

Kansas State University Libraries
New Prairie Press

Conference on Applied Statistics in Agriculture 2007 - 19th Annual Conference Proceedings

USING THE CUMULATIVE-SIZE MECHANISTIC MODEL FOR ANALYZING INSECT DATA

T. I. Matis

J. H. Matis

G .J. Michels Jr.

M. N. Parajulee

See next page for additional authors

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Matis, T. I.; Matis, J. H.; Michels, G .J. Jr.; and Parajulee, M. N. (2007). "USING THE CUMULATIVE-SIZE MECHANISTIC MODEL FOR ANALYZING INSECT DATA," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1107>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

Author Information

T. I. Matis, J. H. Matis, G .J. Michels Jr., and M. N. Parajulee

Using the Cumulative-Size Mechanistic Model for Analyzing Insect Data

T.I. Matis^{*}, J.H. Matis[†], G.J. Michels Jr.[‡], M.N. Parajulee[§]

Abstract

Two data sets of aphid abundance are analyzed using a new cumulative-size based mechanistic model. The first data set pertains to the cotton aphid, and its analysis demonstrates the power of the mechanistic model-based approach. The second data set pertains to greenbug populations, and its analysis shows the key role that birth and death rate coefficients may play in predicting the peak and the cumulative population sizes.

1 Introduction

We have recently investigated a new ‘cumulative-size’ mechanistic model in which the per capita death rate of a local insect population is proportional to the cumulative past size of the population (Matis et al. 2005). One form of this model has an analytical solution, a nonlinear regression model, which is relative easy to fit to data. We have shown previously that this model fits well the observed abundance curves of the pecan, mustard, and cotton aphids (Matis et al. 2006, Matis et al. 2007a, Matis et al. 2007e, respectively)

Notwithstanding the previous successful fitting of the data, the model has met with some resistance among entomologists because of its novelty. The standard approach to population size modeling is to calculate the ‘intrinsic rate of increase’ from life table data (Wyatt and White 1977). Williams et al. (1999) suggest that this birth rate coefficient slows down over time, which leads to a general model for the ascending phase of the growth data. Costamagna et al. (2007) allow the intrinsic rate to become negative, thus yielding a model for the ascending and descending phases of a population size curve. To our knowledge, Prajneshu (1999) proposed the first mechanistic model with a birth and a death rate, in order to describe both the ascending and descending phase of population size curves. We illustrate a generalized, user-friendly form of his model in this paper.

This paper has two parts. The first uses the new nonlinear regression model to analyze a set of cotton aphid data. These data have previously been analyzed with a statistically sophisticated nonlinear mixed effects model, using the nlme routing in R (Matis et al. 2007b), and some ANOVA results have previously been interpreted in physiological terms (Matis et al. 2007e). The objective of the first part of this paper is to supplement these analyses with a transparent, textbook-like analysis which would be readily understood by entomologists. The data are given and interpreted graphically. The first part shows that the traditional analysis in common usage does not give any statistically

^{*}Department of Industrial Engineering, Texas Tech University, Box 43061, Lubbock, Texas 79494-3061

[†]Department of Statistics, Texas A&M University, 3143 TAMU, College Station, Texas 77843-3143

[‡]Texas Agricultural Experiment Station, 2301 Experiment Station Road, Bushland, Texas 79012

[§]Texas Agricultural Experiment Station, 1102 East FM 1294, Lubbock, Texas 79403

significant results for these data, whereas an analysis based on the new mechanistic model yields a number of relevant findings. We suggest that this first part makes a convincing case, with a straightforwardness not previously presented, for the utility of the new model.

The second part of the paper analyzes some new data on greenbug abundance to show a potential use of the new model for forecasting future greenbug abundance. The nonlinear models predicting the peak count and the total cumulative count are linearized using a Taylor series approximation. The resulting linear regression models for these variables, based on only the estimated birth and death rate coefficients, are found to be very accurate for describing greenbug abundance data. Such linear models represent a new approach for predicting population size endpoints directly from growth rate parameters.

2 Analysis of Cotton Aphid Data

2.1 Experiment and Analysis

A study was conducted in 2003 by the Texas Agricultural Experiment Station in Lubbock, TX, on the effect of water and of nitrogen on the population size of the cotton aphid. Three water levels, called Low, Medium, and High, consisting on 65%, 75%, and 85% evapotranspiration replacement, were applied using pivot irrigation. Also, three nitrogen levels, called Zero, Variable Rate, and Blanket Rate, were utilized. These nitrogen treatments were applied at planting and at first-square stage of cotton plant growth. The Blanket Rate treatment applied the same amount of residual nitrate to each experimental unit to bring the *average* level to 134 KgN/ha, whereas the Variable Rate treatment applied the appropriate amount to bring *each* unit to that desired level.

Twentyseven (27) plots, each of approximate 0.2 ha size, were used in the experiment. The nine Water x Nitrogen treatment combinations were deployed in a randomized block split-plot design (Ott and Longnecker 2001) as follows. The 27 plot field was divided into three contiguous blocks, each consisting of nine plots. Each block was subdivided into three distinct areas, or whole plots, consisting of three adjacent individual plots. The three irrigation treatments were randomly assigned to these whole plots within each block. The three nitrogen treatments were then randomly assigned to the individual plots, i.e. to the split plots, within each whole plot.

Cotton aphid abundance was monitored weekly, starting on July 8 with sampling dates of $t = 0, 1, 2, 3,$ and 5 weeks of elapsed time thereafter. Cotton aphids were counted on 10 randomly chosen leaves from the upper half and 10 leaves from the lower half of plants from each plot. For subsequent identification, the three water treatment combinations (Low, Medium, and High) are denoted as $i = 1, 2,$ and 3; the three nitrogen treatments (Zero, Variable Rate, and Blanket Rate) as $j = 1, 2,$ and 3; and blocks for each treatment combinations as $k = 1, 2,$ and 3. The plot with water treatment i and nitrogen treatment j in block k is denoted plot ijk , and the mean count per leaf from this plot at time $l = 0, \dots, 5$ as X_{ijkl} . The complete data set, consisting of 27 observed means for each of the five sampling times, is illustrated in Fig. 1A and is available in Parajulee (2006). Table 1 gives the means for each of the nine (i, j) treatment combinations averaged over the three blocks, for each sampling time.

