

A Nonlinear Statistical Model for Aphid Population Growth

Prajneshu
IASRI, New Delhi
(Received: May, 1997)

SUMMARY

Aphids are recognised as serious pests of cereals, oilseeds, pulses and vegetable crops in our country. It is highly desirable to investigate optimal control policies for controlling this pest. To this end, as a first step, a nonlinear statistical model is developed for describing the dynamics of aphid population growth. The deterministic model, which is expressed in terms of an integro-differential equation, is solved analytically. The corresponding nonlinear statistical model is applied to ten data sets using the Levenberg-Marquardt iterative procedure. Examination of residuals is carried out to study the validity of the underlying assumptions and subsequently the goodness of fit statistics are computed. It is concluded that the proposed model is quite successful in describing the dynamics of the aphid population growth.

Keywords: Nonlinear statistical model, Aphid population growth, Levenberg-Marquardt procedure, Residual analysis.

1. Introduction

Aphid is a small insect which infests almost every plant. It is among the most conspicuous and important pest. In India, aphids are recognized as serious pests of cereals, oilseeds, pulses and vegetable crops. They cause damage directly to host plants by sucking the plant sap and arrest their growth and development. A large number of studies have been conducted dealing with various biological aspects of aphids (see e.g., Minks and Harrewijn [4], Niemczyk and Dixon [5]). Barlow and Dixon [1] developed a detailed deterministic simulation model in discrete time for describing the dynamics of lime aphid population. It predicts changes in population count and growth in size of aphids throughout the season and from one year to the next, while separate submodels also mimic changes in numbers of predators and their alternate prey. Although such types of descriptive models do provide useful information, their main disadvantages are that they tend to be cumbersome, difficult to understand, and specific to individual situations.

The other category of dynamical models is the 'Analytical explanatory models'. Such models incorporate some of the causal relations in the system, but remain simple enough to be analytically tractable. Their generality is their major strength. The level of understanding they provide is high, virtually qualitative, but they are nevertheless truly functional and can clearly be used for predictive purposes. The simplest models with any claim of realism are those describing the changes in single-species populations. Of these, perhaps the best-known is the logistic equation (see e.g., Matis *et al* [3]):

$$dN/dt = rN(1 - N/s) \quad (1)$$

where $N(t)$ denotes the population size at time t . The parameters r and s now have a biological meaning: r is the intrinsic growth rate and s is the carrying capacity. It is possible to predict not only the relation between "initial" and "final" density, but also the temporal changes in the population. In addition, examination of the related model

$$N_{t+1} = N_t \exp(r - rN_t/s) \quad (2)$$

suggests properties of general ecological systems. This simple model can be used to study the behaviour of a population perturbed from its equilibrium; convergence to s ; convergent oscillations; limit cycles; divergent oscillations; divergence or chaos. In this paper, we propose to develop a dynamical model for aphid population growth, along similar lines as Kindlmann [2], and apply the corresponding nonlinear statistical model to aphid data.

2. Deterministic Model

Let the aphid population at time t be denoted by $N(t)$. The rate of change in the population size, viz. dN/dt , is then equal to the difference of the birth rate and the mortality rate. For the first approximation, the simplest form for the birth rate can be chosen as $rN(t)$, where r represents the intrinsic birth rate. In the initial phase, the mortality rate is generally very low, but it increases dramatically subsequently. The reason for this is that honeydew forms a weak cover on the leaf surface preventing the aphids from further sucking and movement and so causes starvation. Evidently, the area covered by the excretion at time t is proportional to the integral $\int_0^t N(s)ds$, which is called cumulative density (Barlow and Dixon [1]). Following these authors, a linear dependence between cumulative density and mortality rate is assumed. Thus the model can be expressed as

$$\frac{dN}{dt} = rN - \frac{N}{C} \int_0^t N(s) ds \quad (3)$$

This equation can also be written as

$$CN^{-1} / (rN - dN/dt) = \int_0^t N(s) ds \quad (4)$$

Let $N(t) \geq 0$ for all $t \geq 0$ and let $N(t)$ be differentiable at least twice. Differentiating eq. (4), we get

$$N \frac{d^2 N}{dt^2} - \left(\frac{dN}{dt} \right)^2 - C^{-1} N^3 = 0 \quad (5)$$

To solve this equation, let

$$\frac{1}{N} \frac{dN}{dt} = u(N) \quad (6)$$

Differentiating w.r.t. t , we get

$$\frac{1}{N} \frac{d^2 N}{dt^2} - \frac{1}{N^2} \left(\frac{dN}{dt} \right)^2 = \frac{du}{dt} \quad (7)$$

