From theory to experimental design - quantifying a trait-based theory of predator-prey dynamics

A.N. Laubmeier[1*,◕], Kate Wootton[2,◕], J.E. Banks[3], Riccardo Bommarco[2], Alva Curtsdotter[2], Tomas Jonsson[2], Tomas Roslin[2], H.T. Banks[1]

1 Center for Research in Scientific Computation, North Carolina State University, Raleigh, NC, USA
2 Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden
3 Undergraduate Research Opportunities Center, California State University Monterey Bay, Seaside, CA, USA

◕These authors contributed equally to this work.
* anjlaubmeier@gmail.com

Abstract

Successfully applying theoretical models to natural communities and predicting ecosystem behavior under changing conditions is the backbone of predictive ecology. However, the experiments required to test these models are dictated by practical constraints, and models are often opportunistically validated against data for which they were never intended. Alternatively, we can inform and improve experimental design by an in-depth pre-experimental analysis of the model, generating experiments better targeted at testing the validity of a theory. Here, we describe this process for a specific experiment. Starting from food web ecological theory, we formulate a model and design an experiment to optimally test the validity of the theory, supplementing traditional design considerations with model analysis. The experiment itself will be run and described in a separate paper. The theory we test is that population dynamics are dictated by species traits, and we study this in a community of terrestrial arthropods. We depart from the Allometric Trophic Network (ATN) model and hypothesize that habitat use, as well as body mass, is necessary to better model trophic interactions. We therefore formulate new terms which account for habitat use as well as intra- and interspecific interference in the ATN model. We design an experiment and an effective sampling regime to validate this model and test the underlying assumptions about the traits dominating trophic interactions. We arrive at a detailed sampling protocol to maximize information content in the empirical data obtained from the experiment and, relying on theoretical analysis of the proposed model, explore potential shortcomings of our design. Consequently, since this is a "pre-experimental" exercise aimed at improving the links between hypothesis formulation, model construction, experimental design and data collection, we hasten to publish our findings before analyzing data from the actual experiment, thus setting the stage for strong inference.

Introduction

Quantifying theory – a challenging process

Models are abstractions of reality, distilling presumed relationships. In empirical science, model credibility is established by predicting outcomes observable in real systems. This is the ultimate test of a model: the better it holds up to tests which should easily prove it wrong, the more confidently we can conclude that it reveals something useful about nature and incorporates important mechanisms dictating system behavior. While this is a fundamental principle of modern science [1], the steps between general theory, formulation of an explicit model, and application of this model to predict the outcome of an empirical experiment are typically challenging. Ecological data tend to be both messy and expensive to obtain. Thus, by intimately linking the development of the model to the design of the experiment testing the model, we are able to ensure that we optimize the amount of useful information collected relative to the amount of work required to collect it. To obtain maximal data for resources invested, we need to cunningly design experiments and sample only as much as necessary. By gauging our approach in advance of performing the actual experiment, we obtain rigor in our approach, minimize sources of uncontrolled variation, and sample only as much as needed. Designing our experiments to target clear-cut predictions and estimating well-defined quantities also increases the conclusiveness of the test. Strong inference requires that experiments are designed specifically to test and distinguish among *a priori* hypotheses [2]. To encourage and support this distinction, some fields, such as psychology and neurology, have started promoting pre-registration of research [3–5]. In addition to encouraging clear *a priori* hypotheses, pre-registration has the advantage of enabling researchers to critically examine their hypotheses and plans before beginning their research and providing the chance to receive feedback and improve experimental design [3]. Here, we carefully perform the steps between ecological theory and experimental testing of that theory, before conducting the actual experiment. The experiment itself, following the design we arrive at here, will be presented in a subsequent publication.

A specific theory in need of testing

Predicting how ecological communities will respond to changes in abundance or composition is a high priority for ecologists and conservationists, requiring the use of models describing the communities and their driving mechanisms. However, it is well understood that "all models are wrong, but some models are useful" [6]. We consider a model useful if it accurately captures the dynamics it is supposed to describe along with a significant portion of the most important mechanisms. Developing such a model requires (i) creating the model with what are thought to be the most important mechanisms, based on the information available, and (ii) exploring that model to determine if it behaves as one would expect and gives robust results. This step includes asking how sensitive the model is to parameter values and whether or not the chosen model formulation is useful. The final step (iii) is testing the model performance and parameterizing the resultant model using empirical data. Establishing a useful predictive model usually requires a number of rounds of (i)-(iii).

We begin with the theory of why and how species interact to create an ecological community, and specifically that these interactions are dictated by species traits. To make accurate predictions about a particular ecological community requires an understanding of its underlying interaction structure and governing dynamics. Empirically obtaining this information is difficult, possibly requiring multiple observations of interactions between predator and prey to determine interaction frequencies as well as experiments to determine the impact of changes to species abundances or presence. There is, however, a growing body of research indicating that species interactions are influenced by their traits, and that these traits can be used, at least to some extent, to predict interaction strength and resulting population dynamics [7, 8]. For example, larger predators have higher metabolic demands and can capture larger prey, while prey can use defenses such as a shell or hard integument to protect themselves from predators [9–12]. These traits show promise in predicting community structure and dynamics as well as, eventually, how communities respond to changing environmental conditions and species compositions [12–14].

Predictions based on species traits can then be used to manage communities subject to environmental change or to manipulate communities to achieve a particular outcome, e.g. to manage for enhanced ecosystem services such as pest control by natural predators [15, 16]. Predators require traits that increase their ability to locate, capture, and consume their prey, while prey require traits which decrease their chances of being consumed [7]. It is, therefore, the match between predator and prey traits which determine the strength of the trophic interaction between them. Understanding and quantifying the nature of this relationship between traits and trophic interactions gives the two-fold advantage of understanding of the mechanisms behind the interactions as well as potentially reconstructing a food web with less empirical information. This allows for the development of models that predict community population dynamics.

The utility of traits in predicting food web dynamics lies in knowing which few, ideally easily measurable traits, are most important for governing interactions. Body size is a prime candidate as it directly relates to the metabolism (and thus ecophysiology) of a species [17] and determines a number of other characteristics and behavior, such as diet generality [18] and range of movement [19]. Additionally, body mass is well-studied as an important trait in governing food web dynamics [20–23] and the Allometric Trophic

Network (ATN) model in particular has been successfully applied to predict the outcome of trophic interactions in microcosm studies of terrestrial arthropods [8, 24]. We therefore begin with the ATN model as used in [8], which models intraspecific competition and can allow for intraguild predation, but which does not support explicit interspecific predator interference. The interplay of different predator species is particularly important in determining trophic interactions, as communities of predators can have both complementary and adversarial effects on one another. In order to capture the complex effect this might have on herbivore populations [25–28], we modify the ATN model to accommodate intraguild interactions beyond predation.

Although body size can explain a large proportion of interactions and is well-studied in a range of systems [8, 12, 29–31], in many cases it is either unimportant or acts primarily as a filter, such that two individuals can, but do not necessarily, have a trophic interaction if they are within a relatively broad size spectrum [32]. In these cases, other traits may be required to explain why an interaction occurs, and efforts have been made to inform models with additional traits [18, 22, 33]. The problem lies in determining which other traits to include. There is little understanding of how different traits might affect interaction strength and resulting population dynamics (but see [13, 15, 34]). Following the ideas explored in [28, 35], we assume that species habitat use is important in determining interactions and formulate terms which account for this trait in the ATN model. Differences in the use of habitat can mean that organisms rarely encounter each other and thus have a weaker than expected interaction.

The specific hypotheses we intend to test are: (i) in the absence of other species, consumption rates will be, to a large extent, determined by body size such that the impact of predators on prey, in simple two-species treatments, will increase with predator-prey body size ratio up to the optimal body size ratio. (ii) habitat overlap between predator and prey affects predation rates such that predators which spend more time in the same habitat as their prey will have a stronger impact on prey population. (iii) presence of other species affects predator behaviour such that predators which are more likely to be intraguild prey to another predator, due to body size ratio and shared habitat use, will spend more time avoiding the other predator and thus have a weaker impact on the shared prey in more complex (species-rich) treatments. (iv) a dynamical model parameterized using predator-prey body size ratios and habitat overlap data will explain more (but not all) of the variation in species abundances than a model based only on body size. To maximize the amount of useful information collected relative to the amount of work required to collect it, we examine how the information content of the data that will be collected depends on the timing and frequency with which it is collected. Below we describe the original model presented in the literature as well as why and how we have altered it to describe missing traits and mechanisms we believe are important. We then describe how we design an experiment to test this model, including choice of predators and prey and traits to include. We follow this with an investigation into how best to sample this experiment to obtain the information that we need to inform the model, using parameter sensitivities and model sensitivity to parameter estimation to determine timing, frequency, and subsampling techniques.

