ABSTRACT

LAUBMEIER, AMANDA NICOLE. A Model-Driven Approach to Experimental Validation of Predator-Prey Dynamics in a System of Terrestrial Arthropods. (Under the direction of H. T. Banks.)

The application of mathematical models to ecological communities in order to describe behavior under varying conditions is fundamental to the field of predictive ecology. Although the data-driven validation of population models is well-practiced, it is an unfortunate reality that many ecological study systems are uncontrolled as well as spatially and temporally expansive. Observing these systems is costly and time intensive, and the information obtained can be messy or driven by unexpected factors. We are therefore interested in the effective validation of mathematical models in population ecology

In this dissertation, we specifically study predator-prey interactions between arthropods (aphids, beetles, and spiders) in an agricultural field. Our interest stems from the potential for the biological control of aphids, an agricultural pest, by natural predators. Although this focus is fairly specific, we still expect to encounter diverse populations of predators in natural communities. We therefore rely on a generalizable model, in which interactions are parameterized by measurable species traits. To evaluate the viability of this model in describing natural communities, we conduct least-squares parameter estimation and fit the model to empirical data.

We begin by introducing the methodology for parameter estimation and motivating our choice of mathematical model. We then detail initial attempts at fitting our model to aphid abundance data in barley fields. The data used in this first attempt at model validation was opportunistically obtained and primarily intended as a study of community complexity instead of temporal dynamics. We demonstrate that insufficient sampling in time, along with mortality driven by unknown factors, prevents the estimation of model parameters with any degree of certainty.

Motivated by the shortcomings of this first attempt, we move on to the design of greenhouse experiments for model validation. We explicitly link the protocol for our experiment to the anticipated mathematical and statistical properties of our parameter estimation problem. We consider the time-dependent sensitivity of the model to parameters in determining the temporal mesh for data collection and the population-level observation error in determining feasible sampling strategies. We attempt to validate our model against the resulting experimental data, with marked improvement over our earlier work.

Returning to our initial question, we again consider the problem of validating our model in an agricultural field. Although we do not have sufficient data to estimate model parameters in the field, we motivate a simple partial differential equation model for predator-prey interactions. We investigate the effect of temporal and spatial sampling protocols on the estimation of model parameters over a single growing season.

A Model-Driven Approach to Experimental Validation of Predator-Prey Dynamics in a System of Terrestrial Arthropods

by
Amanda Nicole Laubmeier

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Applied Mathematics

Raleigh, North Carolina

2018

APPROVED BY:

Lorena Bociu	Kevin Flores
Hien Tran	H. T. Banks Chair of Advisory Committee

DEDICATION

For we put the thought of all that we love
into all that we make. – J. R. R. Tolkien

For my obaachan, who is early mornings and
learning to count and drops of pancake batter.

After that, to everyone who has joined me on this journey,
with my warmest wishes to each of you for the road ahead.

BIOGRAPHY

Amanda Laubmeier was born in Mesa, Arizona on July 12, 1992 and grew up in the valley of the sun. She received her B.S. in Mathematics from the University of Arizona in 2014, with a minor in Physics. That same year, she began her graduate studies at North Carolina State University in Applied Mathematics with doctoral advisor H. T. Banks.

ACKNOWLEDGEMENTS

I would first like to acknowledge the enormity of the contributions by my advisor H. T. Banks and my collaborators John Banks, Riccardo Bommarco, Alva Curtsdotter, Tomas Jonnson, Tomas Roslin, and Kate Wootton. I've been told that it is "better to be lucky than smart, rich, or handsome," and I have been inordinately lucky in my graduate experience. I don't think I could enumerate the things that I have learned from these insightful individuals, and I am immensely grateful to have worked alongside them.

In addition to my collaborators, I owe great thanks to the individuals who made this research possible. Mattias Jonsson and Michael Traugott provided the data studied in Chapter 3. Eve Roubinet provided input on molecular gut content analysis in the same chapter. Riika Kaartinen's previous experiments were invaluable in forming the basis of our work in Chapter 4. Ged Malsher provided data and insight for the design of our experiments in the same chapter.

The team that supported our experiment in Chapter 5 also demonstrated to me the extraordinary commitment required to generate a single data set, and I extend my sincere thanks for their contributions. Ged Malsher, Carol Högfeldt, Carly Miranda, Joe Bliss, Maylis Moro, Elin Ljunggren, Josefin Farrer-Sundberg, Josephine Biro, Lina Wu, and Kevin Cestrierest assisted in most every aspect of the experiment. Sophie Högfeldt, Ola Lundin, and Kirsten Miller also assisted in data collection. Michelle Nordkvist helped collect experimental subjects and Michail Artamonov built the cages which housed our experiments.

For their time and input, I would also like to acknowledge my committee members. In particular, I am thankful to Lorena Bociu and Kevin Flores for offering professional guidance during my job search. I continually extend my gratitude back to my undergraduate mentors - Moysey Brio, Anant Godbole, Karen Rios-Soto, and Bill Velez - who pushed me on to a graduate career.

I would finally like to extend my personal acknowledgements to friends and family. Thank you, Nancy, for keeping me laughing through the worst of it. Thank you, Eric, for your unfaltering positivity. Thank you, Kaska, for showing me the ropes. Thank you, Michael, for talking through every detail. Thank you, Rebecca, for more advice than I could describe here and almost as much fun. Thank you, Nick, for going halvesies on the food and frustration. Thank you, Terry, for selflessly drinking me under the table. Thank you, mom, for your unconditional belief that I could make it. Thank you, dad, for always helping me find my feet.

My graduate work is supported by the National Science Foundation under Research Training Grant DMS-1246991. This collaborative research would not have been possible without support for travel and exchange of ideas, provided by the August T. Larsson Guest Researcher Programme and by a CRSC Fellowship. Additionally, travel to conferences for presentation of ideas was supported by the Leroy B. Martin Fund.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

INTRODUCTION

Faced with shifting global landscapes, there is immediate need for models which describe the effect of changing ecological communities on ecosystem services, either because of species loss or migration [27, 90]. Such models must be generalizable, capable of being applied to communities drastically different from those available for study at present. However, the development and validation of these models is a challenging, and often iterative, process. From empirical observations, we must identify mechanisms which drive community dynamics and formalize them with a mathematical model. From an assumed model for community dynamics, we must analyze predicted behaviors and compare them with empirical observations. In this dissertation, we study a trait-based model for predator-prey dynamics in a system of terrestrial arthropods and the validation of this model against experimental data regarding biological control ecosystem services.

In this chapter, we describe the ecological motivation for this work as well as the mathematical approach we take in pursuing these questions. Chapter 2 outlines the models which we rely on in studying our problem. In Chapter 3, we present preliminary efforts at validating these models against data from an agricultural field. Chapter 4 outlines the design of greenhouse experiments for model validation, and in Chapter 5 we analyze the results from experiments following this design. In Chapter 6, we consider the design of field-level experiments through the use of a spatial model. We offer some concluding remarks and discuss plans for future work in Chapter 7.

1.1 Motivation

Ecosystem services are products of ecological interactions which provide benefits to humans [5, 6], such as crop pollination or water purification. In particular, biological control of pests is an agricultural service driven by predator-prey interactions beween insects in an agro-ecosystem. The value of this service in the US is estimated to be 4.5 billion dollars annually [72], and so there is some interest in farming practices which encourage such interactions [23, 113]. Conventional practices can adversely affect the biodiversity of predator communities in these ecosystems [104, 106, 107, 114], but the strength of predator-prey interactions can be affected in conflicting ways by predator composition [40, 71]. To evaluate the merits of different farming practices, we must quantify the effect of predator diversity on pest control in these systems.

Biological control services rely on the interactions between many species in an ecosystem, and the cumulative effect of these interactions on a pest are complicated. Interactions between biological control agents might be complementary or adversarial [86, 91, 103], so we expect that a community of agents will seldom have an additive effect on the target pest. However, we cannot currently predict which of these contributions will drive biological control as a community [27]. Empirically determining the strength and preference of species interactions in true ecological communities can reveal information about the degree of cooperation or conflict between predators [93, 94], but we must understand the mechanisms underlying these interactions to generalize the expected effects in different communities. Dynamic models for predator-prey interactions at a population level present the opportunity to unravel these effects.

Population models are a well-established tool for the theoretical study of ecosystems, but their true potential lies in the description of real-world dynamics. Recent work has demonstrated the utility of mathematical models in describing the dynamics of ecological foodwebs, in both experimental and natural systems [20, 62, 101]. In such studies, the mathematical formalization of system interactions presents multiple benefits. First, the validation of a mathematical model informs future theoretical study, directly building towards our goal to quantify the ecological impacts of unknown environmental drivers. However, even the shortcomings of a model can provide insight into the aptness of the biological assumptions which informed the model. This information allows for the refinement of our assumed models for system behavior, revealing the conditions under which their underlying descriptions are appropriate. Within this framework, model generalizability is a desirable characteristic, permitting the evaluation of existing models in a variety of ecosystems.

We consider models built on the assumption that population-level interactions are dictated by species traits. Trait-based models rely on descriptions of the biological processes driving species interactions, such as metabolic demands or physical defenses [25, 63, 82, 119]. These assumptions enable us to model arbitrarily large ecosystems using only assumed system traits and a handful of simplifying parameterizations. Such models have been successful in predicting the structure and

dynamics of ecological communities [25, 44, 87], and our eventual goal is to use the same models in judging how these communities might respond to changes in species composition. However, establishing the mechanistic relationships between species traits and community dynamics presents a modelling challenge, in which we must identify traits that are responsible for governing most species interactions and characterize the relationships between these traits and system dynamics.

Empirically obtaining information about these relationships is difficult, possibly requiring multiple observations of predators in experiments targeted at identifying the imact of species traits on these interactions. We instead rely on parameter estimation by least squares minimization [10, 11] to fit these models to empirical observations. In addition to parameterizing the relationships in the model, the analysis of model fits to data provides information about which mechanisms have the strongest effect on populations. As an alternative to fixed-time comparisons between model performance and empirical observations (see, for example, [62, 101]), least squares fitting presents the opportunity for us to evaluate model performance in time. This is a desirable characteristic in ecosystems, where we expect that the factors which drive a population might vary temporally. Additionally, this method of parameter estimation is well-established, with a supporting framework for quantification of uncertainty in parameter estimates and comparisons of model performance.

The work presented in this dissertation is a part of the iterative modelling process [10], in which our attempts to fit models to empirical data inform our future model formulations. Given some empirical data, we estimate parameters which minimize the difference between real dynamics and the proposed model. However, initial attempts at solving such problems naturally lead to questions about the accuracy and information content of empirical data; in evaluating the resulting solutions, we might also find that our mathematical model must be reformulated to accomodate unexpected behaviors. These demands require us to reevaluate the biological assumptions which drove our initial model formulation. Upon revisiting these biological assumptions, we must reformulate our mathematical model to match our new understanding of the system. Eventually, we return to the problem of fitting this new model to observational data; in our work, we pay particular attention to the importance of our model and mathematical methodology in determining appropriate empirical experiments for obtaining this data.

In the remainder of this chapter, we describe the inverse problem methodology as it is used in parameter estimation and model validation. The process of parameter estimation only forms a small part of the aforementioned iterative modelling process. However, since it is the means by which we compare mathematical models with empirical observations, an understanding of the methodology is important in our design of experiments and analysis of model performance.

1.2 Inverse Problem Methodology

Throughout this dissertation, we return to least squares parameter estimation as the means by which we validate a model against experimental data. Although we are interested in the "forward" solution of a model for predator-prey dynamics which describes the effect predator biodiversity has on biological control, we must first solve the "inverse" problem which parameterizes such a model. In this section, we explicitly state the inverse problem for a vector dynamical system, assuming a weighted error statistical model. This problem is given for consistent samples across all states in [8] and alluded to in [10, 11]. We define the problem for inconsistent sampling across system states and outline the algorithm for its solution, as given in [10, 11]. We follow with the corresponding asymptotic approximations for standard errors in our inverse problem solutions and framework for nested model comparison testing.

1.2.1 Iterative Weighted Least Squares: Vector Case

We consider the inverse problem for some vector system given by the mathematical model

$$\frac{d\mathbf{x}}{dt} = \mathbf{g}(t, \mathbf{x}(t), \mathbf{q}), \tag{1.1}$$

with initial condition $\mathbf{x}(t_0) = \mathbf{x}_0$ for $\mathbf{x} \in \mathbb{R}^n$. We denote the solution of this system for a particular parameter set by $\mathbf{x}(t;\theta)$, for $\theta = [\mathbf{q}, \mathbf{x}_0^T]^T$. We assume that there is some observation process

$$\mathbf{f}(t;\theta) = \mathscr{K}(t)\mathbf{x}(t;\theta)$$

such that the time-dependent observation operator $\mathscr{K}(t)$ maps a state $\mathbf{x}(t;\theta) \in \mathbb{R}^n$ to some subset of observed states in $\mathbb{R}^{m(t)}$, where $m(t) \leq n$.

We additionally assume that data from these observations are drawn from realizations of the statistical model,

$$\mathbf{Y}_j = \mathbf{f}\left(t_j;\theta_0\right) + \mathbf{f}^{\gamma}\left(t_j;\theta_0\right) \circ \mathscr{E}_j, \tag{1.2}$$

for $j = 1, \ldots, N$ indexing the times at which data is collected and θ_0 the nominal parameter set which describes the system's underlying dynamics. We note that $\mathbf{f}^{\gamma}\left(t_j;\theta_0\right)$ represents componentwise exponentiation, for which each entry of $\mathbf{f}\left(t_j;\theta_0\right)$ is raised to the power of the corresponding entry in the vector γ, and that the operation $\alpha \circ \beta$ indicates componentwise multiplication between vectors α and β. We assume that $\mathscr{E}_j \in \mathbb{R}^{m(t_j)}$ are random variables corresponding to noise in the data, normally distributed with variance matrix $V_j = \text{diag}\left(\sigma^2_{j,1}, \ldots, \sigma^2_{j,m(t_j)}\right)$. We note that the variance $\sigma^2_{j,k}$ corresponding to the noise in observations of state k at time t_j is constant. However, for some time $t_{\tilde{j}}$, the same state might be at index $\tilde{k}$ in the vector of observations, since our observation operator

is time-dependent. We will therefore have $\sigma^2_{j,k} = \sigma^2_{\tilde{j},\tilde{k}}$, but the variance matrix V_j is necessarily time-dependent to accomodate potential re-indexing of states. For realizations ϵ_j of the noise, the data collected, $\{\mathbf{y}_j\}_{j=1}^N$, is given by

$$\mathbf{y}_j = \mathbf{f}\big(t_j;\theta_0\big) + \mathbf{f}^{\gamma}\big(t_j;\theta_0\big) \circ \epsilon_j. \tag{1.3}$$

We define an estimator Θ for the solution of the inverse problem by

$$\Theta = \arg\min_{\theta\in\Omega_\theta} \sum_{j=1}^{N} \big[\mathbf{Y}_j - \mathbf{f}\big(t_j;\theta\big)\big]^T \big[W_j V_j\big]^{-1} \big[\mathbf{Y}_j - \mathbf{f}\big(t_j;\theta\big)\big], \tag{1.4}$$

where Ω_θ is the space of admissible parameters and W_j is a matrix of weights given by

$$W_j = \mathrm{diag}\big[\mathbf{f}^{\gamma}\big(t_j;\theta_0\big)\big].$$

For V_j the variance matrix of $\mathcal{E}_j$, we therefore have

$$W_j V_j = \mathrm{diag}\Big(f^{2\gamma_1}\big(t_j;\theta_0\big)\sigma^2_{j,1}, \ldots, f^{2\gamma_{m(t_j)}}\big(t_j;\theta_0\big)\sigma^2_{j,m(_j)}\Big). \tag{1.5}$$

Then the parameter estimate, $\hat{\theta}$, corresponding to realizations of the data given in (1.3) is

$$\hat{\theta} = \arg\min_{\theta\in\Omega_\theta} \sum_{j=1}^{N} \big[\mathbf{y}_j - \mathbf{f}\big(t_j;\theta\big)\big]^T \big[W_j V_j\big]^{-1} \big[\mathbf{y}_j - \mathbf{f}\big(t_j;\theta\big)\big].$$

We note that when seeking this minimization, the true values of the diagonal entries in W_j and V_j are unknown. We employ an asymptotic approximation for the entries of the variance matrix corresponding to this same realization of the data,

$$\hat{V}_j(\hat{\theta}_{LS}) = \mathrm{diag}\left(\frac{1}{\kappa_N - \kappa_\theta} \sum_{j=1}^{N} \big[\mathbf{y}_j - \mathbf{f}\big(t_j;\hat{\theta}_{LS}\big)\big]^T \hat{W}_j^{-1} \big[\mathbf{y}_j - \mathbf{f}\big(t_j;\hat{\theta}_{LS}\big)\big]\right), \tag{1.6}$$

for κ_N the number of data points in $\{\mathbf{y}_j\}_{j=1}^N$ and κ_θ the number of parameters estimated in the vector θ. We define

$$\hat{W}_j(\hat{\theta}_{LS}) = \mathrm{diag}\Big(f^{2\gamma_1}\big(t_j;\hat{\theta}_{LS}\big), \ldots, f^{2\gamma_{m(t_j)}}\big(t_j;\hat{\theta}_{LS}\big)\Big) \tag{1.7}$$

and obtain the corresponding parameter estimate

$$\hat{\theta}_{LS} = \arg\min_{\theta\in\Omega_\theta} \sum_{j=1}^{N} \big[\mathbf{y}_j - \mathbf{f}\big(t_j;\theta\big)\big]^T \big[\hat{W}_j \hat{V}_j\big]^{-1} \big[\mathbf{y}_j - \mathbf{f}\big(t_j;\theta\big)\big]. \tag{1.8}$$

However, estimating $\hat{\theta}_{LS}$ requires that we simultaneously solve a coupled system of equations, (1.8) and (1.6). We therefore numerically approximate the parameter in the following iterative process:

1. Define $\hat{W}_j^{(0)} = \hat{V}_j^{(0)} = I_{m(t_j)\times m(t_j)}$ for all j.
2. Obtain the first estimate, $\hat{\theta}_{LS}^{(0)}$, using (1.8) and the matrices $\hat{W}_j^{(0)}$, $\hat{V}_j^{(0)}$.
3. Set $k = 0$.
4. Compute the matrix $\hat{W}_j^{(k)}$ using (1.7) and the parameter estimate $\hat{\theta}_{LS}^{(k)}$.
5. Compute the matrix $\hat{V}_j^{(k)}$ using (1.6) and the parameter estimate $\hat{\theta}_{LS}^{(k)}$.
6. Estimate $\hat{\theta}_{LS}^{(k+1)}$ using (1.8) and the matrices $\hat{V}_j^{(k)}$, $\hat{W}_j^{(k)}$.
7. Set $k = k+1$ and return to step 4. Continue iterating until successive estimates $\hat{\theta}^{(k)}$ and $\hat{\theta}^{(k+1)}$ are sufficiently close, and then set $\hat{\theta}_{LS} = \hat{\theta}_{LS}^{(k+1)}$.

1.2.2 Identifying the Statistical Model

We note that in order to perform the inverse problem as outlined above, we must determine an appropriate statistical model for error in the data. Failure to employ the correct statistical model in solving an inverse problem can result in erroneous conclusions and asymptotic approximations. We assume a relative error model, as defined in (1.2), but the value of γ_k for each state k must be estimated prior to solving the inverse problem. The value of γ_k can be checked against residual plots after solving an inverse problem, or it can be determined by directly applying a differencing scheme to the data before solving an inverse problem. Both methods are appropriate under different circumstances, and we outline their use below.

Residual plots are a traditional tool for identifying mis-specified statistical models, and their utility is described in detail by [10, 11]. We recall from the statistical model that the error for state k at time t_j is given by

$$\mathscr{E}_{j,k} = \frac{Y_{j,k} - f_k(t_j;\theta_0)}{f_k^{\gamma_k}(t_j;\theta_0)}.$$

We additionally assumed that for any time, this error is normally distributed with mean zero and variance σ_k^2. We can verify that these assumed statistical properties hold by examining the modified residuals,

$$r_{j,k}(\gamma_k) = \frac{y_j - f_k(t_j;\hat{\theta}_{LS})}{f_k^{\gamma_k}(t_j;\hat{\theta}_{LS})}.$$

We note that this necessarily requires that we solve the inverse problem for the parameter estimate $\hat{\theta}_{LS}$ using some assumed vector γ in the statistical error model.

To test this choice of γ, we plot the values of $r_{j,k}(\gamma_k)$ against time t_j or the model evaluated at this time, $f_k(t_j;\hat{\theta}_{LS})$. Given the assumed distribution of $\mathcal{E}_{j,k}$, either of these plots should be random. A characteristic "fan" shape in the residuals might be observed if the inverse problem relies on an inappropriate value of γ_k. In Figure 1.1, we demonstrate the "random" behavior we expect in a residual plot, followed by two examples of residual plots with a fan-shaped spread. In practice, if we solve an inverse problem but obtain residuals with this fan shape, then we must revise the assumed value of γ_k until residual plots exhibit the desired random behavior.

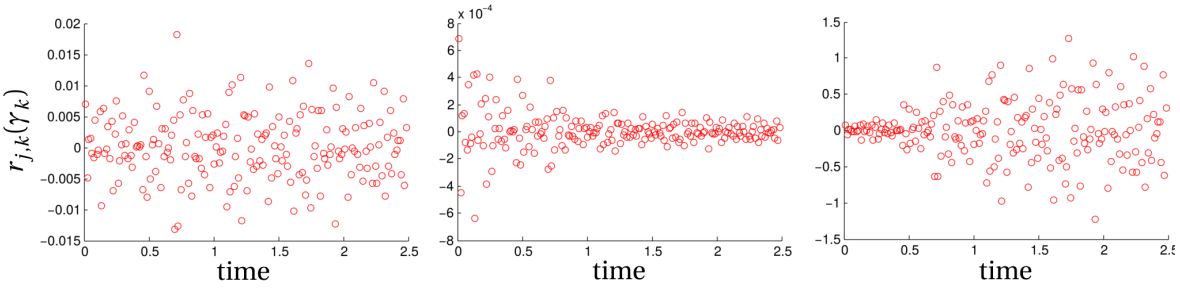

Figure 1.1 Examples of residual plots for an appropriate choice of γ_k (left) and two fan-shaped residual plots characteristic of an inappropriate choice of γ_k (center, right).

We note that there are some drawbacks to this method of statistical model verification. As mentioned above, we must estimate model parameters $\hat{\theta}_{LS}$ to generate residual plots for the corresponding model solution; if solving the inverse problem is computationally intensive, then repeatedly solving the problem under varying values of γ_k will be a time-consuming process. Additionally, the use of residuals as a proxy for observational error in our data is only appropriate when the assumed mathematical model is accurate. However, we seldom expect that a mathematical model will perfectly describe a biological system and must therefore recognize the potential for erroneous conclusions when using residual analysis.

An alternative to analyzing residual plots is the use of difference-based measurement errors, as outlined in [7]. We utilize a second-order difference approximation to compute pseudo-measurement errors directly from the data. This approximation is given by

$$\hat{\epsilon}_{j,k} = \begin{cases} \frac{1}{\sqrt{2}}\left(y_{j+1,k} - y_{j,k}\right) & j = 1 \\ \frac{1}{\sqrt{6}}\left(y_{j-1,k} - 2y_{j,k} + y_{j+1,k}\right) & j = 2,\ldots,N-1 \\ \frac{1}{\sqrt{2}}\left(y_{j,k} - j_{j-1,k}\right) & j = N \end{cases} \tag{1.9}$$

for each state k with N observations. Then, we compute the modified residual at time t_j to be

$$\eta_{j,k}(\gamma_k) = \frac{\hat{\epsilon}_{j,k}}{\left|y_{j,k} - \hat{\epsilon}_{j,k}\right|^{\gamma_k}}.$$

As with model-based residuals, we examine plots of these difference-based residuals against time t_j or observations $y_{j,k}$; an appropriate value of γ_k will yield a random plot with no fan shape.

Utilizing pseudo-measurement errors removes dependence on the mathematical model when identifying the statistical model for error in the data. Additionally, because the difference-based approximations are computed directly from the data, we can investigate statistical model assumptions without solving an inverse problem. However, the mesh between samples at times $t_1, t_2, \ldots, t_N$ determines the accuracy of the difference-based approximation to $\hat{\epsilon}_{j,k}$. If data is collected at an inappropriate temporal resolution, then the resulting pseudo-measurement errors may not be a reliable respresentation of the true error in the data. In this work, we attempt both methods of statistical model verification when handling experimental data for an inverse problem.

1.2.3 Approximating Standard Errors

Although we are interested in the solution to the inverse problem, $\hat{\theta}_{LS}$, the utility of our solution is determined by the accuracy with which the parameters are estimated. We are immediately interested in quantifying the uncertainty in our estimates, which can be impacted by parameter sensitivity, noisy data, or even model misspecification. We therefore consider the approximation of standard errors for each entry in $\hat{\theta}_{LS}$ within the asymptotic framework of the IRLS inverse problem.

Under the conditions for convergence as outlined in [10, 11], the estimator Θ is normally distributed with mean θ_0 and variance Σ_0 in the limit as an infinitely large data set is used to inform the inverse problem. An approximation to Σ_0 for a finite data set with N observations is given by

$$\Sigma_0^N = \left(\sum_{j=1}^{N} D_j^T(\theta_0) \left[W_j V_j \right]^{-1} D_j(\theta_0) \right)^{-1}.$$

$D_j(\theta_0)$ is the sensitivity matrix,

$$D_j(\theta) = \begin{pmatrix} \frac{\partial f_1(t_j;\theta)}{\partial \theta_1} & \frac{\partial f_1(t_j;\theta)}{\partial \theta_2} & \cdots & \frac{\partial f_1(t_j;\theta)}{\partial \theta_{\kappa_\theta}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_{m(t_j)}(t_j;\theta)}{\partial \theta_1} & \frac{\partial f_{m(t_j)}(t_j;\theta)}{\partial \theta_2} & \cdots & \frac{\partial f_{m(t_j)}(t_j;\theta)}{\partial \theta_{\kappa_\theta}} \end{pmatrix},$$

evaluated at θ_0 and $W_j V_j$ is the matrix given by (1.5). In practice, we do not know θ_0 or the associated value of $W_j V_j$. We therefore compute the approximation to Σ_0^N associated with the IRLS estimate

$\hat{\theta}_{LS}$ given by

$$\hat{\Sigma}\left(\hat{\theta}_{LS}\right)=\left(\sum_{j=1}^{N} D_j^T\left(\hat{\theta}_{LS}\right)\left[\hat{W}_j\hat{V}_j\right]^{-1} D_j\left(\hat{\theta}_{LS}\right)\right)^{-1}.$$

The standard error $SE_k(\hat{\theta}_{LS})$ associated with the k^{th} entry of $\hat{\theta}_{LS}$ is therefore given by

$$SE_k\left(\hat{\theta}_{LS}\right)=\sqrt{\hat{\Sigma}_{kk}\left(\hat{\theta}_{LS}\right)},$$

for $\hat{\Sigma}_{kk}$ the $(k,k)^{\text{th}}$ entry of $\hat{\Sigma}$.

The above approximation assumes knowledge of the sensitivity matrix $D_j(\theta)$. However, in the absence of an analytic solution to (1.1), we cannot directly calculate the partial derivatives $\frac{\partial f_i(t_j;\theta)}{\partial \theta_k}$. We can approximate these unknown entries in $D_j(\theta)$ by a numerical differencing scheme, such as the forward difference given by

$$\frac{\partial f_i(t_j;\theta)}{\partial \theta_k}\approx\frac{f_i(t_j;\theta+\mathbf{h}_k)-f_i(t_j;\theta)}{||\mathbf{h}_k||},$$

where $\mathbf{h}_k$ is a vector in $\mathbb{R}^{\kappa_\theta}$, with some small, non-zero quantity in the k^{th} entry and zeroes in all other entries. It is also possible to approximate the partial derivatives by a numerical scheme which eliminates the cancellation error introduced by a differencing scheme [9]. However, in this work, we largely rely on the chain rule to fill the entries of our sensitivity matrix, using the sensitivity equations,

$$\frac{d}{dt}\left(\frac{\partial \mathbf{x}}{\partial \theta}\right)=\frac{\partial \mathbf{g}}{\partial \mathbf{x}}\frac{\partial \mathbf{x}}{\partial \theta}+\frac{\partial \mathbf{g}}{\partial \theta}.$$

We solve these equations for the quantity $\frac{\partial \mathbf{x}}{\partial \theta}(t)$, from which we obtain

$$\frac{\partial f_i(t_j;\theta)}{\partial \theta_k}=\mathcal{K}(t_j)\frac{\partial x_i(t_j;\theta)}{\partial \theta_k}.$$

To solve the sensitivity equations, we define the quantity $\mathbf{s}_{q_k}$ to be

$$\mathbf{s}_{q_k}(t)=\frac{\partial \mathbf{x}}{\partial q_k}(t),$$

for q_k the k^{th} entry of $\mathbf{q}$. We also define the quantity $\mathbf{r}_{x_{0k}}$ to be

$$\mathbf{r}_{x_{0k}}(t)=\frac{\partial \mathbf{x}}{\partial x_{0k}}(t),$$

for x_{0k} the k^{th} entry of the initial condition, $\mathbf{x}_0$. We denote the matrices $\mathbf{s}=\left[\mathbf{s}_{q_1},\mathbf{s}_{q_2},...,\mathbf{s}_{q_{\kappa_q}}\right]$ and $\mathbf{r}=\left[\mathbf{r}_{x_{01}}\mathbf{r}_{x_{02}},...,\mathbf{r}_{x_{0n}}\right]$. Then the desired solution to the sensitivity equations is $\frac{\partial \mathbf{x}}{\partial \theta}=[\mathbf{s},\mathbf{r}]$, for which we

solve the system

$$\begin{aligned}
&\frac{d\mathbf{s}}{dt} = \frac{\partial \mathbf{g}}{\partial \mathbf{x}}(t, \mathbf{x}(t;\theta), \mathbf{q})\mathbf{s}(t) + \frac{\partial \mathbf{g}}{\partial \mathbf{q}}(t, \mathbf{x}(t;\theta), \mathbf{q}),\\
&\frac{d\mathbf{r}}{dt} = \frac{\partial \mathbf{g}}{\partial \mathbf{x}}(t, \mathbf{x}(t;\theta), \mathbf{q})\mathbf{r}(t),\\
&\mathbf{s}(t_0) = \mathbf{0}_{n\times\kappa_q},\\
&\mathbf{r}(t_0) = \mathbf{I}_{n\times n}.
\end{aligned}$$

We note that this system is necessarily coupled with the solution for $\mathbf{x}(t;\theta)$ given by (1.1).

1.2.4 Nested Model Comparison Testing

In finding a mathematical model to describe physical dynamics, we are immediately concerned with the model's ability to "fit" empirical observations. However, an equally important concern is the appropriateness of a model which successfully fits these observations. Increasingly complex models, permitting the estimation of additional parameters, may reduce the error in our model fit to data at the expense of certainty in estimated parameters. In this dissertation, we consider a family of models with varying complexity. Simpler models within this family are special cases of the more complex models, corresponding to some restriction in the parameter space. We seek models which appropriately balance complexity and goodness of fit. We can compare the performance of these nested models with a residual sum of squares model selection criterion, as outlined in [10, 11] and described below.

We define the cost function $J\left(\{\mathbf{Y}_j\}_{j=1}^N, \theta\right)$ to be the weighted least squares cost associated with the inverse problem,

$$J\left(\{\mathbf{Y}_j\}_{j=1}^N, \theta\right) = \sum_{j=1}^{N}\left[\mathbf{Y}_j - \mathbf{f}\left(t_j;\theta\right)\right]^T\left[W_j V_j\right]^{-1}\left[\mathbf{Y}_j - \mathbf{f}\left(t_j;\theta\right)\right].$$

Given a realization $\{\mathbf{y}_j\}_{j=1}^N$ of the data, we have the associated realization of the cost function,

$$J\left(\{\mathbf{y}_j\}_{j=1}^N, \theta\right) = \sum_{j=1}^{N}\left[\mathbf{y}_j - \mathbf{f}\left(t_j;\theta\right)\right]^T\left[\hat{W}_j \hat{V}_j\right]^{-1}\left[\mathbf{y}_j - \mathbf{f}\left(t_j;\theta\right)\right],$$

for which $\hat{\theta}_{LS}$, the realization of the estimator Θ given in (1.4), satisfies

$$\hat{\theta}_{LS} = \arg\min_{\theta\in\Omega_\theta} J\left(\{\mathbf{y}_j\}_{j=1}^N, \theta\right).$$

We consider some restriction, $\Omega_\theta^H \subset \Omega_\theta$, of the admissible parameter space. We assume that for

some matrix $\mathscr{H} \in \mathbb{R}^{\kappa_r \times \kappa_\theta}$ and vector $\mathbf{h} \in \mathbb{R}^{\kappa_r}$, this space is given by

$$\Omega_\theta^H = \{\theta \in \Omega_\theta \mid \mathscr{H}\theta = \mathbf{h}\}.$$

The estimator, Θ^H, associated with this restriction is given by

$$\Theta^H = \arg\min_{\theta \in \Omega_\theta^H} J\left(\{\mathbf{Y}_j\}_{j=1}^N, \theta\right).$$

For realizations of the data $\{\mathbf{y}_j\}_{j=1}^N$, the corresponding estimate is given by,

$$\hat{\theta}_{LS}^H = \arg\min_{\theta \in \Omega_\theta^H} J\left(\{\mathbf{y}_j\}_{j=1}^N, \theta\right).$$

We assert the *null hypothesis* that the nominal parameter set θ_0 which describes system dynamics is within this restricted parameter space, $\theta_0 \in \Omega_\theta^H$. We define the test statistic for this hypothesis,

$$U^N = \frac{N\left[J\left(\{\mathbf{Y}_j\}_{j=1}^N, \Theta^H\right) - J\left(\{\mathbf{Y}_j\}_{j=1}^N, \Theta\right)\right]}{J\left(\{\mathbf{Y}\}_{j=1}^N, \Theta\right)},$$

with the corresponding realization $\hat{U}^N$ defined by,

$$\hat{U}^N = \frac{N\left[J\left(\{\mathbf{y}_j\}_{j=1}^N, \hat{\theta}_{LS}^H\right) - J\left(\{\mathbf{y}_j\}_{j=1}^N, \hat{\theta}_{LS}\right)\right]}{J\left(\{\mathbf{y}\}_{j=1}^N, \hat{\theta}_{LS}\right)}.$$

We use this statistic to determine if the *null hypothesis* is rejected.

We recall that the value of κ_r is equivalently the number of restrictions associated with the reduced parameter space Ω_θ^H. For sufficiently large values of N, the statistic U^N is approximately χ^2-distributed, with κ_r degrees of freedom. Then, for a desired confidence level α, we seek the threshold $\tau = \tau(\alpha)$ such that

$$P\left(\chi^2(\kappa_r) > \tau\right),$$

where $\chi^2(\kappa_r)$ denotes the χ^2 distribution with κ_r degrees of freedom. If a realization of the test statistic satisfies the inequality

$$\hat{U}^N > \tau,$$

then we *reject the null hypothesis* as false with $(1-\alpha)\cdot 100\%$ confidence. If $\hat{U}^n \leq \tau$, we only *fail to reject the null hypothesis*. We note that a stronger conclusion, such as accepting that $\theta \in \Omega_\theta \setminus \Omega_\theta^H$, is not a possible outcome of this nested model comparison test. That is, we can reject a simplification of the parameter space or a special case of our models, but we cannot accept a more complicated model as the "true" description of system dynamics.

CHAPTER

2

MODELLING PREDATOR-PREY DYNAMICS

In this chapter, we discuss classic ordinary- and partial- differential equation models for predator-prey dynamics, beginning with the Lotka-Volterra model. We move on to the variations of this model given by assuming a nonlinear functional response. After, we motivate the Allometric Trophic Network (ATN) model, which we employ in Chapter 3, and a variation of the ATN model which we developed for the work in Chapters 4 and 5. We then state the sensitivity equations for the ATN model, as described in their general form in Chapter 1. Following this, we move on to spatially-explicit models for predator-prey dynamics, as employed in Chapter 6. We conclude with a brief discussion of additional modelling approaches which are appropriate for predator-prey interactions.

2.1 Classic Models for Predator-Prey Dynamics

2.1.1 Lotka-Volterra

We begin with the Lotka-Volterra model, which has historically formed the foundation for many discussions of predator-prey dynamics. A description of the model is readily available in most textbooks on mathematical modelling, and we specifically refer to [42, 66, 79] in the following introduction of the model. We consider the model for a system of two interacting species, some

prey item with population abundance N_1 and a predator item with population abundance N_2. The system dynamics are given by

$$\begin{aligned} \frac{dN_1}{dt} &= rN_1 - aN_1N_2 \\ \frac{dN_2}{dt} &= dN_1N_2 - xN_2 \end{aligned} \tag{2.1}$$

where r, a, d, and x are positive scalar quantities.

The prey population experiences exponential growth with some intrinsic rate r and suffers loss due to predation, as described by the term aN_1N_2. The rate of predation is assumed to be proportional to both the predator and prey abundances, following the description that the rate of predator-prey encounters increases with the product N_1N_2. Then the parameter a can be interpreted as a per-capita attack rate, which scales these encounters to produce the rate of predation on the prey population.

The predator population is assumed to experience exponential death with some intrinsic rate x and to increase with successful consumption of its prey resources at a rate dN_1N_2. This term is defined in the same way as the prey population's loss due to predation, for which we can consider the parameter $d = e \cdot a$. Given the definition of a as an attack rate, d can be interpreted as this rate scaled by some other parameter, e. The value of e translates loss in the prey population to gain in the predator population, as we do not necessarily expect that the consumption of one prey item corresponds to the subsequent birth of one predator item.

The assumption of exponential growth and death in the two populations is, often, obviously incorrect. These terms can be modified as necessary to accomodate logistic or seasonally-varying rates. The underlying assumptions for the rate of predation are not as easily addressed. This term is formulated in accordance with the *law of mass action*, which assumes that the two populations are well-mixed. In true biological systems, we do not expect that such a condition will be satisfied. However, it is difficult to know when the non-homogeneous distribution of true systems is significant enough to invalidate the formulation of this term.

Solutions to the Lotka-Volterra equations can result in mutual extinction, or extinction of only the predator population (reducing to exponential growth of the prey population). However, non-trivial solutions to the Lotka-Volterra equations are typically oscillating limit cycles, as depicted in Figure 2.1. Although there is a coexistence equilibrium for the predator and prey populations, the equilibrium is unstable. Solutions to the differential equation therefore oscillate around this equilibrium, with offset peaks in the predator and prey populations. An important criteria for choosing a mathematical model for a biological system is the realism of its dynamics, and so it is worth noting that we do not expect all ecosystems to mirror these limit cycles. Although there has been some success in matching the Lotka-Volterra model to real systems (see, for example, the

classic “Lynx and Hare” data set [45]), the unstable behavior of this model is often unsuitable for physical applications.

