

使用 SSI 模型描述昆虫温发育速率的 SAS 程序*

时培建^{1, 2**} 戈 峰^{1***}

(1. 中国科学院动物研究所 农业虫害鼠害综合治理研究国家重点实验室, 北京 100101;

2. 南京林业大学 竹类研究所, 南京 210037)

摘 要 Sharpe-Schoolfield-Ikemoto (SSI) 模型是用来描述温度对昆虫发育速率影响的一个重要数学方程, 它既能够很好地拟合温发育速率的数据, 又能提供若干个重要温参数, 特别是内禀最适发育温度 (T_{Φ})。此温度反映了昆虫生存的最佳温环境。然而, 由于 SSI 模型的参数过多, 结构显得较为复杂, 给实际的数据拟合带来了困难。一个基于 R 平台的软件包已经被开发出来, 专门用于执行 SSI 模型的参数拟合。但目前由于国内许多研究者青睐于使用 SAS 统计软件, 所以本文提供了一个基于 SAS 统计软件的程序, 用以快速拟合 SSI 模型的参数。通过拟合棉铃虫 *Helicoverpa armigera* Hübner 蛹期的温发育速率数据, 发现此 SAS 程序具有很好的拟合效果, 同时进一步证明了 SSI 模型对描述温度对昆虫发育速率影响的有效性。

关键词 棉铃虫, 发育时间, 参数拟合, 温环境, 内禀最适发育温度

A SAS procedure for estimating parameters of the SSI model describing temperature-dependent developmental rates in insects

SHI Pei-Jian^{1, 2**} GE Feng^{1***}

(1. State Key Laboratory of Integrated Management of Insect Pests and Rodents in Agriculture, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; 2. Bamboo Research Institute, Nanjing Forestry University, Nanjing 210037, China)

Abstract The Sharpe-Schoolfield-Ikemoto (SSI) model is a very important mathematical equation that is widely used to describe the effect of temperature on the developmental rates of insects. This model is not only a good fit to data on temperature-dependent developmental rates, but also provides estimates of several valuable thermal parameters, especially the intrinsic optimum temperature for development (T_{Φ}) that indicates the optimal thermal environment for a specific insect species. However, the SSI model has considerable difficulty in estimating parameters then making it rather complex to use. A package for estimating these parameters has been developed based on the statistical software R. Because some domestic researchers might prefer to use the statistical software SAS, we here present a practical SAS procedure for quickly estimating SSI model parameters. This procedure exhibited satisfactory goodness of fit to the temperature-dependent developmental rates of male pupae of cotton bollworms. These results further demonstrate the validity of the SSI model for describing the effect of temperature on insect developmental rates.

Key words cotton bollworm, developmental time, parametric fitting, thermal environment, intrinsic optimum temperature for development

温度作为影响昆虫发育一个重要因素, 长期以来备受昆虫研究人员的关注。围绕着温度对昆虫发育时间的影响, 一系列的数学模型被相继提

出 (时培建等, 2011), 其中基于热动力学原理构建的 Sharpe-Schoolfield-Ikemoto 模型 (下文将简称 SSI 模型; Ikemoto, 2005) 是描述温度对

* 资助项目: 国家自然科学基金项目 (31200321); 国家科技支撑计划项目 (2012BAD19B05) 和国家自然科学基金青年基金 (31400348)

** E-mail: shi_peijian@163.com

*** 通讯作者, E-mail: gef@ioz.ac.cn

收稿日期: 2014-09-24, 接受日期: 2014-10-13

昆虫发育速率影响的一个重要的非线性方程,然而由于这个模型的结构十分复杂,能否准确拟合模型的参数成为使用这个模型的瓶颈。Ikemoto 等(2013)提出了一个基于 R 软件的快速拟合程序,然而我国许多研究者更为熟悉 SAS 软件,因而本文提出一个基于 SAS 软件开发的拟合程序,以方便 SSI 模型的推广使用。

1 SSI 模型的基本形式

$$r = \rho_{\phi} \cdot P_2 \cdot \frac{T}{T_{\phi}} \cdot \exp \left[\frac{\Delta H_A^*}{R} \left(\frac{1}{T_{\phi}} - \frac{1}{T} \right) \right]$$