1 Model: ATN with predator avoidance

We introduce a variation on the ATN model used in [8]. This model already depends on body size and we introduce the parameter ν_{ij} to describe spatial overlap, with the hypothesis that species which spend more time in the same space will have a greater impact on each other. We assume that the only source of mortality for the duration of the experiment is predation. Dynamics for the number of individuals N_i of species i are therefore given by

$$\frac{dN_i}{dt} = r_i N_i - \sum_{j \in \mathcal{C}_i} \frac{a_{ij}\nu_{ij}N_iN_j}{1 + \sum_{k \in \mathcal{R}_j} a_{kj}\nu_{kj}h_{kj}N_k + \sum_{l \in \mathcal{C}_j} b_0 a_{jl}\nu_{jl}N_l}. \tag{1}$$

$\mathcal{C}_i$ is the set of species which consume species i, $\mathcal{R}_j$ 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 for the duration of our experiment). The parameters a_{ij} (attack rate), h_{ij} (handling time), and ν_{ij} (spatial overlap, or, similarity in habitat use) are parameterized by measurable traits, as described below.

We also introduce the term $\sum_{l \in \mathcal{C}_j} b_0 a_{jl}\nu_{jl}N_l$, describing the trade off that arises from some species balancing the competing requirements of being both predator and prey. In such cases, the time an organism spends avoiding potential predators will limit the time spent exploiting prey. We sum over the potential attack rates of all species l on a single individual of species j to account for time spent avoiding or evading species l while species j is attempting to capture its own prey. 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 term $c_j(N_j - 1)$ as used in [8] from this version of the ATN model.

As in [8], we assume that for species body masses W_i and W_j (corresponding to predator i and prey j), the allometric parameters are given by

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

We introduce the parameter ν_{ij} to account for the effect of habitat use on encounter rates. For the set Ω of all modes of habitat use, we define the probability measure (μ_i) such that evaluating $\mu_i(\mathcal{A})$ for some habitat $\mathcal{A}$ contained in Ω gives us the likelihood that species i occupies that habitat $\mathcal{A}$. For example, one $\mathcal{A}$ in Ω might be "residing in foliage" and an entirely ground-dwelling predator would therefore have a measure $\mu_i(\mathcal{A}) = 0$. That is, μ_i is a mathematical way to describe observed species habitat use. For species pair i and j, we define the overlap parameter

$$\nu_{ij} = 1 - v_0 TV(\mu_i, \mu_j) = 1 - v_0 \sup_{\mathcal{A} \subset \Omega} |\mu_i(\mathcal{A}) - \mu_j(\mathcal{A})|, \tag{3}$$

where $TV(\mu_i, \mu_j)$ is the total variation distance [36] and $0 \leq v_0 \leq 1$ is a scaling factor for the importance of habitat use in determining spatial overlap. Total variation distance quantifies the difference between two measures, so $TV(\mu_i, \mu_j)$ can be interpreted as the dissimilarity between the spatial distributions μ_i and μ_j of species i and j, and it takes on values between zero (no variation between μ_i and μ_j) and one (maximal variation between μ_i and μ_j). Continuing the example above, an entirely ground-dwelling predator and an entirely foliage-dwelling predator would have $TV(\mu_i, \mu_j) = 1$, since their spatial distributions are completely dissimilar. The corresponding overlap $\nu_{ij} = 1 - v_0$ is a quantification of the similarity of habitat use for the two species in our example, with v_0 indicating how much weight we give to the assumed distributions μ_i and μ_j. If $v_0 = 1$, then we give full consideration to the assumed distributions ($\nu_{ij} = 0$ and the predators do not overlap in our model). If $v_0 = 0$, we ignore the assumed distributions ($\nu_{ij} = 1$ and the predators completely overlap in our model). For a more detailed explanation of this formulation with examples from our experiment, see S1 Appendix.

Our primary goal in choosing traits to model is that they significantly affect population dynamics. However, we must weigh this goal against the model's general relevance for differing communities and environments outside the scope of our experiment, where we eventually hope to evaluate its performance [37]. We must also consider the tractability of our model; a model with too many parameters to feasibly identify or which fluctuates wildly for relatively small changes in model parameters cannot reliably be used to investigate questions about physical processes. Although our formulation of overlap is not driven by physical mechanisms, 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 Designing the experiment

2.1 Choosing the study system

To test this model requires a food web which is easily manipulated in the lab, where generation time is sufficiently brief to observe population dynamics over a short time period, and where predators and prey vary in the traits of interest, body mass and habitat overlap. By selecting both predators and prey which vary in these traits, we can additionally investigate the importance of a match between predator and prey traits.

We use modules from a simplified agricultural food web to test the model, which we manipulate in a replicated cage experiment in greenhouses (Fig 1). In addition to the reasons listed above, we choose this foodweb for its important applied relevance - crop pest control using natural predators - and because we have data from a previous experiment to use as a basis for the predictions outlined here. Furthermore, the prey items (aphids) reproduce rapidly, allowing us to determine the effect of predators on their population growth in a tractable period of time; for this reason, we limit the length of the experiment to 8 days so that aphid growth is neither limited by quality of remaining resources nor complicated by the production of alates.

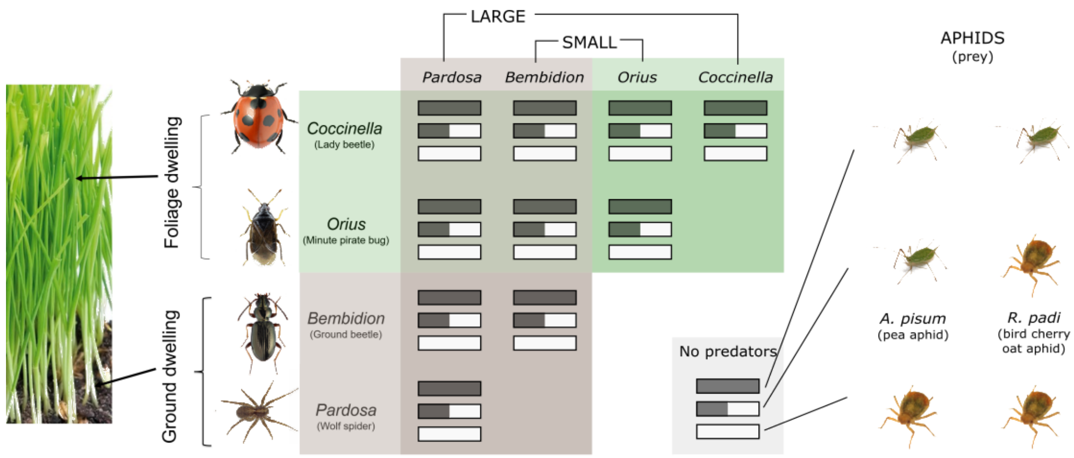

Fig 1. An overview of the experimental design and the hypotheses we test. We select two predators which reside primarily in the foliage and two on the ground. Of those two, one is large and one is small such that each combination of size and habitat use is covered. Each predator is combined with each other predator, including itself, to cover all one- and two-predator combinations. We also select two species of prey, and within each predator combination, each prey combination is represented (shown by horizontal bars of one or two colors). Thus, including controls containing no predators, we have 33 predator-prey combinations.

To determine the effect of body size and habitat overlap, we select predators varying 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*). While these predators primarily reside either in the foliage or on the ground, there is still substantial overlap and all predators can encounter one another. We note in preliminary trials that *Bembidion* is the most restricted in habitat use, while *Pardosa* sometimes climb the plants, and both of the foliage predators spend time on the ground. We choose phylogenetically diverse predators to determine whether the importance of body size and habitat use are valid despite phylogenetic differences.