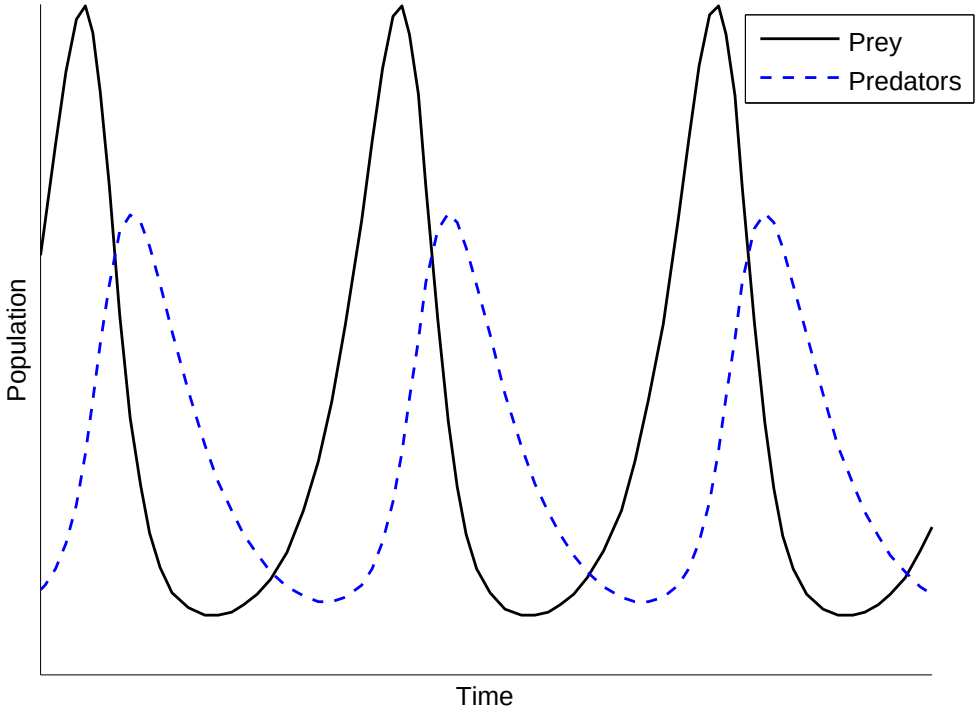

Figure 2.1 An example of a predator-prey limit cycle, obtained by solving (2.1) with initial conditions $N_1(0)=5$ and $N_2(0)=1$ with parameters $r=5$, $a=2$, $d=1$, and $x=3$ for $0\leq t\leq 5$.

2.1.2 Functional Responses

In addition to the aforementioned problems, the Lotka-Volterra model also describes an unrealistic response of predators to abundant prey items. Simple analysis of the model shows that for a given attack rate a, the effect of a fixed predator population scales with prey abundance. That is, a single predator could instantaneously consume $a\cdot 5$ or $a\cdot 5{,}000$ prey items. We therefore consider a common modification to the Lotka-Volterra model, the Holling Type II functional response [58].

The Lotka-Volterra model assumes a Type I functional response, characterized by a linear increase in the per-capita consumption of prey items in the presence of additional prey. Biologically, we expect that there is some maximal number of prey items which a single predator will consume, either due to satiation or physical limitations. We therefore assume a Type II response in the presence of some prey abundance N_1,

$$F(N_1)=1+ahN_1,$$

where a is the attack rate from the Lotka-Volterra model and h is a parameter we will coloquially refer to as "handling time." Modifying attack rates with the response yields the per-capita consumption of prey items,

$$\frac{aN_1}{1+ahN_1}.$$

This quantity saturates in the presence of additional prey items, as illustrated in Figure 2.2. As h increases, the saturating level of the response will decrease; that is, when more time is required to catch or digest a prey item, the number of prey items a predator can possibly consume decreases.

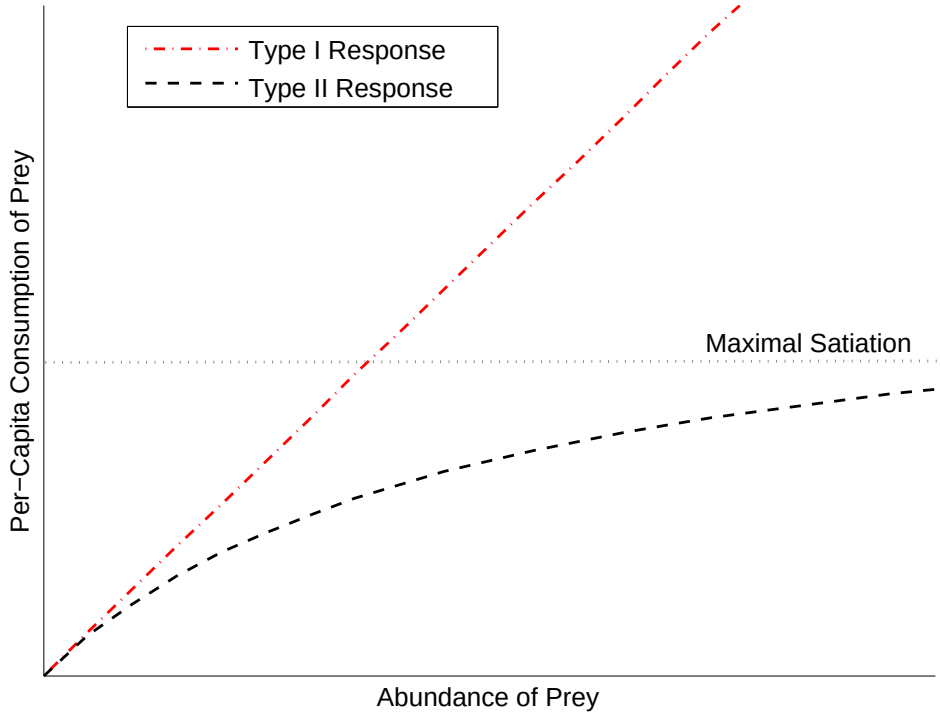

Figure 2.2 The effect of Type I and II functional responses in determining per-capita consumption of prey items as prey abundance increases.

This formulation of functional response is by no means a true description of biological interactions. The true response in many systems exhibits an S-shape at low prey densities [58], which might be attributed to predators learning to hunt sufficiently abundant prey. Moreover, attack rates could justifiably be affected by predator abundance as well as prey abundance; given a single prey item, we do not expect the number of prey consumed to differ in the presence of 500 or 5,000 predators. We therefore consider an update to the Type II functional response, which accounts for intraspecific competition within the predator species, termed the "Beddington-DeAngelis model" [102]. The

resulting model for predator-prey dynamics is

$$\frac{dN_1}{dt} = rN_1 - \frac{aN_1N_2}{1+cN_2+ahN_1}$$
$$\frac{dN_2}{dt} = \frac{dN_1N_2}{1+cN_2+ahN_1} - xN_2$$

where c is the parameter which describes intraspecific competition.

2.2 Allometric Trophic Network (ATN) Model

Although the above treatment of the Lotka-Volterra model only considered two interacting species, true ecosystems involve the interactions of many species, some of which are simultaneously predators and prey. We therefore generalize the Lotka-Volterra model with the Beddington-DeAngelis functional response to describe the dynamics of any number of species with population density N_i,

$$\frac{dN_i}{dt} = N_i\left[r_i + \sum_{j\in\mathscr{R}_i}\frac{e_{ji}a_{ji}N_j}{1+c_iN_i+\sum_{k\in\mathscr{R}_i}a_{ki}h_{ki}N_k} - \sum_{j\in\mathscr{C}_i}\frac{a_{ij}N_j}{1+c_jN_j+\sum_{k\in\mathscr{R}_j}a_{kj}h_{kj}N_k} - x_i\right], \quad (2.2)$$

for $\mathscr{R}_i$ the set of resources (prey) for species i and $\mathscr{C}_i$ the set of consumers (predators) of species i. In this model, r_i and x_i are the intrinsic growth and death rates of the population N_i, a_{ij} is the attack rate of species j on species i, h_{ij} is the handling time for predation on species i by species j, c_j is the parameter for intraspecific competition, and e_{ij} is the parameter for the efficiency with which attacks by species j on species i is converted into gain for species j.

We note that in considering a system of m species which are generalist predators capable of consuming most other species, we must specify up to $3(m^2+m)$ parameters to model these dynamics. Estimating this many parameters through an inverse problem would require a wealth of data that is unlikely to be available for every system we wish to study. Empirically measuring these parameters would similarly require an enormous effort. Moreover, it is not immediately obvious how one might measure e_{ij}, or if a measurement for h_{ij} in a lab setting may differ from the same quantity in a true ecosystem. It is in addressing this problem that the utility of a trait-based model, as described in Chapter 1, becomes clear.

2.2.1 Original ATN Model

Body mass is a trait with extraordinary explanatory power, which is well-studied as a factor governing predator-prey interactions [17, 20, 26, 88]. It is directly related to species metabolism [25], diet generality [51], and range of movement [80]. We therefore rely on the relationship between body mass and predator-prey interactions to parameterize our model. The Allometric Trophic Network (ATN)

model is a trait-based model which utilizes Lotka-Volterra dynamics with a functional response, with model parameters defined in terms of body mass. Previous studies [62, 101] demonstrate this model's success in describing the outcome of predator-prey interactions for controlled experiments in systems of terrestrial arthropods.

We begin by introducing the original ATN model, as described in [101], which assumes the dynamics given in (2.2). The model parameters are given by

$$\begin{aligned}
a_{ij} &= a_0 W_i^{1/4} W_j^{1/4} \left(\frac{W_j/W_i}{R_{opt}} e^{1-\frac{W_j/W_i}{R_{opt}}} \right)^{\phi}, \\
c_j &= c_0 W_j^{1/2}, \\
h_{ij} &= h_0 W_i^{1/4} W_j^{-1/4}, \\
e_{ij} &= .85 \frac{W_i}{W_j}, \\
x_i &= e^{23.05} W_i^{0.695} e^{-\frac{0.686}{(8.617\times 10^{-5})T}} \frac{3}{7W_i}.
\end{aligned}$$

The quantity W_i is the mass of species i, T is the ambient temperature, and the remaining quantites ($a_0, R_{opt}, \phi, c_0, h_0$) are allometric constants. We note that this parameterization significantly reduces the burden of data collection or empirical specification of model parameters. If we can approximate or measure the body mass W_i for each species in our system, then we only require the estimation of 5 allometric parameters to model the dynamics given by (2.2). This is true regardless of the size of our system and a significant improvement over having $3(m^2+m)$ unknown parameters for a system of m species.

The ATN model parameterization is built on the allometric relationships between body mass and biological processes. Parameters which describe interactions between individuals (i.e., attack rates, handling times, and competition coefficients), rely primarily on a relationship between the body mass and ground speed of an individual. Parameters which describe biological processes within an individual (i.e., conversion efficiency and intrinsic death rates), rely on the relationships between body mass and metabolic demands. The definitions of e_{ij} and x_i utilize empirical parameters which are appropriate for systems of terrestrial arthropods, but which might differ for other species. In particular, the constant in e_{ij} follows from [84, 121] and the constants in x_i were obtained from the empirical work in [43].

Following [84], we assume that an individual's speed is proportional to the body mass of that individual raised to the power of $\frac{1}{4}$. This assumption is directly applied in the formulation of handling time; we assume that the amount of time required for species j to handle species i will increase with the speed of the prey item and decrease with the speed of the predator item. Then for h_0 some

normalizing constant, we have

$$h_{ij} = h_0 W_i^{1/4} W_j^{-1/4}.$$

We note that the underlying assumption for this parameterization is a simplication of the biological processes driving predator satiation. "Handling time" might include the time a predator spends processing and metabolising its prey, instead of only accounting for the time required to catch prey. However, in the pursuit of an adequate and tractable model for predator-prey dynamics, we tentatively proceed with this formulation of handling time.

We extend the assumption that individual speed is determined by body mass to an assumption that encounters between individuals are similarly determined by body mass. We assume that the rate at which two individuals encounter one another while traversing a shared space increases with the speed of both individuals. Then the encounter rate between individuals of species i and individuals of species j scales with the quantity $W_i^{1/4} W_j^{1/4}$. In particular, we consider the rate at which an individual of species i encounters members of its own species to be $W_i^{1/2}$. Then, for c_0 a normalizing constant which accounts for the time spent on a competitive encounter, we have

$$c_j = c_0 W_j^{1/4} W_j^{1/4} = c_0 W_j^{1/2}.$$

As with the mass-action term in the Lotka-Volterra model, this relationship between body mass and encounter rates assumes an identical use of "space," which is not explicitly modelled, between individuals in our populations. Moreover, we assume that individuals are in constant random motion, and so this formulation is inappropriate for predators which exhibit "sit and wait" [92] hunting modes or which sense and actively avoid competitive members of their own species.

Despite these potential shortcomings, we use the same relationship between body mass and encounter rates in determining the attack rates, a_{ij}. For a_0 a normalizing constant which also accounts for the frequency with which predator-prey encounters result in an attack, we have the potential attack rate of and individual of species j on an individual of species i, given by

$$a_0 W_i^{1/4} W_j^{1/4}.$$

However, not every attack is successful. The quantity

$$\left(\frac{W_j/W_i}{R_{opt}} e^{1-\frac{W_j/W_i}{R_{opt}}} \right)^{\phi}$$

can be interpreted as the probability that the attack is successful, and therefore the rate of *successful*

attacks by an individual of species j on an individual of species i is given by

$$a_{ij} = a_0 W_i^{1/4} W_j^{1/4} \left(\frac{W_j/W_i}{R_{opt}} e^{1-\frac{W_j/W_i}{R_{opt}}} \right)^{\phi}. \tag{2.3}$$

Following [83], we assume that there is some optimal predator-prey body mass ratio, given by R_{opt}, at which attacks are always successful. The quantity within parentheses in (2.3) decreases as the true predator-prey body mass ratio, given by W_j/W_i, deviates from this value. The sensitivity of success rates to this optimal ratio is tuned by the parameter ϕ, as depicted in Figure 2.3.

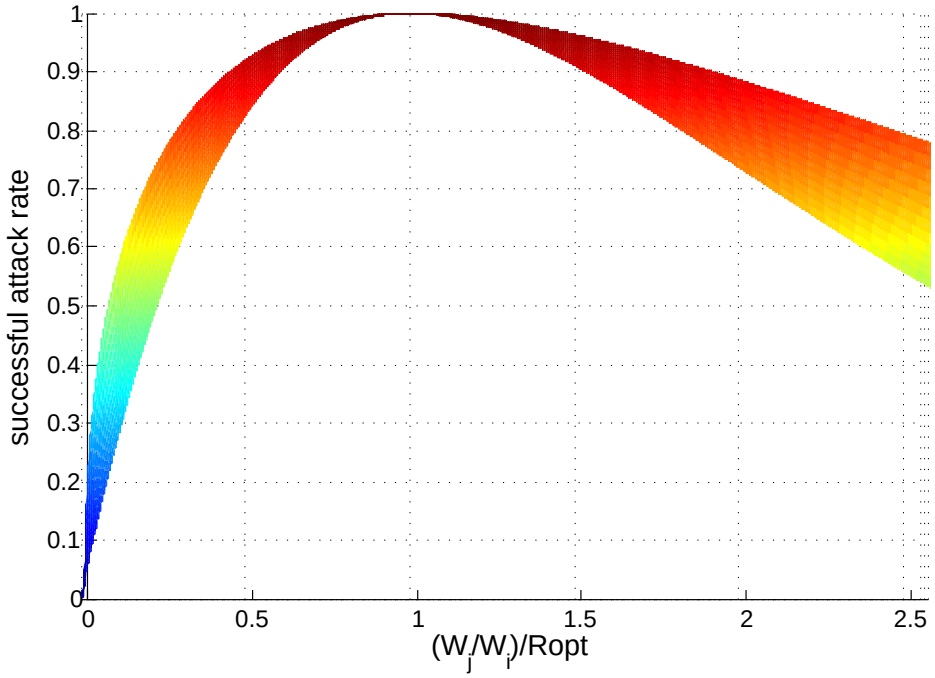

Figure 2.3 The successful attack rate curve, plotted against values of $\frac{W_j/W_i}{R_{opt}}$ along the x-axis. The width of the curve is obtained by taking values of $\phi \in [.4, 1]$.

We plot the probability of a successful attack for values of $\phi \in [.4, 1]$, which gives a band of success rates at each value of $\frac{W_j/W_i}{R_{opt}}$ along the x-axis. The bottom of the band corresponds to the case when $\phi = 1$, and the top of the band corresponds to when $\phi = .4$. As $\phi \to 0$, the curve flattens such that the probability of a successful attack is 1 for all values of $\frac{W_j/W_i}{R_{opt}}$. For $\frac{W_j/W_i}{R_{opt}} < 1$, the predator is smaller than is optimal, and we can see a significant penalty to the probability of a successful attack as this difference becomes extreme and $\frac{W_j/W_i}{R_{opt}} \to 0$. When $\frac{W_j/W_i}{R_{opt}} = 1$, the predator is the ideal size for an attack on its prey, and attacks are always successful. For $\frac{W_j/W_i}{R_{opt}} > 1$, the predator is larger than is

optimal; although the rate of successful attacks does decrease, there is less of a penalty than when the predator is smaller than the optimal size. We note that the effect of ϕ is most pronounced when W_j/W_i differs significantly from R_{opt}, or, that the effect of body mass in determining the success of an attack is more important when predators encounter less-suitable prey.

2.2.2 Short-Term Simplification

In this work, we assume that dynamics occur over a sufficiently short time interval such that there is a negligible effect on the predator population due to natural death. We still permit death by predation in these populations, but we remove the intrinsic death rate x_i from our model. We additionally assume that there is no growth in the predator population under such a short time interval; although predator populations might gain biomass over a short time, we only model abundances or population densities. We therefore remove the terms for predator growth ($e_{ji}a_{ji}$) from our model.

The simplified ATN model that we will use in Chapter 3 is therefore given by

$$\frac{dN_i}{dt} = N_i \left[r_i - \sum_{j \in \mathscr{C}_i} \frac{a_{ij} N_j}{1 + c_j N_j + \sum_{k \in \mathscr{R}_j} a_{kj} h_{kj} N_k} \right],$$

where $r_i = 0$ for any predators, but $r_i \neq 0$ for quickly-growing basal prey populations. The only allometric parameters we require are a_{ij}, h_{ij}, and c_j, given by

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

2.2.3 Modified ATN Model

We next introduce a variation on the ATN model, which forms the basis of our work in Chapters 4 and 5. Although body size is a well-studied trait in determining species interactions, it is not sufficient to describe the mechanisms underlying all systems and efforts have been made to inform models with additional traits [24, 51, 88]. Following the ideas explored in [99, 100], we additionally assume that species habitat use is important in determining interactions; the assumption that interacting species are well-mixed is fundamental to ATN model, but differences in the use of a shared habitat might render this assumption invalid. We therefore formulate terms which account for species' distributions within their habitat in our modification of the ATN model.

We introduce the parameter v_{ij} to describe spatial overlap, with the assumption that species

which spend more time in the same space will have a greater impact on each other. We also introduce a functional response term $\sum_{l\in\mathscr{C}_j} b_0 a_{jl}\nu_{jl}N_l$, describing the trade off that arises from balancing the competing requirements of being both predator and prey. An individual which spends time avoiding potential predators will necessarily limit the time spent exploiting prey items. We sum over the potential attack rates of all species l on an individual of species j to account for time spent avoiding or evading species l while attempting to capture prey items. The parameter b_0 is a scaling constant for the effect of this predator interference on the predation force of species j. For a group of cannibalistic predators, interference from one's own species may not be distinguishable from general predator interference, and so we remove the intraspecific competition parameter c_j from this version of the ATN model.

Dynamics for the number of individuals N_i of species i are therefore given by

$$\frac{dN_i}{dt} = N_i\left[r_i - \sum_{j\in\mathscr{C}_i}\frac{a_{ij}\nu_{ij}N_iN_j}{1+\sum_{k\in\mathscr{R}_j}a_{kj}\nu_{kj}h_{kj}N_k+\sum_{l\in\mathscr{C}_j}b_0a_{jl}\nu_{jl}N_l}\right].$$

As in the original ATN model, $\mathscr{C}_i$ is the set of species which consume species i, $\mathscr{R}_j$ is the set of prey for species j, and r_i is the intrinsic growth rate of species i, which we assume to be zero-valued for predators. The parameters a_{ij} and h_{ij} are allometrically defined, and ν_{ij} is similarly parameterized by the measurable trait "habitat use." For the set Ω of all modes of habitat use, we define the probability measure (μ_i) such that evaluating $\mu_i(\mathscr{A})$ for some habitat $\mathscr{A}$ contained in Ω gives us the likelihood that species i occupies that habitat $\mathscr{A}$. For example, one $\mathscr{A}$ in Ω might be "residing in foliage" and an entirely ground-dwelling predator would therefore have a measure $\mu_i(\mathscr{A})=0$. That is, μ_i is a formalization of observed species habitat use. For species pair i and j, we define the overlap parameter

$$\nu_{ij} = 1-\upsilon_0 TV(\mu_i,\mu_j) = 1-\upsilon_0\sup_{\mathscr{A}\subset\Omega}\left|\mu_i(\mathscr{A})-\mu_j(\mathscr{A})\right|,$$

where $TV(\mu_i,\mu_j)$ is the total variation distance [41] and $0\le\upsilon_0\le 1$ is a scaling factor for the importance of habitat use in determining spatial overlap.

Total variation distance quantifies the difference between two measures, and so $TV(\mu_i,\mu_j)$ quantifies the dissimilarity between the habitat use of individuals of species i and individuals of species j. Total variation takes on values between zero (no difference in the two measures) and one (maximal difference in the two measures). Continuing our previous example, if species i is an entirely ground-dwelling predator and species j is an entirely foliage-dwelling predator, then we would have $TV(\mu_i,\mu_j)=1$, since habitat use is completely dissimilar for the two species. The corresponding overlap $\nu_{ij}=1-\upsilon_0$ is a quantification of the similarity in habitat use, with υ_0 indicating how much weight we give to the assumed distributions μ_i and μ_j. If $\upsilon_0=1$, then we give full consideration to

the assumed distributions; in our example, this results in a value of $\nu_{ij} = 0$, and the two predators never interact in our model). If $\nu_0 = 0$, we ignore the assumed distributions in calculating the overlap parameter. In Chapter 4, we give a more detailed explanation of the parameter's formulation for predators in a cage experiment.

Our primary motivation for the inclusion of this trait in our model is that it might significantly affect population dynamics. However, we must weigh our interest in any additional traits against the tractability and generalizability of our resulting model. We intend to only utilize traits which maintain the model's relevance in describing diverse communities, potentially in environments outside the scope of our current studues, against which we can evaluate the model's performance [96]. Moreover, we cannot justify the use of a model with too many parameters to feasibly identify, or which is so sensitive to model parameters that it cannot be reliably used to investigate questions about physical processes. Our formulation of overlap is not driven by physical mechanisms, but it is amenable to inclusion in the ATN model, which assumes encounters rates increase with mutual distance covered. Inclusion of ν_{ij} scales species encounter rates so that they increase with mutual distance covered *in the same habitat zones*. Additionally, this formulation is flexible and can be adapted to any study system; we can define the set Ω as is most-relevant for the expected species behaviors and measure μ_i in any new habitats.

2.2.4 ATN Sensitivity Equations

We denote the vector of population densities $\mathbf{N} = [N_1, N_2, ..., N_m]^T$ and consider the function $\mathbf{g}(t, \mathbf{N}, \theta)$ as given by the modified ATN model with intraspecific competition,

$$g_i(t\mathbf{N}, \theta) = N_i \left[r_i - \sum_{j \in \mathscr{C}_i} \frac{a_{ij}\nu_{ij}N_iN_j}{1 + c_jN_j + \sum_{k\in\mathscr{R}_j} a_{kj}\nu_{kj}h_{kj}N_k + \sum_{l\in\mathscr{C}_j} b_0 a_{jl}\nu_{jl}N_l} \right].$$

Although we will not use both b_0 and c_j in a single variation of the ATN model, we present this formulation in order to give the sensitivity equations for any variation of the model in a succinct manner. We denote the quantity

$$F_j = \frac{1}{1 + c_jN_j + \sum_{k\in\mathscr{R}_j} a_{kj}\nu_{kj}h_{kj}N_k + \sum_{l\in\mathscr{C}_j} b_0 a_{jl}\nu_{jl}N_l},$$

for brevity and present the sensitivity equations for an ATN model of general size, assuming that $a_{ij} = 0$ whenever species j does not consume species i. Despite this, we retain summations over sets $\mathscr{C}_i$ and $\mathscr{R}_j$ for clarity.

We recall from Chapter 1 that for a system given by

$$\frac{d\mathbf{N}}{dt} = \mathbf{g}(t, \mathbf{N}, \theta),$$

we obtain the sensitivities $\mathbf{s}$ and $\mathbf{r}$ (with respect to model parameters and initial conditions, respectively) by solving the system of differential equations

$$\begin{aligned}
&\frac{d\mathbf{s}}{dt} = \frac{\partial \mathbf{g}}{\partial \mathbf{x}}(t, \mathbf{x}(t;\theta), \mathbf{q})\mathbf{s}(t) + \frac{\partial \mathbf{g}}{\partial \mathbf{q}}(t, \mathbf{x}(t;\theta), \mathbf{q}),\\
&\frac{d\mathbf{r}}{dt} = \frac{\partial \mathbf{g}}{\partial \mathbf{x}}(t, \mathbf{x}(t;\theta), \mathbf{q})\mathbf{r}(t),\\
&\mathbf{s}(t_0) = \mathbf{0}_{n\times \kappa_q},\\
&\mathbf{r}(t_0) = \mathbf{I}_{n\times n}.
\end{aligned}$$

For any i and k, we have

$$\frac{\partial g_i}{\partial N_k} = r_i \frac{\partial N_i}{\partial N_k} - \sum_{j\in \mathscr{C}_i} a_{ij} v_{ij} \left(\frac{\partial N_i}{\partial N_k} N_j F_j + N_i \frac{\partial N_j}{\partial N_k} F_j + N_i N_j \frac{\partial F_j}{\partial N_k} \right),$$

where for $j \neq i$,

$$\frac{\partial F_j}{\partial N_k} = -F_j^2 \left(a_{kj} v_{kj} h_{kj} + b_0 a_{jk} v_{jk} \right),$$

and if $j = i$, then

$$\frac{\partial F_j}{\partial N_k} = -F_j^2 \left(a_{kj} v_{kj} h_{kj} + b_0 a_{jk} a_{jk} + c_j \right).$$

The derivative of f_i with respect to θ is given by

$$\frac{\partial g_i}{\partial \theta} = -\sum_{j\in \mathscr{C}_i} N_i N_j \left(\frac{\partial (a_{ij} v_{ij})}{\partial \theta} F_j + a_{ij} v_{ij} \frac{\partial F_j}{\partial \theta} \right),$$

where

$$\begin{aligned}
&\frac{\partial F_j}{\partial \theta_a} = -F_j^2 \left(\sum_{k\in \mathscr{R}_j} \frac{\partial (a_{kj} v_{kj})}{\partial \theta_a} h_{kj} N_k + \sum_{h\in \mathscr{C}_j} b_0 \frac{\partial (a_{jh} v_{jh})}{\partial \theta_a} N_h \right) \text{ for } \theta_a = a_0, R_{opt}, \phi, v_0,\\
&\frac{\partial F_j}{\partial b_0} = -F_j^2 \sum_{h\in \mathscr{C}_j} a_{jh} v_{jh} N_h,\\
&\frac{\partial F_j}{\partial h_0} = -F_j^2 \sum_{k\in \mathscr{R}_j} a_{kj} v_{kj} \frac{h_{kj}}{h_0} N_k,\\
&\frac{\partial F_j}{\partial c_0} = -F_j^2 \frac{c_j}{c_0} N_j.
\end{aligned}$$

The derivatives of our allometric parameters with respect to a_0, R_{opt}, and ϕ are given by

$$\begin{aligned}
\frac{\partial(a_{ij}v_{ij})}{\partial a_0} &= W_i^{1/4}W_j^{1/4}\left(\frac{W_j/W_i}{R_{opt}}e^{1-\frac{W_j/W_i}{R_{opt}}}\right)^{\phi}\left(1-v_0TV(\mu_i,\mu_j)\right),\\
\frac{\partial(a_{ij}v_{ij})}{\partial v_0} &= -a_0W_i^{1/4}W_j^{1/4}\left(\frac{W_j/W_i}{R_{opt}}e^{1-\frac{W_j/W_i}{R_{opt}}}\right)^{\phi}TV(\mu_i,\mu_j),\\
\frac{\partial(a_{ij}v_{ij})}{\partial R_{opt}} &= a_{ij}v_{ij}\frac{\phi}{R_{opt}^2}\left(W_j/W_i-R_{opt}\right),\\
\frac{\partial(a_{ij}v_{ij})}{\partial \phi} &= a_{ij}v_{ij}\ln\left(\frac{W_j/W_i}{R_{opt}}e^{1-\frac{W_j/W_i}{R_{opt}}}\right) = a_{ij}v_{ij}\left(1-\frac{W_j/W_i}{R_{opt}}+\ln\frac{W_j/W_i}{R_{opt}}\right).
\end{aligned}$$

2.3 Spatial Considerations for Population Dynamics

We next outline some spatial processes, as described by partial differential equations, which we will utilize in considering the spatial design of experiments for validating predator-prey interactions in Chapter 6. We note that due to the aforementioned concerns about the well-mixedness of populations in non-heterogeneous landscapes, describing predator-prey dynamics over large spatial domains might require the use of spatially-explicit models. We introduce here a few spatially-varying mechanisms which might drive predator-prey interactions in such a domain. We leave the formulation of the particular model we will study for Chapter 6.

2.3.1 Diffusion

We first consider diffusion of populations, a common mechanism in spatially-explicit models, and refer to [115] in our discussion. For some spatial domain $\Omega \subset \mathbb{R}^2$, we consider the abundance $N(t,\mathbf{x})$ of some population at time t and position $\mathbf{x}\in\Omega$. The *diffusion equation* for this population is given by

$$\frac{\partial N(t,\mathbf{x})}{\partial t} = \nabla\cdot[D\nabla N(t,\mathbf{x})],$$

where D is the diffusion coefficient which determines the rate at which the population spreads. This relationship can be derived by assuming that the rate at which individuals move across some position $\mathbf{x}$ is proportional to the gradient of the population at that position. That is, the population will move away from areas of high abundance and towards areas of low abundance.

Diffusion is often described in relationship to "random walks," where the path of an individual develops randomly in time. It is therefore an appropriate description of motion for species which exhibit constant motion without a clear directional preference. Models utilizing this assumption have been successfully applied to the behavior of real populations [22, 29], although modifications should be made for species which exhibit more complicated modes of motion. For example, the

diffusion coefficient D might be formulated as a function of time, population abundance N, or the presence of some other state in the system; defining D in this matter permits for diffusion at varying rates, but a more directed movement requires an alternative formulation of spatial dynamics.

2.3.2 Advection

We next consider the advective transport of populations, which describes motion that is driven in a particular direction, again referring to [115]. The advection equation for a population with abundance $N(t,\mathbf{x})$ moving with some velocity $\vec{\mathbf{v}} \in \mathbb{R}^2$ is

$$\frac{\partial N(t,\mathbf{x})}{\partial t} = -\nabla \cdot [\vec{\mathbf{v}} N(t,\mathbf{x})].$$

This relationship is formulated with the assumption that individuals in the population have some preference in the direction of movement, as described by $\vec{\mathbf{v}}$. The velocity of the population can be defined in terms of driving mechanisms; a constant $\vec{\mathbf{v}}$ describes a fixed bias in the population's movement, while formulation of $\vec{\mathbf{v}}$ as a function of system states can describe populations which move towards resources or away from predators. In particular, the formulation of $\vec{\mathbf{v}}$ to describe predators which move towards resources is described as *prey-taxis*. This relationship can be simply linear, but at its most complicated, the quantity $\vec{\mathbf{v}}$ can be specified by its own differential equation.

2.4 Alternative Models for Predator-Prey Dynamics

We finally note that there is a wealth of possible modelling approaches, and given the importance of trophic interactions in ecological research, many have been successfully applied to predator-prey dynamics. An important step in modelling a physical system is identifying an appropriate model for the intended analyses. The data required for validation varies between different types of models, and so model choice might be dictated by available information. Alternatively, the dynamics which are most-important in describing physical behavior might vary at different temporal or spatial scales. The models described below have been appropriate for different analyses of predator-prey dynamics, including in our own intended study systems. However, because of the types of behavior we intend to model and information we hope to obtain, we do not rely on these models in our work.

We focus in this work on deterministic ordinary differential equation models for predator-prey dynamics, but an immediate extension of our overview is to stochastic models for these systems. Stochastic differential equations incorporate random variables which can describes noisy driving processes or parameterizations. Stochastic Lotka-Volterra models are well-studied [30, 75, 89], and permit the inclusion of variability which can drive new modes of species extinction or coexistence. There are numerous ecosystem processes which drive population dynamics, such as temperature and environmental resources, which can present as random effects on the population.

The accomodation of random fluctuations in parameters describing population characteristics is therefore a desirable aspect of these models. However, we begin our work with deterministic differential equations, which may not describe these random effects but are more tractable and can still provide a useful description of the physical processes.

An alternative to partial differential equation models for spatial processes in predator-prey interactions is the use of patch models. Patch models are an intermediate choice between ordinary- and partial- differential equation models, in which the spatial domain of a study system is split into "patches." Dynamics within these patches are described by systems of ordinary differential equations, and populations can move between the patches to approximate spatial interactions. Theoretical analysis of predator-prey interactions in patched environments is well-established [16, 67] and can be used to explore complicated mechanisms for population persistence [97]. Because these models are systems of ordinary differential equations, they are computationally tractable and simpler to analyze than equivalent partial differential equations. Moreover, these models are an intuitive in light of the natural patchiness of real landscapes, particularly at multiple spatial scales [47, 64]. However, because our consideration of spatial dynamics occurs within a relatively-heterogeneous field, we consider partial differential equations in our work, and acknowledge that a patch-based model might be more appropriate for landscape-level interactions.

Another common approach to modelling populations is the use of agent-based (or, individual-based) models. Such models rely on descriptions of behavior at the level of a single individual, simulated simultaneously for the entire population of the study system and are well-studied in ecology (see [35, 77] for a brief overview). These models are particularly useful in ecosystem management and have been successfully used in assessing population-level effects of predation [78] and in describing the behavior of terrestrial arthropods [19, 112]. Additionally, agent-based models for predator-prey dynamics have been utilized in studies of evolutionary processes [53, 76]. Although stochastic variations on the Lotka-Volterra model [109] and classic analysis of Lotka-Volterra differential games [57] have been studied by the mathematical community, agent-based models are an intuitive biological tool. However, the amount of data required to parameterize such models, as well as the limited scope of our own investigations into predator-prey dynamics, make them impractical for our current work.

We finally recognize the utility of entirely statistical, instead of mathematical, models in describing predator-prey interactions. Multivariate autoregressive models (MAR) [36, 56], for example, have been successfully applied to aquatic food webs [48, 60] and can be used to identify the dominant interactions in a foodweb. The benefit to such models is that a strong statistical framework accomodates the consideration of deviations from the model driven by random processes. However, MAR models in particular only support linear species interactions, while we are interested in the nonlinear effects of multiple predators in an ecological community.

CHAPTER

3

FITTING THE ATN MODEL TO DATA: FIELD EXPERIMENT

In this chapter, we describe an initial attempt at fitting the Allometric Trophic Network (ATN) model to field data. The work presented here was originally published in [13] and formed the basis of the ecological discussions in [33]. We begin by describing the experimental data to which we fit the model. We then motivate a few simple variations on the ATN model and discuss their utility in describing the dynamics we observe in the data. We conclude with a discussion of the results and necessary future work.

3.1 Study System

We rely on data from [61, 94], which focused on the biological control of the bird cherry-oat aphid, *Rhopalosiphum padi*, by a community of its natural predators. The data was collected in 2011, from late May through early July, in the province of Uppland, Sweden. The study focused on barley fields managed by five different farmers under two different management strategies, for a total of ten data sets. For each of these fields, we have species abundance data, temperature readings, and the results of molecular gut content analysis for predators. We describe below the necessary assumptions and conversions we used to transform the data prior to its use in a parameter estimation problem.

3.1.1 Species Abundances

Abundance data is collected using different techniques for different types of organisms. In the experiment, the frequency with which data was collected also varied across different categories. We convert all abundance data to a population density, with units [individuals/m^2], for use in the inverse problem.

Aphids were sampled 5-6 times in each field during the study period, approximately once a week. At each field and time, aphids were directly counted on 100 barley plants (with a single exception, the third observation in field JC) by randomly choosing 10 batches of 10 consecutive plants. We convert an observed abundance A to a density N by the conversion $N = AB$, where B is the density of barley plants in the field with value $B = 411$ [plants/m^2] [95]. We assume exponential growth in the aphid population, which is only appropriate if population growth is unhindered by the maturation stage of crop resources [70]; the particular conditions for unimpeded population growth were monitored weekly, and we truncate the aphid population data by excluding points after a field has reached critical crop maturation.

Other plant-dwelling arthropods (Thrips, Diptera, and Coccinella) were sampled 2-4 times during the study period. At each field and time, these arthropds were collected in 50 sweeps with a sweepnet. We obtain a density N from an observed abundance A by the conversion $N = A/a$, where a is the area covered by the 50 sweeps. We compute this area with the expression

$$a = 50 f_c \pi \left((L_a + f_h L_h + d_{net})^2 - (L_a + f_h L_h)^2 \right),$$

for f_c the fraction of a circle that a sweep constitutes, L_a the length of the arm of the person sweeping, L_h the length of the sweepnet handle, f_h a factor accounting for the effective reduction of the sweep circle diameter caused by the handle being held at some angle towards the ground, and d_{net} the diameter of the net. We assume that $f_c = 0.29$, $L_a = 0.65$[m], $f_h = 1/3$, $L_h = 0.85$[m], and $d_{net} = 0.353$[m].

Soil organisms (Collembola and Earthworm) were sampled twice, approximately 4 weeks apart, using 6 soil samples of diameter 0.05[m] for springtails (Collembola) and 4 soil samples of diameter 0.25[m] for earthworms at each field and time. We obtain a density N from an observed abundance A by the conversion $N = A/a$, where a is the area of the soil sample, dependent on the diameter of the sample.

Ground-dwelling predators (Bembidion, Harpalus, Poecilus, Pterostichus, Other Carabid, Linyphiidae, Lycosidae, Tetragnathidae, and Other Spider) were sampled using wet pitfall traps. The traps were left open for the entire study period, with 6 in every field, and emptied 5 times during the study period. There is a known size-bias in abundance data obtained through pitfall catches [4], and so we follow the method from [28] to correct for this over-representation of larger species and obtain "unbiased" abundances for the pitfall catch data, following a negative $\frac{3}{4}$ relationship between body

mass and species abundance [116].

We assume that abundance data $\tilde{A}_i^j$, obtained at time t_j for the i^{th} species in the pitfall trap, should follow the allometric relationship

$$\tilde{A}_i^j \propto e^{\left[-\frac{3}{4}(m_i - M)\right]},$$

for m_i the species body mass and M the average body mass over all species. Using a log-linear regression over all species i and times t_j, we obtain a fit to data that does not follow the negative $\frac{3}{4}$ law, given by

$$\tilde{A}_i^j = C e^{\left[f(m_i - M) + \epsilon_i^j\right]},$$

for ϵ_i^j the error in abundance observation with constants C and f obtained by regression on the data. We use the constant C and errors ϵ_i^j to obtain the "unbiased" abundances A_i^j that follow the negative $\frac{3}{4}$ relationship,

$$A_i^j = C e^{\left[-\frac{3}{4}(m_i - M) + \epsilon_i^j\right]}.$$

We obtain a density N from an observed abundance A by the conversion $N = A/da$ for a the assumed catching range of a pitfall trap, given by $a = \pi 1.5^2 [\mathrm{m}^2]$ [122], and d the number of days the trap was open.

3.1.2 Node Body Masses

We rely on body mass to parameterize predator-prey interactions between broadly characterized "nodes" (groups of species which serve the same function in our foodweb) in the field. However, because we use a model for population-level dynamics, we specify some average body mass for each node which is dependent on the mass of individuals within the node. The calculation of this average mass depends on the taxonomic resolution in, or the species composition of, the abundance data.

