Modelling Pesticide Treatment Effects on *Lygus hesperus* in Cotton Fields

H.T. Banks[1(✉)], J.E. Banks[1,2], Neha Murad[1], J.A. Rosenheim[3], and K. Tillman[1]

[1] Center for Research in Scientific Computation, North Carolina State University, Raleigh, NC 27695-8212, USA
htbanks@ncsu.edu
[2] Undergraduate Research Opportunities Center, California State University, Monterey Bay, Seaside, CA 93955, USA
[3] Department of Entomology and Nematology, Center for Population Biology, University of California, Davis, Davis, CA 95616, USA

Abstract. We continue our efforts on modeling of the population dynamics of herbivorous insects in order to develop and implement effective pest control protocols. In the context of inverse problems, we explore the dynamic effects of pesticide treatments on *Lygus hesperus*, a common pest of cotton in the western United States. Fitting models to field data, we consider model selection for an appropriate mathematical model and corresponding statistical models, and use techniques to compare models. We address the question of whether data, as it is currently collected, can support time-dependent (as opposed to constant) parameter estimates.

Keywords: Inverse problems · Generalized least squares · Model selection · Information content · Residual plots · Piecewise linear splines · Hemiptera · Herbivory · Pesticide

1 Introduction

When addressing questions in fields ranging from conservation science to agricultural production, ecologists frequently collect time-series data in order to better understand how populations are affected when subjected to abiotic or biotic disturbance [12,13,22]. Fitting models to data, which generally requires a broad understanding of both statistics and mathematics, is an important component of understanding pattern and process in population studies. In agricultural ecology, pesticide disturbance may disrupt predator-prey interactions [27,28] as well as impose both acute and chronic effects on arthropod populations. In the past several decades, the focus of many studies of pesticide effects on pests and their natural enemies has shifted away from static measures such as the LC50, as authors have emphasized population metrics/outcomes [17–19,26,29]. Simple mathematical models, parameterized with field data, are often used to then predict the consequences of increasing or decreasing pesticide exposure in the field. Accuracy in parameter estimation and quantification of uncertainty in fitting

data to models, which has recently received increasing attention in ecological circles [20,21], depends critically on the appropriate model selection. In most cases, this includes selection of both **statistical and mathematical models** in fits-to-data – something that is not always fully explicitly addressed in the ecological literature. We first addressed this gap in [1,3] using data from pest population counts of *Lygus hesperus* Knight (Hemiptera: Meridae) feeding on pesticide-treated cotton fields in the San Joaquin Valley of California [23].

In particular, in [1,3] we investigated the effect of pesticide treatments on the growth dynamics of *Lygus hesperus*. This was done by constructing mathematical models and then fitting these models to field data so as to estimate growth rate parameters of *Lygus hesperus* both in the absence and in the presence of pesticide application. Overall, compelling evidence was found for the untreated fields, using model comparison tests, that it may be reasonable to ignore nymph mortality (i.e., just count total number of *L. hesperus* and not distinguish between nymphs and adults). This would greatly simplify the models, *as well as the data collection process.*

In the present effort we further examine the importance of model selection and demonstrate how optimal selection of both statistical and mathematical models is crucial for accuracy in parameter estimation and uncertainty quantification in fitting data to models. This report further investigates these issues by testing different data sets from the same database as in [1,3] but with a varied number of pesticide applications in treated fields.

2   Methods

The data used came from a database consisting of approximately 1500 replicates of L. hesperus density counts, using sweep counts, in over 500 Pima or Acala cotton fields in 1997–2008 in the San Joaquin Valley of California. This data is described more fully in [3]. We selected subsets to analyze using the following criteria:

– In each replicate (corresponding to data collected during one season at one field) we considered data that was collected by pest control advisors (PCAs) between June 1 and August 30.
– We considered data which had the pesticide applications that targeted beet armyworms, aphids, mites as well as Lygus.
– We used only replicates where adult and nymph counts were combined into a total insect count.
– No counts were made on the days of pesticide applications.
– Superposition of pesticide applications has not been incorporated in the algorithm, so we chose samples with at least a week gap between consecutive pesticide applications.

