

Fast fitting of neural ordinary differential equations by Bayesian neural gradient matching to infer ecological interactions from time series data

Willem Bonnaffé^{1,2} & Tim Coulson²

1. Big Data Institute, University of Oxford, Old Road Campus, Oxford OX3 7LF
2. Department of Biology, University of Oxford, Zoology Research and Administration Building, 11a Mansfield Road, Oxford OX1 3SZ

Emails: willem.bonnaffe@nds.ox.ac.uk; tim.coulson@zoo.ox.ac.uk;

Running title: Fast NODEs to infer ecological interactions

Keywords: Artificial neural networks; Ecological dynamics; Ecological interactions; Geber method; Gradient matching; Neural ordinary differential equations; Microcosm; Ordinary differential equations; Prey-predator dynamics; Rotifer; Time series analysis;

Specifications: 257 words in abstract; 7861 words in text; 58 references; 6 figures; 2 tables

Contact: Willem Bonnaffé, 128 Southfield Park, Oxford, OX4 2BA, UK (w.bonnaffe@gmail.com)

Statement of authorship: Willem Bonnaffé designed the method, performed the analysis, wrote the manuscript; Tim Coulson led investigations, provided input for the manuscript, commented on the manuscript.

Abstract

1. Inferring ecological interactions is hard because we often lack suitable parametric representations to portray them. Neural ordinary differential equations (NODEs) provide a way of estimating interactions nonparametrically from time series data. NODEs, however, are slow to fit, and inferred interactions have not been compared to the truth.
2. We provide a fast NODE fitting method, Bayesian neural gradient matching (BNGM), which relies on interpolating time series with neural networks, and fitting NODEs to the interpolated dynamics with Bayesian regularisation. We test the accuracy of the approach by inferring ecological interactions in time series generated by an ODE model with known interactions. We also infer interactions in experimentally replicated time series of a microcosm featuring an algae, flagellate, and rotifer population, as well as in the hare and lynx system.
3. Our BNGM approach allows us to cut down the fitting time of NODE systems to only a few seconds. The method provides accurate estimates of ecological interactions in the artificial system, as linear and nonlinear true interactions are estimated with minimal error. In the real systems, dynamics are driven by a mixture of linear and nonlinear ecological interactions, of which only the strongest are consistent across replicates.
4. Overall, NODEs alleviate the need for a mechanistic understanding of interactions, and BNGM alleviates the heavy computational cost. This is a crucial step availing quick NODE fitting, cross-validation, and uncertainty quantification, as well as more objective estimation of interactions, and complex context dependence, than parametric models.

1 Introduction

The concept of population is central in ecology (Berryman 2002). Ecologists have had a long-standing interest in finding laws that govern population dynamics, namely changes in the number of individuals in the populations (Lawton 1999; Turchin 1999). Population dynamics can be characterised by a logistic growth, or similar forms, limited by ecological interactions with other organisms, and by the state of the environment (Turchin 2001; Berryman 2003). Intra-specific interactions correspond to interactions between individuals of different sex, age or size classes, belonging to the same species (Turchin 2001). Inter-specific interactions are interactions between individuals from different species, be it competitors, preys, predators, or pathogens (Turchin 2001; Berryman 2003). These interactions can cause populations to have lagged effects impacting their own growth, often called feedback effects, mediated by their impact on the other populations they interact with (Berryman and Turchin 1997).

Characterising these interactions has been a longtime challenge. Ecologists started analysing time series data with parametric models (Royama 1984; Kendall et al. 1999; Ives et al. 2003; Gross, Ives, and Nordheim 2005), as time series of population counts are the most commonly collected long-term data in biology (Kendall et al. 1999). Initial analysis involved fitting simple auto-regressive linear models to time series of a single species, leading to contentious interpretations of interactions thereby inferred (e.g. Berryman and Turchin 1997). For instance, Royama et al. interpreted higher order lags as evidence of species interactions (Royama 1984), while Lande et al. interpreted them as age-structure signatures (Lande et al. 2002). Coulson et al. showed they can even be caused by

21 interactions between the sexes (Mysterud, Coulson, and Stenseth 2002). Jonzen et al. added doubt
22 over interpreting lags by demonstrating that autocorrelation in environmental noise could prevent
23 altogether the reliable estimation of lag effects in single species time series data (Jonzén et al.
24 2002). More recent work has investigated time series of multiple species, environmental factors,
25 and has mechanistically modelled various ecological interactions (e.g. Bruijning, Jongejans, and
26 Turcotte 2019; Rosenbaum et al. 2019; Adams et al. 2020). In these models, ecological interactions
27 are quantified explicitly by specific parameters, rather than phenomenologically with lags. This
28 allowed for a more thorough quantification of interactions and comparison of alternative ecological
29 interactions architectures.

30 However, ecologists still face two main obstacles when estimating ecological interactions from time
31 series data. The first is that interactions are highly context-dependent, so that they change in time
32 with the state of the ecosystem and of the environment (Song et al. 2020). Ecological interactions
33 were traditionally considered linear or fixed, yet there is substantial evidence that this is not the
34 case in nature (e.g. Bonsall, Meijden, and Crawley 2003; Gross, Ives, and Nordheim 2005; Kendall
35 et al. 2005; Ushio et al. 2018; Bruijning, Jongejans, and Turcotte 2019; Rosenbaum et al. 2019;
36 Bonnaffé et al. 2021). The effect of the population on itself depends on the density of individuals
37 (e.g. Lingjaerde et al. 2001; Moe et al. 2005; Brook and Bradshaw 2006); while predation rates
38 can depend on the density of the predator (Jost and Ellner 2000; Yoshida et al. 2003). Many
39 vital rates underpinning ecological interactions are age- and size-dependent (Bonnaffé et al. 2018;
40 Bonnaffé et al. 2021), and governed by environmental variables, such as temperature (Brown et al.

41 2004). Interactions also change following evolution of the traits that underpin them (Turchin 2003;
42 Yoshida et al. 2003). This makes it virtually impossible to model the full complexity of ecological
43 interactions (Lawton 1999; Kendall et al. 1999).

44 This leads to the second obstacle, known as structural sensitivity, namely sensitivity of the results
45 to the structure of the model (Wood 2001; Adamson and Morozov 2013). Because of the com-
46 plexity of the interactions, we often lack suitable mathematical representations to portray them
47 (Jost and Ellner 2000; Wood 2001; Ellner, Seifu, and Smith 2002; Wu, Fukuhara, and Takeda
48 2005). Parametric representations of the interactions are assumed *a priori*, which means that any
49 interaction quantified is ultimately contingent on this arbitrary choice, and hence potentially bi-
50 ased (Jost and Ellner 2000; Wood 2001; Ellner, Seifu, and Smith 2002; Wu, Fukuhara, and Takeda
51 2005). Parametric inference of ecological interactions from time series data therefore only provides
52 qualitative evidence, requiring further experimental verification and quantification (Kendall et al.
53 1999).

54 Nonparametric modelling provides a powerful alternative that can help solve these problems (e.g.
55 Jost and Ellner 2000; Wood 2001; Ellner, Seifu, and Smith 2002; Wu, Fukuhara, and Takeda 2005;
56 Pasquali and Soresina 2018). Nonparametric forms give more freedom to researchers wishing
57 to model population dynamics, and allow a test of whether the linear or linearised assumption of
58 standard models is warranted. Interactions are quantified as the sensitivity of the nonparametric ap-
59 proximation of the dynamics with respect to other state variables (Sugihara et al. 2012; Ushio et al.
60 2018). Nonparametric models require minimal assumptions regarding the mathematical nature of

61 ecological interactions (Jost and Ellner 2000; Gross, Ives, and Nordheim 2005), and hence provide
62 interaction estimates that are more robust to model structure (Wood 2001). In particular, arti-
63 ficial neural networks (ANNs) offer a promising, yet underused, nonparametric alternative to linear
64 functional forms. In previous work, we introduced a powerful framework, relying on neural ordi-
65 nary differential equations (NODEs, Chen et al. 2019) to approximate the dynamics of populations
66 nonparametrically, from which we derive ecological interactions (Bonnaffé, Sheldon, and Coulson
67 2021). More specifically, the ANNs embedded in the ODEs learn nonparametrically the shape of
68 the per capita growth rate of the populations and its dependence on the state variables of the system
69 (Bonnaffé, Sheldon, and Coulson 2021). Combined with the Geber method (Hairston et al. 2005),
70 we are able to estimate the direction, strength, and degree of nonlinearity of interactions.

71 One limitation of the approach lies in the computational cost of fitting the NODEs (Chen et al.
72 2019; Bonnaffé, Sheldon, and Coulson 2021). This is due to the fact that NODEs, as with ODEs,
73 need to be simulated over the entire range of the time series in order to compute the likelihood
74 of the trajectories of the model. This can be avoided by using gradient matching, which requires
75 interpolating the time series, and fitting the ODEs directly to the interpolated dynamics (Jost and
76 Ellner 2000; Aarts and Veer 2001; Ellner, Seifu, and Smith 2002). Although a similar approach
77 has been proposed (see Treven et al. 2021), there are no implementations of it to fitting NODEs,
78 in spite of its great potential for cutting down computational costs. In addition, given the novelty
79 of the framework, the accuracy and robustness of NODEs in estimating ecological interactions
80 remain largely unexplored. Most of the work to date is concerned with the accuracy of the fitted

81 trajectories and of the forecasts (Mai, Shattuck, and O’Hern 2016; Treven et al. 2021; Frank 2022),
82 while little attention has been given to the functional form of the processes that are producing
83 the dynamics approximated by NODEs (but see Hu et al. 2020 for a step in this direction). It
84 is important to understand to what extent the neural networks embedded within NODEs carry
85 meaningful biological information (Novak and Stouffer 2021).

86 In this manuscript, we first introduce a novel fitting technique for NODEs, Bayesian neural gradient
87 matching (BNGM). The method extends gradient matching by using neural networks to interpolate
88 the time series data instead of splines (Ellner, Seifu, and Smith 2002), and Bayesian regularisa-
89 tion to fit NODEs to the interpolated dynamics (Cawley and Talbot 2007). This cuts down the
90 fitting time of NODEs to only a few seconds, compared to about 30 minutes in our previous work
91 (Bonnaffé, Sheldon, and Coulson 2021), allowing for efficient cross-validation, and uncertainty
92 quantification. We then demonstrate that NODEs are highly accurate in recovering ecological in-
93 teractions in an artificial three-species prey-predator system where truth is known. Finally, we
94 conclude the work by characterising ecological interactions in three replicates of an experimental
95 three-species prey-predator system with an algae, flagellate, and rotifer (Hiltunen et al. 2013), as
96 well as in the classic hare and lynx time series (Odum and Barrett 1972). We find that only main
97 interactions, between the algae and the rotifer, are conserved across the three replicates, and not
98 the interactions of the flagellate with the other species. We also find that in most cases linear in-
99 teractions are sufficient to explain the dynamics apart from nonlinearity in the effect of the prey on
100 the top predator in both the rotifer and lynx.