For Aphids, Thrips, and Earthworm nodes, the resolution of the abundance data is the same as that of the node; we obtain abundance data for a single species, and that species makes up the entirety of the node. In this case, the node body mass is constant across all fields, and it is an average body mass of the corresponding species. For the remaining nodes, there is greater resolution in the abundance data than the definition of the node, and so the assigned body mass will vary across fields. For these nodes, the species composition is unique to the field in which the abundance data was collected. We therefore take a weighted average of the constituting species' body masses to compute the effective mass of the node in that field. The weight given to each species' mass is the species' relative contribution to node abundance over the entire season. The resulting node body masses are given in Table 3.1, and we refer to [12] for the details of their computation.

Table 3.1 Node body mass in milligram fresh weight. For nodes with field-specific body mass, the range across fields is given.

Node	**Body mass (mg)**
Aphid	0.59
Thrips	0.23
Diptera	26.674-29.687
Collembola	0.0015-0.0021
Earthworm	155
Bembidion	1.276-1.606
Harpalus	53.368-61.770
Poecilus	44.422-45.770
Pterostichus	85.179-103.960
Other Carabid	2.283-21.3030
Linyphiidae	1.036-1.638
Lycosidae	4.802-16.490
Tetragnathidae	1.910-2.263
Other Spider	3.938-31.900
Coccinella	4.470-20.780

3.1.3 Temperature-Dependent Growth Rates

We assume that the intrinsic growth rate, r, of an aphid population depends on temperature. Relying on data regarding the intrinsic growth rate of populations of *R. padi* held at different temperatures (10°C-30°C) in a laboratory setting [34], we obtain the relationship

$$r(t) = 0.024\,T(t) - .089. \tag{3.1}$$

The original data and this linear function are both plotted in Figure 3.1.

Field temperatures $T(t)$ were recorded in each of the fields every 15 minutes (with the exception of field SC, where temperature was collected every 30 minutes). The temperature logging device in field OO disappeared; we assume that data from a nearby field (OC) is an appropriate proxy for temperature in field OO. However, the function $r(t)$ relies on data from populations held at constant temperatures, and we assume that the population growth rate does not instantaneously change with temperature. We therefore take daily averages of the temperature, as depicted in Figure 3.2.

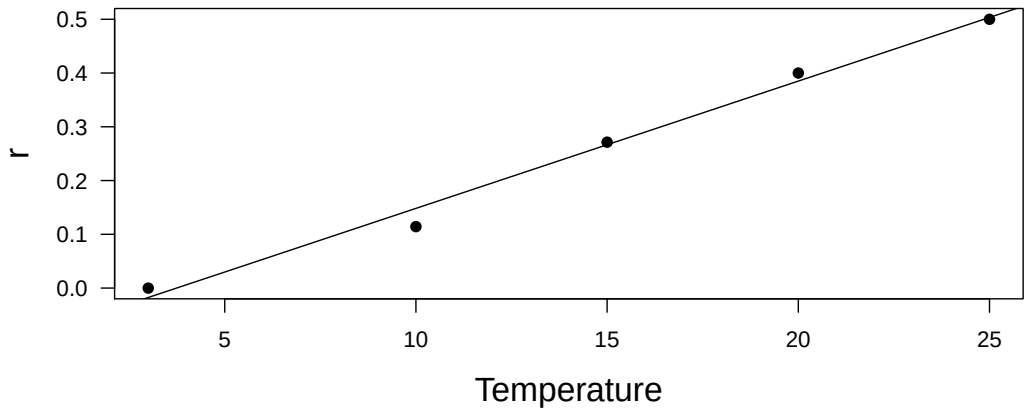

Figure 3.1 Laboratory data and the resulting regression for r(T) on temperature T.

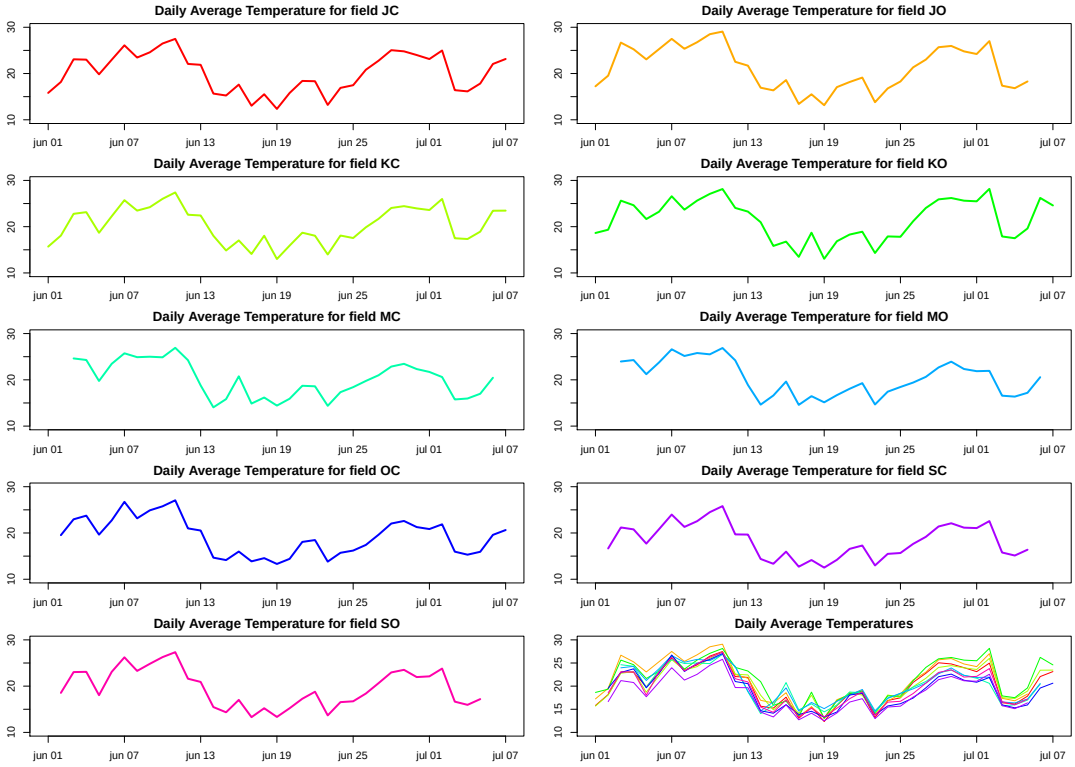

Figure 3.2 Daily average temperatures T(t) with linear interpolation plotted as a function of time.

3.1.4 Foodweb Structure

We must identify which predator-prey interactions occur in our system in order to specify the mathematical model for population dynamics. We rely on observations of true feeding patterns in the field, using a method of *diagnostic multiplex PCR assays* as described in [105]. Every predator caught in a pitfall trap was identified to species and subjected to whole-body DNA extraction, which was then screened for the DNA of prey categories. From this information, we obtain a binary foodweb of predator-prey interactions for each field at each time predators were sampled. This foodweb informs our model structure, specifying which predator-prey interactions are possible, from which model assumptions dictate the frequency with which these interactions occur. Although this is a sophisticated method of identifying true feeding patterns in a field, we must make some assumptions to translate the raw information obtained from the PCR assays into model dynamics.

We assume that potential feeding events are constant across fields and time; we do not expect that the possibility for one species to feed on another changes between May and July. Each of the binary foodwebs obtained from the molecular gut content analysis (MGCA) described above is merely one occurence of some potential feeding events in our study system. The resulting binary foodweb is dependent on which predators were captured in that sample and which species the predators recently consumed. To obtain a description of all potential feeding events for the fields at any time, we generate a single binary foodweb using interactions from all fields and samples. That is, we pool the observed interactions into a foodweb where all feeding events are represented, regardless of the frequency with which they occurred in the MGCA.

Some links cannot be detected by the MGCA; cannibalism, for example, cannot be detected because the DNA of the predator cannot be distinguished from the DNA of the prey items in its gut. Similarly, the prey category “Other Carabid” cannot be detected in the guts of carabids (Bembidion, Harpalus, Poecilus, and Pterostichus) and the prey category “Other Spider” cannot be detected in the guts of spiders (Linyphiidae, Lycosidae, and Tetragnathidae). Additionally, the nodes Tetragnathidae and Other Spider were not analyzed as predators in the MGCA. The results of the MGCA are given in Table 3.2, with * markers indicating undetected interactions.

To fill in the missing interactions, we make the following assumptions. We assume that cannibalistic interactions occur for all predators, since such behavior is known to occur within these groups [1, 32, 55]. Since the carabid and spider categories are generalist predators, we assume that all carabids feed on the Other Carabid category, and that all spiders feed on the Other Spider category. The species in the Tetragnathidae and Other Spider nodes are known to be predators, and we therefore assume potential prey based on their similarities to the other two spider nodes in the food web; the Tetragnathidae node is assumed to have the same set of prey as the Linyphiidae node, and the Other Spider node is assumed to have the same set of prey as the Lycosidae node. The resulting food web structure is given in Table 3.3.

Table 3.2 The incomplete interaction matrix for our system. Each column represents a predator node and each row represents a prey node. If there is a 1 in the $(i, j)^{th}$ entry of the matrix, then node i is prey for node j. If there is a * in the $(i, j)^{th}$ entry, then the gut content analysis did not provide data for that link.

	Predator	Aphid	Thrips	Diptera	Collembola	Earthworm	Bembidion	Harpalus	Poecilus	Pterostichus	Other Carabid	Linyphiidae	Lycosidae	Tetragnathidae	Other Spider	Coccinella
Prey																
Aphid		*	*	*	*	*	1	1	1	1	1	1	1	*	*	1
Thrips		*	*	*	*	*	1	1	1	1	0	1	1	*	*	1
Diptera		*	*	*	*	*	1	1	1	1	1	1	1	*	*	1
Collembola		*	*	*	*	*	1	1	1	1	1	1	1	*	*	1
Earthworm		*	*	*	*	*	1	1	1	1	1	1	1	*	*	1
Bembidion		*	*	*	*	*	*	1	1	1	1	1	1	*	*	1
Harpalus		*	*	*	*	*	1	*	1	1	1	1	1	*	*	1
Poecilus		*	*	*	*	*	1	1	*	1	1	1	1	*	*	1
Pterostichus		*	*	*	*	*	1	1	1	*	1	1	1	*	*	1
Other Carabid		*	*	*	*	*	*	*	*	*	*	1	1	*	*	0
Linyphiidae		*	*	*	*	*	1	1	1	1	1	*	1	*	*	0
Lycosidae		*	*	*	*	*	1	1	1	1	1	1	*	*	*	0
Tetragnathidae		*	*	*	*	*	0	1	1	1	1	1	1	*	*	0
Other Spider		*	*	*	*	*	1	1	1	1	1	*	*	*	*	1
Coccinella		*	*	*	*	*	1	1	1	1	1	0	1	*	*	*

Table 3.3 The interaction matrix for our system, constructed from gut content analysis and additional assumptions. Each column represents a predator node and each row represents a prey node. If there is a 1 in the $(i, j)^{th}$ entry of the matrix, then node i is prey for node j.

	Predator	Aphid	Thrips	Diptera	Collembola	Earthworm	Bembidion	Harpalus	Poecilus	Pterostichus	Other Carabid	Linyphiidae	Lycosidae	Tetragnathidae	Other Spider	Coccinella
Prey																
Aphid		0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Thrips		0	0	0	0	0	1	1	1	1	0	1	1	1	1	1
Diptera		0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Collembola		0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Earthworm		0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Bembidion		0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Harpalus		0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Poecilus		0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Pterostichus		0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Other Carabid		0	0	0	0	0	1	1	1	1	1	1	1	1	1	0
Linyphiidae		0	0	0	0	0	1	1	1	1	1	1	1	1	1	0
Lycosidae		0	0	0	0	0	1	1	1	1	1	1	1	1	1	0
Tetragnathidae		0	0	0	0	0	0	1	1	1	1	1	1	1	1	0
Other Spider		0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Coccinella		0	0	0	0	0	1	1	1	1	1	0	1	0	1	1

3.2 Modelling the System

The experiment from which we draw data addresses a broad study system; even with the course categorization of species into fifteen nodes, we model a diverse ecosystem in each field. Additionally, we consider population changes over more than a month. With the information available to us, we could not specify the natural birth and death dynamics of non-aphid nodes during this time with any certainty. We therefore dynamically model the Aphid node, for which we have a reasonable understanding of seasonal dynamics, and drive the model with density data from the remaining nodes. We describe the models we consider in the sections below, along with the statistical model of error in aphid density data for use in the inverse problem.

3.2.1 ATN Model

We assume that the density of the Aphid node, N_1, adheres to the Allometric Trophic Network (ATN) model as described in Chapter 2,

$$\frac{dN_1}{dt} = N_1 \left[r(t) - \sum_{j \in \mathscr{C}_i} \frac{a_{1j} N_j}{1 + c_j N_j + \sum_{k \in \mathscr{R}_j} a_{kj} h_{kj} N_k} \right]. \tag{3.2}$$

The density N_j of the j^{th} node for all $j \neq 1$ is a linear interpolation of the density data described in Section 3.1.1. The intrinsic growth rate of the Aphid node, $r(t)$, is given by (3.1). We recall that $\mathscr{C}_i$ is the set of predators of the i^{th} node and that $\mathscr{R}_j$ is the set of prey for the j^{th} node. These sets are specified by the interaction matrix in Table 3.3. Model parameters are given by

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

Species body masses, W_i, are determined for each field as described in Section 3.1.2. We must estimate the remaining model parameters (a_0, R_{opt}, ϕ, c_0, and h_0) as well as the initial density of the Aphid node (N_1^0).

3.2.2 Statistical Model

As described in Chapter 1, we utilize a relative error statistical model in which we assume that the density data $N_1^{d,f}$ on day d in field f is a realization of the random variable $Y_{d,f}$ defined by

$$Y_{d,f} = N_1(t_d; \theta_{0,f}) + N_1^{\gamma}(t_d; \theta_{0,f}) \mathscr{E}_{d,f}.$$

The parameter set $\theta_{0,f}$ is the nominal parameter set which describes the system's true dynamics. We allow this "true" parameter set to vary between fields because of the natural variation in field conditions. We assume that $\mathscr{E}_{d,f}$ is independent, normally distributed random noise. The weight γ is unknown and must be identified by residual analysis or pseudo-measurement errors, as outlined in Chapter 1.

We recall that the use of residual plots to determine γ tacitly assumes that our mathematical model accurately describes the system dynamics. Initial treatment of this inverse problem in [12] suggests that the ATN model cannot capture some aspects of population dynamics in the field. We therefore begin by computing difference-based pseudo-measurement errors to test our statistical

model without a mathematical model. We employ the second-order central differencing scheme to compute pseudo-measurement errors, $\hat{\epsilon}_{d,f}$, as given in (1.9). To identify an appropriate value of γ for use in our statistical model, we compute

$$\eta_{d,f,\hat{\gamma}} = \frac{\hat{\epsilon}_{d,f}}{\left|N_1^{d,f} - \hat{\epsilon}_{d,f}\right|^{\hat{\gamma}}}$$

for different values of $\hat{\gamma}$ until we find the value that results in a random distribution of $\eta_{d,f,\hat{\gamma}}$ when plotted against observation times t_d.

We compute pseudo-measurement errors for the density of the Aphid node 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; because deteriorating crop quality prevents comparison between the mathematical model and late-season data, only four or five data points from each field can be used in residual analysis.

For some fields, we can identify an appropriate statistical model using pseudo-measurement errors. The values of $\eta_{d,JC,\hat{\gamma}}$ for Field JC are plotted in Figure 3.3, 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 are obtained in Fields JO, KC, KO, and OO (see the appendix of [13] for corresponding plots). In the remaining fields, sparsity of the data makes it difficult to choose an appropriate value of $\hat{\gamma}$. For example, we plot the results from Field SO in Figure 3.4. 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.

Having insufficient information to suggest otherwise, we take $\gamma = 0$ in all fields and solve the inverse problem with an absolute error formulation. Abundance data may be managed by different individuals in different fields, and so we do not necessarily expect that γ should be the same for all fields. Additionally, intuition about the way that abundance data is collected suggests that $\gamma \neq 0$ is a better description for the error introduced during data collection; the error in a direct count of aphids on a plant should scale with the number of aphids being counted on the plant. However, we would require additional data in order to identify the correct value of γ in each field. We find that $\gamma = 0$ is an appropriate choice for some fields, and do not have enough information to select a different value in the remaining fields.

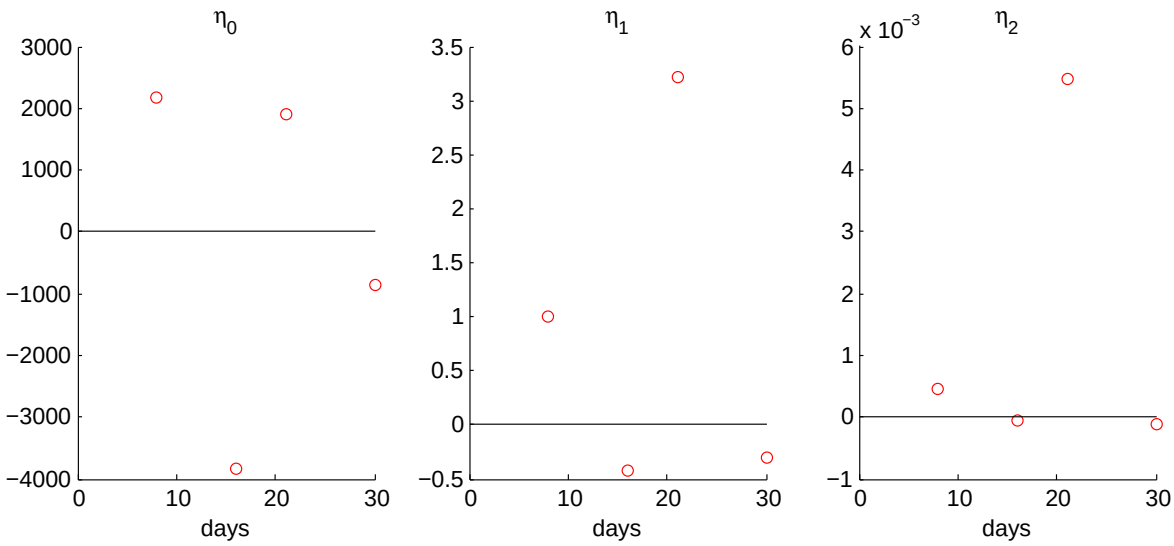

Figure 3.3 Values of $\eta_{d,JC,\hat{\gamma}}$ are plotted against time for $\hat{\gamma} = 0$ (left), $\hat{\gamma} = 1$ (middle), and $\hat{\gamma} = 2$ (right).

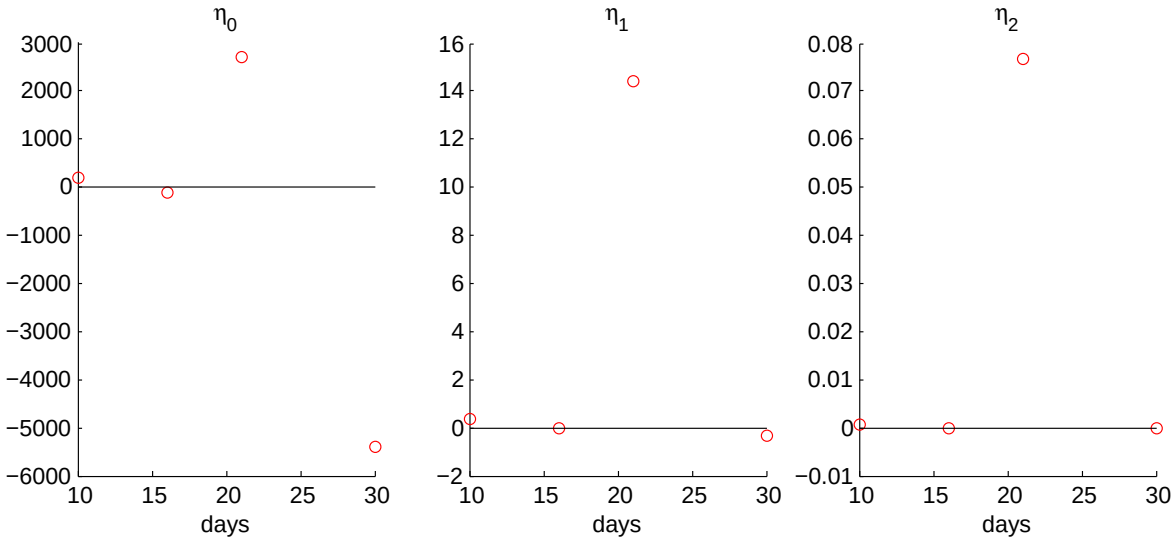

Figure 3.4 Values of $\eta_{d,SO,\hat{\gamma}}$ are plotted against time for $\hat{\gamma} = 0$ (left), $\hat{\gamma} = 1$ (middle), and $\hat{\gamma} = 2$ (right).

3.2.3 Model Variations

We additionally consider three variations on the ATN model. Two of these variations are simplifying restrictions of the model which require the estimation of fewer model parameters. We consider these models because of the sparsity of our data set, and not necessarily because we think that the models

are biologically meaningful. The third variation introduces an unknown mortality event during the season. Although we can only speculate about the potential mechanisms driving mortality, density data for the Aphid node does suggest that such an event occurs in many fields.

We refer to the original ATN model without mortality, as stated in (3.2), as Model 1. We consider the restriction of Model 1 to a parameter space where $h_0 = c_0 = 0$. This results in a linear functional response, or standard Lotka-Volterra type dynamics. The corresponding differential equation is given by,

$$\frac{dN_1}{dt} = N_1 \left[r(t) - \sum_{j \in \mathscr{C}_1} a_{1j}(a_0, R_{opt}, \phi) N_j \right], \tag{3.3}$$

for a_{1j} the attack rate under the ATN model parameterization. For this restriction, we note that there are only three parameters (a_0, R_{opt}, and ϕ) and the initial condition (N_1^0) to be estimated. We denote this as Model 2.

Parameter estimation on Model 2 yields values of $\phi \in [.3, 2]$, so we also consider the restriction of Model 2 to a parameter space where $\phi = 1$. This is equivalent to assuming that the sensitivity of successful attacks to predator-prey body mass ratio does not need to be tuned for the system. The corresponding differential equation is given by,

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

enforcing $\phi = 1$ in the ATN model parameterization. We refer to this as Model 3, where the only parameters to estimate are a_0, R_{opt}, and the initial condition N_0^1.

The aphid population data exhibits unexpected population declines that cannot reasonably be attributed to predation described by the ATN model. We therefore 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_d and the observation collected at t_{d+1}, then we denote the time $T_k = t_d$. We assume that the aphids suffer some extrinsic mortality factor μ over the interval $M_k = [T_k + 1, T_k + 2]$. We denote the set of all such times as $\mathbb{M} = \bigcup_k M_k$ and add this discrete mortality term to (3.3). The corresponding differential equation is given by,

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

where $\chi_{\mathbb{M}}$ is the indicator function for the set $\mathbb{M}$. We refer to this formulation as Model 4, and it requires the specification of four parameters (a_0, R_{opt}, ϕ, and μ) as well as the initial condition

(N_1^0). We note that Model 2 is a restriction of Model 4 to a parameter space where $\mu = 0$.

To compare performance between two models, we follow the methods described in Chapter 1 to test the *null hypothesis* that the nominal parameter set exists in a constrainted parameter space. We denote $J(\theta, Y_f)$ to be the cost functional associated with a parameter θ and observations $Y_f = \{Y_{d,f}\}_{d=1}^n$. For our problem, this reduces from the vector IRLS case previously described to a scalar ordinary least squares cost function,

$$J(\theta, Y_f) = \sum_{d=1}^{n} \big(Y_{d,f} - N_1(t_d;\theta)\big)^2.$$

We respectively define $\hat{\theta}$ and $\hat{\theta}^H$ to be solutions to the inverse problem for a model and its restriction to a subset of the full parameter space. Then the realization of the model comparison test statistic corresponding to the data set $\hat{N}_1^f = \{N_1^{d,f}\}_{d=1}^n$ is

$$\hat{U}_n(\hat{N}_1^f) = n\frac{J(\hat{\theta^H}, \hat{N_1}^f) - J(\hat{\theta}, \hat{N_1}^f)}{J(\hat{\theta}_1, \hat{N_1}^f)}.$$

We reject the *null hypothesis* with $(1-\alpha)\cdot 100\%$ certainty if $\hat{U}_n(\hat{N}_1) > \tau_\alpha$, for τ_α the critical value of the $\chi^2(r)$ distribution and r the additional degrees of freedom the full model has over the parameter restriction. We set the threshold for model rejection at 75%, since we are dealing with rather sparse experimental data sets.

3.3 Results

The fits for the ATN model and its three variations are plotted against experimental data in Figures 3.5 and 3.6. We denote with star markers the density data from each field, with error bars indicating a standard deviation of the data. The date at which crop maturation passed its critical threshold is indicated with a solid horizontal line, after which we no longer fit the models to data. All model fits are plotted with solid lines, with markers to differentiate between models - Model 1 is denoted by a triangle marker ($\triangle$), Model 2 by a circle marker ($\circ$), Model 3 by a square marker ($\square$), and Model 4 by an x marker ($\times$). We note that solutions for all four models are plotted in each figure. However, because of our conditions for a "mortality event," Model 4 does not generate dynamics unique from Model 2 in fields KC, MC, MO, and SO. See Table 3.7 for a list of parameter values for all fields and model variations.

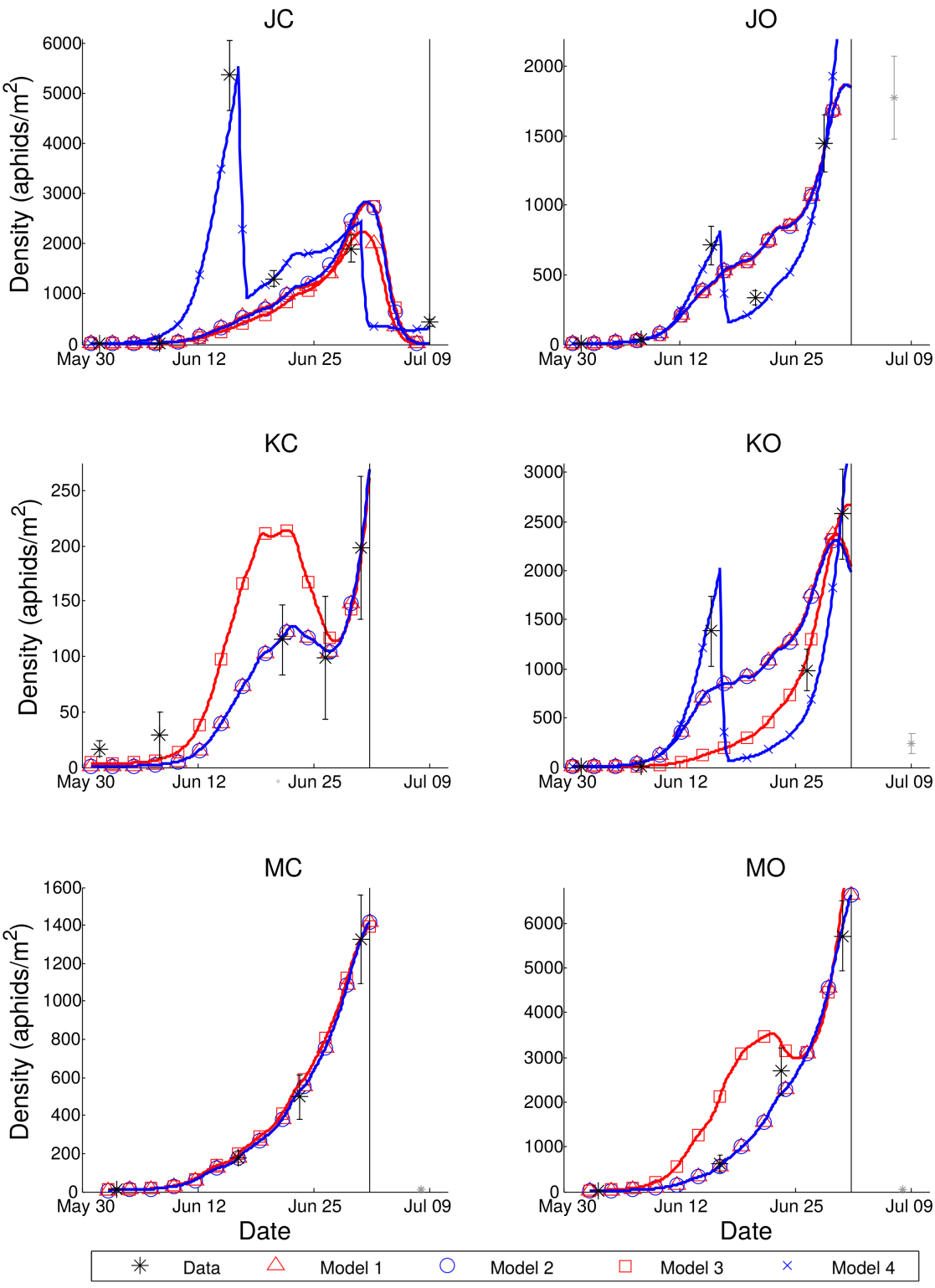

Figure 3.5 Ordinary least squares solutions of the ATN model and its three variations for data from six agricultural fields. Data is plotted with a star marker ($*$) with error bars to indicate the standard deviation in the data and the horizontal line indicates the cutoff date for data to be used in the inverse problem.

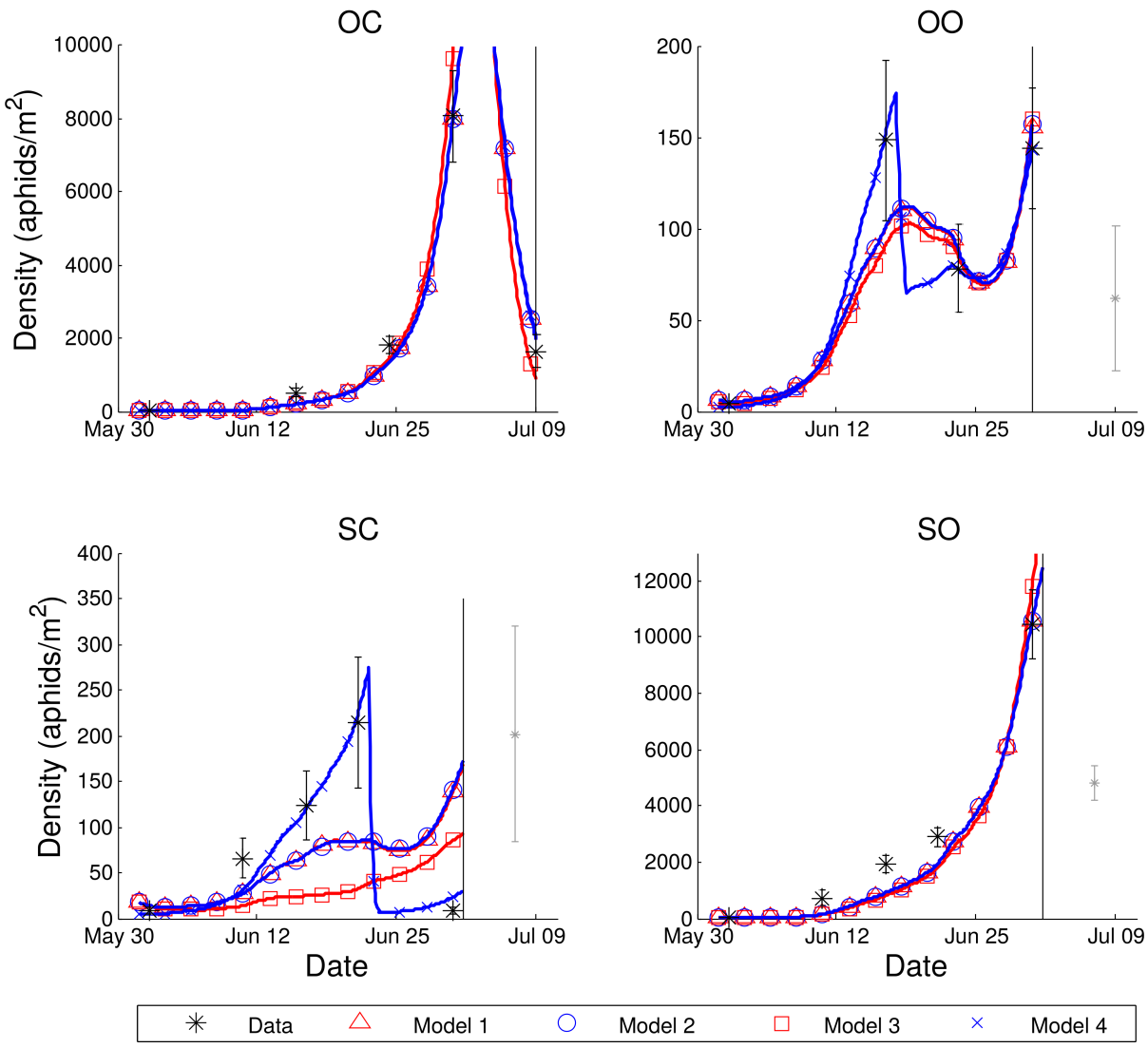

Figure 3.6 Ordinary least squares solutions of the ATN model and its three variations for data from four agricultural fields. Data is plotted with a star marker ($*$) with error bars to indicate the standard deviation in the data and the horizontal line indicates the cutoff date for data to be used in the inverse problem.

3.3.1 Model Comparison Tests

We first compare the performance of Model 1, the full ATN model, to its restriction to a linear functional response, $h_0 = c_0 = 0$, in Model 2. The solution to the inverse problem 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 3.4 a failure to reject the null hypothesis that the solution is contained in the restriction given by Model 2.

Table 3.4 Realizations of the nested model comparison test statistic for Model 2 a restriction of Model 1.

Field	JC	JO	KC	KO	MC	MO	OC	OO	SC	SO
$\hat{U}_n$	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

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. A linear functional response results in a model formulation which does not account for the time predators spend consuming prey, which means that predation force does not saturate in the presence of abundant prey items. This is an unreasonable model assumption, in direct opposition to the system's underlying biology. Mathematically, the linear functional response increases the predation force on the Aphid node. If our model for the population's intrinsic growth rate is too generous or the abundance of predators is underreported, then the linear functional response could be a means of attaining sufficiently slow growth of the Aphid node. With our current data set, there is insufficient information to support the estimation of $h_0, c_0 \neq 0$. However, we caution against drawing conclusions from this statistical result, particularly due to the sparsity of the data set.

We next assert the null hypothesis that the nominal parameters are contained in the restriction given by Model 3, fixing $\phi = 1$, instead of Model 2. The results from the model comparison test are given in Table 3.5, and we reject Model 3 with 75% or higher certainty in all fields except JC and JO. In these two exceptions, the estimated value of ϕ is close to 1, and so it is not surprising that the comparison test fails. We conclude that there is sufficient information in the data set to justify the estimation of ϕ in the ATN model. This parameter tunes sensitivity to R_{opt} in formulation of attack rates, and so it is worth noting that the significance of this parameter can vary with the distribution of predator body masses, which does vary across fields. Moreover, it is likely that the value of R_{opt} varies across guilds of predators - i.e., that the optimal predator-prey body mass ratio differs from beetles to spiders - and so we might estimate more consistent values of ϕ if we had sufficient information to estimate varying values of R_{opt}.

Table 3.5 Realizations of the nested model comparison test statistic for Model 3 a restriction of Model 2.

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%

We finally 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 4. The results from the comparison test, for the six fields that satisfy the requirement that a 25% or greater population decline occur between observations, are given in Table 3.6. 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.

Table 3.6 Realizations of the nested model comparison test statistic for Model 2 a restriction of Model 4.

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

Our primary concern with the estimation of ATN model parameters in these fields is that we do not know if the model *should* reproduce the significant population crashes observed in the data. The mortality might be driven by an abiotic factor outside the scope of our model, such as precipitation or wind, and we cannot verify such correlations without additional information about experimental conditions. However, the mortality might also be driven by predation within the scope of our model, such as by the Coccinella node. It is possible that our model, which is parameterized entirely by body mass, underestimates the effect of these specialist predators on the Aphid node. If there are external factors which can strongly impact the aphid population, then these are certainly worth taking into consideration when designing pest management strategies. However, if our model does not accurately describe the impact of predation on the pest population, then we need to revise the formulation of predator-prey interactions in the model.

Without knowing the cause of the mortality events, and considering the lack of certainty in our parameter estimation due to the sparse data set, it is difficult to say what "next step" in the iterative modelling process these results demand. As demonstrated in Field OC, it is possible to drive significant decline in the Aphid node by predation in the ATN model. However, comparison between Models 1 and 2 demonstrates that our current parameterization of the ATN model is not biologically reasonable, casting some doubt on the validity of our estimated parameters. In addition to requiring additional data to estimate our model parameters with some certainty, we should carefully consider hypotheses about the factors driving aphid dynamics and formulate future iterations of our model with these ideas in mind.

Table 3.7 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 4.

Field	Model	a_0	R_{opt}	ϕ	h_0	c_0	μ	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	2.56	428.92	-	-	-	-	4.62
	Model 4	3.69	0.54	1.01	-	-	1.98	9.99
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	3.04	803.08	-	-	-	-	2.86
	Model 4	2.52	823.56	1.20	-	-	1.84	1.46
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	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	0.88	0.87	-	-	-	-	2.90
	Model 4	0.02	24.93	0.05	-	-	3.68	1.64
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	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	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	0.29	37.78	-	-	-	-	1.00
	Model 4	0.18	46.75	0.21	-	-	0.01	1.41
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	0.95	0.81	-	-	-	-	4.86
	Model 4	0.18	2.12	0.40	-	-	1.08	2.32
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	0.57	146.55	-	-	-	-	18.19
	Model 4	2.07	478.73	2.00	-	-	4.01	4.94
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	1.68	227.45	-	-	-	-	5.00

3.3.2 ATN solution with additional data

For all fields, we find that the inclusion of h_0 and c_0 as unknown parameters in the inverse problem does not add 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 physically admissible range for each parameter is $h_0 \in [0, 0.42]$ and $c_0 \in [0, 0.24]$ in our system [62], and so this is a relatively small perturbation in the parameter value. In its current formulation, 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 3.7 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 3.8.

Table 3.8 Parameter values and initial iterates used in the inverse problem on synthetic data.

Parameter	a_0	R_{opt}	ϕ	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

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 count of the aphids on 100 barley plants. 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 again speculate that there could be some tradeoff between estimates of ϕ and R_{opt}, in which a shift in the value of R_{opt} across our diverse predator

community might be absorbed by an accompanying adjustment in the tuning parameter ϕ. We note that such a problem might be remedied if we were able to model the dynamics of the predator community, instead of only fitting our model to the cumulative effect on one prey item. There is a significant level of consumption within the predator community, and so it is more difficult to mask an erroneous specification of R_{opt} when validating the model against information about predation on *all* nodes in the system.