We select a large and a small predator for both foliage and ground predators to determine the effect of body size (Fig 1). To also compare the importance of a match between predator and prey traits, we use two prey species (both aphids) differing in body size and habitat use. The larger aphid (*Acyrthosiphon pisum*) lives primarily on the underside of bean leaves, while the smaller aphid (*Rhopalosiphum padi*) grows on barley plants and resides low down on the stem where it is available to ground-dwelling predators. The choice of these two prey species necessitates the choice of barley and beans as plants.

Having a range of predator body sizes also allows us to explore intraguild predation and predator avoidance as a result of body size and habitat overlap. We hypothesize that the closer one predator is to the optimal predator-prey body mass ratio of a larger predator, the more likely that the smaller predator will be the victim of intraguild predation and thus spend time avoiding the larger predator instead of consuming aphids. Similarly, we hypothesize that with greater overlap in habitat use, smaller predators are more likely to become intraguild prey and thus spend time avoiding the intraguild predator. To test these hypotheses, we test all combinations of two predators with each of the prey species alone and with both prey species. Each combination of predators and prey is replicated six times in caged mesocosms in the greenhouse, with the number of replicates selected as a result of a previous experiment. Including controls for each of the prey treatments, this yields 33 different predator-prey combinations (Fig 1). We therefore explore how a predator's traits match with different prey as well as how the effect of predators on each other may change their impact on the prey.

2.2 Determining parameter sensitivities and critical sampling times

To effectively estimate parameters for a dynamic model, we must collect sufficient data throughout the experiment, but obtaining data at an appropriate temporal scale is 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 (from which the higher-level parameters a_{ij}, h_{ij}, and ν_{ij} can be computed as given in (2)). We will estimate model parameters using an "inverse problem" methodology, in which model dynamics are fit to time-series data using some predetermined cost function of the collected data. Data from different days can vary in how much information they provide for the solution of the inverse problem, and we want to focus our effort on those days which provide the most information. To ensure that maximal information is present in the collected data, we examine the sensitivity of model output (population densities) to parameter inputs as time progresses. 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 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 [38, 39] which are given in S2 Appendix. To facilitate comparison of sensitivities between treatments, we compute the relative sensitivity (similar to elasticity) [40, 41] $sr_i^\theta(t)$,

$$sr_i^\theta(t) = \frac{\theta}{N_i(t,\theta)} s_i^\theta(t), \tag{4}$$

and to avoid giving undue importance to aphid populations approaching zero, we do not normalize by the population when $N_i(t,\theta) < 1$. 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 cannot know the model parameters prior to running the experiment, and so our scaling parameters are informed by estimates from similar mesocosm experiments [24] and R_{opt} is estimated from personal observations of predators during planning stages of the experiment. However, 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 (no predators) cage "trial runs" during planning, we estimate intrinsic growth rates $r = 0.3007$ for *A. pisum* alone, $r = 0.3211$ for *R. padi*, and $r = 0.2453$, $r = 0.2591$ for *A. pisum* and *R. padi* (respectively) in a combined treatment. We plot in Fig 2 the sensitivities of aphid populations to model parameters for four cage treatments (described in Table 1), which exhibit the types of behavior we see in the sensitivities as a function of time.

Table 1. Key for cage treatments referenced in Fig 2.

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*

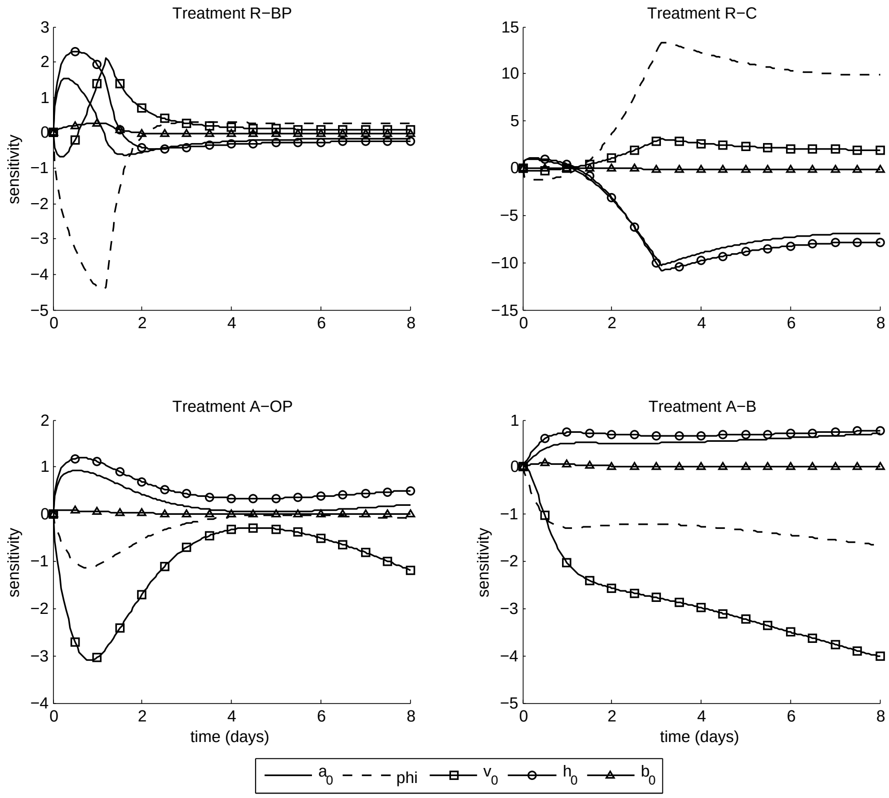

Fig 2. Sensitivities of aphid population abundances with respect to model parameters. Upper left: *R. padi* with predator treatment *Bembidion-Pardosa*. Upper right: *R. padi* with predator treatment *Coccinella*. Lower left: *A. pisum* with predator treatment *Orius-Pardosa*. Lower right: *A. pisum* with predator treatment *Bembidion*. Sensitivities are plotted in time and different parameters are indicated by line style.

The first type (represented by treatment R-BP) is the most common and characterized by high sensitivities on or before $t = 2$, which quickly decrease to low levels for the remainder of the experiment. Cages of this type correspond to aphid populations which the model predicts will be quickly decimated by the predator treatment and do not recover. The second type (represented by treatment R-C) is 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. In these cages, the modeled aphid population does experience eventual decimation, but the decline is sometimes slightly delayed. The third type (represented by treatment A-OP) exhibits high sensitivity early in the experiment, which decreases to low values around $t = 4$ and then increases slightly towards the end of the experiment. The modeled aphid population under these treatments experiences a drastic decline at the beginning of the experiment but manages a slight rebound when predators become scarce later in the experiment. The final type (represented by treatment A-B) has constant or increasing sensitivities for most of the experiment, with sensitivities increasing the most 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 note that our sensitivity results are local to the assumed parameters in model formulation and therefore may change for different parameter inputs. That is, if any parameters estimated from experimental data are significantly different from the values we assume above, the model sensitivities to *all* parameters may differ from what we present here. To minimize this risk, we obtain assumed parameter ranges using estimates 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, 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. We obtain less information with repeated sampling 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.

However, use of the above sampling scheme does not immediately imply that model parameters are realistically identifiable. 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 3, 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 [24] ($h_0^1 = 2/24, h_0^2 = 3/24$). 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. Suppose we were trying to determine the value of h_0 from population trajectories; 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. This problem is inherent to the model and we cannot necessarily anticipate where it may arise prior to parameter estimation. However, we can guard against this problem 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.

