# 沿海蝗区东亚飞蝗(*Locusta migratoria manilensis*)产卵场所选择的 Logistic 回归模型

季 荣[1],高增祥[2],谢宝瑜[3],李典谟[3,*],曾献春[1]

(1. 新疆师范大学生命科学与化学学院,乌鲁木齐 830054;2. 中国海洋大学海洋环境和生态教育部重点实验室,青岛 266100;
3. 中国科学院动物研究所,北京 100080)

**摘要**:以沿海蝗区南大港水库为研究区域,通过 2002 和 2003 两年野外 450, 50 m 规则栅格取样获取东亚飞蝗卵块、植物种类及其密度、土壤特性如含盐量、5 cm 含水量、pH、有机质及地形(阴坡和阳坡)等数据,采用多元 Logistic 回归模型,运用 SAS 软件筛选出与飞蝗产卵场所选择密切相关的变量,建立用于预测飞蝗产卵场所选择的 Logistic 回归模型。结果表明用植株密度(veg_d)和土壤含水量(water)所组建的模型能较好地预测飞蝗产卵选择,$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right)=21.63-\frac{76.23}{\text{water}}-5.43\log(\text{water})-0.86(veg_d)$。利用拟合优度(Goodness of fit)、预测准确性(Predictive accuracy)及模型 $x^2$ 统计(Model chi-square statistic)等指标对模型进行评价的结果表明,所组建的用于预测飞蝗产卵场所选择的 Logistic 回归模型是可靠的,且能较好地预测事件是否发生。研究结果为区域蝗灾早期预警提供了科学依据和方法,对今后预测飞蝗产卵地点选择及防治决策有较高的实用性和应用价值。

**关键词**:二分类因变量;Logistic 回归模型;产卵场所选择;东亚飞蝗;沿海蝗区

**文章编号**:1000-0933(2007)12-5029-09 **中图分类号**:Q141 **文献标识码**:A

# Prediction on oviposition site selection *Locusta migratoria manilensis* by logistic model in coastal areas

JI Rong[1], GAO Zeng-Xiang[2], XIE Bao-Yu[3], LI Dian-Mo[3,*], ZENG Xian-Chun[1]

1 *College of Life Sciences and Chemistry, Xinjiang Normal University, Urumqi* 830054, *China*

2 *Key Laboratory of Marine Environmental Science and Ecology, Ministry of Education, Ocean University of China, Qingdao* 266100, *China*

3 *Institute of Zoology, Chinese Academy of Sciences, Beijing* 100080, *China*

***Acta Ecologica Sinica*, 2007, 27(12): 5029 ~ 5037.**

**Abstract**: The multi-logistic regression method was used to establish a regression model for predicting oviposition site selection in the *Locusta migratoria manilensis* in Nandagang farm, a coastal area Hebei province. Eight potential factors are observed including vegetation species and densities, soil properties (water content at 5 cm depth, salinity, pH and organic matter) and microrelief properties (south slope and north slope) in coastal areas. Data of locust eggpods, vegetation, soil

**基金项目**:国家自然科学基金资助项目(30460028,30660030);国家教育部科技研究重点资助项目(206165);新疆维吾尔自治区高校科研资助项目(XJEDU2005I23, XJEDU2004S20)

**收稿日期**:2006-10-08; **修订日期**:2007-04-29

**作者简介**:季荣(1970 ~),女,江苏泰兴人,博士,副教授,主要从事种群生态学及害虫遥感监测研究. E-mail: jirongxj@yahoo.com.cn

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

**Foundation item**: The project was financially supported by National Natural Sience Foundation of China (No. 30460028, 30660030); Scientific Research Program of Education Ministry of China (No. 206165); Scientific Research Program of the Higher Education Institution of XinJiang (No. XJEDU2005I23, XJEDU2004S20)

**Received date**: 2006-10-08; **Accepted date**: 2007-04-29

**Biography**: JI Rong, female, Ph. D., Associate professor mainly engaged in population ecology. E-mail: jirongxj@yahoo.com.cn

and microrelief were collected from two spatial scales in 2002 and 2003: 450m intervals throughout the study area, and 50m grids for the possible egg-laying areas where locust plague occured or vegetation was sparse. The obtained regression model includs only vegetation density (veg_d) and soil water content at 5 cm (water), $\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right)=21.63-\frac{76.23}{\text{water}}-5.43\log(\text{water})-0.86(\text{veg_d})$. Where $P(Y=1)$ is the probablity of a site selected by *Locusta migratoria manilensis*. After evaluated by methods of Goodness of Fit, Predictive Accuracy and Model Chi-square Statistic, the established regression model is proved to be reliable and can predict eggpods occurrence well. This model can be used for locust plague prediction and early decision-making of its control.