101 **2 Material and Methods**

102 **2.1 Method overview**

103 We provide a nonparametric method for estimating ecological interactions from time series data of
104 species density. We do this by approximating the dynamics of each species with neural ordinary
105 differential equations (NODEs, Bonnaiffé, Sheldon, and Coulson 2021). We then compute ecolog-
106 ical interactions as the sensitivity of these dynamics to a change in the respective species densi-
107 ties (Sugihara et al. 2012; Bonnaiffé, Sheldon, and Coulson 2021). We provide a novel method,
108 Bayesian neural gradient matching (BNGM), allowing us to fit NODE systems in a only a few
109 seconds.

110 **2.2 Neural ordinary differential equation**

111 A NODE is a class of ordinary differential equation (ODE) that is partly or entirely defined as an ar-
112 tificial neural network (ANN) (Chen et al. 2019). They are useful to infer dynamical processes non-
113 parametrically from time series data (Bonnaiffé, Sheldon, and Coulson 2021). We choose NODEs
114 over standard statistical approaches because they offer two advantages. The first is that NODEs
115 approximate the dynamics of populations nonparametrically. NODEs are therefore not subjected
116 to incorrect model specifications (Jost and Ellner 2000; Adamson and Morozov 2013). This pro-
117 vides a more objective estimation of the inter-dependences between state variables. The second
118 advantage is that it is a dynamical systems approach. So that the approach includes lag effects
119 through interacting state variables, not only direct effects between them.

120 We first consider a general NODE system,

$$\frac{dy_i}{dt} = f_p(y, \theta_i), \quad (1)$$

121 where dy_i/dt denotes the temporal change in the i^{th} variable of the system, y_i , as a function of the
 122 other state variables $y = \{y_1, y_2, \dots, y_I\}$. The function f_p is a nonparametric function of the state
 123 variables and its shape is controlled by the parameter vector θ_i . In the context of NODEs, f_p is
 124 an ANN. The most common class of ANN used in NODEs are single-layer fully connected feed-
 125 forward ANNs (e.g. Wu, Fukuhara, and Takeda 2005), also referred to by single layer perceptrons
 126 (SLPs, e.g. Bonnaff , Sheldon, and Coulson 2021),

$$f_p(y, \theta_i) = f_\lambda \left(\theta_i^{(0)} + \sum_{j=1}^J \theta_{ij}^{(1)} f_\sigma \left(\theta_{ij}^{(2)} + \sum_{k=1}^K \theta_{ijk}^{(3)} y_k \right) \right), \quad (2)$$

127 which feature a single layer, containing J neurons, that maps the inputs, here the state variables y ,
 128 to a single output, the dynamics of state variable i , dy_i/dt . The parameter vector θ_i contains the
 129 weights $\theta^{(l)}$ of the connections in the SLPs. SLPs can be viewed as weighted sums of activation
 130 functions f_σ , which are usually chosen to be sigmoid functions $f(x) = 1/(1 + \exp(-x))$. The link
 131 function f_λ allows to map the output of the network to a specific domain, for instance applying
 132 \tanh will constrain the dynamics between -1 and 1, $dy_i/dt \in]-1, 1[$.

133 This general form can be changed to represent biological constraints on the state variables. In
 134 particular for population dynamics, the state variables are strictly positive population densities,

135 $y_i = N_i \in \mathcal{R}^+$. We could hence re-write equation (1) as, $dN_i/dt = f_p(N, \theta_i)N_i$, where the SLPs
 136 approximate the per-capita growth rate of the populations. More details regarding these models
 137 can be found in our previous work (Bonnaiffé, Sheldon, and Coulson 2021).

138 **2.3 Fitting NODEs by Bayesian neural gradient matching**

139 In this section, we describe how to estimate the parameters θ of the NODE system given a set of
 140 time series. Fitting NODEs can be highly computationally intensive, which hinders uncertainty
 141 quantification, cross-validation, and model selection (Bonnaiffé, Sheldon, and Coulson 2021). We
 142 solve this issue by introducing BNGM, a computationally efficient approach to fit NODEs. The ap-
 143 proach involves two steps (Fig. 1). First, we interpolate the state variables and their dynamics with
 144 neural networks (Fig. 1, red boxes). Second, we train each NODE to satisfy the interpolated state
 145 and dynamics (Fig. 1, blue boxes). This bypasses the costly numerical integration of the NODE
 146 system and provides a fully mathematically tractable expression for the posterior distribution of the
 147 parameter vector θ . We coin the term BNGM to emphasise two important refinements of the stan-
 148 dard gradient matching algorithm (Ellner, Seifu, and Smith 2002). The first is that we use neural
 149 networks as interpolation functions, and the second is that we use Bayesian regularisation to limit
 150 overfitting and estimate uncertainty around parameters (Cawley and Talbot 2007).

151 **Interpolating the time series**

152 The first step is to interpolate the time series and differentiate it with respect to time in order to ap-
 153 proximate the state and dynamics of the variables. We perform the interpolation via nonparametric

154 regression of the interpolating functions on the time series data,

$$Y_{it} = \tilde{y}_i(t, \omega_i) + \varepsilon_{it}^{(o)}, \quad (3)$$

155 where Y_{it} is observed value of the state variable i at time t , $\tilde{y}_i(t, \omega_i)$ is the value predicted by the
 156 interpolation function given the parameter vector ω_i , and $\varepsilon_{it}^{(o)}$ is the observation error between the
 157 observation and prediction. The interpolation function is chosen to be a neural network,

$$\tilde{y}_i(t, \omega_i) = f_\lambda \left(\omega_i^{(0)} + \sum_{j=1}^J \omega_{ij}^{(1)} f_\sigma \left(\omega_{ij}^{(2)} + \omega_{ij}^{(3)} t \right) \right), \quad (4)$$

158 where the parameter vector ω_i contains the weights $\omega^{(l)}$ of the network. We can further differentiate
 159 this expression with respect to time to obtain an interpolation of the dynamics of the state variables
 160 (Fig. 1, red boxes),

$$\frac{\partial \tilde{y}_i}{\partial t}(t, \omega_i) = \sum_{j=1}^J \omega_{ij}^{(1)} \omega_{ij}^{(3)} \frac{\partial f_\sigma}{\partial t} \left(\omega_{ij}^{(2)} + \omega_{ij}^{(3)} t \right) \frac{\partial f_\lambda}{\partial t} \left(\omega_i^{(0)} + \sum_{k=1}^J \omega_{ik}^{(1)} f_\sigma \left(\omega_{ik}^{(2)} + \omega_{ik}^{(3)} t \right) \right). \quad (5)$$

161 **Fitting NODEs to the interpolated time series**

162 The second step is to train the NODE system (Eq. 1) to satisfy the interpolated dynamics. Thanks
 163 to the interpolation step, this simply amounts to performing a nonparametric regression of each
 164 NODE (Eq. 1) on the interpolated dynamics (Eq. 5),

$$\frac{\partial \tilde{y}_i}{\partial t}(t, \omega_i) = \frac{dy_i}{dt}(\tilde{y}, \theta_i) + \varepsilon_{it}^{(p)}, \quad (6)$$

where $\varepsilon_{it}^{(p)}$ is the process error, namely the difference between the interpolated dynamics, $\partial \tilde{y}_i / \partial t$ and the NODE, dy_i / dt , given the interpolated state variables $\tilde{y} = \{\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_I\}$ (Fig. 1, blue boxes).

Bayesian regularisation

In the context of standard gradient matching, defining the observation model (Eq. 3) and process model (Eq. 6) would be sufficient to fit the NODE system (Eq. 1) to the time series via optimisation (Jost and Ellner 2000; Ellner, Seifu, and Smith 2002; Wu, Fukuhara, and Takeda 2005). We could find the parameter vector ω_i and θ_i that minimise the sum of squared observation and process errors, $\varepsilon_{it}^{(o)}$ and $\varepsilon_{it}^{(p)}$ (Eq. 3 and 6). However, this approach is prone to overfitting, and does not provide estimates of uncertainty around model predictions. To account for this, we introduce Bayesian regularisation, which allows us to control for overfitting by constraining parameters with prior distributions (Cawley and Talbot 2007), and to root our interpretation of uncertainty in a Bayesian framework.

First, we define a simple Bayesian model to fit the interpolation functions (Eq. 3) to the time series data. We assume normal distributions for the observation error, $\varepsilon_{ij}^{(o)} \sim \mathcal{N}(0, \sigma_i)$, and for the parameters, $\omega_{ij} \sim \mathcal{N}(0, \gamma_{ij})$. Here, we are only interested in interpolating the time series accurately, irrespective of the value of σ_i and γ_{ij} . Therefore, we use the approach developed by Cawley and

182 Talbot to average out the value of the parameters σ_i and γ_{ij} in the full posterior distribution (Caw-
 183 ley and Talbot 2007), assuming gamma hyperpriors $p(\xi) \propto \frac{1}{\xi} \exp\{-\xi\}$ for both parameters. This
 184 yields the following expression for the log marginal posterior density of the parameters,

$$\log P(\omega_i | Y_i) \propto -\frac{J}{2} \log \left(1 + \sum_{j=1}^J \left(\varepsilon_{ij}^{(o)} \right)^2 \right) - \frac{K}{2} \log \left(1 + \sum_{k=1}^K \omega_{ik}^2 \right) \quad (7)$$

185 where P is the marginal posterior density, $\omega_i = \{\omega_{i1}, \omega_{i2}, \dots, \omega_{iK}\}$ is the observation parameter
 186 vector controlling the interpolation function, $Y_i = \{Y_{i1}, Y_{i2}, \dots, Y_{iJ}\}$ corresponds to the sequence of
 187 observations of state variable i at time step j , J is the total number of time steps in the time series,
 188 $\varepsilon_{ij}^{(o)}$ is the observation error at time step j between the interpolated and observed value of variable
 189 i , K is the total number of parameters. More details on how to derive this expression can be found
 190 in a supplementary file (Supplementary A).

191 Then, we define a simple Bayesian model to fit the NODEs to the interpolated dynamics, given the
 192 interpolated states. We assume normal distributions for the observation error, $\varepsilon_{ij}^{(p)} \sim \mathcal{N}(0, \sigma_i)$, and
 193 parameters, $\theta_{ik} \sim \mathcal{N}(0, \delta_{ik})$. This gives the following expression for the log posterior density of
 194 the parameters given the interpolations,

$$\log p(\theta_i | \omega) \propto -\frac{1}{2} \sum_{j=1}^J \left(\frac{\varepsilon_{ij}^{(p)}}{\sigma_i} \right)^2 - \frac{1}{2} \sum_{k=1}^K \left(\frac{\theta_{ik}}{\delta_{ik}} \right)^2 \quad (8)$$

195 where $\theta_i = \{\theta_{i1}, \theta_{i2}, \dots, \theta_{iK}\}$ are the NODE parameters of the i^{th} variable, $\omega = \{\omega_1, \omega_2, \dots, \omega_I\}$ are
 196 the interpolation parameters of each state variable, $\varepsilon_{ij}^{(p)}$ is the process error of variable i at time

step j between the interpolated dynamics and NODE prediction, σ_i is the standard deviation of the likelihood, K is the total number of parameters, δ_{ik} is the standard deviation of the prior distribution of parameter θ_{ik} .