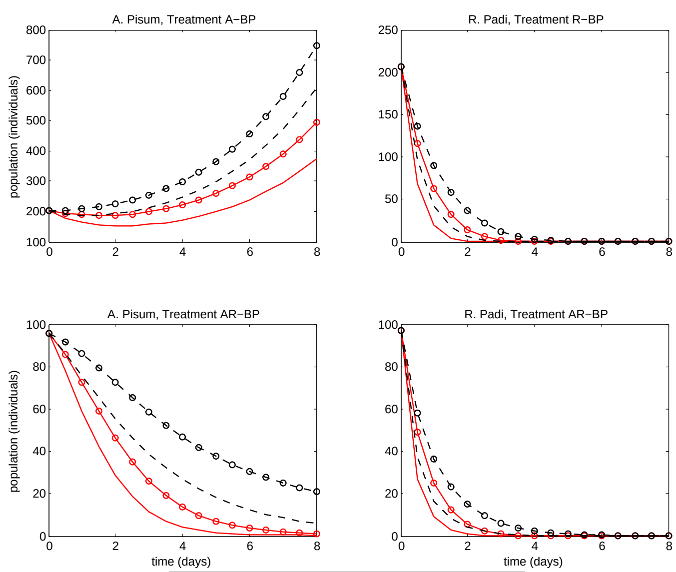

Fig 3. Model trajectories for aphid populations under predator treatment Bembidion-Pardosa. Upper left: *A. pisum* aphid treatment. Upper right: *R. padi* aphid treatment. Lower left: *A. pisum* in combined aphid treatment. Lower Right: *R. padi* in combined aphid treatment. We plot model solutions for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

2.3 Determining feasible sampling strategies for parameter estimation

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 [38, 39]. Although other methods exist for the estimation of model parameters (see, for example, examining interaction strengths as in [8]), solving a least squares inverse problem allows us to fit model dynamics to data in time and preserve maximal information about physical processes.

Convergence of a least squares parameter estimation requires substantial, temporal data. Obtaining this data is a time-consuming and expensive process and, with limited resources, becomes a trade-off between sampling thoroughly at a single time or sampling frequently in time. There is no consistent translation of this requirement to some number of actual data points required to obtain a sensible 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 (especially since we can only theorize that dynamics will be the same for all predators, despite phylogenetic differences). 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. This choice still presents an enormous effort, requiring many hours of work from lab assistants. We therefore consider the potential for subsampling of the aphid population in order to obtain data more quickly (with less effort) on these days; due to resource availability, it is crucial that we establish a subsampling protocol for the experiment.

To subsample the aphid population, N_j, at time t_j in a cage with T tillers (individual plants on which aphid populations are counted), we count the number of aphids on $n < T$ tillers, $\tilde{N}_j^n$, 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. In these cages, we assume strictly exponential growth of the aphids at some rate r_0, and we estimate the average growth rate $\overline{r}$ in six replicate cages to use as an approximation to r_0 in the ATN model. From the per-tiller counts in these six cages, we construct synthetic data for aphids with growth rate $\overline{r}$ (see S3 Appendix for details). We consider subsampling strategies in which $n = 5, 6, ..., 90$ tillers are counted and estimate the growth rate r^n from the approximated population N_j^n. We compute the resulting error $|r^n - \overline{r}|$ for all n, and repeat this process for 200 synthetic data sets.

We plot the average error $|r^n - \overline{r}|$ for each subsampling strategy in Fig 4. The dashed line in the figure indicates the average value of $|r - \overline{r}|$ 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 tillers, the accuracy gained from sampling additional tillers drops significantly around $n = 25$ tillers. Sampling around $n = 40$ tillers results in error that is on the same scale as the average value of $|r - \overline{r}|$. 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 tillers (or, 1-in-2 subsampling) yields sufficiently accurate parameter estimates. If the burden of data collection is too high, sampling 30 out of 90 tillers (1-in-3 subsampling) is justifiable, and sampling as high as 60 out of 90 tillers (2-in-3 subsampling) seems to be a very safe, if intensive, strategy.

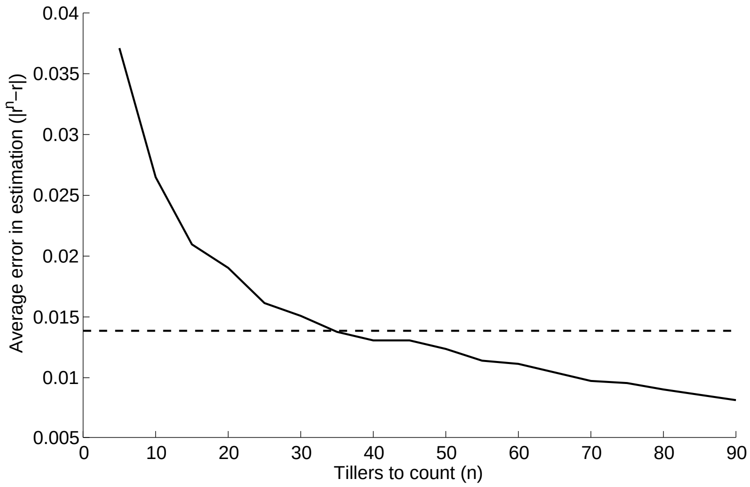

Fig 4. The average value of $|r^n - \bar{r}|$ across 200 synthetic cages. The horizontal dashed line indicates the average value of $|r - \bar{r}|$ for parameter estimates using the true data set.

We next consider the effect of error induced by subsampling on our estimates of ATN model parameters. We obtained per-tiller aphid counts for experimental cages with *R. padi* as the basal species, but sometimes under different predator treatments than our intended study communities [24]. 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 (see S3 Appendix for details) and add this noise to a simulation of cages which we assume perfectly follow ATN dynamics. We present in Fig 5 the scatter plots and histograms of the normalized error in aphid population counts induced by subsampling synthetic data for three categories of aphid densities, corresponding to full-cage counts in the ranges of $[0, 100]$, $[150, 1500]$, and $[4000, 9000]$. We only present results for the 1-in-3 subsampling rate here; at higher subsampling rates, the behavior of the error is similar.

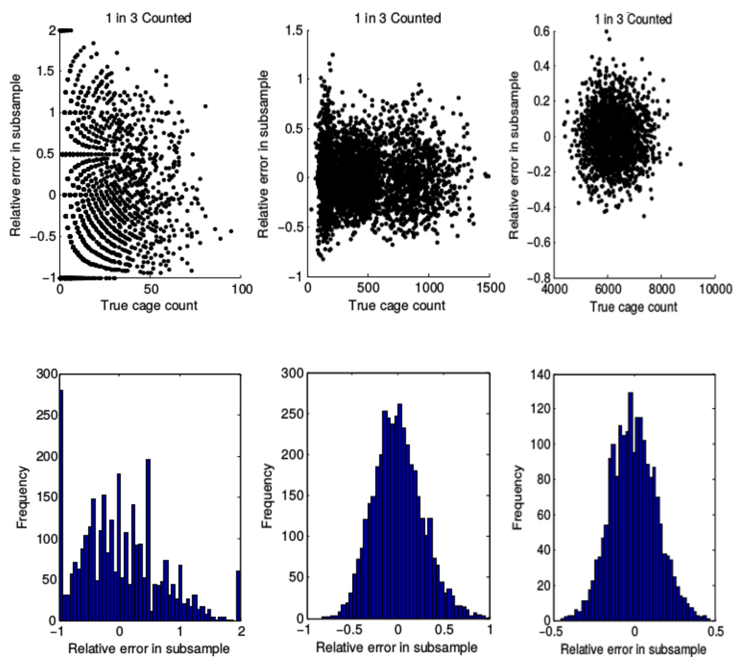

Fig 5. Normalized errors induced by subsampling at a 1-in-3 rate for synthetic data. Left: data in the ranges of $[0, 100]$. Middle: data in the ranges of $[150, 1500]$. Right: data in the ranges of $[4000, 9000]$.

We conclude that for full-cage populations greater than 150 aphids, the normalized error induced by subsampling n tillers 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. There is therefore a clear bias towards the upper and lower limits of the error induced by subsampling in this range. When generating synthetic data with errors sampled from the distributions in Fig 5 and dynamics given by ATN model solutions for use in an inverse problem, visual fits to the true model solution are poor for aphid populations below 150 (results not shown). 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.

An additional challenge in employing this method of parameter estimation is that the accuracy of results is dependent on the study system's adherence to both the *mathematical model of dynamics* and the *statistical model of error* in the data. 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 [39]; model-based analysis in this situation could be wrong without any indication of the underlying issue. To avoid this, the statistical model is often revised iteratively, with methods which either include an assumption of the underlying dynamics or which rely entirely on information present in the data, once experimental data is available (see [42, 43] for implementation of these methods). 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.

A previous cage experiment [24], 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) and only collected aphid abundance data on one additional day during the experiment. 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. Unfortunately, this means 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.