We consider inverse or parameter estimation problems in the context of a parameterized (with vector parameter $q \in \Omega^{\kappa_q} \subset \mathbb{R}^{\kappa_q}$) N-dimensional vector

dynamical system or **mathematical model** given by

$$\frac{d\boldsymbol{x}}{dt}(t) = \boldsymbol{g}(t, \boldsymbol{x}(t), \boldsymbol{q}), \tag{1}$$

$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0, \tag{2}$$

with <u>scalar</u> **observation process**

$$f(t; \boldsymbol{\theta}) = \mathcal{C}\boldsymbol{x}(t; \boldsymbol{\theta}), \tag{3}$$

where $\boldsymbol{\theta} = (\boldsymbol{q}^{\mathsf{T}}, \tilde{\boldsymbol{x}}_0^{\mathsf{T}})^{\mathsf{T}} \in \boldsymbol{\Omega}^{\kappa_{\boldsymbol{\theta}}} \subset \mathbb{R}^{\kappa_{\boldsymbol{q}} + \tilde{N}} = \mathbb{R}^{\kappa_{\boldsymbol{\theta}}}, \tilde{N} \leq N$, and the observation operator $\mathcal{C}$ maps $\mathbb{R}^N$ to $\mathbb{R}^1$. The sets $\boldsymbol{\Omega}^{\kappa_{\boldsymbol{q}}}$ and $\boldsymbol{\Omega}^{\kappa_{\boldsymbol{\theta}}}$ are assumed known restraint sets for the parameters.

We make some standard statistical assumptions (see [7,8,16,24]) underlying our inverse problem formulations.

– (A1) Assume $\mathcal{E}_i$ are independent identically distributed *i.i.d.* with $\mathbb{E}(\mathcal{E}_i) = 0$ and $\mathrm{cov}(\mathcal{E}_i, \mathcal{E}_i) = \sigma_0^2$, where $i = 1, ..., n$ and n is the number of observations or data points in the given data set taken from a time interval $[0, T]$.
– (A2) Assume that there exists a *true or nominal set* of parameters $\boldsymbol{\theta}_0 \in \boldsymbol{\Omega} \equiv \boldsymbol{\Omega}^{\kappa_{\boldsymbol{\theta}}}$.
– (A3) $\boldsymbol{\Omega}$ is a compact subset of Euclidian space of $\mathbb{R}^{\kappa_{\boldsymbol{\theta}}}$ and $f(t, \boldsymbol{\theta})$ is continuous on $[0, T] \times \boldsymbol{\Omega}$.

Denote as $\hat{\boldsymbol{\theta}}$ the estimated parameter for $\boldsymbol{\theta}_0 \in \boldsymbol{\Omega}$. The inverse problem is based on statistical assumptions on the observation error in the data. If one assumes some type of *generalized relative error data model*, then the error is proportional in some sense to the measured observation. This can be represented by a **statistical model** with observations of the form

$$\boldsymbol{Y}_i = f(t_i; \boldsymbol{\theta}_0) + f(t_i; \boldsymbol{\theta}_0)^{\gamma} \mathcal{E}_i, \quad \gamma \in [0, 2], \tag{4}$$

with corresponding realizations

$$y_i = f(t_i; \boldsymbol{\theta}_0) + f(t_i; \boldsymbol{\theta}_0)^{\gamma} \epsilon_i, \quad \gamma \in [0, 2], \tag{5}$$

where the ϵ_i are realizations of the $\mathcal{E}_i$, $i = 1, ..., n$.

For relative error models one should use inverse problem formulations with *Generalized Least Squares (GLS)* cost functional

$$J^n(\boldsymbol{Y}; \boldsymbol{\theta}) = \sum_{i=1}^{n} \left(\frac{\boldsymbol{Y}_i - f(t_i; \boldsymbol{\theta})}{f(t_i; \boldsymbol{\theta})^{\gamma}} \right)^2. \tag{6}$$

The corresponding estimator and estimates are respectively defined by

$$\boldsymbol{\Theta}_{GLS} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Omega}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(\frac{\boldsymbol{Y}_i - f(t_i; \boldsymbol{\theta})}{f(t_i; \boldsymbol{\theta})^{\gamma}} \right)^2, \quad \gamma \in [0, 2], \tag{7}$$

with realizations

$$\hat{\boldsymbol{\theta}}_{GLS} = \operatorname*{argmin}_{\boldsymbol{\theta} \in \boldsymbol{\Omega}} \sum_{i=1}^{n} \left(\frac{y_i - f(t_i; \boldsymbol{\theta})}{f(t_i; \boldsymbol{\theta})^{\gamma}} \right)^2, \quad \gamma \in [0, 2]. \tag{8}$$

