

PARAMETER ESTIMATION  
FOR AN ALLOMETRIC FOOD WEB MODEL

H.T. Banks<sup>1</sup>, J.E. Banks<sup>2</sup>, Riccardo Bommarco<sup>3</sup>,  
Alva Curtsdotter<sup>4</sup>, Tomas Jonsson<sup>5</sup>, A.N. Laubmeier<sup>6</sup>

<sup>1,6</sup>Center for Research in Scientific Computation  
North Carolina State University  
Raleigh, NC 27695, USA

<sup>2</sup>Undergraduate Research Opportunities Center (UROC)  
California State University, Monterey Bay  
Seaside, CA 93955, USA

<sup>3,4,5</sup>Department of Ecology  
Swedish University of Agricultural Sciences  
Uppsala, SE-75007, SWEDEN

<sup>5</sup>School of Bioscience  
University of Skövde  
Skövde, SE-54128, SWEDEN

---

**Abstract:** The application of mechanistic models to natural systems is of interest to ecological researchers. We use the mechanistic Allometric Trophic Network (ATN) model, which is well-studied for controlled and theoretical systems, to describe the dynamics of the aphid *Rhopalosiphum padi* in an agricultural field. We diagnose problems that arise in a first attempt at a least squares parameter estimation on this system, including formulation of the model for the inverse problem and information content present in the data. We seek to establish whether the field data, as it is currently collected, can support parameter estimation for the ATN model.

**AMS Subject Classification:** 92D25, 92D40, 65L09, 93C10

**Key Words:** population models, *Rhopalosiphum padi*, allometric trophic network, parameter estimation, residual analysis, model selection

---

Received: March 27, 2017

Revised: April 20, 2017

Published: April 21, 2017

© 2017 Academic Publications, Ltd.

url: [www.acadpubl.eu](http://www.acadpubl.eu)

## 1. Introduction

The effects of increased biological diversity on species and ecosystems has long been a central area of focus in ecology. Early biodiversity-stability debates[21, 30, 36] have given rise to more general explorations of how biodiversity affects ecosystem functioning and services[27, 1, 31, 29]. One critical ecosystem service, biological control, has been the subject of intense scrutiny for decades, though often in a simplified way that focuses on single species or a handful of species interactions within only one trophic level[18, 28, 34]. A broader understanding of how predator richness affects pest suppression remains elusive[26]; models of multi-species interactions in a food-web context across several trophic levels indicate that biological control can be both positively and negatively affected by biological diversity[35]. We present here a predictive, mechanistic approach to modelling multiple species interactions that allows us to investigate the link between biological control and biodiversity. We use an Allometric Trophic Network (ATN) model, a dynamic food web model which assumes a general relationship between trophic interaction strengths and the body size of predator and prey. The strength of the ATN model, particularly compared to standard Lotka-Volterra and matrix models, is that in parameterizing interactions by body mass, we reduce the number of parameters which must be identified to compute a model solution. The ATN model is well-studied in theoretical ecology [14, 32, 11] and has been applied to controlled microcosms [25, 33]. The model has also been used to describe the average seasonal behaviors of an aquatic food web [12]. Multivariate autoregressive (MAR) models [19, 23] have also been successfully applied to aquatic food webs [24, 22] and can be used with model comparison tests to identify the dominant interactions in a foodweb. Despite having a strong statistical framework which accommodates process-driven deviation from the model, these MAR models only support linear species interactions. In comparison, the ATN model is formulated with a Holling-type functional response[33] and therefore supports dynamic species interactions which can change with availability of alternative prey. In order to test the robustness of our approach, we validate the model by fitting field data to our model.

We consider the use of the ATN model to describe the dynamics of the aphid *Rhopalosiphum padi*, an agricultural pest, subject to population loss by the naturally-occurring community of predators in a single-season barley field. In initial efforts at solving a least squares inverse problem for this model using observational data, we found that the model's complexity, when paired with the sparsity of available data, led to numerical difficulties and poor fits to data [2].

Here, we investigate the limitations present in this first parameter estimation. We explore the statistical model for observational error in the data, formulation of the mathematical model used to describe the system, and the information content necessary to estimate the parameters in the ATN model. Revisiting the underlying statistical and mathematical models used in the formulation of the inverse problem is an expected step in the *iterative modelling process*[10] and allows us to understand the conditions under which the parameters for the study system and model can be estimated with available field data.

## 2. Methods

### 2.1. ATN Model

We have data for the arthropod populations of ten barley fields in Uppland, Sweden, for six weeks from late May to early July, 2011. The arthropods are divided into 15 groups, which we will call nodes, to form the basis for the food web in the barley fields. See [2] for a description of the experimental design and assumptions used in the formulation of the food web.

We attempt to model the dynamics of the bird cherry-oat aphid, *Rhopalosiphum padi*, with an Allometric Trophic Network (ATN) model. The ATN model assumes Lotka-Volterra dynamics with a Holling Type-II response, where parameters are assumed to depend on the body masses of interacting species[33]. The dynamics of the aphid population,  $N_1$ , are given by

$$\frac{dN_1}{dt} = N_1 \left[ r(T(t)) - \sum_{j \in \mathcal{C}_1} \frac{a_{1j}(a_0, R_{opt}, \phi)N_j}{1 + c_j(c_0)N_j + \sum_{k \in \mathcal{R}_j} a_{kj}(a_0, R_{opt}, \phi)h_{kj}(h_0)N_k} \right], \quad (1)$$

where  $N_i$  is the population density of node  $i$ ,  $\mathcal{C}_i$  is the set of consumers for node  $i$ ,  $\mathcal{R}_i$  is the set of prey for node  $i$ ,  $a_{ij}$  is the attack rate of node  $j$  on node  $i$ ,  $c_j$  is the coefficient of intraspecific interference competition for node  $j$ ,  $h_{ij}$  is the handling time for one individual of node  $j$  to catch and consume one individual of node  $i$ . Here,  $r(T(t))$  is the temperature-dependent growth rate of the aphid population,

$$r(T(t)) = .024T(t) - .089.$$

The time-varying temperature,  $T(t)$ , is given by observational data. For all  $j \neq 1$ , the density  $N_j$  is an input to the system from observational data.

The body mass-dependent parameters are

$$\begin{aligned} a_{ij}(a_0, R_{opt}, \phi) &= a_0 W_i^{1/4} W_j^{1/4} \left( \frac{W_j/W_i}{R_{opt}} e^{1 - \frac{W_j/W_i}{R_{opt}}} \right)^\phi, \\ h_{ij}(h_0) &= h_0 (W_i/W_j)^{1/4}, \\ c_j(c_0) &= c_0 W_j^{1/2}, \end{aligned}$$

where  $W_i$  is the body mass of node  $i$ ,  $R_{opt}$  is the optimal body mass ratio between predator and prey,  $\phi$  scales the dependency of a successful attack on predator-prey body mass ratio, and  $a_0$ ,  $h_0$ , and  $c_0$  are normalizing constants. The body masses  $W_i$  are known, but we must estimate the parameter set  $q = [a_0, h_0, c_0, R_{opt}, \phi]$ . We note that the parameters  $a_0$ ,  $h_0$ ,  $c_0$ , and  $\phi$  do not have a simple, physical meaning; experimentally determining empirical values for these parameters would be challenging, in particular for realistic predator-prey species compositions. Additionally, we expect that some of these parameters will vary between study systems; for example, the optimal predator-prey body mass ratio has been found to vary between predator types[13], so we can expect  $R_{opt}$  to vary between systems. For theoretical studies, the parameters for the ATN model have historically been generalized from metabolic theory and known allometric relationships across many species and ecosystems[14, 15, 37]. In empirical studies, parameters are hand-selected by choosing the best-fitting parameters to a data set over many simulations[33, 25]. We use a least squares parameter estimation to find the parameters for the system.

