DOI:10.6660/TESFE.1984016 台灣昆蟲Formosan Entomol. 4: 77-81 (1984) 研究報告 Research report

Formosan Entomologist
Journal Homepage: entsocjournal.yabee.com.tw

A FORECASTING MODEL FOR BROWN PLANTHOPPER POPULATION DENSITY 【Research report】

褐飛蝨族群密度之預測模式1【研究報告】

Tsan-Long Lin and Ta-Hsiu Liu
林燦隆、劉達修

*通訊作者E-mail :

Received: Accepted: Available online: 1984/09/01

Abstract

Box-Jenkins methodology is involved to fit an autoregressive integrated moving average (ARIMA) model for population densities of the brwon planthopper, Nilaparvata lugens. The re- sultant model consists of two density regulating components, the seasonal regulating component and the random regulating component; the former makes the population growth to follow a pattern, but the latter to shift from the pattern and often masks the regularity density change contributed by the former. To improve forecasting, analysis of residuals is necessary. Possibility of improvement is seen in the example of evaluating the predation effect of Lycosa on the insect, based on the residuals from the ARIMA models fitted to these two insects.

摘要

Key words:

關鍵詞:

Full Text: PDF(0.51 MB)

A Forecasting Model for Brown Planthopper Population Density

Tsan-Long Lin
National Taiwan University, Taipei
and
Ta-Hsiu Liu
Taichung Agricultural Improvement Station

ABSTRACT

Box-Jenkins methodology is involved to fit an autoregressive integrated moving average (ARIMA) model for population densities of the brown planthopper, *Nilaparvata lugens.* The resultant model consists of two density regulating components, the seasonal regulating component and the random regulating component; the former makes the population growth to follow a pattern, but the latter to shift from the pattern and often masks the regularity density change contributed by the former. To improve forecasting, analysis of residuals is necessary. Possibility of improvement is seen in the example of evaluating the predation effect of Lycosa on the insect, based on the residuals from the ARIMA models fitted to these two insects.

Introduction

The objective of the study is to develop a model that accounts for the relationships between observations and that will allow prediction of the expected size of the population for the brown planthopper. Since the data are structurally correlated and chronologically ordered, the Box-Jenkins methodology is used in the statistical analysis. The methodology was first used by Hacker *et al.* (1973) in entomology to devleop a forecasting model for mosquito population densities. Saboia (1977) used the method to obtain birth forecasting models and investigated their relationships with classical models for population growth. The methodology was also used by Lin (1979) to establish regression autoregressive integrated moving average models for the light-trap data of rice insects.

Materials

Survey was made at a paddy of 0.1 ha at Taichung Agricultural Improvement Station for eight years from 1975 to 1982. Samples of size varying from 100 to 200 hills of paddy were investigated every one week after transplanting, and mean densities of the brown planthopper (BPH, *Nilaparvata lugens), Lycosa pseudoannulata* and *Oedothorax insecticeps* per hill were obtained.

To avoid getting into difficulties, strings of zero counts obtained at the begining of every crop season are excluded from the study, and only 23 sets of data, of which 12 are obtained at the first and 11 at the second crop season for every year are used in the statistical analysis. Observations are chronologically ordered to form a time series of length 184.

Statistical Methods

Provided that, with a suitable data transformation, the transformed density of an insect population at time t is representable by a linear function of the previous history of the population and the random components closely related to the dynamics of the insect population, we have

$$z_t = b_1 z_{t-1} + b_2 z_{t-2} + \ldots + b_n z_{t-n} + a_t + c_1 a_{t-1} + \ldots + c_m a_{t-m}$$

where z_t is the transformed density at time t,

b_i and c_i are unknown constants to be estimated from the data and $a_t, a_{t-1}, \ldots, a_{t-m}$ are random components assumed to follow the normal distribution with zero mean and a specific variance independently. By using backward shift operator B, where $Bz_t = z_{t-1}$ and $B^h z_t = z_{t-h}$, the model can be rewritten as

$$(1 - b_1 B - \ldots - b_n B^n) z_t = (1 - c_1 B - \ldots - c_m B^m) a_t,$$

which is a member of a fairly general class of models explored in some detail by Box and Jenkins, known as ARIMA processes. The general expression of the process is given by

$$[1 - A_p(B)]\,[1 - C_p(B^S)]\,(1 - B)^d (1 - B^S)^D z_t = [1 - G_q(B)]\,[1 - H_Q(B^S)]\, a_t \quad \ldots\ldots (1)$$

where $[1 - A_p(B)]$, $[1 - C_P(B^S)]$, $[1 - G_q(B)]$, $[1 - H_Q(B^S)]$ are palynomials of backward shift operator B or B^S with order p, P, q and Q respectively, and S is the seasonal lag, which is 23 in this study. Integers d and D are respectively degrees of nonseasonal and seasonal differencing which may be required to produce stationary time series from nonstationary one.

