Running Head: Predator effects on prey density

Title: Evaluating consumptive and nonconsumptive predator effects on prey density using field times series data

J.A. Marino, Jr.[1-4], S.D. Peacor[3], D.B. Bunnell[5], H.A. Vanderploeg[6], S.A. Pothoven[7], A.K. Elgin[7], J.R. Bence[3], J. Jiao[3], and E.L. Ionides[4]

[1]Corresponding Author: jmarino@fsmail.bradley.edu, (309) 677-2352

[2]Department of Biology, Bradley University, 101 Olin Hall, 1501 W. Bradley Ave., Peoria, IL 61625.

[3]Department of Fisheries and Wildlife, Michigan State University, Natural Resources Building 480 Wilson Road, Room 13, East Lansing, Michigan 48824. Peacor email: peacor@msu.edu, Bence email: bence@msu.edu, Jiao email: jiaojin1@msu.edu

[4]Department of Statistics, University of Michigan, 311 West Hall, 1085 South University, Ann Arbor, MI 48109. Ionides email: ionides@umich.edu

[5]Great Lakes Science Center, U.S. Geological Survey, 1451 Green Road, Ann Arbor, MI 48105. Bunnell email: dbunnell@usgs.gov

[6] Great Lakes Environmental Research Laboratory, National Oceanic and Atmospheric Administration, 4840 S. State Rd., Ann Arbor, MI 48108. Vanderploeg email: henry.vanderploeg@noaa.gov

[7]Lake Michigan Field Station, Great Lakes Environmental Research Laboratory, National Oceanic and Atmospheric Administration, 1431 Beach St., Muskegon, MI 49441. Pothoven email: steve.pothoven@noaa.gov, Elgin email: ashley.elgin@noaa.gov

Abstract

Determining the degree to which predation affects prey abundance in natural communities constitutes a key goal of ecological research. Predators can affect prey through both consumptive effects (CEs) and nonconsumptive effects (NCEs), although the contributions of each mechanism to the density of prey populations remain largely hypothetical in most systems. Common statistical methods applied to time series data cannot elucidate the mechanisms responsible for hypothesized predator effects on prey density (e.g., differentiate CEs from NCEs), nor provide parameters for predictive models. State space models (SSMs) applied to time series data offer a way to meet these goals. Here, we employ SSMs to assess effects of an invasive predatory zooplankter, *Bythotrephes longimanus*, on an important prey species, *Daphnia mendotae*, in Lake Michigan. We fit mechanistic models in a SSM framework to seasonal time series (1994-2012) using a recently developed, maximum likelihood-based optimization method, iterated filtering, which can overcome challenges in ecological data (e.g. nonlinearities, measurement error, and irregular sampling intervals). Our results indicate that *B. longimanus* strongly influences *D. mendotae* dynamics, with mean annual peak densities of *B. longimanus* observed in Lake Michigan estimated to cause a 61% reduction in *D. mendotae* population growth rate and a 59% reduction in peak biomass density. Further, the observed *B. longimanus* effect is most consistent with an NCE via reduced birth rates. The SSM approach also provided estimates for key biological parameters (e.g., demographic rates) and the contribution of dynamic stochasticity and measurement error. Our study therefore provides evidence derived directly from survey data that the invasive zooplankter *B. longimanus* is affecting zooplankton demographics and offer parameter estimates needed to inform predictive models that explore the effect of *B. longimanus* under different scenarios such as climate change.

Keywords

Daphnia mendotae, *Bythotrephes longimanus*, nonconsumptive effects, iterated filtering, predator-prey interaction, Laurentian Great Lakes

Introduction

Quantification of the effects of predators on prey abundance is important for understanding ecological systems. Experiments in the field and laboratory can offer insights into potential mechanisms through which predators affect prey, but translating experimental measurements to field-relevant effects is challenging. For instance, in addition to consumption (i.e., consumptive effects, CEs), short-term experimental and observational studies suggest that nonconsumptive effects (NCEs) of predators can strongly affect prey density (Nelson et al. 2004, Matassa and Trussell 2011). However, the realized importance of NCEs in natural systems has recently been called into question (discussed in Kimbro et al. 2017), and the relative contributions of CEs and NCEs to large-scale, long-term prey density patterns remain largely unknown.

Existing field time series data may contain valuable information regarding the influence of predators on prey abundance at field-relevant spatial and temporal scales. In effect, analyzing consecutive points in time series with variable predator and prey abundances might offer information about how each is affecting the other as a function of hypothesized mechanisms. Challenges exist, however, to extract this information. Ecological systems are complex, e.g., due to nonlinearities and stochasticity, and the collection of ecological data is subject to measurement error and other constraints, such as irregular sampling intervals (Turchin and Taylor 1992,

Bjornstad and Grenfell 2001, Scheffer et al. 2001). Further, potentially confounding factors (e.g., seasonality, density dependence) can be difficult to disentangle from predator effects. Fortunately, recent methodological advancements can confront these challenges and provide insights into the contribution of different hypothesized mechanisms (Breto et al. 2009, Ionides et al. 2015). Specifically, mechanistic models of population dynamics can be implemented as state space models (SSMs, also known as partially observed Markov process models or hidden Markov models). SSMs include both a process model representing the true population dynamics and a measurement model representing the generation of the data (Newman et al. 2014). By explicitly accounting for these sources of variation, SSMs allow for testing of mechanistic hypotheses using time series data.

There are extensive time series data collected at multiple trophic levels in the Laurentian Great Lakes for management purposes, and applying SSMs to these data could be useful to address major questions, such as the impact of invasive species. A recent invader to the Great Lakes believed to be having a major impact on the zooplankton community is the large predatory cladoceran, *Bythotrephes longimanus*. For example, *Daphnia retrocurva* and *D. pulicaria*, declined rapidly in Lake Michigan after the introduction of *B. longimanus* in 1986 (Lehman and Caceres 1993, Barbiero and Tuchman 2004). Recent experimental and modeling research suggest that *B. longimanus* could further be affecting the abundance and spatial distribution of current dominant zooplankton species in the Great Lakes. Such effects are of potential importance to fisheries management, because *B. longimanus* effects on zooplankton density and position may reduce food availability for common prey fishes, with potential impacts on growth and recruitment. In turn, effects on prey fishes may affect key fisheries, such as Chinook salmon, that depend on those planktivores (Jacobs et al. 2013, Bunnell et al. 2015).

Simulation and statistical modeling as well as experimental research suggest that *B. longimanus* influences the composition and density of mesozooplankton through both CEs and NCEs. *B. longimanus* is known to prey on zooplankton (Vanderploeg et al. 1993) and bioenergetics models indicate planktivory by *B. longimanus* can be substantial (Bunnell et al. 2011). NCEs are hypothesized to occur when zooplankton prey perceive *B. longimanus* through chemical cues and adopt anti-predatory behavior in response to higher *B. longimanus* densities by migrating to lower depths (Pangle and Peacor 2006, Bourdeau et al. 2011), which reduces predation risk but at the cost of reduced growth rate and reproduction due to colder water at lower depths (Pangle et al. 2007). Previous research has estimated CEs and NCEs on zooplankton population growth rates (Pangle et al. 2007). Consumptive rates measured in the laboratory can be used to estimate consumptive rates the field. NCEs can be estimated from known temperature dependent effects on zooplankton birth rate and field measurements of the effect of *B. longimanus* on zooplankton position (and hence the temperatures that those zooplankton experience). Results yield an estimate of the relative magnitude of NCEs and CEs on demographic rates, and thus serve to highlight potential influence of NCEs through simulations. However, this approach cannot determine if *B. longimanus* is actually affecting the density of zooplankton in the field; e.g., there could be feedback mechanisms or indirect effects which would offset the predicted negative effects. Therefore, while we can predict mechanisms by which *B. longimanus* affects zooplankton population growth rate (e.g., as in Pangle et al. 2007), evaluating the extent to which *B. longimanus* affects zooplankton prey density in the field is a major challenge and could benefit from methods that allow for inference directly from field density data. This problem is not unique to the Great Lakes zooplankton system, as we are aware of many studies that examine the influence of NCEs on prey demographic rates in the field (e.g.,

Peckarsky et al. 2008, Kimbro et al. 2017), but few that examine if NCEs are affecting prey density directly from prey density patterns.

Herein our approach is to use SSMs to test the hypothesis that *B. longimanus* influences the density of an important zooplankton species, *Daphnia mendotae*, in the field through CEs and NCEs. We focus on *D. mendotae* because it composes a relatively high biomass among cladocerans in the community (Vanderploeg et al. 2012) and is consumed by planktivorous fishes (Bunnell et al. 2015). Multiple population models of *D. mendotae*, with different functional dependence on its predator, *B. longimanus*, were implemented as SSMs and fit to time series data via a recently developed, maximum likelihood-based optimization method, iterated filtering. Iterated filtering can fit nonlinear, non-Gaussian, non-stationary SSMs to data and handle complexities associated with ecological data like irregular sampling intervals (Ionides et al. 2006, 2015). Such complexities are intrinsic to complex ecological systems and field survey data, including those available for the Great Lakes. Iterated filtering algorithms are distinguished from other state space model methodology by providing statistically efficient, simulation-based, maximum likelihood inference for general nonlinear state space models (Ionides et al., 2015). Our approach should allow us to estimate key biological rates (e.g., birth and death rates) and the magnitude of predator effects, as well as the contribution of stochasticity to dynamics and the influence of measurement error on variation in the data, which are important to account for in order to successfully address our hypothesis.

We had two goals: 1) Evaluate if, and to what extent, *B. longimanus* affects *D. mendotae* density and, if so, whether such effects are more consistent with CEs or NCEs. 2) Estimate key parameters (e.g., birth and predation rates) needed to model this system, which will be valuable in the future to predict dynamics under different scenarios (e.g., climate change effects).

Methods

Data description

D. mendotae and *B. longimanus* biomass density data were collected as part of a long-term survey of Lake Michigan zooplankton by the NOAA Great Lakes Environmental Research Laboratory (GLERL) at an offshore site near Muskegon, MI (depth = 110 m; 43° 11.99', 086° 34.19'; located about 20 km offshore). The survey quantified the biomass density of crustacean zooplankton 7-16 times per year across 16 years (1994-2003, 2007-2012) using whole water column vertical net tows (details on sampling and biomass density calculations presented in Vanderploeg et al. 2012).

General process model of population dynamics

The process model represents dynamics of *D. mendotae* using a stochastic, seasonally-forced variant of a logistic population growth model. The state variable is *D. mendotae* biomass density, *V* (i.e., the prey zooplankton), and dynamics are represented by the following stochastic differential equation with respect to time, *t*:

$$dV = \left(V\,\beta(t)\left(1-\frac{V}{\kappa}\right)\left(1-\,\eta g(P)\right) - f(V)P - \mu V\right)dt + V\epsilon dW + \rho(t) \qquad (1),$$

where $\beta(t)$ is a function representing prey birth and/or somatic growth rate at low population size, and κ is a prey density dependence term (here affecting prey birth/somatic growth rate). The term $\eta g(P)$ determines the nonconsumptive effect of *B. longimanus* on *D. mendotae* via a proportional reduction in birth rate, with *P* representing *B. longimanus* biomass density treated as a covariate (not dynamically modeled). The functional response $f(V)$ determines the consumptive effect, and μ is the background mortality rate of *D. mendotae* not due to

consumption by *B. longimanus*. The NCE and CE of *B. longimanus* are described in more detail below (see: *Consumptive and nonconsumptive predator effects*). The $V\epsilon dW$ term allows for random variation to occur in *D. mendotae* dynamics (i.e., process error), which can occur due to factors influencing growth rates not specified in the model, such as variation in weather. The standard deviation ϵ scales the process error dW, and this process variation is driven by Brownian motion:

$$dW \sim Normal\,(mean = 0, sd = \sqrt{dt}\,) \quad (2),$$

which is a common way to represent stochasticity in dynamic population models (Panik 2017). The term $\rho(t)$ represents the initiation of *D. mendotae* dynamics each year via emergence from resting eggs. Briefly, $\rho(t)$ is modeled as a pulse that only contributed to the population on the first day of each year's dynamics and is equal to zero on other days (see *Initiation of dynamics each year* for more detail).

Seasonality in prey birth rate

We modeled seasonality in *D. mendotae* birth rate given known strong seasonality in abundance due to factors such as temperature, light levels, and resources that affect birth rate using the equation:

$$\beta(t) = exp\left\{\sum_{i=1}^{N_s} \lambda_i s_i(t)\right\} \quad (3),$$

where $\{s_i(t), i = 1, ..., N_s\}$ is a periodic cubic B-spline basis with 4 bases ($N_s = 4$), a degree of 3, and a period of 1 year; $\{\lambda_i, i = 1, ..., N_s\}$ are parameters that specify the seasonality of the birth rate.

$\beta(t)$ is intended to capture *D. mendotae* seasonality using a function allowing enough flexibility to capture dynamics while avoiding overly complicating the model (i.e., adding unnecessary parameters). A periodic b-spline with $N_s > 3$ provides a more flexible representation of seasonal forcing compared to a sinusoidal, which has been used to represent seasonality in biological parameters. Tests that we performed using $N_s > 4$ suggested that additional parameters result in worse model performance based the Akaike Information Criterion (AIC), a measure of model quality, than $N_s = 4$. Eq. 3 therefore provides a reasonable representation of the seasonality in *D. mendotae* dynamics.

