Fitting Cumulative Size Mechanistic Models to Insect Population Data: A Nonlinear Mixed Effects Model Analysis

James H. Matis, Lan Zhou, Thomas R. Kiffe[1] and Timothy I. Matis[2]

Department of Statistics, Texas A&M University, College Station, TX, 77843-3143, USA

SUMMARY

Recently a new class of mechanistic models for insect population size has been developed. The death rate in the new models is a function of the cumulative size of past generations. Prajneshu (1998) developed the first such model and others have followed. The models have been shown to fit data from diverse aphid populations well. This paper shows that the nonlinear regression model solution for aphid population size may be used as a basis for a nonlinear mixed effects model analysis of designed experiments. The experiment analyzed in this paper consists of a 3×3 factorial, of Water by Nitrogen levels, applied in a randomized block design, repeated two years. The fixed Water treatments are statistically significant in one year and the fixed Nitrogen treatments in the other. The paper outlines possible generalizations.

Key words : Aphids, Logistic models, Nonlinear least squares.

1. INTRODUCTION

A new class of mechanistic models based on a cumulative size concept has recently been developed to describe the growth of insect populations. In Section 2, this paper first establishes an ecological foundation for the new models by reviewing the well-known logistic models. The new models based on cumulative size are then presented. The simplest of these models has a concise analytical solution which serves as a regression model for data. The regression model and some of its properties are given. In Section 3, we note that the new regression model fits diverse aphid population data well. Data for a 3×3 factorial experiment investigating the effects of water and nitrogen on cotton aphids are presented. In Section 4, the nonlinear regression model is used as the basis for a nonlinear mixed effects model analysis using the nlme software for the R package. The analysis indicates that there are significant Water and Nitrogen main effects in the data. In Section 5, the results

from the nlme analysis are compared to results obtained by fitting the individual curves using nonlinear least squares. The results are shown to be similar qualitatively.

Concluding remarks, including possibilities for generalizing both the underlying model and the accompanying statistical analysis, are given in Section 6. In brief, the paper suggests that the new models, combined with the nlme analysis, have broad appeal for describing and analyzing data on local insect populations that collapse in size. Such data are widespread in agriculture.

2. MODEL DEVELOPMENT

2.1 Logistic Models Based on Current Size

In his classic book on ecology, Andrewartha (1951) describes several types of curves for describing the growth of 'local populations'. One class of curves is of the logistic type which increases monotonically in size with an asymptotic equilibrium value called a 'carrying capacity'. Andrewartha suggests that such curves may be "generally true of any local population whose numbers are determined by the stock of some non-expendable resource; non-expendable in the sense that the amount

[1] *Department of Mathematics, Texas A&M University, College Station, TX, 77843-3368, USA*

[2] *Department of Industrial Engineering, Texas Tech University, Lubbock, TX, 79409-3061, USA*

available to the next generation is independent of the amount used by the present generation" (ibid, p. 157). Food and nesting sites are given as examples of possible non-expendable resources.

The well-known Verhulst-Pearl logistic equation, with a rich history dating back to 1838 (Renshaw 1991), provides a mathematical model for the size of such populations. Letting N and N' denote the population size variable and its derivative, this deterministic model may be expressed as

$$N' = (a - bN)N \qquad (1)$$

with a, b > 0. In a simple interpretation of this model, parameter 'a' could be regarded as the 'per capita' birth rate, also called the 'intrinsic rate of natural increase'. The per capita death rate is bN where parameter 'b' is the 'crowding coefficient' which determines the strength of the 'density-dependent' mortality. The non-zero root, K = a/b, is immediately the carrying capacity. The point of inflection occurs at K/2.

In our studies of African bees, an aggressive so-called 'r-strategist', we observed that the point of inflection often exceeded K/2 and the model

$$N' = (a - bN^S)N \qquad (2)$$

with $s \geq 1$ was proposed in Matis *et al.* (1996, 1998). The solution to (2) and (1) is

$$N(t) = \frac{K}{\left[1 + m \exp(-ast)\right]^{1/s}} \qquad (3)$$

where $K = (a/b)^{1/s}$ and $m = (K/N_0)^s - 1$. Model (3) with small integer s = 1, 2,... was fitted successfully to biological invasion data on the muskrat (Matis *et al.* 1996) and on the African bee (Matis *et al.* 2005). For both the bee and the muskrat examples, each current generation leaves, in Andrewartha's terms, sufficient food and habitat resources for the succeeding generations to eventually establish a carrying capacity of size $K = (a/b)^{1/s}$.