Steps of data analysis are summarized as follows:

1. Selecting the order of differencing and transformation

To fulfill additive and stationary conditions of the model (1), a set of $[d, D, \lambda]$ in the expression

$$w_t = (1 - B)^d (1 - B^{23})^D (y_t + 0.1)^\lambda \qquad \lambda \neq 0$$

or

$$w_t = (1 - B)^d (1 - B^{23})^D \ln(y_t + 0.1) \qquad \lambda = 0$$

which minimizes a standarized sum of squares

$$\frac{\Sigma\,(w_t - W)^2}{\left(\Pi_t\, y_t^{\lambda-1}\right)^{2/n}} \quad \ldots\ldots (2)$$

is selected from the total combinations of d = 0, 1, 2, D = 0, 1, 2 and λ = 0, 0.25, 0.33, 0.50, 0.75, 1.0. In (2), y_t represents the desnity observed at time t, and the transformed variable $(y_t + 0.1)^\lambda$ or $\ln(y_t + 0.1)$ is expressed by z_t in the general model (1). Rationale of the transformation was given by Box and Cox (1964).

2. Model building

The Box-Jenkins methodology consists of several steps. The first step involves using historical boservations of the time series to identify a tentative model to be used in forecasting future values of the time series. In this identification process, autocorrelation and partial autocorrelation functions are estimated and used to identify the particular stationary time series model that adequately describes the density of the observed insect population. The second step involves estimating unknown parameters of the tentatively identified model. Non-linear least squares technics are involved. The third step of the Box-Jenkins methodology, called diagnostic checking, involves testing the adequacy of the tentatively identified and efficiently estimated model and, if necessary, suggesting ways to improve the model. Since the modeling process is supposed to account for the relationships between the observation, and if it does, the residuals should be unrelated. Hence, numerically large autocorrelation coefficients in the sample autocorrelation function of the residuals indicate that the fitted model is deemed inadequate. In this case, improvement in the model should be made.

Details of the Box and Jenkins methodology are given in the books of Box and Jenkins (1970) and Bowerman and O'Conell (1979).

3. Farecasting

The minimum mean square error forecast $z_T(\ell)$ for $z_{T+\ell}$, the value to be observed at time $T + \ell$, is given by

$$z_{T(\ell)} = u_1 z_{T+\ell-1} + \ldots + u_{p+PS+d+DS} z_{T+\ell-p-PS-d-DS} + a_{T+\ell} - v_1 a_{T+\ell-1} - \ldots - v_{q+QS} a_{T+\ell-q-QS}$$

where the coefficients in the right are obtained by expanding the general model defined in (1) and $z_{T+\ell-1}, \ldots, z_{T+\ell-p-PS-d-DS}, a_{T+\ell}, a_{T+\ell-1}, \ldots, a_{T+\ell-q-QS}$ are conditional expectations. The forecasts are evaluated by inserting actual z_t's when these are known, forecasted z's for future values, actual a's when these are known, and zeros for future a's. Hence, the one-step ahead forcast at origin T is given by

$$z_{T(1)} = u_1 z_t + \ldots + u_{p+PS+d+DS} z_{T+1-p-PS-d-DS} - v_1 \hat{a}_T - \ldots - v_{q+QS} \hat{a}_{T+1-q-QS}$$

A necessary consequence is that the lead-1 forecast errors are the generating a's in the model, that is

$$\hat{a}_{T+1} = z_{T+1} - z_T(1).$$

Results

The standarized mean squares defined by (2) are given in Table 1. The minimum appears at d=1, D=1, and λ=0 and the second smallest which is only a little larger than the minimum appears at d=1, D=0 and λ=0. Since the effective number of observation for estimating parameters is 160 for the former and is 183 for the latter, the first order differencing of the log-transformed series with added 0.1 is tried to produce stationary series.

Table 1. Mean Square Standarized with the Jacobian of Inverse Transformation.