This approach allows us to limit overfitting by adjusting the constraint on the parameters, which is controlled by the standard deviation of the parameter prior distributions, δ_{ik} (Cawley and Talbot 2007; Bonnaffé, Sheldon, and Coulson 2021). We could set small values of δ to limit the degree of nonlinearity in the response, or to eliminate specific variables from the model by constraining their parameters to be close to zero. We identify the appropriate degree of constraint δ_i on NODE parameters via cross-validation. We train the NODE model on the first half of the interpolated data and predict the remaining half. We repeat this process for increasing values of δ_i , until we find the value that maximises the log likelihood of the test data.

2.4 Inference and uncertainty quantification

Finally, we estimate uncertainty in parameter values by anchored ensembling, which produces approximate Bayesian estimates of the posterior distribution of the parameters (Pearce et al. 2018). This involves sampling a parameter vector from the prior distributions, $\theta_i \sim \mathcal{N}(0, \delta_i)$, and then optimising the posterior distribution from this starting point, $\theta_i^* = \underset{\theta_i}{\operatorname{argmax}} \log p(\theta_i | \omega)$. By repeatedly taking samples, the sampled distribution θ^* approaches the posterior distribution and provides estimates and error around the quantities that can be derived from the models. The expectation and uncertainty around derived quantities can then be obtained by computing the mean and variance of

216 the approximated posterior distributions. The great strength of this approach is that it is unlikely to
 217 get stuck in local maxima hence providing a more robust optimisation of the posterior.

218 **2.5 Analysing NODEs**

219 In this study we are mainly interested in two outcomes of NODEs, namely inferring the direction
 220 (or effect) and strength (or contribution) of interactions between the state variables (Bonnaiffé,
 221 Sheldon, and Coulson 2021). We define the direction of the interaction between variable y_i and y_j as
 222 the derivative of the dynamics of y_i with respect to y_j , and vice versa (Sugihara et al. 2012),

$$e_{ijt} = \frac{\partial}{\partial y_j} \frac{dy_i}{dt}. \quad (9)$$

223 Knowing the direction, however, is not sufficient to determine the importance of a variable for the
 224 dynamics of another. Given the same effects, a variable that fluctuates a lot will have a greater
 225 impact on the dynamics of a focal variable, compared to a variable that remains quasi-constant. We
 226 hence compute the strength of the interaction by multiplying the dynamics of a variable y_j by its
 227 effect on the focal variable y_i , also known as the Geber method (Hairston et al. 2005),

$$c_{ijt} = \frac{dy_j}{dt} \frac{\partial}{\partial y_j} \frac{dy_i}{dt}. \quad (10)$$

228 To summarise results across the entire time series we can compute the mean effects e_{ij} by av-
 229 eraging e_{ijt} across all time steps, $e_{ij} = 1/K \sum_k e_{ijk}$, as well as the relative total contribution, c_{ij} ,
 230 of a variable to the dynamics of another by computing the relative sum of square contributions,

231 $c_{ij} = \left(\sum_{jk} c_{ijk}^2 \right)^{-1} \sum_t c_{ijt}^2$. By computing the direction and strength of interactions between all the
 232 variables in the system we can build dynamically informed ecological interaction networks (e.g.
 233 fig. 5). Other metrics can be computed by analysing the NODEs, such as equilibrium states, these
 234 are discussed in our previous work (Bonnaiffé, Sheldon, and Coulson 2021).

235 **3 Case studies**

236 **3.1 Case study 1: artificial tri-trophic prey-predator oscillations**

237 In this first case study, we aim to demonstrate the accuracy of the NODE fitted by BNGM in
 238 inferring nonlinear per-capita growth rates in a system where ground truth is known. Hence, we
 239 simulate a set of time series from a tri-trophic ODE model with known equations and parameters,
 240 and we compare the fitted NODEs to the actual ODEs.

241 **System**

242 We consider a tri-trophic ODE system consisting of a prey, an intermediate predator, and a top
 243 predator. The system is built on the real tri-trophic system featuring algae, flagellates, and rotifers,
 244 considered in case study 2 (Hiltunen et al. 2013),

$$\begin{aligned}
 \frac{dG}{dt} &= \left(\alpha \left(1 - \frac{G}{\kappa} \right) - \frac{\beta B}{1 + \delta G} - \frac{\gamma R}{1 + \delta G} \right) G \\
 \frac{dB}{dt} &= \left(\frac{\beta G}{1 + \delta G} - \phi R - \mu \right) B \\
 \frac{dR}{dt} &= \left(\frac{\gamma G}{1 + \delta G} + \phi B - \nu \right) R,
 \end{aligned} \tag{11}$$

where G , B , and R , correspond to the prey, intermediate, and top predator population densities, respectively, α is the prey intrinsic growth rate, limited by a carrying capacity κ , β and γ are the predation rates by the intermediate and top predator, δ is the saturation rate of prey predation, which emulates the capacity of the algae to display predator defense at higher algal density (Hiltunen et al. 2013), ϕ is the predation rate of the intermediate predator by the top predator, μ and ν are the intrinsic mortality of the intermediate and top predator.

We simulate a case of invasion, by introducing the top predator at a low density, with a set of parameters that result in dampening prey-predator oscillations, namely $\alpha = 1$, $\beta = 2.5$, $\gamma = 1.5$, $\kappa = 3$, $\delta = \phi = \mu = \nu = 1$. We focus on the middle section of the time series, $t \in [20, 50]$, as in the initial section the top predator is rare, and in the later section populations have attained a fixed equilibrium point. The resulting time series are presented in figure 2.

256 NODE model

In order to nonparametrically learn the per-capita growth rate of each species, and to derive ecological interactions, we define a three-species NODE system,

$$\begin{aligned}\frac{dR}{dt} &= r_R(R, G, B, \beta_R)R \\ \frac{dG}{dt} &= r_G(R, G, B, \beta_G)G \\ \frac{dB}{dt} &= r_B(R, G, B, \beta_B)B,\end{aligned}\tag{12}$$

where the per-capita growth rates r_R , r_G , and r_B are neural network functions of the density R , G , B of each species (function f_p , Eq. 2). We choose a combination of linear and exponential activation

261 functions $f_{\sigma, j \leq J/2}(x) = x$, and $f_{\sigma, j > J/2}(x) = \exp(x)$. This allows us to progressively switch from a
 262 simple linear model to a nonlinear model by releasing the constraint on the exponential section of
 263 the neural network during cross-validation. The number of units in the hidden layer J is chosen to
 264 be 10, as this is a commonly used number for systems of that size (e.g. Wu, Fukuhara, and Takeda
 265 2005; Bonnaffé, Sheldon, and Coulson 2021).

266 **Time series interpolation**

267 We interpolate the time series using the neural network described in section 2.3 (Eq. 4). We set
 268 the number of neurons in the network to $J = 30$. We use sinusoid activation functions, $f_{\sigma}(x) =$
 269 $\sin(x)$, so that the weights $\omega_{ij}^{(1)}$, $\omega_{ij}^{(2)}$, and $\omega_{ij}^{(3)}$ control the amplitude, shift, and frequency of the
 270 oscillations in the time series, respectively. Given that the population densities are strictly positive
 271 $R, G, B \in \mathcal{R}^+$, we use an exponential link function, $f_{\lambda}(x) = \exp(x)$. We then approximate the
 272 marginal posterior distribution of the interpolation parameters, and thereby of interpolated states
 273 and dynamics, by taking 100 samples from the log marginal posterior distribution (Eq. 7) via
 274 anchored ensembling. In practice, the high number of parameters in the neural network equation
 275 may impede the fit of the time series, especially for short time series. We found that dividing the
 276 number of parameters K (Eq. 7) by the number of neurons in the network J (Eq. 2) yields consistent
 277 fitting results. Interpolated states and dynamics are presented in figure 2.

278 **Fitting NODEs to the interpolated time series**

279 We fit the NODE system to the interpolated time series. In practice, we fit the NODE to the expect-
 280 ation of the interpolated state and dynamics, $E(\tilde{y}_i)$ and $E(d\tilde{y}_i/dt)$, by averaging over all sampled

281 interpolation parameters. An alternative approach could be to consider the interpolation that max-
 282 imises the log marginal posterior density, but this may decrease repeatability due to the difficulty of
 283 reliably identifying a global maximum. Averaging across multiple interpolations ensures an overall
 284 smoother and robust interpolation. In addition, we standardise the response and explanatory vari-
 285 ables with respect to their mean and standard deviation (i.e. $Z = (Y - \mu)/\sigma$). This is to facilitate
 286 the training of the NODE by equalizing the scale of the different parameters in the neural network.
 287 Then, we identify the optimal regularisation parameter δ (Eq. 8) by cross-validation. To do that,
 288 we split the data in half, train NODEs on the first half, and calculate the log likelihood of the test
 289 set for increasing values of δ , from 0.05 (linear) to 0.5 (highly nonlinear), by increments of 0.05.
 290 This allows us to identify the maximum degree of nonlinearity, δ , in the per-capita growth rate that
 291 ensures generalisability throughout the time series. Then, we approximate the posterior distribu-
 292 tion of the NODE parameters by taking 30 samples from the posterior distribution (Eq. 8). Finally,
 293 we perform model selection by removing variables that do not result in a significant decrease in
 294 the log likelihood of the model (assessed by comparing log likelihood confidence intervals). We
 295 ensure moderate temporal autocorrelation and normality by visualising the residuals of the models.
 296 We also ensure results repeatability by running the entire fitting process a second time.

297 **Computing ecological interactions**

298 Finally, we analyse the shape of the per-capita growth rates to recover the interaction between the
 299 three species in the system. In particular, we look at the effect and contribution of each species
 300 to the dynamics of the others. The effect is computed as the sensitivity (i.e. the gradient) of the

per-capita growth rate of a given species with respect to the density of the other species (Sugihara et al. 2012; Bonnaiffé, Sheldon, and Coulson 2021). The contribution is computed following the Geber method (Hairston et al. 2005), which consists in multiplying the dynamics of a variable by its effects on the other variables. We further compute the importance of a species in driving the dynamics of another by computing its relative total contribution compared to other species. More details on how to compute these quantities can be found in section 2.5 and in our previous study (Bonnaiffé, Sheldon, and Coulson 2021).

3.2 Case study 2: real tri-trophic prey-predator oscillations

In this second case study, we want to assess the quality of the NODE analysis when performed on a real time series. We are further interested in comparing the direction and strength of uncovered ecological interactions across virtually identical replicated time series.

System

We consider a three-species laboratory microcosm consisting of an algal prey (*Chlorella autophylla*), a flagellate intermediate predator (*Oxyrrhis marina*), and a rotifer top predator (*Brachionus plicatilis*). The algal prey is consumed by the intermediate and top predator, which also consumes the intermediate predator (Arndt 1993). The dynamics of this system, here the daily change in the density of each species, were recorded in three replicated time series experiments performed by Hiltunen and colleagues (Hiltunen et al. 2013). We use their time series because they describe a simple yet biologically realistic ecosystem, and because the quality of the replication of their

microcosm reduces as much as possible observational and experimental error, and rules out environmental variation (Hiltunen et al. 2013). We digitised these time series by extracting by hand the coordinates of every points in the referential of the axis of the graph of the original study, and analysed them.

