

ARIMA 模型在小菜蛾幼虫种群动态中的应用

岑冠军¹, 黄寿山², 肖莉¹, 钟谭卫¹

(1 华南农业大学 理学院, 广东 广州 510642; 2 华南农业大学 昆虫生态研究室, 广东 广州 510642)

摘要:在定义了害虫种群动态的 ARIMA 模型后,应用扩展样本自相关函数法(ESACF)和最小信息准则法(MINIC)对 ARIMA 模型定阶,并按照模型识别、参数估计与检验和模型诊断3个步骤,分别建立了菜心 *Brassica parachinensis* 和芥兰 *Brassica albograbra* 上小菜蛾 *Plutella xylostella* 幼虫种群动态的 ARIMA(2,1,3)模型和 ARIMA(0,1,5)模型,其预测值误差均方分别为 0.213 和 0.188,拟合后的残差序列为白噪声序列.结果表明在应用 ARIMA 模型研究害虫种群动态时,采用扩展样本自相关函数法和最小信息准则法定阶能快速寻找到最优的预测模型.

关键词:ARIMA; 种群动态; 害虫; 应用

中图分类号:O29

文献标识码:A

文章编号:1001-411X(2008)01-0109-05

The Application of ARIMA Model to the Larva's Population Dynamic of *Plutella xylostella*

CEN Guan-jun¹, HUANG Shou-shan², XIAO Li¹, ZHONG Tan-wei¹

(1 College of Sciences, South China Agricultural University, Guangzhou 510642, China;

2 Lab of Insect Ecology, South China Agricultural University, Guangzhou 510642, China)

Abstract: With the definition of ARIMA model of pest population dynamic given, the ESACF and MINIC methods were used to tentatively identify the orders of the ARIMA models. The ARIMA(2,1,3) model on *Brassica parachinensis* and ARIMA(0,1,5) model on *Brassica albograbra* of the larva's population dynamic of *Plutella xylostella* were constructed respectively according to the three steps of identification, parameters estimation and test, and diagnostic checking of models. The RMSEs of the two models were 0.213 and 0.188 respectively and the residuals were white noise. The results showed that tentative order identification by the ESACF and MINIC methods could find adaptive models quickly in the application of ARIMA model to pest population dynamics.

Key words: ARIMA; population dynamics; insect pests; application

害虫种群动态研究中很多数据是以时间序列的形式出现的;如每隔 5 d 调查 1 次的害虫种群密度序列,近年来越来越多的学者应用时间序列分析方法研究害虫种群动态变化并进行预测分析.模型定阶是应用 ARIMA 模型最关键的步骤,Tsay 等^[1]在 1984 年提出了扩展样本自相关函数法(Extended Sample Autocorrelation Function, ESACF)的定阶方法,Choi 等^[2]在 1992 年给出了该方法的算法,Hannan 等^[3]在 1982 提出了最小信息准则法(MINimum Informa-

tion Criterion, MINIC),Box 等^[4]和 Choi 等^[2]分别在 1994 年和 1992 年给出了该方法的算法,目前在应用 ARIMA 模型研究害虫种群动态时,采用上述 2 个方法进行模型定阶的研究鲜见报道.本文以小菜蛾幼虫种群动态为研究对象,在应用 ARIMA 模型时采用上述 2 个方法定阶,并且在 SAS9.0 支持下,严格按照 ARIMA 模型的建模步骤来建立模型,为时间序列模型在害虫种群动态中的应用提供更科学合理的建模方法和步骤.

收稿日期:2007-03-20

作者简介:岑冠军(1980—),男,助理实验师,硕士;通讯作者:钟谭卫(1962—),男,副教授,E-mail:zhongtanwei@scau.edu.cn

1 原理与方法

1.1 种群动态的 ARIMA 模型

ARIMA 模型可称为自回归求和滑动平均 (autoregressive integrated moving average) 模型,是由 Box 等^[4]提出的一种时间序列预测方法.