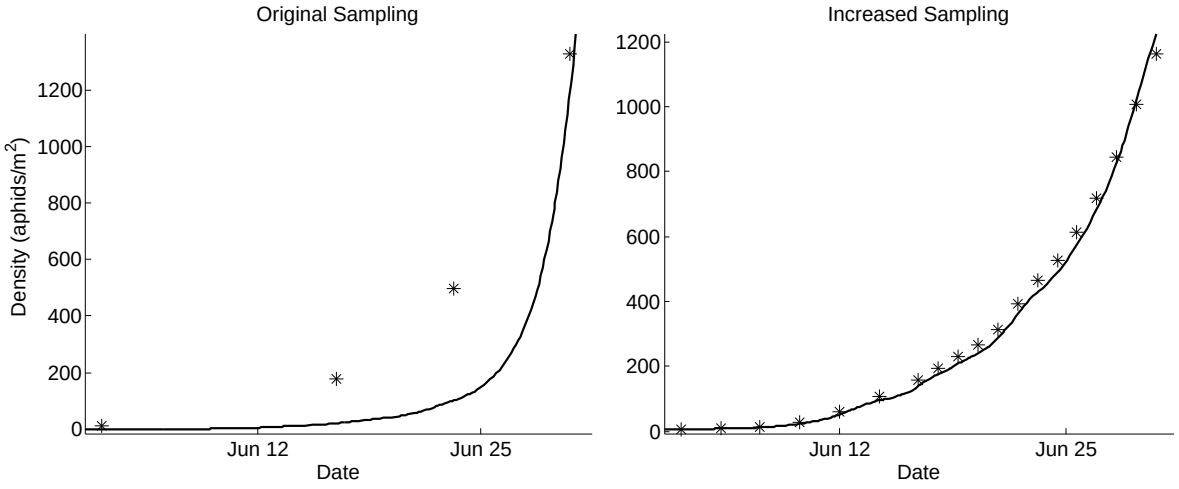

Figure 3.7 The solution to the inverse problem using the full ATN model with the original data (left) and a synthetic data set with increased sampling (right) is plotted with a solid line. The data used in the inverse problem is plotted with a star marker. Take care to note the differently scaled vertical axes.

3.4 Conclusion

Despite our inability to draw concrete conclusions about the effect of a given predator community on the Aphid node, the results presented here are a necessary step in the development of more accurate models for predator-prey interactions in our study system. This is particularly important in the context of our overarching goal to meet recent calls for an integration of food web ecology and ecosystem services [40]. As a proof-of-concept effort, we hoped to verify that our mathematical model and inverse problem methodology could be used to replicate the dynamics of a pest in an agricultural field. Although the inverse problem was not tractable for the original formulation of the ATN model, and efforts must be made to address unknown mortality events in our system, we are optimistic about the viability of our approach in future work addressing this problem.

We have demonstrated that the ill-posedness of the inverse problem can be ameliorated by reducing the number of parameters or increasing the data against which we validate our model.

Moving forward, we propose studying this model in a much simpler study system or collecting more data in the current study system, regarding both species abundances and external factors which might affect population dynamics. These results, paired with further formalization of biological hypotheses about the processes driving system dynamics, will inform our future formulations of a trait-based model for predator-prey dynamics in similar systems and attempts at validating such models against experimental data.

CHAPTER

4

DESIGNING GREENHOUSE EXPERIMENTS FOR MODEL TESTING

In this chapter, we outline the design of greenhouse experiments for validation of the Allometric Trophic Network (ATN) model. The work presented here formed the basis for a manuscript currently under review [69], and the resulting data is used in Chapter 5. We begin by motivating our approach, which explicitly links experimental design to model dynamics. We then describe our choice of study system, as it relates to the trait-based formulation of our model. We additionally formulate a sampling protocol by examining the mathematical model for system dynamics and the statistical model for error in our observations. We conclude by exploring the potential shortcomings of our design.

4.1 Motivation

A significant component of work in the fields of mathematical biology and ecology as a whole concerns the application of theoretical models to physical systems. Tailoring these models to real study systems, either through our approach of least squares minimization or an alternative approach of empirical parameter measurement, requires costly collections of data. Moreover, the validation of these models is an iterative process, in which we must verify that the model is applicable under a variety of conditons before we can confidently conclude that it describes the driving mechanisms

behind system behavior. The experiments required to obtain such data are necessarily dictated by practical constraits, and so it is not uncommon to opportunistically validate a model using data for which the model was never intended (our own work in Chapter 3 being an example of such an occurence). It is therefore vitally important that we inform and improve experimental design by in-depth model analysis, generating experiments better targeted at testing the model.

We propose intimately linking the design of our experiment with the development of our mathematical model to ensure the relevance of collected data. Our primary concern is ensuring that we collect sufficient data for model validation, subject to physical constraints. We supplement traditional design considerations with mathematical and statistical analyses to optimize the information content in our data set, while only sampling as much as is necessary. This approach requires the simulation of different sampling strategies prior to conducting the experiment, which permits a rigorous evaluation of the proposed strategy and highlights potential sources of uncertainty. This is a valuable exercise, as empirical experiments represent a significant investment of time and resources.

Due to the cost associated with experiments and the uncertainty in observational data, empirical scientists are careful to explicitly design experiments in terms of testable hypotheses. The conclusiveness of an experiment is determined by its design, which must target clear-cut predictions and measure well-defined quantities. The goal of our experiment is to test our mathematical model, and this broad goal does not immediately lend itself to specificity in design. However, mathematical models are formalizations of physical and biological interactions, built on our assumptions of a system's behavior. When we validate models against empirical data, we establish credibility by matching model dynamics to the observed outcomes; in this way, data-driven parameter estimation is a type of experiment intended to test the assumptions on which we built our model. We therefore design our experiment to test the underlying assumptions of our model, formally hypothesizing that body mass and habitat use govern predator-prey dynamics.

We note that in addition to guiding our experimental design, the explicit formulation of biological hypotheses sets the stage for stronger scientific conclusions. Experiments designed specifically to test and distinguish among *a priori* hypotheses permit strong inference [85], which is a desirable result. Some empirical sciences encourage this practice by promoting pre-registration of proposed experiments [2, 52, 81]. Pre-registration enables researchers to critically examine their hypotheses and design prior to running an experiment and invites feedback from the scientific community [52]. Following this template, we lay out our hypotheses and experimental design in [69] prior to conducting the analysis of the experimental data.

4.2 Study System

Our intent is to validate the variation on the ATN model described in Chapter 2, which incorporates our hypothesized mechanism for habitat use overlap. The model dynamics for the number of

individuals N_i of species i are given by

$$\frac{dN_i}{dt} = N_i \left[r_i - \sum_{j \in \mathscr{C}_i} \frac{a_{ij} \nu_{ij} N_i N_j}{1 + \sum_{k \in \mathscr{R}_j} a_{kj} \nu_{kj} h_{kj} N_k + \sum_{l \in \mathscr{C}_j} b_0 a_{jl} \nu_{jl} N_l} \right],$$

where r_i is the growth rate of species i, $\mathscr{C}_i$ is the set of species which consume species i, and $\mathscr{R}_j$ is the set of prey for species j. The parameters a_{ij} (attack rate), h_{ij} (handling time), and ν_{ij} (spatial overlap, or, similarity in habitat use) are given by

$$\begin{aligned}
a_{ij} &= a_0 W_i^{1/4} W_j^{1/4} \left(\frac{W_j/W_i}{R_{opt}} e^{1 - \frac{W_j/W_i}{R_{opt}}} \right)^{\phi}, \\
\nu_{ij} &= 1 - \nu_0 TV(\mu_i, \mu_j) = 1 - \nu_0 \sup_{\mathscr{A} \subset \Omega} \left| \mu_i(\mathscr{A}) - \mu_j(\mathscr{A}) \right|, \\
h_{ij} &= h_0 W_i^{1/4} W_j^{-1/4},
\end{aligned}$$

for species body mass W_i a measurable quantity and species habitat use μ_i defined from observations of individual behavior. We recall that for the set Ω of all modes of habitat use, we define the probability measure $\mu_i : \Omega \to \mathbb{R}$ such that evaluating $\mu_i(\mathscr{A})$ for some habitat $\mathscr{A}$ contained in Ω gives us the likelihood that species i occupies that habitat $\mathscr{A}$.

4.2.1 Cage Design

Due to the complications that arose from our field-level experiment, we conduct greenhouse experiments for this attempt at model validation. Although we eventually hope to understand predator-prey dynamics in a real agricultural field, a greenhouse experiment presents the opportunity to examine these dynamics without interference from ambient weather conditions and uncontrolled species introduction. Additionally, the necessarily small scale of a greenhouse experiment permits frequent sampling in time with a lower experimental burden. Our greenhouse experiments are conducted inside of "cages," which are wooden frames with a dirt base and mesh walls which prevent the escape of experimental subjects.

The aphid-based study system of terrestrial arthropods is appropriate for use in a greenhouse. Natural predators for the aphid are sufficiently diverse that we can identify species which vary in body mass and habitat overlap, a necessary condition for testing our experimental hypotheses. Additionally, aphids undergo rapid asexual reproduction, allowing us to run experiments over short periods while still observing the impact of predators on population growth. Due to this rapid growth, we limit the length of the experiment to 8 days; this period is sufficiently long to observe population growth, but sufficiently short to avoid conditions of overcrowding and resource scarcity which trigger complicated modes of reproduction in aphid populations.

To study the effects of body size and habitat use, we choose predators which vary in both of these traits. We choose two predators which are primarily ground predators (Wolf spiders of the genus *Pardosa* and ground beetles of the genus *Bembidion*) and two which primarily reside in the foliage (lady beetle *Coccinella septempunctata* and the minute pirate bug *Orius majusculus*). We additionally note that preliminary observations suggest *Bembidion* is the most restricted in habitat use, while *Pardosa* sometimes climbs the plants, and both of the foliage predators spend time on the ground; there is substantial overlap in habitat use, and all four predators can be expected to encounter one another. Within these two categories, we can further divide the predators into being small-bodied (*Bembidion* and *Orius*) or large-bodied (*Pardosa* and *Coccinella*). There are many species which could be divided into these categories, but we have selected a diverse group of predators (a spider, a beetle, a bug, and a specialist predator) in order to demonstrate that our hypothesized traits are important across different predators. We present this information in Table 4.1, for reference.

Table 4.1 Species traits for the four predators in our experiment. In parentheses, we designate the subject species' unique contribution within the experiment.

	Foliage-Dwelling	Ground-Dwelling
Small-Bodied	*Orius* (bug)	*Bembidion* (beetle)
Large-Bodied	*Coccinella* (specialist)	*Pardosa* (spider)

We additionally explore the effect of body mass and habitat use in determining interactions by including two species of aphid prey items. We choose a large-bodied aphid (*Acyrthosiphon pisum*), which primarily resides on bean leaves with a prefence for the undersides of the plant. We also choose a small-bodied aphid (*Rhopalosiphum padi*), which lives on barley plants with a preference for the bottom of the stem. Although both species of aphid primarily reside in their respective foliage, *R. padi* is more accessible to ground predators due to its location on the plant. We note that individuals of species *A. pisum* will frequently drop from the underside of the bean plant when neighboring aphids are disturbed by a foliage-dwelling predator, making this prey item temporarily accessible to ground-dwelling predators before the aphids reestablish themselves on nearby bean plants.

The use of these two aphids necessitates that we seed our cages with barley and bean plants. We plant the beans and barley in alternating rows within the cage prior to running the experiment, initially growing the plants in abundance before thinning the plant density to a uniform level at the start of the experiment. We establish a population of aphids in the cage for two days, permitting sufficient time for population growth to begin, before introducing predators to the cage. Each cage

has a designated "prey treatment" and "predator treatment." The three types of aphid treatments are: *R. padi* only, *A. pisum* only, or *R. padi* and *A. pisum* in combination. We take all pairwise combinations of our four predators (including a predator matched with itself) for ten predator treatments. We additionally study a predator-free control treatment for all three aphid treatments. Taking each combination of predator and prey treatments, we have 33 unique cage designations, each of which we replicate six times to account for natural variation in the systems.

The body masses and initial abundances of all species in our system are given in Table 4.2. We note that the initial abundance of a species is doubled if it appears individually in a treatment; for example, if *Orius* is the only predator in a cage, then we include two populations of *Orius* for a total of 40 predators.

Table 4.2 The body masses and initial abundances for all species used in the mesocosm experiments.

	A. pisum	*R. padi*	*Bembidion*	*Coccinella*	*Orius*	*Pardosa*
W_i (mg)	0.6706	0.1550	2.145	37.4636	0.58	17.72
$N_{i,0}$ (individuals)	75	75	20	2	20	10

4.2.2 Habitat Use

The parameter ν_{ij} can be interpreted as a measure of similarity in spatial distribution between species in our system. To motivate the mathematical formulation of this parameter, we outline its physical value for species in our experiment, using preliminary observations of habitat use in the greenhouse cages. In Table 4.3, we state the assumed probability that each species will use a particular region of the cage. We obtained these probabilities by observing the cages on multiple occasions and approximating the fraction out of all observations where the species was within each region.

Recall that for calculation of the overlap parameter we must first define the set Ω of all modes of habitat use and the probability measure μ_i that a species utilizes some habitat within that set. From these observations, we define $\Omega = \{P, N, G, PN, PG, NG, PNG\}$ where P indicates use of the plant region, N indicates use of the netting region, and G indicates use of the ground region. We then have PN indicating use of the plant *and* netting regions, PG indicating use of the plant *and* ground regions, and so forth.

We next define the probability measures μ_i of each species i for habitat use within the set Ω. We

state, for example, the definition of μ_A (for species *A. pisum*) on the set Ω:

$$\mu_A(P)=0.9,\ \mu_A(N)=0,\ \mu_A(G)=0.1$$
$$\mu_A(PN)=0.9,\ \mu_A(PG)=1,\ \mu_A(NG)=0.1,\ \mu_A(PNG)=1$$

For comparison, we give the same values of μ_B (for species *Bembidion*):

$$\mu_B(P)=0.05,\ \mu_B(N)=0.2,\ \mu_B(G)=0.75$$
$$\mu_A(PN)=0.25,\ \mu_B(PG)=0.8,\ \mu_B(NG)=0.95,\ \mu_B(PNG)=1$$

Table 4.3 Probability that species from the cage experiments occupy various regions of the cage. We split habitat use into the categories "plant," "netting," and "ground." We note that the "ground" region includes the base of the plants, which can be reached by ground-dwelling predators without climbing plants.

	Plant	Netting	Ground
A. pisum	0.9	0	0.1
R. padi	0.6	0	0.4
Bembidion	0.05	0.2	0.75
Coccinella	0.5	0.25	0.25
Orius	0.5	0.25	0.25
Pardosa	0.05	0.35	0.6

We next compute for all species pairs in the system, the total variation $TV(\mu_i,\mu_j)$ in the habitat use distributions. We recall that the definition of total variation is given by

$$TV(\mu_i,\mu_j)=\sup_{\mathscr{A}\subset\Omega}\left|\mu_i(\mathscr{A})-\mu_j(\mathscr{A})\right|$$

and can be interpreted as dissimilarity between the distributions μ_i and μ_j. For the example we began above, we note that this largest difference occurs when we look at the sets P or NG, since

$$\left|\mu_A(P)-\mu_B(P)\right|=\left|\mu_A(NG)-\mu_B(NG)\right|=0.85,$$

and this is the largest magnitude difference we see over all sets in Ω. The total variations for all species pairs in our system are given in Table 4.4.

We note that a larger value of $TV(\mu_i,\mu_j)$ indicates a higher dissimilarity in habitat use; there exists some mode of habitat use in Ω for which the probability that species i uses that region is

very different from the probability that species j uses that region. Because the values of $\mu_i(\mathscr{A})$ must fall between zero and one by the definition of a probability measure, we will necessarily have $0 \leq TV(\mu_i, \mu_j) \leq 1$. We can therefore see that *A. pisum* and *Bembidion* use the habitat within the cage in a very dissimilar way. However, we see that $TV(\mu_i, \mu_i) = 0$, since a species will always match its own probability of occupying a given region. We finally note that we always have $TV(\mu_i, \mu_j) = TV(\mu_j, \mu_i)$; this is biologically reasonable, since we do not assume that habitat use will asymmetrically affect the two species.

Table 4.4 Total variation $TV(\mu_i, \mu_j)$ for species in the cage experiments. We calculate these values using the probability measures for habitat use defined in Table 4.3

	A. pisum	*R. padi*	*Bembidion*	*Coccinella*	*Orius*	*Pardosa*
A. pisum	0	0.3	0.85	0.4	0.4	0.85
R. padi	0.3	0	0.55	0.25	0.25	0.55
Bembidion	0.85	0.55	0	0.5	0.5	0.15
Coccinella	0.4	0.25	0.5	0	0	0.45
Orius	0.4	0.25	0.5	0	0	0.45
Pardosa	0.85	0.55	0.15	0.45	0.45	0

The final step in computing habitat use *overlap* is to convert the total variation into a measure of *similarity* instead of dissimilarity. Since we know that $0 \leq TV(\mu_i, \mu_j) \leq 1$, we subtract this quantity from 1 to obtain a quantity which increases with similarity of habitat use but still falls between 0 and 1. We introduce some scaling factor $0 \leq \nu_0 \leq 1$ to quantify the effect of assumed habitat use on spatial overlap, which gives the parameter $\nu_{ij} = 1 - \nu_0 TV(\mu_i, \mu_j)$.

4.3 Sampling Protocols

In defining the sampling protocol, we must balance what is a reasonable undertaking with the data required for parameter estimation. Since different methods of data analysis have different criteria for meaningful information content, we must identify the method of analysis when establishing sampling protocol. This choice determines the frequency with which we must collect data and the sampling strategies we can implement. We employ a data-driven least squares minimization to estimate model parameters, as described in Chapter 1. Although other methods exist for the estimation of ATN model parameters (see, for example, examining interaction strengths as in [101]), solving a least squares inverse problem allows us to fit model dynamics to data in time and preserve maximal information about physical processes.

To effectively solve a least squares problem, we must collect sufficient data throughout the experiment, but obtaining data at an appropriate temporal scale is a time-consuming and expensive endeavour. It is therefore advantageous to know the optimal timing for data collection during the experiment, with the goal of using this data to estimate model parameters a_0, R_{opt}, ϕ, h_0, b_0, and v_0. Our method of parameter estimation relies on fitting a mathematical model to time-series data, and observations from different days can vary in how much information they provide about model parameters. In order to focus our effort on the days which provide the most information and ensure maximal information content in the data, we examine the sensitivity of aphid populations (the primary source of data in our experiment) to parameter inputs as time progresses. After we determine the optimal timing for these samples, we will investigate the feasibility of subsampling strategies, which reduce the burden of data collection.

4.3.1 Timing for Population Samples

We compute the sensitivity of a population density N_i with respect to a parameter θ, $s_i^\theta = \frac{dN_i}{d\theta}$, by solving the sensitivity equations [10, 11] as given in Chapter 2. To facilitate comparison of sensitivities between treatments, we compute the relative sensitivity [46, 49], given by

$$sr_i^\theta(t) = \frac{\theta}{N_i(t,\theta)} s_i^\theta(t).$$

To avoid giving undue importance to aphid populations approaching zero, we do not normalize by the population when $N_i(t,\theta) < 1$. We refer to sensitivities of great magnitude (regardless of sign) as "high," indicating that model solutions depend strongly on a given parameter. To effectively estimate parameters for our model, we must collect data on days with high sensitivity; we obtain less information about the parameters by sampling on days with low sensitivity. We are primarily concerned with the sensitivities of aphid population densities with respect to model parameters, since we cannot obtain predator population data during the experiment.

We present sensitivity results for $\theta = a_0, \phi, v_0, h_0, b_0$ using parameter values $a_0 = 24 \times .9$, $\phi = 1$, $v_0 = 1$, $h_0 = 2/24$, $b_0 = h_0$, $R_{opt} = 60$ for *Bembidion*, $R_{opt} = 115$ for *Coccinella*, $R_{opt} = 1$ for *Orius*, and $R_{opt} = 60$ for *Pardosa*. We plot in Fig 4.1 the sensitivities of aphid populations to model parameters for four cage treatments, which exhibit the types of behavior we see in the sensitivities as a function of time. In Table 4.5, we describe the conditions of the four treatments.

We cannot know the model parameters prior to running the experiment, and so our parameters are informed by estimates from similar mesocosm experiments [62]. We estimate R_{opt} from personal observations of predator feeding preferences during planning stages of the experiment; because we manually select values of R_{opt}, we do not include this parameter's sensitivity results in our consideration of sampling protocol. From control-treatment "trial runs" during planning,

we estimate intrinsic growth rates $r = 0.3007$ for *A. pisum* alone, $r = 0.3211$ for *R. padi* alone, and $r = 0.2453$, $r = 0.2591$ for *A. pisum* and *R. padi* (respectively) in a combined treatment.

Table 4.5 Key for cage treatments referenced in Fig 4.1.

Treatment	Prey	Predator 1	Predator 2
R-BP	*R. padi*	*Bembidion*	*Pardosa*
R-C	*R. padi*	*Coccinella*	*Coccinella*
A-OP	*A. pisum*	*Orius*	*Pardosa*
A-B	*A. pisum*	*Bembidion*	*Bembidion*

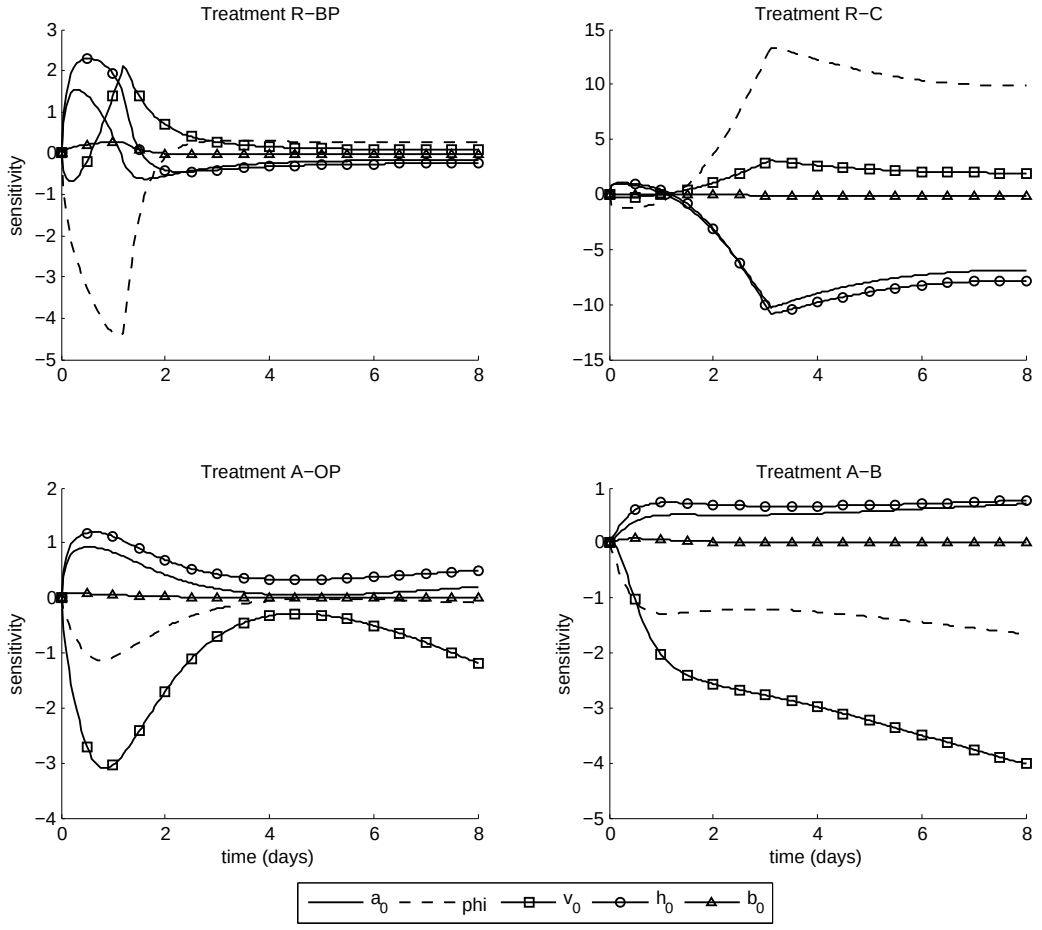

Figure 4.1 Sensitivities of aphid population abundances with respect to model parameters. Sensitivities are plotted in time and different parameters are indicated by line style.

Treatment R-BP represents the most common type of sensitivity behavior we see, characterized by high sensitivities on or before $t = 2$, which quickly decrease to low levels for the remainder of the experiment. Solutions with this type of sensitivity describe aphid populations which are quickly decimated by the predator treatment and do not recover. Treatment R-C represents the second type of behavior for model sensitivities, characterized by low sensitivities until $t = 2$, after which sensitivities quickly increase to their peak around $t = 3$ and remain at high levels for the duration of the experiment. Solutions with this type of sensitivity describe aphid populations which experience eventual decimation, but the decline is slightly delayed. Treatment A-OP represents the third type of behavior for model sensitivities, in which we see high sensitivity early in the experiment and a decrease to low values around $t = 4$, with a slight rebound towards the end of the experiment. Solutions with this type of sensitivity describe aphid populations which experience drastic decline at the beginning of the experiment but manage to recover when predators become scarce later in the experiment. Treatment A-B represents the final type of sensitivity behavior we see, with constant or increasing sensitivities for most of the experiment and the most significant increase in model sensitivities before $t = 2$. Treatments of this type do not result in a drastic decline in the modeled aphid population; the population is increasing for at least several days of the experiment.

We caution that our solutions to the sensitivity equations are local to the assumed parameters in our model formulation, and therefore may change for different parameter inputs. If any parameters estimated from our experiment differ significantly from these assumed values, the model sensitivities to *all* parameters may not follow the patterns we observe here. To minimize the risk of missing the sensitivity of our "true" model solution, we obtain assumed parameter ranges from a related experiment and verify that similar peaks in sensitivity are obtained for different parameter values in this range. We cannot draw conclusions about the specific cages which belong to each of these categories for the experiment, where true model parameters may be significantly different from what is assumed here. Rather, we present an example of the types of behaviors that model sensitivities might exhibit, regardless of the treatments under which they truly occur.

Based on these results, we conclude that sampling on days $t = 2, 4$, and 8 will yield data with high information content related to all model parameters, in at least some treatments. In the interest of obtaining maximal information in treatments with rapid aphid population decline, we suggest additional samples on days $t = 1$ and 3 when possible. We obtain less information with repeated sampling of such treatments later in the experiment, when the population has already been decimated; the rate of decimation, occurring quickly over the first few days, is of greater interest. Under this sampling scheme, we can be as confident as possible that the collected data contains sufficient information for parameter estimation.

Convergence of a least squares parameter estimation requires substantial, temporal data. There is no consistent translation of this requirement to some number of data points required to obtain a good solution to the inverse problem. However, we tentatively assume that to estimate our five

model parameters, we should inform the inverse problem with at least ten data points. Grouping our cages into cohorts of the three aphid treatments, with predator treatments fixed, yields nine data points when using our proposed samples on days $t = 2, 4, 8$. The alternative of grouping our cages into cohorts of the ten predator treatments, with aphid treatments fixed, for 30 data points is perhaps drastic. Since we can only theorize that dynamics will be the same for all predators, despite significant predator diversity, we hesitate to rely on such a broad categorization of treatments. We therefore compromise by adding an additional sample on day $t = 6$, which may not be a day with particularly high parameter sensitivities but does give additional information.

4.3.2 Subsampling Protocol

Our choice of timing for population samples presents an enormous effort, requiring many man-hours of work in the lab. With limited resources, obtaining this data becomes a trade-off between sampling thoroughly at every time or sampling frequently in time. We therefore consider the potential for subsampling of the aphid population in order to obtain data more quickly on these days. To subsample the aphid population, N_j, at time t_j in a cage with T plants, we count the number of aphids on $n < T$ plants, $\tilde{N}_j^n$, and scale by T/n to obtain an estimate of the aphid population, $N_j^n = \frac{T}{n}\tilde{N}_j^n$. We investigate the effect that any error in N_j^n might have on our ability to estimate model parameters. We first consider this problem in predator-free control cages for *R. padi*, for which we had previous data to consult.

We assume exponential growth in predator-free control cages of *R. padi*, with the model solution given by $f(t, r) = 150e^{r_0(t+1)}$, assuming that $t = 0$ one day after aphids are introduced to the cage. We temporarily suppress the species indexing and let N denote the aphid population and N_j the sampled population at some time t_j. In current experiments, the protocol is to obtain some estimate $\hat{r}$ to the growth rate r_0 in predator-free control cages, which is in turn used for the full ATN model in predator-treated cages. We estimate this parameter by seeking the ordinary least squares minimization,

$$\hat{r} = \arg\min_{r \geq 0} \sum_{j=1}^{M} (N_j - f(t_j, r))^2.$$

If we instead use the subsampled approximations to the population, N_j^n, in the least squares formulation, we obtain the estimate

$$\hat{r}^n = \arg\min_{r \geq 0} \sum_{j=1}^{M} (N_j^n - f(t_j, r))^2.$$

To investigate the effect of subsampling on the estimated growth rate, we generate synthetic data with a known growth rate and compare the error in parameter estimation under a subsampling scheme for different values of n. The provided data includes the per-plant aphid counts, y_j^k, for

six cages, collected on days $t_j = 4, 8$. For each cage, we estimate the growth rate $\hat{r}$ from the fully sampled population counts, N_j. The resulting parameter estimates for the six cages are given in Table 4.6.

Table 4.6 Estimated growth rates $\hat{r}$ for each of the six replicate control cages. The average across all growth rates is $\overline{r} = 0.4145$, with an average value of $|\hat{r} - \overline{r}| = 0.0138$.

Cage	Cage 1	Cage 2	Cage 3	Cage 4	Cage 5	Cage 6	Mean
$\hat{r}$	0.4204	0.4045	0.3830	0.4233	0.4233	0.4323	0.4145
$\lvert\hat{r} - \overline{r}\rvert$	0.0059	0.0099	0.0315	0.0088	0.0088	0.0178	0.0138

For a cage with growth rate $\hat{r}$, we compute the normalized difference from model output, ϵ_j^k, at time t_j for each plant k. Assuming that aphids are uniformly distributed across the cage's T plants, we have

$$\epsilon_j^k = \frac{1}{f(t_j, \hat{r})/T} \left(y_j^k - \frac{f(t_j, \hat{r})}{T} \right).$$

For $\overline{r}$ the average of the estimated growth rates across all six cages, we compute the true population curve $f(t, \overline{r}) = 150 e^{\overline{r}(t+1)}$. We generate synthetic cages with $T = 90$ plants which have true population $f(t, \overline{r})/T$ and have been sampled on days $t_j = 4, 8$. We randomly draw observational noise for each plant from the set of all normalized errors ϵ_j^k for a fixed j.

For each synthetic cage, we consider subsampling strategies in which $n = 5, 6, \ldots, 90$ plants are counted. For each strategy, we estimate the growth rate $\hat{r}^n$ from the approximated population N_j^n. We compute the resulting error $|\hat{r}^n - \overline{r}|$ for all n, and repeat this process for 200 synthetic cages. We plot the average error $|\hat{r}^n - \overline{r}|$ for each subsampling strategy in Fig 4.2. The dashed line in the figure indicates the average value of $|\hat{r} - \overline{r}|$ for the estimated growth rates in the six true data sets.

We note that although the strategy prone to the least error is to sample all 90 plants, the accuracy gained from sampling additional plants drops significantly around $n = 25$ plants. Sampling around $n = 40$ plants results in error that is on the same scale as the average value of $|\hat{r} - \overline{r}|$. That is, the error induced by subsampling is no more drastic than the natural variation we see in parameter estimates across controlled replicates; if we are comfortable averaging out the latter and incorporating this parameter in our model, then the former should be similarly manageable. We therefore conclude that sampling 45 out of 90 plants (or, 1-in-2 subsampling) yields sufficiently accurate parameter estimates. If the burden of data collection is too high, sampling 30 out of 90 plants (1-in-3 subsampling) is justifiable, and sampling as high as 60 out of 90 plants (2-in-3 subsampling) seems to be a very safe, if intensive, strategy.

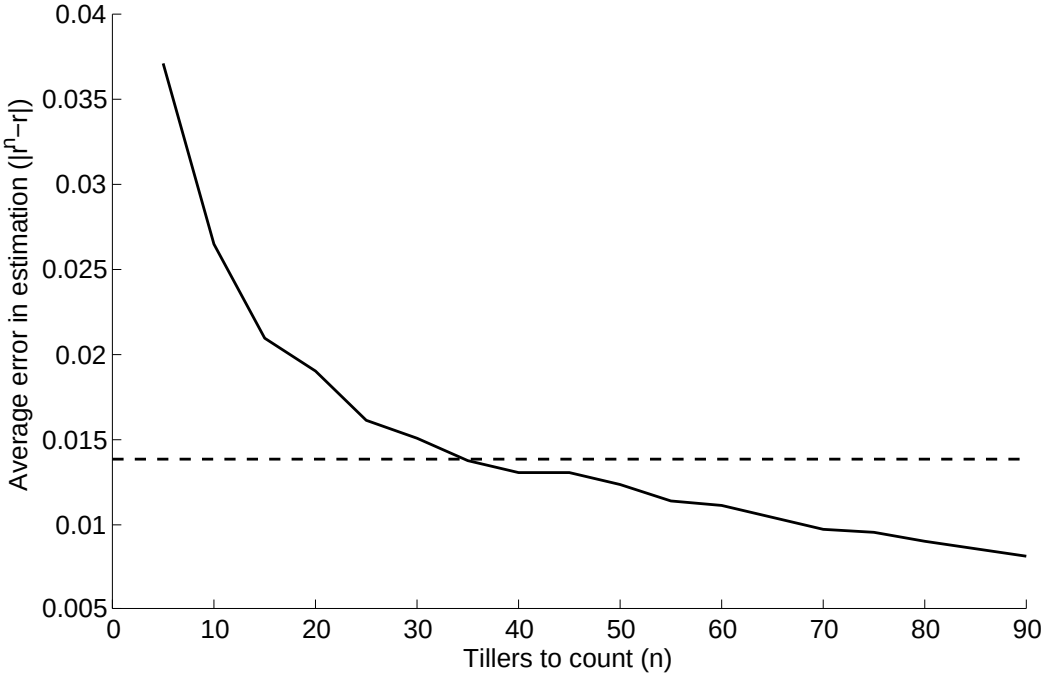

Figure 4.2 The average value of $|\hat{r}^n - \overline{r}|$ across 200 synthetic cages. The horizontal dashed line indicates the average value of $|\hat{r} - \overline{r}|$ for parameter estimates using the true data set.

We next consider the effect of error induced by subsampling on our estimates of ATN model parameters. We obtained per-plant aphid counts for experimental cages with *R. padi* as the basal species, but sometimes under different predator treatments than our intended study communities [62]. Additionally, previous data is not sampled sufficiently in time to permit the estimation of ATN model parameters. We must directly infer from this experimental data the distribution of noise induced by subsampling and add this noise to a simulation of cages which we assume perfectly follow ATN dynamics.

To investigate the effect of subsampling on parameter estimation for the ATN model, we first categorize available mesocosm data by full-cage population values. For all replicate cages of a single treatment and fixed times t_j, we have $N_j \in [0,100]$, $N_j \in [150,1500]$, or $N_j \in [4000,9000]$. We define the set of all plant counts y_j^k which correspond to full-cage population values in the range of $N_j \in [0,100]$ as "Category 1," the set of all plant counts y_j^k which correspond to full-cage population values in the range of $N_j \in [150,1500]$ as "Category 2," and the set of all plant counts y_j^k which correspond to full-cage population values in the range of $N_j \in [4000,9000]$ as "Category 3." We do not make any distinctions between plant counts based on the treatment from which they originated or time at which they were obtained.

We generate synthetic cages for a given category by randomly sampling 90 plant counts, $\{\hat{y}^k\}_{k=1}^{90}$,

from the category. We compute the true population of the synthetic cage,

$$\hat{N} = \sum_{k=1}^{90} \hat{y}^k,$$

and a subsampled approximation to the true population,

$$\hat{N}^n = \frac{90}{n} \sum_{k=1}^{n} \hat{y}^k,$$

for $n = 30, 45$, and 60. We then compute the normalized error induced by subsampling,

$$\hat{\epsilon}^n = \frac{1}{\hat{N}}(\hat{N}^n - \hat{N}),$$

and repeat this process for 1000 synthetic cages. From the synthetic cages, we attempt to identify the distribution of errors $\hat{\epsilon}^n$ for each category.

In Figure 4.3, we present the scatter plots and histograms for the normalized errors induced by the three subsampling procedures for Category 3. We note that the distribution of $\hat{\epsilon}^n$ is close to normal, with the variance of the distribution increasing at lower subsampling rates.

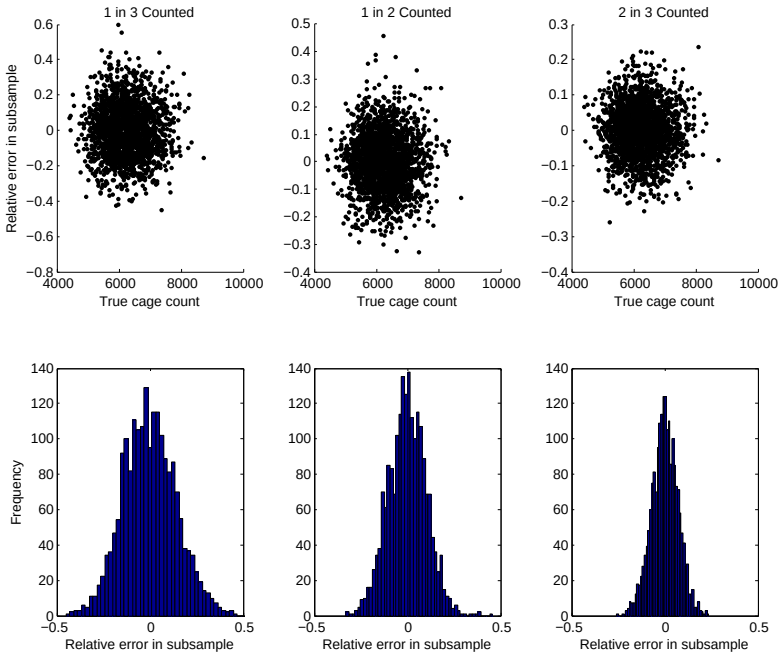

Figure 4.3 Normalized errors induced by subsampling ($\hat{\epsilon}^n$) for synthetic cages generated from plant counts in Category 3. We take $n = 30, 45, 60$ (1-in-3, 1-in-2, and 2-in-3 subsampling strategies, respectively).

In Figure 4.4, we present the scatter plots and histograms for the normalized errors induced by the three subsampling procedures for Category 2. We note that the histograms again indicate a normal distribution, with variances at a given subsampling rate higher than the corresponding variances in Category 3. However, the scatter plots seem to indicate that at the lower-range of populations in Category 2, the relative error induced by subsampling begins to skew.

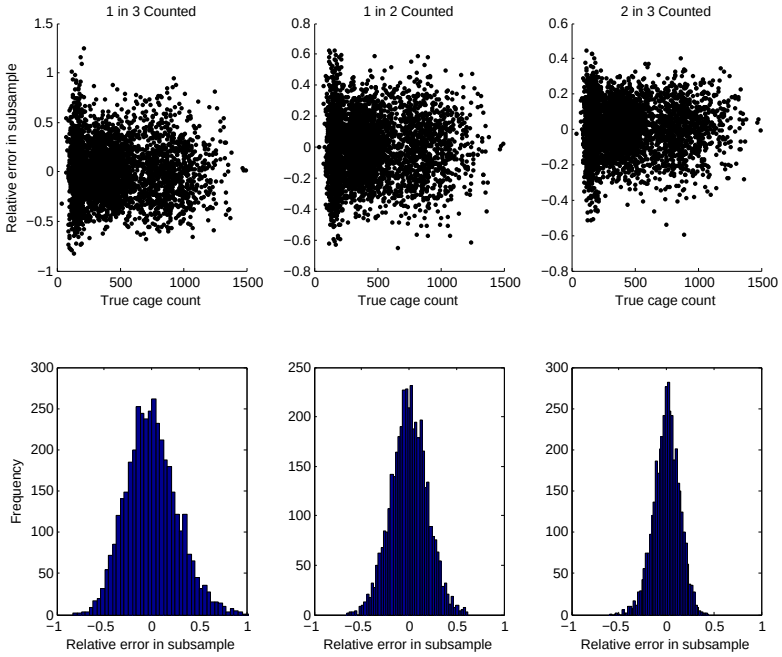

Figure 4.4 Normalized errors induced by subsampling ($\hat{\epsilon}^n$) for synthetic cages generated from plant counts in Category 2. We take $n = 30, 45, 60$ (1-in-3, 1-in-2, and 2-in-3 subsampling strategies, respectively).