2.2 ANOVA Analysis of Standard Variables

The standard approach for the analysis of this common type of insect abundance experiment would be to analyze response variables observed directly from the data. Typically the following three response variables are used (see e.g. Chattopadhyay et al. 2005): the crop age at which the insect first appears

on the crop, denoted Y_1 ; the crop age at peak population size, Y_2 ; and the peak number of insects on the crop in the season, Y_3 . We call this the ‘model-free’ approach.

Our assumed ANOVA model, letting Y_{ijk} denote an individual observation for any one of the three above response variables of interest, is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \rho_k + \alpha\rho_{ik} + \varepsilon_{ijk} \quad (1)$$

where α , β , and ρ denote, respectively, the fixed Water, Nitrogen and Block effects, with $\alpha\beta$ and $\alpha\rho$ two-factor interactions, and random error term ε_{ijk} . In this experiment, the time of first appearance, Y_1 , was not observed. The time of peak, Y_2 , was either the integer 2 or 3, and not surprisingly is not significant for any effect. For these data, Y_3 is the primary standard response variable. It has a clear Block main effect, which is plainly visible in the graphs in Fig. 1A. Table 2 lists some summary results from the ANOVA model (1) analysis of various response variables. Note that neither the Water nor Nitrogen main effect, nor the Water x Nitrogen interaction is significant (at $\alpha = 0.05$) for Y_2 or Y_3 . Hence, the analysis based on the standard model-free variables does not appear to yield *any* treatment effects of interest.

2.3 New Regression Model for Aphid Abundance

A new nonlinear regression model was recently derived in Matis et al. (2007a) that describes the abundance of aphids in a local population. Letting $N(t)$ denote the number of aphids at time t , the model is given by

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

where parameter N_{max} denotes the peak count, t_{max} is the time of peak count, and b is an approximate birth rate. This regression model is the solution to the underlying mechanistic model

$$N(t) = [\lambda - \delta F(t)] N(t) , \quad (3)$$

which makes certain assumptions about aphid population changes. The population’s ‘birthrate’ at any time t is assumed to be $\lambda N(t)$, where λ denotes the ‘per capita’ birthrate or the ‘intrinsic rate of natural increase’. The novelty of the model lies in it’s assumed death rate. In particular, letting $F(t)$ denote the ‘cumulative density’, which is defined mathematically as

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

and conceptually as the total area under the $N(t)$ curve in units of insect-time, the population’s per capita death rate, $\delta F(t)$ is a function of the cumulative size of past generations of aphids. $F(t)$ may be regarded as a measure of past environmental degradation associated with the insect.

Once the observed data are fitted to regression model (2), the least squares estimates of N_{max} , t_{max} , and b may be used to estimate the parameters of mechanistic model (3). In particular, letting

$$d = e^{-(bt_{max})} \quad (5)$$

we have shown (Matis et al. 2007a) that

$$\begin{aligned} \lambda &= b(d - 1)/(d + 1) \\ \delta &= b^2/2N_{max} \\ N_0 &= 4dN_{max}(1 + d)^{-2} \end{aligned} \quad (6)$$

where N_0 is the estimate of the initial population size $N(0)$. Because d in (5) is usually very large, b may be noted from (6) to be an accurate approximation of the assumed birthrate λ in (3). An accurate approximation for the final cumulative density in (4), denoted as $\tilde{F}(\infty)$, is given by

$$\tilde{F}(\infty) = 4N_{max}/\lambda \quad (7)$$

2.4 ANOVA Analysis Based on Mechanistic Model

Model (2) was fitted to each of the 27 abundance curves illustrated in Fig. 1 using a standard nonlinear least squares program (SPSS 2002). We have previously shown that the aphid counts over time in any given plot are heteroscedastic and have weak serial correlation (Matis et al. 2005). However, for simplicity as is Matis et al. (2006), we assume that these observations are independent with constant variances, and so use the simple unweighted least squares option. This choice yields asymptotically unbiased, though not optimal, estimates (Milliken and Milliken-MacKinnon, 1997).

The fitted curves describe the data exceptionally well, as illustrated in Fig. 1B. The modified R^2 ,

$$R^2 = 1 - \text{ResidualSS}/\text{CorrectedSS} ,$$

for this nonlinear model exceeds 0.99 for each case, and the only visible residuals occur at $t=1$. These outstanding fittings may not be surprising in light of the small number (5) of data points. The parameter estimates for each curve are given in Table 3.

Two of the parameters, N_{max} and t_{max} , have immediate natural interpretation, as the corresponding fitted estimates of the observed Y_2 and Y_3 variables. Consider introducing these estimates as new response variables, with $Y_4 = N_{max}$ and $Y_5 = t_{max}$. When these variables are analyzed, again using ANOVA model (1), there are no significant effects for N_{max} , however there is a significant ($p = 0.02$) Water main effect for t_{max} , as given in Table 1. The means for the individual Water and Nitrogen levels are given in Table 4, and a profile plot of the means for t_{max} is given in Fig. 2A. It's apparent both from Table 4 and Fig. 2A that Low water leads to a delayed peak, i.e. to higher mean t_{max} .