此处, r 表示昆虫在绝对温度 T (单位: K) 下完成某一发育阶段的发育速率(即发育时间的倒数)。其中, P_2 表示在不同温度下控制昆虫发育的酶处于自然状态的概率,它是绝对温度 T 的一个函数:

$$P_2 = \frac{1}{1 + \exp \left[\frac{\Delta H_L}{R} \left(\frac{1}{T_{1/2L}} - \frac{1}{T} \right) \right] + \exp \left[\frac{\Delta H_H}{R} \left(\frac{1}{T_{1/2H}} - \frac{1}{T} \right) \right]}$$

SSI 模型的自变量为 T , 因变量为 r , R 表示

通用气体常数 ($R = 1.987 \text{ cal} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$) ($1 \text{ cal} = 4.1868 \text{ J}$), 其余参数为待拟合常数。在 SSI 模型中, 最重要的参数为 T_{ϕ} , 表示能够使 P_2 达到最大值的温度, 即代表了能够使控制昆虫发育的酶活性表达最大化的温度。 T_{ϕ} 与其它参数之间有如下的关系 (Ikemoto, 2005):

$$T_{\phi} = \frac{\Delta H_L - \Delta H_H}{R \cdot \ln \left(-\frac{\Delta H_L}{\Delta H_H} \right) + \frac{\Delta H_L}{T_{1/2L}} - \frac{\Delta H_H}{T_{1/2H}}}$$

2 数据

吴坤君等(2009)详细报道了温度对棉铃虫蛹期发育速率的影响,我们选取了在不同常温下的雄蛹发育时间数据用以展示 SSI 模型的使用。

3 SSI 模型参数的使用步骤

3.1 数据表的建立

首先, 我们需要建立一个适温区域的 Excel 数据表, 用以估算 ΔH_A^* (参看图 1)。

	A	B	C	D	E	F	G	H	I
1	T	D	x	y	(x-mean(x))^2	(y-mean(y))^2	(x-mean(x)) *(y-mean(y))	beta	
2	18	30.12	0.003435	-3.40519	3.99717E-09	0.441782835	-4.20224E-05	-9338.93	
3	19	25.85	0.003423	-3.25231	2.64882E-09	0.261927632	-2.63401E-05		
4	20	21.75	0.003411	-3.07961	1.58328E-09	0.11498326	-1.34926E-05	R	
5	21	19.27	0.0034	-2.95855	7.94876E-10	0.047536009	-6.14697E-06	1.987	
6	22	17.76	0.003388	-2.87695	2.78064E-10	0.018612271	-2.27495E-06		
7	23	16.18	0.003377	-2.78378	2.74025E-11	0.001870906	-2.26423E-07	HA	
8	24	14.01	0.003365	-2.63977	3.75618E-11	0.01015068	-6.17477E-07	18556.45	
9	25	12.66	0.003354	-2.53845	3.03319E-10	0.040834113	-3.51934E-06		
10	26	11.56	0.003343	-2.44755	8.19555E-10	0.085832052	-8.38714E-06		
11	27	10.67	0.003332	-2.36744	1.58125E-09	0.139193071	-1.48358E-05		
12	28	10.03	0.003321	-2.30558	2.5835E-09	0.189173968	-2.21073E-05		
13	29	9.31	0.00331	-2.23109	3.82147E-09	0.259521828	-3.14921E-05		
14									
15			mean(x)	mean(y)	Sxx	Syy	Sxy		
16			0.003371	-2.74052	1.84763E-08	1.611418626	-0.000171463		
17									

图 1 在 Excel 数据表中 HA 的估计

Fig. 1 Estimating HA in a sheet of Excel

T 表示温度 ($^{\circ}\text{C}$), D 表示发育时间 (d), $x = 1/(T + 273.15)$ 表示绝对温度的倒数, $y = \ln(1/D)$ 表示发育速率的自然对数, $S_{xx} = \sum (x_i - \bar{x})^2$, $S_{yy} = \sum (y_i - \bar{y})^2$, $S_{xy} = \sum [(x_i - \bar{x})(y_i - \bar{y})]$ 。如果 $S_{xy} > 0$, $\text{beta} = \sqrt{S_{yy}/S_{xx}}$; 如果 $S_{xy} < 0$, $\text{beta} = -\sqrt{S_{yy}/S_{xx}}$ 。HA = -beta · R。