2.2 Mechanistic Models Based on Cumulative Size

Andrewartha (1951) also discusses local populations which collapse in size. He suggests that some populations become extinct because their "numbers depend on the amount of some diminishing or expendable resource; expendable in the sense that the more that is used by one generation the less there will be for the next" (ibid, p. 158). He also notes that other populations collapse

because although the "animals are rare relative to their stock of food (or other resources)" the preceding aphids increase the likelihood of "a predator, ... the recurrent use of an insecticide, or any other component of environment" which eventually lead to local extinction (ibid, p. 158). We suggest that local aphid populations collapse for both reasons; namely that the preceding aphid generations use up some diminishing resources, specifically food and habitat resources and that the preceding aphid generations also degrade the environment through greater attraction of predators, an increased likelihood of plant and aphid disease and the chemical reaction of the plant to a loss of sap. We call this general dependency of current population growth on past generations the 'cumulative size dependency.'

We suggest two measures of cumulative size. One is the 'cumulative density' defined as

$$F(t) = \int_0^t N(s)ds \qquad (4)$$

in units of insect-time. The other is the 'cumulative count', C(t), defined as the past cumulative insect count. Both measures may be used as indices of past cumulative environmental degradation.

Prajneshu (1998) develops the first cumulative size model that we are aware of, one based on the cumulative density. His model may be written as

$$N' = (\lambda - \delta F)N \qquad (5)$$

The per capita birth rate in (5) is denoted by λ. The per capita death rate is δF which as a contrast to model (1), is a function of cumulative density F. Therefore, the growth rate in (5) obviously is dependent on the cumulative past environmental degradation and hence it belongs to the second type of growth curve discussed by Andrewartha.

Matis *et al.* (2005) developed an analogous deterministic model based on the cumulative count, C(t), calculated as

$$C(t) = \lambda F(t) \qquad (6)$$

Substituting this into (5) yields

$$N' = (\lambda - \mu C)N \qquad (7)$$

where μ is a new death rate parameter defined as

$$\mu = \delta / \lambda \qquad (8)$$

The formulation in (5) is more natural to ecologists who are familiar with the 'area under the population curve', $F(t)$. The formulation in (7), as based on the count data, is more convenient for stochastic modeling and simulation. We call the former a 'cumulative density' based and the latter a 'cumulative count' based population growth model.

Prajneshu (1998) shows that the analytical solution of (5) has form

$$N(t) = ae^{-bt}(1 + de^{-bt})^{-2} \qquad (9)$$

One can show that the time and size of peak value, denoted by t_{max} and N_{max} of model (9) are

$$t_{max} = b^{-1} \log d \qquad (10)$$

and

$$N_{max} = a/(4d)$$

We use these in Matis *et al.* (2007) to reparameterize model (9) as

$$N(t) = 4N_{max}e^{-b(t-t_{max})}[1 + e^{-b(t-t_{max})}]^{-2} \qquad (11)$$

Regression model (11) is a nonlinear regression model which may be fitted to data using standard software. Parameters N_{max} and t_{max} have obvious physical interpretation, in fact are directly measurable and 'b' will be shown to approximate the birth rate λ. Letting

$$d = \exp(bt_{max})$$

one can show that the solutions for the mechanistic parameters in equation (5) in terms of the regression parameters N_{max}, t_{max} and 'b' in (11) are

$$\lambda = b(d-1)/(d+1)$$

$$\delta = b^2/2N_{max} \qquad (12)$$

$$N_0 = 4d N_{max}(1+d)^{-2}$$

where N_0 denotes the initial value $N(0)$.

It is easy to show that $N(t)$ in (9) and (11) collapses to zero as t becomes large. Thus replacing N in the parentheses of the logistic model in (1), for a population with non-expendable resources, with the cumulative density $F(t)$ in (5), for a population with an expendable resource, leads to population collapse. Regression model (11) is therefore a mathematical representation of what

Andrewartha (1951) describes qualitatively (and indeed illustrates in his Figs. 9.02 and 9.03, p. 164) for local populations which collapse. Though such data are abundant in agriculture e.g. for insects, we are not aware of any other simple analytical model for describing such populations.