定义 1 设 N_t 为 t 时刻种群密度值,令 $L_t = \ln(N_t)$, B 表示后移算子, $BL_t = L_{t-1}$, ARIMA(p, d, q) 模型可表示为

$$(1 - B)^d L_t = \mu + \frac{\theta(B)}{\phi(B)} a_t,$$

其中

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p,$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q,$$

p 和 q 分别是自回归部分和滑动平均部分的阶数, d 为差分阶数, μ 为序列均值. 对于波动比较大的或者不平稳的种群动态序列,如果差分后可以变平稳, ARIMA 模型将有着很好的应用.

1.2 种群动态 ARIMA 模型的识别

1.2.1 种群动态数据的预处理 在模型识别阶段,首先对采集到的原始数据进行预处理,这部分主要检验原始数据的平稳性、独立性和周期性. 种群密度制约的时滞效应通常引起种群动态的不平稳^[5-6],环境因素的剧烈变化也会增加种群数量的波动,当种群动态序列有一个单位根时该序列是不平稳的,因此检验时间序列的平稳性可以转化为检验该序列是否存在单位根, Dickey 等^[7]在 20 世纪 70 年代中后期提出了检验单位根是否存在的 ADF (augmented dickey-fuller unit root test) 检验, ADF 检验的零假设是所研究的时间序列为存在单位根的序列,对应的备择假设为该序列为趋势平稳序列,计算统计量 ρ 和 τ ,并对其检验. 在对原始数据序列进行 ADF 检验后,如果原始数据序列是不平稳的,则要进行差分,对差分后的序列再检验其平稳性,直至差分后的序列为平稳序列.

1.2.2 种群动态 ARIMA 模型的定阶 实际时间序列模型的阶次虽是客观存在的,但又往往是未知的,因此用某种方法确定模型的阶次是系统建模的重要一环.

Tasy 等^[1]在 1984 年提出了利用扩展样本自相关函数法 (ESACF). 设 $\{x_t\}$ 是零均值的 ARMA 系列,其自回归部分的阶数是 p ,滑动平均部分的阶数是 q ,我们试探性地假设 $\{x_t\}$ 序列为 ARMA(m, j),记

$$w_t^{(m,j)} = \theta_{(m,j)}(B) a_t = x_t - \sum_{i=1}^m \hat{\phi}_i^{(m,j)} x_{t-i}, \quad (1)$$

式(1)中 $\{w_t^{(m,j)}\}$ 是在假定 $\{x_t\}$ 序列为 ARMA(m, j) 时的滑动平均部分,在此假设下序列 $\{x_t\}$ 的纯自回归部分 $\hat{\phi}_i^{(m,j)}$,记 $r_{j(m)}$ 为序列 $\{w_t^{(m,j)}\}$ 滞后 j 步的自相关函数,即扩展样本自相关函数. 由于序列 $\{x_t\}$ 的真实模型是 ARMA(p, q),因此有

$$\begin{aligned} r_{j(m)} &\approx 0, & (q > j > m - p) \\ r_{j(m)} &\neq 0, & (m - p \geq j - q \geq 0) \end{aligned}$$

因此,如果依次计算自回归部分的阶数 $m = 0, 1, \cdots, p_{\max}$,然后对每一个假定的自回归阶数 m ,依次计算滑动平均部分阶数 $j = 0, 1, \cdots, q_{\max}$ 时扩展样本自相关函数 $r_{j(m)}$,并写成表格的形式,则表格中 $r_{j(m)}$ 为 0 的区域是一个三角形区域,样本扩展自相关函数法就是要找到这个三角形区域的顶点所对应的 m 值和 j 值,分别作为 ARIMA 模型的自回归部分和滑动平均部分的阶. 对于扩展样本自相关函数 $r_{j(m)}$ 是否等于 0,由于随机误差的存在,接近于 0 的数值难以直观确定,因此每一个 $r_{j(m)}$ 进行原假设为 $H_0: r_{j(m)} = 0$ 的检验,如果检验的概率值大于检验水平,则接受原假设,即认为 $r_{j(m)} = 0$.