## 2.2. Formulation of the Inverse Problem

Following [9, 10], we assume a *relative error* statistical or observational model

$$Y_j = f(t_j; \theta_0) + f^\gamma(t_j; \theta_0) \mathcal{E}_j, \quad (2)$$

for  $Y_j$  the observation of our system at a point  $t_j$  and  $\theta_0 = [q_0, y_0]^T$ , the parameter set that we must estimate. Here  $f(t_j; \theta_0)$  is the solution  $N_1(t_j; q_0, y_0)$  of (1) at point  $t_j$  with nominal parameters  $q_0$  and initial condition  $y_0$ ,  $\gamma$  is a constant weight that specifies the parameters of the error's distribution, and  $\mathcal{E}_j$  is independent, identically distributed random noise assumed to be  $\mathcal{N}(0, 1)$ . For an inverse problem with  $n$  observations, the *iterative weighted least squares* (IWLS) (also referred to as *generalized least squares* (GLS)) cost functional gives an estimator  $\Theta$  of the parameter  $\theta_0$  defined by

$$\Theta = \arg \min_{\theta \in \Omega} \sum_{j=1}^n \left( \frac{Y_j - f(t_j; \theta)}{f^\gamma(t_j; \theta)} \right)^2, \quad (3)$$

where  $\Omega$  is the space of admissible parameters. Our experiences [4, 3, 5, 6] with other ecological and population modeling efforts involving *population counts* have suggested that some type of error process depending on the *size* of the population is likely to be most appropriate. This motivates our choice of statistical model in (2).

The collected data  $N_1^j$  is a realization of  $Y_j$ ,

$$N_1^j = f(t_j; \theta_0) + f^\gamma(t_j; \theta_0)\epsilon_j$$

for  $\epsilon_j$  a realization of  $\mathcal{E}_j$ . Then a realization of the minimizing parameter is given by

$$\hat{\theta} = \arg \min_{\theta \in \Omega} \sum_{j=1}^n \left( \frac{N_1^j - f(t_j; \theta)}{f^\gamma(t_j; \theta)} \right)^2. \quad (4)$$

The weight  $\gamma$  is unknown and can generally be identified using the method of residual plots, outlined in [10, 9]. From the statistical model, we know that  $\mathcal{E}_j = [Y_j - f(t_j; \theta_0)]/f^\gamma(t_j; \theta_0)$  and that the realizations of  $\mathcal{E}_j$  should be independent and identically distributed as random noise. That is, we expect a plot of the modified residuals

$$r_{j,\gamma} = \frac{N_1^j - f(t_j; \hat{\theta})}{f^\gamma(t_j; \hat{\theta})}$$

will be randomly distributed for an appropriate choice of  $\gamma$ .

However, the use of residual plots to determine the statistical model tacitly assumes that our mathematical model accurately describes the system dynamics. Initial treatment of the inverse problem in [2] suggested that the ATN model cannot capture some aspects of aphid population dynamics. Therefore, we first use difference-based pseudo-measurement errors, as described in [7], to test our statistical model without tacitly assuming a mathematical model. We compute estimates of the measurement error,  $f^\gamma(t_j; \hat{\theta})\epsilon_j$ , from the observed  $N_1^j$ . We employ the second-order central differencing scheme for pseudo-measurement errors

$$\hat{\epsilon}_j = \frac{1}{\sqrt{6}} \left( N_1^{j-1} - 2N_1^j + N_1^{j+1} \right).$$

To find  $\gamma$  for use in our statistical model, we compute

$$\eta_{j,\hat{\gamma}} = \frac{\hat{\epsilon}_j}{\left| N_1^j - \hat{\epsilon}_j \right|^{\hat{\gamma}}}$$

for different values of  $\hat{\gamma}$  until we find the value that results in a random distribution of  $\eta_{j,\hat{\gamma}}$  when plotted against observations  $t_j$ .

We assume this resulting value of  $\gamma$  to facilitate the use of model comparison tests for our system. After assuming an appropriate mathematical model, we may return to the traditional method of residual plots to verify that our choice of  $\gamma$  is accurate.

### 2.3. Model Comparison

We will refer to the model given by the equation in (1) as Model 1. We will also consider the restriction of (1) to a parameter space where  $h_0 = c_0 = 0$ . This results in a linear functional response

$$\frac{dN_1}{dt} = N_1 \left[ r(T) - \sum_{j \in \mathcal{C}_1} a_{1j}(a_0, R_{opt}, \phi) N_j \right]. \quad (5)$$

In this restriction, we note that there are only three constants,  $[a_0, R_{opt}, \phi]$ , and the initial condition  $N_1^0$  to be estimated. We denote this as Model 2.

The aphid population data exhibits unexpected population declines that cannot reasonably be attributed to the predators in the ATN model. Therefore, we also consider the addition of an as yet unknown source of mortality to the model. We assume that the mortality occurs over a full day and after the last data point before an unexpected population decline. That is, if the aphid population decreases by 25% or more between the observation collected at  $t_j$  and the observation collected at  $t_{j+1}$ , then we denote the time  $T_k = t_j$ . We assume that the aphids suffer some extrinsic mortality factor  $\mu$  over the interval  $M_k = [T_k + 1, T_k + 2]$ . We denote the set of all such times as  $\mathcal{M} = \bigcup_k M_k$  and add this discrete mortality term to (5)

$$\frac{dN_1}{dt} = N_1 \left[ r(T) - \sum_{j \in \mathcal{C}_1} a_{1j}(a_0, R_{opt}, \phi) N_j - \mu \chi_{\mathcal{M}} \right], \quad (6)$$

where  $\chi_{\mathcal{M}}$  is the indicator function for the set  $\mathcal{M}$ . We refer to this as Model 3, and it requires the specification of the constants  $[a_0, R_{opt}, \phi, \mu]$  as well as the initial condition  $N_1^0$ . Note that Model 2 is a restriction of Model 3 to a parameter space where  $\mu = 0$ .