Transformation (λ)	Non-seasonal and seasonal differencing: (d,D) (0,0)	(1,0)	(2,0)	(0,1)	(1,1)	(2 1)
0.00	241.87	108.75	255.01	133.87	105.62	270.91
0.25	351.89	162.66	381.89	213.83	171.92	468.86
0.33	466.25	230.02	550.45	324.56	259.84	720.20
0.50	987.86	574.54	1438.41	903.76	721.92	2030.87
0.75	4339.98	3198.23	8433.70	5338.18	4323.33	12099.90
1.00	25298.20	22241.70	60293.90	36713.80	30565.80	84069.90

The sample autocorrelation and partial autocorrelation functions of the first order differences of the transformed variables are given in Fig. 1. Examination of Fig. 1 indecates that spikes appear at lags 1, 5, 10, 15, 22, 23 and 46. Such an autocorrelation function may arise from the model

$$(1 - C_1 B^{23})(1 - B)z_t = (1 - g_1 B - g_2 B^5)\,a_t,$$

where $z_t = \ell n(y_t + 0.1)$, and y_t the observed density at time t.

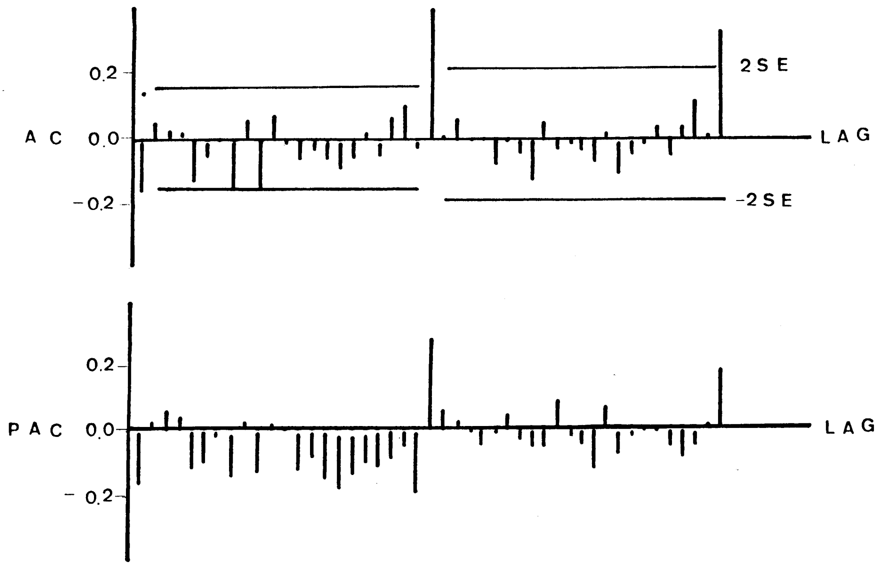

Fig. 1. Autocorrelation coefficient (AC), two standard error of AC (2SE) and partial autocorrelation coefficient (PAC) for the first order differencing series (∇z_t)

Table 2. Summary of Model for BPH

Differencing – 1 Regular differences

Parameter number	Parameter type	Parameter order	Estimated value	95 per cent	
				Lower limit	Upper limit
1	Seasonal autoregressive	23	0.548189	0.421195	0.675183
2	Regular moving average	1	0.386142	0.250203	0.522081
3	Regular moving average	5	0.361649	0.219164	0.504134

Residual sum of squares 284.375
Residual mean square 1.81131

Results of estimation are given in Table 2. The estimated standard deviation is 1.346 with 157 degrees of freedom.

Fig. 2 shows the sample autocorrelation function for the residuals of the estimated model. No significant autocorrelation is revealed. The hypothesis that the residuals are white noise seems acceptable. Also, the confidence intervals in Table 2 indicate that the parameters estimated are significantly different from zero and over-fitting is not detected. The correlation coefficient between observed and calculated is 0.856. Fig. 3 gives the general trends of the predicted and observed.

Discussion

The estimated model can be expanded to yield a difference equation

$$z_t = z_{t-1} + 0.54819\,(z_{t-23} - z_{t-24}) + a_t - 0.386142a_{t-1} - 0.361649a_{t-5}$$

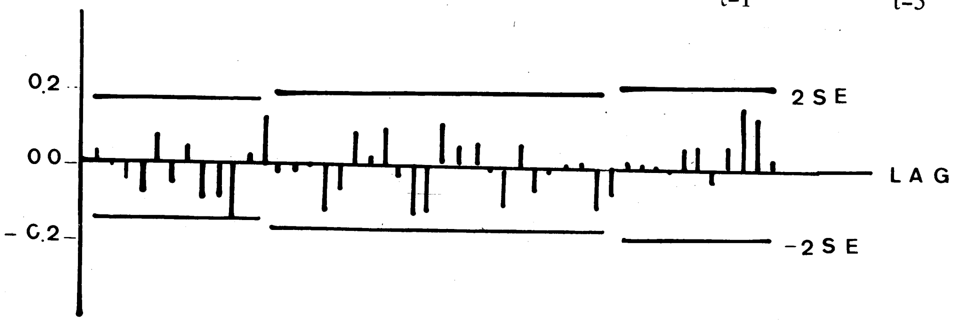

Fig. 2. Autocorrelation coefficient and two standard error of the autocorrelation coefficient (2SE) for the residual series of the brown planthopper.