GLS estimates $\hat{\boldsymbol{\theta}}^n$ and weights $\{\omega_j\}_{j=1}^n$ are found using an iterative method as defined below (see [7]). For the sake of notation, we will suppress the superscript n (i.e., $\hat{\boldsymbol{\theta}}_{GLS} := \hat{\boldsymbol{\theta}}^n_{GLS}$).

1. Estimate $\hat{\boldsymbol{\theta}}_{GLS}$ by $\hat{\boldsymbol{\theta}}^{(0)}$ using the OLS method ((8) with $\gamma = 0$). Set $k = 0$.
2. Compute weights $\hat{\omega}_j = f^{-2\gamma}(t_j, \hat{\boldsymbol{\theta}}^{(k)})$.
3. Obtain the $k + 1$ estimate for $\hat{\boldsymbol{\theta}}_{GLS}$ by $\hat{\boldsymbol{\theta}}^{(k+1)} := \operatorname{argmin} \sum_{j=1}^n \hat{\omega}_j [y_j - f(t_j, \boldsymbol{\theta})]^2$.
4. Set $k := k + 1$ and return to step 2. Terminate when the two successive estimates for $\hat{\boldsymbol{\theta}}_{GLS}$ are sufficiently close.

3   Mathematical Models

Our focus here is on the comparison of two different models for insect (*L. hesperus*) population growth/mortality in pesticide-treated fields. The simplest model (which we denote as model B) is for constant reduced growth due to effects of pesticides versus an added time-varying mortality (denoted by model A) to reflect this decreased total population growth rate. Model B is given by

$$\frac{dx}{dt} = \eta x$$
$$x(t_1) = x_0, \tag{9}$$

where x_0 is defined as initial population count at time t_1 of initial observation and η is the reduced population growth rate in the presence of pesticides.

Model A is given by

$$\frac{dx}{dt} = k(t)x$$
$$x(t_1) = x_0, \tag{10}$$

where t_1 is again the time of the first data point, and $k(t)$ is a time dependent growth rate

$$k(t) = \begin{cases} \eta + p(t) & t \in P_j, j \in \{1, 2, 3, 4\} \\ \eta & \text{otherwise.} \end{cases}$$

Here $p(t)$ is composed of piecewise linear splines as described below, and $P_j = [t_{p_j}, t_{p_j} + 1/4], j = 1, \ldots j^*$ with t_{p_j} as the time point of the jth pesticide application. Observe that these t_{p_j} are *not* the same as the observation or data points t_j. Also note that $|P_j| = 1/4$ which is approximately the length of time of one week when t is measured in months. This reflects the general assumption that pesticides are most active during the 7 days immediately following treatment. Clearly, η is a reduced **constant** growth rate of the total population in the

presence of pesticides. In addition, t = 0 refers to June 1 (as no data is present before June 1 in our database).

Piecewise linear splines [25] were used to approximate $p(t)$ as follows. Consider m linear splines

$$p(t) = \sum_{i=1}^{m} \lambda_i l_i(t),$$

where

$$l_i(t) = 1/h \begin{cases} t - t_{i-1} & t_{i-1} \le t < t_i \\ t_{i+1} - t & t_i \le t \le t_{i+1} \\ 0 & \text{otherwise}, \end{cases}$$