T represents temperature ($^{\circ}\text{C}$), D represents developmental time (d), $x = 1/(T + 273.15)$ representing the reciprocal of absolute temperature, $y = \ln(1/D)$ representing the natural logarithm of developmental rate, $S_{xx} = \sum (x_i - \bar{x})^2$, $S_{yy} = \sum (y_i - \bar{y})^2$, $S_{xy} = \sum [(x_i - \bar{x})(y_i - \bar{y})]$. If $S_{xy} > 0$, $\text{beta} = \sqrt{S_{yy}/S_{xx}}$; if $S_{xy} < 0$, $\text{beta} = -\sqrt{S_{yy}/S_{xx}}$. HA = beta · R.

其次, 我们还需要建立另一个适温区域的数据表, 用以估算 $T_{1/2L}$, 同时我们还可以得到发育起点温度 LDT (Lower developmental threshold) 和有效积温 SET (Sum of effective temperatures) (参看图 2)。

再次, 我们还需要建立一个名为 development.xls 的 Excel 数据表用来存储温度和发育时间的数据 (参看图 3)。注意此表包含了低温区域、适温区域和高温区域的发育时间数据。

3.2 SSI 模型参数拟合的 SAS 程序

第一步, 打开数据表 development.xls, 并且打开 SAS 的操作界面;

第二步, 在 SAS (version $\geq$ V8) 的 Editor 窗口, 录入如下代码:

```
filename a dde 'excel[[development.xls]sheet1!r1c1:r22c2';
data development;
  infile a notab dlm='09'X;
  length T D $10;
  input T D;
run;
```

```
data development;
  set development;
  T=T+273.15;
  r=1/D;
  sqrt_r = sqrt(r);
  R
    = 1.987;
/* The following three values can be changed. */
  HA   = 18556.45;
  LDT  = 13.16511;
  SET  = 150.4971;
run;

proc nlin MAXITER=500 data=development;

  model sqrt_r = SQRT((Tphi-273.15-LDT)/
    SET*T/Tphi*exp(HA/R*(1/Tphi -1/T))/
    (1+exp(HL/R*(1/TL-1/T))+exp(HH/R*(1/(HH/(R*
    log(-HL/HH)+HL/TL+(HH-HL)/Tphi))-1/T))));
  parms Tphi = 293.15
    HL = -100000 to -20000 by 5000
    HH = 30000 to 200000 by 5000
    TL = 288.15 to 298.15 by 1
    ;
run;
```

	A	B	C	D	E	F	G	H	I
	T	D	x	y	$(x-\text{mean}(x))^2$	$(y-\text{mean}(y))^2$	$(x-\text{mean}(x))$ $\times (y-\text{mean}(y))$	LDT	
1									
2	18	30.12	30.12	542.16	182.8580063	29982.65403	2341.488488	13.16511	
3	19	25.85	25.85	491.15	85.60875625	14919.40103	1130.146613		
4	20	21.75	21.75	435	26.54825625	4355.340025	340.0392375	TL	
5	21	19.27	19.27	404.67	7.14225625	1271.992225	95.3147125	286.3151	
6	22	17.76	17.76	390.72	1.35140625	471.541225	25.2436875		
7	23	16.18	16.18	372.14	0.17430625	9.828225	-1.3088625	SET	
8	24	14.01	14.01	336.24	6.69515625	1073.545225	84.7794375	150.4971	
9	25	12.66	12.66	316.5	15.50390625	2756.775025	206.7384375		
10	26	11.56	11.56	300.56	25.37640625	4684.718025	344.7916875		
11	27	10.67	10.67	288.09	35.13525625	6547.237225	479.6236625		
12	28	10.03	10.03	280.84	43.13205625	7773.067225	579.0236375		
13	29	9.31	9.31	269.99	53.10765625	9803.970225	721.5718125		
14									
15			mean(x)	mean(y)	Sxx	Syy	Sxy		
16			16.5975	369.005	482.633425	83650.0697	6347.45255		
17									