There are several properties of model (11) that are of considerable interest. One is that $N(t)$ is symmetric (Matis *et al.* (2005)), which is of interest because aphid population growth curves are typically skewed to the left. This fact has led to two generalizations. The first investigates the following power-law generalization of (5)

$$N' = (\lambda - \delta F^s)N \qquad (13)$$

for integer $s > 1$. The solution has the desired left skewness, however, the model does not have an analytical solution and hence is more challenging for experimenters to use Matis *et al.* (2007). The second Matis *et al.* (2007) adds an initial immigration term, with formulation

$$N' = \begin{cases} \nu & t \leq \tau \\ \nu + (\lambda - \delta F)N & t > \tau \end{cases} \qquad (14)$$

which also has a left-skewed solution. Though both of these generalizations have strong biological justification, model (5) with its simple analytical solution in (11) is typically sufficient for data analysis, particularly for the very common case of data gathered with weekly sampling intervals, as we illustrate subsequently.

As another property of (11), note that the final cumulative density, substituting (9) into (4), is

$$F(\infty) = \frac{a}{b(d+1)} \qquad (15)$$

Accurate approximations for $F(\infty)$ and $C(\infty)$, which are two end-points of primary interest, may be found using (10) and (12) to be Matis *et al.* (2007)

$$\tilde{F}(\infty) = 4N_{max}/\lambda \qquad (16)$$

and

$$\tilde{C}(\infty) = 4N_{max}$$

These simple approximations may be particularly useful in practice as they are based only on biological parameters, λ and N_{max} which may be known or readily estimated by experimenters.

3. EXPERIMENTAL DATA

With mechanistic models (5) and (7) having such compelling biological rationale and with the model solution in (11) being so user-friendly and transparent, an obvious question is whether model (11) also fits aphid data adequately. We answer that question in the affirmative for data on the pecan aphid in Matis *et al.* (2005), the mustard aphid in Matis *et al.* (2007) and the cotton aphid in Matis *et al.* (2007). For brevity we consider only the cotton aphid data in this paper.

Experiments were conducted in years 2003 and 2004 at the Texas A&M Agricultural Extension Center in Lubbock, TX, to investigate the effects of various water and nitrogen treatments on the abundance of the cotton aphid. Three water treatments (Low, Medium and High), denoted i = 1,2,3 were crossed with three nitrogen treatments (Zero, Variable and Blanket), denoted j = 1,2,3 in a 3×3 factorial arrangement. Twenty-seven (27) experimental plots were divided into three blocks, denoted k = 1,2,3 and the nine treatment combinations were randomly assigned within each block. The number of aphids was recorded on five dates in 2003 and on seven dates in 2004 for each plot. The data are available in Matis (2006). Fig. 1 displays the observed aphid counts in 2003 and 2004 for each of the three plots within each of the nine (i,j) treatment combinations.

Experimental data such as these have traditionally been analyzed using a 'model-free' approach. Directly observable response variables, such as the time of first aphid appearance, the observed peak number of aphids and the crop age at the peak number have been analyzed to meet the objectives of such experiments (Chattopadhyay *et al.* (2005)). However, we have shown, using data for the pecan aphid, that a 'model-based' approach incorporating model (11) not only yields new explanatory response variables but may also be more powerful statistically (Matis *et al.* (2006)). In that analysis model (11) was fitted to each plot-level data set individually and an ANOVA was performed on the parameter estimates. This paper illustrates an efficient alternative approach, namely the use of the nonlinear mixed effects model to fit all data simultaneously in Pinheiro and Bates (2000). The data are fitting using the nlme routine in the R software package in R-Project (2006), with results described below.

4. NONLINEAR RANDOM EFFECTS MODEL ANALYSIS

4.1 Model Definition

For each combination of water level i, i = 1, 2, 3 nitrogen level j, j = 1, 2, 3 and block k, k = 1, 2, 3 the outcome variable for each plot was observed at times t_l, l = 1,…, L. Denote the number of aphids observed as Y_{ijkl}. We assume first that the plot-level observation Y_{ijkl} follows the nonlinear regression model

$$Y_{ijkl} = f(t_l, m_{ijk}, T_{ijk}, b_{ijk}) + \varepsilon_{ijkl} \qquad (17)$$

where ε_{ijkl} are independent random errors with identical normal distribution $N(0,s^2)$ and the nonlinear curve f(.) takes the form

$$f(t, m, T, b) = 4e^{m-b(t-T)}[1 + e^{-b(t-T)}]^{-2} \qquad (18)$$

with plot specific parameters m_{ijk}, T_{ijk} and b_{ijk}. Model (18) uses the transformation $m = \log N_{max}$ in (11) in order to stabilize the variance.

