

# Delay differential population models for the decline of *Homalodisca vitripennis* (Hemiptera: Cicadellidae) densities over a ten-year period

Annabel E. Meade<sup>1</sup>, H. T. Banks<sup>1</sup>, John E. Banks<sup>3</sup>, Natalie G. Cody<sup>1</sup>, and Mark S. Hoddle<sup>2</sup>

<sup>1</sup>Center for Research in Scientific Computation, N. C. State University, Raleigh, NC

<sup>2</sup>Department of Entomology, University of California, Riverside, Riverside, CA

<sup>3</sup>Undergraduate Research Opportunities Center, California State University, Monterey Bay, Seaside, CA

September 14, 2018

## Abstract

The glassy-winged sharpshooter, *Homalodisca vitripennis*, is an invasive pest which presents a major economic threat to grape industries in California, as well as Texas [22] and other wine growing regions, because it spreads a disease-causing bacterium, *Xylella fastidiosa*. We continue an earlier investigation [1] into a long-term phenological decline of *H. vitripennis* densities by studying a system of delayed differential equations (DDEs) and analyzing aggregate population data for *H. vitripennis* from a 10-year study in which bi-weekly monitoring of *H. vitripennis* populations decreased significantly. These data present several challenges for modelers. First, they involve truly aggregate population level sampling and hence cannot properly be treated as ordinary longitudinal time series data corresponding to individual level models. The appropriate modeling involves estimation of probability distributions for parameters rather than estimation of the dynamic parameters themselves. Moreover, our analysis reveals that the correct corresponding statistical models involve errors that are observation size dependent (e.g., relative errors should be employed in statistical models). We use these data to test whether DDEs are useful in modeling the observed *H. vitripennis* population decline. To do this, we perform an analysis of variance (ANOVA) type test comparing the glassy-winged sharpshooter (GWSS) model with delay to the model without delay [1]. The model is fit to the aggregate *H. vitripennis* data using iterative reweighted weighted least squares (IRWLS) by estimating probability densities over the delay and one of the birthrate parameters. Results indicate that a positive delay provides improvement with a significance level of  $p < 0.005$ .

## 1 Introduction

The glassy-winged sharpshooter (GWSS), *Homalodisca vitripennis* (Germar) (Hemiptera: Cicadellidae), is a xylem-feeding leafhopper native to the southern United States from Florida to Texas and northeastern Mexico [18]. *H. vitripennis* has a large degree of agricultural importance, because it vectors a xylem-dwelling bacterium, *Xylella fastidiosa*, which causes a variety of often lethal plant diseases (e.g., Pierce's disease of grapes) [19, 20].

*H. vitripennis* exhibits high invasion potential and is a significant pest outside of its native range, including California [20], Hawaii, the Cook Islands, Easter Island, and French Polynesia [13, 14]. In particular, after invading California in 1990, *H. vitripennis* has represented a major economic threat to the wine, table, and raisin grape industries, because it was implicated in a significant increase of the lethal and incurable grape malady, Pierce's disease [18, 20]. This disease costs approximately \$104 million per year in crop damage and resources devoted to mitigating this threat [21].

From March 2002 to August 2012, a long-term monitoring program in citrus, a preferred host plant for *H. vitripennis*, was undertaken in order to study *H. vitripennis* densities in California. Every two weeks ten individual lemon trees in an unsprayed citrus orchard were surveyed for *H. vitripennis* adults and eggs. This large data set consisting of 271 observation dates clearly showed a decline in *H. vitripennis* populations over time. The mymarid egg parasitoid *Cosmocomoidea* (formerly *Gonatocerus*) *ashmeadi* (Girault), which targets *H. vitripennis* exclusively, was found in GWSS eggs collected during the study [1], but the biotic factor of parasitism is not analyzed in this project. The goal of this project is to determine whether abiotic factors such as temperature were responsible for observed population trends (i.e., density independent population regulation). The reproductive and developmental biology of *H. vitripennis* is strongly influence by environmental factors, especially temperature [20]. In an earlier investigation [1] we developed a system of ordinary differential equations (ODEs) to investigate the cause of this observed decrease in *H. vitripennis* populations over time. In this investigation, we present a system of delayed differential equations (DDEs) and test whether this the added delay improves the model tells us more about the GWSS population in the data set.

## 2 Aggregate Data Sets

Biweekly surveys were made of ten Eureka lemon trees in an unsprayed mixed variety citrus plot at the Agricultural Operations Facility at the University of California Riverside (UCR Ag. Ops) from March 13, 2002 to August 8, 2012. Field surveys were conducted at approximately 12:00 to 1:00 PM for each sample date. The north, south, west, and east quadrants of the experimental trees were visually searched for 30 seconds each, and the total number of adult *H. vitripennis* infesting each citrus tree was recorded. After adult counts were completed, each tree was searched for five minutes and all *H. vitripennis* egg masses that were found were removed from trees and returned to the lab where the total number of individual eggs collected on each sampling date were recorded. These egg and adult *H. vitripennis* data can be found in Figure 1A and 1B. Egg masses were also observed over a three week period for the emergence of parasitoids, but these data are not analyzed in this paper. For more information about the parasitoid data collected, see Banks, Banks, Cody, Hoddle, and Meade [1].

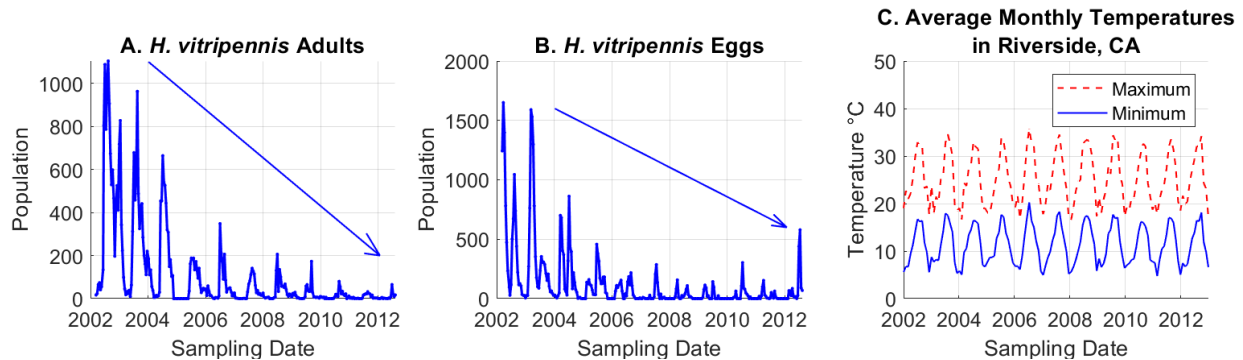


Figure 1: Population phenology data for adult *H. vitripennis* (A) and eggs (B), and (C) the minimum and maximum average monthly temperatures in Riverside, California in citrus at Agricultural Operations, University of California, Riverside.

The reproductive and developmental biology of *H. vitripennis* depends on climatic conditions, especially temperature [20], so temperature is an important part of the GWSS model. The daily minimum and maximum temperatures from January 1, 2002 to December 31, 2012 in Riverside, CA were downloaded from an automated weather station, CIMIS 44, located at UCR Ag. Ops and used to calculate monthly averages (Figure 1D). For the mathematical simulations, we assume a daily sinusoidal curve in temperature fluctuation in which the daily maximum temperature occurs at approximately 4:00 PM, and the daily minimum temperature occurs at approximately 4:00 AM.