图 2 在 Excel 数据表中 TL 的估计

Fig. 2 Estimating TL in a sheet of Excel

T 表示温度 ($^{\circ}\text{C}$), D 表示发育时间 (d) $x = D$ $y = D \cdot T$ $S_{xx} = \sum (x_i - \bar{x})^2$ $S_{yy} = \sum (y_i - \bar{y})^2$ $S_{xy} = \sum [(x_i - \bar{x})(y_i - \bar{y})]$ 。

* 资助项目: 国家自然科学基金项目 (31200321); 国家科技支撑计划项目 (2012BAD19B05) 和国家自然科学基金青年基金 (31400348)

** E-mail: shi_peijian@163.com

***通讯作者, E-mail: gef@ioz.ac.cn

收稿日期: 2014-09-24, 接受日期: 2014-10-13

如果 $S_{xy} > 0$, $LDT = \sqrt{S_{yy}/S_{xx}}$; 如果 $S_{xy} < 0$, $LDT = -\sqrt{S_{yy}/S_{xx}}$ 。 $TL = LDT + 273.15$; $SET = \bar{y} - LDT \cdot \bar{x}$ 。

T represents temperature (°C) , D represents developmental time (d) , $x = D$, $y = D \cdot T$, $S_{xx} = \sum (x_i - \bar{x})^2$, $S_{yy} = \sum (y_i - \bar{y})^2$, $S_{xy} = \sum [(x_i - \bar{x})(y_i - \bar{y})]$. If $S_{xy} > 0$, $LDT = \sqrt{S_{yy}/S_{xx}}$; if $S_{xy} < 0$, $LDT = -\sqrt{S_{yy}/S_{xx}}$. $TL = LDT + 273.15$; $SET = \bar{y} - LDT \cdot \bar{x}$.

	A	B	C
1	15	45.45	
2	16	41.15	
3	17	36.76	
4	18	30.12	
5	19	25.85	
6	20	21.75	
7	21	19.27	
8	22	17.76	
9	23	16.18	
10	24	14.01	
11	25	12.66	
12	26	11.56	
13	27	10.67	
14	28	10.03	
15	29	9.31	
16	30	8.9	
17	31	8.14	
18	32	7.89	
19	33	7.95	
20	34	8.03	
21	35	8.16	
22	36	9.1	
23			

图 3 对棉铃虫雄蛹发育时间的影响

Fig. 3 Effect of temperature on the male pupal developmental time of cotton bollworm

第 1 列中的数据表示温度 (°C) ; 第 2 列中的数据表示对应温度下的发育时间 (d) 。

The data in the first column represent the temperatures (°C) ; and the data in the second column represent the corresponding developmental times (d) .

研究者在使用上述程序时 , 注意要对程序中数据选取的范围进行更改 , 本文温度与发育时间的数据在 development.xls 中第一个表单 sheet1 中的第 1 行至第 22 行、第 1 列至第 2 列 , 所以为 sheet1!r1c1:r22c2 ; 此外 , 研究者还需要根据自己计算的 HA 、 LDT 和 SET 的值来修改程序中对应的内容。

第三步 , 执行上述程序 , 可以得到如下结果 :

The NLIN Procedure				
Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
Tphi	296.1	0.3914	295.3	296.9
HL	-56119.0	9792.2	-76691.4	-35548.5
TL	284.4	1.0655	282.2	286.6
HH	58890.4	2241.8	54180.7	63600.1

第四步 , 根据公式 $\rho_{\Phi} = \frac{T_{\Phi} - 273.15 - LDT}{SET}$

(Shi 等 , 2011) , 计算出 $\rho_{\Phi} = 0.065017$ 。

第五步 , 根据公式

$$T_{1/2H} = \frac{\Delta H_H}{R \cdot \ln \left(-\frac{\Delta H_L}{\Delta H_H} \right) + \frac{\Delta H_L}{T_{1/2L}} + \frac{\Delta H_H - \Delta H_L}{T_{\Phi}}} \quad (\text{参}$$