Consumptive and nonconsumptive predator effects

For the CE, we used a Type I functional response, $f(V) = \alpha V$, where α is *B. longimanus* attack rate on *D. mendotae,* as an approximately linear response is expected at the *D. mendotae* densities found in the survey according to laboratory predation experiments (Pangle and Peacor, unpublished data). We also evaluated an alternative version of the model with a Type II functional response (see: *Evaluation of Type II Functional Response).*

Nonconsumptive effects of *B. longimanus* on *D. mendotae* birth rate are represented by the proportion reduction in birth rate ($\eta g(P)$) according to the equation for $g(P)$:

$$g(P) = 7.601 + \ln(P + 0.0005) \quad (4)$$

We used a logarithmic function based on the log-linear relationship of the behavioral (i.e. vertical migration) response of *D. mendotae* to *B. longimanus* density (e.g., Bourdeau et al. 2015) that leads to an expected reduction in birth rate due to the colder temperatures in deeper water. A correction term (0.0005) was used to account for zero observations equivalent to ½ the smallest observation of *B. longimanus*. The equation for $g(P)$ includes the negative natural log of

the correction term ($-\ln(0.0005) = 7.601$) to be consistent with a reduction in birth rate (i.e., to eliminate the potential for a positive effect of *B. longimanus* biomass density on population growth at low *B. longimanus* densities).

The effects of *B. longimanus* were modeled as forcing functions in which the potential dynamic feedbacks to *B. longimanus* density are not included in the model for two reasons. First, there are likely other factors that affect *B. longimanus* density, including other prey items (e.g., copepods, *Bosmina longirostris*, and other *B. longimanus*), predation by fish, and physical factors (e.g., variable water currents, temperature) (Keeler et al. 2015). Second, treating *B. longimanus* as a state variable would require a substantial increase in the complexity of the model due to the potentially large number of additional parameters needed to model *B. longimanus* dynamics. Adding such additional complexity would substantially increase the challenge of fitting the model, due to having to estimate multiple additional parameters with a limited number of available data points (n = 134).

To reduce the influence of measurement error on estimates for *B. longimanus* (note: the measurement error model in Eq. 7 and 8 below applies only for the *D. mendotae* state variable), which could influence our estimates for predator effects, smoothing was performed by calculating a moving average for *B. longimanus*, *P*. We used a 45-day window for the moving average, which we expected should minimize information lost while reducing the influence of measurement error. This window was chosen because the mean gap between observations (excluding gaps between years) was 21 days, so that the value for the moving average on each day was typically influenced by 2-3 observations. We expected that a shorter window for the moving average would be insufficient given the mean time gap between observations, while a longer window could smooth over too much potentially informative variation in *B. longimanus*

given the typical generation time of *B. longimanus* (7-15 days, Kim and Yan 2010). Further, tests using a longer (e.g., 59-day) and shorter (e.g., 7-day) window for the moving average resulted in worse fits based on maximum likelihood estimates than the 45-day window. Similar tests comparing different durations have been used in other systems to establish the appropriate window for assessing impacts of other important covariates, such as climatic factors (van de Pol et al. 2016). Further, tests we performed using alternative methods of interpolation and smoothing (i.e., ln(+0.0005) transformation of *B. longimanus* data prior to calculation of a moving average or using a moving 45-day median) did not offer improvement in model performance based on AIC, and did not substantially affect our results.

The calculation of the moving average for *B. longimanus* biomass density involved two steps. First, daily estimates of biomass density were interpolated linearly between observations for gaps between observations, with the exception of the gap between the last observation each year and the first observation of the subsequent year. Interpolation is necessary, as the model represents continuous-time dynamics, so that a value for each covariate is required at each time step. The gap between years was treated differently because data were rarely collected during winter and early spring, and *B. longimanus* is typically absent from the water column at that time, while the population is maintained as resting eggs. We therefore assumed that *B. longimanus* was absent for the first 50 days each year (i.e., we set *B. longimanus* biomass density to 0 for those days), prior to the interpolation.

Second, these interpolated values (P_{int}) were then used to calculate a 45-day geometric mean (P). The correction term (0.0005, as for Eq. 4) was used to calculate the geometric mean to account for the presence of 0s in the *B. longimanus* data (otherwise the mean would be 0 for any

time points with a 0 in the 45-day moving average window). The *P* covariate for each time (t) was thus:

$$P(t) = \left(\prod_{i=1}^{45} P_{int}(t - 23 + i) + 0.0005 \right)^{1/45} - 0.0005 \quad (5)$$

Initiation of dynamics each year

Because *D. mendotae* are effectively absent from the water column in winter, we allowed the population in the water column to go extinct each winter and be reseeded via a pulse ($\rho(t)$) representing the emergence from resting eggs each spring occurring 7 days prior to the earliest observation of *D. mendotae* in the data. The size of the pulse is not well understood. In fact, it is plausible that the abundance of neonates emerging from resting eggs is not strongly dependent on the previous year's density given that resting eggs can survive for multiple years (Caceres 1998) and strong variation occurs in physical processes that promote hatching (Kerfoot et al. 2004). We therefore assumed the size of the pulse was random and log-normally distributed:

$$\ln(\rho(t)) \sim \text{Normal}\ (\phi, \psi) \quad (6)$$

φ and ψ represent the mean and standard deviation of the natural log of the pulse, respectively.

Measurement model

A measurement model is used to describe how observations (i.e., the data, which are subject to measurement error) were generated from the prey biomass state variable, which represents the true biomass density; therefore, the observed data are treated as drawn from a distribution around the true state of the system. Measurement error in this sense is general, including any differences between samples collected on different days not attributable to changes

in the true biomass density (e.g., due to differences between two net tows due to small-scale spatial variation or potential short-term fluctuations due to water currents or responses to variation in light levels that could affect individual measurements). We used a left-censored normal ($\text{Normal}_{\text{l-cens}}$) distribution (e.g., Martinez-Bakker et al. 2015, in which the probability of a zero value is treated as a point mass equal to the censored left tail of the normal distribution). Two parameters (σ_a and σ_b) are specified so that the variance (σ^2) scales quadradically with population size:

$$Vobs_{(t)} \sim \text{Normal}_{\,l-cens.}(V_t, \sigma) \tag{7}$$

$$\sigma \sim \sqrt{\sigma_a^2 V_{(t)} + \sigma_b^2 V_{(t)}^2} \tag{8}$$

We used a left-censored distribution to account for zero observations in the data and because negative observations cannot occur. The left-censored model assumes that the observed biomass density at any time point is normally distributed around the true biomass density, with a standard deviation that scales with population size according to Eq. 8, except the left-censored model does not allow observations of negative biomass density.

Model modifications to assess dynamic drivers

To examine the influence of *B. longimanus*, we fit four versions of the model to the data: model a) a null model (i.e., excluding any *B. longimanus* effect by fixing α and η at 0); model b) a model including only the NCE (i.e., fixing α at 0); model c) a model including only the CE (i.e., fixing η at 0); and model d) a model including both the CE and NCE.

Benchmark Statistical Models

A reasonable mechanistic model should perform better than a simple, non-mechanistic benchmark model (King et al. 2008). We therefore compared our mechanistic models to two straightforward benchmark models. First, we used a model assuming observed *D. mendotae* biomass density is independently and identically distributed around a seasonal (monthly) average (model e):

$$Vobs_{(t)} \sim \text{Normal}_{\,l-cens.}(\mathrm{D}_m, \sigma) \quad (9)$$

$$\sigma \sim \sqrt{\sigma_a^2 \mathrm{D}_m + \sigma_b^2 D_m^2} \quad (10)$$

D_m represents mean biomass densities for each month that observations were made, and observations are assumed to follow a left-censored normal distribution, as for models a-d (although model e does not differentiate between measurement and process error). Second, we fit an AR (2) autoregressive model with measurement error to our time series (model f), in which the observed *D. mendotae* biomass density depends linearly on the previous two observations. We used the same measurement model (Eq. 7 and 8) for model f as for models a-d, so as to allow for zero but no negative observations.

Model fitting

Analyses were implemented using the pomp package in R v.3.3.3 (R Core Team 2018), and code is included in the supplement. SSMs (including all models except model e, which was fit using the R optim function) were fit to time series data using iterated filtering via the mif2 algorithm, which is a recently developed algorithm for estimating model parameters via maximum likelihood estimation that offers substantial improvement over other SSM fitting methods (Ionides et al. 2015, King et al. 2016). For each model fit using iterated filtering, we performed 100 runs in which a search through parameter space was initiated using a random set

of starting values for each parameter. Starting values were generated from a uniform distribution bounded by broad plausible values for each parameter. The fit of different models was compared based on the Akaike Information Criterion (AIC) calculated using the maximum likelihood estimate, which provides a measure of model performance that weighs both model complexity based on the number of parameters and fit based on the likelihood (Akaike 1974). A difference of 2 AIC units indicates a substantial improvement in model performance (Burnham and Anderson 2002).

Magnitude of B. longimanus effect

To quantify effects of *B. longimanus* on *D. mendotae* biomass density, we used simulations from the fitted model (model b, the best model based on AIC, see results). We compared biomass densities of *D. mendotae* in 10,000 simulated 1-year data sets including or excluding the effect of *B. longimanus* by setting η to the maximum likelihood estimated value or 0, while all other parameters were fixed at their maximum likelihood estimated values. The simulations used an across-year seasonal mean of smoothed *B. longimanus* biomass density for predator biomass density. We note that these simulations necessarily do not reflect the full range of actual variation in the system (e.g., due to uncertainty in parameter estimates) but provide a straightforward way to quantify and visualize reductions in *D. mendotae* biomass density caused by estimated effects of *B. longimanus*.

Parameter estimates and confidence intervals

To gain further insight into the influence of *B. longimanus* and density dependence on dynamics, we developed confidence intervals for the model estimates of the NCE (η) and density

dependence (κ) parameters using profile likelihood (Hilborn and Mangel 1997). In profile likelihood, the likelihood is maximized and all other parameters are estimated across a fixed plausible range of values of the focal parameter (i.e., η or κ in our case). The result is a profile that shows how the maximum likelihood changes depending on that focal parameter value. The 95% confidence intervals are determined as the range of parameter values for which the log-likelihood is within 1.92 units of the maximum log-likelihood (Hilborn and Mangel 1997).

Evaluation of potential influence of seasonality

We were concerned that seasonality may confound results for two reasons. First, because *B. longimanus* and *D. mendotae* densities vary seasonally, we were concerned that a detected effect of *B. longimanus* was actually due to other seasonal factors that covary with *B. longimanus* but are not included in the model. Second, the NCE in the model is part of an expression that includes a seasonality term ($\beta(t)$), but the CE is part of an expression without seasonality, so that a difference in the influence of the NCE and CE could potential be influenced by the difference in their relationship with seasonality in the model. We therefore performed three additional analyses to examine the influence of seasonality.

First, we wanted to compare the performance of our model using *B. longimanus* as the predator to another species that we would not expect to affect *D. mendotae*. We therefore examined the fit of the best performing model (model b, see Results) substituting the biomass density data for another species, *Limnocalanus macrurus*, as an alternative predator instead of *B. longimanus* (model g). As *L. macrurus* mostly occurs in the hypolimnion and would have limited spatial overlap with *D. mendotae*, we would not expect it to have a detectable effect on *D. mendotae*. However, *L. macrurus* also exhibits strong seasonality in its dynamics (Vanderploeg

et al. 2012), so that treating it in the same manner as *B. longimanus* (i.e., as a predator) in the model provides a useful comparison to evaluate if seasonality itself could be responsible for any detected predatory effect of *B. longimanus*. A test using *L. macrurus* thereby directly addresses whether the methods would have identified a spurious relationship for this particular species.

Second, we calculated a *B. longimanus* biomass density anomaly (deviations from the average seasonal trend across years, i.e., with the seasonal trend removed) and compared how the model performed when using the anomaly compared to the null model (model h; see supplement for details). Because the anomaly excluded the seasonal trend, we would expect that including the anomaly should substantially improve the model AIC over a null model if there is an effect of *B. longimanus* distinct from a seasonal effect.

Third, we examined two additional models to address alternative hypotheses for how seasonality influences *D. mendotae* dynamics: model i) a modified version of the null model (model a) that includes seasonal background mortality, μ; and model j) a modified version of the model with only CEs (model c) that allows seasonal change in *B. longimanus* attack rate, α. In both models, each parameter was allowed to vary seasonally using periodic b-splines in the same manner as birth rate (β) (Eq. 3). We performed these analyses to ensure that our finding of an NCE of *B. longimanus* (see Results) could not be explained by seasonality in background mortality or *B. longimanus* consumption.

Evaluation of Type II Functional Response

In addition, to ensure that our results did not depend on the choice of functional response used in our model, we modified model c to include a Type II functional response for *f(V)*:

$$f(V) = \frac{\alpha V}{1 + \alpha h V} \quad (11),$$

where h represents *B. longimanus* handling time for *D. mendotae* (model k).

Results

The mechanistic SSMs performed substantially better than the benchmark models based on AIC (Table 1).

The models including the NCE of *B. longimanus* on *D. mendotae* outperformed the alternative models based on a comparison of AIC values. In contrast, including the CE did not improve the model performance either in the absence or inclusion of the NCE. Only the model with both the CE and NCE was within 2 AIC units of the best fit model that included the NCE but not the CE (model b). Because the former model included an additional parameter and offered no improvement over the latter model, we moved forward with model b as the best model.