NODE analysis

We apply the same analysis as performed on the artificial tri-trophic prey-predator oscillations. This allows us to recover a nonparametric approximation of the growth rate of each species, and then derive the direction and strength of the ecological interactions that underpin their dynamics. We present detailed results of the analysis of the first time series (Fig. 4), and a summary comparison of the three time series (Fig. 5). Complementary results, including cross-validation plots, and detailed results for the other two replicates can be found in the supplementary material (Supplementary B-E).

3.3 Case study 3: real di-trophic prey-predator oscillations

Finally, we infer ecological interactions by NODE BNGM in the hare-lynx system (Odum and Barrett 1972). This is to provide an example of a longer time series, and to offer a point of comparison with previous and future implementations of NODEs, which commonly use this time series (e.g. Bonnaffé, Sheldon, and Coulson 2021; Frank 2022).

System

The system is described in details in our previous work (Bonnaffé, Sheldon, and Coulson 2021).

339 The data consist in a 90-year long time series of counts of hare and lynx pelts collected by trappers
340 in the Hudson bay area in Canada (Odum and Barrett 1972). The time series displays characteristic
341 10-year long prey-predator oscillations.

342 **NODE analysis**

343 We apply the same analysis as previously described, to the exception that the NODE system only
344 features two variables, H and L , instead of 3. Results are presented in figure 6.

345 **4 Results**

346 **4.1 Model runtimes**

347 We present a breakdown of the runtime of fitting NODEs by BNGM for each system in table
348 1. We find that it takes on average 5.35 minutes to fit NODEs by BNGM. This includes taking
349 390 samples, and thereby performing 390 full optimisations, of the posterior distribution of the
350 interpolation and NODE parameters. This amounts to about 5.37 second to sample each variable
351 of the NODE system once. This is a 335 fold improvement over our previous approach, which took
352 on average 30 minutes (Bonnaffé, Sheldon, and Coulson 2021).

353 **4.2 Case study 1: artificial tri-trophic system**

354 We present the results of fitting NODEs by BNGM to the artificial tri-trophic time series in figure
355 2 and 3. We find that both the interpolation of the state variables and dynamics are highly accurate
356 (Fig. 2), given that they closely match the ground truth, known from the equations of the ODE

357 model that we used to generate the time series (Eq. 11). Similarly, we find that the NODE approx-
 358 imation of the per-capita growth rate of each species also closely matches the ground truth (Fig. 3,
 359 a., d., g.). We find negative nonlinear effects of the two predators on the growth rate of the algae
 360 (Fig. 3, b., blue and purple lines). This nonlinear pattern is mirrored by the effect of the algae
 361 on the growth rate of the predators (Fig. 3, e. and h., red line). The linear interaction between
 362 the two predators is also well-recovered (Fig. 3, e., blue line, and h., purple line). We find that
 363 removing the intra-specific dependence in the growth rate of the predators did not affect the fit of
 364 the model (Fig. e., purple line, and h., blue line). The BNGM approach hence accurately recovers
 365 the dynamical characteristics of the artificial system.

366 **4.3 Case study 2: real tri-trophic prey-predator oscillations**

367 We present the in-depth analysis of the drivers of the dynamics of the algae, flagellate, and rotifer
 368 population in replicate A (Fig. 4). Cross-validation reveals that there is no support for nonlinear
 369 effects in the growth rate of the algae and flagellate for replicate A (Fig. 4, a. and b., d. and e.).
 370 We find negative linear intra-specific density dependence in algal growth (Fig. 4, b., red line), and
 371 negative linear inter-specific effects of the two predators (purple and blue line). We find that the
 372 growth rate of the flagellate is virtually solely driven by predation by the rotifer (Fig. 4, e. and
 373 f., blue line). The rotifer population itself is driven by a positive nonlinear effect of both preys
 374 (Fig. h., red and purple line). There is also evidence for positive nonlinear intra-specific density
 375 dependence (Fig. h., blue line). Overall, comparing results across the three replicates reveals that
 376 the effect of the rotifer population on the flagellate and algae, and the effect of the algae on the

377 rotifer, are the strongest and most consistent interactions (Fig. 5, table 2). The interactions of the
378 flagellate with the algae, and its effect on the rotifer population varies substantially across replicates
379 (Fig. 5, table 2). Interestingly, intra-specific density dependence in rotifer and algae is also found
380 to be inconsistent across the three replicates.

381 **4.4 Case study 3: real di-trophic prey-predator oscillations**

382 Finally, we present the analysis of the drivers of the hare-lynx population dynamics in figure 6.
383 Cross-validation provides support for nonlinear effects in the per-capita growth rate of the hare and
384 lynx. We find that the hare population growth rate is mostly determined by a nonlinear negative
385 effect of the lynx population (Fig. 6, b. and c. blue line), and by weak nonlinear positive density
386 dependence (red line). The lynx growth rate is determined by a positive nonlinear effect of the
387 hare (Fig. 6, e. and f., red line), and to a lesser extent by negative nonlinear intra-specific density
388 dependence (blue line).

389 **5 Discussion**

390 Characterising ecological interactions from time series data is challenging. This is due to the fact
391 that interactions can be highly context-dependent processes (Song and Saavedra 2021), making it
392 difficult to identify parametric models that encapsulate their complexity (Wood 2001). Interactions
393 estimated with parametric models are contingent on the parameterisation arbitrarily chosen by the
394 observer, and hence risk being biased (Wood 2001; Adamson and Morozov 2013). We provide

395 a novel method for estimating ecological interactions nonparametrically, by using neural ordinary
396 differential equations (NODEs) fitted with Bayesian neural gradient matching (BNGM). First, we
397 remove the cost of fitting NODEs by introducing BNGM, which allows for NODE fitting in only
398 a few seconds. The method involves interpolating time series and dynamics with neural networks,
399 and then fitting NODEs to interpolated dynamics with Bayesian regularisation. We further demon-
400 strate that this approach is accurate, as NODEs approximate with minimal error the ecological
401 interactions in artificial time series, where real interactions are known. Finally, we estimate the
402 strength, direction, importance, and nonlinearity of ecological interactions in 4 time series from
403 natural and experimental systems, showing variation in ecological interactions within and across
404 the time series.

405 **Bayesian neural gradient matching**

406 The Bayesian neural gradient matching (BNGM) approach that we propose here extends stan-
407 dard gradient matching, by using artificial neural networks (ANNs) as interpolating functions, and
408 Bayesian regularisation to control the nonlinearity of the processes (Cawley and Talbot 2007). This
409 allows us to accurately fit NODEs within seconds, making it potentially the most efficient current
410 fitting technique available (see also Treven et al. 2021). The use of ANNs as interpolating func-
411 tions sets it apart from the initial approach of Ellner et al., who use splines to interpolate the time
412 series before approximating the ODEs (Ellner, Seifu, and Smith 2002). ANNs are more general
413 and flexible than splines, as well as being easier to manipulate given that they are defined contin-
414 uously on the state space, which is especially useful when handling multiple interactions between

variables. Our approach is related to that of Wu et al., who use ANNs to approximate both the states and ODEs of prey-predator systems (Wu, Fukuhara, and Takeda 2005), as well as that of Treven and colleagues, who developed the Gaussian process equivalent (Treven et al. 2021). In both approaches, they train the interpolation functions at the same time as the NODEs, in order to constrain the interpolation of trajectories such that they can be achieved by the NODE system, which thereby introduces dynamical coupling between state variables. One of the risks of dynamical coupling approaches is that misestimating one of the state variables of the model biases the estimation of the states and dynamics of other variables. To avoid this, we fit each interpolation and NODE independently to each time series. In addition, this makes it possible to parallelise the code, resulting in potentially even faster computation. Our approach removes the main limitation of using NODEs, allowing for quick and extensive model comparison, cross-validation, and uncertainty quantification around estimates.

Accuracy of NODEs in estimating ecological interactions

Our approach relies on approximating population dynamics with NODEs and then computing their sensitivity to a change in the density of the different populations in the system (Bonnaffé, Sheldon, and Coulson 2021). We demonstrate that NODEs accurately recover the dynamics, strength, direction, and nonlinearity of ecological interactions in artificial tri-trophic prey-predator time series, where truth is known. In particular, we find that the interactions between the algae and the two predators are nonlinear, and thereby oscillate throughout the time series, which is consistent with the model, that features a resistance to predation at high algal density. We also recover the

linear interactions between the two predators, which shows that the NODEs are sensitive enough to discriminate between linear and nonlinear interactions within and across time series. To our knowledge, this is the first assessment of the accuracy of NODEs in recovering interactions between variables from time series data, as most of the work focuses on assessing the accuracy of the fitting and forecasting of time series (e.g. Mai, Shattuck, and O'Hern 2016; Chen et al. 2019; Treven et al. 2021; Frank 2022).

Ecological interactions in real prey-predator systems

We further tested NODEs in a real setting, by inferring ecological interactions across three replicated time series of an experimental tri-trophic system of algae, flagellate, and rotifer populations (Hiltunen et al. 2013). Our approach reveals that only stronger interactions, namely the negative effects of the rotifer top predator on the other species, and the positive effect of algae on the rotifer, are conserved across the three replicated time series. We also find evidence for nonlinearity in the dynamics of the rotifer, as the positive effect of the algae on rotifer growth oscillates throughout the time series. This is consistent with the biology of the system, as the algae tends to form anti-predation clumps at higher density, which would dampen the positive effect of algal density on rotifer growth at high algal density (Yoshida et al. 2003; Hiltunen et al. 2013). We find it interesting that the weaker interactions with the flagellate predator are not consistent across time series, given the controlled laboratory conditions. This system is known to evolve rapidly, it is hence possible that fast evolution of the different populations from the onset of the time series may have driven the system onto different attractors (Yoshida et al. 2003; Yoshida et al. 2007; Hiltunen et al. 2013).

455 Additionally, stochasticity in population dynamics may have a similar effect (Dallas et al. 2021).
456 Disentangling these two sources of variation would require refining the modelling framework, for
457 instance by explicitly including evolution in the model (e.g. with the Price equation, Ellner, Geber,
458 and Hairston 2011), and by using neural stochastic differential equations (i.e. NSDEs, Rackauckas
459 et al. 2019) fitted with a particle filter. While these would constitute interesting developments, our
460 method is still a useful first step, identifying differences between the time series, and demonstrating
461 a reasonable amount of deterministic consistency in the dynamics, judging by the cross-validation
462 and fits.