Hannan 等^[3]在 1982 年提出了最小信息准则法 (MINIC). MINIC 的步骤如下:

1) 应用 AIC (Akaike Information Criterion) 准则选择一个足够高阶 AR 模型去近似未知的 ARMA 模型,以此得到随机干扰序列 a_t 的估计 $\hat{a}_t$. 这一步通过 Yule-Walker 方程来估计参数 $\hat{\phi}_i^{(p_a)}$, AIC 准则函数是

$$\text{AIC}(p_a, 0) = \lg \hat{\sigma}_{a(p_a, 0)}^2 + 2(p_a + 0)/N, \quad (2)$$

其中, $\hat{\sigma}_{a(p_a, 0)}^2 = \frac{1}{N} \sum_{t=p_a}^n \hat{a}_t^2$, 如果 $p_a = h$ 时,式(2)取得

最小值,则随机干扰序列为 $\hat{a}_t = x_t - \sum_{i=1}^h \hat{\phi}_i^{(h)} x_{t-i}$.

2) 应用 BIC (bayesian information criterion) 准则来确定模型阶数. 由第 1 步估计的随机干扰序列 $\{\hat{a}_t\}$ 和原序列 $\{x_t\}$ 构建回归方程

$$x_t = \sum_{i=1}^m \hat{\phi}_i^{(m,j)} x_{t-i} + \sum_{k=1}^j \theta_k^{(m,j)} \hat{a}_{t-k} + \varepsilon, \quad (3)$$

这里假定该序列纯自回归部分的阶数 $m = p_{\min}, p_{\min} + 1, \cdots, p_{\max}$, 滑动平均部分阶数 $j = q_{\min}, q_{\min} + 1, \cdots, q_{\max}$, 方程(3)可通过普通最小二乘估计得到参数的解, BIC 准则函数为

$$\text{BIC}(m, j) = \lg \hat{\sigma}_{(m,j)}^2 + 2(m + j) \lg(N)/N, \quad (4)$$

式(4)中, $\hat{\sigma}_{(m,j)}^2 = \frac{1}{n} \sum_{t=h+\max(m,j)}^N (x_t - \sum_{i=1}^m \hat{\phi}_i^{(m,j)} x_{t-i} - \sum_{k=1}^j \theta_k^{(m,j)} \hat{a}_{t-k})^2$, 应用 BIC 准则选择使式(4)的值最

小的 m 和 j 作为 ARMA 模型的阶数。

1.3 种群动态 ARIMA 模型的参数估计和诊断

当确定了 ARIMA 模型各部分的阶数后,就要对模型的参数进行估计和检验,去掉模型中不必保留的参数,同时对模型拟合后的残差进行白噪声检验,如果残差不是白噪声,要进一步挖掘残差中的相关信息。

2 实例分析

2.1 原始数据的预处理

2003-01-05—2004-12-30 在 2 种主要蔬菜菜心 *Brassica parachinensis* 和芥兰 *Brassica albograbra* 上每

隔 5 d 调查 1 次小菜蛾 *Plutella xylostella* 幼虫种群密度动态序列 N_t , 令 $L_t = \ln N_t$ 。

考虑到小菜蛾幼虫种群密度动态可能受到季节性因子的影响,因此在滞后 30、60(或 70)和 120 d 下作单位根检验。对菜心和芥兰上的小菜蛾幼虫种群密度动态序列作阶数为 6、12(或 14)和 24 下的 ADF 检验,表 1 中列出了 ADF 检验采用确定性趋势模型时的统计量值。

从表 1 可以看出,2 种蔬菜上小菜蛾幼虫种群动态序列在滞后 24 步时均出现单位根,2 种蔬菜上小菜蛾幼虫种群动态序列均不平稳,需要差分。

表 1 小菜蛾幼虫种群动态序列的单位根检验

Tab. 1 Seasonal augmented dickey-fuller unit root tests for the larva's population dynamic of *Plutella xylostella*