3 Discussion

We have outlined a theory to test, developed a model which distills that theory into a mathematical relationship, and designed experimental protocol to test that model, including the optimal sampling method to minimize potential difficulties in estimating model parameters. Taxa and treatments are chosen to address our trait-based hypotheses within the framework of our adjusted ATN model, and combined in the design laid out in Fig 1. 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. From the available information, we conclude that population data obtained by subsampling 2-in-3 or 1-in-3 tillers would be sufficiently accurate under most treatments. However, since this analysis is based on data from an experiment with different prey treatments than we use, we fully sample whenever possible. We note that our proposed sampling schemes are intrinsically linked to the assumed ATN model for species behavior and statistical model for error in the data; we cannot make general conclusions about the amount of data necessary for validation of a model with p parameters, or the frequency with which such data should be collected. Instead, sampling schemes must optimize the expected information content in the resulting data set, based on model dynamics and anticipated noise. By designing the experiment explicitly to test the model, we substantially increase our confidence that any variation the model does not capture is due to mechanisms we have missed or misrepresented in the model, as opposed to simply insufficient data or data collected at incorrect temporal resolutions. This can then form the basis for future studies investigating additional traits or mechanisms in the iterative process of refining both the model and experiments testing it [2]. By explicitly presenting our route through this process, we prevent the possibility of short-circuiting it by retrospectively creating hypotheses to explain our data [44].

We first explicitly link the experiment with the model, allowing a clear test of how the traits we seek to study affect food web dynamics. However, it is equally valuable to work in the opposite direction, formulating our model with the knowledge that we will be fitting it to an unrealistic mesocosm experiment. By measuring habitat use throughout the experiment and including it in our formulation of overlap, our model implicitly addresses the fact that species may be forced by the experimental design to share habitat. Our model therefore allows for results to be generalized to communities where species habitat use might be differently defined, which is amenable to our eventual goal of extending an experimentally validated model to naturally occurring communities.

In specifically formulating our model to address a trait-based hypothesis, we guarantee that each parameter's effect on model output is an expression of the importance of an ecological mechanism we seek to study. Solutions to time-varying sensitivity equations therefore reveal intervals over which population data is most-influenced by the mechanisms we have designed this experiment to study, and temporal sampling protocol must be informed by these results. Failure to obtain data on days with high sensitivities might lead to the conclusion that an experiment was not sufficiently thorough to determine the importance of a particular mechanism or, even more alarmingly, that the mechanism does not significantly affect dynamics. For our design, we note that sensitivity peaks across different two-day intervals for a variety of experimental treatments and identify treatments where it might be necessary to sample more frequently at the start of the experiment. By identifying critical sampling times for maximal information prior to running the experiment, we ultimately save time and resources and, most importantly, safeguard against the possibility of collecting data which is useless in model validation [45].

It is an unfortunate reality that unforeseeable circumstances will sometimes result in missing data where experimental results cannot be obtained, making model validation difficult. 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. In addition to identifying a temporal sampling protocol, it is therefore equally 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. Moreover, our quantification of the error induced by subsampling will be necessary in informing any uncertainty quantification for our estimated parameters, if subsampling is implemented. In adapting this hybrid subsampling protocol, we guarantee that the best possible data is collected whenever possible. In the event that we cannot fully sample, we collect imperfect population data with quantifiable error for all treatments; from a mathematical perspective, this is preferable to the alternative of collecting perfect data for a handful of treatments and obtaining no information on the remaining treatments.

The final step is to implement this experiment and use the data to estimate model parameters that yield a best-fitting solution to observed dynamics, which we will present in future work. However, in anticipation of this goal, we explore potential identifiability issues which, as with many ecological models, exist in our formulation of the ATN model. Although we have designed the experiment to maximize information content in the data, we demonstrate that it is necessary to investigate possible non-unique parameter solutions when fitting the model to this data and potentially augment our parameter estimation with some empirically measured parameters. The intent in fitting a model to data from an ecological community is not only to numerically identify a specific parameter, since the parameter value may not be generalizable to other communities. Instead, our goal must be to describe the important mechanisms driving dynamics. If different parameter sets can fit the data equally well but tell different stories about necessary traits in describing trophic interactions, then a failure to recognize these identifiability issues renders conclusions about underlying behaviors questionable.

The future of predictive ecology lies in the successful implementation of models to predict changes in natural communities. In order for these models to produce credible results, they must be based on and tested by empirical experiments. Moreover, these experiments must be optimally designed such that sufficient data can be extracted to inform the model [46]. By presenting this process before conducting the experiment, we have been forced, a priori, to explicitly establish our hypotheses and the rationale behind our experimental design [2, 44, 47]. This ensures that we have designed the experiment in a way which maximizes its relevance to the model. When validating the model with experimental data, we will be able to investigate the importance of modeled traits with maximal confidence and, if necessary, conclude that additional traits must be investigated to sufficiently describe observed behavior. The results of the experiment will inform our efforts to build a model for these dynamics which can be applied to different study systems, regardless of which of our hypotheses are correct.

Supporting information

S1 Appendix - Formulation of Habitat Use The parameter μ_{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 describe its formulation for species in our experiment, using preliminary observations of habitat use in the greenhouse cages. In Table 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 calculating the fraction of observations where the species was within each region out of all observations.

Table 2. Probability that species used in the cage experiments occupy various regions of the cage.

	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 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

Table 3. .

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. For this example, we would have $\Omega = \{P, N, G, PN, PG, NG, PNG\}$ 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 for 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:

$$\begin{aligned} &\mu_A(P) = 0.9, \quad \mu_A(N) = 0, \quad \mu_A(G) = 0.1 \\ &\mu_A(PN) = 0.9, \quad \mu_A(PG) = 1, \quad \mu_A(NG) = 0.1, \quad \mu_A(PNG) = 1 \end{aligned} \tag{5}$$

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

$$\begin{aligned} &\mu_B(P) = 0.05, \quad \mu_B(N) = 0.2, \quad \mu_B(G) = 0.75 \\ &\mu_A(PN) = 0.25, \quad \mu_B(PG) = 0.8, \quad \mu_B(NG) = 0.95, \quad \mu_B(PNG) = 1 \end{aligned} \tag{6}$$

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 $TV(\mu_i, \mu_j) = \sup_{\mathcal{A}\subset\Omega} |\mu_i(\mathcal{A}) - \mu_j(\mathcal{A})|$ and can be interpreted as dissimilarity between the distributions μ_i and μ_j. To compute this value, we take the difference between $\mu_i(\mathcal{A})$ and $\mu_j(\mathcal{A})$ for all sets $\mathcal{A}$ in Ω (such as P or G). We then find the "supremum" over the absolute value of all these differences; that is, we find the largest difference, regardless of sign. For the example we began above, we note that this largest difference occurs when we look at the sets P or NG, since

$$|\mu_A(P) - \mu_B(P)| = |\mu_A(NG) - \mu_B(NG)| = 0.85, \tag{7}$$

and this is the largest magnitude difference we see over all sets in Ω. Then, $TV(\mu_A, \mu_B) = 0.85$. The total variations for all species pairs in our system are given in Table 4.

Table 4. Total variation $TV(\mu_i, \mu_j)$ for species in the cage experiments.

	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

We calculate these values using the probability measures for habitat use defined in Table 3

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 i uses that region is very different from the probability that j uses that region. Because the values of $\mu_i(\mathcal{A})$ must fall between zero and one (by the definition of a "probability" measure), we will have that all differences between $\mu_i(\mathcal{A})$ and $\mu_j(\mathcal{A})$ fall between zero and one. That is, 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 $TV(\mu_A, \mu_A) = TV(\mu_B, \mu_B) = 0$, since a species will always match its own probability of occupying a given region. We finally note that we will always have $TV(\mu_i, \mu_j) = TV(\mu_j, \mu_i)$; because we take the absolute value of differences in computing the total variation, the order in which we subtract the values of $\mu_i(\mathcal{A})$ and $\mu_j(\mathcal{A})$ does not matter.

