发现和初步观察 [J]. 中国水产科学, 2007, 14(6): 1023-1026.
- [4] MARIAPPAN P, BALASUNDARAM C, TRILLES J P. Infection of the isopod *Tachaea spongillicola* on freshwater prawns *Macrobrachium* spp. in southern India [J]. Diseases of Aquatic Organisms, 2003, 55: 259-260.
- [5] HORTON T, OKAMURA B. Cymothoid isopod parasites in aquaculture: A review and case study of a Turkish sea bass (*Dicentrarchus labrax*) and sea bream (*Sparus auratus*) farm [J]. Diseases of Aquatic Organisms, 2000, 46: 181-188.
- [6] KENT M L. Marine netpen farming leads to infections with some unusual parasites [J]. International Journal for Parasitology, 2000, 30: 321-326.
- [7] LEONARDOS I, TRILLES J P. Reproduction of *Mothocya epimerica* (Crustacea: Isopoda: Cymothoidae), parasitic on the sand smelt *Atherina boyeri* (Osteichthyes: Atherinidae) in Greek lagoons [J]. Diseases of Aquatic Organisms, 2004, 62: 249-253.
- [8] LEONARDOS I, TRILLES J P. Host-parasite relationships: Occurrence and effect of the parasitic isopod *Mothocya epimerica* on sand smelt *Atherina boyeri* in the Mesolongi and Etolikon Lagoons (W. Greece) [J]. Diseases of Aquatic Organisms, 2003, 54: 243-251.
- [9] TAMMY H, BETH O. Cymothoid isopod parasites in aquaculture: A review and case study of a Turkish sea bass (*Dicentrarchus labrax*) and sea bream (*Sparus auratus*) farm [J]. Diseases of Aquatic Organisms, 2001, 46: 181-188.
- [10] SIEVERS G, PALACIOS P, INOSTROZA R, et al. Evaluation of the toxicity of 8 insecticides in *Salmo salar* and the *in vitro* effects against the isopode parasite, *Ceratothoa gaudichaudii* [J]. Aquaculture, 1995, 134: 9-16.
- [11] ATHANASSOPOULOU F, BOUBOULIS D, MARTINSEN B. *In vitro* treatments of deltamethrin against the isopod parasite *Ceratothoa oestroides*, a pathogen of sea bass *Dicentrarchus labrax* L [J]. Bull Eur Assoc Fish Pathol, 2001, 21: 26-29.
- [12] PAPAPANAGIOTOU E P, TRILLES J P, PHOTIS G. First record of *Emetha audouini*, a cymothoid isopod parasite from cultured sea bass *Dicentrarchus labrax* in Greece [J]. Dis Aquat Org, 1999, 38: 235-237.
- [13] 张一宾. 农药 [M]. 北京: 中国物资出版社, 1998: 200.

【责任编辑 柴焰】