To visualize the fit of the best model, we generated 10,000 simulated data sets (including the contribution of both process and measurement errors) from the fitted model using the parameter values at the maximum likelihood estimate (Table 2). Quantiles of the resulting simulations are shown to represent the median and 95% simulation intervals (Fig. 1). The clear seasonality of the simulation median suggests strong, predictable seasonality of *D. mendotae* dynamics. In contrast, differences between years are subtler and less predictable. The relatively broad 95% simulation intervals reflect relatively high levels of variation among simulations, attributable to dynamic stochasticity and measurement error. All but four observations fall within the simulation intervals, with the two most notable exceptions being the especially high peaks in the *D. mendotae* data in 2011 and 2012. In these years, *B. longimanus* had especially high density earlier in the season, for which the model would predict lower *D. mendotae* densities than observed those years.

The maximum-likelihood parameter estimates indicate *B. longimanus* can have a profound influence on *D. mendotae* density. Based on the fitted model estimate for η, *D. mendotae* birth rates are reduced by 61% at the mean peak *B. longimanus* across years (Fig. 2a). Simulations from the model generated using the maximum-likelihood estimate compared to simulations generated using the same values for other parameters but excluding the effect of *B. longimanus* (i.e., setting η equal to 0) suggests that the nonconsumptive effect on population growth rate results in as large as a 59% reduction in *D. mendotae* biomass density (difference between height of peaks in Fig. 2b). The likelihood profile for η reveals our level of confidence in our parameter estimate (Fig. 3a, showing 95% confidence intervals). Using the lowest and highest value of eta (at confidence interval bounds), at the mean annual peak of *B. longimanus*, the NCE ranges from a 28% to 82% reduction in growth rate.

The fitted SSM also provides estimates for the contribution of seasonality to *D. mendotae* dynamics. The fitted seasonal function for *D. mendotae* birth rates suggests a peak on Julian day 229 (August 16) in late summer. In the presence of *B. longimanus* at its mean biomass density, the peak both shifts in timing (10 days earlier to Julian day 219) and is reduced due to the NCE (Fig. 2a).

Density dependence also influences *D. mendotae* dynamics, based on parameter estimate and its confidence interval (Table 2, Fig. 3b). The parameter estimate for κ (33 mg x m^{-3}) was within the range of observed *D. mendotae* biomass density (0- 74 mg x m^{-3}), with 6 observations of *D. mendotae* biomass density exceeding the estimated value for κ, suggesting that high conspecific densities may almost entirely suppress positive *D. mendotae* growth under realized conditions in Lake Michigan.

Other parameter estimates provide insights into the contribution of measurement error and process stochasticity. Based on Eq. 7 and 8, the estimates for σ_a and σ_b indicate that the standard deviation of observed biomass at mean *D. mendotae* biomass was approximately 40% of mean, indicating a substantial impact of measurement error. The estimate for the standard deviation of *D. mendotae* growth rate (ϵ) is also large (126% of the maximum seasonal growth rate when at low population size, $\beta(t)$), suggesting the importance of process stochasticity as well. Both process stochasticity and measurement error thus contribute to the high levels of variation in the data (Fig. 1).

Evaluation of potential influence of seasonality

The three tests indicate that the result that *B. longimanus* affected *D. mendotae* through an NCE was not confounded by seasonality. First, using *L. macrurus* biomass density as the predator (model g) had the opposite effect than using *B. longimanus* as it performed worse than the model with no predator effect (model a) based on AIC (Table 1). Second, using the *B. longimanus* anomaly (model h) substantially improved the model fit compared to the model without effects of *B. longimanus*, despite the removal of the across-year seasonal trend, thereby providing further evidence for an effect of *B. longimanus* independent of seasonal factors. If the observed effect of *B. longimanus* was due to other seasonal confounding factors, no improvement would be expected by only using the anomaly. Notably, however, the model using the anomaly did not perform as well as the model using the actual *B. longimanus* biomass density data (model b), suggesting both anomalous and seasonal variation in *B. longimanus* contribute to *D. mendotae* dynamics. Third, if our detection of the NCE was caused by a confounding factor associated with the seasonal nature of the birth rate term, we would expect

that adding seasonality to the mortality or attack rate (models i or j) would have a similar influence to including the NCE. However, models i and j performed substantially worse than model b (Table 1), supporting the importance of the NCE.

Evaluation of Type II Functional Response

Finally, tests using an alternative (Type II) functional response (model k) revealed that our findings were not sensitive to the assumed functional response for the CE.

Discussion

Our analysis provides evidence that *B. longimanus* has strong negative effects on *D. mendotae* population growth rate and density in offshore Lake Michigan and supports the hypothesis that an NCE is the underlying mechanism. Further, our analysis quantifies key demographic rates for *D. mendotae*, including birth and death rates, which can be used in models that forecast the effects of future changes, such as climate change or changes in nutrient concentrations, with implications for overall Lake Michigan food web dynamics and fisheries. Our results demonstrate the utility of developing SSMs and fitting them to field time series data to assess mechanisms by which predators affect prey, despite the challenges intrinsic to ecological systems and data.

Our findings provide evidence of and, for the first time to our knowledge, quantify NCEs derived from field-based time series data in a mechanistic framework. The observed negative effect of *B. longimanus* on *D. mendotae* population growth rate resulted from an NCE in which *B. longimanus* reduced *D. mendotae* birth/somatic growth rates. Of the mechanistic models compared, the model including NCEs but not CEs provided the best fit relative to the number of parameters based on AIC, and greatly reduced AIC relative to the addition of CEs alone.

Whereas, NCEs have received considerable attention, most studies have been performed in a laboratory setting, mesocosms, and enclosures. Further, whereas there is an increasing number of studies performed in the field, very few studies examine the influence on density based on field data (Sheriff et al. in review). For example, previous studies evaluating NCEs of *B. longimanus* on *D. mendotae* (Pangle et al. 2007, Bourdeau et al. 2013) combined laboratory studies that elucidate the behavioral response of *D. mendotae* to *B. longimanus* with field survey data of *D. mendotae* vertical position at different densities of *B. longimanus*. Using temperature-dependent growth models, these studies predicted a large reduction in fitness of *D. mendotae* due to lower temperatures experienced at the lower depths occupied as a result of the anti-predator response to *B. longimanus*. Similarly, other studies that have examined NCEs in the field, have, for example, combined knowledge of predation rates and induced changes in prey behavior to explain hypothesized nonconsumptive effects on spatial variation in prey abundance (e.g., wolf avoidance by elk in Yellowstone, Creel et al. 2005, shark avoidance by marine vertebrates, Heithaus et al. 2009). Our approach to documenting NCEs from field data here is qualitatively different, in that evidence was derived directly from changes in density of prey in relation to changes in predator density, linked through mechanistic models.

We examined the time series data, and the model fits, to interpret why the inclusion of the NCE in the model leads to a large improvement in model performance, but adding the CE does not. Importantly, because *D. mendotae* birth rates peak earlier than peak *B. longimanus* density, the NCE exerts its major influence earlier than when CE effects are maximized. Thus, the model estimates the strongest *B. longimanus* effects in years when *B. longimanus* biomass density reaches high levels early, when *D. mendotae* birth rates would otherwise be high. This contrasts with a CE, which as modeled in Eq. 1 increases mortality the same amount whenever *B.*

longimanus density is high, regardless of time of year. This aspect of the NCE is seen in the temporal patterns in the data. For example, we can calculate a 45-day moving average of *D. mendotae* biomass density ($D_{avg}(t)$) as we did for *B. longimanus* (Eq. 5, using a modified correction factor equal to one half the lowest observation for *D. mendotae*) and then estimate the rate of *D. mendotae* population change (r_{est}) early in the growing season (days 175-225) each year:

$$r_{est} = \ln(D_{avg}(225)/\, D_{avg}(175)) \quad (12)$$

Consistent with the NCE detected by the model, the rate of *D. mendotae* population change between days 175 and 225 was negatively related to *B. longimanus* biomass density during that same period (geometric mean of smoothed *B. longimanus* biomass density + 0.0005 over days 175-225) in the same year (Fig. 4). While it is impossible to entirely rule out that consumption of *D. mendotae* by *B. longimanus* partly contributed to this pattern, model performance including only the CE was substantially poorer than the NCE model, even when we relaxed the assumption of a fixed attack rate by allowing it to vary seasonally (model j). The NCE therefore provides the most parsimonious explanation.

The large magnitude of the estimated effects of *B. longimanus* on *D. mendotae* biomass density here likely have important consequences for the Lake Michigan food web and are also likely relevant for the other four Great Lakes where *B. longimanus* and *D. mendotae* co-occur. For example, planktivorous fishes in Lakes Michigan and Huron have undergone declines in biomass since the 1990s, and these fish are key prey to Chinook salmon *Oncorhynchus tshawytscha* and lake trout *Salvelinus namaycush* that are the foundation of a multi-million dollar recreational fishery (Bunnell et al. 2014). Given that survival of larval planktivorous fish in the first few weeks of life can depend on overlap with zooplankton prey (Beaugrand et al. 2003),

513 understanding the mechanisms that regulate zooplankton densities is critical to improved
514 understanding and prediction of planktivorous fish recruitment. Our model estimates of *D.*
515 *mendotae* vital rates can also be applied to future decision-support tools that explore how future
516 climate or nutrient concentrations (perhaps modeled through modifications to carrying capacity,
517 κ) would affect the dynamics of *D. mendotae*, the most important herbivorous cladoceran in
518 terms of biomass (Vanderploeg et al. 2012).

519 Perhaps surprisingly, including CEs of *B.* longimanus did not substantially improve
520 model fit either alone or in combination with nonconsumptive effects. Experiments demonstrate
521 that *B. longimanus* predation rates on *D. mendotae* can be high (Vanderploeg et al. 1993, Pangle
522 and Peacor 2009), and thus one might expect high CEs in the field. Migration in response to *B.*
523 *longimanus* chemical cues (Pangle et al 2006) could be expected to reduce *B. longimanus*
524 consumption, although some studies still show spatial overlap between *B. longimanus* and *D.*
525 *mendotae* for at least a portion of the *D. mendotae* population (Bourdeau et al. 2015, Nowicki et
526 al. 2017). Nevertheless, we found little evidence for a substantial effect of consumption here.
527 One possible explanation is that our model for *B. longimanus* predation (i.e., Type I functional
528 response) may exclude key biological realism; for example, explicitly incorporating potentially
529 critical covariates that can influence predation rates, such as light levels (Pangle and Peacor
530 2009) and temperature (Yurista et al. 2010), could be explored in future models and may allow
531 for improved estimation of CEs.

532 Distinguishing between CEs and NCEs from observational data, as we have done here,
533 depends on assumed functional relationships. However, an advantage of SSMs is that
534 assumptions are made explicit in the equations and can be further tested in future work or
535 compared to experimental findings. For instance, a key difference between how CEs and NCEs

536 are modeled here is that we assume that the NCE affects birth rate or somatic growth rate, which
537 we model with a seasonal functional form, given known seasonal effects of temperature and food
538 resources on birth rate. Thus, the per capita NCE of *B. longimanus*, (ηg(P)), varies seasonally in
539 magnitude in proportion to *D. mendotae* birth rate as modeled, unlike the CE, which contributes
540 additively to mortality (i.e., proportional to *B. longimanus*). These different functional forms
541 thereby allowed us to at least partially differentiate between a CE and an NCE. Evidence for the
542 latter was then strengthened by additional tests under different assumptions (e.g., allowing
543 seasonal variation in consumptive effects in model j) and comparisons to prior work that also
544 suggest the importance of NCEs (e.g., Pangle and Peacor 2006).

545 Fish predation is also an important consideration for *D. mendotae-B. longimanus*
546 dynamics, although we do not expect fish effects to confound our results. In fact, *B. longimanus*
547 is susceptible to fish predation from alewife (*Alosa pseudoharengus*) and other species (Bunnell
548 et al. 2015), and so more *B. longimanus* may be associated with overall lower fish predation on
549 zooplankton. That we saw declines in *D. mendotae* biomass density associated with higher *B.*
550 *longimanus* despite potentially reduced risk from planktivorous fish at these times thus provides
551 further support that effects of *B. longimanus* are important for *D. mendotae* dynamics, and that
552 *B. longimanus* may be an important competitor with fish for zooplankton prey.