where h is the step size, $h = \frac{|P_j|}{(m+1)}$. Piecewise linear spline representations are simple, yet flexible in that they allow the modeler to avoid assuming a certain shape to the curve being approximated. Incorporating a time-dependent component such as $p(t)$ is useful when modeling a system with discontinuous perturbations (such as the removal of a predator, or the application of an insecticide). The addition of more splines ($m > 3$) provides a finer approximation, but demands more terms in the parameter estimates. We assume here that it is likely that $m = 3$ is sufficient. (Our subsequent findings suggest that perhaps even $m = 2$ is sufficient!). In our analysis, we first estimated the initial condition x_0 using model B (as this data point precedes any pesticide applications and provides a good estimate for x_0), and then fixed this parameter in all subsequent parameter estimates. Therefore, the parameters to be estimated in model A are $\boldsymbol{\theta} = \boldsymbol{q} = \{\eta, \lambda_1, \lambda_2, \lambda_3\}$ whereas for model B we must only estimate $\boldsymbol{\theta} = \boldsymbol{q} = \{\eta, 0, 0, 0\}$ since model A reduces to model B when applying the constraint $p(t) \equiv 0$, i.e. $\lambda_i = 0$ for $i = 1, 2, 3$.

4 Parameter Estimation

Using the model information provided in [3] we try to estimate parameters for new data sets and determine whether the fit-to-data provided by model A does provide a statistically significantly better fit than the fit provided by model B. A big part of the parameter estimation process is the minimization of the respective cost functions for both model A and model B. The constrained nonlinear optimization solver in Matlab, fmincon was initially being used for minimization of GLS cost functionals in model A while fminsearch was being used for minimization of cost functionals for model B. We later switched to lsqnonlin which gave faster and better results. Since model A is stiff in nature, Matlab solver ode15s was used whereas for model B ode45 was used. Both fmincon and fminsearch require an initial guess of parameters. While for model B fminsearch was able to find a minimum fairly quickly, the initial guess of $\boldsymbol{\theta} = \{\eta, \lambda_1, \lambda_2, \lambda_3\}$ for model A involved a fairly detailed process given below:

1. Create a trial file of data selected based on a specific set of rules.
2. Choose a parameter space $\boldsymbol{\Omega} = [\epsilon, K] \times [-K, -\epsilon]^3$.

3. Choose a constant $\gamma \in [0, 1.5]$. (Note that best way to choose (see [7,8]) gamma is to consider the plot of residuals using both residual vs time and residual vs model plots to ascertain whether the scatter of the error appears to violate the statistical assumption of being *i.i.d.*).
4. Choose an initial condition x_0 to the exponential model by considering the plot of the data and model and visually estimating x_0. We observe that this did not produce acceptable results so we eventually solved a separate inverse problem to estimate intrinsic growth rate ignoring the effects of pesticides and initial condition.

5 Model Comparsion: Nested Restraint Sets

Here we summarize the use of *statistically based model comparison tests*. These residual sum of squares model comparison tests as developed in [5], described in [7,8] and extended in [9] to GLS problems is used in the same manner as used in [1,3]. This test is used to determine which of several nested models is the best fit to the data; therefore, this test can be applied to the comparison of models A and B. In these examples below we are interested in questions related to whether the data will support a more detailed or sophisticated model to describe it. In the next section we recall the fundamental statistical tests to be employed here.

5.1 Statistical Comparison Tests

In general, assume we have an inverse problem for the model observations $f(t, \boldsymbol{\theta})$ and are given n observations. As in (6), we define

$$J^n(\boldsymbol{Y}; \boldsymbol{\theta}) = \sum_{i=1}^{n} \left(\frac{\boldsymbol{Y}_i - f(t_i; \boldsymbol{\theta})}{f(t_i; \boldsymbol{\theta})^\gamma} \right)^2,$$

where our statistical model has the form (4). Here, as before, $\boldsymbol{\theta}_0$, is the nominal value of $\boldsymbol{\theta}$ which we assume to exist. We use $\boldsymbol{\Omega}$ to represent the set of all the admissible parameters $\boldsymbol{\theta}$. We make some further assumptions.

– (A4) Observations are taken at $\{t_j\}_{j=1}^n$ in $[0, T]$. There exists some finite measure μ on $[0, T]$ such that

$$\frac{1}{n} \sum_{j=1}^{n} h(t_j) \longrightarrow \int_0^T h(t) d\mu(t)$$

as $n \to \infty$, for all continuous functions h.
– (A5) $J_0(\boldsymbol{\theta}) = \int_0^T (f(t; \boldsymbol{\theta}_0) - f(t; \boldsymbol{\theta}))^2 d\mu(t) = \sigma^2$ has a unique minimizer in $\boldsymbol{\Omega}$ at $\boldsymbol{\theta}_0$.