We then assume the parameters m_{ijk}, T_{ijk} and b_{ijk} follow some linear regression models that account for covariate effects of water, nitrogen and block. We assume the following analysis of variance (ANOVA) model for m_{ijk}

$$m_{ijk} = \mu^m + \alpha_i^m + \beta_j^m + \gamma_{ij}^m + \rho_k^m + \xi_{ijk}^m \qquad (19)$$

where $\mu^m, \alpha_i^m, \beta_j^m, \gamma_{ij}^m$ and ρ_k^m are fixed effects and ξ_{ijk}^T are random effects representing plot deviations from the population mean satisfying $E\left(\xi_{ijk}^m\right) = 0$. Here the fixed effects have the usual interpretation: μm is the grand mean, $\alpha_i^m, \beta_j^m, \rho_k^m$ are the main effects of water, nitrogen and block respectively and γ_{ij}^m are the water by nitrogen interactions. We assume similar ANOVA models for T_{ijk} and b_{ijk} specifically

$$T_{ijk} = \mu^T + \alpha_i^T + \beta_j^T + \gamma_{ij}^T + \rho_k^T + \xi_{ijk}^T \qquad (20)$$

and

$$b_{ijk} = \mu^b + \alpha_i^b + \beta_j^b + \gamma_{ij}^b + \rho_k^b + \xi_{ijk}^b \qquad (21)$$

As in the ANOVA model for m_{ijk} in (19), $\mu^T, \alpha_i^T, \beta_j^T, \gamma_{ij}^T$ and ρ_k^T are fixed effects in (20) and $\mu^b, \alpha_i^b, \beta_j^b, \gamma_{ij}^b$ and ρ_k^b are fixed effects in (21). The random effects satisfy $E\left(\xi_{ijk}^T\right) = 0$ and $E\left(\xi_{ijk}^b\right) = 0$.

4.2 Model Results

The data for the year 2003 and 2004 were analyzed separately due to the differences in the sampling designs. The data for 2003 were fitted to a reduced model, produced by dropping the interaction terms $\gamma_{ij}^{m}, \gamma_{ij}^{T}$ and γ_{ij}^{b} in the ANOVA model (19) - (21) for m, T, and b because of the limited number of sampling dates per plot in 2003. The parameter estimates, averaged over the three blocks for each treatment combination are given for the reduced model in 2003 and the full model in 2004 in Table 1. The corresponding fitted curves are illustrated for each treatment combination in Fig. 1. The curves appear to fit the data averaged over the blocks well.

Table 1. Estimates of parameters (N_{max}, t_{max}, b) averaged over blocks for each of nine treatment combinations of Water by Nitrogen in 2003 and 2004

Treatment		2003			2004		
Water Level	Nitrogen Level	N_{max}	t_{max}	b	N_{max}	t_{max}	b
1	1	28.41	2.74	3.00	23.42	5.41	2.82
1	2	30.94	2.77	2.97	18.01	5.52	1.89
1	3	32.05	2.74	3.15	24.55	5.41	2.45
2	1	23.58	2.62	2.98	20.19	5.58	2.20
2	2	25.68	2.66	2.95	20.27	5.50	1.96
2	3	26.60	2.63	3.13	22.41	5.37	2.46
3	1	22.87	2.62	3.18	21.70	5.48	2.51
3	2	24.90	2.66	3.15	20.94	5.45	2.16
3	3	25.79	2.63	3.33	18.96	5.47	2.09

The means for the five end-point variables of interest, namely N_{max}, t_{max}, λ, δ and $F(\infty)$ are given for each level of the Water and Nitrogen main effects in 2003 in Table 2A, along with p-values testing the significance of the main effects. Note that the Water main effect is significant (p < .05) for the N_{max}, t_{max}, δ and $F(\infty)$ variables. In particular, as the level of water increases

1. N_{max} and $F(\infty)$ decrease consistently and substantially (20-25%)
2. t_{max} decreases slightly (~5%) and
3. δ increases consistently and substantially (~40%)

Table 2. Factor level means from the nlme analysis for main effects, Water and Nitrogen, for five endpoints variables N_{max}, t_{max}, λ, δ and $F(\infty)$ with p-values for testing equality