*Homalodiscs vitripennis* are strong fliers and can travel long distances to find optimal host plants [17].

Thus, it is reasonable to assume that each data point is a different subsample of the total population, making this an *aggregate* [7] data set (i.e., each sample may possibly consist of differing individuals from the overall population). Because of this, we assume that *H. vitripennis* eggs and adults from the data set have been sampled longitudinally from the aggregate population. Therefore, parameters regulating population fluctuation are likely described by some probability distribution [7] instead of a single value. This is vital and changes the inverse problem when comparing aggregate population data to a mathematical model meant to describe an individual population.

### 3 Mathematical Model: Delayed Differential Equations (DDEs)

To model the individual population of *H. vitripennis*, we use a continuous time age structured population growth model [15] with an added delay. *H. vitripennis* is known to begin life as an egg and undergo five nymphal stages before developing into sexually mature adults. For simplicity, we model the egg and adult stage with a delay to represent the nymphal stage in

$$\frac{dx_e(t)}{dt} = b(T_{t-180})x_A(t) - r(T_{t-180})x_e(t) - d_e x_e(t) \quad (1)$$

$$\frac{dx_A(t)}{dt} = r(T_{t-(\tau+180)})x_e(t-\tau) - d_A x_A(t). \quad (2)$$

where  $t$  is time in days and  $T_t$  is temperature in  $^{\circ}\text{C}$  at time  $t$ . Eggs,  $x_e$ , can die at rate  $d_e$  or hatch at rate  $r(T_{t-180})$ . Parameter  $\tau$  represents the interval from egg emergence to oviposition or the combined nymphal time and preoviposition (i.e., interval from adult emergence to the first oviposition) time. Thus, hatched eggs become adults after  $\tau$  days at rate  $r(T_{t-(\tau+180)})$ , and adults,  $x_A$ , lay more eggs at rate  $b(T_{t-180})$  and die at rate  $d_A$ .

The temperature dependent developmental rate is defined by the Lactin model [16],

$$r(T) = e^{\rho T} - e^{(\rho T_u - [T_u - T]/\Delta)} + \lambda. \quad (3)$$

We let the birthrate  $b(T) = R_0(T)/T_c(T)$  from Pilkington, Lewis, Jeske, and Hoddle [20], so that

$$b(T) = \frac{R_0(T)}{T_c(T)} = \frac{a_0 + a_1 T + a_2 T^2}{b_0 + b_1 T + b_2 T^2} \quad (4)$$

is the average number of offspring produced per female per day at temperature  $T$ . The parameter values of these temperature-dependent rates are taken from Pilkington, et al. [20] and listed in Table 1.

Temperature-dependent rates are calculated using temperatures from 6 months or 180 days in the past, because without this shift there is a disconnect between the peaks of the model and that of the actual population. We let the maximum delay be  $\tau = 180$ , and we have temperature data beginning on January 1, 2002, so with a  $\tau + 180 = 360$  day shift we cannot model populations earlier than March 8, 2003. Thus, the first 22 dates in the time series cannot be modeled, bringing the total number of usable *H. vitripennis* egg and adult data points from 544 to 522, which is still sufficient for our purposes (usually 30-40 data points [6] are suggested per parameter being estimated). In order to solve the DDEs (1) and (2), we define an initial condition function using the removed data points in

$$x_e(t) = x_{e0}(t) = \sum_{j=-21}^0 x_{e_j} l_j(t), \quad -\tau \leq t \leq 0 \quad (5)$$

$$x_A(t) = x_{A0}(t) = \sum_{j=-21}^0 x_{A_j} l_j(t), \quad -\tau \leq t \leq 0, \quad (6)$$

where  $[x_{e_j}, x_{A_j}]$  are the first 22 egg and adult data points we removed from the data set, and  $l_k(t)$  are piecewise linear splines defined on the corresponding removed time points  $t_j$  for  $j = -21, \dots, 0$ .

Table 1: Parameters and descriptions for equations (1-4). **Bolded parameters are temperature-dependent equations.**

| $\theta$                   | Definition                                      | Chosen Values | Units             | Source      |
|----------------------------|---|---------------|-------------------|-------------|
| <b><math>r(T)</math></b>   | <b>Developmental rate of eggs into adults</b>   | —             | <b>1/day</b>      | <b>[16]</b> |
| $T_u$                      | Temperature when $r(T) = \lambda$               | 47.1          | $^{\circ}C$       | [20]        |
| $\rho$                     | Fitted in Logan Model                           | 0.002         | —                 | [20]        |
| $\Delta$                   | Fitted in Logan Model                           | 3.61          | —                 | [20]        |
| $\lambda$                  | Determines upper temperature threshold          | -1.03         | 1/day             | [20]        |
| <b><math>R_0(T)</math></b> | <b>Number of daughters produced/female</b>      | —             | <b>eggs/adult</b> | <b>[12]</b> |
| $a_0$                      | Quadratic parameter (constant)                  | -701.19       | —                 | [20]        |
| $a_1$                      | Quadratic parameter                             | 57.48         | —                 | [20]        |
| $a_2$                      | Quadratic parameter                             | -1.113        | —                 | [20]        |
| <b><math>T_c(T)</math></b> | <b>Mean generation time</b>                     | —             | <b>day</b>        | <b>[12]</b> |
| $b_0$                      | Quadratic parameter (constant)                  | -1616.2       | —                 | [20]        |
| $b_1$                      | Quadratic parameter                             | 147.92        | —                 | [20]        |
| $b_2$                      | Quadratic parameter                             | -3.0264       | —                 | [20]        |
| $d_e$                      | Death rate of eggs                              | 0.02          | 1/day             | [15]        |
| $d_N$                      | Death rate of nymphs                            | 0.02          | 1/day             | [15]        |
| $d_A$                      | Death rate of adults                            | 0.02          | 1/day             | [15]        |
| $\tau$                     | Time spent in nymphal and preoviposition period | 0-180         | day               | —           |

This model is an individual model, so it is not sufficient to model the aggregate data described in Section 2. Thus, we use the individual model to create an aggregate model by assuming that a probability distribution exists over one or another of the parameters in Table 1. In [1] we compared estimations of various parameters such as  $a_1, a_2$ , etc. Note that if  $\tau = 0$ , the system of DDEs in equation (1) and (2) becomes a system of ordinary differential equations (ODEs). In the following sections, we compare two aggregate versions of this model to the aggregate data: one model in which the delay  $\tau = 0$ , and one in which the delay  $\tau \geq 0$ . For these comparisons we used an estimated distribution for the birthrate parameter  $a_1$  as described below, because in [1] the egg and adult observations are most sensitive to this parameter.