**Key Words**: dichotomous dependent variable; Logistic regression model; oviposition site selection; *Locusta migratoria manilensis*

东亚飞蝗 *Locusta migratoria manilensis* 产卵于土壤中,并对环境有一定的选择能力。已有研究报道飞蝗产卵场所的选择与环境有密切关系,如东亚飞蝗产卵多选择在植被稀疏(覆盖度通常 <50%)的向阳坡,且有最适宜的土壤含水量(砂土 10% ~12%,壤土 15% ~18%,粘土 18% ~20%)和含盐量范围(0.2% ~1.2%),当土壤含盐量 >2%,土壤湿度 >25% 或 <5% 时,飞蝗不再选择产卵或产卵数量显著降低[1,2]。但环境各变量如植被、土壤和地形等对飞蝗产卵地点选择的协同作用及如何利用环境变量预测飞蝗最有可能的产卵地点等尚未见报道。若能依据环境参数准确预测某一地点卵块的存在与否,这不仅对快速定位蝗虫孳生地、实时掌握蝗卵发育进程、及时进行早期点、片防治很重要,而且还可通过遥感技术预测飞蝗适宜的产卵场所,进而实现蝗灾早期预警。因此,在实际应用与早期防治决策中,卵块在蝗区内某一点出现概率的相对高低成为最为关键的问题。

本研究以目前占我国蝗区面积比例最大、蝗灾发生最严重的沿海蝗区为例[3],以国家一类蝗区(即常年重点防治地区)——河北黄骅南大港为研究区域,通过 2002 和 2003 两年野外数据调查,旨在确定环境因素包括植被种类及其覆盖度、地形(阴坡和阳坡)和土壤特性如含水量、含盐量、pH、有机质等对飞蝗产卵的综合作用,建立用于预测飞蝗产卵生境选择的 Logistic 回归模型,为区域蝗灾早期预警和防治提供科学依据和方法。

## 1 材料与方法

### 1.1 实验区描述

南大港水库(38°28.04′ ~38°33.54′N, 117°25.74′ ~117°32.78′E)位于河北省黄骅市南大港国营农场境内,东临渤海,属于泻洪入海口区域,面积约 4 700 $hm^2$[4]。年均气温 11.9℃,年均降水量 627.6 mm,库区内主要是壤性土质,植被以芦苇 *Phragmites communis* 为主,伴生有盐地碱蓬 *Suaeda salsa*、碱蓬 *S. glauca*、狗尾草 *Setaria viridis* 等[5]。研究区域内东亚飞蝗 1 年 2 代,以卵越冬。自 20 世纪 80 年代以来,由于降水量少、地下水利用过度、人工蓄水不足等原因,一方面造成库区内除低洼地和防火沟里有季节性的积水外,常年无水;另一方面 1991 ~2003 年间南大港农场有 6a 蝗灾大发生,其中 2002 年仅夏蝗发生面积就超过 2 万 $hm^2$,芦苇被 3 ~5 龄群居型蝗蝻危害致只剩茎杆的面积就超过 1 300 $hm^2$,蝗蝻最高密度达 8 000 头 $\cdot m^{-2}$[5]。

南大港农场是典型的沿海蝗区,也是历史上有名的老蝗区及东亚飞蝗的主要发生基地之一[6]。南大港农场已被列为国家一类蝗区,即常年防治的重点地区[3]。

### 1.2 数据采集与处理

越冬卵块数据(无或及卵块数 $\cdot m^{-2}$)的调查分别在 2002 和 2003 年秋蝗产卵盛期或刚结束进行(10 月中、下旬或 11 月初,尽可能保证在没任何形式的降水之前)。首先对整个研究区域采取 450 m 的栅格取样共计 289 个样点(图 1)[7],再对可能产卵的区域(裸露地、稀疏植被地、坡和沟所在地、芦苇严重受害地)采用 50 m 的规则网格取样(2601 个样点)。卵块调查时每一样方大小为 100 cm 长 ×50 cm 宽 ×10 cm 深。现场记录的环境变量(样方大小 1 $m^2$)有植物种类及其密度、土壤 5 cm 的湿度(TSC-Ⅱ™智能野外湿度仪)、地形(坡向用 GPS,坡度通

过测量三边再计算其夹角而得),然后取约 100 g 土样装袋、密封和标记,带回室内测量土壤粒径、总盐量、有机质和 pH。每一样点的经纬度信息用 GPS(Magellan GPS315)准确记录。2002 年所有调查点在 2003 年重复调查,所获数据求其平均值以进一步分析。

将数据整理后作如下处理:

① $y=1$ 有卵,$y=0$ 无卵;

② 植物种类划分为两类,芦苇 $x_1=1$,其它 $x_1=0$;

③ 植株密度(棵 $\cdot m^{-2}$)划分为 4 级,0 级:$x_2 \leqslant 50$(覆盖度≤10%),1 级:$50<x_2 \leqslant 100$(覆盖度介于 10% ~30%),2 级:$100<x_2 \leqslant 150$(覆盖度 30% ~50%),3 级:$x_2>150$(覆盖度≥50%);

④ 地形分为两类,阴坡或沟的阴面 $x_3=1$,阳坡或沟的阳面 $x_3=2$。

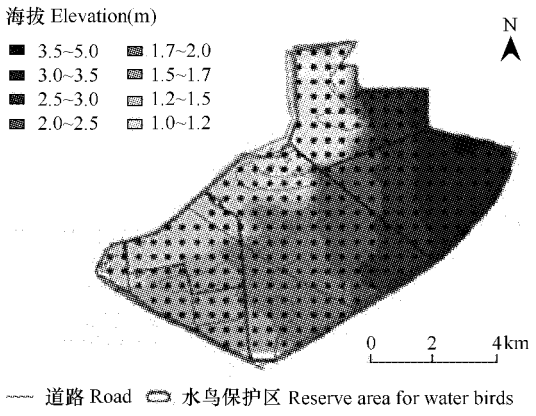


图 1　样点与南大港水库高程图层叠加

Fig. 1　Sketch map of the study area with elevation and locations of 289 samples for 450m grid