Parameter estimation on Model 2 yields values of  $\phi \in [.3, 2]$ , so we also

consider the restriction of (5) to a parameter space where  $\phi = 1$

$$\frac{dN_1}{dt} = N_1 \left[ r(T) - \sum_{j \in \mathcal{C}_1} a_{1j}(a_0, R_{opt}, 1) N_j \right]. \tag{7}$$

This is equivalent to assuming that the sensitivity of a successful attack to predator-prey body mass ratio does not need to be tuned for the system. We will refer to this as Model 4, where the only parameters to estimate are the constants  $[a_0, R_{opt}]$  and the initial condition  $N_0^1$ .

To test if the fit obtained by a particular model is a significant improvement over ones of its restrictions, we follow the methods developed in [8] and carefully summarized in [9, 10]. We denote  $J(Y, \theta)$  to be the cost functional associated with a parameter  $\theta$  and observations  $Y$ , that is,

$$J(Y, \theta) = \sum_{j=1}^n \left( \frac{Y_j - f(t_j; \theta)}{f^\gamma(t_j; \theta)} \right)^2.$$

We let  $\Theta_1$  be the estimator of the nominal parameter  $\theta_0$ , as defined in (3), drawn from an admissible parameter space  $\Omega_1$ . That is,

$$\Theta_1 = \arg \min_{\theta \in \Omega_1} J(Y, \theta).$$

For any  $\Omega_2 \subset \Omega_1$ , we let  $\Theta_2$  be the estimator of  $\theta_0$  drawn from an admissible parameter space  $\Omega_2$ , i.e.,

$$\Theta_2 = \arg \min_{\theta \in \Omega_2} J(Y, \theta).$$

We test the *null hypothesis* that the nominal parameter exists in the constrained parameter space, or, that  $\theta_0 \in \Omega_2$ . We define the test statistic

$$U_n(Y) = n \frac{J(Y, \Theta_1) - J(Y, \Theta_2)}{J(Y, \Theta_1)}.$$

Here,  $n$  is the number of data points used in the inverse problem. For realizations of  $Y_j$ , denoted  $\hat{N}_1 = \{N_1^j\}_1^n$ , we let  $\hat{\theta}_1$  and  $\hat{\theta}_2$  be corresponding realizations of  $\Theta_1$  and  $\Theta_2$ , as defined in (4). Then we have a realization of the test statistic given by

$$\hat{U}_n(\hat{N}_1) = n \frac{J(\hat{N}_1, \hat{\theta}_1) - J(\hat{N}_1, \hat{\theta}_2)}{J(\hat{N}_1, \hat{\theta}_1)}.$$

Given a realization  $\hat{U}_n(\hat{N}_1)$ , we reject the null hypothesis with  $(1 - \alpha) \cdot 100\%$  certainty if  $\hat{U}_n(\hat{N}_1) > \tau_\alpha$ . Here  $\tau_\alpha$  is specified by the critical value of the  $\chi^2(r)$  distribution, for  $r$  the additional degrees of freedom that the model associated with  $\Omega_1$  has over the restriction given by  $\Omega_2$ . We set the threshold for model rejection at 75%, since we are dealing with rather sparse experimental data sets.

### 3. Results

#### 3.1. Selecting $\gamma$

We compute pseudo-measurement errors for the aphid population densities in the ten fields where data was collected. We note that the data sets for all fields were sparse, with five or six observations for each field. After central differencing, we are left with three or four estimates of the pseudo-measurement error, and it is difficult to identify randomness in so small a set of points. However, the sparsity of the data is problematic even when using residuals from the model. As in [2], we cannot compare the mathematical model with data when crop quality deteriorates late in the season, and only four or five data points from each field can be used in residual analysis.

For some fields, we successfully identify the appropriate statistical model using pseudo-measurement errors. The values of  $\eta_{j,\hat{\gamma}}$  for Field JC are plotted in Figure 1, where we can see that  $\hat{\gamma} = 0$  gives a distribution that appears random, while  $\hat{\gamma} = 1$  and  $\hat{\gamma} = 2$  do not perform as well. Similar results were obtained in Fields JO, KC, KO, and OO, plotted in Figures 5, 6, 7, and 11 in the appendix. In the remaining fields, the data sparsity makes it difficult to choose an appropriate value of  $\hat{\gamma}$ . For example, we plot the results from Field SO in Figure 2. No choice of  $\hat{\gamma}$  appears to give a random distribution of  $\eta_{j,\hat{\gamma}}$ , so we cannot choose a value of  $\hat{\gamma}$  for the statistical model. We reach the same conclusion for Fields MC, MO, OC, and SC (see Figures 8, 9, 10, and 12).

For simplicity, and having insufficient data to assume otherwise, we take  $\gamma = 0$  and plot the solutions to the inverse problem with absolute error formulation in Figure 3. Although we do not necessarily expect that  $\gamma$  will be the same for all fields, we require additional data to identify the correct value for the statistical model. In summary, we find that  $\gamma = 0$  is an appropriate choice for some fields, and do not have sufficient information to select a different appropriate value of  $\gamma$  in the remaining fields.



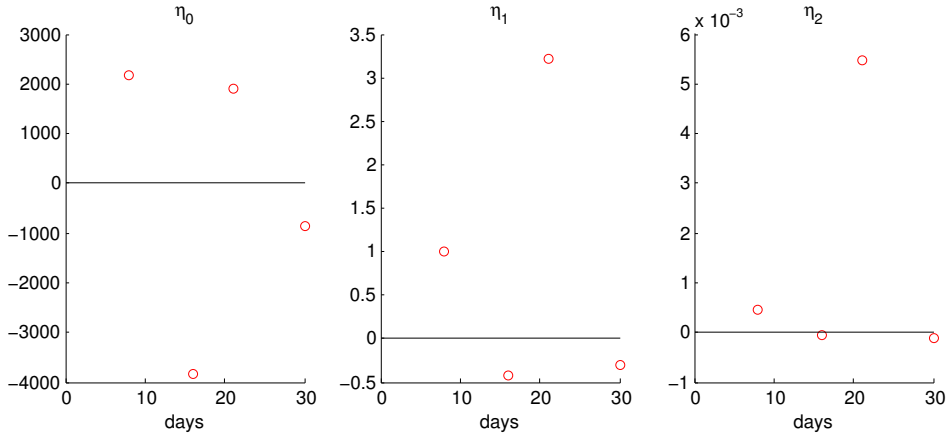


Figure 1: Values of  $\eta_{j,\hat{\gamma}}$  are plotted against time for  $\hat{\gamma} = 0, 1,$  and  $2$  for data from field JC.

### 3.2. Model comparison test

We first compare the performance of Model 1, the full ATN model, to its restriction to a linear functional response in Model 2, which assumes that  $h_0 = c_0 = 0$ . The results of the model comparison test are given in Table 1, while the model solutions are plotted in Figure 3. The least squares minimization for the full model is to take  $h_0 = c_0 = 0$  in almost all fields, and so there is little variation between the solutions to Models 1 and 2. As expected, we see in Table 1 a failure to reject the null hypothesis, that the solution is contained in the restriction given by Model 2.