## 4 Aggregate Parameter Estimation Methodology

### 4.1 Aggregate Model

Equations (1) and (2) model individual populations, where all individual glassy-winged sharpshooters are assumed to have the same growth, birth, and death rates. However, this assumption does not apply to the aggregate data which consists of frequent sampling of the *changing populations*. Thus, the aggregate population  $u$  at time  $t$  corresponding to an individual population  $x$  is given by

$$u(t; P) = \int_{\mathcal{G}} x(t; \theta) dP(\theta),$$

where  $x(t; \theta)$  is distributed over parameters  $\theta$ ,  $\mathcal{G}$  is the collection of admissible parameter values, and  $P$  is a probability measure on  $\mathcal{G}$ . Under the assumption that the probability distribution,  $P$ , possesses a density, the population count is given by

$$u(t; P) = \int_{\mathcal{G}} x(t; \theta) p(\theta) d\theta,$$

where the density  $P' = \frac{dP}{d\theta} = p(\theta)$ . We are interested in determining the probability density  $P' = p(\theta)$  which gives the best fit of the underlying model to the aggregate data. However, this parameter estimation problem involves an infinite dimensional parameter space (the space  $\mathcal{P}$  of probability measures). Instead of using a specific probability density function in the aggregate model, we use finite approximations. Based on [3, 4, 6, 9], we are guaranteed convergence in the Prohorov metric. (In a separate publication [9], we have given a rather complete theoretical foundation - existence of estimators in classes of probability distributions,

convergence of approximations, and consistency of estimators - for the techniques we use here.) Thus, the finite approximation  $\mathcal{P}^M(\mathcal{G})$  to the probability measure space  $\mathcal{P}(\mathcal{G})$  can be defined using linear splines and given by

$$\mathcal{P}^M(\mathcal{G}) = \left\{ P \in \mathcal{P}^M(\mathcal{G}) \mid \frac{dP}{d\theta} = p(\theta) = \sum_{k=0}^M \Theta_k^M l_k^M(\theta), \sum_{k=0}^M \Theta_k^M \int_{\mathcal{G}} l_k^M(\theta) d\theta = 1 \right\},$$

where the piecewise linear splines are represented by  $l_k^M$  for  $k = 0, \dots, M$ . We assume  $\mathcal{G} = [\theta_l, \theta_u]$  is a closed interval. Thus, the aggregate populations corresponding to observations  $x_e$  and  $x_A$  are given by

$$u_e(t; \{\Theta_{1k}^M\}) = \sum_{k=0}^M \Theta_{1k}^M \int_{\theta_l}^{\theta_u} x_e(t; \theta) l_k^M(\theta) d\theta, \quad (7)$$

$$u_A(t; \{\Theta_{2k}^M\}) = \sum_{k=0}^M \Theta_{2k}^M \int_{\theta_l}^{\theta_u} x_A(t; \theta) l_k^M(\theta) d\theta, \quad (8)$$

where  $\Theta_{1k}^M$  and  $\Theta_{2k}^M$  for  $k = 0, \dots, M$  are coefficients for the linear splines of the egg and adult populations, respectively.

Similarly, we can assume that the aggregate data is best described when the individual population is distributed over two parameters. This is done by taking the aggregate population model in equations (7) and (8) which is distributed over parameter  $\theta$  and apply the same principles to assume a probability distributed over  $\omega$ . Thus, an aggregate populations distributed over two parameters,  $\theta$  and  $\omega$ , corresponding to  $x_e$  and  $x_A$  are given by

$$u_e^*(t; \{\Theta_{1k}^M\}, \{\Omega_{1k^*}^{M^*}\}) = \sum_{k^*=0}^{M^*} \Omega_{1k^*}^{M^*} \int_{\omega_l}^{\omega_u} \left( \sum_{k=0}^M \Theta_{1k}^M \int_{\theta_l}^{\theta_u} x_e(t; \theta, \omega) l_k^M(\theta) d\theta \right) l_{k^*}^{M^*}(\omega) d\omega, \quad (9)$$

$$u_A^*(t; \{\Theta_{2k}^M\}, \{\Omega_{2k^*}^{M^*}\}) = \sum_{k^*=0}^{M^*} \Omega_{2k^*}^{M^*} \int_{\omega_l}^{\omega_u} \left( \sum_{k=0}^M \Theta_{2k}^M \int_{\theta_l}^{\theta_u} x_A(t; \theta, \omega) l_k^M(\theta) d\theta \right) l_{k^*}^{M^*}(\omega) d\omega, \quad (10)$$

where  $l_{m^*}^{M^*}$  and piecewise linear splines and  $\Omega_{1k^*}^{M^*}$  and  $\Omega_{2k^*}^{M^*}$  are coordinates for the linear splines of the egg and adult populations, respectively, for  $k^* = 0, \dots, M^*$ . The probability densities of the egg and adult populations over parameters  $\theta$  and  $\omega$  can be described using the  $2 \times (M+1)$  and  $2 \times (M^*+1)$  matrices

$$\Theta^M = \begin{bmatrix} \Theta_{10}^M & \Theta_{11}^M & \dots & \Theta_{1M}^M \\ \Theta_{20}^M & \Theta_{21}^M & \dots & \Theta_{2M}^M \end{bmatrix} \quad \text{and} \quad \Omega^{M^*} = \begin{bmatrix} \Omega_{10}^{M^*} & \Omega_{11}^{M^*} & \dots & \Omega_{1M^*}^{M^*} \\ \Omega_{20}^{M^*} & \Omega_{21}^{M^*} & \dots & \Omega_{2M^*}^{M^*} \end{bmatrix}, \quad (11)$$

respectively. All integrals are approximated using the composite trapezoidal rule.

## 4.2 Parameter Estimation Methodology

We follow techniques from Banks, Hu, and Thompson [7] and Banks, Bekele-Maxwell, Everett, Stephenson, Shao, and Morgenstern [2] to estimate parameters in our mathematical model. We perform two inverse problems in order to estimate a probability densities over two parameters described in equations (9) and (10). First, we estimate the probability density over parameter  $\theta$  while holding  $\omega$  fixed. Next, we fix the probability density over  $\theta$  and estimate the probability density over  $\omega$ . The techniques used for both problems are described below.

For the first inverse problem, we choose to estimate a probability density over parameter  $\theta$  and consider a general  $n$ -dimensional dynamical system,

$$\frac{d\vec{x}}{dt}(t) = \vec{g}(t, \vec{x}(t), \theta), \quad (12)$$

$$\vec{x}(t_0) = \vec{x}_0, \quad (13)$$

with an  $m$ -dimensional aggregate observation process  $\vec{f}(t; \Theta^M) = [u_e(t; \{\Theta_{1k}^M\}), u_A(t; \{\Theta_{2k}^M\})]^T$  from equations (7) and (8), since our data set consists of egg and adult aggregate populations.