Consider now transforming the estimated parameters to three additional variables of interest, namely the birth rate coefficient, λ , the death rate coefficient, δ , and the cumulative density, $\tilde{F}(\infty)$, from (6) and (7). We denote these as Y_6, Y_7 , and Y_8 , respectively. The resulting means of these variables for the levels of Water and Nitrogen are given in Table 4, from whence it is apparent that both δ and $F(\infty)$ have a significant Water main effect. The profile plots of corresponding means are illustrated in Figs. 2B and 2C. As water increases, it is clear that the death rate tends to increase and that the cumulative density tends to decrease. The latter appears to be visible by inspection of Fig. 1B, though it does not appear to be humanly possible to observe the effect on the death rate visually in the figure.

2.5 Interpretation of Results for Cotton Aphid Data

Outstanding among the interpretations of these data are the following:

1. The exceptional fitting of model (2) to the data is expected for, as indicated previously, the model provides an outstanding description also for pecan, mustard, and greenbug aphid data
2. Most aphid field data are observed approximately weekly, as opposed to daily monitoring. Consequently, the observed time of peak, Y_2 , is restricted to only values 2 or 3 for these data, and hence is far less sensitive than the estimated counterpart $Y_5 = t_{max}$. Similarly, one would not expect the peak count to occur precisely at time 2.00 or 3.00, and hence the observed peak

at either one of those exact times is arguably an underestimate of the true peak. By way of contrast, the estimated peak, $Y_4 = N_{max}$, has flexibility to combine mechanistically the data observed at other times for a more realistic estimate of the peak. Though this data set is based on only 5 times of observation, a more extensive study of 192 curves for the pecan aphid observed over 11 weeks (Matis et al. 2006) had the same qualitative finding, namely that the *estimated size and time of peak from the model were more sensitive to treatment effects than the directly observed size and time of peak*. More generally, we submit that this example is a simple, compelling example of the merits of using a mechanistic model-based analysis. The standard model-free analysis failed to find any effects of interest, however the model-based analysis not only detected significant effects, but detected effects on variables which explain the changes mechanistically.

3. Year 2003 was a relatively dry year, which explains the Water main effect. Matis et al. (2007e) analyzes data also from year 2004 which had moderate precipitation. No significant Water main effect is found for any of the variables for 2004, however there are significant Nitrogen main effects for $Y_4 = N_{max}$, $Y_6 = \lambda$, and $Y_7 = \delta$. The paper also gives physiological interpretations for the main effects on the birth and death rates for both years, and for their relationship to the peak count.
4. The cumulative density, in this case in units of insect-weeks, is a key endpoint variable of practical interest. In some cases, one could estimate this integral directly from data using the trapezoidal rule. However in cases, as in this one, where a point is missing and/or where the dates are unequally spaced, the trapezoidal rule becomes more difficult computationally. The model based $\bar{F}(\infty)$ is relatively simple to obtain from model (2) even for unequally spaced data.
5. A statistically more sophisticated analysis of these data was performed using the nonlinear mixed effects model using the nlme routine in R (Matis et al. 2007b). In the analysis in the present paper, each of the 27 data sets is fitted separately to nonlinear model (2), and the resulting estimates of Y_4, \dots, Y_8 were analyzed using the ANOVA model (1) for each response variable. Whilst this is valid statistically, the nlme procedure utilizes a nonlinear ANOVA model based on (2) for the directly observed X_{ijkl} counts with a single combined error assumption. The results from the nlme analysis are near identical qualitatively to those in this paper. Though it is more elegant to use the combined model, the nlme procedure requires special software, and the procedure does not extend to generalizations of (3) which have only numerical solutions, as opposed to an analytical solution such as (2). A second alternative approach would be to generalize the use of ANOVA model (1) for comparing the Y_4, \dots, Y_8 parameters to the ‘beta-hat model’, in which the observations are assumed to be independent, but their individual standard errors are incorporated (Milliken and Johnson, 2002). This was not done, but we expect that the results in Table 2 would not change.
6. Insect management procedures tend to focus on altering the underlying dynamics of the population, e.g. to introduce treatments which would either increase or decrease the birth and/or death rates as desired. This model based analysis hence addresses the key variables for insect *control*. The next section addresses the use of λ and δ for population size *prediction*.

3 Analysis of Greenbug Data

3.1 Data and Model

Data were gathered by the Texas Agricultural Experiment Station on greenbug, corn leaf aphid, and ladybeetle abundance at various irrigated and dryland sorghum fields in a study from 1988 through 2000 (Michels 2007). The fields were sampled approximately weekly, unless there was inclement weather. At each sampling date, 12 randomly selected 50' long strips were marked in each field. In each strip, 12 plants were randomly selected, cut off at ground level, and examined for corn leaf aphid and greenbugs, hence 144 plants were sampled at each location on each date.

Thirtythree fields were sampled in this study. The most frequent sampling occurred in the irrigated fields at the Bushland Experiment Station. The mean greenbug counts in these fields are illustrated in Fig. 3 for the $n = 12$ available years. Samples were also taken for four years from Bushland rainfed fields, six years from Etter dryland fields, seven years from Etter rainfed fields, and from two additional dryland and irrigated fields, for the total of 33 fields. We made the simplifying assumption that all data were gathered at *exact* weekly intervals, as indicated in Fig. 3. Hence, the following results must be regarded as preliminary, and a more extensive analysis using the exact daily sampling intervals is in progress.

Each individual data set was fitted to model (2). Through the model fitted each case successfully, six cases were disregarded, as the largest count in these cases was either the first or the last observation, with resulted in an unreliable estimate of the peak, N_{max} . The remaining 27 cases, each case being data from one field for one year, had modified R^2 values ranging from 0.8 to 0.999, with a mean of 0.96. As an illustration, the fitted curves for the 12 years at the irrigated Bushland fields are displayed in Fig. 3.