In Figure 4.5, we present the scatter plots and histograms for the normalized errors induced by the three subsampling procedures for Category 1. We observe in the histograms a truncated statistical distribution for all three sampling strategies and cannot identify a distribution for the error induced by subsampling for cages in this category. There is a clear bias towards the upper and lower limits of $\hat{\epsilon}^n$ when the aphid population is low, which we intuitively explain by considering the example of a single aphid in a cage where we subsample n of T plants. The only possible approximations are $\hat{N}^n = T/n$ or 0, and unless $n = T$, we will not accurately report the population in the cage. This causes the tendency towards fixed values of $\hat{\epsilon}^n$ in populations close to zero.

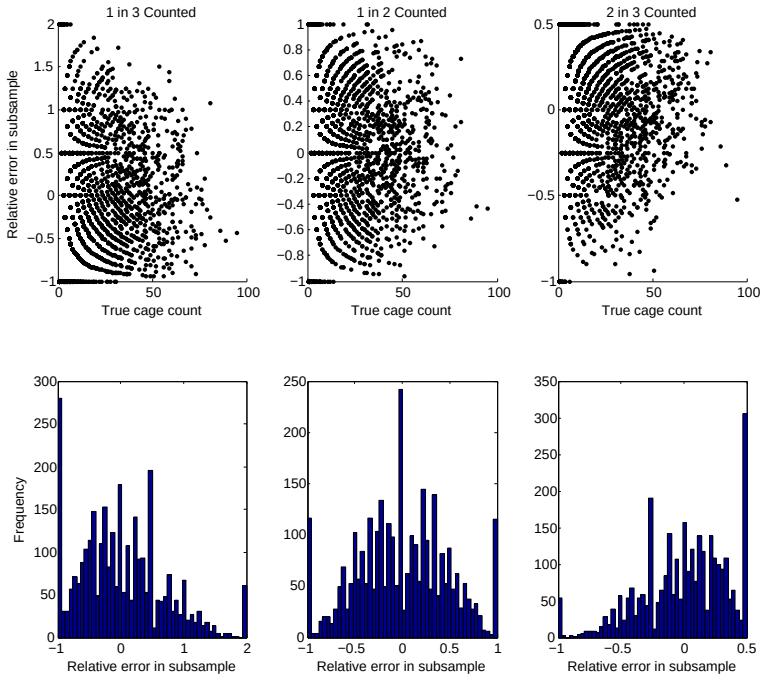

Figure 4.5 Normalized errors induced by subsampling ($\hat{\epsilon}^n$) for synthetic cages generated from plant counts in Category 1. We take $n = 30, 45, 60$ (1-in-3, 1-in-2, and 2-in-3 subsampling strategies, respectively).

We conclude that for full-cage populations greater than 150 aphids, the normalized error induced by subsampling n plants is normally distributed, with variance decreasing as the aphid population increases. However, we note that when the aphid population is at very low values, it is easier to make the worst possible over- and under-estimates of the population when subsampling. When generating synthetic data for aphid densities in the range of $(150, \infty)$, we assume that the relative error induced by each subsampling strategy is normally distributed, with the standard deviation a function of aphid density. Using the distributions seen in categories 2 and 3, we compute the standard devitiation as a linear function of population for our synthetic data. For aphid populations in $[0, 150]$, we generate synthetic data by manually sampling from the distribution of possible relative errors. We generate the distribution using 10000 synthetic cages, and construct at most 6 synthetic data sets with 8 points each.

We generate synthetic data with dynamics given by ATN model solutions and error induced by subsampling for use in an inverse problem in which we incorrectly specify ϕ and R_{opt} and estimate a_0, h_0, v_0, and b_0. We plot two sample outcomes in Figure 4.6 and note that visual fits to the true model solution are poor for aphid populations below 150 (Pardosa-treated populations). This problem occured regardless of subsampling strategy and did not have a significant effect on the error in the estimated parameters. Although the error in our parameter estimation did not vary

considerably in these cases, we caution against implementing a subsampling strategy for cages with low population densities until more information is available.

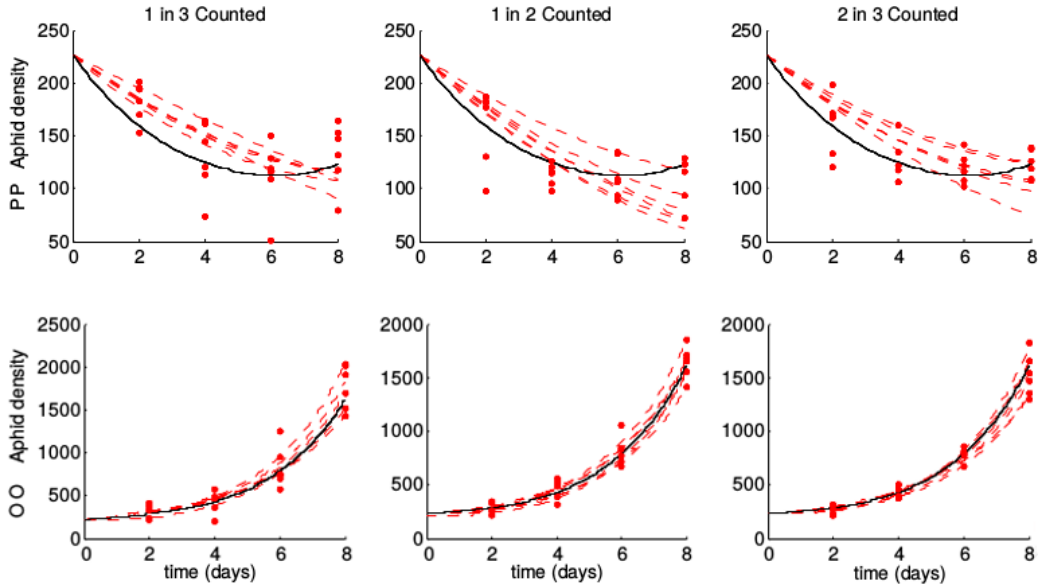

Figure 4.6 The true model solution (solid line) is plotted against six replicates of synthetically subsampled data (red markers) and the resulting estimated model solutions (dashed lines) for each subsampling scheme of *R. padi* populations in Pardosa-treated cages (top, labelled PP) and Orius-treated cages (bottom, labelled OO).

4.3.3 Terminal Sampling Protocol

A previous cage experiment [62], which we have used as the basis for this work, employed destructive sampling on the final day of data collection to obtain estimates of aphid and predator abundances. When a cage is destructively sampled, each plant is cut at ground level and removed from the cage before aphids are counted. Aphids that are obscured from view within the plants can be found with destructive sampling, and we therefore expect that destructive sampling yields a more accurate count of the population. However, it is not possible to destructively sample a cage more than once, meaning we cannot sample this way before the end of the experiment.

We recall the importance of a study system's adherence to both the *mathematical model of dynamics* and the *statistical model of error* in the data in estimating model parameters. In particular, misspecification of the statistical model can result in inverse problem solutions which appear to fit the data without matching underlying behavior. This is a problematic outcome, as it is impossible to tell from standard errors in our estimates and visual fit alone if the data is accurately described by the models [10]; model-based analysis in this situation could be wrong without any indication

of the underlying issue. When establishing the experimental sampling protocol, we must consider the underlying statistical model of error that such a protocol suggests and how we might be able to identify this model.

Unfortunately, we cannot describe the relationship between the error in non-destructive and destructive samples at all aphid density levels without destructively sampling over a range of predator treatments for the duration of the experiment. Without the space and resources to destroy a set of cages on every day that data is collected, we refrain from destructively sampling the aphid population on the final day. We are therefore able to statistically model error with fewer unquantifiable assumptions, despite trading away the added accuracy of a destructive count on a single day.

4.4 Discussion

4.4.1 A Cautionary Example

When solving the inverse problem, we may find that some parameters are difficult to identify simultaneously; despite describing different mechanisms, these parameters might have similar effects on prey dynamics and therefore be indistinguishable when validating the model against experimental data. To illustrate this, we present a contrived example of a potential identifiability issue in Fig 4.7, where we consider model solutions for a single predator treatment (*Bembidion-Pardosa*) using different parameter values. We choose two reasonable values of h_0 based on the results from [62] ($h_0^1 = 2/24, h_0^2 = 3/24$). Since we introduced b_0 to the model, we do not have prior results to consult for its value; we consider $b_0^1 = 2/24$ and $b_0^2 = 9/24$ to capture a broad range of resulting model dynamics.

If $b_0 = b_0^2$, we see a significant difference in the *A. pisum* population for $h_0 = h_0^1$ (solid line, circle marker) and $h_0 = h_0^2$ (dashed line, circle marker) on day 8. If populations were close to 450 under treatment A-BP and close to 0 under treatment AR-BP when $b_0 = b_0^2$, we would conclude that $h_0 = h_0^1$. However, suppose we did not know that $b_0 = b_0^2$ and instead assumed that $b_0 = b_0^1$. Since error in the data is as likely to over- as under-estimate the aphid population, we might erroneously conclude that $h_0 = h_0^2$ (dashed line, no marker) instead of $h_0 = h_0^1$ (solid line, no marker) based on a noisy observation of the true solution (solid line, circle marker).

Given the noise inherent to experimental data (in particular, noise which will increase with the size of the population we sample), it is unlikely that we could distinguish between these two population trajectories. We cannot necessarily anticipate where similar problems may arise prior to parameter estimation, but we can guard against them by sampling sufficiently often in time to outweigh experimental noise or, if such sampling is not feasible, by comparing parameter results across different treatment types.

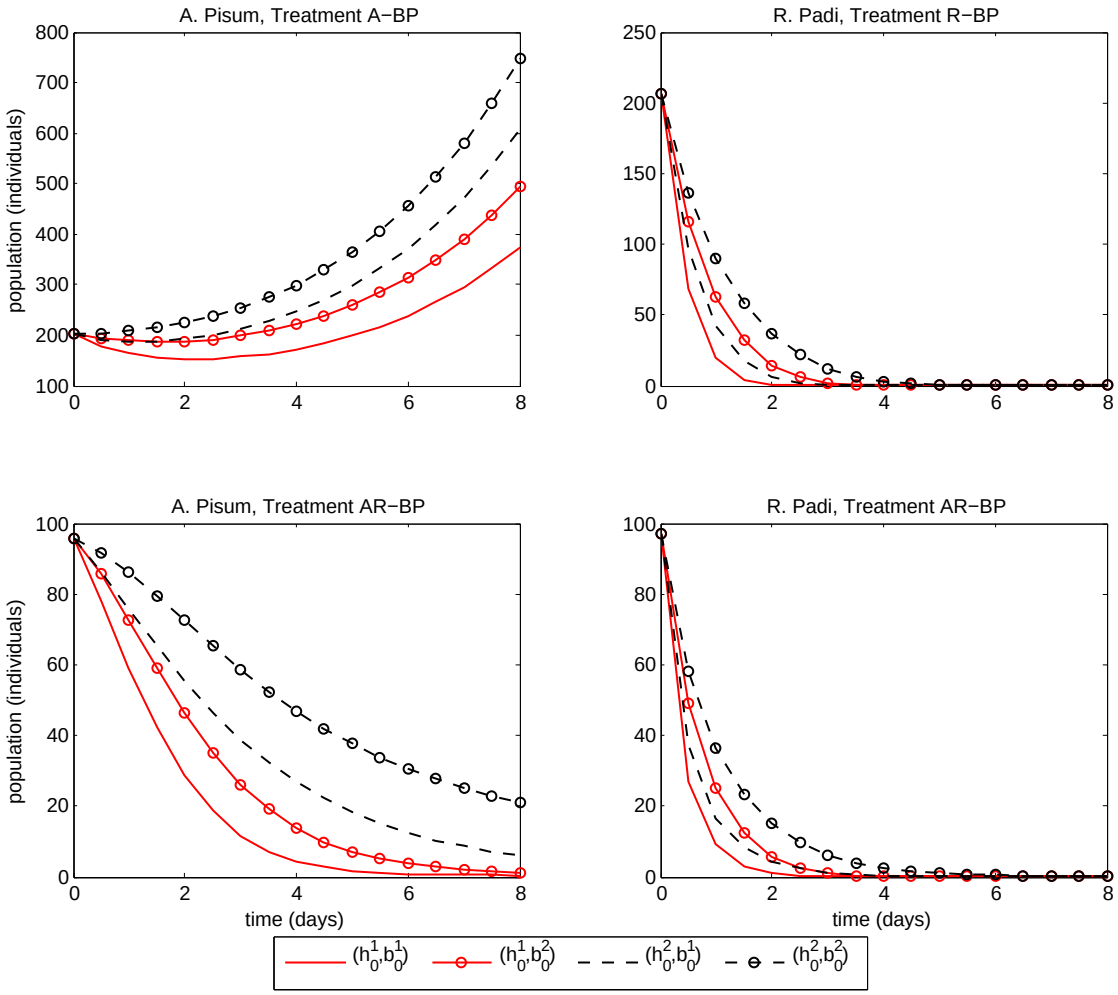

Figure 4.7 Model trajectories for aphid populations under predator treatment Bembidion-Pardosa. Upper left: *A. pisum* aphid treatment. Upper right: *R. padi* aphid treatment. Lower left: *A. pisum* in combined aphid treatment. Lower Right: *R. padi* in combined aphid treatment. We plot model solutions for all combinations of parameters $h_0^1 = 2/24$, $h_0^2 = 3/24$, $b_0^1 = 2/24$, and $b_0^2 = 9/24$.

4.4.2 Conclusion

In order to attain sufficient information for model validation, we sample on days 2, 4, 6, and 8. If we are constrained by resources and cannot sample at this rate, we expect that aphid populations will be most sensitive to model parameters on days 2, 4 and 8. For treatments under which the aphid population is rapidly decimated, it may be necessary to sample more frequently at the start of the experiment. In these cases, we may also sample on days 1 and 3. In our experiment, we cannot guarantee that there will be sufficient time in the day to obtain the necessary data at the experiment's peak, when populations have reached maximal values. It is therefore important that we determine a

viable subsampling protocol before starting the experiment. We conclude that once the population reaches a threshold value, subsampling is a viable option. At low population densities, subsampled populations might poorly estimate the true population size, and so we prioritize fully sampling these populations when there is insufficient time for the full sampling of all populations.

We note that our proposed experimental protocol is intrinsically linked to the assumed ATN model for aphid dynamics and statistical model for error in the data. We cannot make general conclusions about the amount of data required to validate a model with p parameters, or the frequency with which such data must be collected. Instead, we propose that sampling schemes should optimize the expected information content in an experimental data set, conditioned on the anticipated model dynamics and statistical noise. Because we have designed this experiment to explicitly test our model, we substantially increase our confidence that any observed dynamics which we fail to capture are based on mechanisms we have missed or misrepresented in the model, as opposed to simply being caused by insufficient data.

CHAPTER

5

FITTING THE ATN MODEL TO DATA: MESOCOSM EXPERIMENT

In this chapter, we attempt to validate the ATN model using data from controlled experiments conducted in a greenhouse, as outlined in Chapter 4. We begin with a brief description of the experimental methodology and resulting data set. We also describe our estimation of parameters for the modified ATN model as well as a version which ignores habitat use. We conclude with a discussion of our model in light of these results. The results from this chapter appear in [120], which contains a detailed description of experimental methodology and ecological analysis of these results.

5.1 Study System

5.1.1 Experimental Methods

We continue our study of the ATN model applied to a system of terrestrial arthropods by monitoring aphid dynamics in a greenhouse mesocosm experiment. The experiment is conducted inside mesh cages, in which we create miniature foodwebs of plants, aphids, and predators. Aphids are added to the cages when plants, their primary food supply, are sufficiently grown. The aphids are given two days to establish in the cages and begin reproduction before predators are added to the cage. For eight days, we monitor the aphid population in the presence of its predators. After this, plant quality

is sufficiently deteriorated that we must terminate the experiment, and the cages are deconstructed. The mesh cages are kept inside a greenhouse, permitting control of ambient temperature and light.

To explore the importance of body mass in determining aphid dynamics, we use differently sized species of aphid, *Acyrthosiphon pisum* and *Rhopalosiphum padi*. The three possible aphid treatments are (i) *A. pisum*, (ii) *R. padi*, or (iii) *A. pisum* and *R. padi*. We use four predators, selected for variation in body size and habitat use: small foliage-dwelling *Orius*, large foliage-dwelling *Coccinella*, small ground-dwelling *Bembidion*, and large ground-dwelling *Pardosa*. See Figure 5.1 for a depiction of species size differences.

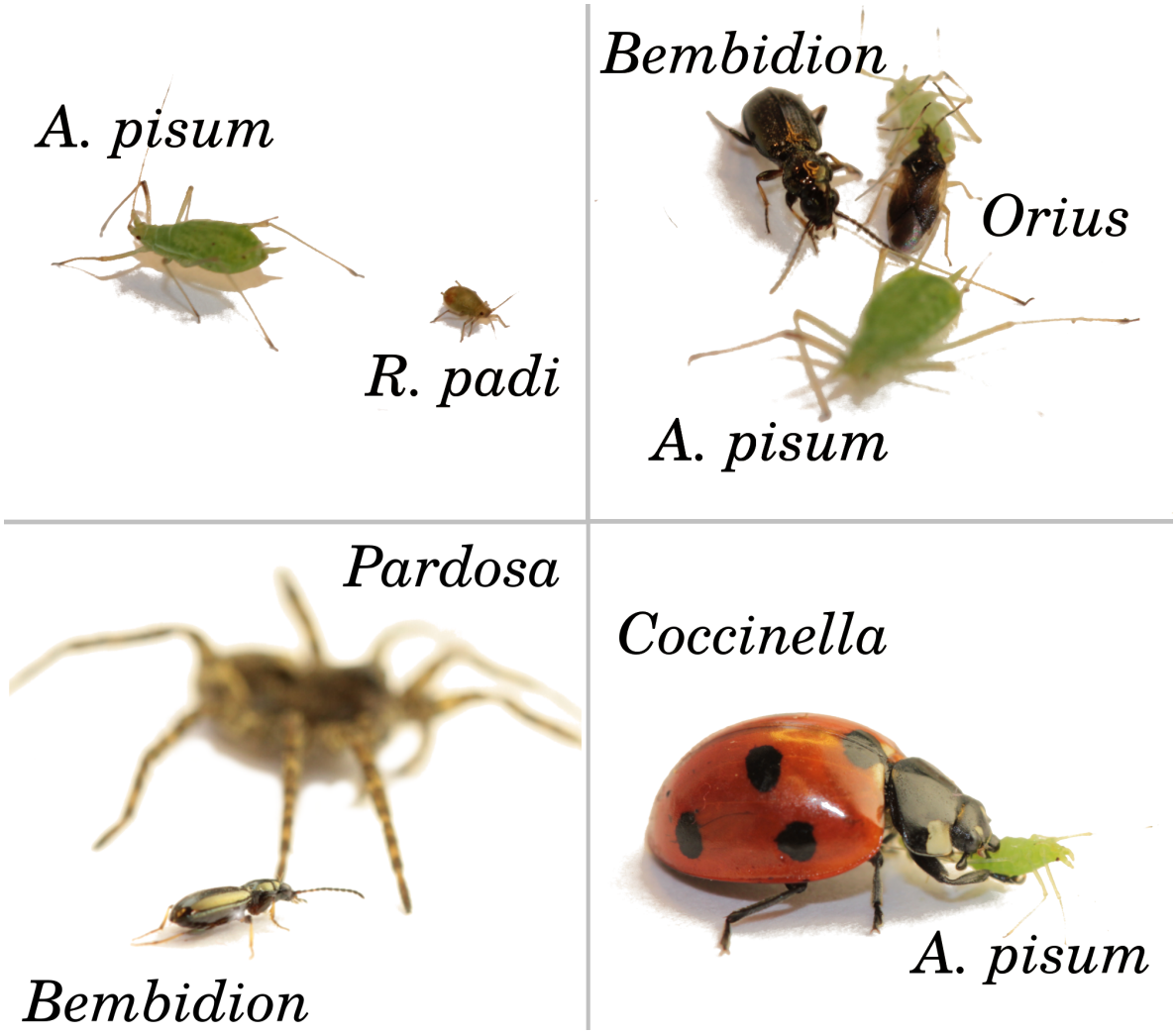

Figure 5.1 Relative sizes of subject organisms in our greenhouse experiments. Photographs courtesy Kate Wootton.

Taking pairwise combinations of all predators, including a predator paired with itself, we have ten possible predator treatments and a control treatment with no predators. We match each predator treatment with each aphid treatment, for a total of 33 treatments, each of which is replicated six times. Due to the number of cages required to do this, our experiment is conducted in batches, and replicates are randomly distributed across batches. That is, the first replicate of two different treatments do not necessarily occur at the same time, and the fifth replicate of either of these treatments could occur before the first.

As motivated in Chapter 4, aphid population counts are taken on days 2, 4, 6, and 8 of the experiment. Additionally, aphids are counted on days 1 and 3 in cages with *Coccinella*, where the population is quickly decimated. Predators are only counted on day 8 of the experiment, after the cages have been deconstructed. We note that in practice, the smaller predators (*Orius* and *Bembidion*) are difficult to recover from the cages, as they can burrow in the dirt. Additionally, we note that *Coccinella* and *Pardosa* regularly escape from cages through small gaps in the mesh walls. Because *Coccinella* is not reasonably a prey item for any other species in the experiment, we replace *Coccinella* whenever it goes missing from a cage, but we do not replace *Pardosa*.

5.1.2 Model Assumptions

We recall that the modified ATN model is given by

$$\begin{aligned}
\frac{dN_i}{dt} &= r_i N_i - \sum_{j\in\mathscr{C}_i} \frac{a_{ij}\nu_{ij}N_iN_j}{1+\sum_{k\in\mathscr{R}_j} a_{kj}\nu_{kj}h_{kj}N_k + \sum_{l\in\mathscr{C}_j} b_0 a_{jl}\nu_{jl}N_l},\\
a_{ij} &= a_0 W_i^{1/4} W_j^{1/4}\left(\frac{W_j/W_i}{R_{opt}} e^{1-\frac{W_j/W_i}{R_{opt,j}}}\right)^{\phi},\\
h_{ij} &= h_0 W_i^{1/4} W_j^{-1/4},\\
\nu_{ij} &= 1-\nu_0 TV(\mu_i,\mu_j) = 1-\nu_0 \sup_{\mathscr{A}\subset\Omega}\left|\mu_i(\mathscr{A})-\mu_j(\mathscr{A})\right|,
\end{aligned}$$

where N_i is the abundance of species i, W_i is the mass of species i, $\mathscr{C}_i$ is the set of predators for species i, and $\mathscr{R}_i$ is the set of prey for species i. For the set Ω of all modes of habitat use, we define the probability measure $\mu_i : \Omega \to \mathbb{R}$ for the likelihood that species i occupies some $\mathscr{A} \subset \Omega$. We must specify W_i, μ_i, $\mathscr{C}_i$, and $\mathscr{R}_i$ from empirical observations and assumptions about the study system. The parameters r_i, a_0, $R_{opt,j}$, ϕ, h_0, and ν_0 must be estimated from experimental data.

We assume that all species in the experiment adhere to the above dynamics, with the exception of *Coccinella*. Due to the aforementioned problem with escaping *Coccinella*, we do not model this population and instead directly input its value from our records in driving all other populations' dynamics. Although we could model the escape and replacement of *Coccinella* with a switching birth and death rate, we reason that this would be largely guesswork, as there is no way to know when

Coccinella escape the cage. This would result in an unnecessarily complicated model, particularly because we do not expect *Coccinella* to be prey to other species in the experiment.

The body masses and initial abundances of all species in our system are given in Table 5.1. We note that the initial abundance of a species is doubled if it appears individually in a treatment; for example, if *Orius* is the only predator in a cage, then we include two populations of *Orius* for a total of 40 predators. We assume that the initial abundance of a predator species is accurate. However, because aphids reproduce quickly and establish themselves in the cages with varying success, we assume that there is some uncertainty in the initial aphid abundance. We therefore choose to estimate the initial condition $N_{i,0}$ for all aphid populations in the experiment, while taking the initial condition $N_{i,0}$ from Table 5.1 for all predators. We consider $t = 0$ to be the day predators are introduced to the cages, and therefore the estimated value of $N_{i,0}$ for aphids is the population after two days in the cage.

Table 5.1 The body masses and initial abundances for all species used in the mesocosm experiments.

	A. pisum	*R. padi*	*Bembidion*	*Coccinella*	*Orius*	*Pardosa*
W_i (mg)	0.6706	0.1550	2.145	37.4636	0.58	17.72
$N_{i,0}$ (individuals)	75	75	20	2	20	10

In determining the sets $\mathscr{C}_i$ and $\mathscr{R}_i$, we begin with the assumption that aphids do not consume any other species and that predators consume all other species. We then remove links between predators and prey where we expect that a feeding interaction would not reasonably occur, due to mechanisms outside the scope of the current model. For example, we assume that *Pardosa* would not consume *Coccinella* due to its hard shell, and because this interaction is not driven by body mass or habitat use, we explicitly remove this link. Following similar reasoning, we also assume that *Coccinella* does not consume *Coccinella* or *Bembidion*. Due to numerical problems associated with our inability to dynamically model *Coccinella*, we additionally remove *Coccinella* as a prey item for *Orius* and *Bembidion*. Although these interactions might reasonably be driven near zero through parameterization in the ATN model, we do not fit *Coccinella* populations to data and therefore have no way to push estimations away from unrealistic assumptions about consumption of *Coccinella* by potential predators. The resulting table of species interactions is given in Table 5.2, where predators are listed along the columns and their potential prey items in the rows. An entry of 0 indicates that the feeding interaction does not occur, and a 1 indicates the interaction can occur, although the strength of the interaction is determined by the ATN model.

To determine μ_i for the species, we partition the cage into six habitat regions: under bean leaf, bottom 2cm of bean plant, rest of bean plant, bottom 2cm of barley plant, rest of barley, and ground.

The bean plant is the primary host for *A. pisum* and the barley plant is the primary host for *R. padi*. We additionally consider the walls of the cage, which predators might climb, to be an extension of the ground. We observe the aphids in each of the control cages on days 2 and 6 of the experiment and count the number of aphids in each of the habitat regions. We observe on days 2, 4, 6, and 8 of the experiment the number of predators in each habitat region for single-predator treated cages. We use these counts to obtain the probability that each species occupies a given habitat region and compute the total variation distances between their habitat use measures accordingly. In general, we assume that there is no overlap between different habitat regions in the cages. However, we do assume that there is some overlap between the floor of the cage and the base of a plant, and that the magnitude of this overlap is dependent on the size of the predator on the ground. For example, a predator on the ground with body length 1 centimeter would overlap with 50% of the 2 centimeter base of a plant. The resulting values of $TV(\mu_i, \mu_j)$ are given in Table 5.3.

Table 5.2 Assumed interaction matrix for species in the cage experiments. Potential feeding interactions are indicated with a 1, with predators in the columns and their prey items down the rows.

Prey \ Predator	*A. pisum*	*R. padi*	*Bembidion*	*Coccinella*	*Orius*	*Pardosa*
A. pisum	0	0	1	1	1	1
R. padi	0	0	1	1	1	1
Bembidion	0	0	1	0	1	1
Coccinella	0	0	0	0	0	0
Orius	0	0	1	1	1	1
Pardosa	0	0	1	0	1	1

Table 5.3 Total variation $TV(\mu_i, \mu_j)$ for species in the cage experiments.

	A. pisum	*R. padi*	*Bembidion*	*Coccinella*	*Orius*	*Pardosa*
A. pisum	0	1	0.9945	0.8700	0.7760	0.9670
R. padi	1	0	0.8870	0.6850	0.7775	0.7670
Bembidion	0.9945	0.8870	0	0.2350	0.6050	0.0450
Coccinella	0.8700	0.6850	0.2350	0	0.4140	0.1970
Orius	0.7760	0.7775	0.6050	0.4140	0	0.5600
Pardosa	0.9670	0.7670	0.0450	0.1970	0.5600	0

5.2 Estimating ATN Model Parameters

We implement an iterative reweighted least squares parameter estimation, as described in Chapter 1, to obtain estimates of r_i, a_0, $R_{opt,j}$, ϕ, h_0, and υ_0 along with the initial conditions $N_{i,0}$ for all aphid populations. To preserve variability between replicates and treatments, we estimate different initial aphid abundances for all cages. Solution of the inverse problem requires more data than can possibly be collected in a single cage, for which there are often more parameters to estimate than days on which data is collected. We therefore solve the problem simultaneously over several different treatment types, using all replicates within these treatments.

We define θ to be the vector of parameters and initial conditions we will estimate. Recall that the iterative reweighted least squares inverse problem corresponds to a relative error statistical model in which we assume that the population data $y_{i,c,d}$ for species i observed at time t_d in replicate c is a realization of the random variable $Y_{i,c,d}$ defined by

$$Y_{i,c,d} = N_i(t_d, \theta_0) + N_i^{\gamma_i}(t_d, \theta_0)\mathscr{E}_{c,d}.$$

$N_i(t_d, \theta_0)$ is the solution to the ATN model for the "true" parameter set θ_0 and $\mathscr{E}_{c,d}$ is normally distributed noise associated with measurement error. We let $\mathscr{A}$ and $\mathscr{B}$ be the set of all aphid and predator populations we use to estimate model parameters and $\mathscr{D}_i$ the days over which species i is sampled. Following the iterative reweighted least squares formulation of this problem, we define the cost functional for realizations $\{y_{i,c,d}\}_{i \in \mathscr{A} \cup \mathscr{B}}$ of the collected data,

$$J(\theta, \vec{y}) = \sum_{c=1}^{6} \left[\sum_{i \in \mathscr{A}} \sum_{d \in \mathscr{D}} N_i^{-2\gamma_i}(t_d, \theta)\big(y_{i,c,d} - N_i(t_d, \theta)\big)^2 + \sum_{j \in \mathscr{B}} N_j^{-2\gamma_j}(8, \theta)\big(y_{j,c,8} - N_j(8, \theta)\big)^2 \right].$$

In the following subsections, we use this cost function to estimate model parameters using three different subsets of cages from the experiment.

5.2.1 Control Cages - Estimating r_i and γ_i

In control treatment cages, there is no predator influence on the aphid population and the ATN model reduces to exponential growth. We use the data from these cages to estimate the growth rates r_i of the two aphid species, which are then fixed in the ATN model for predator-treated cages. We assume that the aphid species have different growth rates, but that across all three aphid treatments (individual or combined) and the replicates of these treatments, the growth rate should be the same within each species. We therefore solve two inverse problems, one for each aphid species, in which the above cost functional reduces to the scalar case of iterative reweighted least squares. We estimate that $r_A = 0.29136$ and $r_R = 0.34231$.

We additionally use the model fit in these cages to determine a preliminary value of γ_i for aphid populations. We assume that the underlying statistical model for observational noise in aphid abundance data is independent of the predator treatment in each cage. We therefore conduct residual testing for the exponential fit in control cages to inform our choice of γ_i when solving the inverse problem in predator-treated cages. We again consider the two aphid species separately, using data from all three aphid treatments and the replicates of these treatments.

In Figure 5.2, we plot the residuals from our iterative reweighted least squares estimation in control *R. padi* cages for different values of γ_i. We note that because the population grows exponentially, the size of the population increases in time. We can see that for $\gamma_R = 0.1$, residuals have a wide spread early in the experiment and a narrow spread at the end of the experiment. At this value of γ_R, we conclude that the residuals are not identically distributed in time, because the solution to the inverse problem minimizes relative error at the end of the experiment, when the aphid population is at its peak. Conversely, for $\gamma_R = 0.6$, residuals have a narrow spread early in the experiment and wide spread later in the experiment. We conclude that $\gamma_R = 0.4$ gives the most-balanced spread of residuals over the course of the experiment. We therefore take $\gamma_R = 0.4$ as a preliminary choice for observations of *R. padi* in the remaining predator-treated cages.

In Figure 5.3, we plot the residuals from our iterative reweighted least squares estimation in *A. pisum* cages for different values of γ_A. Using similar criteria as above, we conclude that $\gamma_A = 0.5$ is the most appropriate choice for the statistical model, although there is less distinction between these residuals and those obtained for $\gamma_A = 0.4$. There is insufficient data to conduct residual tests for predator populations, since samples are obtained only on one day during the experiment. However, we reason that because smaller predators are more difficult to recover, and smaller predators occur in higher abundance in the experiment, there is a relationship between the observational error and population size. Having no other information, we take $\gamma_i = 0.5$ for predators in the experiment; this choice is largely arbitrary, but *A. pisum* is closer in size to the predators.

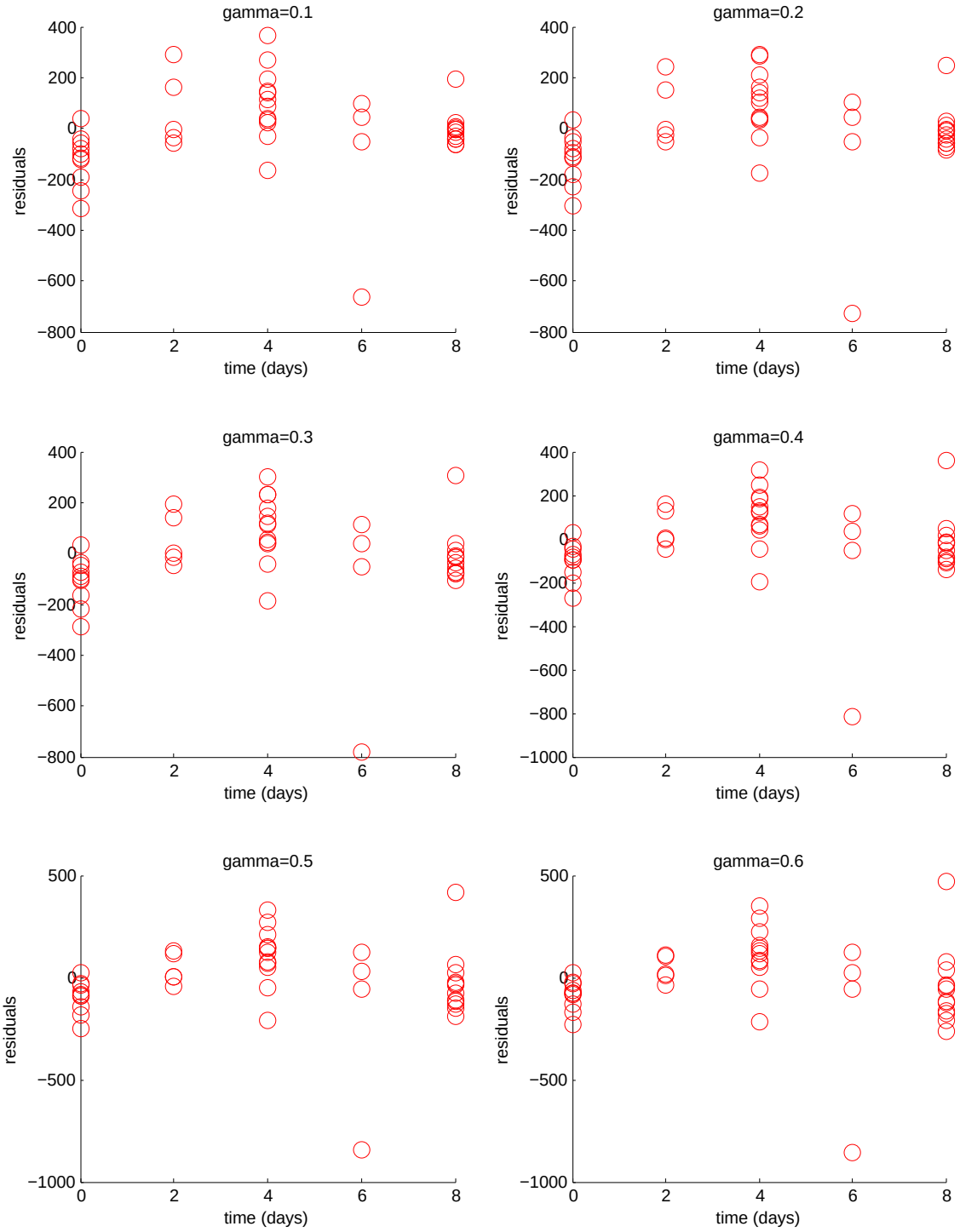

Figure 5.2 Residuals from an exponential fit to all control data for *R. padi* plotted in time, for different values of γ_R.

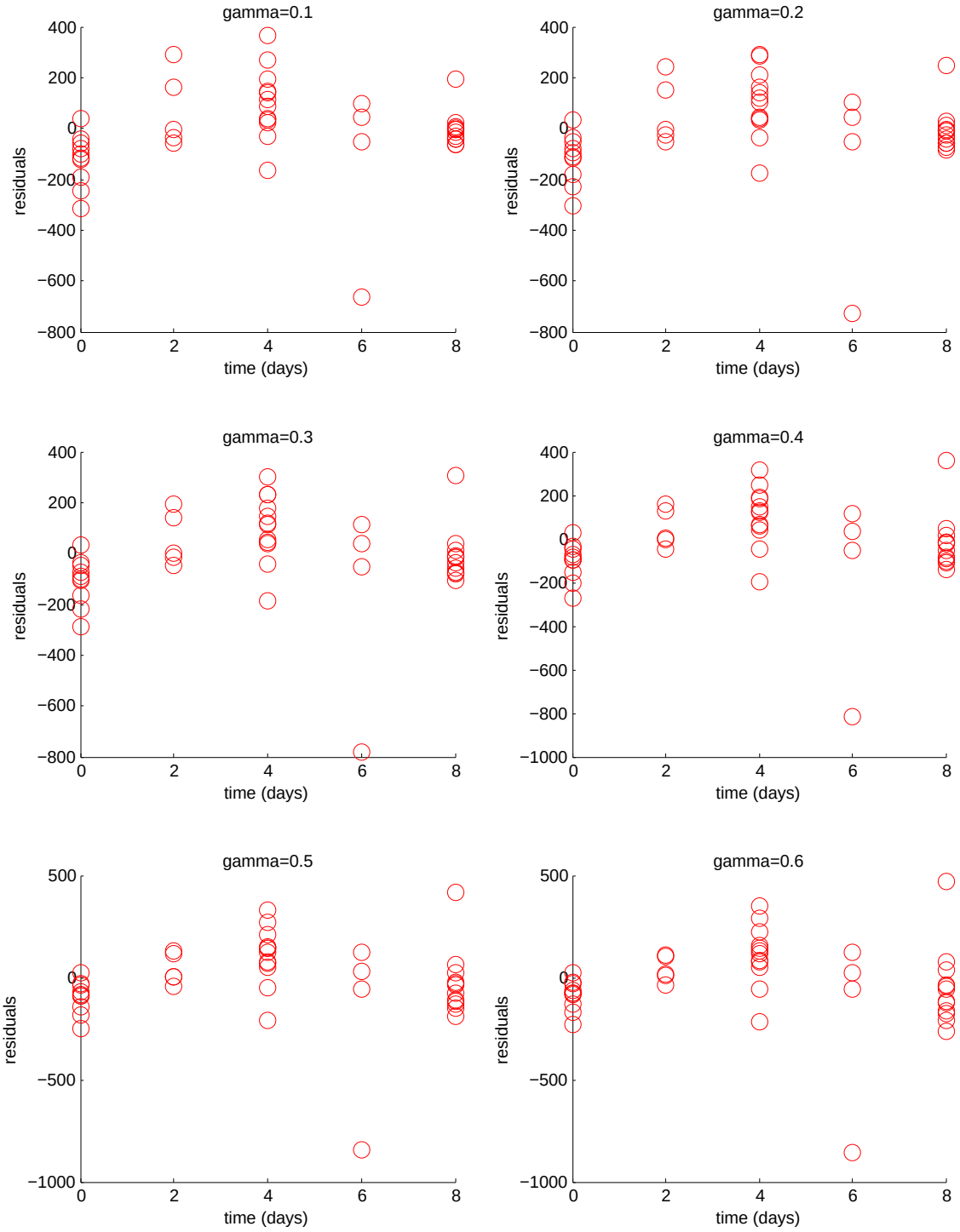

Figure 5.3 Residuals from an exponential fit to all control data for *A. pisum* plotted in time, for different values of γ_A.