Using this, eq. (5) reduces to

$$N^2 du/dt + C^{-1} N^3 = 0 \quad (8)$$

Further, eq. (6) gives

$$\frac{du}{dt} = \frac{du}{dN} \cdot \frac{dN}{dt} = Nu \frac{du}{dN} \quad (9)$$

Substituting in eq. (8), we get

$$u du/dN + C^{-1} = 0 \quad (10)$$

Separation of variables gives on integration

$$u(N) = (K - 2N/C)^{1/2} \quad (11)$$

where K is the constant of integration. At $t = 0$, eq. (3) gives

$$u = N^{-1} dN/dt = r$$

Therefore from eq. (11),

$$K = r^2 + 2N_0/C$$

Now eq. (11) can be written as

$$\frac{dN}{N(K - 2N/C)^{1/2}} = dt \quad (12)$$

To integrate this equation, make the substitution

$$(K - 2N/C)^{1/2} = Z \quad (13)$$

Then eq. (12) becomes

$$2 dZ/(K - Z^2) = - dt \quad (14)$$

which, on integration, yields

$$\frac{1}{\sqrt{K}} \ln \left[\frac{\sqrt{K} + Z}{\sqrt{K} - Z} \right] = -t + L \quad (15)$$

where L is the constant of integration. At $t = 0$, eq. (13) gives

$$Z_0 = (K - 2N_0/C)^{1/2} = r$$

Therefore, from eq. (15),

$$L = \frac{1}{\sqrt{K}} \ln \left[\frac{\sqrt{K} + r}{\sqrt{K} - r} \right]$$

Substituting the expression from eq. (13) and carrying out a little algebra, the final expression for $N(t)$ can be written as

$$N(t) = ae^{-bt}(1 + de^{-bt})^{-2} \quad (16)$$

where a , b and d are given by

$$b = (r^2 + 2N_0 C^{-1})^{1/2}$$

$$d = (b + r)(b - r)^{-1}$$

$$a = 2(r^2 C + 2N_0)(b + r)(b - r)^{-1}$$

Finally, the values of the original parameters r , C , and initial population-size N_0 in terms of a , b , d are

$$C = a(2b^2 d)^{-1}$$

$$N_0 = a(1 + d)^{-2}$$

$$r = (b^2 - 2N_0 C^{-1})^{1/2}$$

3. Fitting the Model to Data

The dynamical model for aphid population growth given by eq. (16) has been developed deterministically. To fit this model to data, an error term ϵ is added to the R.H.S. of this equation. Thus

$$N(t) = ae^{-bt} (1 + de^{-bt})^{-2} + \epsilon \quad (17)$$

where ϵ is assumed to be independently and identically distributed following normal distribution. It may be noted that eq. (17) represents a nonlinear statistical model and accordingly nonlinear iterative procedures have to be employed for estimation of parameters. Three main methods of this kind are: (i) Linearization (or Taylor series) method, (ii) Steepest Descent method, and (iii) Levenberg-Marquardt method. The details of these methods along with their relative merits and demerits have been given, for example, by Seber and Wild [9]. In short, Levenberg-Marquardt procedure combines successfully the best features of the other two methods and, at the same time, also avoids their serious disadvantages. Accordingly, this is the most widely used procedure for computing nonlinear least squares estimates. Most of the standard software packages contain computer programs to fit nonlinear statistical models based on this procedure. For example, SPSS has NLR option, SAS has NLIN option, IMSL has RNSSQ option to accomplish the task. In this paper, as an illustration, NLR option on the SPSS-PC version 3.0 available at I.A.S.R.I., New Delhi is used for fitting the model to aphid population data for five years (1976-77 to 1980-81) for the two varieties of mustard crop, viz, Yellow seeded (YSS-8) and Brown seeded (Pusa Kalyani), as reported by Phadke and Prasad [6].

A number of initial values were tried for each data set to ensure that the convergence is 'global' and not merely 'local'. However, before taking a final decision about the appropriateness of a model, examination of residuals need to be carried out in order to check the validity of the assumptions made about the error term. The two main assumptions of randomness and normality of residuals are examined by using the well-known run test and Shapiro-Wilk test respectively. Furthermore, goodness of fit statistics also needs to be computed. The details are given, for example, by Prajneshu [7].

The results obtained are presented in Tables 1 and 2. A perusal of these tables indicate that for each of the ten data sets, the calculated value of run test statistic $|Z|$ is less than the tabulated value 1.960 at 5% level. Similarly, for each of the ten data sets, the calculated values of Shapiro-Wilk statistic lie in the acceptance region at 5% level. Thus, none of the assumptions of randomness and normality of residuals, is violated for any data set. Further, the calculated values of R^2 for all the data sets are coming out to be quite

high. Hence, it may be concluded that the model provides a good fit to all the data sets.