A. Results for 2003

Level	N_{max}	t_{max}	λ	δ	$F(\infty)$
For Water					
1	30.43	2.75	3.04	0.152	40.11
2	25.26	2.64	3.02	0.181	33.52
3	24.49	2.64	3.22	0.212	30.47
p-value	0.03	0.004	0.58	0.006	0.001
For Nitrogen					
1	24.84	2.66	3.05	0.189	32.81
2	28.02	2.66	3.20	0.185	35.24
3	27.05	2.70	3.02	0.170	36.05
p-value	0.41	0.54	0.64	0.55	0.41

B. Results for 2004

Level	N_{max}	t_{max}	λ	δ	$F(\infty)$
For Water					
1	21.79	5.45	2.39	0.131	37.09
2	20.93	5.48	2.21	0.117	38.17
3	20.50	5.47	2.25	0.123	36.61
p-value	0.74	0.87	0.54	0.59	0.80
For Nitrogen					
1	21.73	5.49	2.51	0.145	34.84
2	21.85	5.42	2.33	0.124	37.61
3	19.70	5.49	2.00	0.102	39.43
p-value	0.33	0.43	0.011	0.009	0.14

These trends are noticeable in Fig. 1A, i.e. the peaks and areas under the curve for column 1 are larger than those for column 2, and these values for column 2 exceed those of column 3. The Nitrogen main effect is not significant for any of the five variables. Year 2003 was a very dry year, which explains the significance of the water treatments on many variables.

The comparable results for the 2004 data are given in Table 2B. The results are very different from 2003, most notably that the Water main effect is not significant for any of the five variables. This is due to the ample rainfall in 2004. However, the Nitrogen main effect is significant for two response variables; in particular as the level of nitrogen increased from Zero to Blanket