Although we fail to show that the nominal parameters do not satisfy  $h_0 = c_0 = 0$ , we cannot readily conclude from this test that the functional response in the ATN model should be linear. Our data set is sparse, and there may be unobserved dynamics which require specification of  $h_0, c_0 \neq 0$ . In other words, the current data do not support with strong statistical significance the inclusion of the more complicated dynamics of Model 1 over the simpler dynamics of Model 2. For these data sets, we take  $h_0 = c_0 = 0$ , having no statistical motivation to do otherwise, but we must revisit the nonlinear formulation of the functional response if we acquire new, hopefully more extensive, data sets.

We consider the addition of a mortality term by asserting the null hypothesis that the nominal parameters are contained in the restriction given by Model 2 instead of the model with mortality, Model 3. The results from the comparison test, for the six fields that satisfy the requirement that a 25% or greater

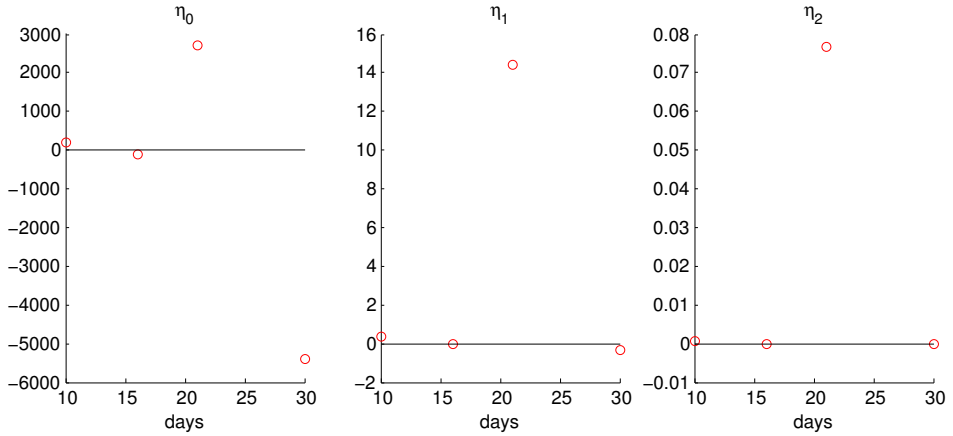


Figure 2: Values of  $\eta_{j,\hat{\gamma}}$  are plotted against time for  $\hat{\gamma} = 0, 1,$  and  $2$  for data from field SO.

Field	JC	JO	KC	KO	MC	MO	OC	OO	SC	SO
$\tilde{U}_n$	0.126	0.001	0	0.012	0.023	0	-0.0003	0.010	0.121	0
Reject?	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail

Table 1: Realizations of the test statistic for Model 2 compared to Model 1.

population decline occur between observations, are given in Table 2. We reject the model without mortality with 99.9% confidence in all but Field OC. We conclude that in Fields JC, JO, KO, OO, and SC, the data supports the estimation of some additional mortality event in the aphid population.

Field	JC	JO	KO	OC	OO	SC
$\tilde{U}_n$	127.58	54.443	13.246	-0.621	582.07	117.74
Reject?	99.9%	99.9%	99.9%	Fail	99.9%	99.9%

Table 2: Realizations of the test statistic for Model 2 compared to Model 3.

We last assert the null hypothesis that the nominal parameters are contained in the restriction given by Model 4, that is  $\phi = 1$ , instead of Model 2. The results from the model comparison test are given in Table 3, and we reject Model 4 with 75% or higher certainty in fields except JC and JO. In these two exceptions, the estimated value of  $\phi$  is close to 1, and so it is not surprising that the comparison test fails. We conclude that there is sufficient information

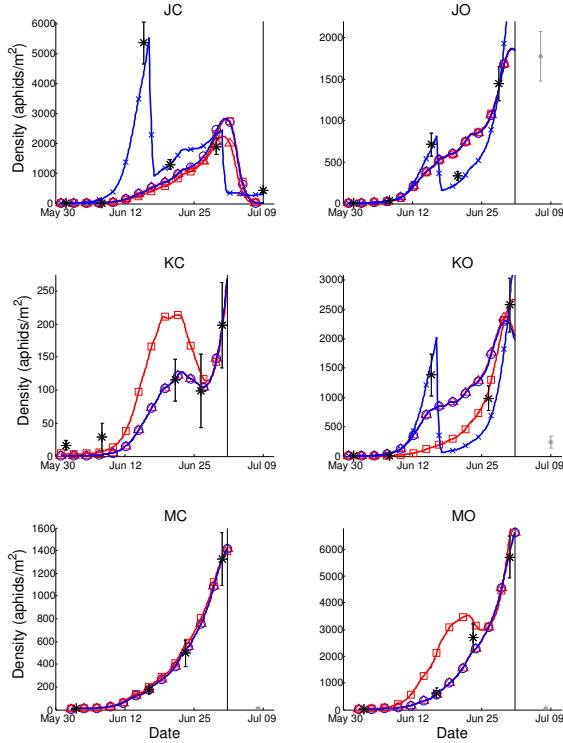


Figure 3: Least squares solutions for Models 1-4 with  $\gamma = 0$ . Data is plotted with a star marker with error bars to indicate the standard deviation in the data and the horizontal line indicates the cutoff date for data to be used in the inverse problem. See Table 5 in the appendix for a list of the parameters used in each field.

in the data set to justify the estimation of  $\phi$  in the ATN model.

### 3.3. ATN solution with additional data

For all fields, we find that the available data supports the estimation of the ATN model with a linear functional response; the inclusion of  $h_0$  and  $c_0$  as unknown parameters in the inverse problem seldom adds value to the estimates. Additionally, the numerical minimization necessary for the inverse problem using the full model is sensitive to the initial iterate supplied to the program. We only find that  $h_0 = c_0 = 0$  by starting sufficiently close to the solution, with initial iterates for each parameter on the order of  $10^{-4}$  or  $10^{-3}$  across fields. The

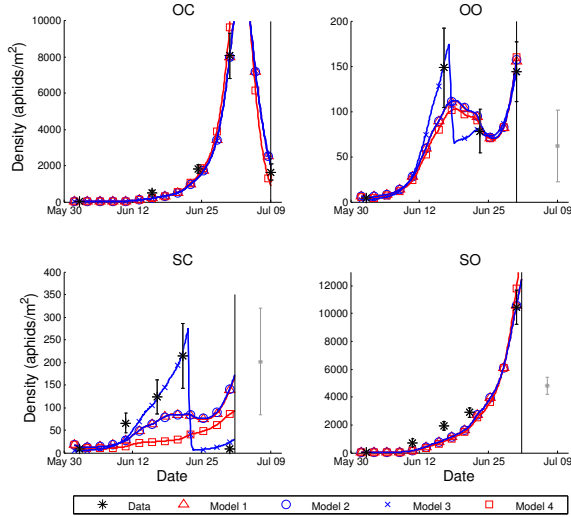


Figure 3: (Continuation) Least squares solutions for Models 1-4 with  $\gamma = 0$ . Data is plotted with a star marker with error bars to indicate the standard deviation in the data and the horizontal line indicates the cutoff date for data to be used in the inverse problem. See Table 5 in the appendix for a list of the parameters used in each field.