蔬菜 vegetable sample	阶数 lags	ρ	P	τ	P	F	P
菜心 <i>Brassica parachinensis</i>	6	582.579 2	0.999 9	-5.42	<0.000 1	14.72	0.001 0
	12	443.101 1	0.999 9	-3.62	0.031 7	6.83	0.036 6
	24	26.352 9	0.999 9	-2.60	0.280 6	3.54	0.470 4
芥兰 <i>Brassica albograbra</i>	6	-34.558 0	0.002 0	-3.13	0.104 0	4.94	0.191 0
	12	-59.811 0	0.001 0	-2.85	0.184 0	4.34	0.309 0
	24	61.018 0	0.999 0	-2.46	0.349 0	3.14	0.549 0

对菜心和芥兰上小菜蛾幼虫种群动态序列进行一阶滞后一步差分,则差分后的序列为 $L'_t = (1 - B)L_t$,对差分后的序列采用 ADF 检验,表 2 中列出了 ADF 检验的采用单均值模型时的统计量值。

从表 2 可以看出,2 种蔬菜上小菜蛾幼虫种群动

态序列差分后的序列是平稳的序列。

2.2 模型定价与参数估计

2.2.1 ARIMA 模型在菜心上小菜蛾幼虫种群动态中的应用 对差分后的平稳序列计算扩展样本自相关函数及其对应的概率值,结果见表 3。

表 2 差分后小菜蛾幼虫种群动态序列的单位根检验

Tab. 2 Augmented dickey-fuller unit root tests for the larva's population dynamic of *Plutella xylostella* after differencing

蔬菜 vegetable sample	阶数 lags	ρ	P	τ	P
菜心 <i>Brassica parachinensis</i>	6	-117.466 0	0.000 1	-10.30	<0.000 1
	12	-69.178 8	0.001 3	-6.82	<0.000 1
	24	-44.794 7	0.001 3	-4.77	<0.000 1
芥兰 <i>Brassica albograbra</i>	6	-142.079 0	0.000 1	-13.07	<0.000 1
	12	-95.743 8	0.001 3	-9.43	<0.000 1
	24	-55.174 2	0.001 3	-5.86	<0.000 1

表 3 $m=1,2,\dots,8; j=1,2,\dots,8$ 时扩展样本自相关函数的概率值

Tab. 3 ESACF probability values with $m=1,2,\dots,8; j=1,2,\dots,8$

m	j								
	0	1	2	3	4	5	6	7	8
0	0.043 5	0.238 7	0.747 4	0.917 8	0.403 3	0.921 9	0.015 4	0.662 3	0.634 9
1	<0.000 1	0.583 0	0.582 5	0.966 8	0.150 5	0.957 1	0.011 9	0.297 3	0.942 9
2	<0.000 1	0.219 8	0.010 9	0.911 9	0.904 5	0.938 8	0.046 1	0.372 5	0.718 3
3	0.000 9	0.637 3	0.844 3	0.650 4	0.732 0	0.985 8	0.169 0	0.849 6	0.795 6
4	0.005 3	<0.000 1	0.767 3	0.648 9	0.864 9	0.889 9	0.162 3	0.652 3	0.947 6
5	<0.000 1	0.000 5	0.045 5	0.968 5	0.218 2	0.789 4	0.185 5	0.702 7	0.994 3
6	0.192 3	<0.000 1	0.687 2	0.694 8	0.037 2	0.007 7	0.274 6	0.712 4	0.903 6
7	0.000 3	<0.000 1	0.013 4	0.299 3	0.091 9	0.541 5	0.358 5	0.304 0	0.570 3
8	<0.000 1	<0.000 1	0.151 1	0.417 7	0.331 2	0.358 0	0.969 1	0.174 0	0.694 7

从表3中可以看出三角形区域顶点为对应的 $m=2, j=3$, 对差分后的序列选择 ARMA(2,3) 模型, 参数估计结果见表4.