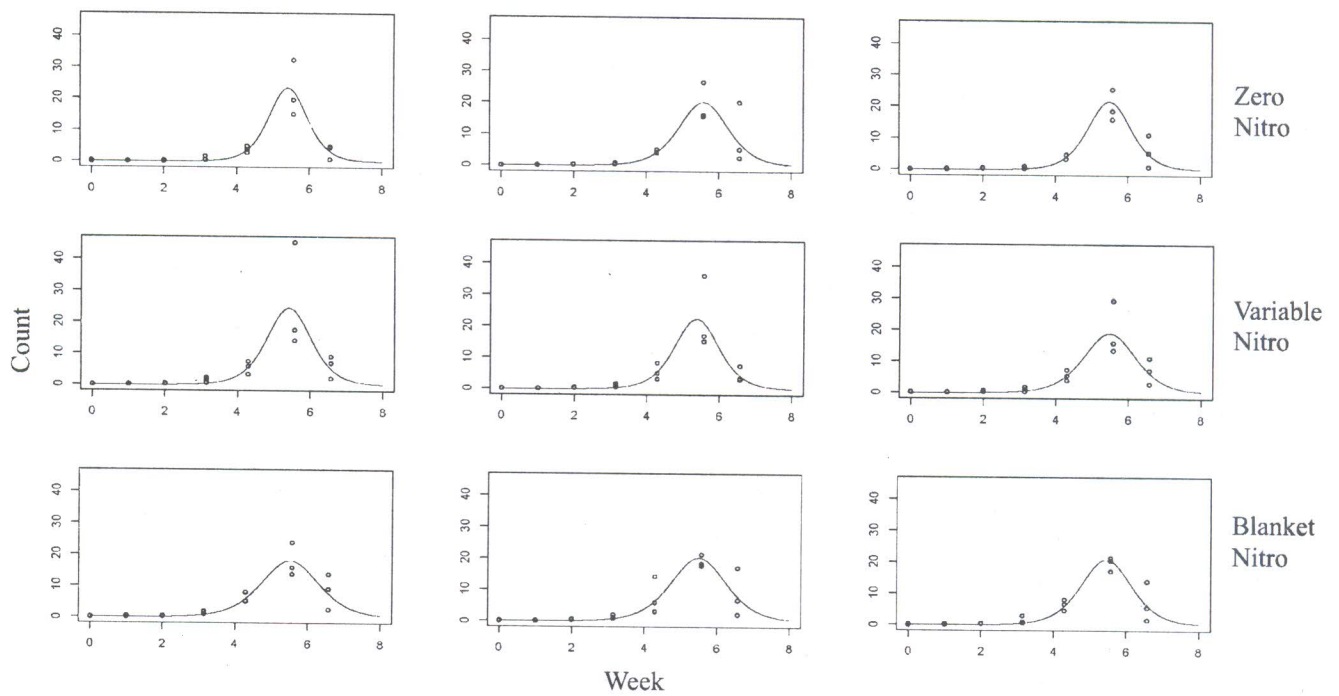A. Year 2003
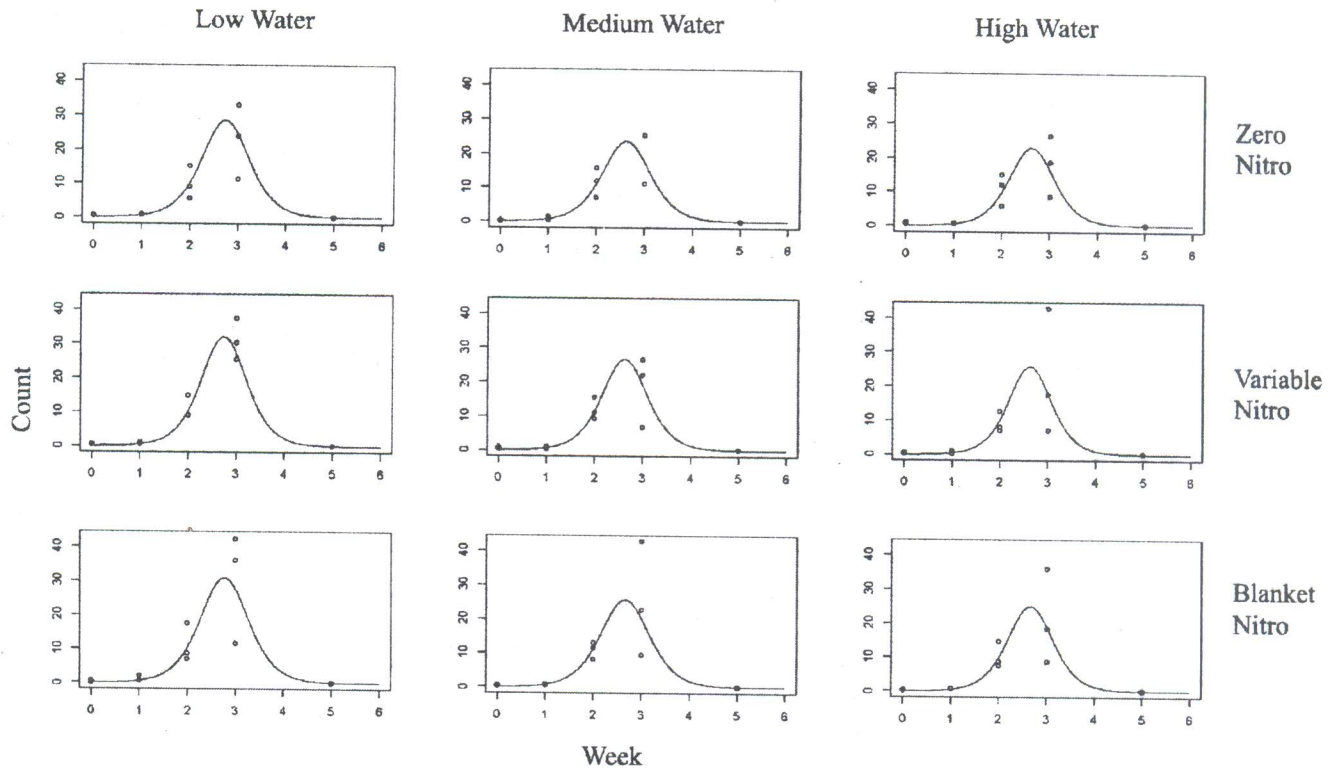
B. Year 2004

Fig. 1. Graphs of experimental data. Each box illustrates the aphid counts within the three blocks at each sampling time. Each column represents a Water level and each row a Nitrogen level. The fitted curves are calculated for each treatment combination from the parameter values given in Table 1. A. Data for 2003. B. Data for 2004

1. the birthrate λ decreases substantially (~20%)

2. the deathrate δ also decreases substantially (~30%)

These two effects largely offset one another and hence no significant changes occur in the levels of N_{max}, t_{max} or $F(\infty)$. Consequently, these effects in the kinetic rates are not visually noticeable neither in the raw data nor in the fitted curves in Fig. 1B. This demonstration that the regression model in (11) can detect treatment effects in the underlying kinetic parameters of the mechanistic model (5), effects which are not apparent in a model-free analysis, is a strong testament to the utility of the model-based analysis.