463 We also analysed the hare-lynx time series (Odum and Barrett 1972), as it is a common benchmark
464 in the field of time series analysis, and provides a comparison point with our previous implemen-
465 tation of NODEs (Bonnaiffé, Sheldon, and Coulson 2021). As in our previous study, we found a
466 predatory inter-specific interaction between lynx and hare, and negative intra-specific density de-
467 pendence in the lynx. Evidence for positive density dependence in the hare was more limited than
468 previously found. We also found stronger evidence for nonlinearity, as intra- and inter-specific ef-
469 fects oscillated throughout the time series, as a result of density dependence. This difference with
470 our previous study is due to the fact that our previous implementation of NODEs was based on sim-
471 ulating the full NODE system, and hence imposed dynamical coupling between the variables. This
472 dynamical coupling comes at a cost, if one variable is not explained well by the model, it will bias
473 the interactions and dynamics of other variables. Here, the time series of lynx and hare are analysed
474 independently, each state variable is interpolated as closely as desired, its effects on the dynamics

of other variables are hence even more robust to model misspecification than before.

Overall, our approach provides a novel and powerful way of estimating interactions nonparametrically from time series data. The benefit of using NODEs is that they make no assumptions about the nature of the ecological interactions that drive the dynamics of the species (Chen et al. 2019; Bonnaffé, Sheldon, and Coulson 2021). Hence, we have a better chance at estimating the actual value of the interactions, knowing that it is not subjected to potential incorrect model specifications (Jost and Ellner 2000; Ellner, Seifu, and Smith 2002; Wu, Fukuhara, and Takeda 2005; Kendall et al. 2005; Adamson and Morozov 2013). This approach is similar to Sugihara’s maps (S-maps, Sugihara et al. 2012), which estimate interactions in time series by approximating the Jacobian matrix nonparametrically via locally linear approximations of the state space (Deyle et al. 2015). However, because S-maps are locally linear, they do not assume the existence of a latent trajectory generated by an overarching model. This creates two caveats, the first being that they are more sensitive to noise in the time series (Cenci and Saavedra 2018), the second being that they have no theoretical grounding given that they are at heart linear functions defined piecewise on the state space. NODEs remain in essence deterministic ODE models, assuming an overarching model driving the populations through the entire state-space, which can hence incorporate parametric assumptions regarding the driving processes (Bonnaffé, Sheldon, and Coulson 2021). For instance, we model the per capita growth rate of populations explicitly in NODEs, while S-maps approximate the population-level growth. Overall, this makes NODEs more suitable than S-maps for fitting noisy data or exploring theory by testing specific assumptions.

495 **Limits and prospects**

496 One of the main difficulty in quantifying ecological interactions is to identify potential context de-
497 pendences on other state variables (Song and Saavedra 2021), for example, whether predation rates
498 are affected by temperature. Our approach allows for the quantification of context dependence,
499 which shows as nonlinear fluctuations of interactions throughout the time series. In the present
500 work, we only report nonlinearity as evidence for context dependence in the interactions, but we
501 do not attempt to understand what it is attributable to. For instance, we identify nonlinear density
502 dependence in the effect of the algae on the rotifer, but we do not know whether this is due to a
503 change in the effect with algae density or rotifer density, or both. In order to disentangle these
504 higher order effects we could compute the Hessian of the system, namely the second order deriva-
505 tive of the dynamics with respect to the different state variables. Though this procedure is simple
506 mathematically, it would result in 27 second order effects to analyse for the simple 3 species system
507 considered here. This type of analysis would get rapidly out of hand for larger systems. Further
508 work should hence consider how to handle these higher order effects, as a way to unveil context
509 dependence in ecological interactions.

510 One further issue is that some interactions may depend on variables that are not observed. For
511 instance, some population dynamics are strongly determined by their demographic state (Lande et
512 al. 2002; Coulson et al. 2004), which would call for time series of the relevant demographic stages.
513 In the system considered here, the dynamics of algae in the rotifer system are most likely coupled
514 with that of nitrogen, for which no time series was available (Hiltunen et al. 2013). Our method

only accounts for observed variables, so that time series for all important variables are required, though unaccounted variables are captured to some extent by nonlinear fluctuations in interactions. One interesting prospect would hence be to incorporate unobserved/latent state variables into the NODE system (Dupont, Doucet, and Teh 2019; Zhang et al. 2019; Frank 2022). Careful thought has to be given here as whether to use an ODE or NODE for the latent states given that they are not constrained by observations.

We consider NODEs, which are only defined along the time dimension. The framework could easily be extended to any other dimension by considering partial differential equations instead (Rackauckas et al. 2019). For instance, in a spatial ecology context we could model the dynamics of populations along two additional spatial dimensions. In an evolutionary context, we could model the dynamics of populations in phenotype space, by adding phenotypic traits as an additional dimension. The BNGM method could be instrumental in fitting these models, which are notoriously expensive to stimulate.

Conclusion

We provide a method, BNGM, which allows for NODE fitting in a matter of seconds. This is a crucial step for efficient model selection and uncertainty quantification in NODEs. We also demonstrate that NODEs allow for accurate estimation of the direction, strength, and nonlinearity of ecological interactions, in a system where truth is known. Finally, we estimate ecological interactions in real prey predator systems, showing that system dynamics are driven by a mixture of linear and nonlinear interactions, of which only strong ones seem to be generalisable across

time series. Our study allows for efficient NODE fitting, and confirms the power of NODEs in identifying dynamical coupling between populations.

Acknowledgments

We thank warmly the Ecological and Evolutionary Dynamics Lab and Sheldon Lab Group at the department of Zoology for their feedback and support. We thank Ben Sheldon for insightful suggestions on early versions of the work. The work was supported by the Oxford-Oxitec scholarship and the NERC DTP.

Data accessibility

All data and code will be made fully available at <https://github.com/WillemBonnafe/NODER/BNGM>, as well as on <https://datadryad.org/stash/dataset/doi:xxx>.

Statement of authorship

Willem Bonnaffé designed the method, performed the analysis, wrote the manuscript; Tim Coulson led investigations, provided input for the manuscript, commented on the manuscript.

References

- Aarts, L. P. and P. V. D. Veer (2001). “Neural network method for solving partial differential equations”. In: *Neural Processing Letters* 14 (3), pp. 261–271.
- Adams, M. P. et al. (Apr. 2020). “Informing management decisions for ecological networks, using dynamic models calibrated to noisy time-series data”. In: *Ecology Letters* 23 (4), pp. 607–619.

553 Adamson, M. W. and A. Y. Morozov (2013). “When can we trust our model predictions? Un-
 554 earthing structural sensitivity in biological systems”. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 469 (2149), pp. 1–19.

556 Arndt, H. (1993). “Rotifers as predators on components of the microbial web (bacteria, heterotrophic
 557 flagellates, ciliates) - a review”. In: *Hydrobiologia* 255-256 (1), pp. 231–246.

558 Berryman, A. and P. Turchin (1997). “Detection of density dependence: comment”. In: *Ecology* 78
 559 (1), pp. 318–320.

560 Berryman, A. A. (2002). “Population: a central concept for ecology?” In: *Oikos* 97 (3), pp. 439–
 561 442.

562 Berryman, A. A. (2003). “On principles, laws and theory in population ecology”. In: *Oikos* 103
 563 (3), pp. 695–701.

564 Bonnaiffé, W., S. Legendre, A. Danet, and E. Edeline (2021). “Comparison of size-structured and
 565 species-level trophic networks reveals antagonistic effects of temperature on vertical trophic
 566 diversity at the population and species level”. In: *Oikos*, pp. 1–14.

567 Bonnaiffé, W., M. Martin, M. Mugabo, S. Meylan, and J. F. L. Galliard (Dec. 2018). “Ontogenetic
 568 trajectories of body coloration reveal its function as a multicomponent nonsenescent signal”. In:
 569 *Ecology and Evolution* 8 (24), pp. 12299–12307.

570 Bonnaiffé, W., B. C. Sheldon, and T. Coulson (2021). “Neural ordinary differential equations for
 571 ecological and evolutionary time series analysis”. In: *Methods in Ecology and Evolution* 2, pp. 1–
 572 46.

573 Bonsall, M. B., E. V. D. Meijden, and M. J. Crawley (2003). “Contrasting dynamics in the same
574 plant-herbivore interaction”. In: *Proceedings of the National Academy of Sciences of the United*
575 *States of America* 100 (25), pp. 14932–14936.

576 Brook, B. W. and C. J. A. Bradshaw (2006). “Strength of evidence for density dependence in
577 abundance time series of 1198 species”. In: *Ecology* 87 (6), pp. 1445–1451.

578 Brown, J. H., J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B. West (2004). “Toward a metabolic
579 theory of ecology”. In: *Ecology* 85 (7), pp. 1771–1789.

580 Bruijning, M., E. Jongejans, and M. M. Turcotte (2019). “Demographic responses underlying eco-
581 evolutionary dynamics as revealed with inverse modelling”. In: *Journal of Animal Ecology* 88
582 (5), pp. 768–779.

583 Cawley, G. C. and N. L. C. Talbot (2007). “Preventing over-fitting during model selection via
584 bayesian regularisation of the hyper-parameters”. In: *Journal of Machine Learning Research* 8,
585 pp. 841–861.

586 Cenci, S. and S. Saavedra (Oct. 2018). “Uncertainty quantification of the effects of biotic interac-
587 tions on community dynamics from nonlinear time-series data”. In: *Journal of the Royal Society*
588 *Interface* 15 (147).

589 Chen, R. T. Q., Y. Rubanova, J. Bettencourt, and D. Duvenaud (2019). “Neural Ordinary Differen-
590 tial Equations”. In: *arXiv*, pp. 1–19.

591 Coulson, T., F. Guinness, J. Pemberton, and T. Clutton-Brock (2004). “The demographic conse-
592 quences of releasing a population of red deer from culling”. In: *Ecology* 85 (2), pp. 411–422.

593 Dallas, T., B. A. Melbourne, G. Legault, and A. Hastings (2021). “Initial abundance and stochas-
594 ticity influence competitive outcome in communities”. In: *Journal of Animal Ecology*, pp. 1–
595 26.

596 Deyle, E. R., R. M. May, S. B. Munch, and G. Sugihara (Jan. 2015). “Tracking and forecasting
597 ecosystem interactions in real time”. In: *Proceedings of the Royal Society B: Biological Sciences*
598 283, pp. 1–9.

599 Dupont, E., A. Doucet, and Y. W. Teh (2019). “Augmented Neural ODEs”. In: *arXiv*, pp. 1–11.

600 Ellner, S. P., M. A. Geber, and N. G. J. Hairston (2011). “Does rapid evolution matter? Measuring
601 the rate of contemporary evolution and its impacts on ecological dynamics”. In: *Ecology Letters*
602 14 (6), pp. 603–614.

603 Ellner, S. P., Y. Seifu, and R. H. Smith (2002). “Fitting Population Dynamic Models to Time-Series
604 Data by Gradient Matching”. In: *Ecology* 83 (8), p. 2256.

605 Frank, S. A. (2022). “Automatic differentiation and the optimization of differential equation models
606 in biology”. In: *arXiv*, pp. 1–10.

607 Gross, K., A. R. Ives, and E. V. Nordheim (2005). “Estimating fluctuating vital rates from time-
608 series data: A case study of aphid biocontrol”. In: *Ecology* 86 (3), pp. 740–752.

609 Hairston, N. G. J., S. P. Ellner, M. A. Geber, T. Yoshida, and J. A. Fox (2005). “Rapid evolution
610 and the convergence of ecological and evolutionary time”. In: *Ecology Letters* 8 (10), pp. 1114–
611 1127.