5.2.2 Individual Predator Treatments - Estimating $R_{opt,j}$

We use data from the four single-predator treatments to estimate the value of $R_{opt,j}$ for each predator species j. With only one predator, we reason that predator interference (and to a lesser degree, habitat use) is less important in determining species interactions in these treatments. We therefore use the information from these cages to establish a baseline for the predator-specific parameter $R_{opt,j}$. We assume that the optimal predator-prey body mass ratio will vary across predators, but that this parameter should be constant in the presence of different prey items. We use data from all three aphid treatments within each single-predator treatment to obtain an estimate of this parameter. For each predator treatment, this corresponds to data from 24 aphid populations, each sampled on at least four days, and a single count of 18 predator populations.

In Table 5.4, we list the estimated model parameters for each of the single-predator treatments. We also give the standard deviation of each parameter in parentheses, using the asymptotic approximation from Chapter 1.

Table 5.4 Estimated model parameters using data from single-predator treatments. Standard errors are given in parentheses.

	Bembidion	*Coccinella*	*Orius*	*Pardosa*
a_0	0.03945 (0.00318)	0.2146 (.0181)	0.001 (1.24)	0.00311 (0.0168)
$R_{opt,j}$	52 (26)	300 (2×10^6)	15 (768)	14 (15)
ϕ	1.031 (.0161)	0.900 (.0448)	1.10 (7.28)	0.2227 (3.99)
h_0	0.01000 (.0476)	0.08 (.00051)	0.09 (110)	0.09 (0.464)
v_0	0.9213 (.0161)	1.00 (1×10^3)	1.00 (174)	1.00 (0.061)
b_0	0.090 (2.89)	–	0.09 (1×10^5)	0.001 (23.6)

The results from *Bembidion* treatment are the most straightforward, and model fits to data are fair for cages in this treatment (not shown for brevity). The standard deviation for b_0 and $R_{opt,B}$ is high, meaning that we are less certain in the values of these parameters. However, the model is not so insensitive to $R_{opt,B}$ that the estimate is arbitrary, and so we fix this value when solving the inverse problem in the combined-predator treatments. The high standard deviation for b_0 can be explained by model insensitivity to this parameter, which is caused by the single-predator treatment. Because *Bembidion* is assumed to be a cannibal, other predators in the cage are both alternative prey items and potential predators, and the attack rate is the same for both interactions. The contributions

from $k, l = B$ in the functional response for *Bembidion*, given by

$$F_B = 1 + \sum_{k \in \mathscr{R}_B} a_{kB} v_{kj} h_{kB} N_k + \sum_{l \in \mathscr{C}_B} b_0 a_{Bl} v_{Bl} N_l,$$

therefore reduce to a common term $a_{BB} v_{BB}(h_{BB} + b_0)N_B$. The model is theoretically sensitive to b_0 because h_0 independently occurs in the term $a_{AB} v_{AB} h_{AB} N_A + a_{RB} v_{RB} h_{RB} N_R$ for alternative aphid prey items. However, in practice, we find that contribution to the functional response due to b_0 is insignificant when compared to h_0, particularly at the low estimated value.

The high standard deviations in our estimates from the *Coccinella* treatment are largely due to the fact that *Coccinella* is not its own predator, and so aphid dynamics are insensitive to many model parameters. We cannot meaningfully estimate b_0 in this treatment, since there are no predators present to cause interference. Additionally, there is a tradeoff between parameters a_0, R_{opt}, ϕ, and v_0 in determining the magnitude of attack rates. Intuitively, we are attempting to specify the magnitude of two effective attack rates (a_{AC} and a_{RC}) using four parameters, and the resulting insensitivity of some of these parameters is unsurprising. In particular, we note that $TV(\mu_C, \mu_A) = 0.870$ and $TV(\mu_C, \mu_R) = 0.685$, and so $v_0 = 1$ results in a preference for attacks on *R. padi*. Similarly, $W_C/W_A = 55.9$ and $W_C/W_R = 241.7$, which means that $R_{opt,C} = 300$ yields the same preference. The magnitude of ϕ tunes the strength of the preference, determining the ratio of attacks made on either species of aphid, and the value of a_0 scales the frequency of attacks as necessary.

The estimates from *Orius* and *Pardosa* treatments have the highest uncertainty in estimated parameters, without a clear cause. In general, these predators have a smaller impact on aphid populations, compared to *Coccinella* and (to a lesser degree) *Bembidion*. The estimated value of a_0 results in a minimal attack rate on aphid populations, and so aphid dynamics are not very sensitive to model parameters. In some cages, this makes sense; the presence of *Orius* or *Pardosa* can permit growth comparable to what we see in some control cages. However, there is significant variation in the aphid population's growth across replicates; for example, see Figure 5.4, where we plot the population data from the six replicates of only *R. padi* in an *Orius* treatment. The final aphid populations in replicates 2, 4, and 6 are close to those seen in most control *R. padi* populations in a combined treatment and one control *R. padi* population in an individual treatment. However, the aphid populations in replicates 1, 3, and 5 are clearly impacted by the *Orius* treatment. We additionally note that there is no clear correlation between the final population of *Orius* and *R. padi* in these treatments. The solution to the inverse problem fits the high growth cages under *Orius* and *Pardosa* treatments well, but the estimated initial condition must be very low to fit populations similar to those seen in replicates 1, 3, and 5.

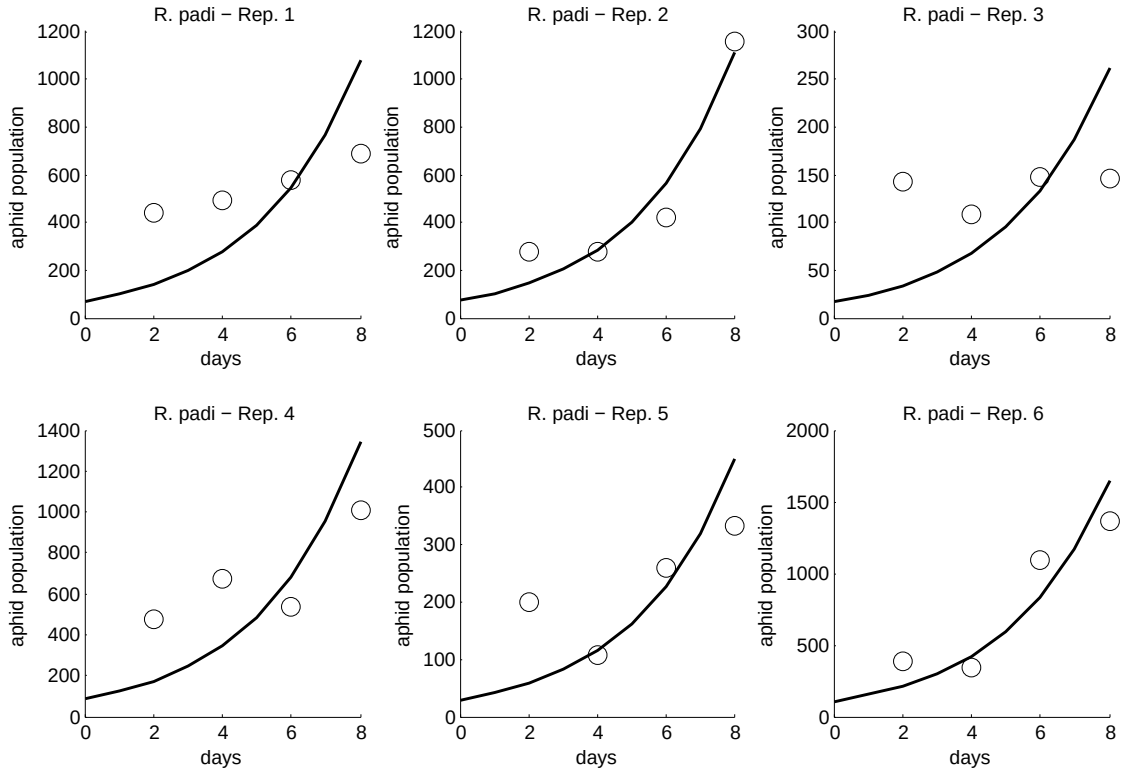

Figure 5.4 Population of *R. padi* for replicates of *Orius*-treated cages and *R. padi* the only aphid species. Data is plotted with a circle marker (∘) and model fit to data with a solid line.

We additionally use these same data sets to estimate model parameters assuming that $v_0, b_0 = 0$. We list in Table 5.5 the estimated values of a_0, $R_{opt,j}$, ϕ, and h_0 in this case, as well as the standard deviation of each parameter in parentheses. In general, the standard deviation is lower for estimated parameters; this is due to an increased sensitivity to model parameters, with the redundancies discussed above removed from the inverse problem. The exception to this pattern is the solution to the inverse problem for *Pardosa* treated cages, which we discuss below.

Table 5.5 Estimated model parameters using data from single-predator treatments, assuming $v_0, b_0 = 0$. Standard deviations are given in parentheses.

	Bembidion	*Coccinella*	*Orius*	*Pardosa*
a_0	0.00165 (0.00006)	0.2146 (0.0181)	0.001 (.0112)	0.00287 (2×10^7)
$R_{opt,j}$	27.4 (2)	269 (6)	30 (433)	1.007 (2×10^9)
ϕ	1.20 (0.0255)	0.8639(0.0245)	1.00 (1.25)	0.838 (7×10^9)
h_0	0.089 (.0063)	0.05512 (0.00216)	0.09 (.003)	0.0739 (5×10^9)

In Figure 5.5, we plot the model fit to data for the first replicate of all three aphid treatments under the individual *Pardosa* treatment. The data for *R. padi* is plotted with a circle marker (∘), the data for *A. pisum* is plotted with a square marker (□), and the data for *Pardosa* is plotted with a triangle marker (△). The model fit for *R. padi* is plotted with a solid line, and the fit for *A. pisum* is plotted with a dashed line. We plot the solutions with v_0, b_0 as well as the solutions when $v_0, b_0 = 0$, indicating the latter with dotted lines; for the aphid populations, these fits are indistinguishable from one another. We plot the model fit with v_0, b_0 to the predator population using a solid line and the fit for $v_0, b_0 = 0$ using a dotted line; these fits differ by only 1 or 2 individuals over the course of the experiment. We observe the same behavior for other replicates (not plotted for brevity). Despite the similarity between model fits, as well as comparable values of a_0, the standard deviation of estimated parameters when $v_0, b_0 = 0$ is unreasonably high. The cause of this seems to be a lack of sensitivity in some cages; at the value of $R_{opt,P}$ estimated when $v_0, b_0 = 0$, some model solutions have no sensitivity to ϕ.

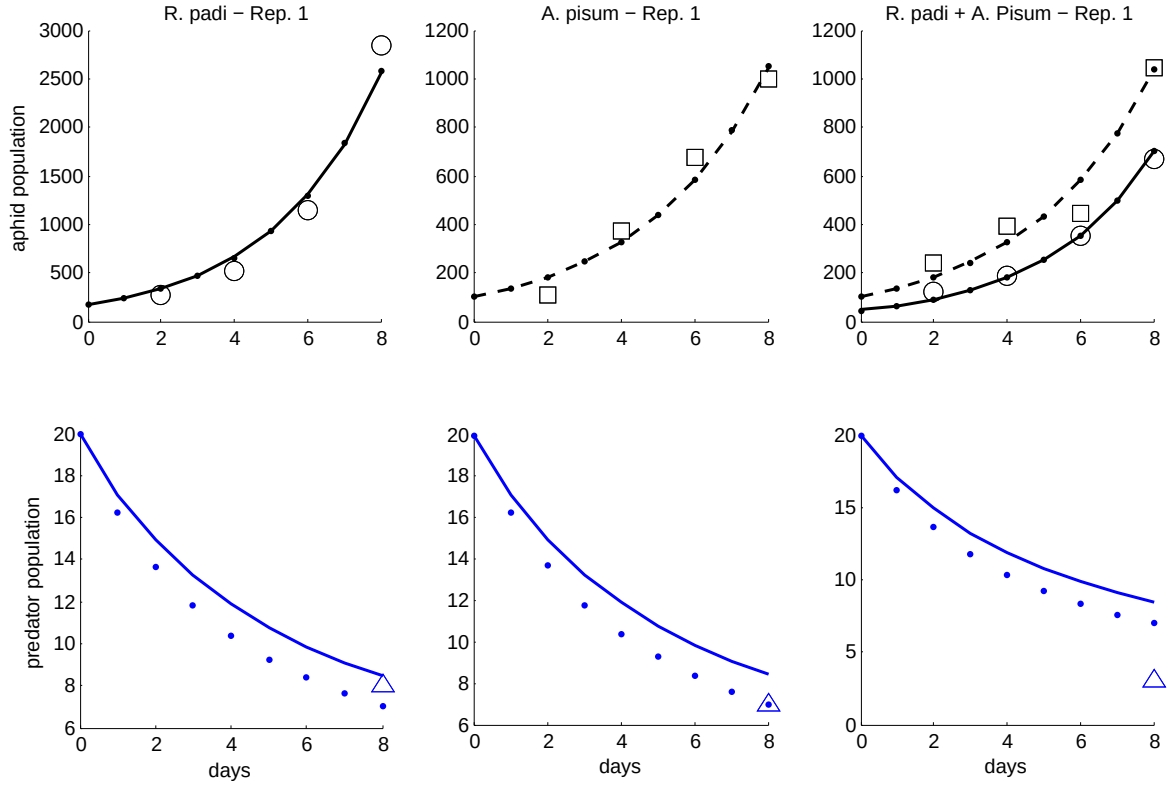

Figure 5.5 Model fit to data for a replicate of all aphid treatments under *Pardosa* single-predator treatment. Model fit to data is plotted with lines, and dots indicate the fit when $v_0, b_0 = 0$. We use markers to indicate the abundance data for *R. padi* (∘), *A. pisum* (□), and *Pardosa*(△).

5.2.3 Combined Predator Treatments - Estimating Scaling Constants

We finally use the remaining data, consisting of the three aphid treatments in the presence of distinct two-predator combinations, to estimate the remaining ATN model parameters, a_0, ϕ, h_0, v_0, and b_0. This corresponds to data from 18 treatment types, each with six replicates, for a total of 144 aphid populations and 216 predator populations. We fix values of $R_{opt,j}$ as estimated in the previous section and consider the case where v_0, b_0 are freely estimated as well as when we enforce the constraint $v_0, b_0 = 0$. In Table 5.6, we list the estimated parameters for both cases, with standard deviations in parentheses. In the figures following the table, we show representative plots of the full ATN model with habitat use for each combined predator treatment.

We first note that standard deviations are significantly lower than in the single predator treatments. Although model fit to data is sometimes poor in this larger data set, we find that the solution of the inverse problem is more sensitive to all parameters under many treatments, resulting in lower estimated values of the standard deviation. However, we note that standard deviations are still lower when we fix $v_0, b_0 = 0$ than when we estimate these parameters. In the case of a_0, this may be an unfair comparison; the estimated value of a_0 when $v_0, b_0 = 0$ is significantly lower than when we estimate v_0, b_0. The inclusion of habitat use in our model reduces attack rates on aphid populations by their predators, and so a higher value of a_0 is required to achieve the same level of predation on the aphid population.

Table 5.6 Estimates of ATN model parameters and standard deviations in combined-predator treatments.

	a_0	ϕ	h_0	v_0	b_0
v_0, b_0 free	0.11 (0.0005)	0.4 (0.019)	0.09 (0.0013)	0.9457 (0.0004)	0.09 (0.0120)
$v_0, b_0 = 0$	0.00841 (0.00002)	0.5 (0.0083)	0.099 (0.0007)	–	–

We plot the full model fit to data in *Coccinella-Bembidion* treatments for the second replicate of all three aphid treatments in Figure 5.6. The model fit to the individual *R. padi* treatment (top-left plot) is typical for decreasing aphid populations. The ATN model fit to all data does not capture the devastating effect of *Coccinella* on *R. padi*. The model can match *R. padi* abundance data from some *Coccinella*-treated cages, but this is only possible when the *R. padi* population is growing, albeit at low rates. The parameters which determine appropriate attack rates for other predators in the experiment are not sufficiently high to match the force of predation in *Coccinella* cages. Additionally, model solutions for the *Bembidion* population are usually too low at the end of the experiment, as seen in the three bottom plots as well as Figures 5.10 and 5.7. We note that the model solution when $v_0, b_0 = 0$ generally yields higher values of *Bembidion*, with aphid populations largely unchanged.

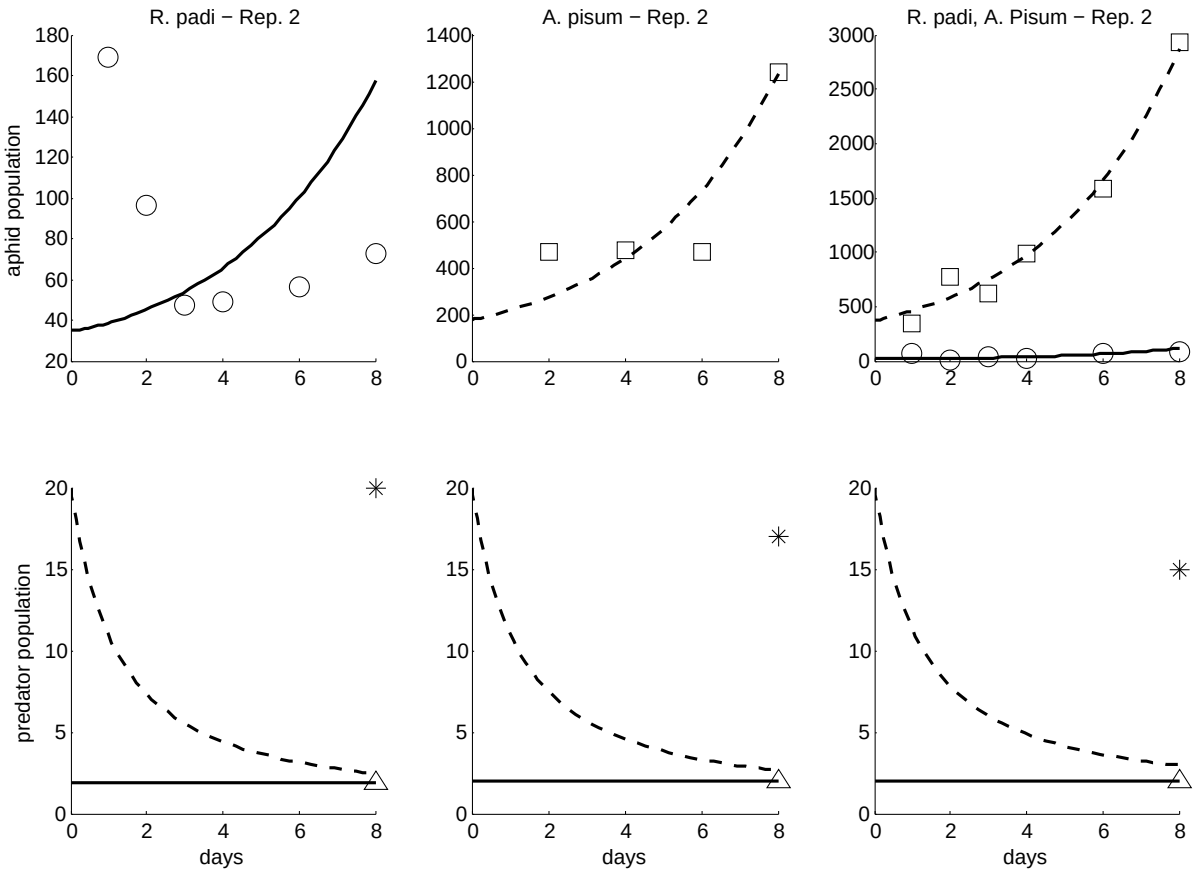

Figure 5.6 ATN model fit to data for a replicate of all aphid treatments under *Coccinella-Bembidion* combined-predator treatment. Model solutions for populations of *Coccinella* and *R. padi* are indicated with solid lines and *Bembidion* and *A. pisum* are indicated with dashed lines. We use markers to indicate the abundance data for *R. padi* ($\circ$), *A. pisum* ($\square$), *Coccinella* ($\triangle$), and *Bembidion* ($*$).

In Figures 5.7 and 5.8, we plot the full model fit to data for the sixth and fourth replicates of the three aphid treatments under *Pardosa-Bembidion* and *Pardosa-Coccinella* combinations, respectively. We again note that the fit to the individual *R. padi* population is significantly worse in the treatment with *Coccinella*, but that fits to increasing aphid populations are closer to the observed abundances. When $v_0, b_0 = 0$, the solutions to the inverse problem are similar, except for the aforementioned difference in *Bembidion* populations.

We compare the model solution for *Bembidion* and *Pardosa* treatments at the end of the experiment (Figure 5.7). The estimated model solutions for the *Pardosa* populations are much closer to the experimental values than we generally see for *Bembidion*. However, in the cage with *Coccinella* (Figure 5.8), model solutions for the *Pardosa* population are much lower than the observed values. This is because *Coccinella* is not a prey item for *Pardosa*; our model predicts low attack rates on

aphid populations by *Pardosa* and in the absence of alternative prey items to increase the functional response, the force of predation on *Pardosa* is higher. We can see that approximations to *Pardosa* populations are improved in Figure 5.11, where *Orius* is available as an alternative prey item. We note that the *Pardosa* population is not always underestimated by the model solution, although the example we selected does exhibit this behavior. Additionally, for the restricted model (v_0, $b_0 = 0$), we see some change in the modelled *Pardosa* populations, but the resulting fit to data is not noticably different.

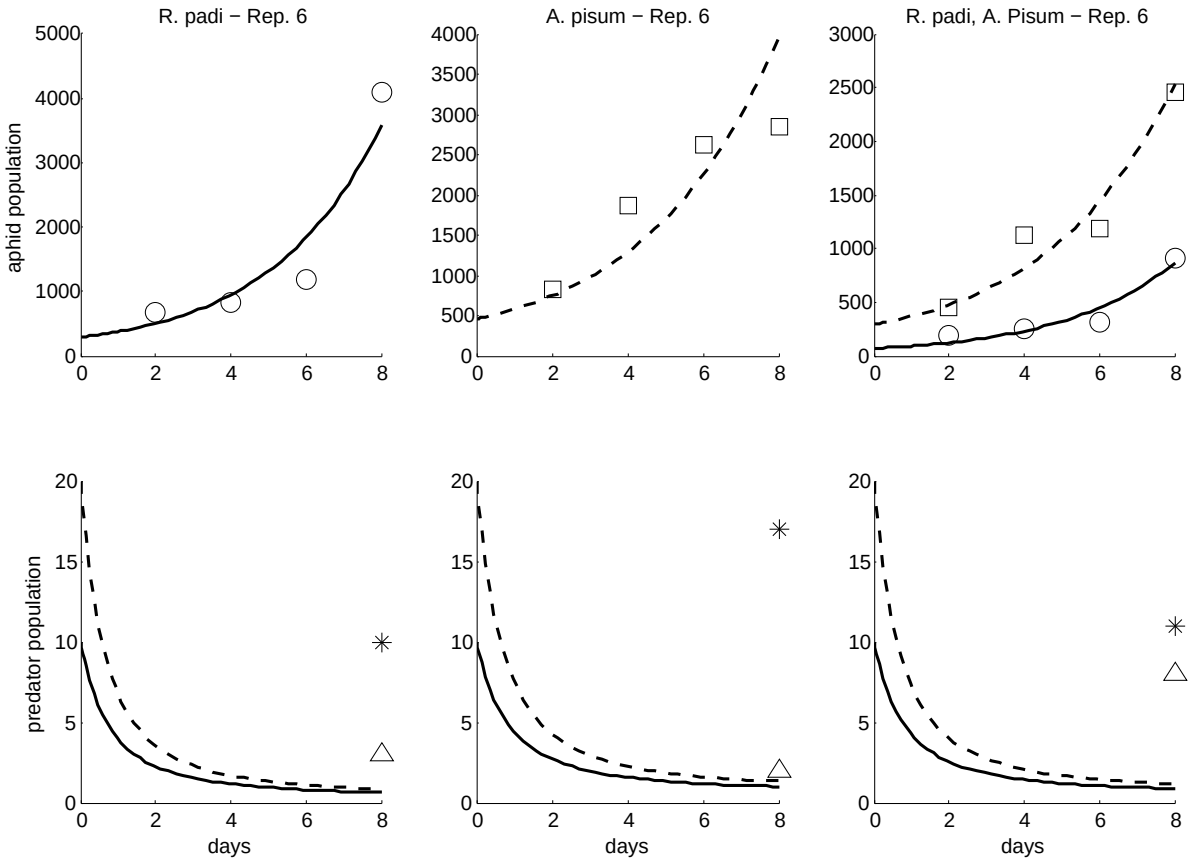

Figure 5.7 ATN model fit to data for a replicate of all aphid treatments under *Pardosa-Bembidion* combined-predator treatment. Model solutions for populations of *Pardosa* and *R. padi* are indicated with solid lines and *Bembidion* and *A. pisum* are indicated with dashed lines. We use markers to indicate the abundance data for *R. padi* (∘), *A. pisum* (□), *Pardosa* (△), and *Bembidion* (∗).

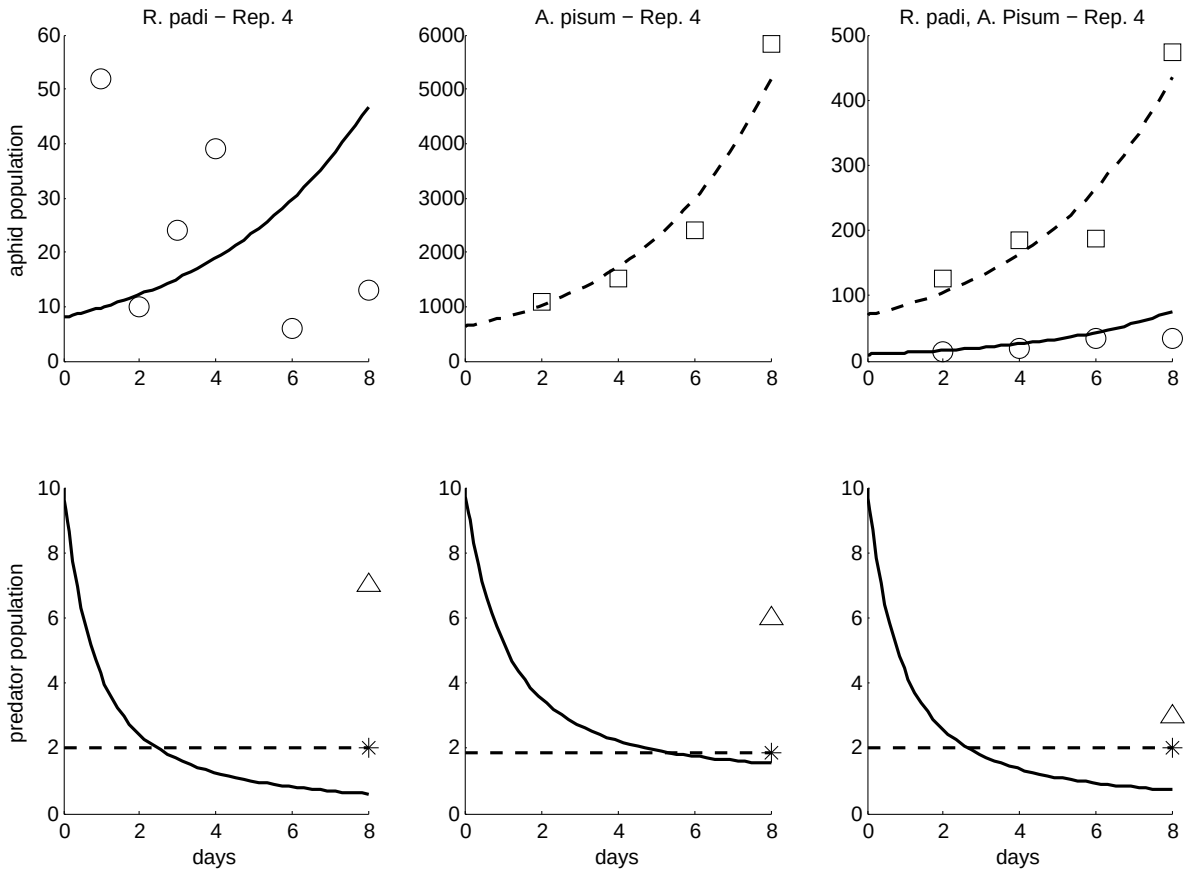

Figure 5.8 ATN model fit to data for a replicate of all aphid treatments under *Pardosa*-*Coccinella* combined-predator treatment. Model solutions for populations of *Pardosa* and *R. padi* are indicated with solid lines and *Coccinella* and *A. pisum* are indicated with dashed lines. We use markers to indicate the abundance data for *R. padi* ($\circ$), *A. pisum* ($\square$), *Pardosa* ($\triangle$), and *Coccinella* ($*$).

We next plot sample solutions of the full ATN model for three aphid treatments when subjected to *Orius*-*Coccinella* and *Orius*-*Bembidion* treatments, in Figures 5.9 and 5.10 respectively. In both of these examples, the individual *R. padi* population is slowly increasing, or close to constant with significant noise in the data. As mentioned above, the model approximation matches the general trend of the data under these conditions. However, the estimated initial condition for *R. padi* is very low in these cages (the same type of solution can be seen above in Figure 5.8). Although the visual fit to data in these cages is improved, the estimated initial condition may not be realistic. This initial condition is much lower than what we estimate in cages with larger aphid populations. Instead of beginning at a very low value, the population may have been higher at $t = 0$, before rapidly decreasing after the introduction of predators and increasing later in the experiment. When estimating model parameters on smaller subsets of the data, we can recreate dynamics that match

this expected behavior. However, our goal is to demonstrate general applicability of the model to the bulk of our experimental data. We therefore trade potential accuracy in some individual descriptions of population dynamics for a single estimate of ATN model parameters which provides an adequate description of most populations' dynamics.

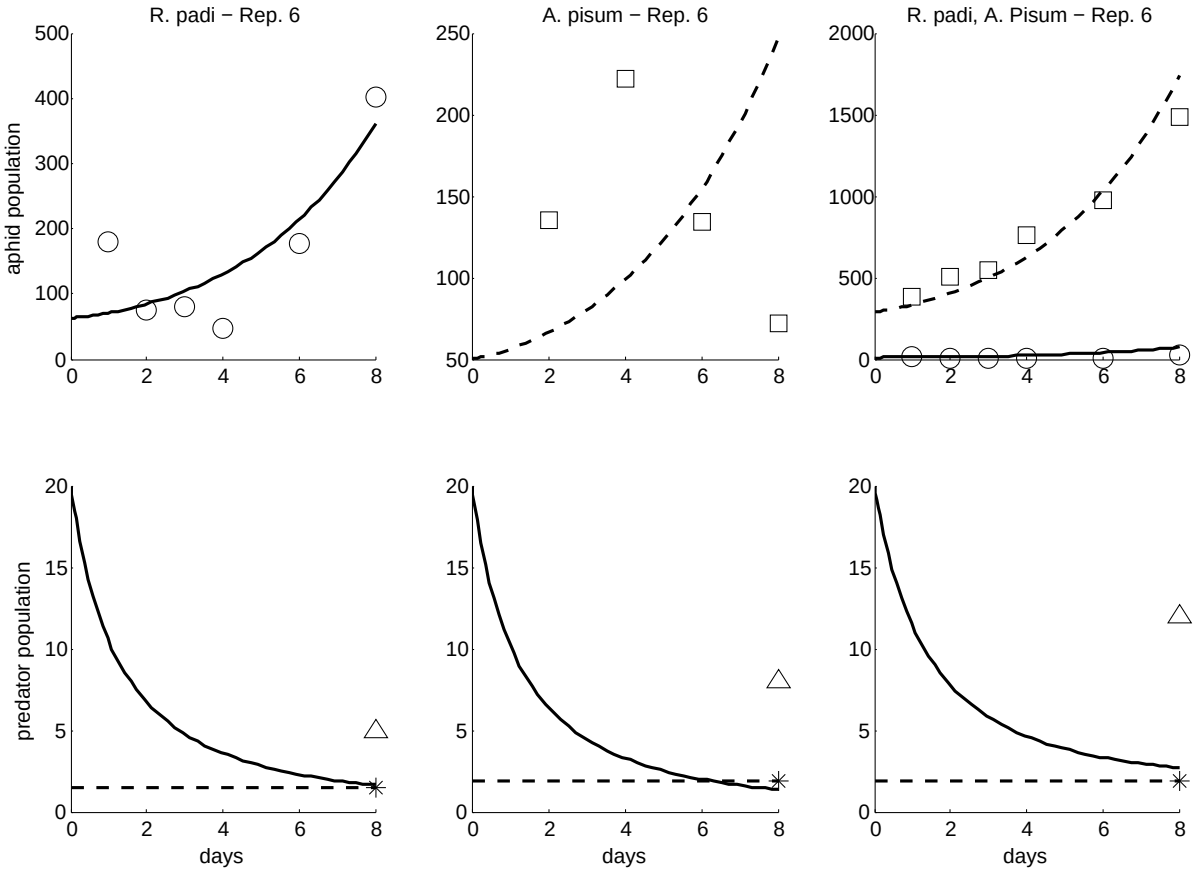

Figure 5.9 ATN model fit to data for a replicate of all aphid treatments under *Orius*-*Coccinella* combined-predator treatment. Model solutions for populations of *Orius* and *R. padi* are indicated with solid lines and *Coccinella* and *A. pisum* are indicated with dashed lines. We use markers to indicate the abundance data for *R. padi* (∘), *A. pisum* (□), *Orius* (△), and *Coccinella* (∗).

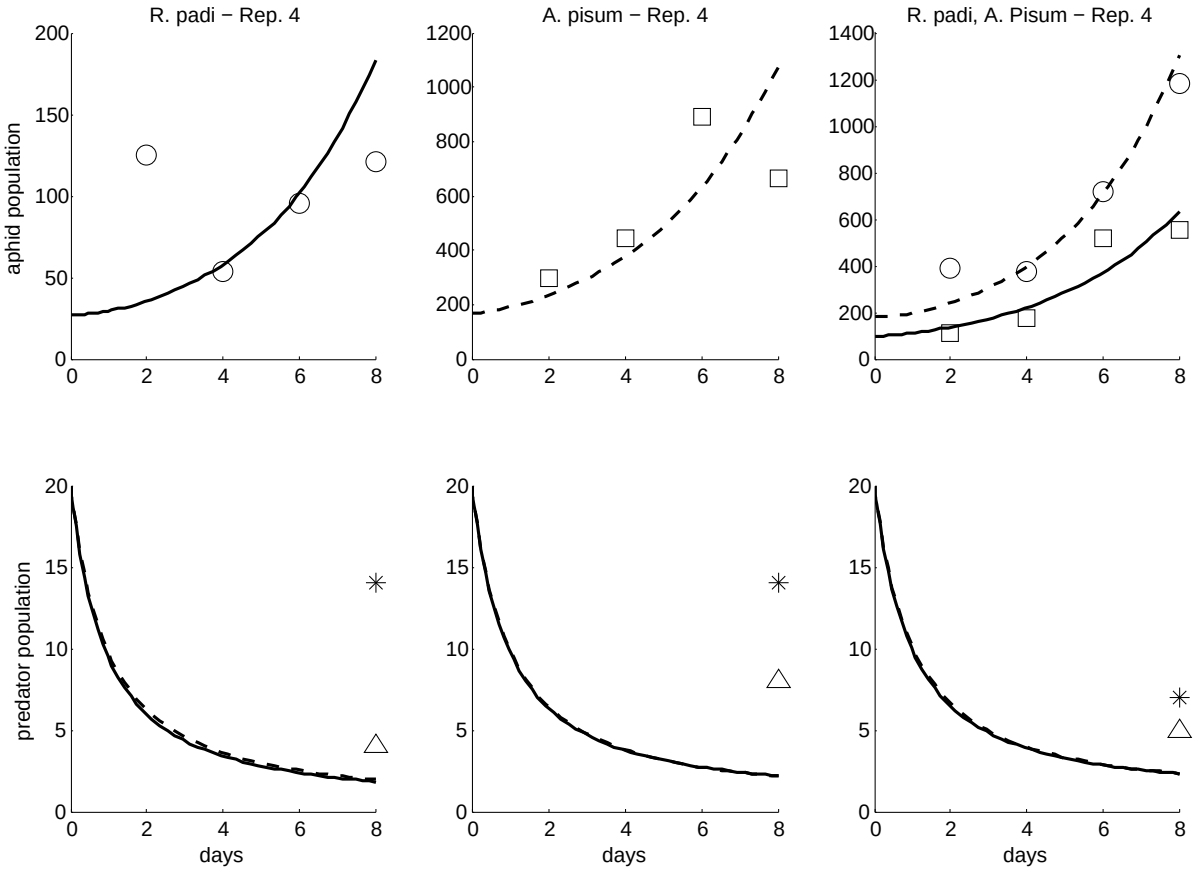

Figure 5.10 ATN model fit to data for a replicate of all aphid treatments under *Orius-Bembidion* combined-predator treatment. Model solutions for populations of *Orius* and *R. padi* are indicated with solid lines and *Bembidion* and *A. pisum* are indicated with dashed lines. We use markers to indicate the abundance data for *R. padi* ($\circ$), *A. pisum* ($\square$), *Orius* ($\triangle$), and *Bembidion* ($*$).

The last combined predator treatment is *Pardosa-Orius*, for which we plot model solutions of the full ATN model for a single replicate of the three aphid treatments in Figure 5.11. As in the individual-predator treatments, these two predators often have a minimal effect on aphid populations. However, we note that the model solutions for *Orius* abundance at the end of the experiment are close to observed values. We see that, to a lesser extent, this is also true in Figures 5.9 and 5.10. However, when we restrict $v_0, b_0 = 0$, the fit to *Orius* populations is generally worse. This is particularly true in cages with *Pardosa*; overall reduced attack rates result in low predation on *Orius*, which is close in size to the aphid species when compared to *Pardosa*. Due to difficulty in recovering the predators, there is significant uncertainty in final *Orius* abundances, and if anything, the observed data is an underestimate of the true population. We therefore hesitate to draw any conclusions about the significance of this difference.