从统计学角度出发,对蝗虫产卵场所选择产生影响的各变量作为自变量,产卵与否则作为二分类因变量,由于模型的因变量是定性的(即蝗卵的有与无),因此多元线性回归使用定性因变量违反了其自身的假设条件,故采用 Logistic 回归模型来研究事件发生的概率与一组自变量间的关系[8]。

### 1.3　Logistic 回归模型组建

设 $P$ 为某事件发生的概率,取值范围为[0,1],$1-P$ 即为该事件不发生的概率,将两者比值取自然对数 $\log(P/(1-P))$,$P$ 为因变量,$x_k$为自变量,建立回归方程:

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right)=\alpha_j+\beta_1x_1+\beta_2x_2+\cdots+\beta_kx_k \tag{1}$$

(1)式即为 Logistic 回归模型,$\alpha$ 为常数,$\beta_i(i=1,2,\cdots,k)$ 为逻辑回归系数。等号左边为某一地点飞蝗卵块有与无的对数比。对于 Logistic 回归模型,观测到的不是 $P$ 本身,而是在自变量取不同值时属性变量出现某一结果的百分率或频率。

#### 1.3.1　变量选择(Selection of variables)

本研究共选出 8 个潜在影响东亚飞蝗产卵选择的环境变量:植物种类、植株密度、地形(阴坡,阳坡)、土壤含水量、含盐量、有机质及 pH。

#### 1.3.2　主效应预备模型(Preliminary main effect model)的组建

为保证(1)式右边项中的各协变量间相互独立,首先须对 1.3.1 中的各变量进行共线性诊断分析(不包括名义变量:植物种类和序次变量:植株密度、地形),然后分别采用最优子集(best-subsets)、逐步回归(stepwise method)以及 Hosmer & Lemeshow[9]推荐的方法选择主效应模型中的侯选变量。

(1)变量的相关和共线性(Multicollinearity)　共线性的协变量将显著增加公式(1)中系数估计的方差[10,11],并使得系数估计对样本和模型都非常敏感[12]。随着变量之间的相关程度提高,系数估计的标准差将显著增加,即两个相关程度很高的变量不能同时出现在一个模型中。容忍度(Tolerance)小于 0.2 是多元共线性存在的标志,小于 0.1 说明多元共线性很严重,若各变量的容忍度均大于 0.2,则可不考虑变量的多元共线性问题[13]。本研究亦采用容忍度指标分析多变量的共线性问题。有关计算用 SAS 8.0 的 CORR 过程进行分析。

(2)变量的最优子集选择　采用基于预测误差的 $C_P$统计量作为最优子集的选择准则。

(3)Hosmer & Lemeshow 协变量筛选方法　名义变量和序次变量采用列联表分析法,然后进行似然比 $x^2$ 检验(SAS 8.0,FREQ 过程);对连续变量则通过拟合单变量 Logistic 模型取得变量的 Wald 统计量显著性检验。为了不遗漏某些潜在可能的重要变量,本研究以 0.25 的显著水平为变量选择的标准,即检验通过 $P<$

0.25 的变量均作为多元模型的候选变量[9]。

(4)逐步回归方法 采用正向逐步选择(forward stepwise)、反向逐步选择 (backward stepwise)和混合逐步选择 (combined stepwise) 3 种方法进行模型的变量选择。变量加入和保留的标准取 SAS 的默认值($P$ = 0.05),指定加入变量和保留变量的标准为 $P<0.25$,

(5)主效应预备模型 为了最大程度保留潜在的重要变量,共拟出 14 个多元模型(见表2)。

### 1.3.3 主效应模型(Main effect model)的组建

1.3.2 得到的预备主效应模型仅仅说明协变量是影响蝗虫产卵选择的最主要参数,但模型中协变量间是线性还是非线性仍不能确定,因此需采用分数多项式法(Method of fractional polynomials)分析每个变量的具体形式[9,14]。

### 1.3.4 准最终模型(Preliminary final model)构建

检验主效应模型中各协变量以及未被包含在该模型里,但对飞蝗产卵选择有潜在影响的各变量(即1.3.1中的变量)间是否存在相互作用。于是列出 7 个具有一定实际或潜在生态意义的不同变量的相互作用项(即植株密度-阳坡、植株密度-阴坡、植被种类-阳坡、植被种类-阴坡、植株密度-植被种类、土壤含水量-阳坡、土壤含水量-阴坡)依次加入到主效应模型中,采用似然比$\chi^2$ 检验评价其显著性以确定是否能加入到主效应模型中,最终得到用于预测飞蝗产卵场所选择的 Logistic 模型。

## 1.4 模型评价

评价所组建的 Logistic 模型主要从 3 个方面[7]:

(1)拟合优度(Goodness of fit)统计指标 主要有皮尔逊$\chi^2$(Pearson $\chi^2$)、偏差(Deviance)、Hosmer-Lemeshow 拟合优度指标和信息测度指标(Information measures)。较小的$\chi^2$ 统计量意味着预测值与观测值之间没有显著差别,说明该模型能较好拟合数据。反之,如果$\chi^2$ 统计量很大,统计检验显著,表明拟合数据不佳。偏差 D 值越大,表示所建模型拟合越差,越小则说明所建模型拟合的越好。在其它条件相同时,模型的 AIC 和 SC 值(信息测度指标)越小,说明模型拟合的越好。

(2)预测准确性(Predictive accuracy) 本研究采用预测概率与观测值间的关联法评价所建 Logistic 模型的预测准确性。如果一个模型中 4 个序次相关指标(Somers' D,Gamma,Tau-a,C)的值较高,表明其有较强的预测能力,即预测概率与观测反应变量之间的关联较为紧密。反之,序次相关指标较低,说明模型的预测能力较差。

(3)模型$\chi^2$ 统计(Model chi-square statistic) 对模型$\chi^2$ 检验是为了对所建模型进行有意义的解释,即模型中所包含的自变量必须对因变量有显著的解释能力。如果模型$\chi^2$ 的统计性显著,则拒绝零假设,认为自变量所提供的信息能更好地预测事件是否发生。

# 2 结果与分析

## 2.1 东亚飞蝗产卵场所选择的 Logistic 回归模型

变量相关和共线性的分析结果表明土壤含盐量分别与土壤含水量和有机质之间存在极显著 ($P<0.001$) 的相关关系(表1),表明含盐量与土壤含水量、含盐量和有机质分别只能有一个变量作为公式(1)右边项中的协变量。

**表1 东亚飞蝗产卵场所选择 Logistic 回归模型所选变量的相关矩阵**

**Table 1 Correlation marix for selected soil variables in the Logistic model of oviposition site selection by *L. m. manilensis***

| 土壤特性 Soil properties | 含盐量 Salinity | 5 cm 含水量 Water content at 5 cm depth | pH | 有机质 Organic matter |
|---|---|---|---|---|
| 含盐量 Salinity | 1.00000 | 0.54471 | 0.13844 | -0.35065 |
| | | <0.0001 | 0.0166 | <0.0001 |
| 含水量 Water content | | 1.00000 | 0.0430 | -0.15846 |
| | | | 0.4589 | 0.0060 |
| pH Soil pH value | | | 1.0000 | -0.07288 |
| | | | | 0.2089 |
| 有机质 Organic matter | | | | 1.0000 |

模型 12 的 AIC 和 SC 值分别高达 314.779 和 325.880,且模型中的两个变量(土壤含水量和植株密度)的 $P$ 值均 <0.0001(表 2),再综合比较各模型中变量的 Wald 统计量和模型的预测准确性 $R$-square 指标后,将模型 12 作为预备主效应模型。然后用分数多项式法分析土壤含水量(water)和植株密度 (veg_d) 的具体形式,结果表明将土壤含水量取非线性形式合适(植株密度为序次变量不作检验),且检验结果表明,土壤含水量取如下形式拟合效果最好。$\beta_1 \times \left(\frac{1}{\text{water}}\right) + \beta_2 \times \log(\text{water})$,其中 $\beta_1$ 和 $\beta_2$ 为模型参数。