612 Hiltunen, T., L. E. Jones, S. P. Ellner, and N. G. J. Hairston (2013). “Temporal dynamics of a simple
613 community with intraguild predation: an experimental test”. In: *Ecology* 94 (4), pp. 773–779.

614 Hu, P., W. Yang, Y. Zhu, and L. Hong (2020). “Revealing hidden dynamics from time-series data
615 by ODENet”. In: *arXiv*, pp. 1–17.

616 Ives, A. R., B. Dennis, K. L. Cottingham, and S. R. Carpenter (2003). “Estimating community
617 stability and ecological interactions from time-series data”. In: *Ecological Monographs* 73 (2),
618 pp. 301–330.

619 Jonzén, N., P. Lundberg, E. Ranta, and V. Kaitala (Feb. 2002). “The irreducible uncertainty of
620 the demography - Environment interaction in ecology”. In: *Proceedings of the Royal Society B:
621 Biological Sciences* 269 (1488), pp. 221–225.

622 Jost, C. and S. P. Ellner (2000). “Testing for predator dependence in predator-prey dynamics: A
623 non-parametric approach”. In: *Proceedings of the Royal Society B: Biological Sciences* 267
624 (1453), pp. 1611–1620.

625 Kendall, B. E. et al. (1999). “Why do populations cycle? A synthesis of statistical and mechanistic
626 modeling approaches”. In: *Ecology* 80 (6), pp. 1789–1805.

627 Kendall, B. E. et al. (2005). “Population cycles in the pine looper moth: Dynamical tests of mech-
628 anistic hypotheses”. In: *Ecological Monographs* 75 (2), pp. 259–276.

629 Lande, R, S. Engen, B.-E. Saether, F. Filli, E. Matthysen, and H. Weimerskirch (2002). “Estimating
630 Density Dependence from Population Time Series Using Demographic Theory and Life-History
631 Data”. In: *American Naturalist* 159, pp. 321–337.

632 Lawton, J. H. (1999). “Are There General Laws in Ecology ?” In: *Oikos* 84 (2), pp. 177–192.

633 Lingjaerde, O. C. et al. (2001). “Exploring the density-dependent structure of blowfly populations
634 by nonparametric additive modeling”. In: *Ecology* 82 (9), pp. 2645–2658.

- 635 Mai, M., M. D. Shattuck, and C. S. O’Hern (2016). “Reconstruction of Ordinary Differential Equations From Time Series Data”. In: *arXiv*, pp. 1–15.
- 636
- 637 Moe, S. J., A. B. Kristoffersen, R. H. Smith, and N. C. Stenseth (2005). “From patterns to processes and back: Analysing density-dependent responses to an abiotic stressor by statistical and mechanistic modelling”. In: *Proceedings of the Royal Society B: Biological Sciences* 272 (1577), pp. 2133–2142.
- 638
- 639
- 640
- 641 Mysterud, A., T. Coulson, and N. C. Stenseth (2002). “The role of males in the dynamics of ungulate populations”. In: *Journal of Animal Ecology* 71, pp. 907–915.
- 642
- 643 Novak, M. and D. B. Stouffer (Nov. 2021). “Geometric Complexity and the Information-Theoretic Comparison of Functional-Response Models”. In: *Frontiers in Ecology and Evolution* 9.
- 644
- 645 Odum, E. P. and G. W. Barrett (1972). “Fundamentals of Ecology”. In: *The Journal of Wildlife Management* 36 (4), p. 1372.
- 646
- 647 Pasquali, S. and C. Soresina (2018). “Estimation of the mortality rate functions from time series field data in a stage-structured demographic model for *Lobesia botrana*”. In: *arXiv*, pp. 1–15.
- 648
- 649 Pearce, T., F. Leibfried, A. Brintrup, M. Zaki, and A. Neely (2018). “Uncertainty in Neural Networks: Approximately Bayesian Ensembling”. In: *arXiv*, pp. 1–10.
- 650
- 651 Rackauckas, C., M. Innes, Y. Ma, J. Bettencourt, L. White, and V. Dixit (Feb. 2019). “DiffEqFlux.jl - A Julia Library for Neural Differential Equations”. In: *arXiv*, pp. 1–17.
- 652
- 653 Rosenbaum, B., M. Raatz, G. Weithoff, G. F. Fussmann, and U. Gaedke (2019). “Estimating parameters from multiple time series of population dynamics using bayesian inference”. In: *Frontiers in Ecology and Evolution* 6 (234), pp. 1–14.
- 654
- 655

656 Royama, T (1984). “Population Dynamics of the Spruce Budworm *Choristoneura fumiferana*”. In:
 657 *Ecological Monographs* 54 (4), pp. 429–462.

658 Song, C., S. V. Ahn, R. P. Rohr, and S. Saavedra (May 2020). “Towards a Probabilistic Under-
 659 standing About the Context-Dependency of Species Interactions”. In: *Trends in Ecology and*
 660 *Evolution* 35 (5), pp. 384–396.

661 Song, C. and S. Saavedra (July 2021). “Bridging parametric and nonparametric measures of species
 662 interactions unveils new insights of non-equilibrium dynamics”. In: *Oikos* 130 (7), pp. 1027–
 663 1034.

664 Sugihara, G. et al. (2012). “Detecting causality in complex ecosystems”. In: *Science* 338 (6106),
 665 pp. 496–500.

666 Treven, L., P. Wenk, F. Dörfler, and A. Krause (2021). “Distributional Gradient Matching for Learn-
 667 ing Uncertain Neural Dynamics Models”. In: *arXiv*, pp. 1–14.

668 Turchin, P. (1999). “Population Regulation: A Synthetic View”. In: *Oikos* 84 (1), pp. 153–159.
 669 – (2001). “Does population ecology have general laws?” In: *Oikos* 94, pp. 17–26.
 670 – (2003). “Evolution in population dynamics”. In: *Nature* 424, pp. 257–258.

671 Ushio, M. et al. (Feb. 2018). “Fluctuating interaction network and time-varying stability of a natural
 672 fish community”. In: *Nature* 554 (7692), pp. 360–363.

673 Wood, S. N. (2001). “Partially specified ecological models”. In: *Ecological Monographs* 71 (1),
 674 pp. 1–25.

675 Wu, J., M. Fukuhara, and T. Takeda (2005). “Parameter estimation of an ecological system by
676 a neural network with residual minimization training”. In: *Ecological Modelling* 189 (3-4),
677 pp. 289–304.

678 Yoshida, T., S. P. Ellner, L. E. Jones, B. J. M. Bohannan, R. E. Lenski, and N. G. J. Hairston (2007).
679 “Cryptic population dynamics: Rapid evolution masks trophic interactions”. In: *PLoS Biology* 5
680 (9), pp. 1868–1879.

681 Yoshida, T., L. E. Jones, S. P. Ellner, G. F. Fussmann, and N. G. J. Hairston (2003). “Rapid evo-
682 lution drives ecological dynamics in a predator – prey system”. In: *Nature* 424 (July), pp. 303–
683 306.

684 Zhang, H., X. Gao, J. Unterman, and T. Arodz (July 2019). “Approximation Capabilities of Neural
685 ODEs and Invertible Residual Networks”. In: *arXiv*, pp. 1–11.

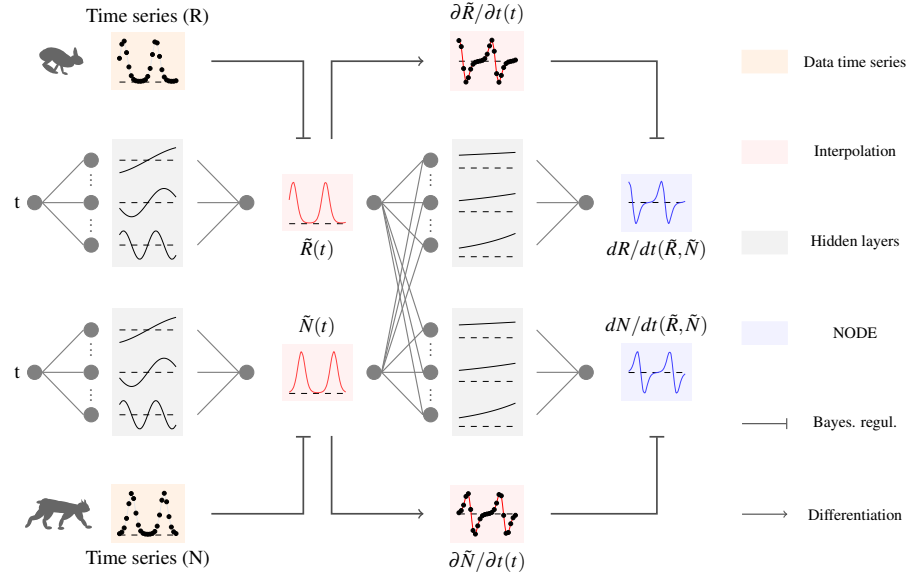


Figure 1: Overview of fitting neural ordinary differential equations (NODE) by Bayesian neural gradient matching (BNGM). In a first step we compute a continuous time approximation (interpolation) of each state variables, here the prey $\tilde{R}(t)$ and predator density $\tilde{N}(t)$ (red boxes). To do that we fit an ANN, that takes time as input, to each time series, via Bayesian regularisation. Interpolated dynamics of populations can then be computed by taking the derivative of the ANN with respect to time, $\partial \tilde{R} / \partial t$ and $\partial \tilde{N} / \partial t$. In a second step, we fit each NODE, dR/dt and dN/dt (blue boxes), to the interpolated dynamics. To do that we fit an ANN, which takes as input the interpolated variables $\tilde{R}(t)$ and $\tilde{N}(t)$, to the interpolated dynamics $\partial \tilde{R} / \partial t$ and $\partial \tilde{N} / \partial t$, via Bayesian regularisation.