3.2 Linear Model Approximations of Key Endpoints

The two primary endpoint variables of interest to entomologists are the peak count and the cumulative density. As noted, both of these may be estimated from N_{max} and $\tilde{F}(\infty)$ available from (2) and (7), however the role of growth rate parameters, λ and δ , are not immediately apparent in these calculations. We have suggested previously (Matis et al. 2007d) that a linear Taylor series approximation be used to estimate the peak size and the cumulative density directly as linear functions of the birth and death rate coefficients, λ and δ . The accuracy of these approximations is examined in this section, and its potential use for population size prediction is discussed in the next section.

The three directly observed, model-free response variables from an individual data set were previously denoted Y_1 , Y_2 , and Y_3 . For present convenience, we let the observed peak counts, previously Y_2 , now be called Z_1 . Let Z_2 denote the observed cumulative count, which for equally spaced data may be calculated using the trapezoidal rule as the sum of all observations in the individual data set. For example, for the 12 years of data displayed in Fig. 3, the smallest of these responses are $Z_1 = 14$ and $Z_2 = 25$ for year 1999. The largest responses are $Z_1 = 1665$ and $Z_2 = 2321$ for year 1990, which has a sharp peak. As a contrast, the second largest Z_2 is $Z_2 = 2252$ observed in year 1997, which has a much smaller peak size of $Z_1 = 829$, but also has a much wider spread. The issue is whether one could describe these attributes of the population size curve directly from a *linear* model based on λ and δ .

The estimated birth and death rate coefficients for the above extreme cases are 2.497 and 0.2195 for 1999, 3.728 and 0.0034 for 1990, and 1.479 and 0.0013 for 1997. Fig. 4 illustrates 3-D plots of $\log(Z_1)$ and $\log(Z_2)$ versus estimated λ and $\log(\delta)$ for all $n = 27$ cases.

Consider first simple linear models for $\log(Z_1)$. Fig. 5A and 5B illustrate the simple linear regression lines, with $\log(\delta)$ and λ as independent variables, and they have $R^2 = 0.84$ and 0.15 respectively. The first-order multiple regression equation model, a plane in space as illustrated in Fig. 6A, has equation

$$\log(Z_1) = -0.48 + 0.60\lambda - 0.48 \log(\delta)$$

with $R^2 = 0.98$.

Fig. 5C and 5D illustrate the simple linear regression lines for observed variable Z_2 as functions of $\log(\delta)$ and λ , and they have $R^2 = 0.92$ and 0.06 , respectively. The multiple regression model, illustrated in Fig. 6B, has equation

$$\log(Z_2) = 0.68 + 0.38\lambda - 0.98 \log(\delta)$$

with $R^2 = 0.99$. These are obviously very high R^2 values.

3.3 Interpretation of Results for Greenbug Data

Current practice is to predict the population endpoint variables, such as Z_1 and Z_2 , *directly* through sophisticated linear regression models based on extensive present and past weather (e.g. Chattopadhyay et al. 2005) or through extensive simulation models based on such variables as aphid fecundity, mortality, effect of temperature, photoperiod, biological control, plant quality, and density dependency (e.g. Parajulee et al. 2004). The results in this paper suggest a different, more mechanistic approach to prediction, namely using all past and present ambient conditions to predict the birth and death rates, which then give near perfect predictions of Z_1 and Z_2 . The coefficient λ may be regarded as a sensitive index to determine the effect of local weather, environmental and management conditions on aphid reproduction, and the coefficient δ of the effect of local conditions on aphid ‘mortality’ (including emigration). In short, the results suggest that λ and $\log(\delta)$ are implicit functions of these local conditions, after all only 2% or less of the variability in $\log(Z_1)$ and $\log(Z_2)$ are due to conditions which are not implicitly incorporated into the birth and death rate coefficients.

At present, these are just conceptual concepts, however we can envision that in time experimenters could develop regression relationships to predict the growth rate coefficients, λ and δ , from local ambient conditions. A previous finding in Section 2.4 is strong evidence to support this notion, as it demonstrates that the death rate coefficient, δ , is a function of water abundance. Research is currently in progress to relate the λ and δ parameters for these 27 fields 1) to the abundance of the ladybug predator and the corn leaf aphid competitor and 2) to the recorded weather data. We are not aware of any previous efforts to use this prediction approach based on these underlying parameters of population growth.

4 Summary

A new cumulative-size based mechanistic model has recently been developed for describing aphid population size on the basis of aphid birth and death rates. The model has an analytical solution, which has been shown to fit abundance curves for a number of aphid species. The analysis of the data set on the cotton aphid demonstrates that an analysis based on this solution may be far more informative and explanatory than the current standard model-free analysis. The analysis of the greenbug data set demonstrates the utility of using linear regression models based on the birth and death rate coefficients for predicting peak and cumulative aphid population size.

Some aphid abundance curves are noticeable skewed, which is particularly true when the local aphid population increases initially due to immigration from other areas rather than from local reproduction. Research is in progress to add an initial immigration component to mechanistic model (3) (Matis et al. 2007c). Though we have fitted the models only to aphid data, we are confident that, due to their very general mechanistic nature, the models could describe the abundance curves for many other insect species as well.

5 Acknowledgement

We would like to express our appreciation to the reviewer of this paper whose suggestions helped improve this work.