**表 2　东亚飞蝗产卵场所选择多元变量 Logistic 回归模型拟合结果**

**Table 2　Results of fitting multivariable models**

| 模型 Model | 变量 Variables | 系数 Coefficient | 标准误 SE | Wald Chi-square:Z | $P>\|Z\|$ | AIC | SC | −2logL | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 土壤含盐量 Salinity | −0.41998 | 0.2355 | 3.1801 | 0.0745 | 248.027 | 273.836 | 234.047 | 0.2988 |
| | 土壤 pH Soil pH value | −1.57437 | 0.9895 | 2.5314 | 0.1116 | | | | |
| | 阴坡 North slope | −12.8713 | 297.9 | 0.0019 | 0.9655 | | | | |
| | 阳坡 South slope | 16.1676 | 325.9 | 0.0025 | 0.9604 | | | | |
| | 植被种类 Veg_sp | 0.32757 | 0.4236 | 0.5978 | 0.4394 | | | | |
| | 植株密度 Veg_d | −0.95317 | 0.2627 | 13.1672 | 0.0203 | | | | |
| 2 | 土壤含水量 Water Content | −0.0394 | 0.0139 | 8.0799 | 0.0045 | 243.087 | 268.896 | 229.087 | 0.3104 |
| | 土壤 pH Soil pH value | −1.7866 | 1.0053 | 3.1578 | 0.0755 | | | | |
| | 阴坡 North slope | −12.8164 | 300.2 | 0.0018 | 0.9659 | | | | |
| | 阳坡 South slope | 16.0402 | 325.6 | 0.0024 | 0.9607 | | | | |
| | 植被种类 Veg_sp | 0.2833 | 0.4267 | 0.4407 | 0.5068 | | | | |
| | 植株密度 Veg_d | −0.9495 | 0.2623 | 13.1060 | 0.003 | | | | |
| 3 | 土壤含盐量 Salinity | −0.0441 | 0.2394 | 3.5262 | 0.0604 | 252.078 | 274.280 | 240.078 | 0.2967 |
| | 土壤 pH Soil pH value | −1.8343 | 0.9540 | 3.6974 | 0.0545 | | | | |
| | 阴坡 North slope | −12.8892 | 293.8 | 0.0019 | 0.9650 | | | | |
| | 阳坡 South slope | 16.1089 | 320.2 | 0.0025 | 0.9599 | | | | |
| | 植株密度 Veg_d | −0.9986 | 0.2607 | 14.670 | 0.0001 | | | | |
| 4 | 土壤含水量 Water content | −0.0398 | 0.0138 | 8.3441 | 0.0039 | 247.299 | 269.501 | 235.299 | 0.3078 |
| | 土壤 pH Soil pH value | −2.0214 | 0.9684 | 4.3571 | 0.0369 | | | | |
| | 阴坡 North slope | −12.8199 | 296.8 | 0.0019 | 0.9659 | | | | |
| | 阳坡 South slope | 15.9902 | 319.4 | 0.0025 | 0.9601 | | | | |
| | 植株密度 Veg_d | −0.9900 | 0.2585 | 14.6638 | 0.0001 | | | | |
| 5 | 土壤含水量 Water content | −0.0384 | 0.0136 | 7.9092 | 0.0049 | 244.247 | 266.368 | 232.247 | 0.3030 |
| | 阴坡 North slope | −12.9416 | 302.3 | 0.0018 | 0.9659 | | | | |
| | 阳坡 South slope | 16.0929 | 325.9 | 0.0024 | 0.9609 | | | | |
| | 植株密度 Veg_d | −0.8877 | 0.2356 | 11.9692 | 0.0005 | | | | |
| | 植被种类 Veg_sp | 0.4569 | 0.4135 | 1.2208 | 0.2692 | | | | |
| 6 | 土壤含水量 Water content | −0.0397 | 0.0137 | 8.3907 | 0.0038 | 252.136 | 270.638 | 242.136 | 0.2918 |
| | 土壤 pH Soil pH value | −2.1946 | 0.9624 | 5.1997 | 0.0226 | | | | |
| | 阳坡 South slope | 16.0735 | 318.4 | 0.0026 | 0.9567 | | | | |
| | 植株密度 Veg_d | −1.0115 | 0.2505 | 16.3107 | <0.0001 | | | | |
| 7 | 土壤含水量 Water content | −0.0387 | −0.0135 | 8.2006 | 0.0042 | 249.672 | 268.174 | 239.62 | 0.2976 |
| | 阴坡 North slope | −12.9651 | 299.2 | 0.0019 | 0.9655 | | | | |
| | 阳坡 South slope | 16.0674 | 318.2 | 0.0026 | 0.9597 | | | | |
| | 植株密度 Veg_d | −0.9320 | 0.2542 | 13.4467 | 0.0002 | | | | |
| 8 | 土壤含水量 Water content | −0.0381 | 0.0135 | 7.9588 | 0.0048 | 249.722 | 268.157 | 239.722 | 0.2851 |
| | 阳坡 South slope | 16.1963 | 324.4 | 0.0025 | 0.9602 | | | | |
| | 植被种类 Veg_sp | 0.4560 | 0.4102 | 1.2358 | 0.2663 | | | | |
| | 植株密度 Veg_d | −0.9169 | 0.2479 | 13.6764 | 0.0002 | | | | |
| 9 | 土壤含水量 Water content | −0.0384 | 0.0134 | 8.2126 | 0.0042 | 255.360 | 270.162 | 247.360 | 0.2793 |
| | 阳坡 South slope | 16.1659 | 316.9 | 0.0026 | 0.9593 | | | | |

续表

| 模型 Model | 变量 Variables | 系数 Coefficient | 标准误 SE | Wald Chi-square:Z | $P>\|Z\|$ | AIC | SC | -2logL | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|
| | 植株密度 Veg_d | -0.9515 | 0.2457 | 14.9927 | 0.0001 | | | | |
| 10 | 土壤 pH Soil pH value | -2.0986 | 0.9406 | 4.9784 | 0.0257 | 259.159 | 273.961 | 251.159 | 0.2701 |
| | 阳坡 South slope | 16.3163 | 319.7 | 0.0026 | 0.9593 | | | | |
| | 植株密度 Veg_d | -0.915 | 0.2410 | 14.0987 | 0.0002 | | | | |
| 11 | 阴坡 North slope | -12.9669 | 301.7 | 0.0018 | 0.9567 | 256.476 | 271.277 | 248.476 | 0.2766 |
| | 阳坡 South slope | 16.3171 | 319.7 | 0.0026 | 0.9593 | | | | |
| | 植株密度 Veg_d | -0.8423 | 0.2456 | 11.7581 | 0.0006 | | | | |
| 12 | 土壤含水量 Water content | -0.0479 | 0.0123 | 16.4220 | <0.0001 | 314.779 | 325.880 | 308.779 | 0.1150 |
| | 植株密度 Veg_d | -0.9028 | 0.2112 | 18.2677 | <0.0001 | 13 | | | |
| | 阳坡 South slope | 16.4027 | 318.7 | 0.0026 | 0.9590 | 262.168 | 273.270 | 256.168 | 0.2578 |
| | 植株密度 Veg_d | -0.8518 | 0.2370 | 12.9193 | 0.0196 | | | | |
| 14 | 土壤含水量 Water content | -0.0294 | 0.0126 | 5.4516 | 0.0196 | 272.273 | 283.374 | 266.273 | 0.2323 |
| | 阳坡 South slope | 16.0866 | 334.7 | 0.0073 | 0.9617 | | | | |