看 Shi 等 , 2011) , 计算出 $T_{1/2H} = 308.34$ 。

至此 , SSI 模型中所有的参数都已经被估计出来 (表 1) 。其中 T_{Φ} 表示可使控制昆虫发育的酶活性表达最大化的温度 ; ρ_{Φ} 表示在 T_{Φ} 温度下的平均发育速率 (假设控制昆虫发育的酶活性表达最大化时的概率值为 1) ; ΔH_A^* 、 ΔH_L 、 ΔH_H 可视作为待拟合参数 ; $T_{1/2L}$ 和 $T_{1/2H}$ 分别代表昆虫发育的下限和上限温度 ; LDT 表示昆虫完成某一虫态的发育起点温度 (单位 : °C) , 若用单位 K 表示则等于 $T_{1/2L}$ (也即 : $T_{1/2L} = LDT + 273.15$) ; SET 表示昆虫完成某一发育阶段所需的有效积温。

表 1 SSI 模型参数的估计值

Table 1 Parameter estimates in the SSI model

标准公式中的参数名 Parametric name in the standard formula	SAS 程序中对应的参数名 Parametric name in the SAS procedure	估计值 Estimate
T_{Φ}	Tphi	296.10 (K)
ρ_{Φ}	—	0.065017 (1/d)
ΔH_A^*	HA	18 556.45 (cal·mol ⁻¹)
ΔH_L	HL	- 56 119 (cal·mol ⁻¹)
ΔH_H	HH	58 890.4 (cal·mol ⁻¹)
$T_{1/2L}$	TL	284.40 (K)
$T_{1/2H}$	—	308.34 (K)

LDT	LDT	13.17 (°C)
SET	SET	150.50 (degree-days)

图 4 展示了发育速率的观测值和 SSI 模型预测值的对比情况。在图 4 (A) 中, 曲线表示使用 SSI 模型预测的发育速率; 小圆圈表示观测的发育速率, 其中加 X 号的小圆圈表示在曲线拟合中未被使用的观测值; 直线表示在适温区域根据 Ikemoto-Takai 方程 (Ikemoto and Takai, 2000) 拟合得到的发育速率直线。在图 4 (B) 中, 曲线表示不同温度下控制昆虫发育的酶处于自然状态下的概率, 虚线对应的温度即 T_{Φ} 。