553 Another concern with analyses of field data relevant to our study is disentangling the
554 influence of seasonality from other dynamical drivers, such as the effects of *B. longimanus*. We
555 chose a flexible approach to incorporate seasonality in the system (periodic b-splines), and the
556 additional tests we performed (i.e., using *L. macrurus*, the anomaly, or allowing other terms to
557 vary seasonally) offered further support that other seasonal factors were not responsible for the
558 observed effect of *B. longimanus*. Similar rigorous tests should be a broadly useful approach to

559 disentangle seasonality from other drivers in many systems using SSMs. By using these tests, our
560 approach here was conservative in attempting to rule out a confounding effect of seasonality; in
561 fact, beyond the NCE we detected, it is plausible that *B. longimanus* effects on *D. mendotae* may
562 also actually contribute to the estimated effect of seasonal forcing. We may therefore be
563 underestimating a CE or an NCE if they are attributed to and therefore subsumed by the seasonal
564 model terms; explicitly considering some seasonal factors (e.g., temperature, resources) in future
565 may allow better resolution of these effects. In particular, future models including additional data
566 for spatial variation in *D. mendotae*, *B. longimanus*, resources, and temperature may allow better
567 resolution of the relative contribution of seasonality, CEs, and NCEs, as water column structure
568 likely plays an important role in mediating *B. longimanus* effects.
569 Our approach was also useful to quantify the influence of other drivers of *D. mendotae*
570 dynamics, including seasonality and density dependence. Model results reflect how *D. mendotae*
571 birth rates and biomass density change with Julian day (Fig. 3), likely due to seasonal variation
572 in temperature, food resources, water column structure, or other factors. Similarly, the estimated
573 density dependence term (κ) and its confidence interval indicate that *D. mendotae* population
574 growth is substantially density dependent under field conditions in Lake Michigan, potentially
575 due to competition for food resources. Further, estimates of density dependence will be vital for
576 predicting impacts of ongoing changes in the lower food web (Fahnenstiel et al. 2010). Our
577 findings thus motivate future work to investigate the underlying mechanisms driving seasonality
578 and density dependence and implications to other parts of the food web.
579 Our findings also provide estimates for the substantial contribution of both measurement
580 error (i.e., variation introduced during measurement) and process error (i.e., uncertainty in the
581 actual dynamics that cannot be explained by the deterministic components of the current model)

582 to variation in the data. Estimates of these sources of variation are critical to quantify uncertainty
583 for prediction of ecological dynamics and design sampling efforts (e.g., frequency of sampling
584 within and across years) to maximize the information gained. Explicit inclusion of measurement
585 error (represented by σ in Eq. 7 and 8) and process error (here in both birth rate represented by ϵ
586 and the seasonal pulse represented by Ψ) allowed us to quantify the amount of variation among
587 observations that is attributable to these sources of error. Simulations illustrate that, based on our
588 model, process and measurement variation can lead to a wide range of possible observed values
589 under the conditions of any given year. Although incorporating additional covariates or added
590 realism into the model in future may offer some reduction in the breadth of the simulation
591 intervals, much of this uncertainty may be irreducible given available information. Nevertheless,
592 our results indicate that the data contain important information about predictable changes in the
593 dynamics of the populations, such as the effects of *B. longimanus,* seasonal forcing, and density
594 dependence.

595 The models fit to time series here are relatively simple and yet have provided new
596 insights into interactions among zooplankton in Lake Michigan. Nevertheless, additional realism
597 could likely improve model fit (e.g., better capture the outlier observations in 2011 and 2012)
598 and the strength of inferences gained from the model. For instance, our models only included one
599 prey species, while future models may attempt to incorporate multiple prey species
600 simultaneously and potential interspecific competition or apparent competition mediated by *B.*
601 *longimanus*. Our ability to distinguish between increasingly complex models is limited by
602 available data (i.e., number of observations and years included), although continuing data
603 collection may allow for inference using more complex models. Future work should endeavor to
604 examine the limits to our SSM fitting approach to provide insights under different limitations

that are at play in this and many other systems (e.g., sampling frequency, number of data points, levels of measurement error). Additional data collected as a part of the NOAA GLERL Long Term Research program should also provide the opportunity to confirm estimated effects here and test additional drivers of dynamics.

Our application of mechanistic models here thus demonstrates how SSMs can provide useful insights into classic questions in ecology, such as the contribution of predators and other drivers to animal population dynamics, which for many systems remains largely hypothetical. In some cases, time series analysis of field data may be the only approach to address such questions at the relevant spatial scale. Fitting of models to data, as we have done here, allows for more direct tests of such fundamental ecological questions in spite of the complex factors involved, including nonlinearities, measurement error, seasonal forcing, and irregular measurement (Bjornstad and Grenfell 2001), which are seldom considered simultaneously. Our findings thus demonstrate the utility of using SSMs and provide a framework for advancing ecological understanding in a mechanistic framework. Further, our results provide novel and valuable example of quantifying NCEs over long timescales at a field scale, providing further evidence for their importance in ecological systems.

Finally, the insights gained from testing these hypotheses are vital to understanding and predicting consequences of ongoing large-scale environmental changes, such as the ecosystem-scale shifts caused by invasive species in the Great Lakes. In light of the suite of challenges facing key natural resources globally, advancing understanding of mechanisms for invasive species impacts in the field represents an important step forward.

Acknowledgments

We thank Dao Nguyen for assistance with analysis and Craig Stow and Aaron King for helpful discussion. We thank the National Oceanic and Atmospheric Administration Great Lakes Environmental Research Laboratory for sharing data. This work was supported by a National Science Foundation Postdoctoral Research Fellowship in Biology to JAM (DBI-1401837), Michigan DNR funding to JRB, and the Great Lakes Fishery Commission (44066). This is NOAA GLERL Contribution No. 1904 and QFC Publication number 2018-21. SDP acknowledges support from AgBioResearch of Michigan State University. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Literature Cited

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19:716–723.

Barbiero, R. P., and M. L. Tuchman. 2004. Changes in the crustacean communities of Lakes Michigan, Huron, and Erie following the invasion of the predatory cladoceran *Bythotrephes longimanus*. Canadian Journal of Fisheries and Aquatic Sciences 61:2111–2125.

Beaugrand, G., K. M. Brander, J. A. Lindley, S. Souissi, and P. C. Reid. 2003. Plankton effect on cod recruitment in the North Sea. Nature 426:661–664.

Bjornstad, O., and B. Grenfell. 2001. Noisy clockwork: Time series analysis of population fluctuations in animals. Science 293:638–643.

649 Bourdeau, P. E., K. L. Pangle, and S. D. Peacor. 2011. The invasive predator *Bythotrephes*
650 induces changes in the vertical distribution of native copepods in Lake Michigan.
651 Biological Invasions 13:2533–2545.
652 Bourdeau, P. E., K. L. Pangle, and S. D. Peacor. 2015. Factors affecting the vertical distribution
653 of the zooplankton assemblage in Lake Michigan: The role of the invasive predator
654 *Bythotrephes longimanus*. Journal of Great Lakes Research 41:115–124.
655 Bourdeau, P. E., K. L. Pangle, E. M. Reed, and S. D. Peacor. 2013. Finely tuned response of
656 native prey to an invasive predator in a freshwater system. Ecology 94:1449–1455.
657 Breto, C., D. H. He, E. L. Ionides, and A. A. King. 2009. Time series analysis in mechanistic
658 models. Annals of Applied Statistics 3:319–348.
659 Bunnell, D. B., R. P. Barbiero, S. A. Ludsin, C. P. Madenjian, G. J. Warren, D. M. Dolan, T. O.
660 Brenden, R. Briland, O. T. Gorman, J. X. He, T. H. Johengen, B. F. Lantry, B. M. Lesht,
661 T. F. Nalepa, S. C. Riley, C. M. Riseng, T. J. Treska, I. Tsehaye, M. G. Walsh, D. M.
662 Warner, and B. C. Weidel. 2014. Changing ecosystem dynamics in the Laurentian Great
663 Lakes: Bottom-up and top-down regulation. BioScience 64:26–39.
664 Bunnell, D. B., B. M. Davis, M. A. Chriscinske, K. M. Keeler, and J. G. Mychek-Londer. 2015.
665 Diet shifts by planktivorous and benthivorous fishes in northern Lake Michigan in
666 response to ecosystem changes. Journal of Great Lakes Research 41:161–171.
667 Bunnell, D. B., B. M. Davis, D. M. Warner, M. A. Chriscinske, and E. F. Roseman. 2011.
668 Planktivory in the changing Lake Huron zooplankton community: *Bythotrephes*
669 consumption exceeds that of *Mysis* and fish. Freshwater Biology 56:1281–1296.

Burnham, K. P., and D. A. Anderson. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer - Verlag, New York, New York, USA.

Caceres, C. E. 1998. Interspecific variation in the abundance, production, and emergence of *Daphnia* diapausing eggs. Ecology 79:1699–1710.

Creel, S., J. Winnie, B. Maxwell, K. Hamlin, and M. Creel. 2005. Elk alter habitat selection as an antipredator response to wolves. Ecology 86:3387–3397.

Fahnenstiel, G., T. Nalepa, S. Pothoven, H. Carrick, and D. Scavia. 2010. Lake Michigan lower food web: Long-term observations and *Dreissena* impact. Journal of Great Lakes Research 36:1–4.

Heithaus, M. R., A. J. Wirsing, D. Burkholder, J. Thomson, and L. M. Dill. 2009. Towards a predictive framework for predator risk effects: the interaction of landscape features and prey escape tactics. Journal of Animal Ecology 78:556–562.

Hilborn, R., and M. Mangel. 1997. The Ecological Detective: Confronting Models with Data. Princeton University Press, Princeton, New Jersey, USA.

Ionides, E. L., C. Breto, and A. A. King. 2006. Inference for nonlinear dynamical systems. PNAS 103:18438–18443.

Ionides, E. L., D. Nguyen, Y. Atchade, S. Stoev, and A. A. King. 2015. Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. PNAS 112:719–724.

Jacobs, G. R., C. P. Madenjian, D. B. Bunnell, D. M. Warner, and R. M. Claramunt. 2013. Chinook salmon foraging patterns in a changing Lake Michigan. Transactions of the American Fisheries Society 142:362–372.

Keeler, K. M., D. B. Bunnell, J. S. Diana, J. V. Adams, J. G. Mychek-Londer, D. M. Warner, D. L. Yule, and M. R. Vinson. 2015. Evaluating the importance of abiotic and biotic drivers on *Bythotrephes* biomass in Lakes Superior and Michigan. Journal of Great Lakes Research 41:150–160.

Kerfoot, W. C., J. W. Budd, B. J. Eadie, H. A. Vanderploeg, and M. Agy. 2004. Winter storms: Sequential sediment traps record *Daphnia* ephippial production, resuspension, and sediment interactions. Limnology and Oceanography 49:1365–1381.

Kim, N., and N. D. Yan. 2010. Methods for rearing the invasive zooplankter *Bythotrephes* in the laboratory. Limnology and Oceanography: Methods 8:552–561.

Kimbro, D. L., J. H. Grabowski, A. R. Hughes, M. F. Piehler, and J. W. White. 2017. Nonconsumptive effects of a predator weaken then rebound over time. Ecology 98:656–667.

King, A. A., E. L. Ionides, M. Pascual, and M. J. Bouma. 2008. Inapparent infections and cholera dynamics. Nature 454:877-U29.

King, A. A., D. Nguyen, and E. L. Ionides. 2016. Statistical inference for partially observed Markov processes via the R package pomp. Journal of Statistical Software 69.

Lehman, J. T., and C. A. Caceres. 1993. Food-web responses to species invasion by a predator invertebrate - *Bythotrephes* in Lake Michigan. Limnology and Oceanography 38:879–891.

Martinez-Bakker, M., A. A. King, and P. Rohani. 2015. Unraveling the transmission ecology of polio. PLOS Biology 13.

Matassa, C. M., and G. C. Trussell. 2011. Landscape of fear influences the relative importance of consumptive and nonconsumptive predator effects. Ecology 92:2258–2266.

715 Nelson, E. H., C. E. Matthews, and J. A. Rosenheim. 2004. Predators reduce prey population
716 growth by inducing changes in prey behavior. Ecology 85:1853–1858.
717 Newman, K., S. T. Buckland, B. Morgan, R. King, D. L. Borchers, D. Cole, P. Besbeas, O.
718 Gimenez, and L. Thomas. 2014. Modelling Population Dynamics: Model Formulation,
719 Fitting and Assessment using State-Space Methods. Springer-Verlag, New York, New
720 York, USA.
721 Nowicki, C. J., D. B. Bunnell, P. M. Armenio, D. M. Warner, H. A. Vanderploeg, J. F.
722 Cavaletto, C. M. Mayer, and J. V. Adams. 2017. Biotic and abiotic factors influencing
723 zooplankton vertical distribution in Lake Huron. Journal of Great Lakes Research
724 43:1044–1054.
725 Pangle, K. L., and S. D. Peacor. 2006. Non-lethal effect of the invasive predator *Bythotrephes*
726 *longimanus* on *Daphnia mendotae*. Freshwater Biology 51:1070–1078.
727 Pangle, K. L., and S. D. Peacor. 2009. Light-dependent predation by the invertebrate planktivore
728 *Bythotrephes longimanus*. Canadian Journal of Fisheries and Aquatic Sciences 66:1748–
729 1757.
730 Pangle, K. L., S. D. Peacor, and O. E. Johannsson. 2007. Large nonlethal effects of an invasive
731 invertebrate predator on zooplankton population growth rate. Ecology 88:402–412.
732 Panik, M. J. 2017. Stochastic Population Growth Models. Pages 167–191 Stochastic Differential
733 Equations. John Wiley & Sons, Inc., Hoboken, New Jersey, USA.
734 Peckarsky, B. L., P. A. Abrams, D. I. Bolnick, L. M. Dill, J. H. Grabowski, B. Luttbeg, J. L.
735 Orrock, S. D. Peacor, E. L. Preisser, O. J. Schmitz, and G. C. Trussell. 2008. Revisiting
736 the classics: Considering nonconsumptive effects in textbook examples of predator-prey
737 interactions. Ecology 89:2416–2425.

van de Pol, M., L. D. Bailey, N. McLean, L. Rijsdijk, C. R. Lawson, and L. Brouwer. 2016. Identifying the best climatic predictors in ecology and evolution. Methods in Ecology and Evolution 7:1246–1257.

R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Scheffer, M., D. Straile, E. H. van Nes, and H. Hosper. 2001. Climatic warming causes regime shifts in lake food webs. Limnology and Oceanography 46:1780–1783.