References

- 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., Mehta, S. C. 2005. Forecasting of *Lipaphis erysimi* on oilseed Brassicas in India a case study. *Crop Protection* 24:1042-1053.
- Costamagna, A. C., van der Werf, W., Bianchi, F. J. J. A., Landis, D. A. 2007. An exponential growth model with decreasing r captures bottom-up effects on the population growth of *Aphis glycines* Matsumura (Hemiptera: Aphididae). *Agricultural and Forest Entomology*, in press.
- Matis, J. H., Kiffe, T. R., Matis, T. I., Stevenson, D. E. 2005. Nonlinear stochastic modeling of aphid population growth. *Mathematical Biosciences* 198:81-92.
- Matis, J. H., Kiffe, T. R., Matis, T. I., Stevenson, D. E. 2006. Application of population growth models based on cumulative size to pecan aphids. *J. Agricultural, Biological, and Environmental Statistics* 11:425-449.
- Matis, J. H., Kiffe, T. R., Matis, T. I., Jackman, J. A., Singh, H. 2007a. Population size models based on cumulative size, with application to aphids. *Ecological Modelling* 205: 81-92.
- Matis, J. H., Zhou, L., Kiffe, T. R., Matis, T. I. 2007b. Fitting cumulative size mechanistic models to insect population data: A nonlinear mixed effects model analysis. *J. Indian Society Agricultural Statistics*, in press.
- Matis, J. H., Kiffe, T. R., Matis, T. I., Chattopadhyay, C. 2007c. Generalized insect population growth models with immigration and cumulative-size dependent dynamics. Manuscript.
- Matis, J. H., Kiffe, T. R., Matis, T. I., Jackman, J. A., Grant, W. E., Singh, H. 2007d. On the use of growth rate parameters for projecting population sizes: Application to aphids. Manuscript.
- Matis, T. I., Parajulee, M. N., Matis, J. H., Shrestha, R. B. 2007e. Analysis of cotton aphid population dynamics data based on a cumulative size model. Manuscript.
- Michels, G. J. Jr. 2007. Greenbug Data, Texas Agricultural Experiment Station Farms, Bushland and Etter, Texas, 1988 through 2000, <http://webpages.acs.ttu.edu/tmatis/datapage/GreenbugData.xls>

- Milliken, G.A., Milliken-MacKinnon, A.J. 1997. Analysis of repeated measures data using nonlinear models, In *Mathematical Models in Microbiology*, Chapman & Hall, New York, NY.
- Milliken, G.A., Johnson, D.E. 2001. *Analysis of messy data, Vol. 3: Analysis of covariance*, Chapman & Hall, New York, NY.
- Ott, R. L., Longnecker, M. 2001. *An introduction to statistical methods and data analysis*. 5ed. Duxbury, Pacific Grove, CA.
- Parajulee, M. N., Slosser, J. E., Wilson, L. T. 2004. Cotton aphid dynamics modeling. pp. 1584-88. In *Proceedings, Beltwide Cotton Conferences*, National Cotton Council, Memphis, TN.
- Parajulee, M. N. 2007. Cotton Aphid Data, Agricultural Complex for Advanced Research and Extension Systems Farm, Lamesa, Texas, 2003 and 2004, <http://webpages.acs.ttu.edu/tmatis/datapage/CottonAphidData.xls>
- Prajneshu. 1998. A nonlinear statistical model for aphid population growth. *J. Indian Society Agricultural Statistics* 51: 7380.
- SPSS. 2002. *SPSS 11.0 for Macintosh*. SPSS Inc., Chicago, IL.
- Williams, I. S., van der Werf, W., Dewar, A. M., Dixon, A. F. G. 1999. Factors affecting the 12 relative abundance of two coexisting aphid species on sugar beet. *Agricultural and Forest Entomology* 1:119-125.
- Wyatt, I. J., White, P. F. 1977. Simple estimation of intrinsic rates for aphids and tetranychid mites. *J. Applied Ecology* 14:757-766.

Table 1: Mean aphid count per leaf, averaged over blocks, for each Water by Nitrogen treatment combination at each sampling time.

Water Level	Nitrogen Level	Sampling Time				
		0	1	2	3	5
1	1	0.56	0.79	9.93	22.76	0.11
1	2	0.44	1.15	11.07	29.99	0.11
1	3	0.53	0.85	10.99	31.11	0.12
2	1	0.29	0.91	11.63	20.59	0.03
2	2	0.42	0.69	10.87	24.97	0.05
2	3	0.57	0.61	11.98	18.48	0.11
3	1	0.70	0.66	10.95	17.81	0.15
3	2	0.32	0.75	10.27	20.96	0.16
3	3	0.45	0.56	9.23	22.48	0.14

Table 2: Significance of Water and Nitrogen main effects and of Water x Nitrogen interaction for ANOVA on seven individual response variables. (ns denotes nonsignificant)

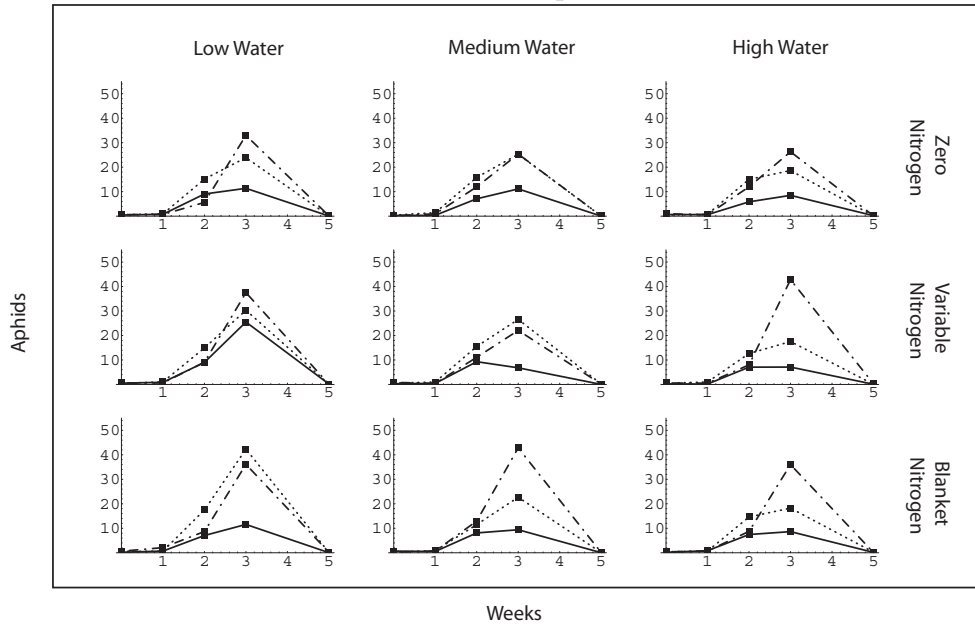
Variable	Water (α)	Nitrogen (β)	Water x Nitrogen ($\alpha\beta$)
Y_2 , Obs. Time of Peak	ns	ns	ns
Y_3 , Obs. Peak Count	ns	ns	ns
N_{max} , Est. Peak Count	ns ($p = .06$)	ns	ns
t_{max} , Est. Time of Peak	$p = .02$	ns	ns
λ , Birth Rate Coeff.	ns	ns	ns
δ , Death Rate Coeff.	$p = .02$	ns	ns
$\tilde{F}(\infty)$, Est. Cum. Density	$p = .05$	ns	ns