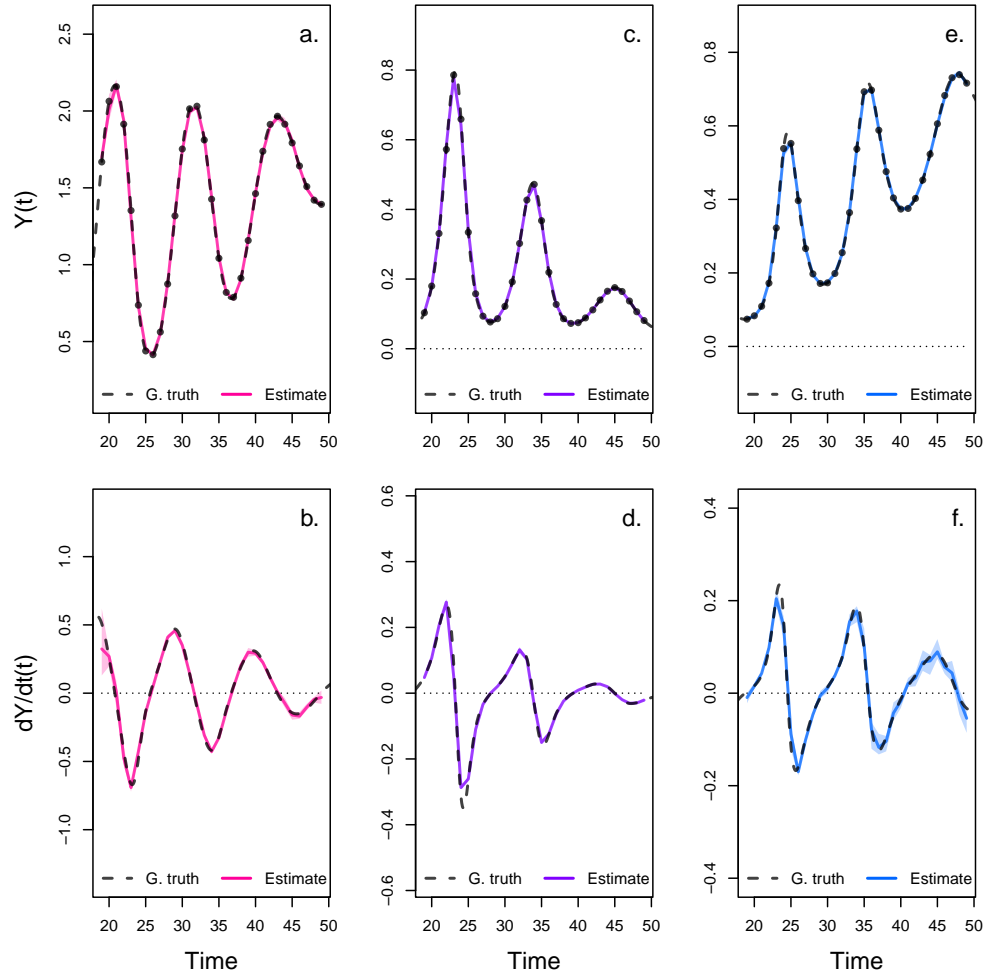


Figure 2: Interpolated density and dynamics of prey, intermediate, and top predators in the artificial system. This figure corresponds to the first step in the overview figure (Fig. 1). It shows the accuracy of the interpolated densities of prey (a.), intermediate (c.), and top predators (e.). We obtain interpolated densities by fitting observed densities (black dots) with ANNs that take time as input. The observed densities were obtained by sampling a tri-trophic prey-predator ODE model at regular time steps. We then derive interpolated dynamics (b., d., f.) by computing the temporal derivative of the interpolated densities with respect to time. In all graphs, the dashed line represents the ground truth, namely trajectories generated by the ODE model. The solid lines correspond to the interpolations. The shaded area shows the 90% confidence interval, obtained by approximately sampling the marginal posterior distributions.

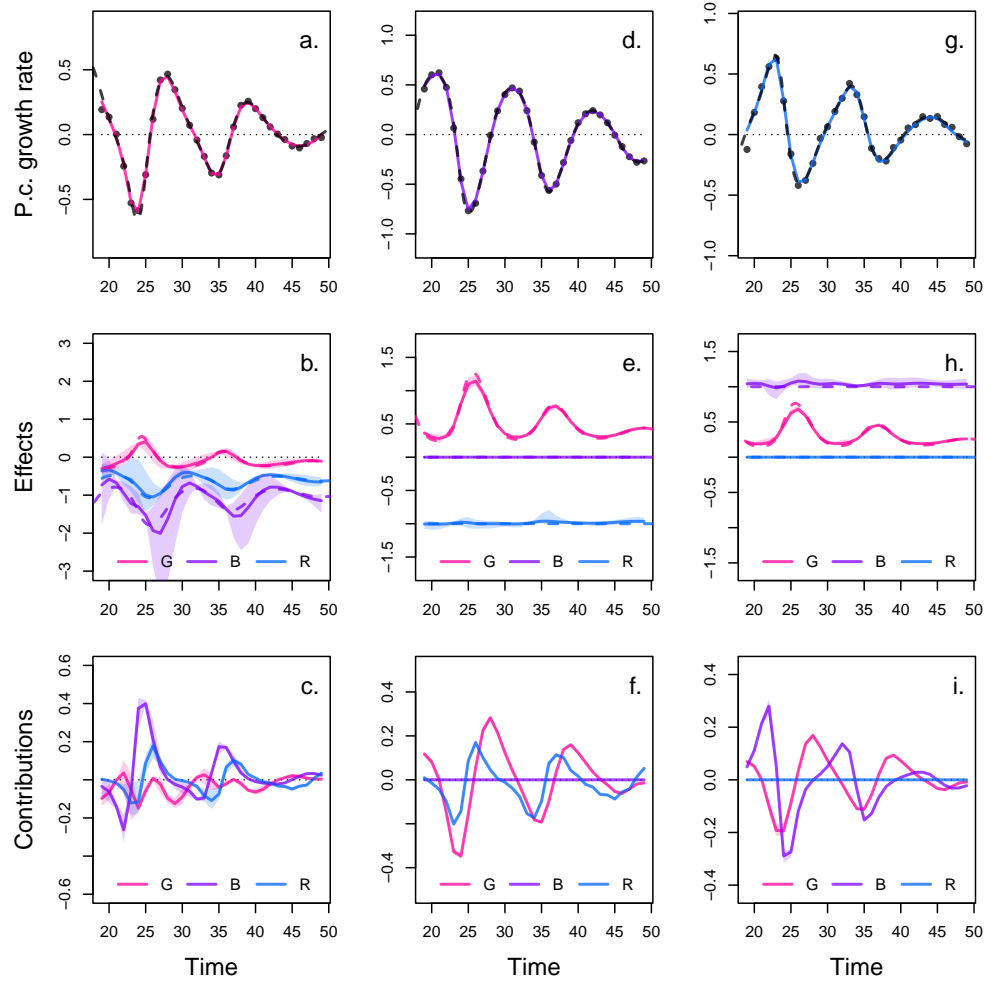


Figure 3: Drivers of dynamics of prey, intermediate, and top predator in the artificial system.

This figure corresponds to the second step in the overview figure (Fig. 1). It displays the NODE nonparametric approximations of the per-capita growth rate of prey (a., b., c.), intermediate (d., e., f.), and top predators (g., h., i.). We obtain the NODE approximations (a., d., g., solid line) by fitting the interpolated per-capita growth rates (black dots) with ANNs that take population densities as input. We then estimate the direction of ecological interactions (effects, b., e., h.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, c., f., i.) by multiplying the interpolated dynamics of each population (fig. 1, b., d., f.) with its effects. Dashed lines correspond to ground truth, obtained from the original trajectories of the tri-trophic ODE model. The shaded area shows the 90% confidence interval, obtained by approximately sampling the posterior distributions.

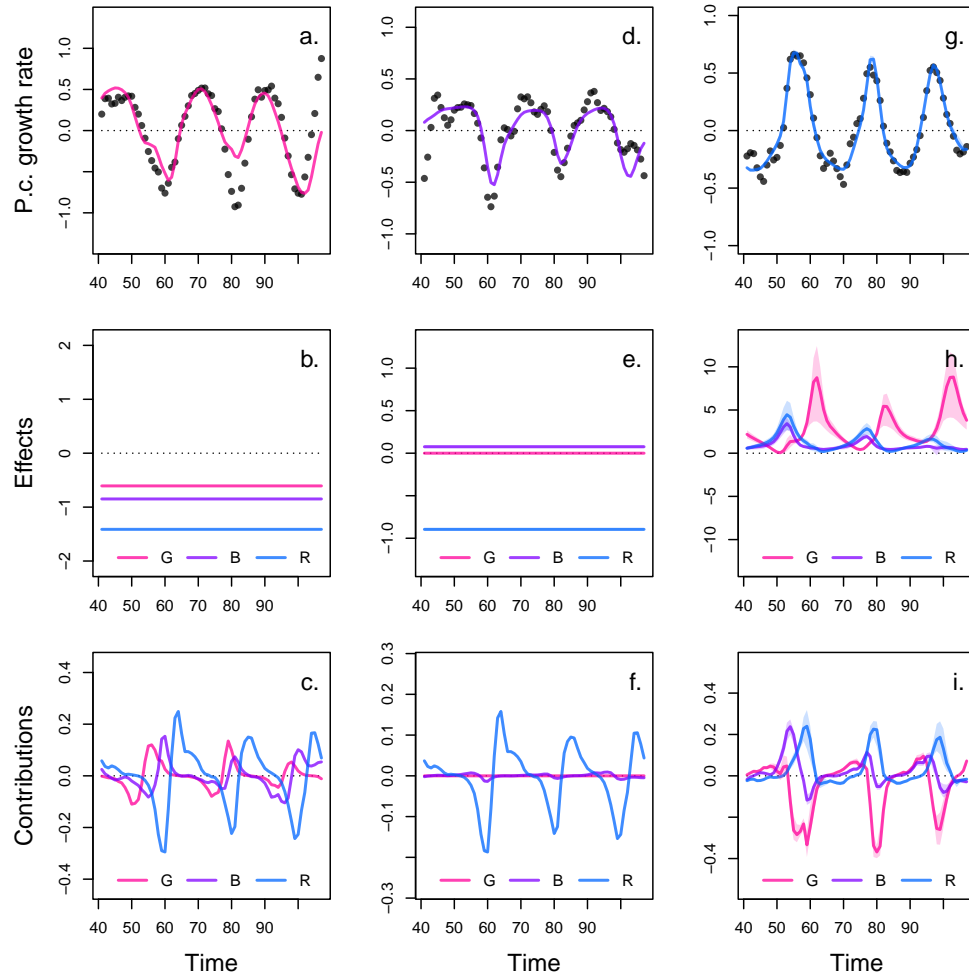


Figure 4: Drivers of dynamics of algae, flagellate, and rotifer in replicate A. This figure displays the NODE nonparametric approximations of the per-capita growth rate of algae (a., b., c.), flagellate (d., e., f.), and rotifer (g., h., i.). We obtain the NODE approximations (a., d., g., solid line) by fitting the interpolated per-capita growth rates (black dots) with ANNs that take population densities as input. We then estimate the direction of ecological interactions (effects, b., e., h.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, c., f., i.) by multiplying the interpolated dynamics of each population with its effects. The shaded area shows the 90% confidence interval, obtained by approximately sampling the posterior distributions. The replicated time series were obtained by digitising the time series in Hiltunen et al. (2013).