We only need to estimate the set of  $2(M + 1)$  spline coordinates in the matrix  $\Theta^M$  from equation (11). We note that there is some discrepancy between the actual phenomenon, which is represented through the data, and in the above observation process. This uncertainty is accounted for in the statistical model

$$\underbrace{\begin{bmatrix} U_j^1 \\ U_j^2 \end{bmatrix}}_{\tilde{U}_j} = \underbrace{\begin{bmatrix} f_1(t_j; \Theta_0^M) \\ f_2(t_j; \Theta_0^M) \end{bmatrix}}_{\tilde{U}(t_j; \Theta_0^M)} + \underbrace{\begin{bmatrix} f_1^{\gamma_1}(t_j; \Theta_0^M) \\ f_2^{\gamma_2}(t_j; \Theta_0^M) \end{bmatrix}}_{\text{error}} \circ \begin{bmatrix} \mathcal{E}_j^1 \\ \mathcal{E}_j^2 \end{bmatrix},$$

where  $\vec{\gamma} = [\gamma_1, \gamma_2] \geq 0$  and  $\Theta_0^M$  is the nominal probability density approximation. Here  $f_1(t; \Theta_0^M) = u_e(t; \Theta_0^M)$  and  $f_2(t; \Theta_0^M) = u_A(t; \Theta_0^M)$ . The  $n \times 1$  random error vectors  $\vec{\mathcal{E}}^1$  and  $\vec{\mathcal{E}}^2$  respectively are assumed to be independent and identically distributed (*i.i.d.*) with mean zero and  $\text{Var}(\mathcal{E}_j^1) = \sigma_{01}^2$  and  $\text{Var}(\mathcal{E}_j^2) = \sigma_{02}^2$ . The corresponding realizations are

$$\underbrace{\begin{bmatrix} u_j^1 \\ u_j^2 \end{bmatrix}}_{\tilde{u}_j} = \underbrace{\begin{bmatrix} f_1(t_j; \Theta_0^M) \\ f_2(t_j; \Theta_0^M) \end{bmatrix}}_{\tilde{u}(t_j; \Theta_0^M)} + \underbrace{\begin{bmatrix} f_1^{\gamma_1}(t_j; \Theta_0^M) \\ f_2^{\gamma_2}(t_j; \Theta_0^M) \end{bmatrix}}_{\text{error}} \circ \begin{bmatrix} \epsilon_j^1 \\ \epsilon_j^2 \end{bmatrix}.$$

This multiplicative structure of the observational error in the above statistical model exists because often in biological models the size of the observation error is proportional to the size of the observations. For  $\vec{\gamma} \geq 0$ , a generalized least squares method or an iterative reweighted weighted least squares (IRWLS) method [2] as used below is appropriate to perform the inverse problem. In order to estimate  $\Theta_0^M \approx \Theta^M$ , we want to minimize the distance between the collected data and aggregate mathematical model, where the observables are weighted according to their variability and, for each observable, the observations over time are weighted unequally.

The iterative weighted least squares estimate,  $\hat{\Theta}^M$ , is numerically determined by iteratively solving the following system:

$$\hat{\Theta}^M = \arg \min_{\Theta^M} \sum_{j=1}^n [\tilde{u}_j - \tilde{u}(t_j; \Theta^M)]^T \hat{V}_j^{-1} [\tilde{u}_j - \tilde{u}(t_j; \Theta^M)] \quad (14)$$

$$\hat{V}_j = \text{diag} \left( \frac{1}{n-p} \hat{W}_j \left( \sum_{j=1}^n [\tilde{u}_j - \tilde{u}(t_j; \hat{\Theta}^M)] [\tilde{u}_j - \tilde{u}(t_j; \hat{\Theta}^M)]^T \hat{W}_j^{-1} \right)_{ii} \right) \quad (15)$$

for  $n$  data points. The matrix  $\hat{W}_j = \text{diag}(f_1^{2\gamma_1}(t_j; \hat{\Theta}^M), f_2^{2\gamma_2}(t_j; \hat{\Theta}^M))$  is made up of the error weights, and  $\hat{V}_j$  is the estimated covariance matrix at data point  $j = 1, \dots, n$ . We use the following iterative procedure [7, 8]:

1. Estimate  $\hat{\Theta}^{M(0)}$  using (14) with  $\hat{V}_j = I$ . Set  $l = 0$ .
2. Compute weight matrices  $\hat{W}_j^{(l)} = \text{diag}(f_1^{2\gamma_1}(t_j; \hat{\Theta}^{M(l)}), f_2^{2\gamma_2}(t_j; \hat{\Theta}^{M(l)}))$ .
3. Solve for  $\hat{V}_j^{(l)}$  using  $\hat{\Theta}^{M(l)}$  and  $\hat{W}_j^{(l)}$  in equation (15).
4. Estimate  $\hat{\Theta}^{M(l+1)}$  using  $\hat{V}_j^{(l)}$  in equation (14).
5. Set  $l := l + 1$  and return to step 2. Terminate when two successive estimates for  $\hat{\Theta}^M$  are sufficiently close.

Note that this is not the same as taking the derivative of the argument in the right side of (14) and setting it equal to zero. We iteratively want to minimize

$$J(\hat{\Theta}^M) = \sum_{j=1}^n [\tilde{u}_j - \tilde{u}(t_j; \hat{\Theta}^M)]^T \hat{V}_j^{-1} [\tilde{u}_j - \tilde{u}(t_j; \hat{\Theta}^M)]. \quad (16)$$

The estimated variances for each observable  $\sigma_{01}^2$  and  $\sigma_{02}^2$  are approximated by the following:

$$[\hat{\sigma}_{01}^2, \hat{\sigma}_{02}^2] = \frac{1}{n-1} \sum_{j=1}^n \text{diag}(\hat{V}_j). \quad (17)$$

If we assume  $\vec{\gamma} = [\gamma_1, \gamma_2] = [0, 0]$ , then our statistical model is called an absolute error model and an ordinary least squares method is appropriate for parameter estimation. However, we believe it is more biologically realistic to assume the observation error is proportional to the size of the observed quantity for our data sets and models investigated here.

For the second inverse problem, we estimate a probability density over parameter  $\omega$  by fixing the spline coefficients of the density over  $\theta$  at  $\Theta^M = \hat{\Theta}^M$ . When we estimate a probability density over parameter  $\omega$ , the observation process becomes  $\vec{f}(t; \Omega^{M*}) = [u_e^*(t; \{\hat{\Theta}_{1k}^M\}, \{\Omega_{1k}^{M*}\}), u_A^*(t; \{\hat{\Theta}_{2k}^M\}, \{\Omega_{2k}^{M*}\})]^T$  from equations (9) and (10). For this inverse problem, we only need to estimate the set of  $2(M^* + 1)$  spline coefficients in matrix  $\Omega^{M*}$ . A similar iterative process described above estimates values of  $J^*(\hat{\Omega}^{M*})$ ,  $\hat{\sigma}_{01}^{*2}$ , and  $\hat{\sigma}_{02}^{*2}$  for the new estimated distribution,  $\Omega^{M*}$ . In [1], we use a second order difference-based method to determine the correct statistical model to be defined by the weighted error coefficients  $\vec{\gamma} = [0.3, 0.6]$ .