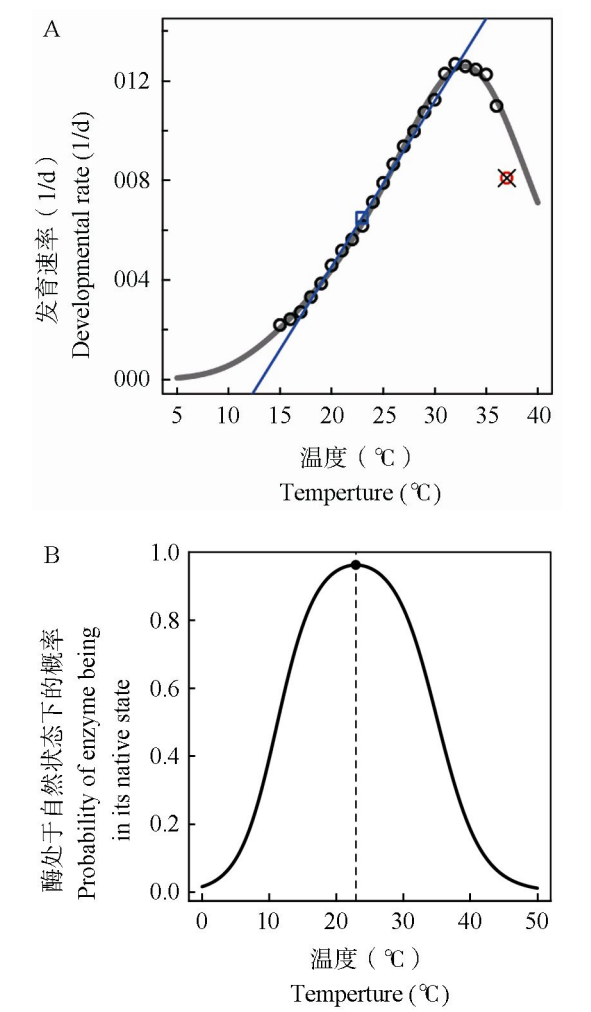


图 4 SSI 模型对雄蛹发育速率的拟合结果
Fig. 4 Fitted results for the temperature-dependent

developmental rates of male pupae
A. 发育速率曲线和适温区的直线。开放性圆点为发育速率的观测值, 其中 37°C 对应的发育速率数据被排除, 因为在此温度下, 雄蛹的死亡率非常高; B. 控制发育的酶处于自然状态下的概率。
A..Developmental rate curve and straight line in the mid-temperature range. Small open circles represent the observations of developmental rates. Among these data points, the datum at 37°C was excluded because the mortality was too high; B. Probability of enzyme controlling development being in its native state.

4 讨论

4.1 程序的优缺点

使用此文推荐的 SAS 程序, 对于熟悉 SAS 统计平台的研究者将非常方便, 并且拟合的参数更倾向于为全局最优解。但不是全自动程序, 相比于 Ikemoto 等 (2013) 提出的 R 函数程序 (只需要输入温度和对应的发育历期及其拟合线性模型和非线性模型所需的有效数据范围, 所有模型参数将被程序自动计算求出), 需要手动在 Excel 计算部分参数, 然后再使用 SAS 程序估计其余参数。如果是处理批量数据, 使用本文提供的程序较为繁琐。然而, 如果只是处理少量数据, 则此程序可以提供更为精确的参数值, 因为本文提供的算法是基于非线性最小二乘的牛顿-高斯算法, 较最小卡方的最优化算法, 参数估计更为稳定。因为 SAS 是商业软件, 运行环境也较 R 更为可靠。

4.2 应用时应注意的问题

首先, 在使用 Excel 数据表计算 SSI 模型中的部分参数时, 一定要注意 S_{xy} 的正负号, 因为这决定着减少主轴法 (Ikemoto and Takai, 2000) 所计算斜率的正负号。其次, 在运行 SAS 程序前, 一定要先打开 Excel 数据表。再次, 程序中涉及参数的初始值, 要根据 Excel 数据表计算的结果做相应的调整。然后, 还需要注意各个 Excel 表使用的数据所在温区的范围有所不同。计算 HA、LDT 和 SET 的那两个数据表使用了适温区域的数据, 而直接用于执行 SAS 程序的那个数据表则使用了所有温区的数据。最后, 未能被

Excel 数据表和 SAS 程序所计算的参数 ρ_{Φ} 和 $T_{1/2H}$, 还需要自行计算。

参考文献 (References)

- Ikemoto T, 2005. Intrinsic optimum temperature for development of insects and mites. *Environmental Entomology*, 34(6): 1377–1387.
- Ikemoto T, Takai K, 2000. A new linearized formula for the law of total effective temperature and the evaluation of line-fitting methods with both variables subject to error. *Environmental Entomology*, 29(4): 671–682.
- Ikemoto T, Kurahashi I, Shi P, 2013. Confidence interval of intrinsic optimum temperature estimated by using thermodynamic SSI model. *Insect Science*, 20(3): 420–428.
- Shi P, Ikemoto T, Egami C, Sun Y, Ge F, 2011. A modified program for estimating the parameters of the SSI model. *Environmental Entomology*, 40(2): 462–469.
- 时培建, 池本孝哉, 戈峰, 2011. 温度与昆虫生长发育关系模型的发展与应用. 应用昆虫学报 2011, 48(5): 1149–1160. [Shi PJ, IKEMOTO Takaya, Ge F, 2011. Development and application of models for describing the effects of temperature on insects' growth and development. *Chinese Journal of Applied Entomology*, 2011, 48(5): 1149–1160.]
- 吴坤君, 龚佩瑜, 阮永明, 2009. 用非线性模型估测恒温 and 变温下棉铃虫蛹的发育率. 昆虫学报, 52(6): 640–650. [Wu KJ, Gong PY, Ruan YM, 2009. Estimating developmental rates of *Helicoverpa armigera* (Lepidoptera: Noctuidae) pupae at constant and alternating temperatures by nonlinear models. *Acta Entomologica Sinica*, 52(6): 640–650.]