Let $\boldsymbol{\Theta}^n = \boldsymbol{\Theta}_{GLS}^n(\boldsymbol{Y})$ be the GLS estimator for J^n as defined in (7) so that

$$\boldsymbol{\Theta}_{GLS}^n(\boldsymbol{Y}) = \operatorname*{argmin}_{\boldsymbol{\theta} \in \boldsymbol{\Omega}} J^n(\boldsymbol{Y}; \boldsymbol{\theta})$$

and
$$\hat{\boldsymbol{\theta}}_{GLS}^n = \underset{\boldsymbol{\theta} \in \boldsymbol{\Omega}}{\operatorname{argmin}} \, J^n(\boldsymbol{y}; \boldsymbol{\theta}),$$
where as above $\boldsymbol{y}$ is a realization for $\boldsymbol{Y}$.

One can then establish a series of useful results (see [5,7,9] for detailed proofs).

Result 1: Under (A1) to (A5), $\frac{1}{n}\boldsymbol{\Theta}^n = \frac{1}{n}\boldsymbol{\Theta}_{GLS}^n(\boldsymbol{Y}) \longrightarrow \boldsymbol{\theta}_0$ as $n \to \infty$ with probability 1.

We will need further assumptions to proceed (these will be denoted by (A7)–(A11) to facilitate reference to [5,7]). These include:

- (A7) $\boldsymbol{\Omega}$ is finite dimensional in R^p and $\boldsymbol{\theta}_0 \in \mathrm{int}\boldsymbol{\Omega}$.
- (A8) $f : \boldsymbol{\Omega} \to C[0,T]$ is a C^2 function.
- (A10) $\mathcal{J} = \frac{\partial^2 J_0}{\partial \boldsymbol{\theta}^2}(\boldsymbol{\theta}_0)$ is positive definite.
- (A11) $\boldsymbol{\Omega}_H = \{\boldsymbol{\theta} \in \boldsymbol{\Omega} | H\boldsymbol{\theta} = c\}$ where H is an $r \times p$ matrix of full rank, and c is a known constant.

In many instances, including the examples discussed here, one is interested in using data to question whether the "nominal" parameter $\boldsymbol{\theta}_0$ can be found in a subset $\boldsymbol{\Omega}_H \subset \boldsymbol{\Omega}$ which we assume for discussions here is defined by the constraints of assumption (A11). Thus, we want to test the *null hypothesis* H_0: $\boldsymbol{\theta}_0 \in \boldsymbol{\Omega}_H$, i.e., that the constrained model provides an adequate fit to the data.

Define then
$$\boldsymbol{\Theta}_H^n(\boldsymbol{Y}) = \underset{\boldsymbol{\theta} \in \boldsymbol{\Omega}_H}{\operatorname{argmin}} \, J^n(\boldsymbol{Y}; \boldsymbol{\theta})$$
and
$$\hat{\boldsymbol{\theta}}_H^n = \underset{\boldsymbol{\theta} \in \boldsymbol{\Omega}_H}{\operatorname{argmin}} \, J^n(\boldsymbol{y}; \boldsymbol{\theta}).$$

Observe that $J^n(\boldsymbol{y}; \hat{\boldsymbol{\theta}}_H^n) \geq J^n(\boldsymbol{y}; \hat{\boldsymbol{\theta}}^n)$. We define the related non-negative test statistics and their realizations, respectively, by
$$T_n(\boldsymbol{Y}) = J^n(\boldsymbol{Y}; \boldsymbol{\theta}_H^n) - J^n(\boldsymbol{Y}; \boldsymbol{\theta}^n)$$
and
$$\hat{T}_n = T_n(\boldsymbol{y}) = J^n(\boldsymbol{y}; \hat{\boldsymbol{\theta}}_H^n) - J^n(\boldsymbol{y}; \hat{\boldsymbol{\theta}}^n).$$

One can establish asymptotic convergence results for the test statistics $T_n(\boldsymbol{Y})$–see [5]. These results can, in turn, be used to establish a fundamental result about much more useful statistics for model comparison. We define these statistics by
$$U_n(\boldsymbol{Y}) = \frac{nT_n(\boldsymbol{Y})}{J^n(\boldsymbol{Y}; \boldsymbol{\theta}^n)}, \tag{11}$$
with corresponding realizations
$$\hat{u}_n = U_n(\boldsymbol{y}).$$

We then have the asymptotic result that is the basis of our analysis-of-variance–type tests.