### 4.3 Uncertainty Quantification: Standard Errors

In order to quantify uncertainty in estimating the spline coefficients  $\Theta^M$  and  $\Omega^{M*}$ , respectively, the standard errors and confidence intervals can be computed using standard asymptotic theory for the IRWLS spline coefficient estimators  $\Theta_{ik}^M$  and  $\Omega_{ik}^{M*}$ , respectively, for  $i = 1, 2$  [5, 7, 8]. For each time point  $t_j$ ,  $j = 1, \dots, n$ , the  $2 \times (M + 1)$  and  $2 \times (M^* + 1)$  sensitivity matrices for the  $i = 1, 2$  observations are

$$\left( D_j(\hat{\Theta}^M) \right)_{ik} = \frac{\partial f_i(t_j; \hat{\Theta}^M)}{\partial \Theta_{ik}} = \int_{\theta_l}^{\theta_u} x_i(t_j; \theta) l_k^M(\theta) d\theta \quad (18)$$

$$\left( D_j(\hat{\Omega}^{M*}) \right)_{ik^*} = \frac{\partial f_i(t_j; \hat{\Omega}^{M*})}{\partial \Omega_{ik^*}} = \int_{\omega_l}^{\omega_u} \left( \sum_{k=0}^M \hat{\Theta}_{ik}^M \int_{\theta_l}^{\theta_u} x_i(t_j; \theta, \omega) l_k^M(\theta) d\theta \right) l_{k^*}^{M*}(\omega) d\omega \quad (19)$$

$$(20)$$

where  $x_1 = x_e$  and  $x_2 = x_A$  is defined in equations (7) and (8). The  $(M + 1) \times (M + 1)$  and  $(M^* + 1) \times (M^* + 1)$  Fisher Information Matrices,  $\mathcal{F}$  and  $\mathcal{F}^*$ , respectively, are approximated using the sensitivity matrices,  $D_j(\hat{\Theta}^M)$  and  $D_j(\hat{\Omega}^{M*})$  in

$$\mathcal{F} \text{ (or } \mathcal{F}^*) \approx \left( \sum_{j=1}^n D_j^T \hat{V}_j^{-1} \hat{W}_j D_j \right)^{-1}. \quad (21)$$

where  $\hat{V}_j$  (or  $\hat{V}_j^*$ ) is the covariance matrix from equation (15) and  $\hat{W}_j = \text{diag}(f_1^{2\gamma_1}(t_j; \hat{\Theta}^M), f_2^{2\gamma_2}(t_j; \hat{\Theta}^M))$  (or  $\hat{W}_j^* = \text{diag}(f_1^{2\gamma_1}(t_j; \hat{\Omega}^{M*}), f_2^{2\gamma_2}(t_j; \hat{\Omega}^{M*}))$ ) is made up of error weights.

In the following sections, first we estimate the probability density of aggregate egg and adult GWSS populations over parameter  $\theta = a_1$  while fixing  $\tau = 0$  and all other parameters at values from Table 1. Thus, the first inverse problem analyzes a system of ODEs. Then, we fix the probability density of  $a_1$  at this estimated density and compute a second inverse problem to estimate the probability density of aggregate egg and adult GWSS populations over parameter  $\omega = \tau$  using equations (9) and (10). The second inverse problem utilizes the DDEs by letting  $\tau \geq 0$ .

## 5 Results

We first set  $\tau = 0$  and carry out estimations of the birthrate parameter distribution for  $a_1$  using  $M = 2, 3, 4, 5$  spline elements which yields relatively stable results as depicted in Table 2. Then we chose the distribution over  $a_1$  corresponding to  $M = 4$  to carry out the estimations of the delay parameter distribution for  $\tau$  using  $M^* = 4, 6, 8, 10, 12, 14, 16, 18, 20$  spline elements which also yields relatively stable results as depicted in Table 3. Lastly, we perform an analysis of variance (ANOVA) test to compare the model with no delay ( $\tau = 0$ ) to the model with a delay ( $\tau \geq 0$ ).

### 5.1 Probability Density over Birthrate Parameter $a_1$

Below, we graph the results of conducting the aggregate inverse problem over  $a_1$  while fixing  $\tau = 0$ . We choose parameter  $a_1$ , because the individual observations  $x_e$  and  $x_A$  from equation (1) and (2) are very sensitive to  $a_1$ . The parameter range  $[\theta_l, \theta_u] = [56, 57.7]$  is chosen so that the birthrate, defined in equation (4) makes biological sense and does not become unstable. The optimal cost value  $J$  from equation (14), the estimated variances of the residuals of eggs and adults from equation (15), and the condition number of the Fisher Information Matrix from equation (21) are shown in Table 2.

| $M$ | $J$ | $\hat{\sigma}_{01}^2$ | $\hat{\sigma}_{02}^2$ | $\kappa(\mathcal{F})$ |
|-----|-----|-----------------------|-----------------------|-----------------------|
| 2   | 498 | $9.6230 \times 10^7$  | $3.9589 \times 10^7$  | $2.1321 \times 10^2$  |
| 3   | 498 | $9.0193 \times 10^6$  | $4.1480 \times 10^7$  | $9.6931 \times 10^3$  |
| 4   | 498 | $8.6485 \times 10^6$  | $2.8241 \times 10^7$  | $2.4841 \times 10^5$  |
| 5   | 498 | $7.9681 \times 10^6$  | $2.0838 \times 10^7$  | $2.0838 \times 10^6$  |

Table 2: Optimal cost value ( $J$ ), estimated variances for eggs and adults respectively ( $\hat{\sigma}_{01}^2$  and  $\hat{\sigma}_{02}^2$ ), and the condition number of the Fisher Information Matrix,  $\mathcal{F}$  when computing the inverse problem for  $a_1$ .

The optimal cost value,  $J$ , appears to have converged for values of  $M$ . The estimated probability densities are graphed in Figure 2A and 2B. The resulting aggregate population model for eggs and adults when we let  $M = 4$  is plotted against the corresponding data points in Figure 2C and 2D, and the modified residuals are plotted in Figure 2E and 2F. The modified residuals of the adults do not appear to be independent, but this is to be expected since there is no delay ( $\tau = 0$ ).