Turchin, P., and A. Taylor. 1992. Complex dynamics in ecological time-series. Ecology 73:289–305.

Vanderploeg, H. A., S. A. Pothoven, G. L. Fahnenstiel, J. F. Cavaletto, J. R. Liebig, C. A. Stow, T. F. Nalepa, C. P. Madenjian, and D. B. Bunnell. 2012. Seasonal zooplankton dynamics in Lake Michigan: Disentangling impacts of resource limitation, ecosystem engineering, and predation during a critical ecosystem transition. Journal of Great Lakes Research 38:336–352.

Vanderploeg, H., J. Liebig, and M. Omair. 1993. *Bythotrephes* predation on Great Lakes zooplankton measured by an in situ method - Implications for zooplankton community structure. Archiv Fur Hydrobiologie 127:1–8.

Yurista, P. M., H. A. Vanderploeg, J. R. Liebig, and J. F. Cavaletto. 2010. Lake Michigan Bythotrephes prey consumption estimates for 1994-2003 using a temperature and size corrected bioenergetic model. Journal of Great Lakes Research 36:74–82.

760 **Table 1:** Model ΔAIC values relative to best model (lowest AIC).

Model	**Maximum Log-Likelihood**	**Parameters**	**AIC**	**ΔAIC**
a. No *B. longimanus* effect	-213.3	11	448.6	6.9
b. *B. longimanus* nonconsumptive effect	-208.9	12	441.7	0.0
c. *B. longimanus* consumption	-212.5	12	449.1	7.3
d. Consumption and nonconsumptive effect	-208.7	13	443.4	1.7
e Monthly average I.I.D.	-336.3	13	698.5	256.8
f. AR (2) with measurement error	-369.4	6	750.7	309.0
g. *Limnocalanus* nonconsumptive effect	-212.5	12	449.1	7.3
h. *B. longimanus* anomaly	-210.2	12	444.4	2.6
i. Seasonal birth and background death	-210.4	14	448.9	7.2
j. Seasonal birth and attack rate	-210.1	15	450.3	8.5
k. Type II functional response	-212.0	13	450.0	8.9

761

762 **Table 2:** Values of model terms at maximum likelihood estimate for best fit model (b).

Parameter	Description	Estimate	Units
λ_1	Seasonal birth rate	-10.0	$\ln(day^{-1})$
λ_2	Seasonal birth rate	-3.4	$\ln(day^{-1})$
λ_3	Seasonal birth rate	-1.2	$\ln(day^{-1})$
λ_4	Seasonal birth rate	0.32	$\ln(day^{-1})$
κ	Density dependence term	32.5	mg x m^{-3}
μ	Background mortality	0.048	day^{-1}
α	Attack rate	NA	(mg *B. longimanus*)$^{-1}$ x day^{-1}
η	Induced proportional birth reduction	0.089	(ln (mg *B. longimanus*))$^{-1}$
ϵ	Standard deviation of growth rate	0.26	day^{-1}
φ	ln (Spring pulse mean)	-3.2	mg x m^{-3}
Ψ	Standard deviation of ln (Spring pulse)	1.7	mg x m^{-3}
σ_a	Measurement error (scales with $V_{(T)}$)	0.22	mg x m^{-3}
σ_b	Measurement error (scales with $V_{(T)}^2$)	0.39	mg x m^{-3}

763

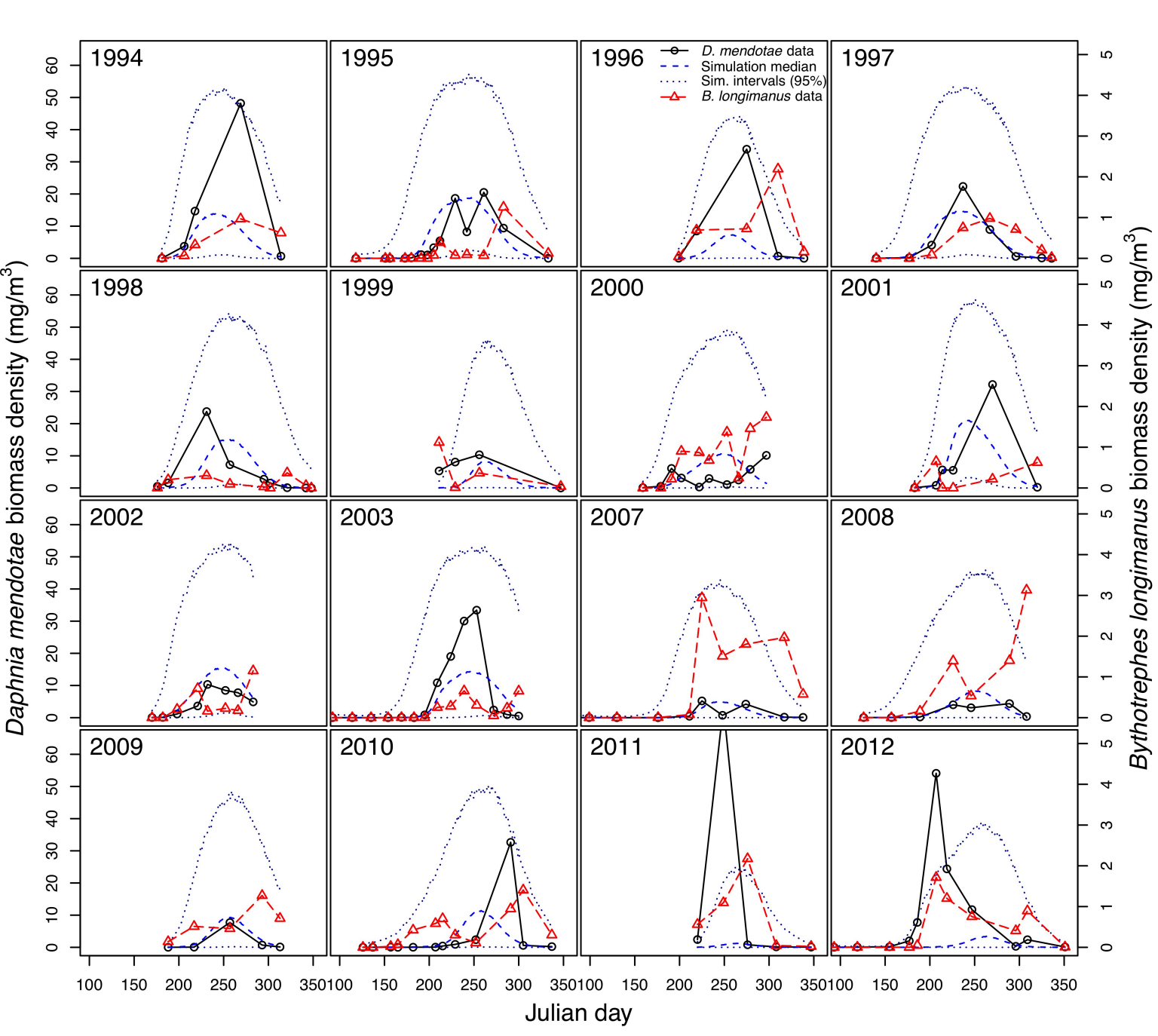
Figure Legends

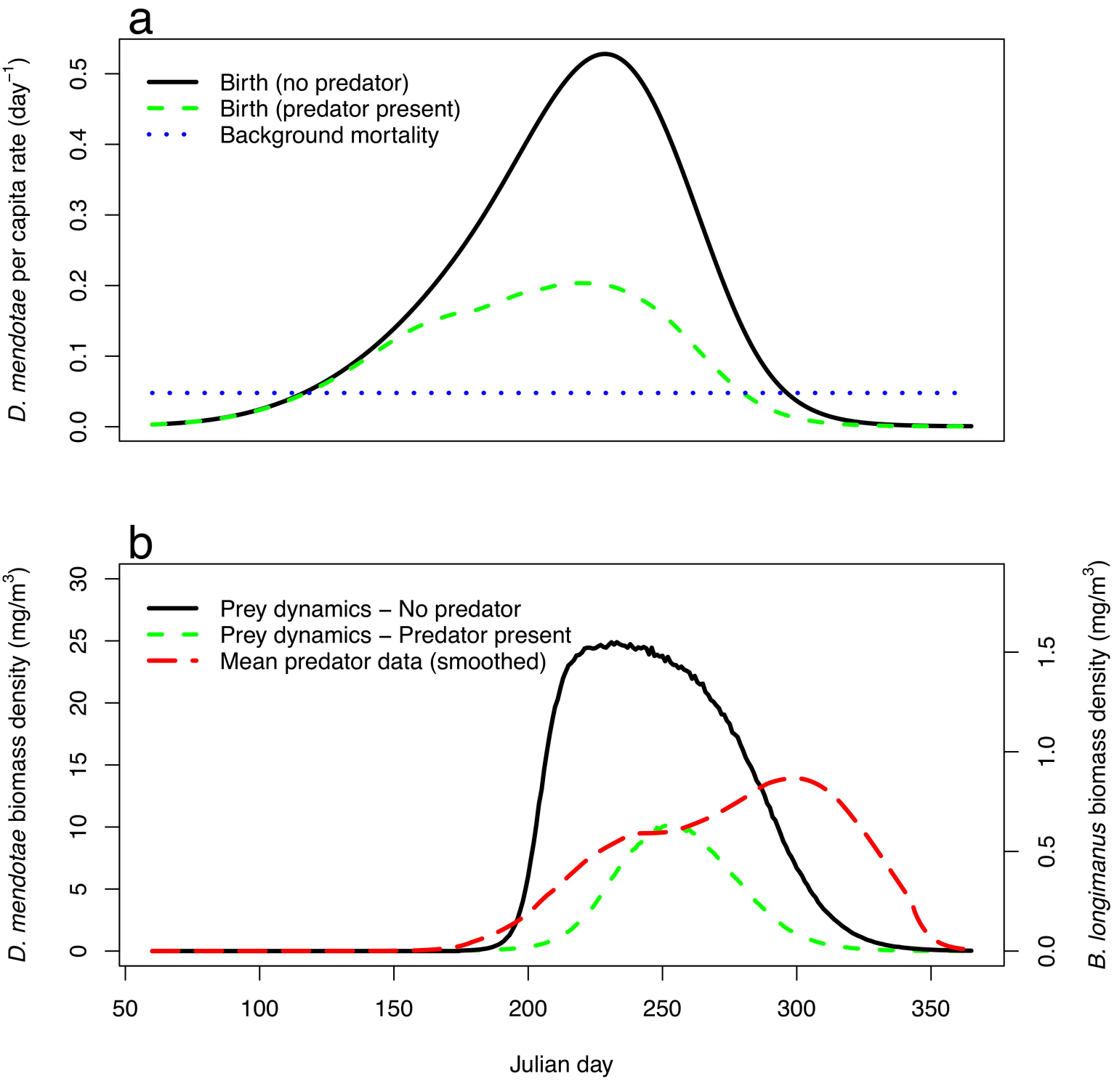
Figure 1: Simulated *Daphnia mendotae* biomass density (mg/m^3) from fitted model compared to *D. mendotae* and *Bythotrephes longimanus* time series data in Lake Michigan from 1994-2012. Median and 95% simulation intervals for the model that only includes nonconsumptive effects (model b); Black solid line: *D. mendotae*; Red dashed line: *B. longimanus*; blue dashed line: median simulated *D. mendotae* biomass density; dark blue dotted line: 95% simulation intervals. The first observations in 2007 and 2012 and the *D. mendotae* peak in 2011 are cut off from the plot.