Table 3: Estimates of regression parameters (N_{max}, t_{max}, b) for each of 27 observed curves (3 Water x 3 Nitrogen x 3 Block)

Water Level, i	Nitrogen Level, j	Block, k	N_{max}	t_{max}	b
1	1	1	15.89	2.57	2.76
1	1	2	32.8	2.62	3.03
1	1	3	33	3	3.06
1	2	1	26.62	2.84	2.69
1	2	2	36.2	2.71	2.88
1	2	3	38.53	2.89	3.03
1	3	1	14.01	2.66	2.64
1	3	2	54.77	2.69	3.38
1	3	3	36.07	2.99	2.67
2	1	1	16.25	2.61	3.22
2	1	2	31.57	2.65	2.72
2	1	3	39.37	2.72	2.81
2	2	1	15.26	2.43	3.39
2	2	2	34.73	2.64	2.98
2	2	3	30.19	2.65	3.32
2	3	1	14.46	2.54	2.96
2	3	2	27.46	2.69	2.89
2	3	3	46.99	2.8	3.13
3	1	1	10.1	2.65	2.39
3	1	2	30.13	2.55	3.21
3	1	3	32.18	2.7	3.07
3	2	1	16.46	2.5	3.94
3	2	2	23.79	2.6	2.81
3	2	3	42.91	2.97	3.07
3	3	1	11.79	2.55	2.55
3	3	2	30.17	2.55	3.29
3	3	3	36.79	2.9	2.99

Figure 1: Mean aphid counts per leaf in designed cotton aphid study in 2003. The columns identify the water treatments and the rows the nitrogen treatments. The blocks are identified by the different lines (solid, dashed, mixed).

A. Observed mean aphid counts



B. Fitted aphid curves

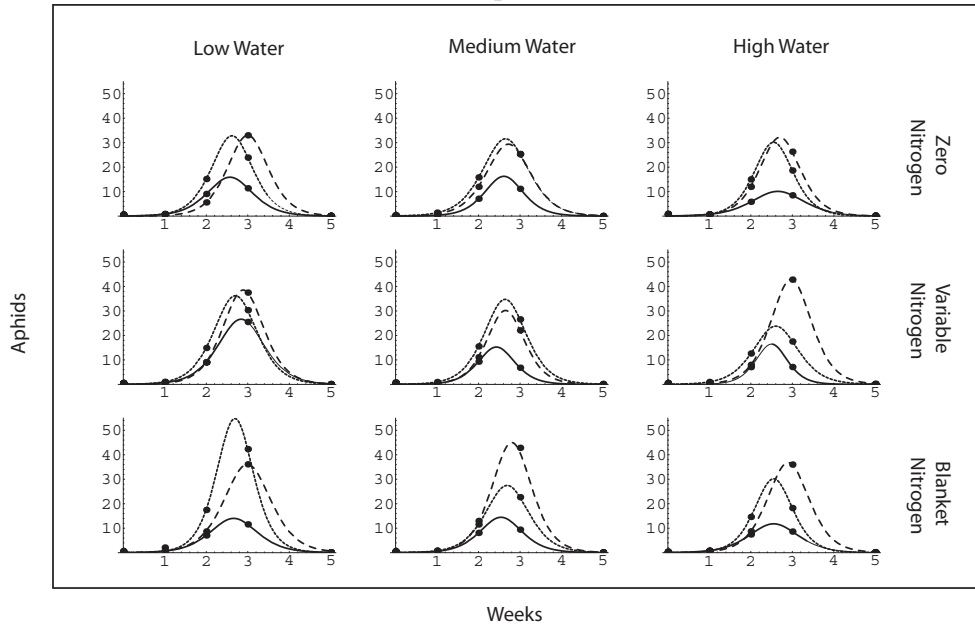
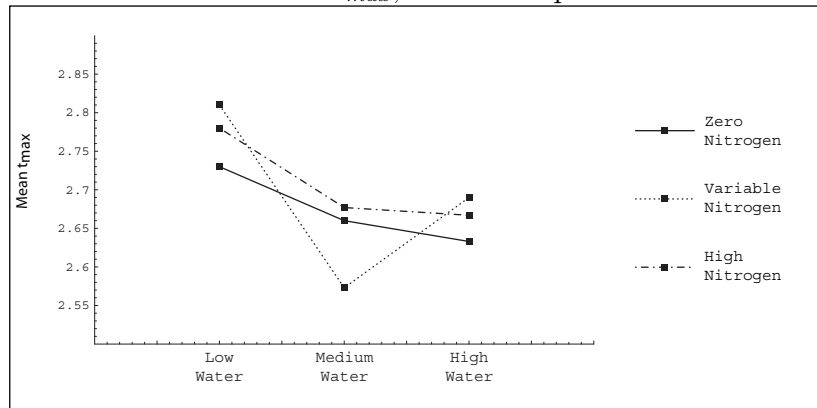
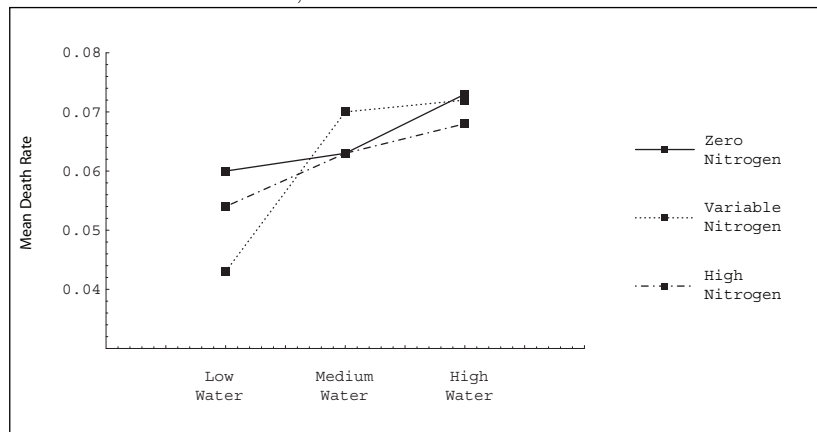