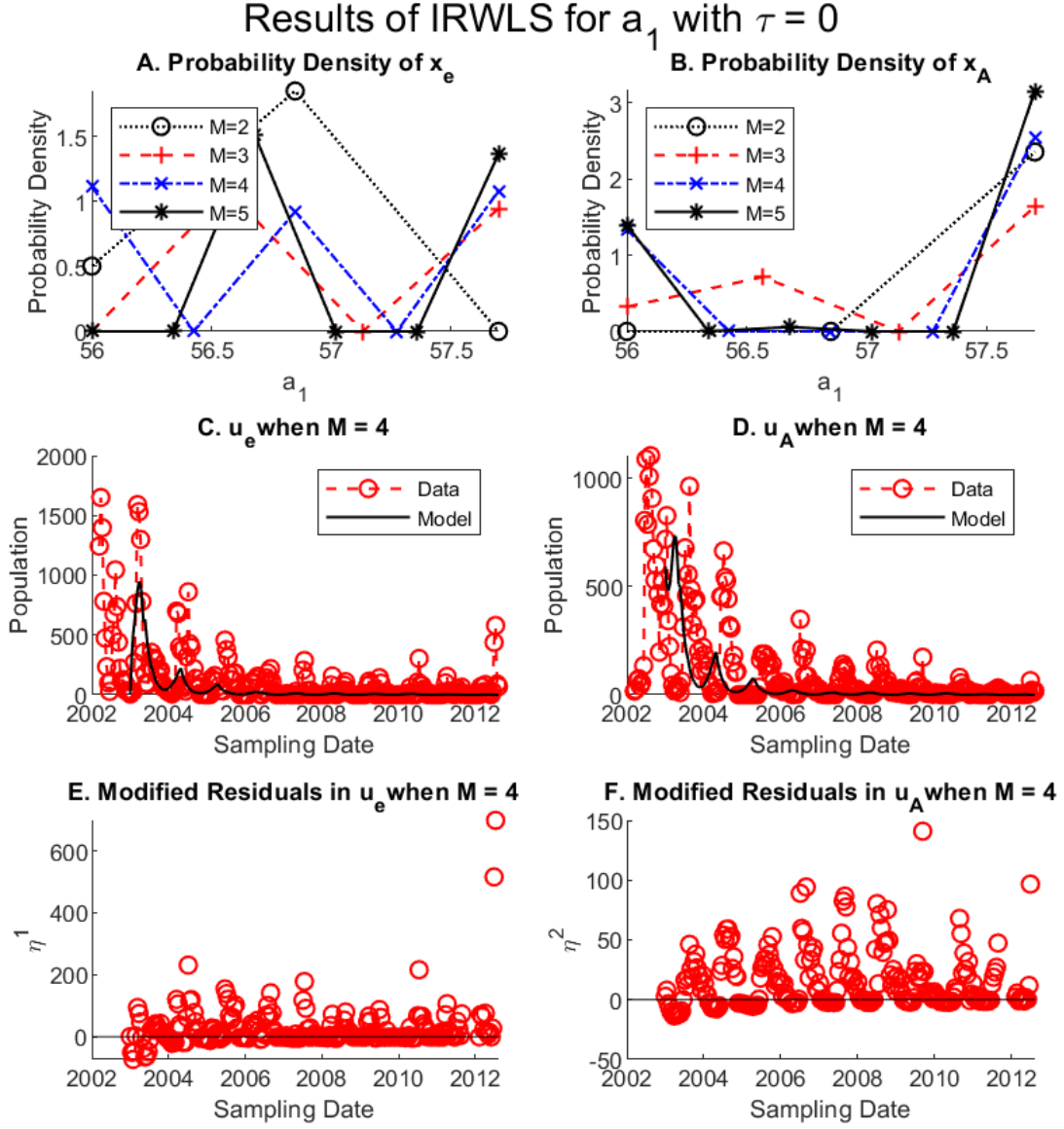


Figure 2: Estimated probability density coefficients of eggs (A) and adults (B) for  $M = 2, 3, 4, 5$  while varying parameter  $a_1$ ; aggregate egg (C) and adult (D) model solutions to equations (7) and (8) at estimated densities for  $M = 4$  and population phenology data; and modified residual errors for egg (E) and adult (F) observations.

## 5.2 Probability Density over Delay $\tau$

Below, we graph the results of conducting the aggregate inverse problem over  $\tau$  while fixing the probability density of  $a_1$  at the estimate  $\hat{\Theta}^4$  ( $M = 4$ ) in Figure 3. The optimal cost value  $J^*$ , the estimated variances of the eggs and adults, and the condition number of the Fisher Information Matrix for parameter  $\tau$  is shown in Table 3.

| $M^*$ | $J^*$  | $\hat{\sigma}_{01}^{*2}$ | $\hat{\sigma}_{02}^{*2}$ | $\kappa(\mathcal{F}^*)$ |
|-------|--------|--------------------------|--------------------------|-------------------------|
| 4     | 498.00 | $4.3336 \times 10^6$     | $2.0788 \times 10^6$     | $9.9987 \times 10^2$    |
| 6     | 497.99 | $4.0701 \times 10^6$     | $2.0296 \times 10^6$     | $1.0236 \times 10^4$    |
| 8     | 497.72 | $3.9197 \times 10^6$     | $2.0208 \times 10^6$     | $3.4799 \times 10^4$    |
| 10    | 496.44 | $3.8242 \times 10^6$     | $2.0371 \times 10^6$     | $1.3454 \times 10^5$    |
| 12    | 491.98 | $3.7574 \times 10^6$     | $2.0883 \times 10^6$     | $2.6122 \times 10^5$    |
| 14    | 482.62 | $3.7089 \times 10^6$     | $2.1616 \times 10^6$     | $5.6538 \times 10^5$    |
| 16    | 481.79 | $3.6747 \times 10^6$     | $2.2098 \times 10^6$     | $1.2025 \times 10^6$    |
| 18    | 482.65 | $3.6488 \times 10^6$     | $2.2038 \times 10^6$     | $1.4729 \times 10^6$    |
| 20    | 482.09 | $3.6329 \times 10^6$     | $2.2387 \times 10^6$     | $1.9585 \times 10^6$    |

Table 3: Optimal cost value ( $J^*$ ), estimated variances for eggs and adults respectively ( $\hat{\sigma}_{01}^{*2}$  and  $\hat{\sigma}_{02}^{*2}$ ), and the condition number of the Fisher Information Matrix,  $\mathcal{F}^*$  when computing the inverse problem for  $\tau$  while fixing the distribution over  $a_1$  at the estimate when  $M = 4$ .