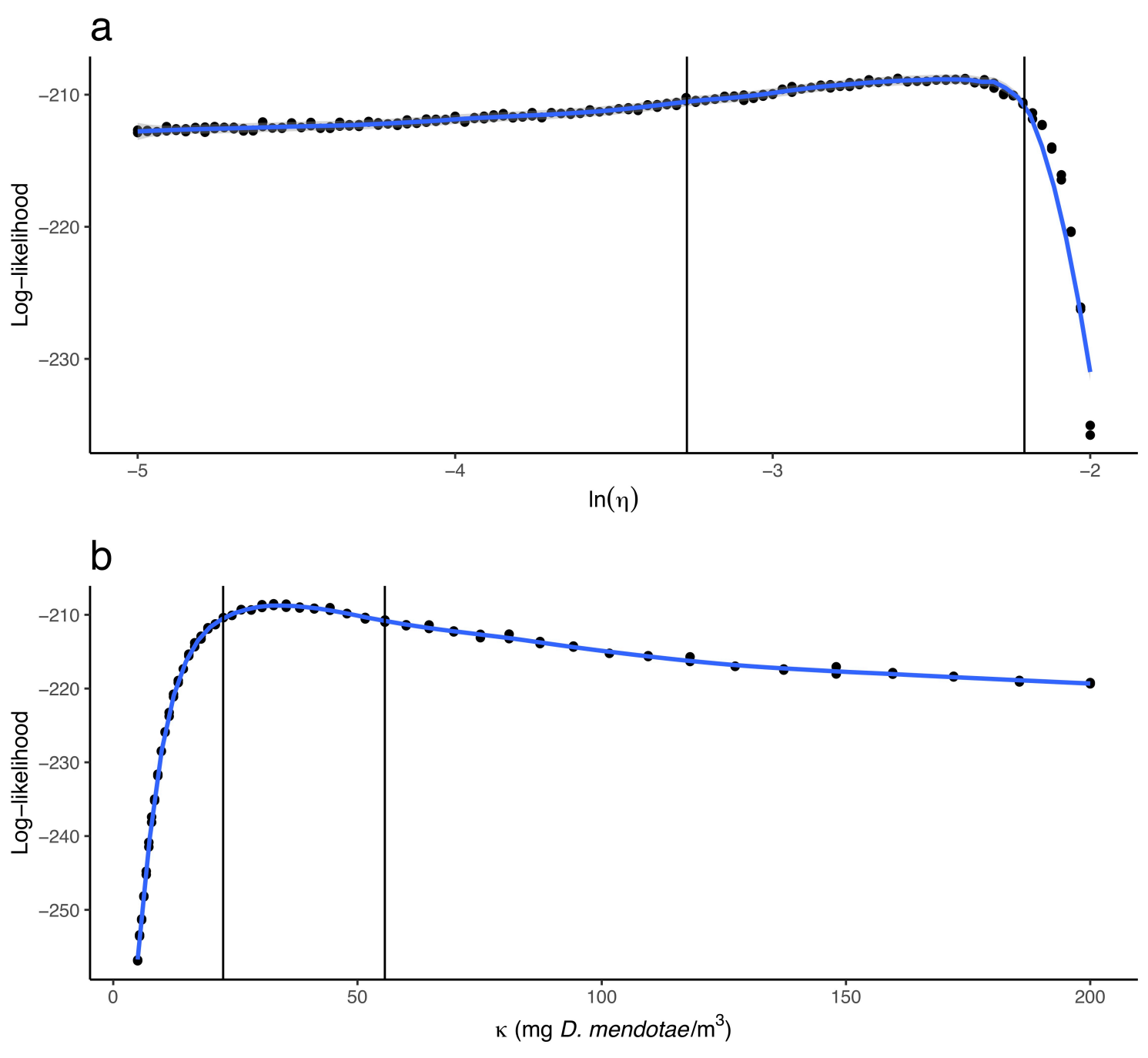
Figure 2: For the fitted model (model b, which only includes nonconsumptive effects): a) estimated seasonal birth rate and b) simulated biomass density (from 10,000 simulations) of *Daphnia mendotae* in the presence (green dashed line) or absence (black solid line) of *Bythotrephes longimanus*. Growth rates and simulated density were determined using across-year averages of smoothed *B. longimanus* biomass density (red dashed line in plot b) for each Julian day. Estimated background mortality rate is indicated by the blue dotted line in (a).

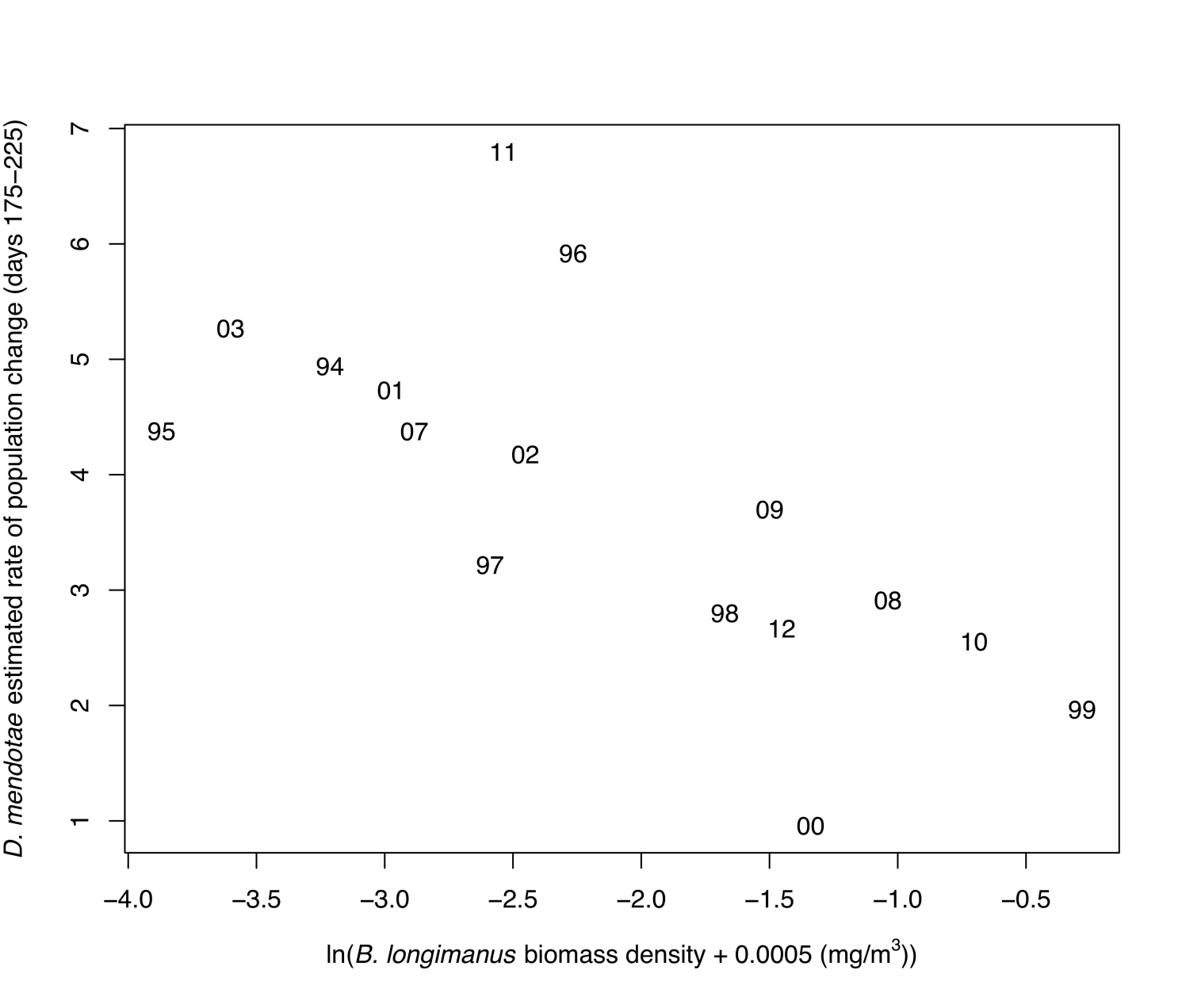
Figure 3: Likelihood profiles for a) η (reduction in *Daphnia mendotae* birth rate in response to *Bythotrephes longimanus*) and b) κ (density dependence) parameters. Black vertical lines indicate 95% confidence intervals (η: 0.038-0.11 (ln (mg *B. longimanus*))$^{-1}$; κ: 22.5-55.6 mg *D. mendotae* per m^3). Black points show the two highest maximum likelihood estimates from the searches performed at each parameter value for each profile, blue lines show a loess smoothed curve fit to those points, and gray shading (approximately the width of the points) indicates confidence intervals for the loess fit.

Figure 4: Estimated rate of change in *Daphnia mendotae* population early in growing season (days 175-225, calculated via Eq. 12) vs. smoothed *Bythotrephes longimanus* biomass density (geometric mean of smoothed *B. longimanus* + 0.005 over days 175-225) each year. Points are shown as 2-digit numbers representing each year.


1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2007
2008
2009
2010
2011
2012
D. mendotae data
Simulation median
Sim. intervals (95%)
B. longimanus data
Daphnia mendotae biomass density (mg/m3)
Bythotrephes longimanus biomass density (mg/m3)
Julian day
0 10 20 30 40 50 60
0 1 2 3 4 5
100 150 200 250 300 350


a
Birth (no predator)
Birth (predator present)
Background mortality
D. mendotae per capita rate (day−1)
0.0
0.1
0.2
0.3
0.4
0.5
b
Prey dynamics – No predator
Prey dynamics – Predator present
Mean predator data (smoothed)
D. mendotae biomass density (mg/m3)
0
5
10
15
20
25
30
B. longimanus biomass density (mg/m3)
0.0
0.5
1.0
1.5
50
100
150
200
250
300
350
Julian day


a
−210
−220
−230
Log−likelihood
−5
−4
−3
−2
ln(η)
b
−210
−220
−230
−240
−250
Log−likelihood
0
50
100
150
200
κ (mg D. mendotae/m3)


11
96
03
94
01
95
07
02
09
97
08
98
12
10
99
00
7
6
5
4
3
2
1
−4.0
−3.5
−3.0
−2.5
−2.0
−1.5
−1.0
−0.5
D. mendotae estimated rate of population change (days 175–225)
ln(B. longimanus biomass density + 0.0005 (mg/m3))

Title: Evaluating consumptive and nonconsumptive predator effects on prey density using field times series data

Authors: J.A. Marino, Jr., S.D. Peacor, D.B. Bunnell, H.A. Vanderploeg, S.A. Pothoven, A.K. Elgin, J.R. Bence, J. Jiao, and E.L. Ionides

Journal: *Ecology*

Appendix S1

Table of Contents

Licensed under the Creative Commons attribution-noncommercial license, http://creativecommons.org/licenses/by-nc/3.0/. Please share and remix noncommercially, mentioning its origin.

Model description and code

We describe the model and pomp code for a Lake Michigan zooplankton system. The state variable is a victim, V (i.e., *Daphnia mendotae*). The predator, P (i.e., *Bythotrephes longimanus*, a large predatory zooplankter) is treated as a covariate. Dynamics are represented by the following stochastic differential equations in which Brownian motion (dW) drives stochasticity in prey growth.

$$dV = (V\beta(t)(1 - V/\kappa)(1 - \eta g(P)) - \alpha P - \mu V)\, dt + V\epsilon\, dW + \rho(t)$$

$$dW \sim \text{Normal}(0, sd = \sqrt{dt})$$

where $\beta(t)$ represents prey brith and/or somatic growth rate at low population size, κ represents density dependence for the prey population, η is a proportional reduction in *D. mendotae* birth rate caused by *B. longimanus* (i.e., a nonconsumptive effect), α is predator attack rate, μ is prey background mortality, ϵ is the standard deviations of the prey population growth rate, and $\rho(t)$ is a pulse term that initiates prey dynamics each year. Prey birth rate changes seasonally according to:

$$\beta(t) = \exp\left\{\sum_{i=1}^{N_s} \lambda_i \, s_i(t)\right\}$$

where $\{s_i(t), i = 1, \ldots, N_s\}$ is a periodic cubic B-spline basis; $\{\lambda_i, i = 1, \ldots, N_s\}$ model seasonality of the birth rate. We set $N_s = 4$.

$\rho(t)$ represents the initiation of the population each year via *D. mendotae* emergence from resting eggs. The size of this pulse is assumed to be random and log-normally distributed, with parameters ϕ and ψ representing the natural log of the mean and standard deviation of the pulse (ρ(t)), respectively:

$$\ln(\rho(t)) \sim \text{Normal}(\phi, \psi)$$

The pulse is assumed to occur 7 days prior to the first observation of *D. mendotae*. On other dates, $\rho(t)$ = 0.

We build the rpocess model according to the above equations.

```
Daphnia_rprocess <- Csnippet("
  double births;
  double deaths;
  double betaz;
  double dw;

  if (error_count != 0.0) return;

  betaz = exp(dot_product(4,&seas_1,&log_betaz1));

  dw = rnorm(0,sqrt(dt));  // white noise for prey growth

  births = (betaz*(1-eta*(7.601+log(byth + 0.0005)))*dt)*prey*(1-prey/kappa)
+
  prey*sdbeta*dw + pulse*rlnorm(mean_rho,sigma_rho);    // prey births
  deaths = prey*(alpha*byth*0 + mu)*dt; // prey deaths due to predation

  prey += births - deaths;
  noise += dw;

  // check for violations of positivity constraints
```

```
  if (prey <= 0.0)  {
    prey = 0;
  }
  //reset population to 0 at beginning of each year
  if (nocoup < 1)  {
    prey = 0;
  }
  ")
```

The data consist of measurements of *D. mendotae* biomass density:

```
#need to import data
zoop<-read.table("GLERL_M110_Zoop_1994-2012.txt",header=TRUE)
zoop$day<-as.numeric(1+(as.Date(zoop$Date, format="%m/%d/%y")-rep(as.Date("19
94-01-01"),length(zoop$Date))))
zoopsubset<-subset(zoop,Prior_to_1st_Daphnia_obs==0) #generate subset excludi
ng early zero observations
```