Figure 2: Profile plots of mean responses for Water x Nitrogen levels

A. Means for t_{max} , estimated peak time



B. Means for δ , estimated death rate coefficient



C. Means for $\tilde{F}(\infty)$, estimated cumulative density

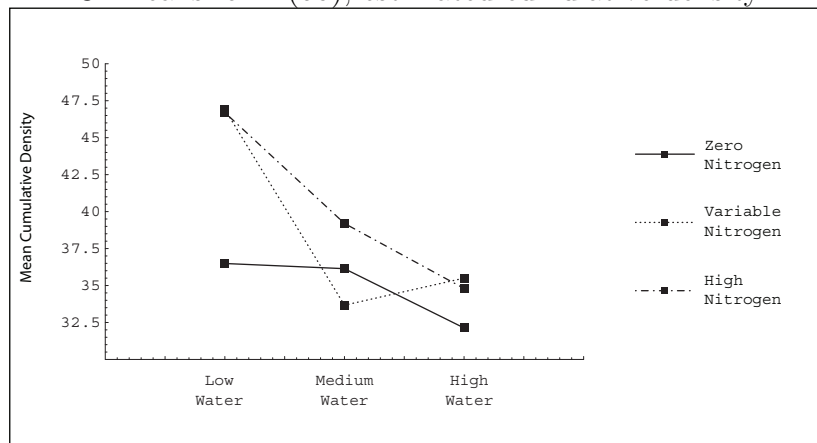


Table 4: Means and standard errors of response variables in 2003 data for each Water and Nitrogen treatment, with p-value for testing equality

Measure	N_{max}	t_{max}	λ	δ	$\tilde{F}(\infty)$
<u>A. For Water Levels</u>					
Means for Low	31.99	2.77 ¹	2.95	0.149 ¹	43.4 ¹
Means for Medium	27.36	2.64 ²	3.02	0.202 ²	36.36 ¹²
Means for High	26.04	2.66 ²	3.02	0.214 ²	34.17 ²
St. Error of Mean	2.27	0.03	0.12	0.018	2.52
p-value	0.18	0.02*	0.66	0.05*	0.05*
<u>B. For Nitrogen Levels</u>					
Means for Zero	25.7	2.67	2.92	0.188	34.95
Means for Variable	29.41	2.69	3.12	0.200	38.71
Means for Blanket	30.28	2.71	2.94	0.176	40.26
St. Error of Mean	2.27	0.03	0.12	0.018	2.52
p-value	0.34	0.77	0.44	0.65	0.33

* $p \leq 0.05$. Means without common superscript are significantly different, $\alpha = 0.05$ LSD tests.

Figure 3: Observed mean counts and fitted curves for weekly sampling intervals for 12 years of greenbug data on irrigated fields at Bushland Station

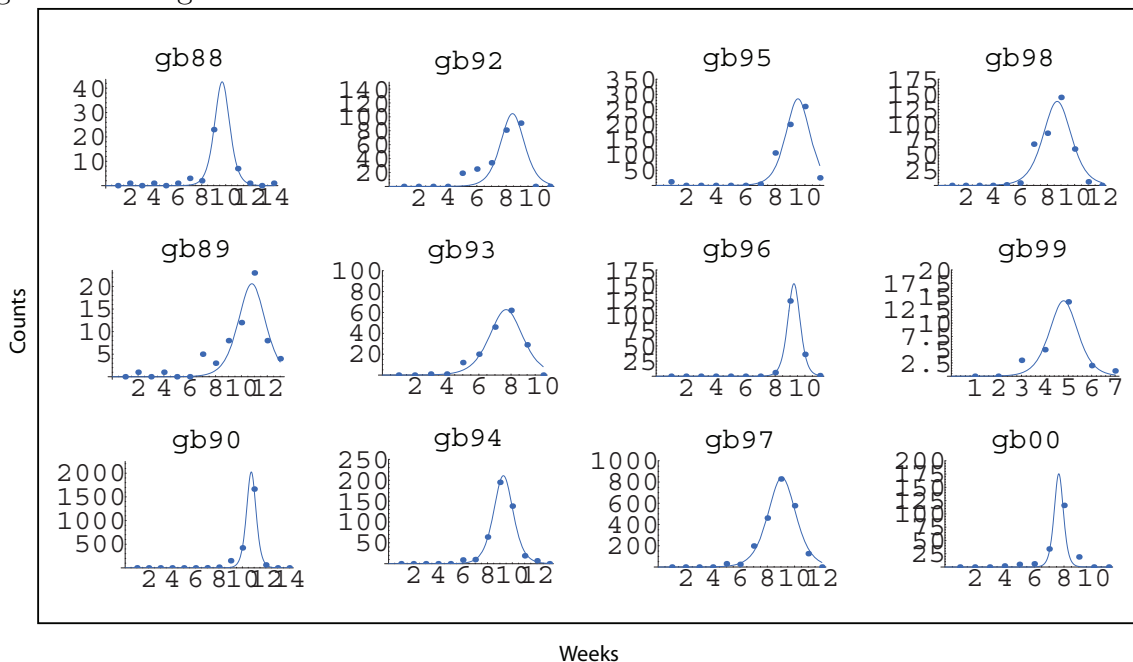


Figure 4: 3-D plots of aphid population endpoints versus estimated λ and $\log(\delta)$. A. Plot of log peak size. B. Plot of log cumulative density