## Results of IRWLS for $\tau$ with $a_1$ fixed at $M = 4$ Probability Density

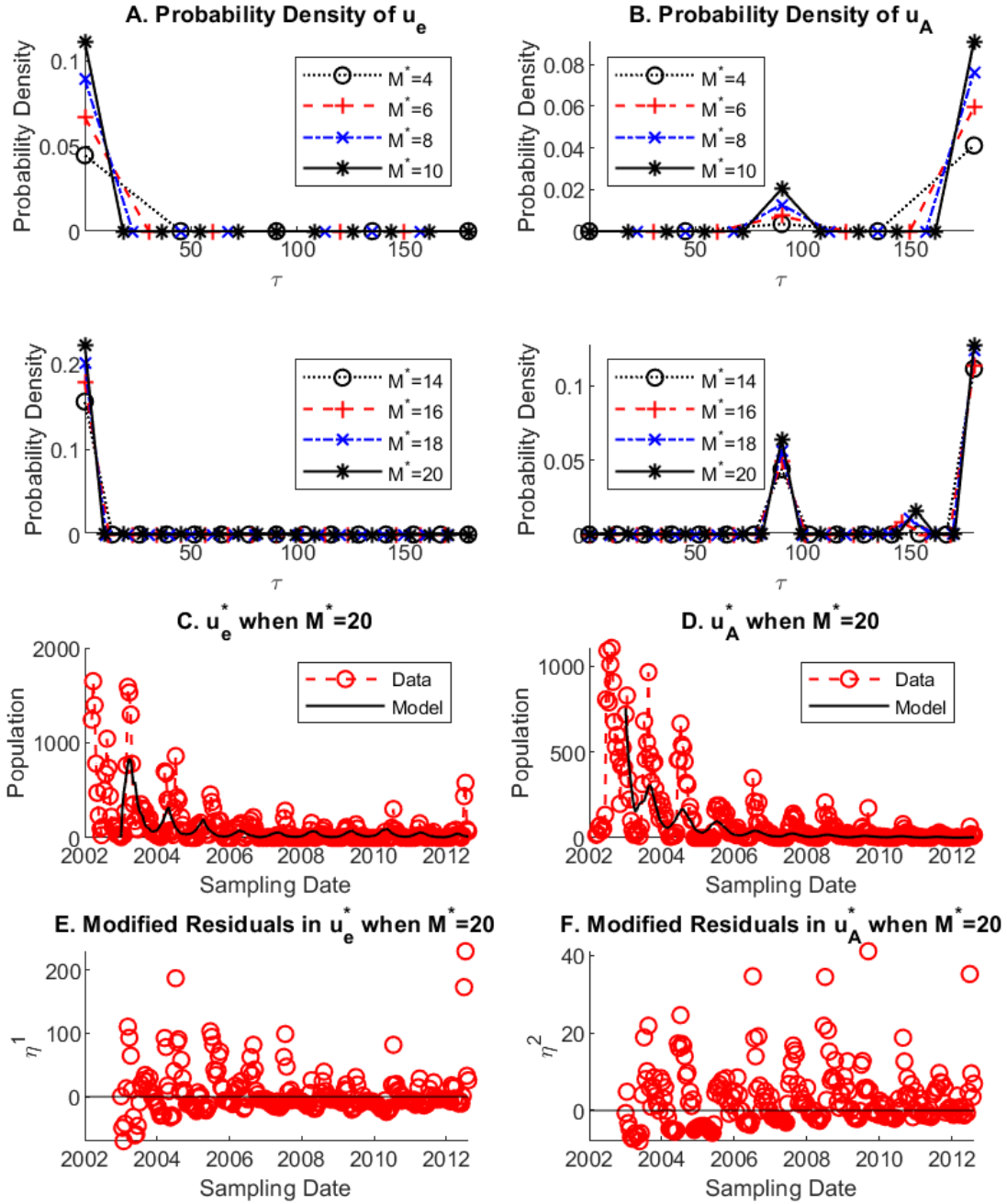


Figure 3: Estimated probability density coefficients of eggs (A) and adults (B) for  $M^* = 4, 6, 8, 10$  and  $M^* = 14, 16, 18, 20$  while varying parameter  $\tau$  and fixing  $a_1$  at the probability density when  $M = 4$ ; aggregate egg (C) and adult (D) model solutions to equations (9) and (10) at estimated densities for  $M^* = 20$  and population phenology data; and modified residual errors for egg (E) and adult (F) observations.

## 6 Model Comparison

We use an analysis of variance (ANOVA) type test based on asymptotic distributional results for estimators and residuals to compare the GWSS model without a delay (i.e., the aggregate model assuming a distribution over birthrate parameter  $a_1$  with  $M = 4$  spline elements) to the GWSS model with a delay (i.e., the aggregate model assuming a distribution over the delay  $\tau$  while fixing the probability density of  $a_1$  at the estimate for  $M = 4$ ). This statistical test is taken from Banks and Fitzpatrick [10]. We can calculate the test statistic realizations,  $\mathcal{T} = n(J - J^*)$  where  $n = 250$  is the number of time points used in the IRWLS problem. This test statistic is used to determine a more useful test statistic realization,  $\mathcal{U} = \mathcal{T}/J^*$ . Using the  $\chi^2(1)$  distribution, we can reject or fail to reject the following null hypothesis:

$H_0$ : The fits provided by the aggregate GWSS model with the delay  $\tau \geq 0$  is not significantly different from the fit with the aggregate GWSS model when  $\tau = 0$ .

$H_A$ : The aggregate GWSS model with the delay  $\tau \geq 0$  provides a significantly better fit than when restricted to the constant value  $\tau = 0$ .

The three test statistics,  $\mathcal{T}$ ,  $\mathcal{U}$ , and the  $p$ -value of the  $\chi^2$  test for different values of  $M^*$  are shown below in Table 4. With a confidence value of  $\alpha = 0.01$ , we reject the null hypothesis  $H_0$  for  $M^* = 14, 16, 18, 20$ . In Figure 3, for smaller values of  $M^* = 4, 6, 8, 10$  the probability density over the delay approaches values of  $\tau = 90, 180$  for the aggregate adult populations. However, for larger values of  $M^* = 14, 16, 18, 20$ , the probability density over the delay approaches values of  $\tau = 90, 150, 180$  for the aggregate adult populations. This added feature of the adult density causes optimal cost,  $J^*$ , to decrease significantly.

Table 4: Values of  $\mathcal{U}$  when comparing the GWSS model without a delay,  $\tau = 0$ , to the GWSS model with a distribution over the delay  $\tau$ .

| Test Statistic | $M^*$       |        |        |        |        |        |        |        |        |
|----------------|-------------|--------|--------|--------|--------|--------|--------|--------|--------|
|                | 4           | 6      | 8      | 10     | 12     | 14     | 16     | 18     | 20     |
| $\mathcal{T}$  | $< 10^{-3}$ | 3.4066 | 70.765 | 390.60 | 1506.2 | 3845.1 | 4051.8 | 38.36  | 3977.2 |
| $\mathcal{U}$  | $< 10^{-5}$ | 0.0068 | 0.1422 | 0.7868 | 3.0615 | 7.9672 | 8.4098 | 7.9495 | 8.2499 |
| $p$ -value     | $\approx 1$ | 0.9341 | 0.7061 | 0.3751 | 0.0802 | 0.0048 | 0.0037 | 0.0048 | 0.0041 |

## 7 Discussion and Conclusion

By assuming a probability distributions over parameters in least squares problems, we can use a simple mathematical model to capture sophisticated information from aggregate data sets. Without the delay ( $\tau = 0$ ), the model laid out in equations (1) and (2) is a simple linear system of ODEs with two age classes, and with the delay ( $\tau > 0$ ), the system is the simplest form of DDEs, with one constant delay. However, by taking the aggregate nature of the data into account and assuming weighted observational errors, we find that in this population of *H. vitripennis* large proportions of eggs are dying out, and the combined nymphal and preoviposition periods of the populations vary between 90, 150, and 180 days.

In the first IRWLS problem, we estimate the probability distribution over one of the birthrate parameters,  $a_1$ . When  $a_1 = 56$ , the birthrate is zero for most temperature values, except optimal values. When  $a_1 = 56.9$  and  $57.7$ , the birthrates are larger and nonzero in the range of 20-30°C. Thus, since our results show that aggregate adult *H. vitripennis* populations have a probability density centered around  $a_1 = 56$  and  $57.7$ , only the best and worst egg populations are being modeled. Thus, in this population of glassy-winged sharpshooters, a large proportion of *H. vitripennis* eggs are dying out before they reach adulthood.

In the second IRWLS problem, we estimate the probability distribution over the delay,  $\tau$ , and get significantly better results when we fix the birthrate parameter  $a_1$  at the probability density with  $M = 4$ . The results indicate that the aggregate egg *H. vitripennis* populations hatch from the adult populations with no delay, while the adults in this population of *H. vitripennis* undergo a nymphal and preoviposition period of 90, 150, and 180 days. The proportion of adults with  $\tau = 150$  are essential when comparing the GWSS model with no delay to the GWSS model with a delay. The fit only becomes significantly better ( $p \leq 0.005$ )

when  $M^*$  is large enough to account for this proportion of adults. Thus, we have shown that the driving factors of this population of *H. vitripennis* are the differing combined nymphal and preoviposition periods.

There are a few drawbacks and improvements to be made with this model. When we fix  $\tau = 0$ , we assume that a nymphal and preoviposition period does not exist and that a GWSS adult can emerge from the egg fully capable of sexual reproduction. This is not biologically reasonable, since the combined nymphal and preoviposition period for GWSS is known to last from 50 to 140 days [20]. In Figure 2, the modified residuals graphed are clearly not independent, so this is not the correct mathematical model. Since the GWSS ODE model (the model with  $\tau = 0$ ) is not correct, the estimated distribution over  $a_1$  may not be accurate. Thus, it may be useful to estimate the probability densities over  $a_1$  and  $\tau$  at the same time.