We generate the *B. longimanus* covariate by assuming that *B. longimanus* is absent from the water column during the first 50 days each year beginning January 1 and then linearly interpolating remaining data points from observed biomass density of *B. longimanus*. We then use those interpolated values to calculate a 45-day moving average.

```
zoop$cases<-complete.cases(zoop$D.mendotae);zoopcomplete<-subset(zoop,cases==
TRUE)
require(timeSeries)
Byth<-rep(c(rep(0,50),rep(NA,315)),19);
Byth[zoopcomplete$day]<-zoopcomplete$Bythotrephes; Byth[1]<-0;Byth[6935]<-0
Byth<-interpNA(Byth, method = "linear")
Byth_corrected<-Byth+0.0005

#calculate 45 day moving average
BythMA45<-rep(0.0005,6935)

for (i in 23:6913) {
  BythMA45[i]<-(Byth_corrected[i-22]*Byth_corrected[i-21]*
    Byth_corrected[i-20]*Byth_corrected[i-19]*Byth_corrected[i-18]*
    Byth_corrected[i-17]*Byth_corrected[i-16]*Byth_corrected[i-15]*
    Byth_corrected[i-14]*Byth_corrected[i-13]*Byth_corrected[i-12]*
    Byth_corrected[i-11]*Byth_corrected[i-10]*Byth_corrected[i-9]*
    Byth_corrected[i-8]*Byth_corrected[i-7]*Byth_corrected[i-6]*
    Byth_corrected[i-5]*Byth_corrected[i-4]*Byth_corrected[i-3]*
    Byth_corrected[i-2]*Byth_corrected[i-1]*Byth_corrected[i]*
    Byth_corrected[i+1]*Byth_corrected[i+2]*Byth_corrected[i+3]*
    Byth_corrected[i+4]*Byth_corrected[i+5]*Byth_corrected[i+6]*
    Byth_corrected[i+7]*Byth_corrected[i+8]*Byth_corrected[i+9]*
    Byth_corrected[i+10]*Byth_corrected[i+11]*Byth_corrected[i+12]*
    Byth_corrected[i+13]*Byth_corrected[i+14]*Byth_corrected[i+15]*
    Byth_corrected[i+16]*Byth_corrected[i+17]*Byth_corrected[i+18]*
    Byth_corrected[i+19]*Byth_corrected[i+20]*Byth_corrected[i+21]*
```

```
    Byth_corrected[i+22])^(1/45)
}

#need to back-transform
BythMA45<-BythMA45-0.0005;BythMA45<-round(BythMA45,10)
```

We build the periodic B-spline basis for $\beta(t)$. We also generate the covariate representing an influx of *D. mendotae* (set to occur 1 week prior to the observation of *D. mendotae* each year) to determine the timing of the pulse ($\rho(t)$).

```
#set up covariate to reset population to 0 each year
nocoup<-rep(c(rep(1,364),0),19)

#figure out first date of daphnia observation each year
#need to determine first observation of Daphnia each year
zoopcomplete$year<-year(as.Date(zoopcomplete$Date, format="%m/%d/%y"))
surveyyears<-c(1994:2003,2007:2012)
firstdaph<-rep(0,16)
firstdaphjday<-rep(0,16)
firstdaphdens<-rep(0,16)

for (i in 1:16) {
  zoopyear<-subset(zoopcomplete,year==surveyyears[i])
  zoopyearno0<-subset(zoopyear,D.mendotae>0)
  firstdaph[i]<-zoopyearno0$day[which.min(zoopyearno0$JulianDay)]
  firstdaphjday[i]<-zoopyearno0$JulianDay[which.min(zoopyearno0$JulianDay)]
  firstdaphdens[i]<-zoopyearno0$D.mendotae[which.min(zoopyearno0$JulianDay)]
}

#set start equal to 7 days before each observation
daphstart<-firstdaph-7

#create vector with 1s at time of first observation to indicate when pulse oc
curs
pulse<-rep(0,6935)
pulse[daphstart]<-1



covar.dt <- 1
t0 <- 0
nbasis <- 4
tcovar <- seq(from=t0,to=6934,by=covar.dt)
yr <- 1:365

covartable <- data.frame(
  time=tcovar,
  seas=periodic.bspline.basis(tcovar,nbasis=nbasis,degree=3,period=365),
  byth=BythMA45,
  nocoup,
```

```
  pulse
)
```

Here are some candidate parameter values:

```
params.init <- c(sigmaa=0.21907755,sigmab=0.38558873,
                 log.betaz1=-10.04763994,log.betaz2=-3.41093475,
                 log.betaz3=-1.06310127,log.betaz4=0.31538577,
                 alpha=0,mu=0.04784941,eta=0.08924933,
                 mean_rho=-3.15200907,sigma_rho=1.69402220,
                 kappa=32.54473250,sdbeta=0.25785123,
                 prey.0=0,noise.0=0,error_count.0=0)
```

The state variable V is linked to observed measures of *D. mendotae* biomass density using a measurement model. The measurement model is specified as a left-censored normal model to handle the zero observations, with demographic and environmental scale measurement noise terms (σ_a and σ_b) .

$$Vobs_{(t)} \sim \text{Normal}(V_{(t)}, \sigma)$$

$$\sigma = \sqrt{\sigma_a^2 V_{(t)} + \sigma_b^2 V_{(t)}^2}$$

if $Vobs_{(t)} > 0$. Otherwise, $Vobs_{(t)} = 0$

```
Daphnia_rmeasure <- Csnippet("
  double sigma, daphnia;
  sigma=sqrt(sigmaa*sigmaa*prey + sigmab*sigmab*prey*prey);
  if ((error_count > 0.0) || (!(R_FINITE(prey)))) {
    D_mendotae = R_NaReal;
  } else {
    daphnia = rnorm(prey,sigma);
  if (daphnia<=0) {
    D_mendotae=0;
  }
  else {
    D_mendotae=daphnia;
  }
  }
")

Daphnia_dmeasure <- Csnippet("
  double sigma, tol = 1.0e-18;
  sigma=sqrt(sigmaa*sigmaa*prey + sigmab*sigmab*prey*prey);
  double f = 0.0;
  if ((error_count > 0.0) || (!(R_FINITE(prey)))) {
    lik = tol;
  } else {
  if (D_mendotae==0) {
    f += pnorm(0,prey,sigma,1,1)+tol;
```

```
  }
  else {
    f += dnorm(D_mendotae, prey, sigma, 1)+tol;
  }
  lik = (give_log) ? f : exp(f);
  }
")
```

We also build some parameter transformations. These are only needed for optimization, so could be added later when we get to mif2 (iterated filtering).

```
Daphnia_untrans <- Csnippet("
  Tsigmaa = log(sigmaa);
  Tsigmab = log(sigmab);
  Tkappa = log(kappa);
  Tsdbeta = log(sdbeta);
  Tmu = log(mu);
  Teta = log(eta);
  Tsigma_rho = log(sigma_rho);
  ")

Daphnia_trans <- Csnippet("
                      Tsigmaa = exp(sigmaa);
                      Tsigmab = exp(sigmab);
                      Tkappa = exp(kappa);
                      Tsdbeta = exp(sdbeta);
                      Tmu = exp(mu);
                      Teta = exp(eta);
                      Tsigma_rho = exp(sigma_rho);
                      ")
```

We then construct the pomp object.

```
Daphnia_pomp <- pomp(
  data=subset(zoopsubset,select=c("D.mendotae","day")),
  times="day",
  t0=0,
  params=params.init,
  rprocess = euler.sim(step.fun = Daphnia_rprocess, delta.t=1),
  rmeasure= Daphnia_rmeasure,
  dmeasure = Daphnia_dmeasure,
  covar=covartable,
  tcovar="time",
  obsnames = c("D.mendotae"),
  zeronames = c("error_count"),
  statenames = c("prey","noise","error_count"),
  paramnames = c("sigmaa","sigmab","log.betaz1","kappa","alpha","mu","eta","s
dbeta","mean_rho","sigma_rho","prey.0","noise.0","error_count.0"),
  covarnames = c("seas.1","byth","nocoup","pulse"),
  fromEstimationScale=Daphnia_trans,
```

```
  toEstimationScale=Daphnia_untrans
)
plot(Daphnia_pomp@times,Daphnia_pomp@data[1,],type="l",xlab="Time (days)",yla
b=expression(paste(italic("D. mendotae"), " biomass density (", "mg/m"^"3", "
)")))
```

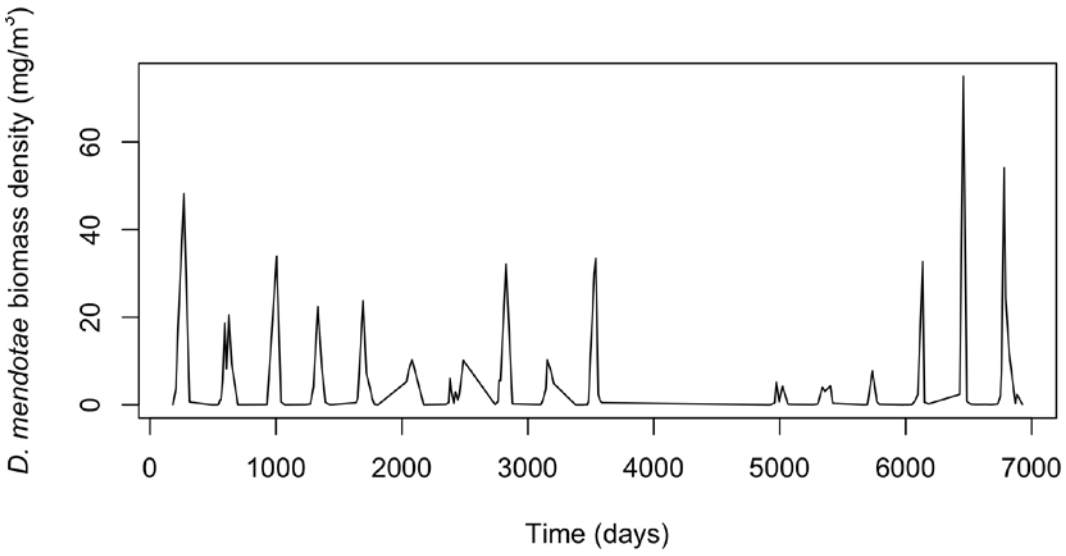


Figure S1: *D. mendotae* biomass density data.

The model can then be simulated to test the coding:

```
sim1 <- simulate(Daphnia_pomp,seed=3)
plot(sim1@times,sim1@data[1,],type="l", xlab="Time (days)", ylab=expression(p
aste("Simulated ",italic("D. mendotae"), " biomass density (", "mg/m"^"3", ")
")))
```

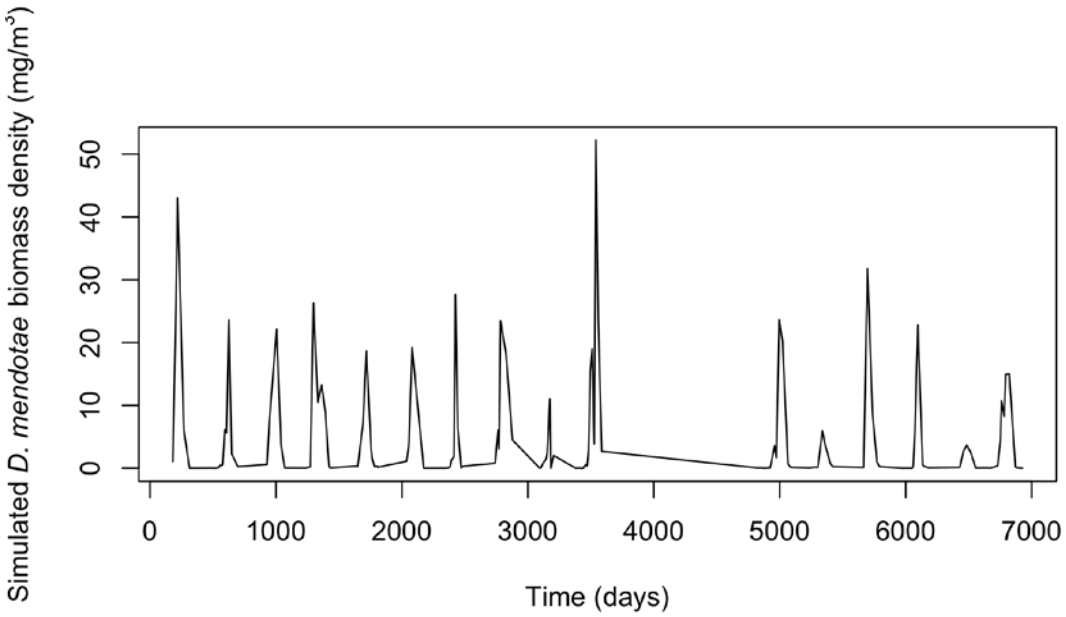


Figure S2: Simulation of *D. mendotae* data from pomp object.

Comparison with benchmark models

A reasonable mechanistic model should outperform some simple benchmarks. First, we compared the mechanistic model to a model assuming that *D. mendoate* biomass density is independently and identically distributed around a seasonal (monthly) average.

```
zoopsubset$month<-month(as.Date(zoopsubset$Date, format="%m/%d/%y"))


#create function for monthly mean (11 parameters because no Jan observations)
and two sigmas
truncnormiid<- function (x) {
  f<-rep(0,134);
  for (i in 1:134) {
    prey<-exp(x[zoopsubset$month[i]-1]);
    D_mendotae<-zoopsubset$D.mendotae[i]; #
    sigma<-sqrt(x[12]^2*prey + x[13]^2*prey^2); #two parameters control error
    if (D_mendotae==0) {
      f[i]<-pnorm(0,prey,sigma,1,1); }
    else {
      f[i]<- dnorm(D_mendotae, prey, sigma, 1);
    }
  }
  -sum(f)
}

testparams<-c(rep(0.0001,11),1,1)
maxf<-optim(par=testparams,fn=truncnormiid,method="BFGS")
```

As an alternative benchmark, we can use an AR (2) model represented as a state-space model with measurement error, using the same measurement model as the mechanistic model.

```
AR_rprocess_one <- Csnippet("
       double Ez;

       if (error_count != 0.0) return;

       Ez=rnorm(prey,epsilonz);
       prey = alphaz*prey + betaz*preylag + Ez;
       preylag = prey;

       // check for violations of positivity constraints
       // nonzero error_count variable signals violation
       if (prey <= 0.0)  {
       prey = 0.00000001;
       }
        if (preylag <= 0.0)  {
        preylag = 0.0;
         error_count += 1;
         }
         ")
```

Both benchmark models fit substantially worse than our mechanistic models (Table 1), strengthening our confidence that our mechanistic models provide a reasonable fit.