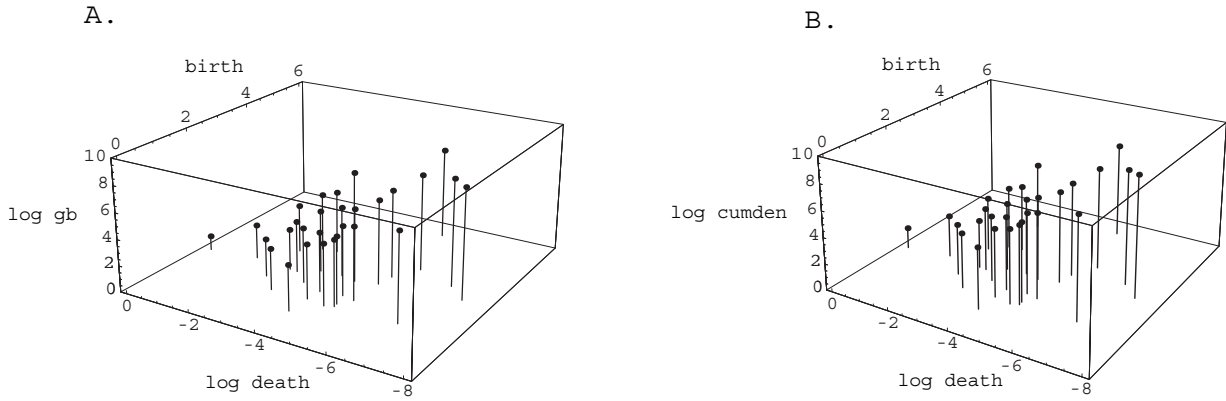


Figure 5: Scatter plots and fitted linear regression lines. A. Log peak size versus $\log(\delta)$. B. Log peak size versus λ . C. Log cumulative density versus $\log(\delta)$. D. Log cumulative density versus λ .

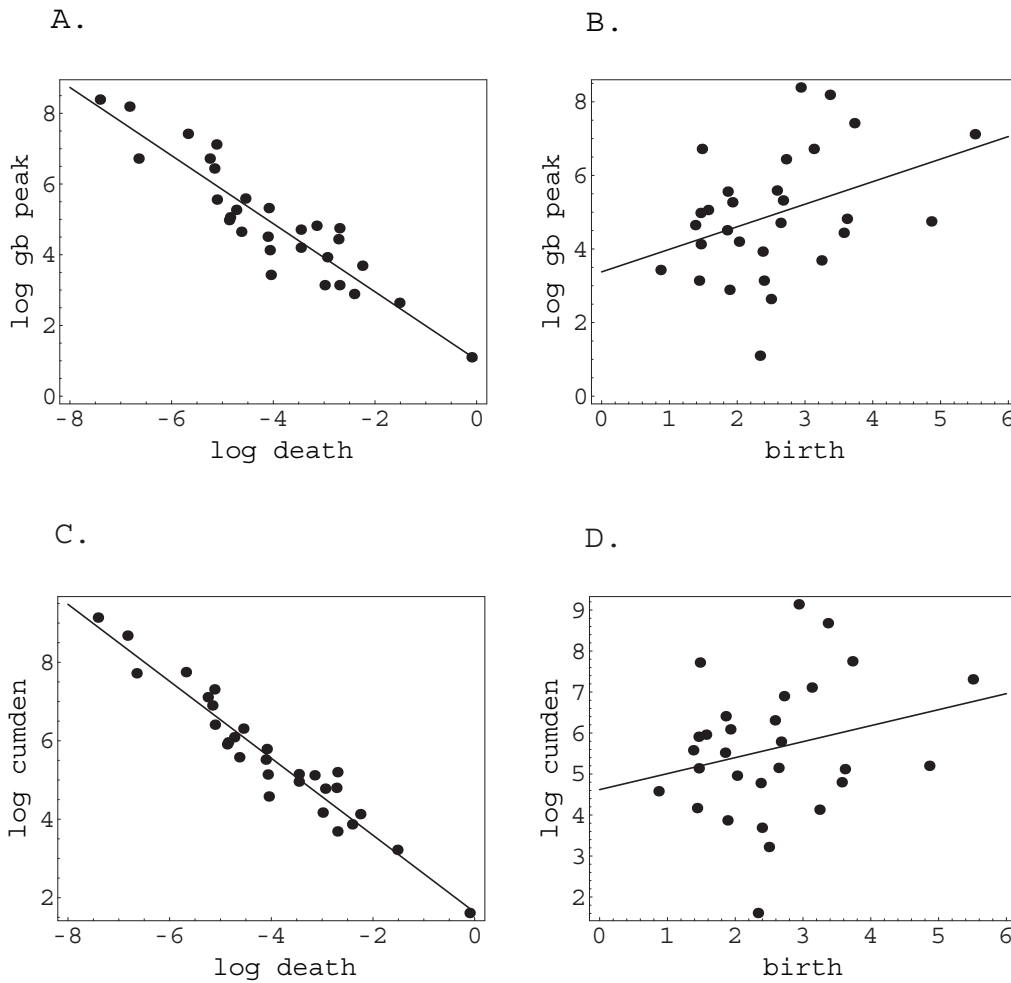


Figure 6: 3-D plots of aphid population endpoints and fitted first order model with $\log(\delta)$ and λ as independent variables. A. Log peak size. B. Log cumulative density