Results 2: Under the assumptions (A1)–(A5) and (A7)–(A11) above and assuming the null hypothesis H_0 is true, then $U_n(\mathbf{Y})$ converges in distribution (as $n \to \infty$) to a random variable $U(r)$, i.e.,

$$U_n \xrightarrow{\mathcal{D}} U(r),$$

with $U(r)$ having a chi-square distribution $\chi^2(r)$ with r degrees of freedom.

In any graph of a χ^2 density there are two parameters (τ, α) of interest. For a given value τ, the value α is simply the probability that the random variable U will take on a value greater than τ. That is, Prob$\{U > \tau\} = \alpha$ where in hypothesis testing, α is the *significance level* and τ is the *threshold*.

We then wish to use this distribution $U_n \sim \chi^2(r)$ to test the null hypothesis, H_0, that the restricted model provides an adequate fit to represent the data. If the test statistic, $\hat{u}_n > \tau$, then we *reject H_0 as false* with confidence level $(1 - \alpha)100\%$. Otherwise, we *do not reject H_0*. For our examples below, we use a $\chi^2(3)$ table, which can be found in any elementary statistics text, online or the partial summary below. Typical confidence levels of interest are $90\%, 95\%, 99\%, 99.9\%$, with corresponding (α, τ) values given in Table 1 below.

To test the null hypothesis H_0, we choose a significance level α and use χ^2 tables to obtain the corresponding threshold $\tau = \tau(\alpha)$ so that Prob$\{\chi^2(r) > \tau\} = \alpha$. We next compute $\hat{u}_n = \overline{\tau}$ and compare it to τ. If $\hat{u}_n > \tau$, then we reject H_0 as false; otherwise, we do not reject the null hypothesis H_0.

We use a $\chi^2(3)$ for our comparison tests as summarized here.

Table 1. Chi-square table: $\chi^2(3)$

α	.10	.05	.01	.001
τ	6.251	7.815	11.345	16.266

We can then formulate the null and alternative hypotheses:

H$_0$: The fit provided by model A *does not* provide a statistically significantly better fit to the data than the fit provided by model B.

H$_A$: The fit provided by model A *does* provide a statistically significantly better fit to the data than the fit provided by model B.

We considered such comparison tests for a number of data sets with a varying no. of pesticides applications among the fields. These included

– Replicate number 296 (1 pesticide application at t $=$ 0.5 months)
– Replicate number 350 (2 pesticide applications at times t $=$ 1 and t $=$ 1.7 months)
– Replicate number 277 (3 pesticide applications at t $=$ .4, .7, 1.9 months)
– Replicate number 178 (4 pesticide applications at t $=$ .13, .77, 1.9, 2.27 months)
– Replicate number 174 (4 pesticide applications at t $=$ .13, .77, 1.83, 2.27 months)

We carried out multiple inverse problems with varying values of γ for these data sets (see [2]). We visually examined the resulting residual plots (residual vs time and residual vs observed output) and determined whether the scatter of the error appear to be i.i.d. On examining the plots for a wide range of γ we observed that the statistical i.i.d. assumptions were approximately satisfied for γ values ranging around 0.7 to 0.8. We therefore used these values of γ in the results reported here and in [2].

We note here that one could, as an alternative to use of residual plots, include the parameter γ as a parameter to be estimated along with θ as often done in statistical formulations [16] for the joint estimation of $\beta = (\theta, \gamma)$, or one could attempt to estimate the form of the statistical model (i.e., the value of γ directly from the data itself as suggested in [6]. Both of these methods offer some advantages but include more complex inverse problem analysis. We have therefore chosen here to use the simpler but less sophisticated analysis of residuals in our approach.

6 Results

We examine the importance of model selection and how optimal selection of both statistical and mathematical models is crucial for accuracy in parameter estimation and fitting data to models testing different data sets from the same database with varied number of pesticide applications. Tables 2, 3 and 4 contain summaries of the results for the investigated replicates.