2.2.2 ARIMA 模型在芥兰上小菜蛾幼虫种群动态中的应用 对差分后的平稳序列采用最小信息准则定价, 误差序列模型为 AR(14), SAS 给出的最小信息值见表5.

从表5可以看出, 在误差序列模型为 AR(14), 当 $m=5$ 或 $m=6$ 阶、 $j=0$ 时, BIC 值均比取其他阶数小, 分别为 -0.965 和 -0.970 . 根据初步试探, 且 $m=5$ 时模型的自回归部分参数均通过检验, 因此对差分后的序列选择模型 ARMA(0,5), 参数估计见表6.

表5 芥兰上小菜蛾幼虫种群动态最小信息准则的 BIC 值

Tab. 5 Minimum information criterion for the larva's population dynamic of *Plutella xylostella* on *Brassica albograbra*

m	j								
	0	1	2	3	4	5	6	7	8
0	-0.828	-0.822	-0.842	-0.848	-0.826	-0.945	-0.945	-0.915	-0.890
1	-0.800	-0.827	-0.820	-0.814	-0.791	-0.933	-0.929	-0.915	-0.884
2	-0.794	-0.805	-0.805	-0.789	-0.761	-0.899	-0.896	-0.883	-0.855
3	-0.800	-0.795	-0.791	-0.756	-0.729	-0.909	-0.896	-0.864	-0.841
4	-0.780	-0.765	-0.759	-0.725	-0.701	-0.874	-0.863	-0.831	-0.814
5	-0.965	-0.937	-0.932	-0.899	-0.874	-0.841	-0.830	-0.797	-0.781
6	-0.970	-0.953	-0.928	-0.893	-0.864	-0.831	-0.796	-0.762	-0.756
7	-0.936	-0.918	-0.901	-0.868	-0.837	-0.804	-0.769	-0.752	-0.731
8	-0.911	-0.893	-0.875	-0.841	-0.808	-0.775	-0.741	-0.725	-0.712

表6 芥兰上小菜蛾幼虫种群动态的 ARIMA 模型参数估计

Tab. 6 Parameter estimation of ARIMA model for the larva's population dynamic of *Plutella xylostella* on *Brassica albograbra*

参数	参数估计	标准误	t	p
parameter	estimate	standard error		
μ	0.006	0.030	0.200	0.842
ϕ_1	-0.185	0.084	-2.200	0.028
ϕ_2	-0.197	0.083	-2.360	0.018
ϕ_3	-0.199	0.084	-2.380	0.017
ϕ_4	-0.220	0.084	-2.610	0.009
ϕ_5	-0.175	0.085	-2.060	0.040

2.3 模型诊断

从表4、表6可以看出, 确定的菜心和芥兰上小菜蛾幼虫种群动态序列的 ARIMA 模型中, 均值以外各参数对应 t 值的 p 值都小于 0.05, 均通过检验, 对2种蔬菜上小菜蛾幼虫种群动态序列应用 ARIMA 模型后残差的白噪声检验, 表示是滞后 6、12、24 步时残差白噪声检验(表7).

从表7可以看出, 对2种蔬菜上小菜蛾幼虫种群动态序列应用 ARIMA 模型后, 残差序列 χ^2 检验的

表4 菜心上小菜蛾幼虫种群动态 ARIMA 模型参数估计

Tab. 4 Parameter estimation of ARIMA model for the larva's population dynamic of *Plutella xylostella* on *Brassica parachinensis*

参数	参数估计	标准误	t	p
parameter	estimate	standard error		
μ	0.004	0.033	0.120	0.904 0
θ_1	-1.513	0.229	-6.600	<0.000 1
θ_2	-0.375	0.170	-2.210	0.027 0
θ_3	0.338	0.162	2.080	0.037 0
ϕ_1	-1.663	0.070	-23.800	<0.000 1
ϕ_2	-0.808	0.068	-11.970	<0.000 1

P 值均大于 0.05, 接受白噪声检验的原假设, 即拟合后的残差序列为白噪声.