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 $0 \leq TV(\mu_i, \mu_j) \leq 1$, subtracting this quantity from 1 will give us a number that is *still* between 0 and 1. However, the new value is smaller for larger total variations (that is, closer to zero when there is high variation in habitat use) and larger for smaller total variations (closer to one for very similar habitat use). We introduce some scaling factor $0 \leq v_0 \leq 1$ to quantify the effect of assumed habitat use on spatial overlap. If $v_0 = 0$, we do not consider the total variation between spatial distributions at all in computing overlap; that is, we assume all species in the system are equally likely to encounter one another, regardless of their assumed spatial distributions. If $v_0 = 1$, we give full weight to the total variation between species distributions. The final formula for computing overlap, as given in the body of the manuscript, is

$$\nu_{ij} = 1 - v_0 TV(\mu_i, \mu_j). \tag{8}$$

S2 Appendix - ATN Sensitivity Equations We present the sensitivity equations for an ATN model of general size, assuming that $a_{ij} = 0$ whenever node j does not predate on node i. Despite this, we retain summations over sets $\mathcal{C}_i$ and $\mathcal{R}_j$ for clarity. We recall that for a system given by $\frac{d\mathbf{N}}{dt} = \mathbf{f}(t, \mathbf{N}, \theta)$, we obtain the sensitivities $\mathbf{s}^\theta = \frac{d\mathbf{N}}{d\theta}$ by solving the differential equation

$$\frac{d\mathbf{s}^\theta}{dt} = \frac{\partial \mathbf{f}}{\partial \mathbf{N}}\mathbf{s}^\theta + \frac{\partial \mathbf{f}}{\partial \theta}, \tag{9}$$

with initial condition $\mathbf{s}^\theta(0) = \mathbf{0}$. For $\mathbf{f}(t, \mathbf{N}, \theta)$ as given by the ATN model and any i, k, we have

$$\begin{aligned}
\frac{\partial f_i}{\partial N_k} &= r_i \frac{\partial N_i}{\partial N_k} - \sum_{j \in \mathcal{C}_i} a_{ij} \left(\frac{\partial N_i}{\partial N_k} N_j F_j + N_i \frac{\partial N_j}{\partial N_k} F_j + N_i N_j \frac{\partial F_j}{\partial N_k} \right), \\
\frac{\partial F_j}{\partial N_k} &= -F_j^2 \left(a_{kj} h_{kj} + b_0 a_{jk} \right).
\end{aligned} \tag{10}$$

The derivative of f_i with respect to θ is given by

$$\frac{\partial f_i}{\partial \theta} = -\sum_{j \in \mathcal{C}_i} N_i N_j \left(\frac{\partial a_{ij}}{\partial \theta} F_j + a_{ij} \frac{\partial F_j}{\partial \theta} \right), \tag{11}$$

where

$$\begin{aligned}
\frac{\partial F_j}{\partial \theta_a} &= -F_j^2 \left(\sum_{k \in \mathcal{R}_j} \frac{\partial a_{kj}}{\partial \theta_a} h_{kj} N_k + \sum_{h \in \mathcal{C}_j} b_0 \frac{\partial a_{jh}}{\partial \theta_a} N_h \right) \text{ for } \theta_a = a_0, R_{opt}, \phi, v_0, \\
\frac{\partial F_j}{\partial b_0} &= -F_j^2 \sum_{h \in \mathcal{C}_j} a_{jh} N_h, \\
\frac{\partial F_j}{\partial h_0} &= -F_j^2 \sum_{k \in \mathcal{R}_j} a_{kj} \frac{h_{kj}}{h_0} N_k, \\
\frac{\partial F_j}{\partial W_i} &= -F_j^2 \left(\sum_{k \in \mathcal{R}_j} \left(\frac{\partial a_{kj}}{\partial W_i} h_{kj} + a_{kj} \frac{\partial h_{kj}}{\partial W_i} \right) N_k + \sum_{h \in \mathcal{C}_j} b_0 \frac{\partial a_{jh}}{\partial W_i} N_h \right).
\end{aligned} \tag{12}$$

The derivatives of our allometric parameters with respect to a_0, R_{opt}, ϕ, W_i for $i \neq j$ are given by

$$\begin{aligned}
\frac{\partial a_{ij}}{\partial a_0} &= \frac{a_{ij}}{a_0}, \\
\frac{\partial a_{ij}}{\partial v_0} &= -a_0 W_i^{1/4} W_j^{1/4} \left(\frac{W_j/W_i}{R_{opt}} e^{1 - \frac{W_j/W_i}{R_{opt}}} \right)^{\phi} TV(\mu_i, \mu_j), \\
\frac{\partial a_{ij}}{\partial R_{opt}} &= a_{ij} \frac{\phi}{R_{opt}^2} \left(W_j/W_i - R_{opt} \right), \\
\frac{\partial a_{ij}}{\partial \phi} &= a_{ij} \ln \left(\frac{W_j/W_i}{R_{opt}} e^{1 - \frac{W_j/W_i}{R_{opt}}} \right) = a_{ij} \left(1 - \frac{W_j/W_i}{R_{opt}} + \ln \frac{W_j/W_i}{R_{opt}} \right), \\
\frac{\partial a_{ij}}{\partial W_i} &= \frac{a_{ij}}{4W_i} \left(1 - 4\phi + \frac{4\phi W_j}{R_{opt} W_i^2} \right), \\
\frac{\partial a_{ij}}{\partial W_j} &= \frac{a_{ij}}{4W_j} \left(1 + 4\phi - \frac{4\phi}{R_{opt} W_i} \right), \\
\frac{\partial a_{jj}}{\partial W_j} &= \frac{a_{jj}}{2W_j}, \\
\frac{\partial h_{ij}}{\partial W_i} &= \frac{1}{4} \frac{h_{ij}}{W_i}, \\
\frac{\partial h_{ij}}{\partial W_j} &= -\frac{1}{4} \frac{h_{ij}}{W_j}, \\
\frac{\partial h_{jj}}{\partial W_j} &= 0.
\end{aligned} \tag{13}$$

S3 Appendix - Expanded Subsampling Results 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. Because we only consider aphid population, 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 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,

$$r = \arg\min_{r \geq 0} \sum_{j=1}^{M} (N_j - f(t_j, r))^2. \tag{14}$$

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

$$r^n = \arg\min_{r \geq 0} \sum_{j=1}^{M} (N_j^n - f(t_j, r))^2. \tag{15}$$

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.

We begin with the control cage data for *R. padi* obtained in a prior microcosm experiment. The provided data includes the per-tiller aphid counts, y_j^k, for six cages, collected on days $t_j = 4, 8$. For each cage, we estimate the growth rate r from the fully sampled population counts, N_j. The resulting parameter estimates for the six cages are given in Table 5 and the model trajectories are plotted in S1 Fig.

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

Table 5. Estimated growth rates 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 $\lvert r - \overline{r}\rvert = 0.0138$.

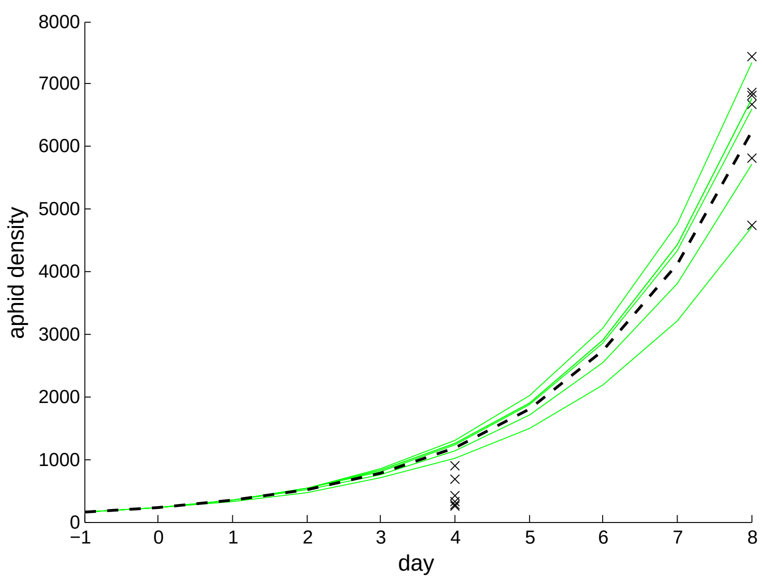

Fig 6. Fit of exponential model to control data using an OLS parameter estimation. Full-cage population counts are indicated with x-markers and the estimated exponential growth for each cage is plotted with a solid green line. Exponential growth at rate $\overline{r}$, the average across the six replicates, is plotted with a dashed black line.