Field	JC	JO	KC	KO	MC	MO	OC	OO	SC	SO
$\hat{U}_n$	0.19	0.03	49.35	4.30	17.97	41.34	34.75	1.63	1.58	5.10
Reject?	Fail	Fail	99.9%	95%	99.9%	99.9%	99.9%	75%	75%	95%

Table 3: Realizations of the test statistic for Model 4 compared to Model 2.

physically admissible range for each parameter is  $h_0 \in [0, 0.42]$  and  $c_0 \in [0, 0.24]$  when converted to the units for our system [25]. With the current data, the inverse problem for the full ATN model is too sensitive to initial conditions to be tractable. We conjecture that with data collected at a higher sampling rate, the problem could be solved with less sensitivity to the initial conditions of the numerical minimizer.

We consider this assumption for Field MC, where the aphid population does not experience a mortality event and is well-fit by the inverse problem using the reduced ATN model with initial iterate  $a_0 = .8$ ,  $R_{opt} = 150$ , and  $\phi = 1$ , yielding parameters  $a_0 = 0.4207$ ,  $R_{opt} = 160$ , and  $\phi = 0.8570$ . We construct a synthetic data set with these parameters and solve the inverse problem for the full model, using initial iterate  $a_0 = .4$ ,  $R_{opt} = 160$ , and  $\phi = 0.8$ , with an

increasing number of data points added between each existing observation. The resulting solutions are plotted in Figure 4 and the parameters used to generate the synthetic data, initial iterate for the inverse problem using the synthetic data, and resulting parameter estimates are given in Table 4.

Parameter	$a_0$	$R_{opt}$	$\phi$	$h_0$	$c_0$
Synthetic Value	0.4207	160	0.8570	0	0
Initial Iterate	0.4	160	0.8	0.1	0.1
Estimated Value	0.3459	140	0.7772	0.0005	0.0005

Table 4: Parameter values and initial iterates used in the inverse problem on synthetic data.

The inverse problem for the full ATN model in Field MC requires five additional observations between each existing point to visually fit the data, but the solution does not give the same parameters as we used to generate the synthetic data. This rate of sampling would require population measurements approximately three times a week, where a single measurement of the aphid population requires a sweepnet count of the aphids on 100 barley tillers. This would be a significant undertaking, and even then we do not claim that this rate of data collection would support parameter estimation for the full ATN model in future experiments. We only demonstrate that the current limitations of the inverse problem on the full ATN model can be remedied by increased collection of data; we can not readily generalize these results to a statement about the amount of data necessary to estimate the parameters for the full ATN model in a given study system.

#### 4. Conclusion

The results presented here are an important first step in developing more accurate models for describing the dynamics of species interactions in food webs. Despite recent calls for better integration of food web ecology and ecosystem functioning and services[20], there has been little theory developed to date[20, 16, 17]. Agroecosystems, with their simplified dynamics and species diversity, are an ideal system for testing these models especially for identifying and establishing the conditions under which the parameters for the dynamic models can be estimated with some degree of confidence. The present results show that although the inverse problem was not tractable as originally formulated for the ATN model, this can be ameliorated by reducing the number of parameters to be estimated or increasing data available for least squares

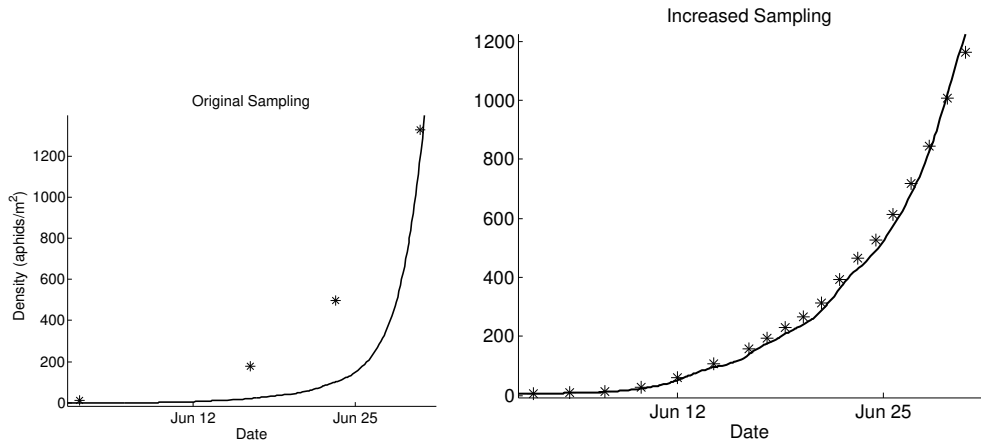


Figure 4: The solution to the inverse problem using the full ATN model with the original data (left) and a synthetic data set with increased sampling (right) is plotted with a solid line. The data used in the inverse problem is plotted with a star marker.

minimization. We find that simplifying the model by fixing some parameters at constant values or generating synthetic data with higher sampling rates allows for parameterization of the ATN model. The results from this paper and ongoing efforts in information content analysis should better-inform future discussions about data collection practices and model formulation for experiments on this study system.

## 5. Acknowledgements

The authors are grateful to Mattias Jonsson for kindly providing the data used in this report. This research was supported in part by the Air Force Office of Scientific Research under grant numbers AFOSR FA9550-12-1-0188 and AFOSR FA9550-15-1-0298, and in part by the National Science Foundation under Research Training Grant (RTG) DMS-1246991. Funding was also provided by the August T. Larsson guest researchers program at the Swedish University of Agricultural Sciences, by the Swedish research council FORMAS, by the Biodiversa-FACCE project “Enhancing biodiversity-based ecosystem services to crops through optimized densities of green infrastructure in agricultural landscapes - ECODEAL,” and the Biodiversa-ERANet project “Assessment and valuation of pest suppression potential through biological control in European

agricultural landscapes - APPEAL.”