Table 2. Parameter estimates for models A and B

Replicate no.	No. of pest. apps	Par. Est. for model A $(\eta, \lambda_1, \lambda_2, \lambda_3)$	Par. Est. for model B $(\eta, \lambda_1, \lambda_2, \lambda_3)$
296 (n = 19)	1	$(1.3551, -7.8782, -7.1632, -9.9904)$	$(0.3548, 0, 0, 0)$
350 (n = 19)	2	$(3.0150, -16.7393, -17.9356, -8.580)$	$(0.7683, 0, 0, 0)$
277 (n = 19)	3	$(3.1255, -3.5165, -6.2870, -7.1868)$	$(1.7284, 0, 0, 0)$
178 (n = 22)	4	$(1.8201, -6.6352, -2.5786, -4.1605)$	$(0.6335, 0, 0, 0)$
174 (n = 20)	4	$(2.8600, -7.9867, -13.2780, -4.6992)$	$(0.4110, 0, 0, 0)$

Table 3. Cost functional values for models A and B

Replicate no.	Cost functional model A	Cost functional model B	Range of γ values	x_0
296	3.9577	7.3663	0.5 to 1.0	0.39
350	7.4100	18.223	0.5 to 0.8	0.28
277	5.5290	8.075	0.7 to 0.9	0.05
178	5.7596	8.371	0.7 to 1.0	0.37
174	6.7480	9.390	0.0 to 0.8	0.34

Table 4. Statistical comparison test results for degrees of freedom $r = 3$

Replicate no.	Test statistic $\hat{u}_n$	Threshold τ	Signif. level α	Null hyp. (H_0) accept or reject	Confd. level
296	16.36	16.26	0.001	Rejected	99.9%
350	27.73	16.26	0.001	Rejected	99.9%
277	8.7458	7.815	0.05	Rejected	95%
178	9.9715	7.815	0.05	Rejected	95%
174	7.8281	7.815	0.05	Rejected	95%

7   Concluding Remarks

The above results strongly support the notion that time varying reduced growth/mortality rates as opposed to constant rates provide substantially better models at the population levels for the description of the effects of pesticides on the growth rates. It is interesting to note that our findings hold consistently across the differing levels of pesticide applications, even in the case of only one pesticide application. Another interesting observation is that this is consistent even when one uses total pest counts as opposed to individual nymph and adult counts. This, of course, has significant implications for data collection procedures. The model comparison techniques employed here are just one of several tools that one can use to determine aspects of information content in support of model sophistication/complexity. Of note are the use of the Akiake Information Criterion (AIC) and its variations [1,4,10,11,14,15] for model comparison.

References

1. Banks, H.T., Banks, J.E., Link, K., Rosenheim, J.A., Ross, C., Tillman, K.A.: Model comparison tests to determine data information content. Appl. Math. Lett. **43**, 10–18 (2015). CRSC-TR14-13, North Carolina State University, Raleigh, NC, October 2014
2. Banks, H.T., Banks, J.E., Murad, N., Rosenheim, J.A., Tillman, K.: Modelling pesticide treatment effects on Lygus hesperus in cotton fields. CRSC-TR15-09, Center for Research in Scientific Computation, North Carolina State University, Raleigh, NC, September 2015
3. Banks, H.T., Banks, J.E., Rosenheim, J., Tillman, K.: Modeling populations of Lygus hesperus on cotton fields in the San Joaquin Valley of California: the importance of statistical and mathematical model choice. J. Biol. Dyn. **11**, 25–39 (2017)
4. Banks, H.T., Doumic, M., Kruse, C., Prigent, S., Rezaei, H.: Information content in data sets for a nucleated-polymerization model. J. Biol. Dyn. **9**, 172–197 (2015). doi:10.1080/17513758.2015.1050465