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

$$\epsilon_j^k = \frac{1}{f(t_j, r)/T}\left(y_j^k - \frac{f(t_j, r)}{T}\right). \tag{16}$$

For $\overline{r}$ the average of the estimated growth rates across all six cages, we compute the true population curve $f(t, \overline{r}) = 150e^{\overline{r}(t+1)}$. We generate synthetic cages with $T = 90$ tillers 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 tiller from the set of all normalized errors ϵ_j^k for a fixed j.

For each synthetic cage, we consider subsampling strategies in which $n = 5, 6, ..., 90$ tillers are counted. For each strategy, we estimate the growth rate r^n from the approximated population N_j^n. We compute the resulting error $|r^n - \overline{r}|$ for all n, and repeat this process for 200 synthetic cages. We discuss the results of this in the main text, where we plot the average error $|r^n - \overline{r}|$ for each subsampling strategy in Fig 4.

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 tiller 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 tiller 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 tiller 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 tiller 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 tiller counts, $\{\hat{y}^k\}$, from the category. We compute the true population of the synthetic cage,

$$\hat{N} = \sum_{k=1}^{90} \hat{y}^k, \tag{17}$$

and a subsampled approximation to the true population,

$$\hat{N}^n = \sum_{k=1}^{n} \hat{y}^k, \tag{18}$$

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

$$\hat{\epsilon}^n = \frac{1}{\hat{N}}(\hat{N}^n - \hat{N}), \tag{19}$$

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 S2 Fig, S3 Fig, and S4 Fig, we present the scatter plots and histograms for the normalized errors induced by the three subsampling procedures for Categories 1, 2, and 3 respectively.

We note that the distribution of $\hat{\epsilon}^n$ is close to normal for cages in Categories 2 and 3. The variance of the distribution in a given category is higher at lower subsampling rates, and the variance of distributions in Category 2 for a given n are higher than for the corresponding distributions in Category 3. We therefore conclude that for full-cage populations greater than 150 aphids, the normalized error induced by subsampling n tillers is normally distributed, with variance decreasing as the aphid population increases.

As discussed in the main text, we cannot identify a distribution for the $\hat{\epsilon}^n$ for cages in Category 1. If there is a single aphid in a cage and we subsample n of T tillers, then the possible approximations to the population are $\hat{N}^n = T/n, 1$, or 0. Because of this, there is a clear bias towards the upper and lower limits of $\hat{\epsilon}^n$ when the aphid population is low, and we cannot choose a simple distribution to describe the error induced by subsampling in Category 1.

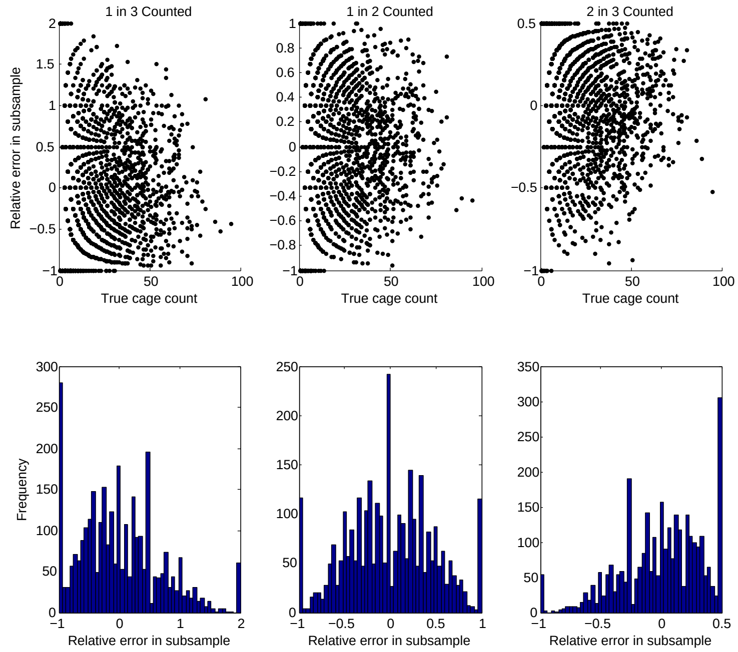

Fig 7. Normalized errors induced by subsampling ($\hat{\epsilon}^n$) for synthetic cages generated from tiller counts in Category 1. We take $n = 30, 45, 60$ (1-in-3, 1-in-2, and 2-in-3 subsampling strategies, respectively).

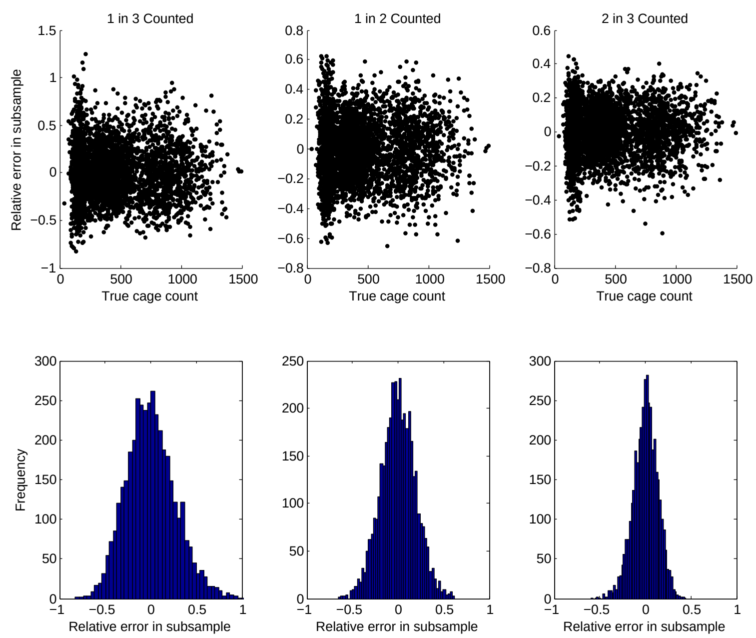

Fig 8. Normalized errors induced by subsampling ($\hat{\epsilon}^n$) for synthetic cages generated from tiller counts in Category 2. We take $n = 30, 45, 60$ (1-in-3, 1-in-2, and 2-in-3 subsampling strategies, respectively).

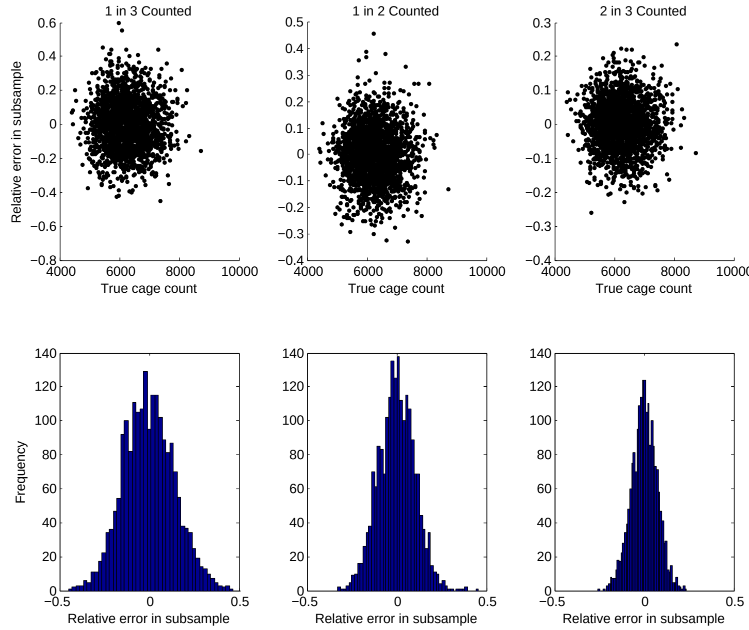

Fig 9. Normalized errors induced by subsampling ($\hat{\epsilon}^n$) for synthetic cages generated from tiller counts in Category 3. We take $n = 30, 45, 60$ (1-in-3, 1-in-2, and 2-in-3 subsampling strategies, respectively).