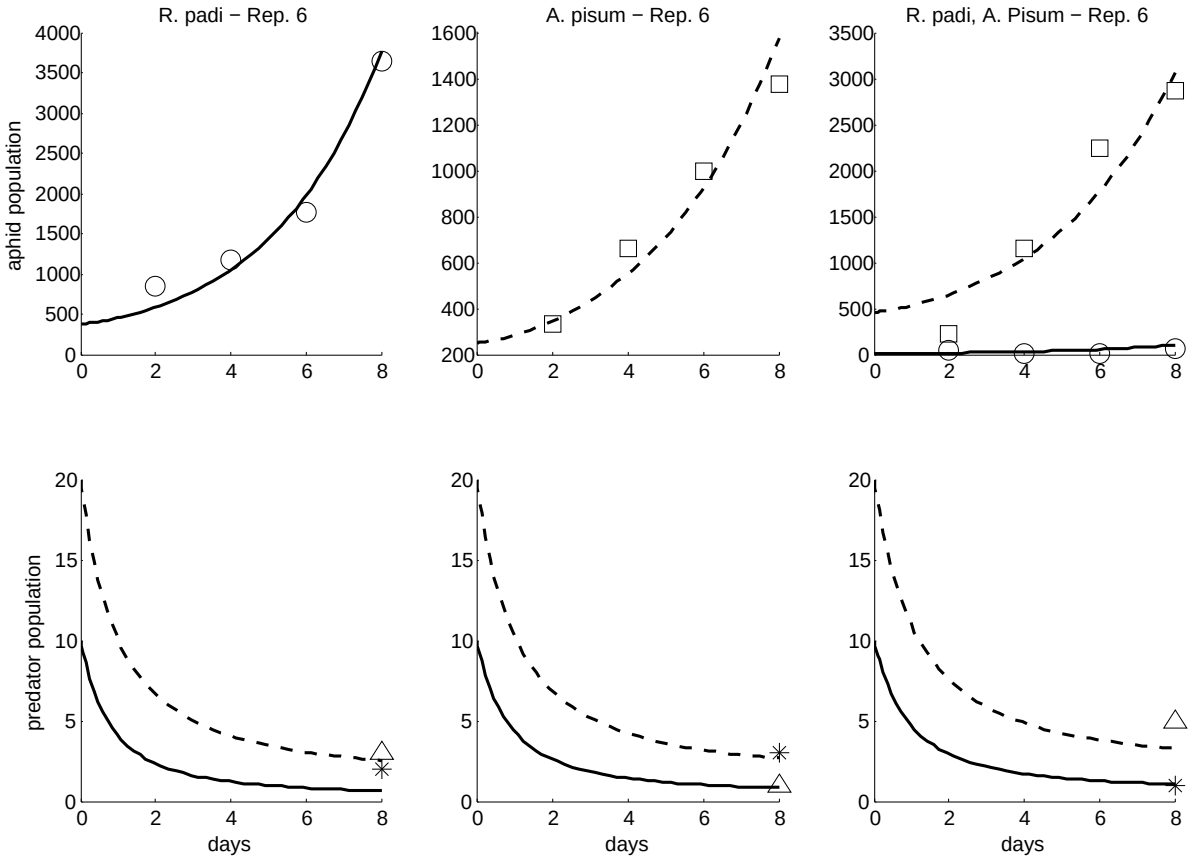

Figure 5.11 ATN model fit to data for a replicate of all aphid treatments under *Pardosa*-*Orius* combined-predator treatment. Model solutions for populations of *Pardosa* and *R. padi* are indicated with solid lines and *Orius* and *A. pisum* are indicated with dashed lines. We use markers to indicate the abundance data for *R. padi* ($\circ$), *A. pisum* ($\square$), *Pardosa* ($\triangle$), and *Orius* ($*$).

We chose not to plot estimated model solutions for the restriction of $\upsilon_0, b_0 = 0$ in the above examples. Population trajectories were generally similar, although sufficient variation existed to reduce graph readability. However, these small differences in model solutions accumulated over the entirety of the data set. In Table 5.7, we list the final iterative reweighted least squares cost associated with the full ATN model and its restriction to $\upsilon_0, b_0 = 0$. We additionally list the number of data points used to obtain our iterative least squares parameter estimates. Following the description of residual sum of squares-based model selection criterion for nested models from Chapter 1, we test the null hypothesis that the "true," or nominal, ATN model parameter values for combined-predator treatments adhere to the restriction $\upsilon_0, b_0 = 0$.

We respectively define $\hat{\theta}^H$ and $\hat{\theta}$ to be the estimated ATN model parameters given in Table 5.6 for the restricted case $\upsilon_0, b_0 = 0$ and for freely varying υ_0, b_0 during parameter estimation. Then for

$J(\theta, \vec{y})$ as defined in 5.2, the realization of our test statistic for the null hypothesis is

$$\hat{U} = \frac{M\left(J(\theta^H, \vec{y}) - J(\theta, \vec{y})\right)}{J(\theta, \vec{y})} = 59.19.$$

Since the null hypothesis restricts the value of two variables, we consult a table of critical values for the χ^2 distribution with two degrees of freedom. We conclude that we can reject the null hypothesis (that $v_0, b_0 = 0$) with 99.9% confidence. That is, the nominal ATN model parameter values for populations subjected to combined-predator treatments do not necessarily satisfy $v_0, b_0 = 0$. We note that due to the number of parameters used in solving the inverse problem, we would have rejected the null hypothesis for any relative difference in iterative reweighted least squares cost for these two cases greater than 1.5%.

Table 5.7 Total variation $TV(\mu_i, \mu_j)$ for species in the cage experiments.

IRLS cost with v_0, b_0 free	$(J(\hat{\theta}, \vec{y}))$	9.06×10^4
IRLS cost with $v_0, b_0 = 0$	$(J(\hat{\theta}^H, \vec{y}))$	9.68×10^4
Number of data points	(M)	865

5.3 Conclusion

We note that in individual-predator treatments, there is insufficient information to estimate parameters for the modified ATN model. We theorize that this is not caused by scarcity of data, but instead by an inappropriate foodweb configuration; the mechanisms described by the modified ATN model are less relevant in a single-predator system, compared to a two-predator system. In contrast, we can estimate the parameters for the modified ATN model in combined-predator treatments with a reasonable degree of confidence. We note that habitat use (as quantified by v_0) seems to be more important in describing aphid dynamics than intraguild predator interference (as quantified by b_0). In future work, we may return to the classic ATN model description of intraspecific competition instead of general predator interference, while retaining our formulation of habitat use overlap.

The modified ATN model fits the data with less error than the model without habitat use or predator interference, and we can reject this simpler model with 99.9% confidence. However, the description of aphid dynamics is similar between the two models, and we can see that the effect of habitat use overlap is absorbed by the optimal predator-prey body mass ratios, R_{opt}, as well as the scaling parameter for attack rates a_0. It remains to be seen if the additional complexity presented by the modified ATN model is necessary to describe predator-prey dynamics in more complicated study systems. Additionally, our specification of habitat use may not be appropriate for *Coccinella*,

which is more voracious than model predictions; we note that instead of randomly moving through its habitat zones, *Coccinella* actively hunts while in proximity to aphids and rests at other times. In future work, we hope to investigate the appropriateness of our formulation of attack rates between *Coccinella* and its prey.

CHAPTER

6

FIELD-LEVEL DESIGN OF EXPERIMENTS

In this chapter, we consider the design of field-level experiments for the validation of predator-prey models for dynamics in a system of terrestrial arthropods. We motivate the use of a partial differential equation model for the dynamics of three populations in a field, as well as a 1-dimensional reduction of this model. We then investigate the estimation of model parameters using synthetically generated data and compare these results to parameter estimation in an analagous ordinary differential equation model. The results from this chapter are an initial proof-of-concept exercise, which we will expand in a future publication.

6.1 Introduction

Throughout this work, we have focused on the validation of ordinary differential equation models for predator-prey dynamics. This type of model has been sufficient for quantifying the interactions between predators and prey on a small scale, but the biological control of pest species in an agricultural field will be driven by spatial processes outside the scope of our current model. Due to the seasonally varying nature of agroecosystems, landscape composition affects the diversity and abundance of predator communities [18, 108, 113] which in turn affects the level of biological control in the field [50, 68, 103]. In particular, landscape use at the border of agricultural fields affects the migration of beneficial predators into a field [54, 110, 111, 118]. We must therefore understand the spatial processes underlying population dynamics in order to describe the predator community

in an agricultural field.

In order to investigate the contribution of these factors to biological control, we develop a simple partial differential equation model for predator-prey dynamics in an agricultural field over a single season. In contrast to our previous modelling efforts, we additionally consider crop dynamics in this model; our primary motivation for this decision is that pest growth is limited by crop availability, but an additional benefit to this decision is our ability to monitor crop viability (and therefore the economic benefits of biological control ecosystem services). We again consider a single population of aphid pests which feed on this crop, but we do not preserve the species-level resolution of the predator community from our previous work. We instead consider a single predator population in this preliminary work, as a multiple-predator model would be complicated and we are primarily interested in understanding spatial effects on the population. Our eventual goal is to formalize questions about biological control as a boundary control problem, since the predator community in an agroecosystem is partially determined by landscape use at the edge of the field.

In this work, we consider the implications of this model's simple spatial dynamics on the design of field-level experiments. We are primarily interested in the utility of this model for theoretical study of biological control under varying conditions, and so we seek an understanding of the data necessary for this model's validation in real-world settings. However, we also study this model within the framework of our attempts at validating the Allometric Trophic Network (ATN) model against field-level dynamics. Regardless of the model to which we fit it, data from a field is inherently dependent on spatial processes. We are therefore interested in the effect of these assumed spatial dynamics on our ability to validate an ordinary differential equation model for population dynamics.

6.2 Spatial Model

We consider the domain of the field to be some $\Omega \subset \mathbb{R}^2$, with boundary $\partial\Omega$. We define the quantities $A(\mathbf{x}, t)$, $B(\mathbf{x}, t)$, and $C(\mathbf{x}, t)$ to be the population densities of the crop, pest, and predator respectively for $\mathbf{x} \in \Omega$ and $t \geq 0$. We assume that the crop grows logistically at some rate r_A with carrying capacity K_A and is consumed at some rate μ_A by the pest, with dynamics given by

$$\frac{d}{dt}A(\mathbf{x}, t) = r_A\left(1 - \frac{A(\mathbf{x}, t)}{K_A}\right)A(\mathbf{x}, t) - \mu_A A(\mathbf{x}, t)B(\mathbf{x}, t).$$

We note that although the density of the crop is defined on the 2-D region Ω, the crop does not have spatial dynamics. We therefore only specify an initial condition for the population, $A(\mathbf{x}, 0) = \mathscr{A}(\mathbf{x})$.

We assume that the pest grows at some rate r_B as crop resources are consumed. In reality, we expect pest growth to be limited; as aphids become crowded in a resource-scarce environment, they switch to the production of flying *alatae* which permit migration to a resource-rich environment [38, 39]. However, because this is a spatial process which does not result in the loss of aphids in the

field, we model this as a diffusion from high-density areas at a rate d_B. In the presence of predators, we assume an additional difffusion at some rate d_{BC}. Although we do not expect aphids to actively evade predators, it has been noted that aphids are knocked from a plant with increasing frequency when disturbed by predators [37, 73, 74]. In greenhouse experiments, we additionally noted that such disturbances resulted in increased movement towards neighboring plants by more-mobile aphids (of the species *A. pisum*). This type of diffusion may not be appropriate for all aphid species or terrains, but we assume that it is possible in our preliminary treatment of this model. We finally assume that the pest is consumed by predators at some rate μ_B.

The dynamics of the pest are therefore given by

$$\frac{\partial}{\partial t}B(\mathbf{x},t) = r_B A(\mathbf{x},t)B(\mathbf{x},t) + \nabla\cdot[(d_B + d_{BC}C(\mathbf{x},t))\nabla B(\mathbf{x},t)] - \mu_B B(\mathbf{x},t)C(\mathbf{x},t).$$

We assume that the initial condition of the pest in the field is given by $B(\mathbf{x},0) = \mathscr{B}(\mathbf{x})$ and that there is no movement of pests beyond the boundary of the field; aphids migrate by flight [38, 39], and so movement from the field would occur over the entirety of the domain, but we assume that the net effect of migration between fields is negligible for this preliminary model. The boundary condition for the population is therefore given by $\nabla B(\mathbf{x},t)|_{\mathbf{x}\in\partial\Omega} = 0$. We note, however, that landscape composition might affect pest migration patterns [14, 117] and that this condition may be affected by land-use around the field; moreover, movement of aphids within a field might be affected by growing different types of crops simultaneously, and the theoretical work of [15] indicates that this might have a strong impact on pest abundance.

We assume that the predator does not experience growth for the duration of the season. Although this may not be a reasonable assumption for true ecosystems, the timing of reproduction and effect on within-field dynamics varies across predator species; in the interest of developing a tractable model for theoretical investigation, we temporarily accept this erroneous assumption. We first assume that predators diffuse at some rate d_C; this is a well-studied assumption for randomly moving foragers [59, 115], although modifications for more complicated behavior are necessary. We additionally assume that predators undergo directed movement with velocity $\vec{V}(\mathbf{x},t)$. This directed movement accounts for "prey-taxis" [65], or the movement of predators towards prey populations, at some rate d_{VB}. This velocity also experiences smoothing with some diffusion constant d_V, which quantifies the effects of intraspecific competition on predator movement [3].

We follow the model development of [3] in our formulation of $\vec{V}(\mathbf{x},t)$, and the resulting dynamics for the predator and its velocity are given by

$$\begin{aligned}\frac{\partial}{\partial t}C(\mathbf{x},t) &= \nabla\cdot[d_C\nabla C(\mathbf{x},t)] + \nabla\cdot[\vec{V}(\mathbf{x},t)C(\mathbf{x},t)],\\ \frac{\partial}{\partial t}\vec{V}(\mathbf{x},t) &= d_{VB}\nabla B(\mathbf{x},t) + d_V\Delta\vec{V}(\mathbf{x},t).\end{aligned}$$

We assume that the initial condition of the predator in the field is given by $C(\mathbf{x},0) = \mathscr{C}(\mathbf{x})$ and that the initial velocity of the predator is given by $\vec{V}(\mathbf{x},0) = \vec{0}$. As with the pest species, we assume that there is no movement of predators beyond the boundary of the field. We eventually hope to incorporate boundary-value dynamics in this model, but the flux of predators at the field boundary depends on the quality of prey resources both within and outside the field. We begin with the assumption that this field is isolated from another predator community. The boundary conditions for the predator and its velocity are therefore given by $\nabla C(\mathbf{x},t)|_{\mathbf{x}\in\partial\Omega} = 0$ and $\vec{V}(\mathbf{x},t)\big|_{\mathbf{x}\in\partial\Omega} = \vec{0}$.

6.2.1 1-D Reduction

For this preliminary work, we consider the reduction of our partial differential equation (PDE) to one dimension. Although we are eventually interested in answering questions on a 2-D domain, we begin by considering a simpler case. We therefore define $\Omega = [0, L]$ for L the length of an agricultural field and have $x \in \mathbb{R}$. The quantities $A(x,t)$, $B(x,t)$ and $C(x,t)$ can therefore be thought of as the population densities of the crop, pest, and predator along a single "slice" of the agriultural field (see Figure 6.1 for a schematic of this domain).

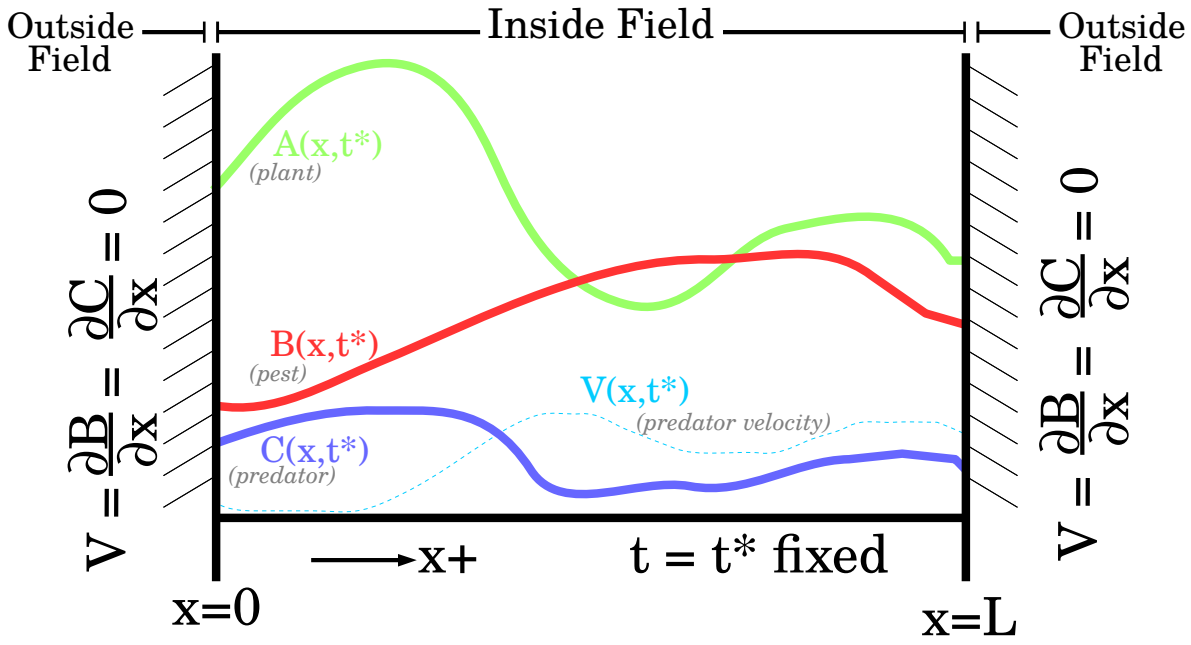

Figure 6.1 The values of the crop, pest, and predator populations are indicated in green, red, and blue respectively at a fixed time t^* along the length of the field, $x \in [0, L]$.

We suppress the functional notation of our states for simplicity, letting $A = A(x,t)$, $B = B(x,t)$,

and $C = C(x,t)$. We additionally note that in one dimension, the velocity $\vec{V}$ can be written as a scalar quantity, for which positive values correspond to motion in the positive x-direction (increasing values of x and negative values correspond to motion in the opposite direction. We denote this quantity by $V = V(x,t)$. The system of differential equations corresponding to this reduction is given by

$$\begin{aligned}
\frac{dA}{dt} &= r_A\left[1-\frac{A}{K_A}\right]A-\mu_A AB,\\
\frac{\partial B}{\partial t} &= r_B AB+[d_B+d_{BC}C]\frac{\partial^2 B}{\partial x^2}+d_{BC}\frac{\partial C}{\partial x}\frac{\partial B}{\partial x}-\mu_B BC,\\
\frac{\partial C}{\partial t} &= d_C\frac{\partial^2 C}{\partial x^2}+\frac{\partial V}{\partial x}C+V\frac{\partial C}{\partial x},\\
\frac{\partial V}{\partial t} &= d_{VB}\frac{\partial B}{\partial x}+d_V\frac{\partial^2 V}{\partial x^2}.
\end{aligned} \tag{6.1}$$

The initial conditions for our system are given by

$$A(x,0)=\mathscr{A}(x),\ \ B(x,0)=\mathscr{B}(x),\ \ C(x,0)=\mathscr{C}(x),\ \ V(x,0)=0,$$

and our states satisfy the boundary conditions

$$\begin{aligned}
\frac{\partial B}{\partial x}(0,t)&=\frac{\partial B}{\partial x}(L,t)=0,\\
\frac{\partial C}{\partial x}(0,t)&=\frac{\partial C}{\partial x}(L,t)=0,\\
V(0,t)&=V(L,t)=0.
\end{aligned}$$

We note that in order to have $V(0,t)=V(L,t)=0$ for all t, we must necessarily have

$$\frac{\partial V}{\partial t}(0,t)=\frac{\partial V}{\partial t}(L,t)=0.$$

However, because $\frac{\partial B}{\partial x}(0,t)=\frac{\partial B}{\partial x}(L,t)=0$, substitution into the PDE for $\frac{\partial V}{\partial t}$ yields an additional boundary condition,

$$\frac{\partial^2 V}{\partial x^2}(0,t)=\frac{\partial^2 V}{\partial x^2}(L,t)=0.$$

We will utilize this condition in our numerical approximation scheme.

6.2.2 Numerical Approximation

To numerically evaluate this PDE, we utilize the method of lines [98] and reduce the model to a system of ordinary differential equations by substituting discretized approximations for spatial

derivatives. We discretize our domain into a uniform grid with mesh size h. The points along this spatial grid are $\{x_k\}_{k=1}^{M}$ given by

$$x_1 = 0,\ x_2 = h,\ \ldots,\ x_M = h(M-1) = L$$

for $M = L/h+1$ the number of grid points. We then define for all $k = 1, \ldots, M$ the states $A^k(t) = A(x_k, t)$, $B^k(t) = B(x_k, t)$, $C^k(t) = C(x_k, t)$, and $V^k(t) = V(x_k, t)$, which vary only in time. We then utilize the central difference approximations for first and second spatial derivatives, given by

$$\frac{\partial F}{\partial x} \approx \frac{F(x+h) - F(x-h)}{2h} \quad \text{and} \quad \frac{\partial^2 F}{\partial x^2} \approx \frac{F(x+h) - 2F(x) + F(x-h)}{h^2}$$

for a general function $F(x)$.

Substituting these approximations for the spatial derivatives in (6.1) for our system of discrete states yields the ordinary differential equations,

$$\begin{aligned}
\frac{dA^k}{dt} &= r_A\left(1 - \frac{A^k}{K_A}\right)A^k - \mu_A A^k B^k, \\
\frac{dB^k}{dt} &= r_B A^k B^k + \left[d_B + d_{BC} C^k\right]\frac{B^{k+1} - 2B^k + B^{k-1}}{h^2}, \\
&\qquad + d_{BC}\frac{\left[C^{k+1} - C^{k-1}\right]\left[B^{k+1} - B^{k-1}\right]}{4h^2} - \mu_B B^k C^k, \\
\frac{dC^k}{dt} &= d_C\frac{C^{k+1} - 2C^k + C^{k-1}}{h^2} + \frac{V^{k+1} - V^{k-1}}{2h}C^k + V^k\frac{C^{k+1} - C^{k-1}}{2h}, \\
\frac{dV^k}{dt} &= d_{VB}\frac{B^{k+1} - B^{k-1}}{2h} + d_V\frac{V^{k+1} - 2V^k + V^{k-1}}{h^2}.
\end{aligned}$$

The initial conditions are given by

$$A^k(0) = \mathscr{A}(x_k),\ \ B^k(0) = \mathscr{B}(x_k),\ \ C^k(0) = \mathscr{C}(x_k),\ \ V^k(0) = 0.$$

We note that for $k = 0, M$ we rely on the boundary conditions of our PDE to define the quantities at x_{k-1} and x_{k+1}, respectively. These conditions are given by

$$\begin{aligned}
&B^{-1} = B^1,\ \ B^{M+1} = B^{M-1}, \\
&C^{-1} = C^1,\ \ C^{M+1} = C^{M-1}, \\
&V^0 = V^M = 0,\ \ V^{M+1} = -V^{M-1},\ \ V^{-1} = -V^1.
\end{aligned}$$

6.2.3 Analagous Ordinary Differential Equation Model

We finally consider the analagous ordinary differential equation (ODE) model for population dynamics over the entire domain. That is, we define the quantities

$$A_I(t) = \int_0^L A(x,t)dx \text{ and } B_I(t) = \int_0^L B(x,t)dx \tag{6.2}$$

and seek the differential equation which describes the dynamics of these states. We note that because the predator does not experience growth and does not cross the boundary of the field, the total predator population is fixed at $\overline{C} = \int_0^L \mathscr{C}(x)dx$ for all time. The dynamics for A_I and B_I analagous to the dynamics assumed by the PDE model are

$$\begin{aligned}
\frac{dA_I}{dt} &= r_A\left[1-\frac{A_I}{K_A}\right]A_I - \mu_A A_I B_I, \\
\frac{dB_I}{dt} &= r_B A_I B_I - \mu_B B_I \overline{C},
\end{aligned}$$

with initial conditions

$$A_I(0) = \int_0^L \mathscr{A}(x)dx \text{ and } B_I(0) = \int_0^L \mathscr{B}(x)dx.$$

We note that in practice, the solutions to this differential equation with fixed model parameters (r_A, K_A, μ_A, r_b, and μ_B) do not satisfy the desired equalities given in (6.2) for $A(x,t)$ and $B(x,t)$ solutions to our PDE model with the same parameters. Instead, given a set of true parameters r_A, K_A, μ_A, r_b, and μ_B, there are best-fitting parameters $r_{A,I}$, $K_{A,I}$, $\mu_{A,I}$, $r_{B,I}$, and $\mu_{B,I}$ such that

$$A_I(t) \approx \int_0^L A(x,t)dx \text{ and } B_I(t) \approx \int_0^L B(x,t)dx$$

for A_I, B_I solutions to the ODE model with parameters $r_{A,I}$, $K_{A,I}$, $\mu_{A,I}$, $r_{B,I}$, and $\mu_{B,I}$ and A, B solutions to the PDE model with parameters r_A, K_A, μ_A, r_b, and μ_B.

6.3 Synthetic Experiments

We consider the duration of a season to be $0 \leq t \leq 90$ days and we define the length of the field to be $L = 1$, scaling our parameters for spatial motion accordingly. We consider the average size of an agricultural field to be 3 hectares, which corresponds to a field length of approximately 170 meters. We refer to [31] for the expected rate of diffusion d_C of the predator community, and in the absence of data on aphid diffusion, we assume that the prey diffuses ten-times more slowly than the

more-mobile predator population. We assume that directed movement of the predators d_{BV} occurs ten-times more quickly than natural diffusion and that the movement of aphids when disturbed by predators occurs ten-times more quickly than natural diffusion. We therefore take the parameter values

$$d_B = 0.0001, d_{BC} = 0.001, d_C = 0.001, d_{VB} = 0.01, \text{ and } d_V = 0.01.$$

We nominally fix the carrying capacity of the crop at $K_A = 1$, and specify the growth rate $r_A = 0.08$, which permits the crops to grow to maturity over the duration of our season. We use the prey growth rate $r_B = 0.75$, following our observations from greenhouse experiments. We set the rates of predation to be $\mu_A = 0.9$ and $\mu_B = 0.7$, which did not result in eradication of the plants or pests for the duration of the season. Following the goal of our previous work, we will assume that the synthetic experiment is conducted with the goal of estimating the predation rate on the pest species, μ_B, but that other parameters are unknown and must be estimated as well.

Specifically, we assume that the parameters μ_A, μ_B, d_B, and d_C must be estimated. We assume that the parameters related to plant and aphid growth (r_A, K_A, and r_B) can be specified from other sources and are not estimated. We additionally assume that the relationships $d_{BC} = 10d_B$ and $d_{VB} = d_V = 10d_C$ are given when solving the inverse problem. Although in future work we may consider estimating additional parameters or incorrectly specifying some of our assumed parameters, we take this idealized case in our preliminary investigations.

We assume that the initial condition of the crop is uniformly $A(x,0) = 0.05$ for all $0 \le x \le L$ and that the initial condition of the prey population is $B(x,0) = 0.1$ for all $0 \le x \le L$. We assume that the predator population is non-uniform at the start of the season, with the highest abundances at the edge of the field. We therefore take $C(0,0) = C(L,0) = .25$ and $C(x,0) = .1$ for all $0 < x < L$. When solving the inverse problem, we let A_0 and C_0 be fixed and only estimate B_0; again, this is aligned with our approach to parameter estimation for similar study systems in our previous work.

We plot in Figure 6.2 the solution for the aphid population which adheres to our model's dynamics under these conditions. On the left, we plot a heat map of the aphid population, with position in the field on the horizontal axis and time increasing along the vertical axis. The population is uniformly distributed to start, before peaking at the edges of the field mid-season. This might seem counter-intuitive, as the predator population is initially concentrated at the edges of the field; the predators begin moving toward the un-exploited aphid population in the middle of the field during the first half of the season, causing the edges to be neglected. By the end of the season, the predators again cover the edge of the field and the aphid population is at its highest in the middle of the field.

On the right, we plot the total aphid population in the field along the vertical axis, against time on the horizontal axis. We indicate with a solid line the true total aphid population, obtained by numerically integrating the solution to the PDE over the spatial domain. With a dashed line, we indicate the closest-fitting ODE approximation to the true total population. We fix $r_{A,I} = r_A$,

$K_{A,I} = K_A$, and $r_{B,I} = r_B$ at the same values as the PDE, but we estimate values of $\mu_{A,I}$ and $\mu_{B,I}$ which yield the desired population levels. We obtain

$$\mu_{A,I} = 0.8960 \text{ and } \mu_{B,I} = 0.6693,$$

which results in an approximated population which closely matches the true solution, with a slight over-estimation at the population's peak.

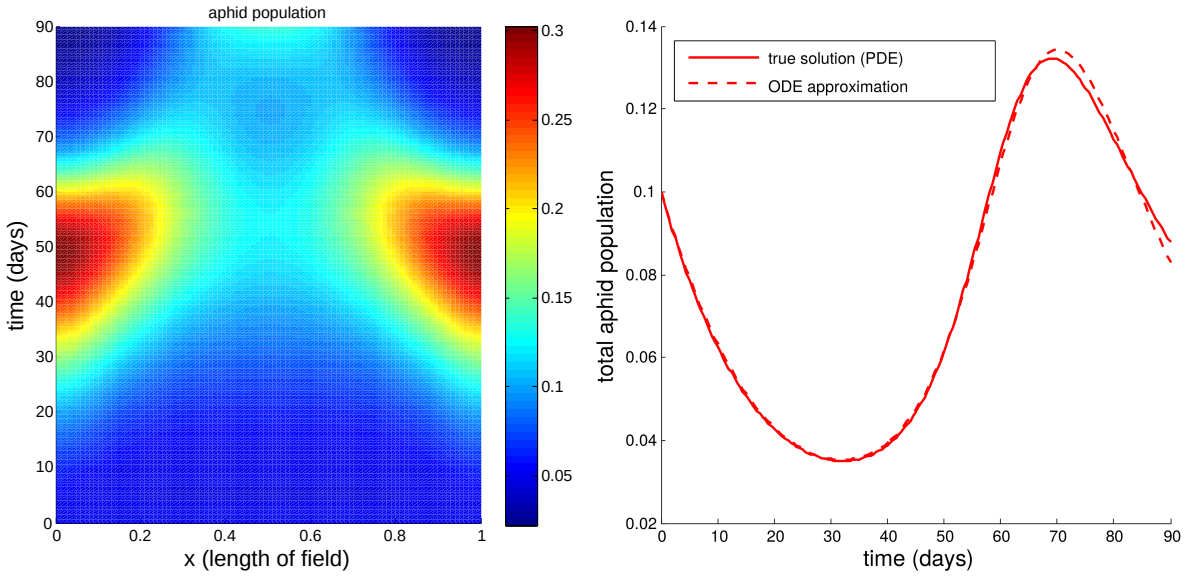

Figure 6.2 The "true" aphid population utilized in our synthetic experiments, plotted against both field length and time on the left. The total aphid population in the field is plotted against time on the right.

We generate synthetic data which is sampled at points $x_j = \Delta x \cdot j$ for $j = 1, 2, ..., L/\Delta x + 1$ and times $t_k = \Delta t \cdot k$ for $k = 1, 2, ..., 90/\Delta t + 1$. We consider spatial grids corresponding to

$$\Delta x = 0.05\,, 0.1,\ 0.15,\ 0.2,\ 0.25,\ 0.3,\ \text{and } 0.5$$

and temporal grids corresponding to waiting

$$\Delta t = 3,\ 7,\ 10,\ 14,\ 21,\ \text{and } 28$$

days between samples. We assume that it is possible to collect data on the pest and predator populations ($B(x,t)$ and $C(x,t)$), but that there is no data on the plant population; although it is certainly possible to monitor plant growth over the course of the experiment, our previous work has

not relied on this type of data. We therefore generate synthetic data $\{\hat{B}_j^i, \hat{C}_j^i\}_{x_j,t_i}$ sampled at every combination of these spatial and temporal grid sizes.

We introduce an absolute error to the data which is normally distributed with mean zero and variance $\sigma^2 = 0.01$ or $\sigma^2 = 0.05$. Data which is noisier than this becomes problematic in estimating parameters in the PDE model; because we are interested in comparing performance between sampling strategies, we choose sufficiently low noise to give an appropriate solution to the inverse problem at some sampling strategies. We use two criteria to evaluate the performance of the inverse problem on synthetic data; we first consider our ability to estimate the parameter μ_B, which is related to the biological control of the pest species, and we additionally consider the accuracy with which we fit the underlying model dynamics.

We directly estimate PDE model parameters by a least squares minimization over the synthetic data, fitting the numerical approximation of our model to the set of all $\{\hat{B}_j^i, \hat{C}_j^i\}_{x_j,t_i}$. To estimate ODE model parameters, we must numerically integrate our synthetic data over our spatial domain. That is, the data for the parameter estimation is given by

$$\hat{B}_I^i = \sum_{j=1}^{L/\Delta x+1} \frac{1}{\Delta x} \hat{B}_j^i,$$

against which we fit the model for $B_I(t)$. We note that because the predator population is assumed to be constant in this model, we only use data from the pest population in estimating model parameters.

6.3.1 Effect of Spatial Grid on Parameter Estimates

We first consider the effect of different values of Δx on our ability to estimate model parameters. Using the PDE model, we hope to identify the value of μ_B in the synthetic experiment. However, in the ODE model, we can only hope to identify the value of $\mu_{B,I}$; it is therefore important to note that the value of $\mu_{B,I}$ introduces an error of approximately 0.04 to the true value of μ_B, and so low error in the estimate of $\mu_{B,I}$ still represents a poor approximation to μ_B. We average the error in our estimate for μ_B in the PDE and $\mu_{B,I}$ in the ODE across all temporal sampling schemes and plot the resulting error as a function of the spatial sampling grid in Figure 6.3. We note that the schemes which introduce the most error for the estimation of $\mu_{B,I}$ in the ODE are the very schemes which introduce the least error for the estimation of μ_B in the PDE. This is caused by our uniform sampling grid.

Grids which do not include the very center of the field ($\Delta x = 0.15$ and $\Delta x = 0.3$) result in poor performance of the ODE in fitting model dynamics. This is because there are periods over the course of the season during which the peak of the aphid population is in the center of the field. By not collecting data at this point, we underestimate the total population in the field at these times. The population then shifts symmetrically so that there are peaks of the population at the two edges

of the field, with a significantly lower population in the middle of the field. After this shift, the approximation of the data for use in the ODE parameter estimation is an overestimation of the true population.

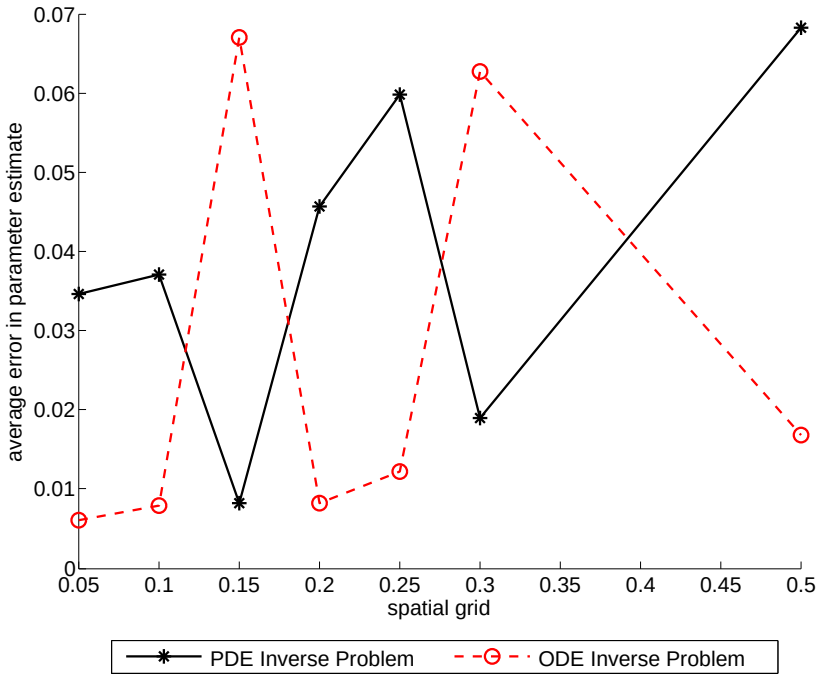

Figure 6.3 The average error in the estimated value of μ_B for the PDE and $\mu_{B,I}$ for the ODE over all temporal sampling strategies. We average over three replicates of the experiment using noise drawn from a normal distribution with standard error $\sigma^2 = 0.01$, plotted against sampling strategy.

These same grids result in the best performance of the PDE in fitting model dynamics. By sampling away from the center of the field, we obtain more information about the rate at which the aphid population shifts across the field. We therefore note that the optimal design of a field-level experiment is linked to the type of dynamics we hope to validate in the experiment. To obtain information about a spatial process, such as diffusion, we seek samples at points which see significant change in the movement of populations, even if we do not obtain information about the points which see the most significant changes in the density of populations.

6.3.2 Error in Parameter Estimates

We next consider the effect of temporal sampling and noise in the data on our estimated parameters. We consider the spatial sampling strategies which corresponded to the best solutions of the inverse problem in the ODE (excluding $\Delta x = 0.15, 0.3$); this permits more cases to study than when only

considering strategies which correspond to the best solutions of the inverse problem for the PDE (excluding $\Delta x = 0.05, 0.1, 0.2, .0.25, 0.5$). In Figure 6.4, we plot the average error in our estimate of μ_B and $\mu_{B,I}$ over the three synthetic experiments with noise drawn from a normal distribution with standard deviation $\sigma^2 = 0.01$ for our range of spatial and temporal grids.

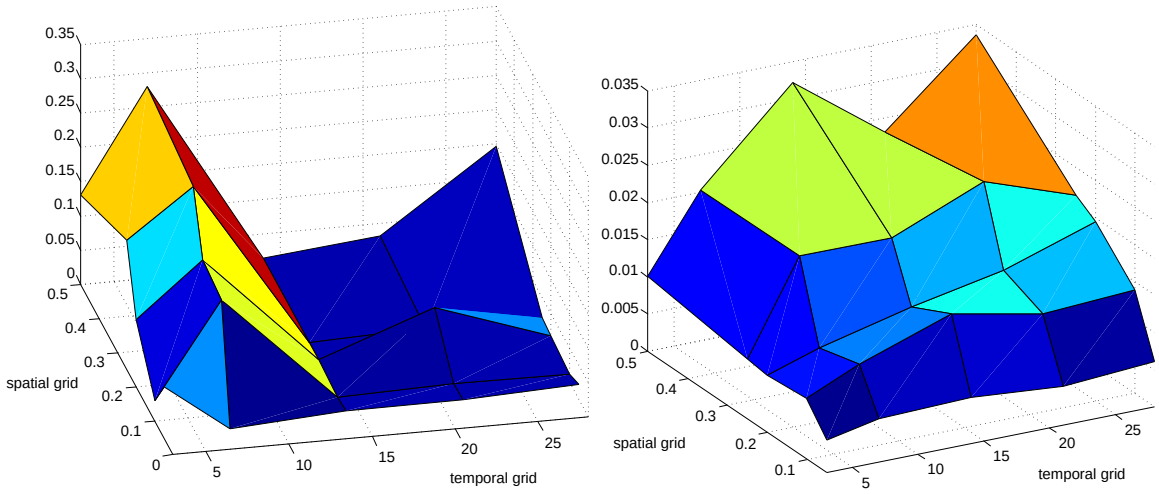

Figure 6.4 The average error in estimation of μ_B using the PDE (left) and $\mu_{B,I}$ using the ODE (right. We consider the average over three synthetic experiments with noise drawn from a normal distribution with standard deviation $\sigma^2 = 0.01$.

We again caution against comparing the scale of the error between the PDE and ODE parameter estimates. Estimates of $\mu_{B,I}$ are often closer to their true value than estimates of μ_B, but this "true" value is not an accurate quantification of the predation rate on the pest population.

We note that the error in parameter estimation for the ODE increases as we increase the size of our temporal and spatial grids. However, contrary to our expectations, the highest error in parameter estimation for the PDE actually occurs for fine temporal grids, when we sample every 3 or 7 days. At these temporal resolutions, we can only accurately estimate the model parameters for the finest spatial grids ($\Delta x = 0.05$). We theorize that pairing a fine temporal grid with a sparse spatial grid results in a fit to noise in our data with insufficient information about the diffusion process driving our PDE. Because the fit to the ODE model averages across our spatial grid, the parameter estimation is not as strongly impacted by this problem.