5. LEAST SQUARES ANALYSIS

5.1 Model Definition

As an alternative to fitting all the data simultaneously to model (17), we consider fitting the data for each of the 27 individual cases, one for each (i, j, k) combinations of water, nitrogen and block, for each of the two years directly to the nonlinear regression model (11). The random errors for each (i, j, k) case are assumed to have independent normal $N(0, \sigma_{ijk}^2)$ distributions, hence differing from model (17) in which each case has a common $N(0, \sigma^2)$ error distribution. The model was fitted to the data using the nonlinear least squares option in SPSS (2002). The individual effects are incorporated into the five response variables namely N_{max} and t_{max} directly from (11) and λ, δ and $F(\infty)$ from (12) and (16).

Let Z_{ijk} denote any one of the five response variables for case (i, j, k). As before, the data for each year was analyzed separately due to the different environmental conditions and sampling schedules. The assumed ANOVA model for each year is

$$Z_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \rho_k + \varepsilon_{ijk} \qquad (22)$$

where α, β, γ and ρ denote the fixed water, nitrogen, water by nitrogen, and block effects and ε_{ijk} is the random error term. This is a restated form of ANOVA models (19)-(21).

5.2 Model Results

The analysis for the 2003 data included the γ_{ij} interaction terms, which is different from the nlme analysis. The factor level means for the new analysis are given in Table 3A. Though the means are slightly

different numerically, the results are similar qualitatively. The Water main effects are significant for three variables t_{max}, δ and $F(\infty)$. In particular, as the level of water increases over the three levels, $F(\infty)$ decreases by 21%, t_{max} decreases slightly by 4% and δ increases substantially by 44% all of which are very close to the previous results. The one different result is that although N_{max} increases by 19% with increasing water, this new result is not substantially significant (with $p = 0.18$) as it was previously. As before, the Nitrogen main effect is not significant for any of the five variables, and neither is the Water by Nitrogen interaction which is included in this model.

Table 3. Factor level means from the least squares analysis for main effects, Water and Nitrogen, for five endpoint variables N_{max}, t_{max}, λ, δ and $F(\infty)$ with p-values for testing equality

A. Results for 2003

Level	N_{max}	t_{max}	λ	δ	$F(\infty)$
For Water					
1	31.99	2.77	2.95	0.149	43.40
2	27.36	2.64	3.02	0.202	36.36
3	26.04	2.66	3.02	0.214	34.17
p-value	0.18	0.02	0.66	0.05	0.05
For Nitrogen					
1	25.70	2.67	2.92	0.188	34.95
2	29.41	2.69	3.12	0.200	38.71
3	30.28	2.71	2.94	0.176	40.26
p-value	0.34	0.77	0.44	0.65	0.33

B. Results for 2004

Level	N_{max}	t_{max}	λ	δ	$F(\infty)$
For Water					
1	23.77	5.45	2.42	0.135	39.52
2	22.03	5.48	2.25	0.129	41.49
3	21.29	5.46	2.30	0.132	38.56
p-value	0.26	0.90	0.31	0.77	0.49
For Nitrogen					
1	23.34	5.48	2.59	0.151	36.68
2	23.55	5.43	2.32	0.132	40.24
3	20.20	5.48	2.05	0.114	42.66
p-value	0.08	0.70	0.003	0.008	0.10

The results for the analysis of the 2004 data are given in Table 3B, and they too are qualitatively similar to those from the nlme analysis. As before, the Water main effect is not significant for any of the five endpoint variables nor is the Water by Nitrogen interaction significant. As the level of nitrogen increased from Zero to Blanket, the two kinetic rates decreased significantly, the birthrate l decreasing by 19% and the death rate d by 25%.

The results from this least squares analysis are presented in detail in Matis *et al.* (2007). The discussion includes testing contrasts for linear and for pairwise comparisons. Biological interpretations of the results are also given. For present purposes, however, the chief conclusion is that the results of the two analyses are similar qualitatively, as one might expect.

6. DISCUSSION

The principle objective of this paper was to investigate whether the nonlinear regression model (11) could be used as the basis for a nonlinear random effects model analysis of a designed experiment. The model has previously been shown to fit diverse aphid abundance curves satisfactorily. This paper demonstrates the utility of the elegant nlme analysis based on this new model. Though there is obviously a learning curve involved with utilizing the sophisticated nlme software, the analysis reduces the effort required in fitting each individual curve with standard, simple nonlinear least squares software.