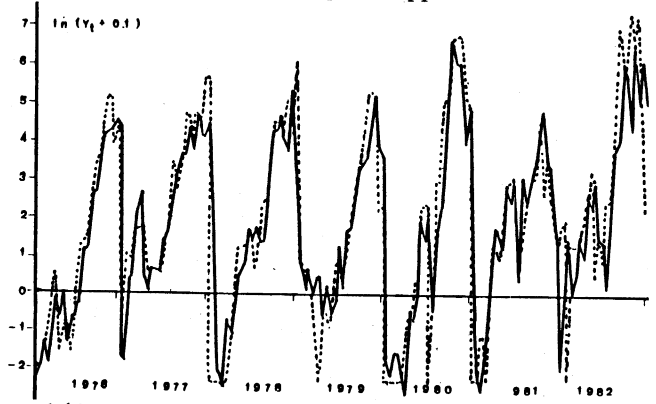

Fig. 3. Observed (dashed line) and fitted (solid line) series of the brown planthopper.

It consists of two population density regulating components; the seasonal regulating component and the random regulating component. The seasonal component represented by the term C. $(z_{t-s} - z_{t-s-1})$ with positive C adds a negative or a positive contribution to z_t, according to the population trend in the same season last year. The random regulating component represented by $-C_1a_{t-1} - C_2a_{t-5}$ with positive C_1 and C_2 provids a feedback contribution to z_t. Provided that the model is adequate, a_{t-b} represents the deviation of the actual population growth from the expected, and hence, the population growth b weeks later will be slowed down if a_{t-b} is positive and will be prompted if it is negative. This kind of feedback regulation may partly be attributable to the effect of predation by Lycosa. The results of the regression analysis for the residual on that from the ARIMA fitted to Lycosa reveal that the regression coefficient at lag 0 is significant and positive and that at lag 5 is also significant but negative. A positive regression coefficient at lag 0 indicates that the predator tends to aggregate about the paddy hill where the prey is relatively abundant. The consequence may increase the rate of predation and thus decrease the population size of one week later and suppress the population growth of the next generation. The details of the residual analysis will be discussed eleswhere.

Once an appropriate model has been obtained, it is used to forecast future values. To get an accurate forecast is very difficult as seen from Fig. 3. It has been revealed that the expanded difference equation consists of two density regulating components. Although the coefficients of these two density regulating componets are nearly equal numerically, it is obvious that the unpredictable change in population density contributed by the random density regulating component often masks the regularity change offered by the seasonal density regulating component, if the size of the standard deviation of the residuals is taken into consideration. Besides, in evaluating the minimum mean square error forecast, all the unestimable residuals are set to zero, that is, all the environmental effects important to the population growth during critical period are neglected. To see the possibility of forecasting future population densities solely based on the relationships between the densities observed, the parameters in the model are newly estimated with the data obtained from 1975 to 1981, and is used to produce 23 one-week-ahead forecasts of 1982. The sample standard deviation of the residuals from the model newly fitted is 1.23, and the correlation coefficient between these 23 one-week-ahead forecasts and the actually observed is about 0.7. To improve forecasting, analysis of residuals seems necessary and importand. Possibility is shown in the example of evaluating predation effect of Lycosa, given above.

REFERENCES

1. Bowerman, B. L. and R. T. O'Conell. 1979. *Time Series and Forecasting, Duxbury Press.*
2. Box, G.E.P. and D.R. Cox. 1964. An analysis of transformations. *J. Roy. Statist. Soc. B,* **26,** 211-243.
3. Box, G.E.P. and G.M. Jenkins. 1970. *Time Series Analysis, Forecasting and Control, Holden-Day*.
4. Hacker, C. S., D.W. Scott, and J.R. Thompson. 1973. A forecasting model for mosquito population densities. *J. Med. Ent.,* **10,** 544-551.
5. Lin, T.L. 1979. Fitting regression - ARIMA models to the light-trap data of rice insect. (in Chinese with an English abstract). *Natl. Sci. Counc. Monthly, ROC,* 7(2), 180-200.
6. Saboia, J.L.M. 1977. Autoregressive integrated moving average (ARIMA) models for birth forecasting. *J. Am Stat. Asso.,* **72,** 264-270.