Even with the delay, the GWSS model fails to capture the second yearly spike in the *H. vitripennis* egg population. It is known that invasive populations of GWSS located in Southern California usually have two periods of oviposition per year: once in early spring, and again in mid-to-late summer [11]. This behavior may need to be modeled with a more complicated system of DDEs. In [1] we simply ignored the first oviposition period when the GWSS adult population is not very large compared to the egg population and modeled the population with a simple system of ODEs, which resulted in much larger errors in the GWSS egg observations.

The most important change to be made with the GWSS model with respect to this data set is to take egg parasitism into account. Egg parasitoids including *Cosmocomoidea ashmeadi* and *Cosmocomoidea morrili* were observed and counted in the *H. vitripennis* egg populations. [1] This biotic factor may play a crucial role in the decline of the *H. vitripennis* population over the ten year period and needs further analysis.

## 8 Acknowledgement

This research was supported in part by the Air Force Office of Scientific Research (HTB) under grant numbers AFOSR FA9550-15-1-0298 and AFOSR FA9550-18-1-0457, in part by the National Science Foundation (AEM) under grant number RTG/DMS - 1246991. The GWSS survey work was supported, in part, by funds awarded to MSH by the California Department of Food and Agriculture Pierce's Disease Research Program and the University of California Pierce's Disease Research Program.

## References

- [1] H. T. Banks, John E. Banks, Natalie G. Cody, Mark S. Hoddle, and Annabel E. Meade. Population model for the decline of *Homalodisca vitripennis* (Hemiptera: Cicadellidae) over a ten-year period, CRSC-TR18-06, N. C. State University, Raleigh, NC, June, 2018; *J. Biological Dynamics*, submitted.
- [2] H. T. Banks, K. Bekele-Maxwell, R. A. Everett, L. Stephenson, S. Shao, and J. Morgenstern. Dynamic modeling of problem drinkers undergoing behavioral treatment, CRSC-TR16-12, N. C. State University, Raleigh, NC, October, 2016. *Bulletin of Mathematical Biology*, **79**(6):1254–1273, 2017.
- [3] H. T. Banks and K. L. Bihari. Modelling and estimating uncertainty in parameter estimation. *Inverse Problems*, **17**:95–111, 2001.
- [4] H. T. Banks, D. M. Bortz, G. A. Pinter, and L. K. Potter. Modeling and imaging techniques with potential for application in bioterrorism, CRSC-TR03-02, N. C. State University, Raleigh, NC, January, 2003. In *Bioterrorism: Mathematical Modeling Applications in Homeland Security*, pages 129–154. SIAM, 2003.
- [5] H. T. Banks and J. L. Davis. Quantifying uncertainty in the estimation of probability distributions with confidence bands, CRSC-TR07-21, N. C. State University, Raleigh, NC, December, 2007. *Mathematical Biosciences and Engineering: MBE*, **5**(4):647–667, 2008.
- [6] H. T. Banks, K. B. Flores, I. G. Rosen, E. M. Rutter, Melike Sirlanci, and W. Clayton Thompson. The prohorov metric framework and aggregate data inverse problems for random PDEs, CRSC-TR18-05, N. C. State University, Raleigh, NC, June, 2018. *Communications in Applied Analysis*, **22**:415–446, 2018.

- [7] H. T. Banks, Shuhua Hu, and W. C. Thompson. *Modeling and inverse problems in the presence of uncertainty*. CRC Press, 2014.
- [8] H. T. Banks and Hien T Tran. *Mathematical and experimental modeling of physical and biological processes*. CRC Press, 2009.
- [9] H. T. Banks and W. C. Thompson. Random delay differential equations and inverse problems for aggregate data problems, CRSC-TR18-07, N. C. State University, Raleigh, NC, July, 2018. *Eurasian J. Mathematical and Computer Applications*, 2018, submitted.
- [10] H. T. Banks and B. G. Fitzpatrick. Statistical methods for model comparison in parameter estimation problems for distributed systems. *Journal of Mathematical Biology*, **28**(5):501–527, 1990.
- [11] M. J. Blua, P. A. Phillips, and R. A. Redak. A new sharpshooter threatens both crops and ornamentals. *Plant Health Progress*, **1**(1):4, 2000.
- [12] J. R. Carey. The multiple decrement life table: a unifying framework for cause-of-death analysis in ecology. *Oecologia*, **78**(1):131–137, 1989.
- [13] J. Grandgirard, M. S. Hoddle, J. N. Petit, G. K. Roderick, and N. Davies. Engineering an invasion: classical biological control of the glassy-winged sharpshooter, *homalodisca vitripennis*, by the egg parasitoid *gonatocerus ashmeadi* in tahiti and moorea, french polynesia. *Biological Invasions*, **10**(2):135–148, 2008.
- [14] J. Grandgirard, M. S. Hoddle, J. N. Petit, G. K. Roderick, and N. Davies. Classical biological control of the glassy-winged sharpshooter, *homalodisca vitripennis*, by the egg parasitoid *gonatocerus ashmeadi* in the society, marquesas and austral archipelagos of french polynesia. *Biological Control*, **48**(2):155–163, 2009.
- [15] J. N. Kapur. A continuous time age structured population growth model. *Applied Mathematical Modelling*, **3**(6):455–458, 1979.
- [16] D. J. Lactin, N. J. Holliday, D. L. Johnson, and R. Craigen. Improved rate model of temperature-dependent development by arthropods. *Environmental Entomology*, **24**(1):68–75, 1995.
- [17] R. F. Mizell, C. Tipping, P. C. Andersen, B. V. Brodbeck, W. B. Hunter, and T. Northfield. Behavioral model for *Homalodisca vitripennis* (hemiptera: Cicadellidae): optimization of host plant utilization and management implications. *Environmental Entomology*, **37**(5):1049–1062, 2008.
- [18] California Department of Food and Agriculture. Pierce’s disease control program, 2018. <http://www.cdffa.ca.gov/pdcp/>. Online; accessed 9 March 2018.
- [19] L. J. Pilkington and M. S. Hoddle. Reproductive and developmental biology of *Gonatocerus ashmeadi* (hymenoptera: Mymaridae), an egg parasitoid of *Homalodisca coagulata* (hemiptera: Cicadellidae). *Biological Control*, **37**(3):266–275, 2006.
- [20] L. J. Pilkington, M. Lewis, D. Jeske, and M. S. Hoddle. Calculation and thematic mapping of demographic parameters for *Homalodisca vitripennis* (hemiptera: Cicadellidae) in california. *Annals of the Entomological Society of America*, **107**(2):424–434, 2014.
- [21] K. Tumber, J. Alston, and K. Fuller. Pierce’s disease costs california \$104 million per year. *California Agriculture*, **68**(1):20–29, 2014.
- [22] J-M Yoon, V. Hryniv, L. Morano, A-T Nguyen, S. Wilder, and F. Mitchell. Mathematical modeling of Glassy-winged Sharpshooter population. *Mathematical Biosci. and Engineering*, **11**(3):667–677, 2014.