### References

- [1] P. Balvanera, A.B. Pfisterer, N. Buchmann, J.S. He, T. Nakashizuka, D. Raffaelli, and B. Schmid. Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecology Letters*, 9:1146–1156, 2006.
- [2] H.T. Banks, J.E. Banks, Riccardo Bommarco, Alva Curtsdotter, Tomas Jonsson, and A.N. Laubmeier. Parameter estimation for an allometric food web model. CRSC Technical Report, CRSC-TR16-03, May 2016.
- [3] H.T. Banks, J.E. Banks, J. Rosenheim, and K. Tillman. Modeling populations of *Lygus Hesperus* on cotton fields in the San Joaquin Valley of California: The importance of statistical and mathematical model choice. CRSC-TR15-04, Center for Research in Scientific Computation, N. C. State University, Raleigh, NC, May, 2015; *J. Biological Dynamics*, 11:25–39, 2017. DOI: 10.1080/17513758.2016.1143533.
- [4] H.T. Banks, J.E. Banks, Neha Murad, J. A Rosenheim, and K. Tillman. Modelling pesticide treatment effects on *Lygus hesperus* in cotton fields. CRSC-TR15-09, Center for Research in Scientific Computation, N. C. State University, Raleigh, NC, September, 2015; *Proceedings, 27 th IFIP TC7 Conference 2015 on System Modelling and Optimization*, L. Bociu et al (Eds.) CSMO 2015 IFIP AICT 494:1–12, 2017. Springer : DOI: 10.1007/978-3-319-55795-3-8.
- [5] H.T. Banks, E. Collins, K. Flores, P. Pershad, M. Stemkovski, and L. Stephenson. Statistical error model comparison for logistic growth of green algae (*Raphidocelis subcapitata*). CRSC-TR16-10, Center for Research in Scientific Computation, N. C. State University, Raleigh, NC, August, 2016; *Applied Mathematical Letters*, 64:213–222, 2017.
- [6] H.T. Banks, R.A. Everett, S. Hu, N. Murad, and H.T. Tran. Mathematical and statistical model misspecifications in modeling immune response in renal transplant recipients. CRSC-TR16-14, Center for Research in Scientific Computation, N. C. State University, Raleigh, NC, November, 2016; *Inverse Problems in Science and Engineering*, to appear.
- [7] H.T. Banks, Jared Catenacci, and Shuhua Hu. Use of difference-based methods to explore statistical and mathematical model discrepancy in inverse problems. *Journal of Inverse and Ill-posed Problems*, 23:491–509, 2015.
- [8] H.T. Banks and B.G. Fitzpatrick. Statistical methods for model comparison in parameter estimation problems for distributed systems. *Journal of Mathematical Biology*, 28:501–527, 1990.
- [9] H.T. Banks, S. Hu, and W.C. Thompson. *Modeling and Inverse Problems in the Presence of Uncertainty*. CRC Press, Boca Raton, 2014.
- [10] H.T. Banks and H.T. Tran. *Mathematical and Experimental Modeling of Physical and Biological Processes*. CRC Press, New York, 2009.
- [11] Eric L. Berlow, Jennifer A. Dunne, Neo D. Martinez, Phillip B. Stark, Richard J. Williams, Ulrich Brose, and Simon A. Levin. Simple prediction of interaction strengths in complex food webs. *Proceedings of the National Academy of Sciences of the United States of America*, 106:187–191, 2009.

- [12] Alice Boit, Neo D. Martinez, Richard J. Williams, and Ursula Gaedke. Mechanistic theory and modelling of complex food-web dynamics in Lake Constance. *Ecology Letters*, 15:594–602, 2012.
- [13] U. Brose, R.B. Ehnes, B.C. Rall, O. Vucic-Pestic, E.L. Berlow, and S. Scheu. Foraging theory predicts predator-prey energy fluxes. *Journal of Animal Ecology*, 77:1072–1078, 2008.
- [14] Ulrich Brose, Richard J. Williams, and Neo D. Martinez. Allometric scaling enhances stability in complex food webs. *Ecology Letters*, 9:1228–1236, 2006.
- [15] James H. Brown, James F. Gillooly, Andrew P. Allen, Van M. Savage, and Geoffrey B. West. Toward a metabolic theory of ecology. *Ecology*, 85:1771–1789, 2004.
- [16] B. Cardinale, E. Duffy, S. Srivastava, M. Loreau, M. Thomas, and M. Emmerson. Towards a food web perspective on biodiversity and ecosystem functioning. In S. Naeem, D.E. Bunker, A. Hector, M. Loreau, and C. Perrings, editors, *Biodiversity, Ecosystem Functioning, and Human Wellbeing*, chapter 8, pages 105–120. Oxford University Press, Oxford, 2009.
- [17] B.J. Cardinale, J.E. Duffy, A. Gonzalez, C. Hooper, D.U. Perrings, P. Venail, A. Narwani, G.M. Mace, D. Tilman, D.A. Wardle, A.P. Kinzig, G.C. Daily, M. Loreau, J.B. Grace, A. Larigauderie, D.S. Srivastava, and S. Naeem. Biodiversity loss and its impact on humanity. *Nature*, 486:59–67, 2012.
- [18] G.C. Chang. Comparison of single versus multiple species of generalist predators for biological control. *Environmental Ecology*, 25:207–212, 1996.
- [19] Brian Dennis, José Miguel Ponciano, Subhash R. Lele, Mark L. Taper, and David F. Staples. Estimating density dependence, process noise, and observation error. *Ecological Monographs*, 76:323–341, 2006.
- [20] J.E. Duffy, B.J. Cardinale, K.E. France, P.B. McIntyre, E. Thebault, and M. Loreau. The functional role of biodiversity in ecosystems: incorporating trophic complexity. *Ecology Letters*, 10:522–538, 2007.
- [21] C.S. Elton. *The Ecology of Invasions by Plants and Animals*, Volume 18. Methuen, London, 1958.
- [22] Tessa B. Francis, Elizabeth M. Wolkovich, Mark D. Scheuerell, Stephen L. Katz, Elizabeth E. Holmes, and Stephanie E. Hampton. Shifting regimes and changing interactions in the Lake Washington, U.S.A., plankton community from 1962–1994. *PLOS ONE*, 9, 2014.
- [23] Stephanie E. Hampton, Elizabeth E. Holmes, Lindsay P. Scheef, Mark D. Scheuerell, Stephen L. Katz, Daniel E. Pendleton, and Eric J. Ward. Quantifying effects of abiotic and biotic drivers on community dynamics with multivariate autoregressive (MAR) models. *Ecology*, 94:2663–2669, 2013.
- [24] A.R. Ives, B. Dennis, K.L. Cottingham, and S.R. Carpenter. Estimating community stability and ecological interactions from time-series data. *Ecological Monographs*, 73:301–330, 2003.
- [25] Riikka Kaartinen, Tomas Jonsson, Mattias Jonsson, and Riccardo Bommarco. Intraguild interactions modify the effect of body size on trophic interactions. Unpublished manuscript.