Veg_sp 和 Veg_d 分别代表植被种类和植株密度 Veg_sp and Veg_d indicate vegetation species and vegetation density respectively

**表 3 包含相互作用项的多变量 Logistic 回归模型拟和结果**

**Table 3 Results of fitting multivariable models containing interaction items**

| 新增变量 Added variables | 系数 Coefficient | 标准误 SE | Wald Chi-Square:Z | $P>\|Z\|$ | AIC | -2LogL | $R^2$ |
|---|---|---|---|---|---|---|---|
| 阳坡-植株密度 SS-Veg_d | 11.7434 | 173.9 | 0.0046 | 0.9462 | 249.029 | 239.079 | 0.4366 |
| 阳坡-植被种类 SS-Veg_sp | -9.9439 | 132.9 | 0.0056 | 0.9403 | 291.634 | 281.634 | 0.2800 |
| 阴坡-植株密度 NS- Veg_d | 16.0169 | 368.2 | 0.0019 | 0.9613 | 255.512 | 245.512 | 0.3968 |
| 阳坡-植被种类 NS-Veg_sp | -12.9723 | 335.4 | 0.0015 | 0.9691 | 289.485 | 279.485 | 0.2665 |
| 植被种类-密度 Veg_sp- d | 0.2956 | 0.1653 | 3.1985 | 0.0737 | 292.239 | 282.239 | 0.2552 |
| 土壤含水量-阳坡 WC- SS | 1.3121 | 17.682 | 0.0055 | 0.9411 | 249.079 | 239.079 | 0.4366 |
| 土壤含水量-阴坡 WC- NS | -1.2185 | 15.3813 | 0.0063 | 0.9369 | 291.634 | 281.634 | 0.2800 |

SS 和 NS 分别表示阳坡和阴坡,Veg_sp 和 Veg_d 分别代表植被种类和植株密度,WC 表示土壤含水量 SS,NS,Veg_sp,Veg_d and WC indicate south slope and north slope,vegetation species,vegetation density and soil water content respectively

表 3 结果表明相互作用项的估计系数均不显著,因而均不能加入到模型中。于是,得到用于预测东亚飞蝗产卵场所选择的准最终模型:

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right)=21.63-\frac{76.32}{\text{water}}-5.43\log(\text{water})-0.86(\text{veg_d}) \quad (2)$$

准最终模型(2)式的极大似然估计结果表明 Logistic 回归模型中的两个协变量(土壤含水量和植株密度)的效应极显著($P<0.001$)(表 4),说明用所建立的 Logistic 回归模型预测飞蝗产卵选择与这两个环境变量的关系是合理的。

**表 4 东亚飞蝗产卵场所选择的 Logistic 回归模型的极大似然估计**

**Table 4 MLE of Logistic model for *Locusta migratoria manilensis* oviposition site selection**

| 变量 Variables | DF | 参数估计 Parameter estimate | 标准误 SE | Wald Chi-Square | Pr> Chi-Square | Standardized Estimate |
|---|---|---|---|---|---|---|
| 截距 Intercept | 1 | 21.6291 | 4.2304 | 28.1402 | <0.001 | |
| 土壤含水量 a Water content a | 1 | -76.3173 | 17.4726 | 19.0778 | <0.001 | -1.5346 |
| 土壤含水量 b Water content b | 1 | -5.4342 | 1.0775 | 25.4351 | <0.001 | -1.7494 |
| 植株密度 Veg_density | 1 | -0.8586 | 0.2154 | 15.8248 | <0.001 | -0.3985 |

土壤含水量 a 和土壤含水量 b 分别表示公式(2)中的 1/water 和 log(water) Water content a and water content b indicate1/water and log(water) of the formula (2) respectively

## 2.2　所组建 Logistic 回归模型的评价

预测值与观测值之间统计检验结果($\chi^2$ 和 D 值)不显著(表 5),说明所建模型能很好地拟合数据。从 Hosmer-Lemeshow 检验指标结果可看出(表 6),*HL* 指标值为 8.6725,统计不显著($P=0.3707$),故不能拒绝关于模型拟合数据很好的假设,即模型能很好地拟合了数据。从表 7 结果看出,$C=0.764$,Somers' D = 0.529,由此判断所组建的用于预测飞蝗产卵选择的 Logistic 回归模型公式(2)具有较好的预测能力。表 8 是对飞蝗产卵选择的 Logistic 回归模型的 $\chi^2$ 统计结果,$P<0.0001$ ($\chi^2=345.296-290.968=54.328$, DF $=4-1=3$) 表明模型 $\chi^2$ 的统计性显著,说明自变量所提供的信息能较好地预测事件是否发生。