We next evaluate the performance of our parameter estimation in the three synthetic experiments with noise drawn from a normal distribution with standard deviation $\sigma^2 = 0.05$ for our range of spatial and temporal grids. The average error in our parameter estimates is plotted in Figure 6.5. In this case, we note that the accuracy of the parameter estimation in the PDE model is almost

entirely-determined by our choice of spatial grid. In contrast, we see that for estimates of μ_B in the ODE model, we have accurate results for relatively sparse spatial grids when using temporal grids of 3 or 7 days between experiments. We note that in reality, such a fine temporal grid may not be feasible; if we take a sufficiently fine spatial grid, we can estimate model parameters for both the ODE and PDE with data collected every 14 or 21 days.

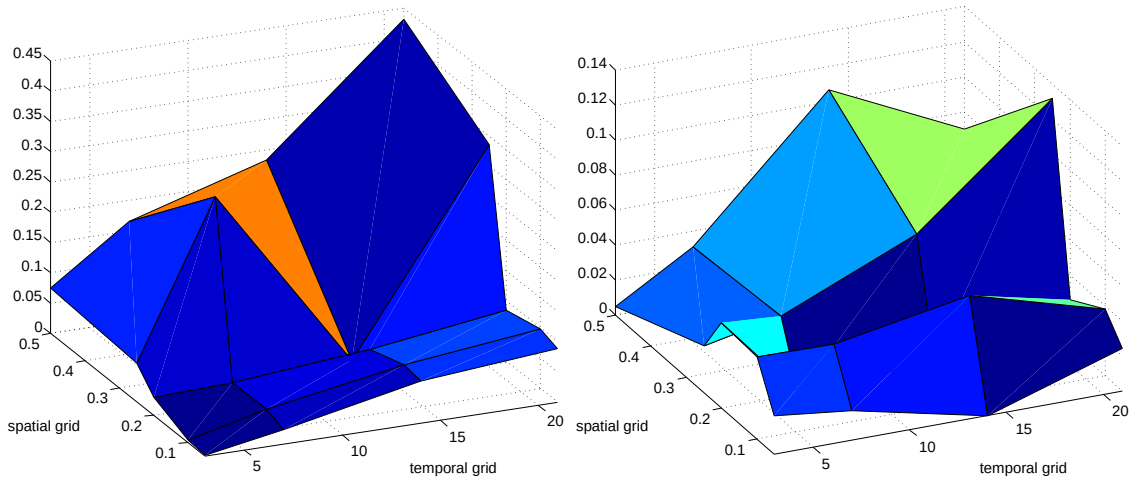

Figure 6.5 The average error in estimation of μ_B using the PDE (left) and ODE (right) over three synthetic experiments with noise drawn from a normal distribution with standard deviation $\sigma^2 = 0.05$.

6.3.3 Error in Population Approximation

We finally consider the accuracy with which our estimated model parameters describe the pest population in the field. In Figures 6.6 and 6.7, we plot the approximation to the aphid population from data with noise drawn from a normal distribution with standard deviation $\sigma^2 = 0.01$. We compare the approximation to the true distribution of the aphid population for temporal samples obtained every seven days and three different spatial sampling strategies, $\Delta x = 0.05,\ 0.25$, and $\Delta x = 0.50$. As we saw in our analysis of the error in estimated parameters, we note that the approximation closely follows the true population for the fine sampling grid ($\Delta x = 0.05$), but that the approximation is poor over a coarse sampling grid ($\Delta x = 0.50$).

Although the results for the coarse grid seems discouraging at a glance, we note that the finer spatial grids are feasible sampling strategies. For the assumed size of our synthetic field, a distance of $\Delta x = 0.05$ between samples is approximately 8.5 meters, while a distance of $\Delta x = 0.25$ is approximately 42.5 meters. In consulting the literature, we find that pitfall traps for predators can be placed with anywhere from 10 meters between traps to 50 meters between traps [21]. The grid over which we can accurately estimate model parameters and recreate the true distribution of prey in the field

is only slightly finer than sampling strategies reported in the literature. Samples of prey populations are often obtained over a finer grid than this, although this might occur over only a subset of the field.

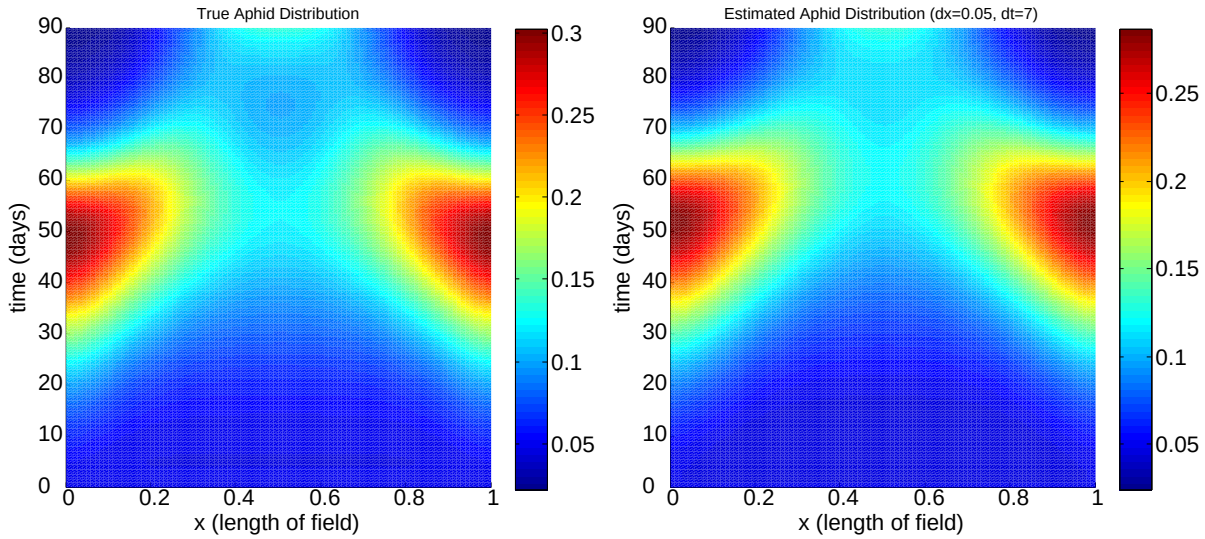

Figure 6.6 The solution for the true pest population (left) and the approximated pest population using a spatial grid of $\Delta x = .05$ and 7 days between samples (right).

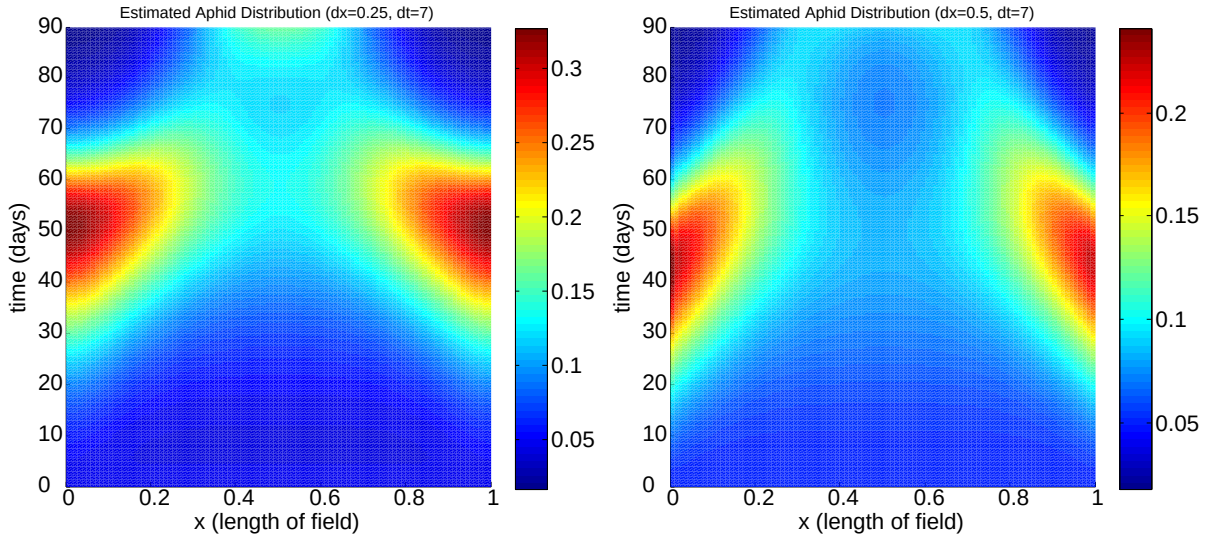

Figure 6.7 The solutions for the approximated pest population using 7 days between samples and $\Delta x = .25$ (left) or $\Delta x = .50$ (right).

We additionally plot the performance of the ODE in matching the true population in the field for these two cases in Figure 6.8. We note that for both the coarse and fine spatial sampling grid, we accurately capture the true dynamics of the prey population. Because the estimate of $\mu_{B,I}$ is not an accurate description of predator feeding rates, it is particularly important that the ODE model fit the true population in the field. There is little information to be gained from estimating ODE model parameters, and so our primary goal is to obtain an appropriate approximation to the pest population. Our results indicate that, for the current PDE model formulation, this is possible with many sampling strategies. Moreover, we can obtain such an approximation with less data than is required to validate the PDE model. We are cautiously optimistic about the implications this might have for efforts to validate field-level populations against ODE models, although we must consider different PDE model formulations to describe the breadth of spatial processes which might affect the population.

We note that the accuracy of the ODE model approximation is sensitive to noise in the data. In Figure 6.9, we plot the performance of the ODE in matching the true population in the field for data with noise drawn from a normal distribution with standard deviation $\sigma^2 = 0.05$, utilizing the same temporal and spatial strategies as we previously explored. Although the PDE performs similarly with the higher level of noise (results not shown), the ODE approximation to prey dynamics is worse in this case.

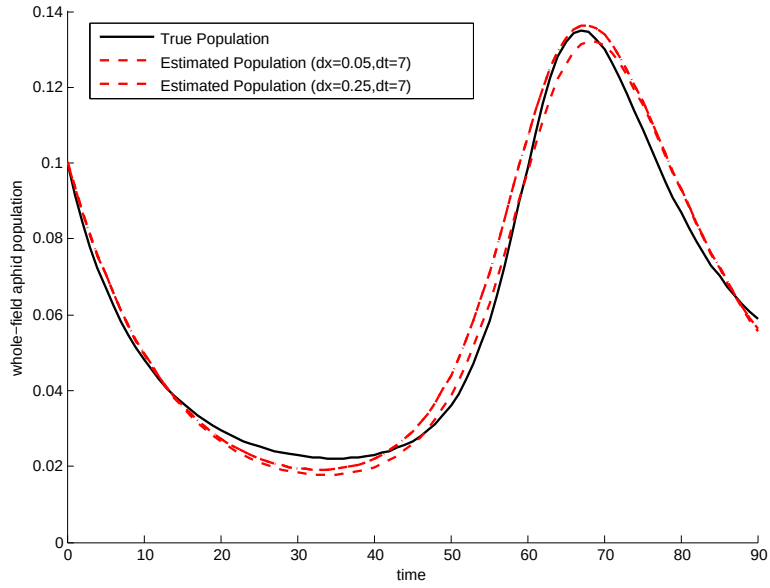

Figure 6.8 The solution for the true pest population and the approximated pest population using a spatial grids of $\Delta x = .05, .25$ and 7 days between samples and noise with standard deviation $\sigma^2 = 0.01$

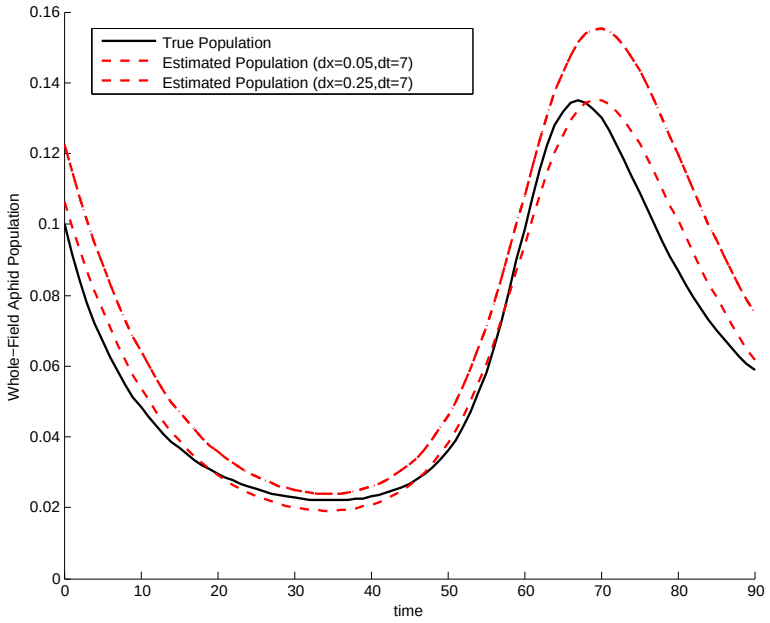

Figure 6.9 The solution for the true pest population and the approximated pest population using a spatial grids of $\Delta x = .05, .25$ and 7 days between samples and noise drawn with standard deviation $\sigma^2 = 0.05$.

6.4 Directions for Future Work

The preliminary work we have presented here is specific to our assumed boundary conditions and parameters. We are therefore interested in first studying the effect that weaker or stronger spatial processes, as described by the values of d_B and d_C, might have on sampling protocols. In particular, we would like to identify the degree to which spatial processes affect our ability to match the ODE model to the total population in the field as described by PDE dynamics. We then hope to loosen our restrictions on boundary conditions in order to accomodate realistic descriptions of alternative farming practices. Different types of field boundaries can affect aphid and predator migration, and we expect that non-insulated boundary conditions (corresponding to fields with natural foliage at the borders) will present a challenge when fitting the ODE model to field data.

Another interesting avenue for future research is the consideration of pitfall traps in experimental design. Ground-dwelling predators are sampled through the use of pitfall traps, which are cups buried in the ground and left open for a designated period (for example, several days). The contents of the cups are collected after a period, and an approximation to the predator population is obtained from the number of individuals caught in the cup. There are different transformations of the data for this approximation, but our current work has assumed that the results from pitfall traps are a near-perfect report of the predator population. Moreover, *directional* pitfall traps exist, which distinguish between predators travelling in opposite directions; we are interested in the implications

that this knowledge might have on the estimation of model parameters.

Finally, we hope to expand our model back to the two-dimensional case and repeat similar analyses. In this more-realistic case, we can investigate the effect of radial or grid-like sampling of the predator population. We can also simulate a realistic aphid sampling strategy, in which populations are highly sampled in small, discrete regions within the field. In the two-dimensional case, we additionally hope to conduct preliminary analysis of the biological control of the aphid as it relates to the assumed boundary conditions for the field; these conditions can affect both the migrating predator population and the seasonal colonization of the field by the aphid population. Although we must validate our model against real data before we can make statements about true biological control, we can begin by testing the sensitivity of model solutions to different descriptions of the field boundary.

CHAPTER

7

CONCLUDING REMARKS

Throughout this work, we have relied on the interplay between the biological and mathematical sciences, with the eventual goal of addressing real-world ecological questions in a quantitative manner. We have employed least squares parameter estimation as the means by which we validate models for biological processes against empirical data, and we have consistently referred back to an ecological understanding of these processes in our model development. We were initially motivated by an interest in modelling the expected level of biological control in an agricultural ecosystem by a pest's natural predators. However, our attempts to address this question became an exercise in the iterative process of modelling, in which both our mathematical models and biological hypotheses developed over time.

Our own work began many iterations into this process, as there were already well-established models for predator-prey dynamics appropriate for our study system. We started with an existing trait-based model for these dynamics and attempted to fit the model to experimental data from ten true agro-ecosystems. We quickly discovered that, although we chose a simple model for these initial eforts, there was insufficient data for the reliable estimation of model parameters. Relying on our biological understanding of the study system, we concluded that the mechanisms described by the estimated parameters were physically unreasonable. Within the constraints of the existing data set, we were unable to validate our model or draw conclusions about its deficiencies.

In pursuit of a suitable data set for model validation, we returned to our understanding of the inverse problem methodology. We designed controlled experiments against which we hoped to test

the model, studying subsets of this same agro-ecosystem inside a greenhouse. Because of our heavy reliance on experimental data in the iterative modelling process, it is important that we consider our means of model validation in experimental design. We considered model sensitivities and characteristics of population data in similar experiments to determine appropriate protocols. By linking our model and mathematical methodology for validation with the design of the experiment, we safeguarded against potential shortcomings in the resulting data set.

We were able to fit the simple model from previous work, as well as a more-complicated model which was formulated to reflect new biological hypotheses, to the the experimental data with varying degrees of success. We evaluated model performance under different conditions and – because of the link between our biological assumptions, the mathematical model, and the experiment designed to test that model – we were able to determine the relevance of our biological hypotheses. We identified future work with information from greenhouse and field experiments which could further our understanding of these agro-ecosystems.

We finally considered some questions of experimental design in the field. We employed a simple spatial model for ecological interactions in a preliminary effort to explore the effect of landscape-level interactions on the population data used in model validation. Our results should be verified against existing data and more realistic models before informing experimental design. However, we concluded that the interaction between temporal and spatial sampling strategies is critically important in determining an appropriate design for field experiments. Our efforts in this dissertation have set the stage for future work modelling biological control agents in an agricultural ecosystem, both with and without spatial considerations.

BIBLIOGRAPHY

[1] Agarawala, B. & Dixon, A. "Laboratory study of cannibalism and interspecific predation in ladybirds". *Ecological Entomology* **17** (1992), pp. 303–309.

[2] Allinson, M. *Royal Society Open Science launches Registered Reports*. Ed. by Society, T. R. 2015.

[3] Arditi, R., Tyutyunov, Y., Murgulis, A., Govorukhin, V. & Senina, I. "Directed movement of predators and the emergence of density-dependence in predator-prey models". *Theoretical Population Biology* (2001), pp. 207–221.

[4] Arneberg, P. & Andersen, J. "The energetic equivalence rule rejected because of a potentially misleading sampling error: evidence from carabid beetles". *Oikos* **101** (2003), pp. 367–375.

[5] Assessment, M. E. "Ecosystems and human well-being: biodiversity synthesis". *World Resources Institute* (2005).

[6] Assessment, T. U. N. E. "The UK National Ecosystem Assessment Technical Report". *UNEP-WCMC* (2011).

[7] Banks, H. T., Catenacci, J. & Hu, S. "Use of difference-based methods to explore statistical and mathematical model discrepancy in inverse problems". *Journal of Inverse and Ill-posed Problems* **24** (2016), pp. 413–433.

[8] Banks, H. T. et al. "Dynamic modelling of problem drinkers undergong behavioral treatment". *Bulletin of Mathematical Biology* **79** (2017), pp. 1254–1273.

[9] Banks, H., Bekele-Maxwell, K., Bociu, L., Noorman, M. & Tillman, K. "The complex-step method for sensitivity analysis of non-smooth problems arising in biology". *Eurasian Journal of Mathematical and Computer Applications* **3** (2015), pp. 16–68.

[10] Banks, H., Hu, S. & Thompson, W. *Modeling and Inverse Problems in the Presence of Uncertainty*. Boca Raton: CRC Press, 2014.

[11] Banks, H. & Tran, H. *Mathematical and Experimental Modeling of Physical and Biological Processes*. New York: CRC Press, 2009.

[12] Banks, H. et al. "Parameter estimation for an allometric food web model". CRSC Technical Report, CRSC-TR16-03. 2016.

[13] Banks, H. et al. "Parameter estimation for an allometric food web model". *International Journal of Pure and Applied Mathematics* **114** (2017), pp. 143–160.

[14] Banks, J. E. "The scale of landscape fragmentation affects herbivore response to vegetation heterogeneity". *Oecologia* **117** (1998), pp. 239–246.

[15] Banks, J. E. & Ekbom, B. “Modelling herbivore movement and colonization: pest management potential of intercropping and trap cropping”. *Agricultural and Forest Entomology* **1** (1999), pp. 165–170.

[16] Beretta, E., Solimano, F. & Takeuchi, Y. “Global stability and periodic orbits for two-patch predator-prey diffusion-delay models”. *Mathematical Biosciences* **85** (1987), pp. 153–183.

[17] Berlow, E. L. et al. “Simple prediction of interaction strengths in complex food webs”. *Proceedings of the National Academy of Sciences of the United States of America* **106** (2009), pp. 187–191.

[18] Bianchi, F., Booij, C. & Tscharntke, T. “Sustainable pest regulation in agricultural landscapes: a review on landscape composition, biodiversity and natural pest control”. *Proceedings of the Royal Society B* **273** (2006), pp. 1715–1727.

[19] Bilde, T. & Topping, C. “Life history traits interact with landscape composition to influence population dynamics of a terrestrial arthropod: A simulation study”. *Ecoscience* **11** (2004), pp. 64–73.

[20] Boit, A., Martinez, N. D., Williams, R. J. & Gaedke, U. “Mechanistic theory and modelling of complex food-web dynamics in Lake Constance”. *Ecology Letters* **15** (2012), pp. 594–602.

[21] Bommarco, R. “Reproduction and energy reserves of a predatory carabid beetle relative to agroecosystem complexity”. *Ecological Applications* **8** (1998), pp. 846–853.

[22] Bommarco, R. & Fagan, W. F. “Influence of crop edges on movement of generalist predators: a diffusion approach”. *Agricultural and Forest Entomology* **4** (2002), pp. 21–30.

[23] Bourget, D. & Guillemaud, T. “The hidden and external costs of pesticide use”. *Sustainable Agriculture Reviews* **39** (2016), pp. 35–120.

[24] Brose, U. et al. “Foraging theory predicts predator-prey energy fluxes”. *Journal of Animal Ecology* (2008).

[25] Brose, U. “Body-mass constraints on foraging behaviour determine population and food-web dynamics”. *Functional Ecology* **24** (2010), pp. 28–34.

[26] Brose, U., Williams, R. J. & Martinez, N. D. “Allometric scaling enhances stability in complex foodwebs”. *Ecology Letters* **9** (2006), pp. 1228–1236.

[27] Cardinale, B. et al. “Biodiversity loss and its impact on humanity”. *Nature* **486** (2012), pp. 323–330.

[28] Clough, Y. et al. “Predator body masses determine aphid biological control by natural predator communities”. Unpublished manuscript.

[29] Codling, E. A., Plank, M. J. & Benhamou, S. “Random walk models in biology”. *Journal of the Royal Society Interface* **5** (2008), pp. 813–834.

[30] Cohen, J., Łuczak, T, Newman, C. & Zhou, Z. “Stochastic structure and nonlinear dynamics of food webs: qualitative stability in a Lotka-Volterra cascade model”. *Proc. R. Soc. Lond. B.* Vol. 240. 1990, pp. 607–627.

[31] Corbett, A. & Plant, R. E. “Role of movement in the response of natural enemies to agroecosystem diversification: A theoretical evaluation”. *Environmental Entomology* **22** (1993), pp. 591–531.

[32] Currie, C., Spence, J. & Niemelä, J. “Competition, cannibalism and intraguild predation among ground beetles (Coleoptera:Carabidae): A laboratory study”. *The Coleopterists Bulletin* **50** (1996), pp. 135–148.

[33] Curtsdotter, A. et al. “Confronting dynamic food web models with population time series data to understand ecosystem functioning in natural predator-prey”. 2016.

[34] Dean, G. “Effect of temperature on the cereal aphids *Metopolophium dirhodum* (Wlk.), *Rhopalosiphum padi* (L.) and *Macrosiphum avenue* (F.)(Hem., Aphididae)”. *Bulletin of Entomological Research* **63** (1974), pp. 401–409.

[35] DeAngelis, D. L. & Grimm, V. “Individual-based models in ecology after four decades”. *F1000prime reports* **6** (2014).

[36] Dennis, B., Ponciano, J. M., Lele, S. R., Taper, M. L. & Staples, D. F. “Estimating density dependence, process noise, and observation error”. *Ecological Monographs* **76** (2006), pp. 323–341.

[37] Dixon, A. F. G. “The escape responses shown by certain aphids to the presence of the coccinellid *Adalia decempunctata (L.)*” *Ecological Entomology* **110** (1958), pp. 319–334.

[38] Dixon, A. F. G. “The life-cycle and host preferences of the bird cherry-oat aphid, *Rhopalosiphum padi* L., and their bearing on the theories of host alternation in aphids”. *Annals of Applied Biology* **68** (1971), pp. 135–147.

[39] Dixon, A. F. G. “Aphid ecology: life cycles, polymorphism, and population regulation”. *Annual Review of Ecology and Systematics* **8** (1977), pp. 329–353.

[40] Duffy, J. et al. “The functional role of biodiversity in ecosystems: incorporating trophic complexity”. *Ecology Letters* **10** (2007), pp. 522–538.

[41] Durrett, R. *Probability: Theory and Examples*. Third. Belmond, California: Brooks and Cole, 2005, p. 137.

[42] Edelstein-Keshet, L. *Mathematical Models in Biology*. New York: Random House, 1988.

[43] Ehnes, R., Rall, B. & Brose, U. “Phylogenetic grouping, curvature, and metabolic scaling in terrestrial invertebrates”. *Ecology Letters* **14** (2011), pp. 993–1000.

[44] Eklöf, A. et al. “The dimensionality of ecological networks”. *Ecology Letters* **16** (2013), pp. 577–583.

[45] Elton, C. & Nicholson, M. “The Ten-Year Cycle in Numbers of the Lynx in Canada”. *Journal of Animal Ecology* **11** (1942), pp. 215–244.

[46] Eslami, M. *Theory of Sensitivity in Dynamic Systems: An Introduction*. Springer-Verlag, Berlin, 1994.

[47] Fauchald, P., Erikstad, K. E. & Skarsfjord, H. “Scale-dependent predator–prey interactions: the hierarchical spatial distribution of seabirds and prey”. *Ecology* **81** (2000), pp. 773–783.

[48] Francis, T. B. et al. “Shifting regimes and changing interactions in the Lake Washington, USA, plankton community from 1962–1994”. *PloS one* **9** (2014).

[49] Frank, P. M. *Introduction to System Sensitivity Theory*. New York: Academic Press, 1978.

[50] Gardiner, M. et al. “Landscape diversity enhances biological control of an introduced crop pest in the north-central USA”. *Ecological Applications* **19** (2009), pp. 143–154.

[51] Gilljam, D. et al. “Seeing double: size-based and taxonomic views of food web structure”. *Advances in Ecological Research* **45** (2011), pp. 67–133.

[52] Gonzales, J. E. & Cunningham, C. A. *The promise of pre-registration in psychological research*. http://www.apa.org/science/about/psa/2015/08/pre-registration.aspx. Accessed: 2017-07-27.

[53] Gras, R., Golestani, A., Hendry, A. P. & Cristescu, M. E. “Speciation without pre-defined fitness functions”. *PloS one* **10** (2015).

[54] Griffiths, G. J., Winder, L., Holland, J. M., Thomas, C. G. & Williams, E. “The representation and functional composition of carabid and staphylinid beetles in different field boundary types at a farm scale”. *Biological Conservation* **135** (2005), pp. 145–152.

[55] Hagen, K. “Biology and ecology of predaceous coccinellidae”. *Annual Review of Entomology* (1962), pp. 289–327.

[56] Hampton, S. E. et al. “Quantifying effects of abiotic and biotic drivers on community dynamics with multivariate autoregressive (MAR) models”. *Ecology* **94** (2013), pp. 2663–2669.

[57] Hofbauer, J. & Sigmund, K. “Evolutionary game dynamics”. *Bulletin of the American Mathematical Society* **40** (2003), pp. 479–519.

[58] Holling, C. S. “Some Characteristics of Simple Types of Predation and Parasitism”. *The Canadian Entomologist* **91** (1959), pp. 385–398.

[59] Holmes, E., Lewis, M., Banks, J. & Veit, R. “Partial differential equations in ecology: spatial interactions and population dynamics”. *Ecology* (1994), pp. 17–29.

[60] Ives, A., Dennis, B., Cottingham, K. & Carpenter, S. “Estimating community stability and ecological interactions from time-series data”. *Ecological monographs* **73** (2003), pp. 301–330.

[61] Jonsson, M. “Food webs in barley fields: implications for biological control”. *IOBC-WPRS Bulletin* **122** (2017), pp. 94–97.

[62] Jonsson, T., Kaartinen, R., Jonsson, M. & Bommarco, R. “Predictive power of food web models based on body size decreases with trophic complexity”. Accepted - Ecology Letters.

[63] Kalinoski, R. M. & DeLong, J. P. “Beyond body mass: how prey traits improve predictions of functional response parameters”. *Oecologia* **180** (2016), pp. 543–550.

[64] Kareiva, P. “Habitat fragmentation and the stability of predator–prey interactions”. *Nature* **326** (1987).

[65] Kareiva, P. & Odell, G. “Swarms of predators exhibit" preytaxis" if individual predators use area-restricted search”. *The American Naturalist* **130** (1987), pp. 233–270.

[66] Kot, M. *Elements of Mathematical Ecology*. New York: Cambridge University Press, 2001.

[67] Kuang, Y. & Takeuchi, Y. “Predator-prey dynamics in models of prey dispersal in two-patch environments”. *Mathematical Biosciences* **120** (1994), pp. 77–98.

[68] Landis, D. A., Wratten, S. D. & Gurr, G. M. “Habitat management to conserve natural enemies of arthropod pests in agriculture”. *Annual Review of Entomology* **45** (2000), pp. 175–201.

[69] Laubmeier, A. N. et al. “From theory to experimental design - quantifying a trait-based theory of predator-prey dynamics”. *Submitted to PLOS ONE* (2017).

[70] Leather, S. & Dixon, A. “The effect of cereal growth stage and feeding site on the reproductive activity of the bird-cherry aphid, Rhopalosiphum padi”. *Annals of Applied Biology* **97** (1981), pp. 135–141.

[71] Letourneau, D., Jedlicka, J., Bothwell, S. & Moreno, C. “Effects of natural enemy biodiversity on the suppression of arthropod herbivores in terrestrial ecosystems”. *Annual Review of Ecology, Evolution, and Systematics* **40** (2009), pp. 573–592.

[72] Losey, J. & Vaughan, M. “The economic value of ecological services provided by insects”. *BioScience* **56** (2006), pp. 311–323.

[73] Losey, J. E. & Denno, R. F. “Positive predator-predator interactions: enhanced predation rates and synergistic suppression of aphid populations”. *Ecology* **79** (1998), pp. 2143–2152.

[74] Losey, J. E. & Denno, R. F. “The escape response of pea aphids to foliar-foraging predators: factors affecting dropping behaviour”. *Ecological Entomology* **23** (1998), pp. 53–61.

[75] Mao, X., Sabanis, S. & Renshaw, E. “Asymptotic behaviour of the stochastic Lotka-Volterra model”. *Journal of Mathematical Analysis and Applications* **287** (2003), pp. 141–156.

[76] Mashayekhi, M. & Gras, R. “Investigating the effect of spatial distribution and spatiotemporal information on speciation using individual-based ecosystem simulation”. *GSTF Journal of Computing* **2** (2012), pp. 98–103.

[77] McLane, A. J., Semeniuk, C., McDermid, G. J. & Marceau, D. J. “The role of agent-based models in wildlife ecology and management”. *Ecological Modelling* **222** (2011), pp. 1544–1556.

[78] Morrison, A. E. & Addison, D. J. “Assessing the role of climate change and human predation on marine resources at the Fatu-ma-Futi site, Tutuila Island, American Samoa: an agent based model”. *Archaeology in Oceania* **43** (2008), pp. 22–34.

[79] Murray, J. D. *Mathematical Biology I: An Introduction*. 3rd. 2002.

[80] Ottaviani, D., Cairns, S., Oliverio, M. & Boitani, L. “Body mass as a predictive variable of home-range size among Italian mammals and birds”. *Journal of Zoology* **269**.3 (2006), pp. 317–330.

[81] Pain, E. *Register your study as a new publication option*. Ed. by Science. 2015.

[82] Peckarsky, B. L., McIntosh, A. R., Àlvarez, M. & Moslemi, J. M. “Disturbance legacies and nutrient limitation influence interactions between grazers and algae in high elevation streams”. *Ecosphere* **6** (2015), pp. 1–15.

[83] Persson, L., Leonardsson, K., Roos, A. M. de, Gyleenberg, M. & Christensen, B. “Ontogenetic scaling of foraging rates and the synamics of a size-structured consumer-resource model”. *Theoretical Population Biology* **54** (1998), pp. 270–293.

[84] Peters, R. H. *The Ecological Eimplications of Body Size*. New York: Cambridge University Press, 1983.

[85] Platt, J. “Strong Inference”. *Science* **146** (1964), pp. 347–353.

[86] Polis, G. A., Myers, C. A. & Holt, R. D. “The ecology and evolution of intraguild predation: potential competitors that eat each other”. *Annual Review of Ecology and Systematics* **20** (1989), pp. 297–330.

[87] Rall, B. C., Vucic-Pestic, O., Ehnes, R. B., Emmerson, M. & Brose, U. "Temperature, predator-prey interaction strength and population stability". *Global Change Biology* **16** (2010), pp. 2145–2157.

[88] Rall, B. C., Kalinkat, G., Vucic-Pestic, O. & Brose, U. "Taxonomic versus allometric constraints on non-linear interaction strengths". *Oikos* **120** (2011), pp. 483–492.

[89] Reichenbach, T., Mobilia, M. & Frey, E. "Coexistence versus extinction in the stochastic cyclic Lotka-Volterra model". *Physical Review E* **74** (2006).

[90] Rockström, J. et al. "A safe operating space for humanity". *Nature* **461** (2009), pp. 472–475.

[91] Rosenheim, J. A., Wilhoit, L. R. & Armer, C. A. "Influence of intraguild predation among generalist insect predators on the suppression of an herbivore population". *Oecologia* **96** (1993), pp. 439–449.

[92] Ross, C. T. & Winterhalder, B. "Sit-and-wait versus active-search hunting: A behavioral ecological model of optimal search mode". *Journal of Theoretical Biology* **387** (2015), pp. 76–87.

[93] Roubinet, E., Jonsson, T. & Jonsson, M. "Low level of specialization and seasonal variation characterizes the predator-prey food web structure in agroecosystems". Unpublished manuscript.

[94] Roubinet, E. et al. "The diet of generalist predators reflects effects of cropping season and farming system on extra- and intraguild prey". *Ecological Applications* **27** (2017), pp. 1167–1177.

[95] Roubinet, E. et al. "Additive effects of predator diversity on pest control caused by few interactions among predator species". *Ecological Entomoloy* **40** (2015), pp. 362–371.

[96] Rykiel Jr., E. J. "Testing ecological models: the meaning of validation". *Ecological Modelling* **90** (1996), pp. 229–244.

[97] Sabelis, M., Diekmann, O. & Jansen, V. "Metapopulation persistence despite local extinction: predator-prey patch models of the Lotka-Volterra type". *Metapopulation Dynamics: Empirical and Theoretical Investigations*. 1991, pp. 267–283.

[98] Schiesser, W. E. *The numerical method of lines: integration of partial differential equations*. Elsevier, 2012, pp. 10–18, 98–99.

[99] Schmitz, O. J. "Predator diversity and trophic interactions". *Ecology* **88** (2007), pp. 2415–2426.

[100] Schmitz, O. J., Krivan, V. & Ovadia, O. "Trophic cascads: the primacy of trait-mediated indirect interactions". *Ecology Letters* **7** (2004).

[101] Schneider, F. D., Scheu, S. & Brose, U. “Body mass constraints on feeding rates determine the consequences of predator loss”. *Ecology Letters* **15** (2012), pp. 436–443.

[102] Skalski, G. & Gilliam, J. “Functional responses with predator interference: viable alternatives to the Holling type II model”. *Ecology* **82** (2001), pp. 3083–3092.

[103] Snyder, W. E., Snyder, G. B., Finke, D. L. & Straub, C. S. “Predator biodiversity strengthens herbivore suppression”. *Ecology Letters* **9** (2006), pp. 789–796.

[104] Stark, J., Vargas, R. & Banks, J. “Incorporating ecologically relevant measures of pesticide effect for estimating the compatibility of pesticides and biocontrol agents”. *Journal of Economic Entomology* **100** (2007), pp. 1027–1032.

[105] Staudacher, K., Jonsson, M. & Traugott, M. “Diagnostic PCR assays to unravel food web interactions in cereal crops with focus on biological control of aphids”. *Journal of Pest Science* **89** (2016), pp. 281–293.

[106] Stiling, P. “Why do natural enemies fail in classical biological control programs?” *American Entomologist* **39** (1993), pp. 31–37.

[107] Straub, C., Finke, D. & Snyder, W. “Are the conservation of natural enemy biodiversity and biological control compatible goals?” *Biological Control* **45** (2008), pp. 225–237.

[108] Sunderland, K. & Samu, F. “Effects of agricultural diversification on the abundance, distribution, and pest control potential of spiders: a review”. *Entomologia Experimentalis et Applicata* **95** (2000), pp. 1–13.

[109] Takeuchi, Y., Du, N., Hieu, N. & Sato, K. “Evolution of predator–prey systems described by a Lotka–Volterra equation under random environment”. *Journal of Mathematical Analysis and applications* **323** (2006), pp. 938–957.

[110] Thomas, C., Brown, N. & Kendall, D. “Carabid movement and vegetation density: Implications for interpreting pitfall trap data from split-field trials”. *Agriculture, Ecosystems and Environment* **113** (2006), pp. 51–61.

[111] Thomas, C. & Marshall, E. “Arthropod abundance and diversity in differently vegetated margins of arable fields”. *Agriculture, Ecosystems and Environment* **72** (1999), pp. 131–144.

[112] Topping, C. J. & Lagisz, M. “Spatial dynamic factors affecting population-level risk assessment for a terrestrial arthropod: An agent-based modeling approach”. *Human and Ecological Risk Assessment: An International Journal* **18** (2012), pp. 168–180.

[113] Tscharntke, T. et al. “Conservation biological control and enemy diversity on a landscape scale”. *Biological Control* **43** (2007), pp. 294–309.

[114] Tuck, S. et al. “Land-use intensity and the effects of organic farming on biodiversity: a hierarchical meta-analysis”. *Journal of Applied Ecology* **51** (2014), pp. 746–755.

[115] Turchin, P. *Quantitative analysis of movement*. Sunderland, MA: Sinauer Associates, Inc., 1998.

[116] White, E., Ernest, S., Kerkoff, A. & Enquist, B. "Relationships between body size and abundance in ecology". *Trends in Ecology and Evolution* **22** (2007), pp. 323–330.

[117] Wissinger, S. A. "Cyclic colonization in predictably ephemeral habitats: a template for biological control in annual crop systems". *Biological Control* **10** (1997), pp. 4–15.

[118] Woltz, J. M., Isaacs, R. & Landis, D. A. "Landscape structure and habitat management differentially influence insect natural enemies in an agricultural landscape". *Agriculture, Ecosystems, and Environment* **152** (2012), pp. 40–49.

[119] Woodward, G., Speirs, D. C. & Hildrew, A. G. "Quantification and resolution of a complex, size-structured food web". *Advances in Ecological Research* **36** (2005), pp. 85–135.

[120] Wootton, K. et al. "Improving model predictions of foodweb dynamics: using mesocosm studies to link species traits with interaction strength". Unpublished Manuscript.

[121] Yodzis, P. & Innes, S. "Body size and consumer-resource dynamics". *American Naturalist* **139** (1992), pp. 1151–1175.

[122] Zhao, Z.-H., Shi, P.-J., Hui, C., Ge, F. & Li, B.-L. "Solving the pitfalls of pitfall trapping: a two-circle method for density estimation of ground-dwelling arthropods". *Methods in Ecology and Evolution* **4** (2013), pp. 865–871.