The model assumed that the blocks constituted a fixed effect, in part because there were so few of them. In the case where there are more blocks, a convenient generalization would be to assume a random block effect, where the block effects would have an independent $N(0, \sigma_b^2)$ distribution. This would add a mild assumption, which could be checked, but would retain a low number of parameters. This generalization is simple to implement with the nlme software.

The nlme analysis may also be based on an underlying linear differential equations model, instead of on an explicit solution such as regression model (11). However, the differential equations for the mechanistic models discussed in this paper, including the basic model

(5) and the generalized models (13) and (14), are all nonlinear. Consequently at the present time, any application of the generalized models (13) or (14) to data from a designed experiment would require first fitting the models to individual data sets using numerical integration, as demonstrated in Matis *et al.* (2007) and Matis *et al.* (2007), and then proceeding with the standard analysis outlined in Section 5.

Most of the local insect populations of interest in agricultural production collapse in size. This class of mechanistic models including (5), (7), (13), and (14) contains a mechanism based on cumulative size which explains mathematically how such populations might collapse. Given also that we are not aware of any competing analytical models for single population collapse, we are confident that this new class of models will be widely applied, not only to other aphid species but to many other insect pest populations as well.

REFERENCES

Andrewartha, H.G. (1951). *Introduction to the Study of Animal Populations*. University of Chicago Press, Chicago, IL.

Chattopadhyay, C., Agrawal, R., Kumar, A., Singh, Y.P., Roy, S.K., Khan, S.A., Bhar, L.M., Chakravarthy, N.V.K., Srivastava, A., Patel, B.S., Srivastava, B., Singh, C.P. and Mehta, S.C. (2005). Forecasting of *Lipaphis erysimi* on oilseed brassicas in India – A case study. *Crop Protection*, **24**, 1042-1053.

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. *J. Ind. Soc. Agril. Statist.*, **49**, 139-156.

Matis, J.H., Kiffe, T.R. and Parthasarathy, P.R. (1998). On the cumulants of population size for the stochastic power law logistic model. *Theor. Popul. Bio.*, **53**, 16-29.

Matis, J.H. and Kiffe, T.R. (2005). Predicting the Africanized bee invasion. In : *Statistics: A Guide to the Unknown*,

4th ed., R. Peck, G. Casella, G. Cobb, R. Hoerl, D. Nolan, R. Starbuck and H. Stern (eds.), Duxbury, Belmont, CA.

Matis, J.H., Kiffe, T.R., Matis, T.I. and Stevenson, D.E. (2005). Nonlinear stochastic modeling of aphid population growth. *Math. Biosci.,* **198**, 148-168.

Matis, J.H., Kiffe, T.R., Matis, T.I. and Stevenson, D.E. (2006). Application of population growth models based on cumulative size to pecan aphids. *J. Ag. Biol. Environ. Stat.*, **11**, 425-449.

Matis, J.H., Kiffe, T.R., Matis, T.I. and Stevenson, D.E. (2007). Stochastic modeling of aphid population growth with nonlinear, power-law dynamics. *Math Biosci.* (in press).

Matis, J.H., Kiffe, T.R., Matis, T.I., Jackman, J.A. and Singh, H. (2007). Population size models based on cumulative size, with application to aphids. *Ecol. Modelling*, **205**, 81-92.

Matis, J.H., Kiffe, T.R., Matis, T.I. and Chattopadhyay, C. (2007). Generalized insect population growth models with immigration and cumulative – size dependent dynamics. Manuscript.

Matis, T.I. 2006. http://webpages.acs.ttu.edu/tmatis/ datapage/ CottonAphidData.xls.

Matis, T.I., Parajulee, M.H., Matis, J.H. and Shrestra, R.B. (2007). Analysis of cotton aphid population dynamics data based on a cumulative size model. Manuscript.

Pinheiro, J.C. and Bates, D.M. (2000). *Mixed Effects Models in S and S-Plus*. Springer, New York.

Prajneshu (1998). A nonlinear statistical model for aphid population growth. *J. Ind. Soc. Agril. Statist.,* **51**, 73-80.

Renshaw, E. (1991). *Modeling Biological Populations in Space and Time*. Cambridge University Press, New York.

R-Project (2006). An Introduction to R (available from www.r-project.org).

SPSS (2002). *SPSS 11.0 for Macintosh*. SPSS Inc., Chicago, IL.