综合以上结果,即模型 $\chi^2$ 的统计性检验显著而拟合优度统计性不显著,说明所组建的用于预测飞蝗产卵选择的 Logistic 回归模型是可靠的。

## 3　结果与讨论

正如线性回归模型一样,Logistic 回归模型也能获得事件发生的预测概率。如果已知参数估计 $\alpha$ 和 $\beta_k$,且已知 $x_1$ 到 $x_2$ 的所有值,即可利用 Logistic 回归公式,计算出预测概率。本研究通过筛选研究区域内可能影响飞蝗产卵场所选择的环境变量,结果表明植株密度和土壤 5 cm 含水量是影响飞蝗产卵场所选择的最重要因素。通过用不同方法对飞蝗产卵场所选择的 Logistic 模型的评价结果表明,所组建的模 S 型是合理可靠的。

建立模型时,某些变量统计上不显著并不表示其对所研究的实际问题的影响不明显。在本研究中,土壤含盐量和向阳坡虽未能选入模型,但其对蝗虫产卵选择的影响却是不可忽视的。两年的野外调查发现,多数飞蝗卵块集中分布在植被稀疏、土壤含水量、含盐量适宜的向阳坡区域。分析两者未被选入模型的原因:(1)统计结果表明,土壤含盐量与 5 cm 含水量存在高度相关性($P<0.0001$),而模型要求所包含的变量之间须相互独立,而土壤含盐量采用 4 种方法筛选均未被选入模型中;(2)土壤含水量亦是微地形(microrelief)差异的反映。野外调查发现,阳坡和沟的阳面、植被稀疏的平地土壤含水量一般较低,而阴坡和沟的阴面、植株密度较高的地块则伴随着较高的土壤含水量。因此,是土壤含水量而不是含盐量被选入准最终模型应该是合理的。

本研究在建立蝗虫产卵选择的 Logistic 模型过程中,采用 $C_P$ 最优子集、逐步回归等多种方法选择变量,但不同方法的选择结果并不完全一致,特别是采用 SAS 本身默认变量进入或保留在模型中的显著水平后($P<0.05$),就遗漏了某些重要变量或根本无法选出有效变量。因此,本文调整了选进与保留变量的阈值($P<0.25$)。本研究结果及已有报道[15]表明,建立模型必须基于多种变量选择方法,并对不同方法所选的变量加以比较,才能使有重要意义的变量不被删除或没有意义的变量却被包含在模型当中。

**表 5　东亚飞蝗产卵场所选择 Logistic 回归模型 $x^2$ 和偏差统计量**

**Table 5　$x^2$ and Deviance of Logistic model for *Locusta migratoria manilensis* oviposition site selection**

| 标准 Criterion | *df* | 值 Value | Value/DF | Pr > Chi-square |
|---|---|---|---|---|
| Pearson | 248 | 277.9417 | 1.1298 | 0.0790 |
| Deviance | 248 | 254.9239 | 1.0363 | 0.3348 |
| Number of unique profiles: 250 | | | | |

**表 6　东亚飞蝗产卵场所选择 Logistic 回归模型的 Hosmer-lemeshow 拟合优度指标**

**Table 6　Hosmer and Lemeshow goodness-of-fit test for Logistic model**

| | | *Y*=1 | | *Y*=0 | |
|---|---|---|---|---|---|
| 组 Group | Total | Observed | Expected | Observed | Expected |
| 1 | 30 | 2 | 1.12 | 28 | 28.88 |
| 2 | 30 | 4 | 2.17 | 26 | 27.83 |
| 3 | 31 | 3 | 3.14 | 28 | 27.86 |
| 4 | 30 | 2 | 4.10 | 28 | 25.90 |
| 5 | 31 | 3 | 5.72 | 28 | 25.28 |
| 6 | 30 | 8 | 8.24 | 22 | 21.76 |
| 7 | 30 | 10 | 10.35 | 20 | 19.65 |
| 8 | 30 | 14 | 13.09 | 16 | 16.91 |
| 9 | 31 | 20 | 15.68 | 11 | 15.32 |
| 10 | 26 | 13 | 15.38 | 13 | 10.62 |
| | Chi-Square | DF | Pr > ChiSq | | |
| | 8.6725 | 8 | 0.3707 | | |

**表 7　东亚飞蝗产卵场所选择 Logistic 回归模型预测概率与观测值之间的关联**

**Table 7　Association of predicted probabilities and observed responses for Logistic model**

| | |
|---|---|
| Concordant = 76.3 | Somers'D = 0.529 |
| Discordant = 23.4 | Gamma = 0.531 |
| Tied = 0.3 | Tau - a = 0.206 |
| Pairs = 17380 | C = 0.764 |

表 8 东亚飞蝗产卵场所选择 Logistic 回归模型的 $\chi^2$ 统计结果

Table 8 Chi-Square test of Logistic regression model

| Model Fitting Information and Testing Global Null Hypothesis BETA = 0 模型检验,无效假设为总体 $\beta = 0$ | | | |
|---|---|---|---|
| Criterion | Intercept Only | Intercept and Covariates | Chi-Square for Covariates |
| AIC | 347.296 | 298.968 | · |
| SC | 350.997 | 313.769 | · |
| -2LOG L | 345.296 | 290.968 | · |
| Likelihood ratio | · | · | 54.3288 with 3 DF ($P < 0.0001$) |
| Score | · | · | 47.7008 with 3 DF ($P < 0.0001$) |
| Wald | · | · | 40.6158 with 3 DF ($P < 0.0001$) |
| R-square = 0.1661 | Max-rescaled R-Square = 0.2426 | | |