Anomaly analysis

To disentangle the contribution of seasonal and inter-annual variation in *D. mendotae* biomass to our results, we performed a loess regression of *B. longimanus* biomass density vs. Julian day. The fitted points represented a seasonal average and the residuals represented the deviations from that average (i.e., an anomaly).

```
#calculate anomaly
jday<-rep(1:365,19)
lo.B<-loess(BythMA45 ~ jday,span=0.25)
bythseas<-lo.B$fitted
bythanom<-lo.B$residuals
#plot(1:6935,bythanom,type="l",xlab="day")
plot(60:365,bythanom[60:365],xlab="Julian day", ylab=expression(paste(italic(
"B. longimanus"), " anomaly (", "mg/m"^"3", ")")),type="l",
     xlim=c(150,365),ylim=c(min(bythanom),max(bythanom)))
legend("topleft",legend=c("annomaly","mean anomaly"),lty=c(1,1),col=c("black"
,"red"))
sampleyear.no<-c(2:10,14:19)
for (i in 1:15) {
  lines(60:365,bythanom[60:365+365*(sampleyear.no[i]-1)])
}
anoavg<-rep(0,365)
```

```
for (i in 1:16) {
  anomavg<-bythanom[1:365 + 365*(i-1)]
}
lines(1:365,anomavg,col="red")
```

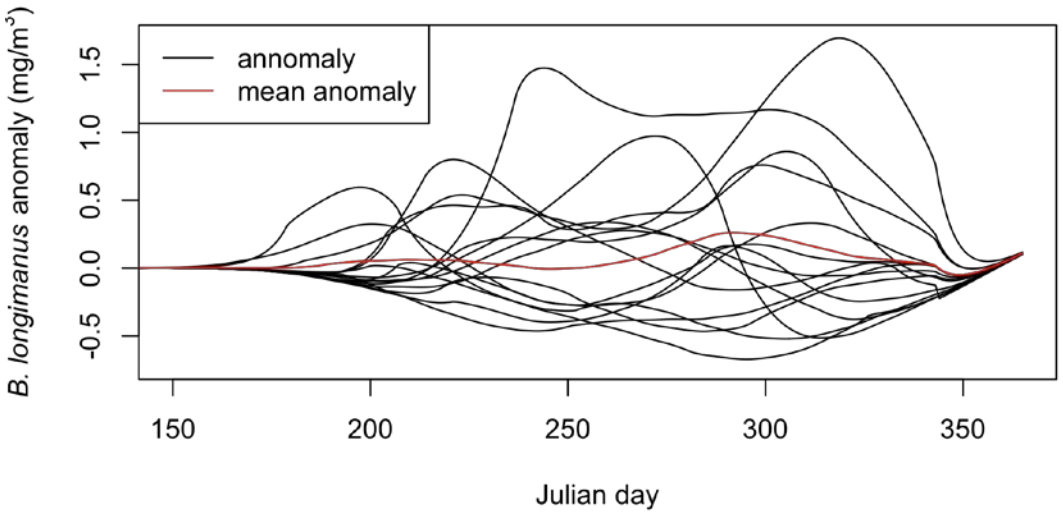


Figure S3: *B. longimanus* anomaly for different years (black line) and averaged across years (red line).

We then can substitute the sum of the seasonal average and anomaly term for total *B. longimanus* biomass density as the predator covariate and test the consequences for model fit.

```
Daphnia_rprocess_anom <- Csnippet("
       double births;
       double deaths;
       double betaz;
       double dw;

       if (error_count != 0.0) return;

       betaz = exp(dot_product(4,&seas_1,&log_betaz1));

       dw = rnorm(0,sqrt(dt));  // white noise for prey birth

       births = (betaz*(1-eta*(0.163 + log(bythanom + 0.85)))*dt)*prey*(1-pre
y/kappa) + prey*sdbeta*dw + pulse*rlnorm(mean_rho,sigma_rho);  // prey births
       deaths = prey*mu*dt; // prey deaths due to predation

       prey += births - deaths;
       noise += dw;
```

```
      // check for violations of positivity constraints
      // nonzero error_count variable signals violation
      if (prey <= 0.0)  {
      prey = 0;
      }
      if (nocoup < 1)  {
      prey = 0;
       }
      ")

covartable_anom <- data.frame(
  time=tcovar,
  seas=periodic.bspline.basis(tcovar,nbasis=nbasis,degree=3,period=365),
  bythseas=bythseas,
  bythanom=bythanom,
  nocoup,
  pulse
)

Daphnia_pomp_anom <- pomp(
  data=subset(zoopsubset,select=c("D.mendotae","day")),
  times="day",
  t0=0,
  params=params.init,
  rprocess = euler.sim(step.fun = Daphnia_rprocess_anom, delta.t=1),
  rmeasure= Daphnia_rmeasure,
  dmeasure = Daphnia_dmeasure,
  covar=covartable_anom,
  tcovar="time",
  obsnames = c("D.mendotae"),
  zeronames = c("error_count"),
  statenames = c("prey","noise","error_count"),
  paramnames = c("sigmaa","sigmab","log.betaz1","kappa","alpha","mu","eta","s
dbeta","mean_rho","sigma_rho","prey.0"),
  covarnames = c("seas.1","bythseas","bythanom","nocoup","pulse"),
  all.state.names=c("prey","noise", "error_count"),
  comp.names=c("prey"),
  comp.ic.names=c("prey.0"),
  fromEstimationScale=Daphnia_trans,
  toEstimationScale=Daphnia_untrans,
  initializer = function (params, t0, comp.ic.names, comp.names, all.state.na
mes, ...) {
    states <- numeric(length(all.state.names))
    names(states) <- all.state.names
    frac <- params[comp.ic.names]
    states[comp.names] <- frac
    states
  }
)
```

The model including the *B. longimanus* anomaly performed substantially better than the null model ($\Delta AIC > 4$), suggesting that interaannual variation contributes to the observed *B. longimanus* effect.

Model fitting

For each model, we performed 100 searches using mif2 function in the pomp R package in which a search through parameter space was initiated using a random set of starting values for each parameter. Starting values were generated from a uniform distribution bounded by plausible values for each parameter.

```
param.tab <- read.table("params.csv",sep=",",row.names=1,header=TRUE)
kable(t(param.tab[1:13]),caption = "Parameter ranges used to generate random starting point for each search")
```

Parameter ranges used to generate random starting point for each search

	lower.bound	upper.bound
sigmadem	0.0010	2.0
sigmaenv	0.0010	2.0
log.betaz1	-11.0000	11.0
log.betaz2	-11.0000	11.0
log.betaz3	-11.0000	11.0
log.betaz4	-11.0000	11.0
alpha	-0.1000	0.1
eta	0.0001	1.0
mean_rho	-10.0000	0.0
sigma_rho	0.0010	10.0
mu	0.0010	1.0
kappa	5.0000	1000.0
sdbeta	0.0010	1.0

```
CORES<-10 ##update for flux
JOBS<-100 ##update for flux

require(doParallel)
registerDoParallel(CORES)

tic <- Sys.time()
mpar <- foreach(
  i=1:JOBS,
  .packages=c('pomp'),
  .inorder=FALSE) %dopar% {
    Sys.sleep(i*.1)
    NMIF<-200 ##update for flux
    NP<-10000 ##update for flux
```

```
    METHOD="mif2"
    param.tab <- read.table("params.csv",sep=",",row.names=1,header=TRUE)
    LV.pars <- c("sigmaa","sigmab","log.betaz1","log.betaz2",
                 "log.betaz3","log.betaz4","alpha", "mu","kappa",
                 "eta","sdbeta","mean_rho","sigma_rho")
    LV.ivps <- c("prey.0")
    LV.rw.sd<- rw.sd(sigmaa=0.02, sigmab=0.02,log.betaz1=0.02,
                     log.betaz2=0.02,log.betaz3=0.02,log.betaz4=0.02,
                     alpha=0.02, mu=0.02,kappa=0.02,eta=0.02,sdbeta=0.02,
                     mean_rho=0.02,sigma_rho=0.02)

    LV.hyperparams <-
      list(min=unlist(param.tab["lower.bound",]),max=unlist(param.tab["upper.
bound",]))

    LV.rprior <- function(hyperparams, ...)
    {
      r<-runif(length(hyperparams$min),min=hyperparams$min,max=hyperparams$ma
x)
      names(r) <- names(hyperparams$min)
      return(r)
    }
    set.seed(8100+i)
    Sys.sleep(i*0.1)
    th.draw <-LV.rprior(LV.hyperparams)
    m<-try(mif2(Daphnia_pomp,
                Nmif=NMIF,
                start=th.draw, # we will initialize
                rw.sd=LV.rw.sd,
                Np=NP,
                cooling.type='geometric',
                cooling.fraction= 0.3,
                max.fail=200,
                transform=TRUE
    ))
    list(pomp=m,start=th.draw)
  }
m.out <- rbind(
      pf.lik = sapply(mpar,function(x){
        if(class(x$pomp)=="mif2d.pomp") logLik(x$pomp) else NA
      }),
      sapply(mpar,function(x) {
        if(class(x$pomp)=="mif2d.pomp") coef(x$pomp) else rep(NA,length(coef(
Daphnia_pomp)))
      }),
      sapply(mpar,function(x)x$start)
    )
toc <- Sys.time()
print(toc-tic)
print(m.out[1,])
```

Profile likelihood

Confidence intervals for parameters of interest were generated via profile likelihood, in which the likelihood is maximized across a fixed range of values for the parameter of interest while estimating all other parameters (Hilborn and Mangel 1997). Code to generate the profile for η is below.

```
CORES<-14 ##update for flux
require(doParallel)
registerDoParallel(CORES)

source("Daphnia.R")

estpars <- setdiff(names(params.init),c("eta"))

theta.t <- partrans(Daphnia_pomp,params.init,"toEstimationScale")

theta.t.hi <- theta.t.lo <- theta.t
theta.t.lo[estpars] <- theta.t[estpars]-log(2)
theta.t.hi[estpars] <- theta.t[estpars]+log(2)

profileDesign(
  eta=seq(from=-3,to=-2,length=50),
  lower=theta.t.lo,upper=theta.t.hi,nprof=100
) -> pd

dim(pd)

pd <- as.data.frame(t(partrans(Daphnia_pomp,t(pd),"fromEstimationScale")))

bake("eta-profile1.rds",{

  foreach (p=iter(pd,"row"),
           .combine=rbind,
           .errorhandling="remove",
           .inorder=FALSE,
           .options.mpi=list(chunkSize=1,seed=1598260027L,info=TRUE)
  ) %dopar% {

    tic <- Sys.time()

    require(magrittr)
    require(plyr)
    require(reshape2)
    require(pomp)

    options(stringsAsFactors=FALSE)
    dat<-subset(zoopcomplete,select=c("D.mendotae","day"))
```

```
    dat %>%
      pomp(
  times="day",
  t0=0,
  params=params.init,
  rprocess = euler.sim(step.fun = Daphnia_rprocess, delta.t=1),
  rmeasure= Daphnia_rmeasure,
  dmeasure = Daphnia_dmeasure,
  covar=covartable,
  tcovar="time",
  obsnames = c("D.mendotae"),
  zeronames = c("error_count"),
  statenames = c("prey","noise","error_count"),
  paramnames = c("sigmaa","sigmab","log.betaz1","kappa","alpha","mu","eta","s
dbeta","mean_rho","sigma_rho","prey.0","noise.0","error_count.0"),
  covarnames = c("seas.1","byth","nocoup","pulse"),
  fromEstimationScale=Daphnia_trans,
  toEstimationScale=Daphnia_untrans
        ) %>%
      mif2(start = unlist(p),
           Nmif = 75,
           rw.sd = rw.sd(sigmaa=0.02,sigmab=0.02,
                         log.betaz1=0.02,log.betaz2=0.02,
                         log.betaz3=0.02,log.betaz4=0.02,
                         mu=0.02,mean_rho=0.02,
                         sigma_rho=0.02,alpha=0,
                         kappa=0.02,sdbeta=0.02,
                         prey.0=ivp(0)),
           Np = 2000,
           cooling.type = "geometric",
           cooling.fraction.50 = 0.1,
           max.fail=200,
           transform = TRUE) %>%
      mif2() -> mf

    ## Runs 10 particle filters to assess Monte Carlo error in likelihood
    pf <- replicate(10, pfilter(mf, Np = 2000))
    ll <- sapply(pf,logLik)
    ll <- logmeanexp(ll, se = TRUE)
    nfail <- sapply(pf,getElement,"nfail")

    toc <- Sys.time()
    etime <- toc-tic
    units(etime) <- "hours"

    data.frame(as.list(coef(mf)),
               loglik = ll[1],
               loglik.se = ll[2],
```

```
                nfail.min = min(nfail),
                nfail.max = max(nfail),
                etime = as.numeric(etime))
  }
}) -> eta_prof
```

Note on reproducibility

To enhance the reproducibility of this work, this appendix was generated using Rmarkdown (https://rmarkdown.rstudio.com/). An advantage of Rmarkdown is that it shows the development of the model combining the mathematical model specification, model code and data analysis code in a single, reproducible document.

References

Hilborn, R., and M. Mangel. 1997. *The Ecological Detective: Confronting Models with Data*. Princeton University Press, Princeton, New Jersey, USA.