4 Acknowledgments

We thank Gerard Malsher for providing data from previous mesocosm experiments, as well as Riikka Kaartinen who designed the protocol for the same experiments. This research was supported in part by the National Science Foundation under Research Training Grant (RTG) DMS-1246991, by a CRSC/Lord Fellowship, by the Air Force Office of Scientific Research under grant number AFOSR FA9550-15-1-0298, by the August T. Larsson guest researchers programme at the Swedish University of Agricultural Sciences, and by the Swedish Research Council (VR) under grant number 2016-04580 and the Swedish Research Council FORMAS under grant number 2016-01168.

References

1. Lakatos I. Criticism and the Methodology of Scientific Research Programmes. Proceedings of the Aristotelian Society. 1968;69:149–186.

2. Platt J. Strong Inference. Science. 1964;146:347–353.

3. Gonzales JE, Cunningham CA. The promise of pre-registration in psychological research;. http://www.apa.org/science/about/psa/2015/08/pre-registration.aspx.

4. Allinson M. Royal Society Open Science launches Registered Reports; 2015. Available from: `https://blogs.royalsociety.org/publishing/registered-reports/`.

5. Pain E. Register your study as a new publication option; 2015. Available from: `http://www.sciencemag.org/careers/2015/12/register-your-study-new-publication-option`.

6. Box GE. Sampling and Bayes' inference in the advancement of learning. Wisconsin University - Madison Mathematics Research Center; 1979.

7. Boukal DS. Trait- and size-based descriptions of trophic links in freshwater food webs: current status and perspectives. Journal of Limnology. 2014;73:171–185.

8. Schneider FD, Scheu S, Brose U. Body mass constraints on feeding rates determine the consequences of predator loss. Ecology Letters. 2012;15:436–443. doi:10.1111/j.1461-0248.2012.01750.x.

9. Peckarsky BL, McIntosh AR, Àlvarez M, Moslemi JM. Disturbance legacies and nutrient limitation influence interactions between grazers and algae in high elevation streams. Ecosphere. 2015;6:1–15.

10. Woodward G, Speirs DC, Hildrew AG. Quantification and resolution of a complex, size-structured food web. Advances in Ecological Research. 2005;36:85–135.

11. Kalinoski RM, DeLong JP. Beyond body mass: how prey traits improve predictions of functional response parameters. Oecologia. 2016;180:543–550.

12. Brose U. Body-mass constraints on foraging behaviour determine population and food-web dynamics. Functional Ecology. 2010;24:28–34.

13. Eklöf A, Jacob U, Kopp J, Bosch J, Castro-Urgal R, Chacoff NP, et al. The dimensionality of ecological networks. Ecology Letters. 2013;16:577–583.

14. Rall BC, Vucic-Pestic O, Ehnes RB, Emmerson M, Brose U. Temperature, predator-prey interaction strength and population stability. Global Change Biology. 2010;16:2145–2157.

15. Rusch A, Birkhofer K, Bommarco R, Smith HG, Ekbom B. Predator body sizes and habitat preferences predict predation rates in an agroecosystem. Basic and Applied Ecology. 2015;16:250–259.

16. Tscharntke T, Bommarco R, Clough Y, Crist TO, Kleijn D, Rand TA, et al. Conservation biological control and enemy diversity on a landscape scale. Biological control. 2007;43:294–309.

17. Brown JH, Gillooly JF, Allen AP, Savage VM, West GB. Towards a metabolic theory of ecology. Ecology. 2004;85:1771–1789.

18. Gilljam D, Thierry A, Edwards FK, Figueroa D, Ibbotson AT, Jones JI, et al. Seeing double: size-based and taxonomic views of food web structure. Advances in Ecological Research. 2011;45:67–133.

19. Ottaviani D, Cairns SC, Oliverio M, Boitani L. Body mass as a predictive variable of home-range size among Italian mammals and birds. Journal of Zoology. 2006;269(3):317–330.

20. Brose U, Williams RJ, Martinez ND. Allometric scaling enhances stability in complex foodwebs. Ecology Letters. 2006;9:1228–1236.

21. Berlow EL, Dunne JA, Martinez ND, Stark PB, Williams RJ, Brose U, et al. Simple prediction of interaction strengths in complex food webs. Proceedings of the National Academy of Sciences of the United States of America. 2009;106:187–191.

22. Rall BC, Kalinkat G, Vucic-Pestic O, Brose U. Taxonomic versus allometric constraints on non-linear interaction strengths. Oikos. 2011;120:483–492.

23. Boit A, Martinez ND, Williams RJ, Gaedke U. Mechanistic theory and modelling of complex food-web dynamics in Lake Constance. Ecology Letters. 2012;15:594–602.

24. Kaartinen R, Jonsson T, Jonsson M, Bommarco R. Intraguild interactions modify the effect of body size on trophic interactions;.

25. Polis GA, Myers CA, Holt RD. The ecology and evolution of intraguild predation: potential competitors that eat each other. Annual Review of Ecology and Systematics. 1989;20:297–330.

26. Rosenheim JA, Wilhoit LR, Armer CA. Influence of intraguild predation among generalist insect predators on the suppression of an herbivore population. Oecologia. 1993;96:439–449.

27. Snyder WE, Snyder GB, Finke DL, Straub CS. Predator biodiversity strengthens herbivore suppression. Ecology Letters. 2006;9:789–796.

28. Schmitz OJ. Predator diversity and trophic interactions. Ecology. 2007;88:2415–2426.

29. Jonsson T, Cohen JE, Carpenter SR. Food webs, body size, and species abundance in ecological community description. Advances in ecological research. 2005;36:1–84.

30. Deraison H, Badenhausser I, Loeuille N, Scherber C, Gross N. Functional trait diversity across trophic levels determines herbivore impact on plant community biomass. Ecology Letters. 2015;18:1346–1355.

31. Cohen. Body sizes of hosts and parasitoids in individual feeding relationships. PNAS. 2005;102:684–689.

32. Klecka J, Boukal DS. Foraging and vulnerability traits modify predator-prey body mass allometry: freshwater macroinvertebrates as a case study. Journal of Animal Ecology. 2013;82:1031–1041.

33. Brose U, Ehnes RB, Rall BC, Vucic-Pestic O, Berlow EL, Scheu S. Foraging theory predicts predator-prey energy fluxes. Journal of Animal Ecology. 2008;.

34. Klecka J, Boukal DS. Foraging and vulnerability traits modify predator-prey body mass allometry: freshwater macroinvertebrates as a case study. Journal of Animal Ecology. 2013;82:1031–1041.

35. Schmitz OJ, Krivan V, Ovadia O. Trophic cascades: the primacy of trait-mediated indirect interactions. Ecology Letters. 2004;7.

36. Durrett R. Probability: Theory and Examples. 3rd ed. Belmond, California: Brooks and Cole; 2005.

37. Rykiel Jr EJ. Testing ecological models: the meaning of validation. Ecological Modelling. 1996;90:229–244.

38. Banks HT, Tran HT. Mathematical and Experimental Modeling of Physical and Biological Processes. New York: CRC Press; 2009.

39. Banks HT, Hu S, Thompson WC. Modeling and Inverse Problems in the Presence of Uncertainty. Boca Raton: CRC Press; 2014.

40. Eslami M. Theory of Sensitivity in Dynamic Systems: An Introduction. Springer-Verlag, Berlin; 1994.

41. Frank PM. Introduction to System Sensitivity Theory. New York: Academic Press; 1978.

42. Banks HT, 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. 2016;24:413–433.

43. Banks HT, Everett RA, Hu S, Murad N, Tran HT. Mathematical and statistical model misspecifications in modelling immune response in renal transplant recipients. Inverse Problems in Science and Engineering. 2017;.

44. Kerr NL. HARKing: Hypothesizing After the Results are Known. Personality and Social Pscyhology Review. 1998;2:196–217.

45. Blanchet FG, Legendre P, He F. A new cost-effective approach to survey ecological communities. Oikos. 2016;125:975–987.

46. Caswell H. The Validation Problem. In: Patten BC, editor. Systems Analysis and Simulation in Ecology. 4th ed. New York, New York: Academic Press; 1975. p. 601.

47. Bosco FA, Aguinis H, Field JG, Pierce CA, Dalton DR. HARKing's threat to organizational research: evidence from primary and meta-analytic sources. Personnel Psychology. 2010;69:709–750.