- [26] D.K. Letourneau, J.A. Jedlicka, S.G. Bothwell, and C.R. Moreno. Effects of natural enemy biodiversity on the suppression of arthropod herbivores in terrestrial ecosystems. *Annual Review of Ecology, Evolution, and Systematics*, 40:573–592, 2009.
- [27] M. Loreau, S. Naeem, P. Inchausti, J. Bengtsson, J.P. Grime, A. Hector, D.U. Hooper, M.A. Huston, D. Raffaelli, B. Schmid, and D. Tilman. Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science*, 294:804–808, 2001.
- [28] J.E. Losey and R.F. Denno. Positive predator-predator interactions: enhanced predation rates and synergistic suppression of aphid populations. *Ecology*, 79:2143–2152, 1998.
- [29] G.M. Mace, K. Norris, and A.H. Fitter. Biodiversity and ecosystem services: a multilayered relationship. *Trends in Ecology and Evolution*, 27:19–26, 2012.
- [30] R.M. May. *Stability and Complexity in Model Ecosystems*. Volume 6. Princeton University Press, 1973.
- [31] Shahid Naeem, Daniel E. Bunker, Andy Hector, Michel Loreau and Charles Perrings, Eds. *Biodiversity, Ecosystem Functioning, and Human Wellbeing: An Ecological and Economic Perspective*. Oxford University Press, 2009.
- [32] B.C. Rall, G. Kalinkat, O. Vucic-Pestic, and U. Brose. Taxonomic versus allometric constraints on non-linear interaction strengths. *Oikos*, 120:483–492, 2011.
- [33] Florian Dirk Schneider, Stefan Scheu, and Ulrich Brose. Body mass constraints on feeding rates determine the consequences of predator loss. *Ecology Letters*, 15:436–443, 2012.
- [34] W.E. Snyder and A.R. Ives. Interactions between specialist and generalist natural enemies: parasitoids, predators, and pea aphid biocontrol. *Ecology*, 84:91–107, 2003.
- [35] E. Thebault and M. Loreau. Food-web constraints on biodiversity - ecosystem functioning relationships. *Proceedings of the National Academy of Sciences*, 100:14949–14954, 2003.
- [36] D. Tilman. The ecological consequences of changes in biodiversity: a search for general principles. *Ecology*, 80:1455–1474, 1999.
- [37] P. Yodzis and S. Innes. Body size and consumer-resource dynamics. *The American Naturalist*, 139:1151–1175, 1992.

### Appendix

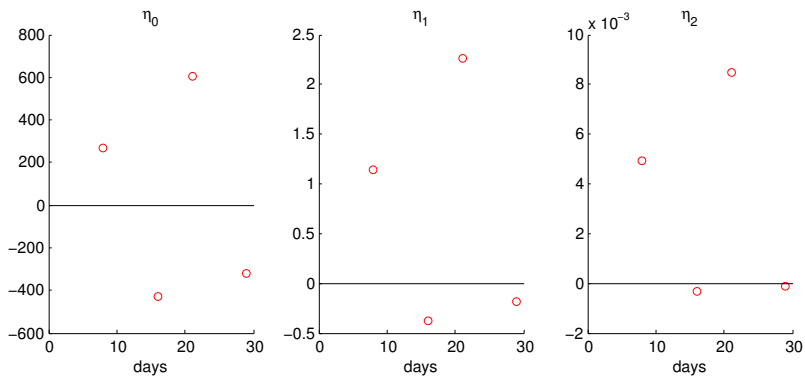


Figure 5: Values of  $\eta_{j,\hat{\gamma}}$  are plotted against time for  $\hat{\gamma} = 0, 1,$  and  $2$  for data from field JO.

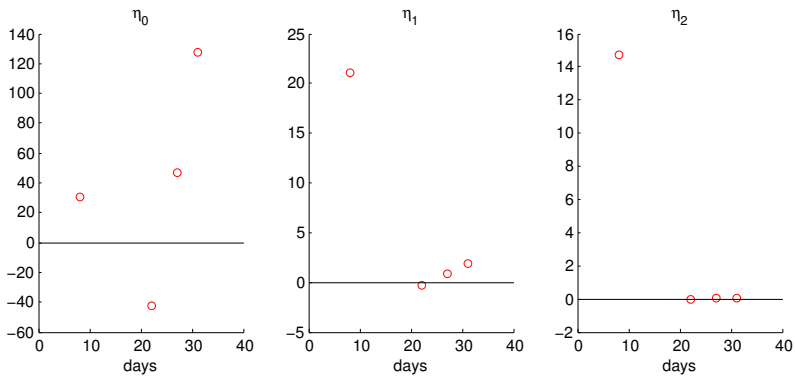


Figure 6: Values of  $\eta_{j,\hat{\gamma}}$  are plotted against time for  $\hat{\gamma} = 0, 1,$  and  $2$  for data from field KC.

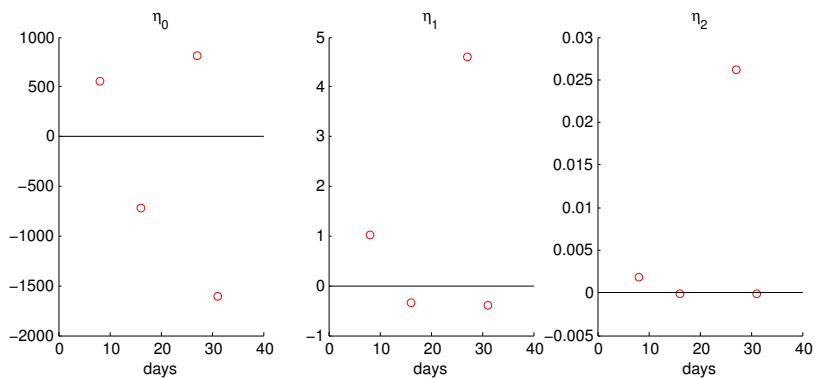


Figure 7: Values of  $\eta_{j,\hat{\gamma}}$  are plotted against time for  $\hat{\gamma} = 0, 1,$  and  $2$  for data from field KO.

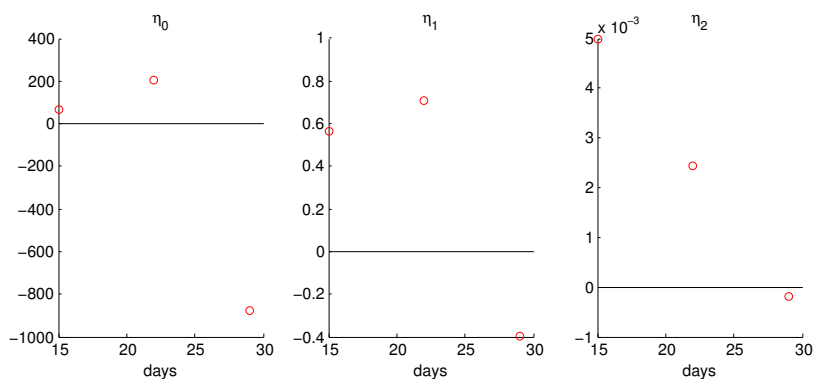


Figure 8: Values of  $\eta_{j,\hat{\gamma}}$  are plotted against time for  $\hat{\gamma} = 0, 1,$  and  $2$  for data from field MC.

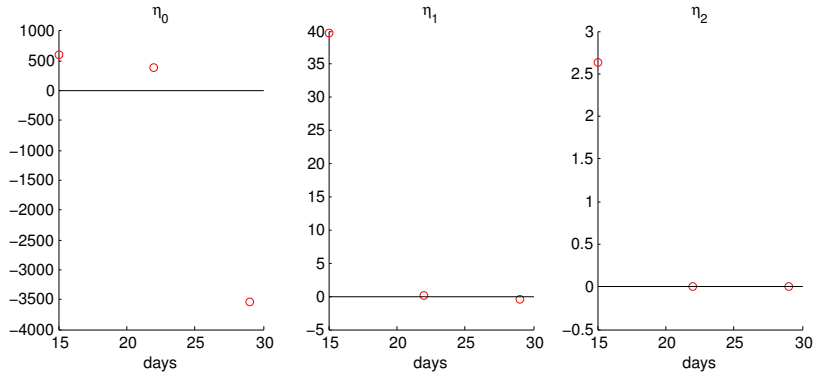


Figure 9: Values of  $\eta_{j,\hat{\gamma}}$  are plotted against time for  $\hat{\gamma} = 0, 1, \text{ and } 2$  for data from field MO.