4. Concluding Remarks

As an application of the model, optimal time for insecticidal spray can be determined. Evidently, an insecticide should be sprayed when the rate of growth of the aphid population i.e. dN/dt , is maximum. Differentiation of (16) twice yields

$$\frac{d^2 N}{dt^2} = \frac{ab^2 e^{-bt} [-1 - 3d(1+ab)e^{-bt} + d^2(-2+3ab)e^{-2bt}]}{(1+de^{-bt})^4} \quad (18)$$

A straightforward algebra shows that the rate of growth of aphid population is maximum when t satisfies

$$t^* = -\frac{1}{b} \ln \left[\frac{3(1+ab) + (9a^2 b^2 + 30ab + 1)^{1/2}}{2d(-2+3ab)} \right] \quad (19)$$

It may be noticed that, for all the data sets, the estimates of the parameters a and d are very large as compared to that of b . Therefore, for the present data sets, an approximation to the optimum time of insecticidal spray can be taken as

$$t^{**} = b^{-1} \ln d \quad (20)$$

By chance, this value is also the same as the one for which the aphid population attains its maximum. A perusal of Tables 1 and 2 shows that, although the estimates of the parameters a , b , d vary markedly, yet the value of optimum time for spraying is almost constant over different years. Nevertheless, wide variation in the parameter estimates may perhaps be due to the fact that these parameters may no longer be constants but may be either time-dependent, or random variables, or even stochastic processes. This issue needs to be further investigated in depth before arriving at a final conclusion.

Table 1. Fitting of dynamical model to aphid data on yellow seeded (YSS-8) mustard

STATISTIC\YEAR	1976-77	1977-78	1978-79	1979-80	1980-81
<i>(i) Parameters:</i>					
$10^{-7} a$	4890739	2071	798	38714	97656
b	1.559	.783	.857	1.010	1.104
d	322580600	322581	163934	2702703	31250000
<i>(ii) Examination of residuals:</i>					
Run-test ($ Z $)	.520	1.495	1.825	1.495	1.157
Shapiro-Wilk test (W)	.930	.861	.900	.935	.812
<i>(iii) Goodness of fit:</i>					
R^2	.979	.895	.879	.971	.931
<i>(iv) Optimum time for spraying:</i>					
t^*	12.567	16.199	14.011	14.663	15.632
t^{**}	12.567	16.199	14.011	14.663	15.632

Table 2. Fitting of dynamical model to aphid data on brown seeded (Pusa Kalyani) mustard

STATISTIC\YEAR	1976-77	1977-78	1978-79	1979-80	1980-81
<i>(i) Parameters:</i>					
$10^{-7} a$	4844375	208	1250	45000	2829335
b	1.537	.726	.916	1.050	1.353
d	125000000	109890	416667	5000000	1063830000
<i>(ii) Examination of residuals:</i>					
Run-test ($ Z $)	1.536	1.825	1.703	1.157	.538
Shapiro-Wilk test (W)	.900	.756	1.450	.892	.898
<i>(iii) Goodness of fit:</i>					
R^2	.967	.783	.906	.982	.975
<i>(iv) Optimum time for spraying:</i>					
t^*	12.130	15.988	14.127	14.690	15.362
t^{**}	12.130	15.988	14.127	14.690	15.362

REFERENCES

- [1] Barlow, N.D. and Dixon, A.F.G., 1980. *Simulation of lime aphid population dynamics*. Centre for Agricultural Publishing and Documentation, Wageningen.
- [2] Kindlmann, P., 1985. A model of aphid population with age structure, *Lecture notes in biomathematics*, 57, 72-77.
- [3] Matis, J.H., Kiffe, T.R. and Parthasarathy, P.R., 1996. Using density-dependent birth-death migration models for analyzing muskrat spread in the Netherlands. *Jour. Ind. Soc. Agri. Statist.*, 49, 139-156.
- [4] Minks, A.K. and Harrewijn, P., 1987. *Aphids - their biology, natural enemies and control (Three volumes)*, Elsevier.
- [5] Niemczyk, E. and Dixon, A.F.G., 1987. *Ecology and effectiveness of aphidophaga*. SPB Acad. Publication.
- [6] Phadke, K.G. and Prasad, S.K., 1984. Investigation on forecasting the outbreaks of pests and diseases of selected agricultural crops - Project Report, IARI (Entomology Division).
- [7] Prajneshu, 1991. Cautionary note about nonlinear models in fisheries. *Ind. J. Fisheries*, 38, 231-233.
- [8] Seber, G.A.F. and Wild, C.J., 1989. *Nonlinear regression*. John Wiley and Sons.