建立统计模型时,通常需要较多的样本量,否则,较少的样本量会使得模型估计的风险增高[16]。在野外研究中,所得观测数据一般有限,很难达到建立模型所需的合适样本数量,进而使检验变得较为困难。本研究采取了以尽量多的样本数量(450 m 栅格取样 289 个样点;50 m 栅格取样 2601 个样点)弥补时间尺度上的(2a)不足,在理论与方法上都是可行的。

研究结果为利用遥感手段监测蝗虫孳生地(产卵地)提供了科学依据。在秋蝗产卵盛期,通过遥感数据准确反演蝗区土壤含水量和植被覆盖度,就可能事先绘制出飞蝗最有可能的产卵区域,这不仅有利于科学合理地进行飞蝗卵块的野外调查,更能为在蝗灾发生早期采取因地制宜的防治措施提供依据,从而达到蝗灾早期遥感监测与预警的目的。

**References:**


[1] You D S, Ma S J. Relationship of *Locusta migratoria manilensis* oviposition and hatch to soil salinity. Acta Phytophylacica Sinica, 1964, 3 (4): 333 - 344.

[2] You Q, Guo F, Chen Y L, *et al*. Habits of *Locusta migratoria manilensis* (Meyen). Acta Entomologica Sinica, 1958, 8 (2): 119 - 135.

[3] Zhu E L, Occurrence and Management of the Oriental Migratory Locust in China. Beijing: China Agriculture Press, 1999. 3 - 38.

[4] Li Y D. Hydraulic Annals of Nandagang State Farm. Tianjin: Tianjin People Press, 1993.

[5] Ji R, Xie BY, Li Z, *et al*. Reasons and characteristics of outbreak of the Oriental migratory locust *Locusta migratoria manilensis* plague in Nandagang, Hebei Province. Entomological Knowledge, 2002, 39 (6): 430 - 433.

[6] Ma S J, *et al*. Studies on Breeding Areas of the Oriental Migratory Locust in China. Beijing: Science Press, 1965. 1 - 335.

[7] Ji R, Xie B Y, Li Z, Li D M, *et al*. Spatial distribution of the Oriental Migratory Locust (Orthoptera: Acrididae) egg pods studied with GIS and GS. Acta Entomologica Sinica, 2006, 49 (3): 410 - 415.

[8] Wang J C, Guo Z G. Logistic Regression Model-Method and Application. Beijing: Higher Education Press, 2001.

[9] Hosmer D W, Lemeshow S. Applied Logistic Regression, 2nd ed. New York: John Wiley and Sons, 2000.

[10] Weisberg S. Applied Linear Regression. New York: John Wiley and Sons, 1985.

[11] Berry W D, Feldman S. Multiple Regression in Practice. Newbury Park, CA: Sage Publication, 1985.

[12] Hanushek E A, Jackson J E. Statistical Methods for Social Scientists. Orlando: Academic Press, 1977.

[13] Menard S. Applied Logistic Regression Analysis. California: Sage Publication, 1995. 116 - 173.

[14] Royston P, Altman D G. Regression using fractional polynomials of continuous covariates: Parsimonious parametric modeling (with discussion). Appl. Stat., 1994, 43 (4): 429 - 467.

[15] Gao Z X. Outbreak mechanism of *Helicoverpa armigera* (H bner) in Tongzhou Jiangsu Province. Ph. D dissertation of Institute of Zoology, Chinese Academic of Sciences, 2002, 80 - 104.

[16] Long S J. Regression Model for Categorical and Limited Dependent Variables. California: Sage Publication, 1997.

**参考文献:**

[1] 尤端淑, 马世骏. 东亚飞蝗产卵及蝗卵孵化与土壤含盐量的关系. 植物保护学报,1964,3(4):333~344.

[2] 尤其儆, 郭郛, 陈永林, 等. 东亚飞蝗的生活习性. 昆虫学报, 1958, 8(2):119~135.

[3] 朱恩林. 中国东亚飞蝗发生与治理. 北京:中国农业出版社, 1999. 3~38.

[4] 李贻铎. 南大港农场水利志. 天津:天津人民出版社, 1993. 41~48.

[5] 季荣, 谢宝瑜, 李哲, 等. 河北省南大港农场 2002 年夏蝗发生特点及原因浅. 昆虫知识, 2002,39(6):430~433.

[6] 马世骏,等. 中国东亚飞蝗蝗区的研究. 北京:科学出版社, 1965. 1~335.

[7] 季荣,谢宝瑜,李哲,等. 基于 GIS 和 GS 的东亚飞蝗卵块空间格局的研究. 昆虫学报,2006, 49(3):410~415.

[8] 王济川, 郭志刚. Logistic 回归模型——方法与应用. 北京:高等教育出版社, 2001.

[15] 高增祥. 江苏通州棉铃虫 *Helicoverpa armigera* (Hübner) 种群暴发的机理研究. 中国科学院研究生院博士学位论文, 2002. 80~104.