5. Banks, H.T., Fitzpatrick, B.G.: Statistical methods for model comparison in parameter estimation problems for distributed systems. J. Math. Biol. **28**, 501–527 (1990)
6. Banks, H.T., Catenacci, J., Hu, S.: Applications of difference-based variance estimation methods. J. Inverse Ill-Posed Prob. **24**, 413–433 (2016)
7. Banks, H.T., Hu, S., Thompson, W.C.: Modeling and Inverse Problems in the Presence of Uncertainty. Taylor/Francis-Chapman/Hall-CRC Press, Boca Raton (2014)
8. Banks, H.T., Tran, H.T.: Mathematical and Experimental Modeling of Physical and Biological Processes. CRC Press, Boca Raton (2009)
9. Banks, H.T., Kenz, Z.R., Thompson, W.C.: An extension of RSS-based model comparison tests for weighted least squares. Int. J. Pure Appl. Math. **79**, 155–183 (2012)
10. Bozdogan, H.: Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. Psychometrika **52**, 345–370 (1987)
11. Bozdogan, H.: Akaike's information criterion and recent developments in information complexity. J. Math. Psychol. **44**, 62–91 (2000)
12. Brabec, M., Honěk, A., Pekár, S., Martinková, Z.: Population dynamics of aphids on cereals: digging in the time-series data to reveal population regulation caused by temperature. PLoS ONE **9**, e106228 (2014)
13. Brawn, J.D., Robinson, S.K., Thompson, F.R.: The role of disturbance in the ecology and conservation of birds. Annu. Rev. Ecol. Syst. **32**(2001), 251–276 (2001)
14. Burnham, K.P., Anderson, D.R.: Model Selection and Inference: A Practical Information-Theoretical Approach, 2nd edn. Springer, New York (2002)
15. Burnham, K.P., Anderson, D.R.: Multimodel inference: understanding AIC and BIC in model selection. Sociol. Methods Res. **33**, 261–304 (2004)
16. Davidian, M., Giltinan, D.M.: Nonlinear Models for Repeated Measurement Data. Chapman and Hall, London (2000)
17. Forbes, V.E., Calow, P.: Is the per capita rate of increase a good measure of population-level effects in ecotoxicology? Environ. Toxicol. Chem. **18**, 1544–1556 (1999)
18. Forbes, V.E., Calow, P., Sibly, R.M.: The extrapolation problem and how population modeling can help. Environ. Toxicol. Chem. **27**, 1987–1994 (2008)
19. Forbes, V.E., Calow, P., Grimm, V., Hayashi, T., Jager, T., Palmpvist, A., Pastorok, R., Salvito, D., Sibly, R., Spromberg, J., Stark, J., Stillman, R.A.: Integrating population modeling into ecological risk assessment. Integr. Environ. Assess. Manag. **6**, 191–193 (2010)
20. Hilborn, R.: The Ecological Detective: Confronting Models with Data, vol. 28. Princeton University Press, Princeton (1997)
21. Motulsky, H., Christopoulos, A.: Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting. Oxford University Press, Oxford (2004)
22. Pickett, S., White, P.: Natural disturbance and patch dynamics: an introduction. In: Pickett, S.T.A., White, P.S. (eds.) The Ecology of Natural Disturbance and Patch Dynamics, pp. 3–13. Academic Press, Orlando (1985)
23. Rosenheim, J.A., Steinmann, K., Langelloto, G.A., Link, A.G.: Estimating the impact of Lygus hesperus on cotton: the insect, plant and human observer as sources of variability. Environ. Entomol. **35**, 1141–1153 (2006)
24. Seber, G.A.F., Wild, C.J.: Nonlinear Regression. Wiley, Hoboken (2003)
25. Schultz, M.H.: Spline Analysis. Prentice-Hall, Englewood Cliffs (1973)

26. Stark, J.D., Banks, J.E.: Population-level effects of pesticides and other toxicants on arthropods. Annu. Rev. Entomol. **48**, 505–519 (2003)
27. Stark, J.D., Banks, J.E., Vargas, R.I.: How risky is risk assessment? The role that life history strategies play in susceptibility of species to pesticides and other toxicants. Proc. Nat. Acad. Sci. **101**, 732–736 (2004)
28. Stark, J.D., Vargas, R., Banks, J.E.: Incorporating ecologically relevant measures of pesticide effect for estimating the compatibility of pesticides and biocontrol agents. J. Econ. Entomol. **100**, 1027–1032 (2007)
29. Stark, J.D., Vargas, R., Banks, J.E.: Incorporating variability in point estimates in risk assessment: bridging the gap between LC50 and population endpoints. Environ. Toxicol. Chem. **34**, 1683–1688 (2015)