表7 残差白噪声检验

Tab. 7 Tests for white noise residuals

蔬菜	阶数	χ^2	自由度	P
vegetable sample	lags		ν	
菜心 <i>Brassica parachinensis</i>	6	4.79	1	0.068 6
	12	11.71	7	0.110 6
	24	18.08	19	0.517 2
芥兰 <i>Brassica albograbra</i>	6	0.16	1	0.685 0
	12	6.46	7	0.487 2
	24	10.58	19	0.937 0

2.4 结果分析

根据表4和表6得到菜心和芥兰2种蔬菜上小菜蛾幼虫种群动态序列的 ARIMA(2,1,3) 模型和 ARIMA(0,1,5) 模型, 具体结果见表8.

从表8可以看出, 利用扩展样本自相关函数法和最小信息准则法对 ARIMA 模型定阶, 按照时间序列模型建模的3个步骤建立的小菜蛾幼虫种群动态的 ARIMA 模型, 在预测菜心和芥兰上小菜蛾幼虫种

表 8 小菜蛾幼虫种群动态序列 ARIMA 模型的拟合

Tab. 8 The simulation of applying ARIMA to the larva's population dynamic of *Plutella xylostella*

蔬菜 vegetable sample	ARIMA 模型 ARIMA model ($L_t = \ln N_t$)	定价方法 methods for identifying the orders	均方根误差 RMSE
菜心 <i>Brassica parachinensis</i>	$(1 - B)L_t = 0.004 + \frac{1 + 1.513B + 0.375B^2 - 0.338B^3}{1 + 1.663B + 0.809B^2}a_t$	ESCAF	0.213
芥兰 <i>Brassica albograbra</i>	$(1 - B)L_t = 0.006 + \frac{1}{1 + 0.185B + 0.197B^2 + 0.199B^3 + 0.220B^4 + 0.175B^5}a_t$	MINIC	0.188

群数量时的均方根误差分别为 0.213 和 0.188,有着较高的预测准确性.

3 讨论

模型定阶是建立 ARIMA 模型的重要环节,在模型定阶时,选择不同阶数的自回归部分,对应的滑动平均部分的阶数也不相同,利用样本的自相关函数(SACF)和偏自相关函数(SPACF)的特性来确定模型结构,该方法对于确定纯 AR 模型和纯 MA 模型是行之有效的,但对于辨识混合 ARMA 模型则相当困难,扩展样本自相关函数法和最小信息准则法同时考察自回归部分和滑动平均部分,获得 ARMA 模型的结构,与其他方法相比有明显的优势,但有时扩展样本自相关函数法形成的零三角阵并不明显,因此很难确定 ARMA 模型的结构.同时 ARIMA 模型也是一类线性模型,由于害虫种群动态内在的非线性^[8],使其预测准确性受到很大的限制,对差分后仍不平稳的序列很难适用^[9],而且 ARIMA 模型进行预测时没有考虑到环境因素的影响,这些都是今后要进一步研究的方向.

参考文献:

[1] TSAY R S, TIAO G C. Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models [J]. JASA, 1984, 79 (385): 84-96.

[2] CHOI, BYOUNGSEON. ARMA model identification[M]. New York: Springer-Verlag, 1992: 129-132.

[3] HANNAN E J, RISSANEN J. Recursive estimation of mixed autoregressive moving average order[J]. Biometrika, 1982, 69 (1): 81-94.

[4] BOX G E P, JENKINS G M, REINSEL G C. Time series analysis: forecasting and control[M]. 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1994: 197-199.

[5] MAY R M. Biological populations with non-overlapping generations: stable points, stable circles and chaos[J]. Sci, 1974, 186(4164): 645-647.

[6] NISBET R M, GURNEY W S C. The systematic formulation of population models for insects with dynamically varying instar- duration[J]. Theor Pop Biol, 1983, 23: 114-135.

[7] DICKEY D A, FULLER W A. Distribution of the estimators for autoregressive time series with a unit root[J]. Journal of the American Statistical Association, 1979, 74 (366): 427-431.

[8] TURCHIN P. Rarity of density dependence or population regulation with lags [J]. Nature, 1990, 344: 660-663.

[9] SWAMY P A. Correcting for omitted-variable and measurement-error bias in autoregressive model estimation with panel data [J]. Computational Economics, 2003, 22: 225-253.

【责任编辑 李晓卉】