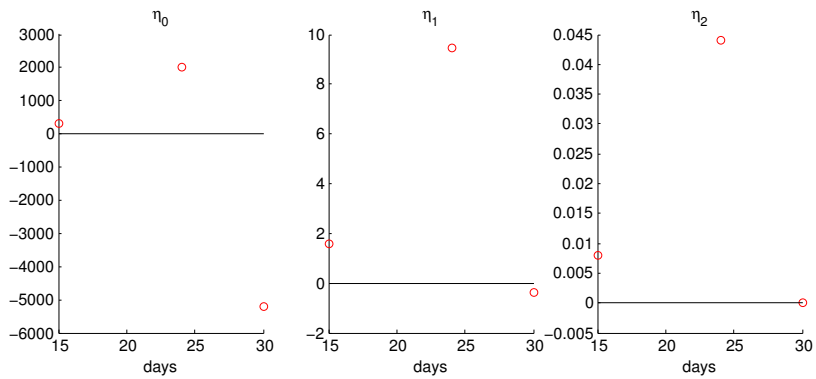


Figure 10: Values of  $\eta_{j,\hat{\gamma}}$  are plotted against time for  $\hat{\gamma} = 0, 1, \text{ and } 2$  for data from field OC.

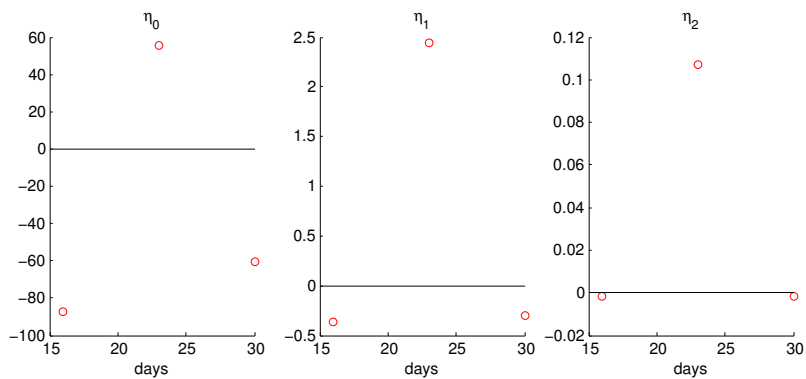


Figure 11: Values of  $\eta_{j,\hat{\gamma}}$  are plotted against time for  $\hat{\gamma} = 0, 1,$  and  $2$  for data from field OO.

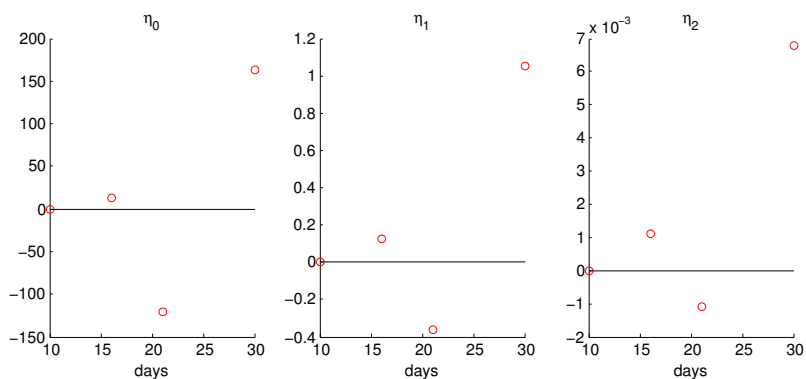


Figure 12: Values of  $\eta_{j,\hat{\gamma}}$  are plotted against time for  $\hat{\gamma} = 0, 1,$  and  $2$  for data from field SC.

Field	Model	$a_0$	$R_{opt}$	$\phi$	$h_0$	$c_0$	$\mu$	$N_0^1$
JC	Model 1	3.10	387.37	1.07	0	0	-	7.43
	Model 2	2.97	515.64	0.96	-	-	-	7.30
	Model 3	3.69	0.54	1.01	-	-	1.98	9.99
	Model 4	2.56	428.92	-	-	-	-	4.62
JO	Model 1	2.57	750.28	0.94	0	0	-	3.13
	Model 2	2.58	750.31	0.94	-	-	-	3.13
	Model 3	2.52	823.56	1.20	-	-	1.84	1.46
	Model 4	3.04	803.08	-	-	-	-	2.86
KC	Model 1	0.81	184.75	1.98	0	0	-	0.94
	Model 2	0.81	184.75	1.98	-	-	-	0.94
	Model 3	-	-	-	-	-	-	-
	Model 4	0.83	518.41	-	-	-	-	3.93
KO	Model 1	0.12	37.49	1.43	0	0	-	1.71
	Model 2	0.12	38.73	1.52	-	-	-	1.71
	Model 3	0.02	24.93	0.05	-	-	3.68	1.64
	Model 4	0.88	0.87	-	-	-	-	2.90
MC	Model 1	0.42	160.09	0.86	0	0	-	7.00
	Model 2	0.42	160.09	0.86	-	-	-	7.00
	Model 3	-	-	-	-	-	-	-
	Model 4	0.51	161.19	-	-	-	-	7.61
MO	Model 1	0.16	1000	0.22	0	0	-	18.18
	Model 2	0.16	1000	0.22	-	-	-	18.18
	Model 3	-	-	-	-	-	-	-
	Model 4	0.52	0.90	-	-	-	-	18.17
OC	Model 1	0.18	46.86	0.21	0	0	-	1.45
	Model 2	0.18	46.86	0.21	-	-	-	1.45
	Model 3	0.18	46.75	0.21	-	-	0.01	1.41
	Model 4	0.29	37.78	-	-	-	-	1.00
OO	Model 1	1.15	0.75	0.99	0	0	-	6.02
	Model 2	1.16	0.75	0.99	-	-	-	6.01
	Model 3	0.18	2.12	0.40	-	-	1.08	2.32
	Model 4	0.95	0.81	-	-	-	-	4.86
SC	Model 1	0.37	0.73	0.41	0	0	-	18.18
	Model 2	0.34	0.70	0.35	-	-	-	18.18
	Model 3	2.07	478.73	2.00	-	-	4.01	4.94
	Model 4	0.57	146.55	-	-	-	-	18.19
SO	Model 1	3.30	239.33	1.57	0	0	-	5.01
	Model 2	3.30	239.33	1.57	-	-	-	5.01
	Model 3	-	-	-	-	-	-	-
	Model 4	1.68	227.45	-	-	-	-	5.00

Table 5: Estimated parameters for Models 1, 2, 3, and 4 over the ten fields. In Fields KC, MC, MO, and SO, the data did not meet the requirements to estimate the parameters for Model 3.