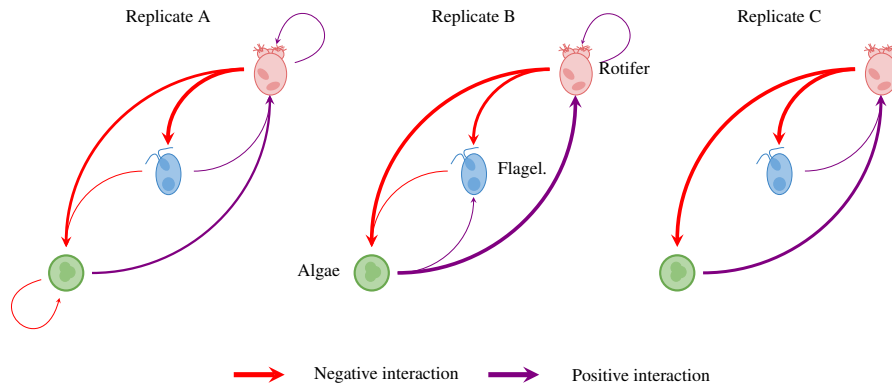


Figure 5: Interaction networks inferred from 3 replicated time series of algae, flagellate, and rotifers. This figure shows the direction and strength of ecological interactions inferred from 3 replicated sets of time series of algae, flagellate, and rotifer, using NODEs fitted by BNGM. The replicates B and C were analysed in the same way as replicate A (see fig. 4 for details). Red and purple arrows correspond to negative or positive mean effects. We estimated mean effects by averaging effects (i.e. derivative of NODE approximated per-capita growth rates with respect to each population density) across the time series. The width of the arrows is proportional to the relative strength of the ecological interaction. We compute the relative strength as the % of total contributions attributable to either algae, flagellate, or rotifer, obtained from summing the square of contributions of each species throughout the time series. For instance in replicate A, the relative strength of the effect of rotifer on algae is found by summing the square of the blue line in fig. 4 c., and comparing it to the sum of square of all contributions (Fig. 4 c., red, purple and blue lines). We provide the value of the mean effects and relative strengths in table 2. The replicated time series were obtained by digitising the time series in Hiltunen et al. (2013).

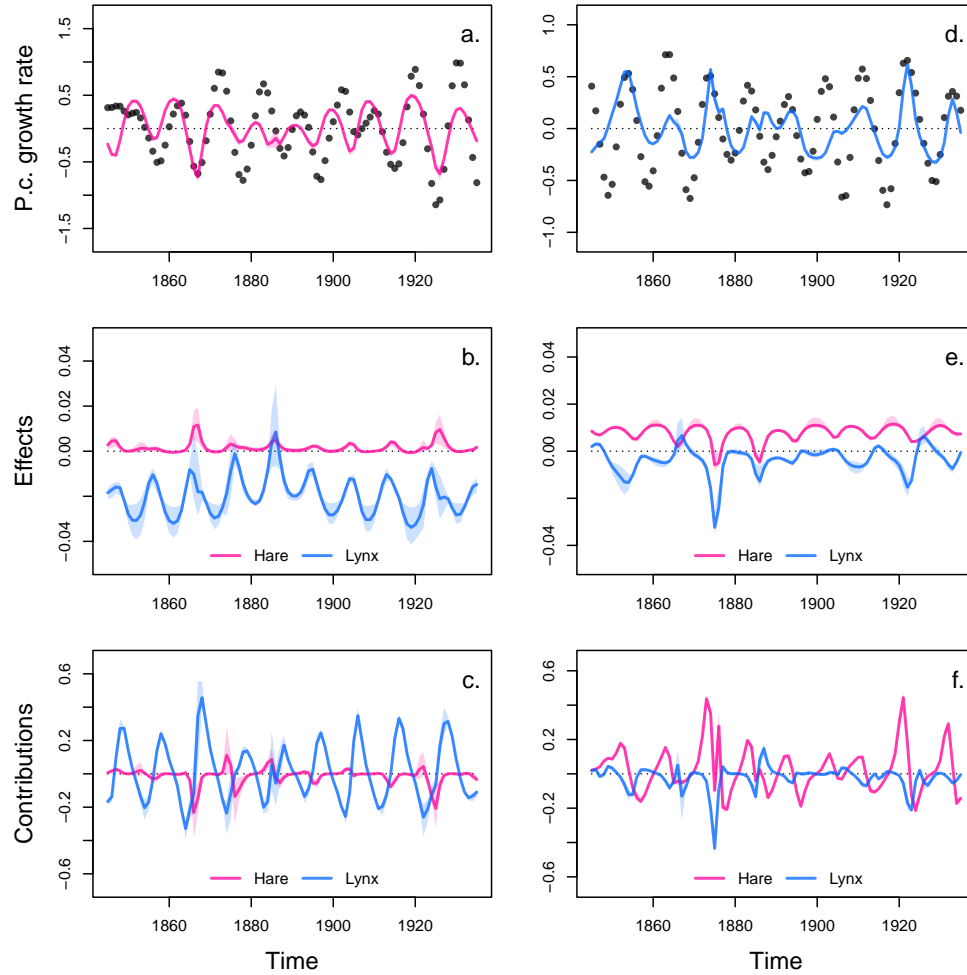


Figure 6: Drivers of dynamics of hare and lynx in the Odum and Barrett pelt count time series. This figure displays the NODE nonparametric approximations of the per-capita growth rate of hare (a., b., c.), and lynx (d., e., f.). We obtain the NODE approximations (a., d., solid line) by fitting the interpolated per-capita growth rates (black dots) with ANNs that take population densities as input. We then estimate the direction of ecological interactions (effects, b., e.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, c., f.) by multiplying the interpolated dynamics of each population with its effects. The shaded area shows the 90% confidence interval, obtained by approximately sampling the posterior distributions.

Table 1: Summary of model runtimes. We measured the time required to perform 100 interpolations and 30 NODE fits to each variable in the systems. Replicate A, B, and C correspond to each replicated time series of the aglae, flagellate, and rotifer tri-trophic system (Hiltunen et al. 2013). The Hare-Lynx system correspond to the 90 years long time series of hare and lynx pelt counts (Odum and Barrett 1972). The number of time steps (N steps) is given for each time series. The total time per fit is obtain by dividing the total time in seconds by the number of fits (i.e. 130). It takes on average 5.35 minutes for the 130 NODE fits NODE, which amounts to 5.37 seconds per sample taken. This is 335 times faster than the 30 minutes fitting times obtained in a previous study (Bonnaffé, Sheldon, and Coulson 2021). These results were obtained on a macbook pro M1 MAX 2022, in base R (v4.0.2), with non-optimised code.

System	N var.	N steps	Interpolation		NODE fit		total	total p. fit
			N fits	time (s)	N fits	time (s)		
Replicate A	3	66	100	239.47	30	129.41	368.88	6.71
Replicate B	3	66	100	233.59	30	133.13	366.72	6.77
Replicate C	3	40	100	136.51	30	74.01	210.52	3.83
Hare-lynx	2	90	100	303.64	30	33.56	337.20	4.16

Table 2: Comparison of the direction and strength of ecological interactions estimated by BNGM across 3 replicated tri-trophic microcosms. Mean effects are obtained by averaging the effect of one species on the growth rate of another throughout the time series. The % of total contributions is obtained by summing the square of contributions of one species density to the growth of the other at each time step throughout the time series, then by computing the proportion of total change that it accounts for. The variables *G*, *B*, and *R* correspond to the population density of algae, flagellate, and rotifer respectively. r^2 corresponds to the r squared of the NODE nonparametric approximation of the pre-capita growth rate of the three species.

		G	B	R
<hr/>				
Replicate A	r^2	0.3	0.47	0.94
Mean effects	on G	-0.61	-0.85	-1.41
	on B	0.00	0.08	-0.90
	on R	2.84	0.93	1.23
% of total contributions	to G	0.13	0.15	0.73
	to B	0.00	0.00	1.00
	to R	0.60	0.16	0.25
<hr/>				
Replicate B	r^2	0.65	0.85	0.47
Mean effects	on G	0.00	-0.56	-1.13
	on B	0.34	0.00	-0.58
	on R	0.87	0.00	0.19
% of total contributions	to G	0.00	0.06	0.94
	to B	0.23	0.00	0.77
	to R	0.95	0.00	0.05
<hr/>				
Replicate C	r^2	0.93	0.29	0.87
Mean effects	on G	-0.14	0.13	-2.31
	on B	-0.05	-0.09	-0.72
	on R	2.46	0.49	-0.09
% of total contributions	to G	0.02	0.02	0.96
	to B	0.00	0.01	0.99
	to R	0.79	0.18	0.03

686 **6 Supplementary**

687 **A Bayesian regularisation**

688 The fitting of the models is performed in a Bayesian framework, considering normal error structure
689 for the residuals, and normal prior density distributions on the parameters

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta) \quad (13)$$

690 where θ is the parameter vector of the model, and \mathcal{D} the evidence, namely the data that the model
691 is fitted to. Assuming a normal likelihood for the residuals given the evidence we get

$$p(\mathcal{D}|\theta) = \prod_{i=1}^I \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{e_i(\mathcal{D}, \theta)^2}{2\sigma^2} \right\} \quad (14)$$

692 where $e_i(\mathcal{D}, \theta)$ are the residuals of the model given the parameters, and the evidence. In the case of
693 the interpolation, the residuals correspond to the observation error $\varepsilon^{(o)}$ (equation 3). In the case of
694 the NODE approximation, they correspond to the process error $\varepsilon^{(p)}$ (equation 7). I is the number
695 of data points, either observations in the case of the interpolation, or interpolated points in the case
696 of the NODE fitting.

697 The prior probability density functions for the parameters are given by

$$p(\theta) = \prod_{j=1}^J \frac{1}{\sqrt{2\pi\delta_j^2}} \exp \left\{ -\frac{\theta_j^2}{2\delta_j^2} \right\} \quad (15)$$

698 where J is the number of parameters in the models. The parameter δ_j controls the dispersion of the
699 priors, and thereby the complexity/level of constraint of the model.

700 Bayesian regularisation simply amounts to constraining the values of the parameters in the model
701 to be close to a desired value. Usually, parameters are constrained by choosing normal priors
702 centered about 0. In this case, the standard deviation of the normal priors governs the range of
703 values that the parameters can take, and hence constrains more or less strongly the behaviour of the
704 model (Cawley and Talbot 2007). There is no standard approach for choosing δ . Low values of
705 dispersion may increase constraint on parameters too drastically, which would lead to underfitting,
706 and result in a reduction of the variance of parameter estimates and bias mean estimates towards
707 0. In contrast, too high values of dispersion may lead to overfitting, by allowing for more complex
708 shapes. To account for this, we optimise models on the second-level of inference. This means that
709 we are finding the optimal value of δ , in addition to optimising the model parameters.

710 In practice, choosing the level of constraint is difficult, Cawley and Talbot hence developed a
711 criterion to perform model selection on the second level of inference. They proposed to optimise the
712 marginal posterior distribution by averaging out the dispersion of the priors. With an appropriate
713 choice of prior, the dispersion can be integrated out, leaving us with a formula for the posterior that
714 only depends on the parameters of the model,

$$\log P(\theta|\mathcal{D}) \propto -\frac{I}{2} \log \left(\sum_{i=1}^I e_i(\mathcal{D}, \theta)^2 \right) - \frac{J}{2} \log \left(\sum_{j=1}^J \theta_j^2 \right) \quad (16)$$

715 where $P(\theta|\mathcal{D})$ denotes the marginal posterior density, \mathcal{D} denotes the evidence, I and J denote
 716 the number of data points and parameters, respectively, e_i denote the residuals, and θ denote the
 717 parameters of the model. The construction is elegant because it is not sensitive to the choice of
 718 prior hyperparameters, and simple as it amounts to optimising the log of the sum of squares, rather
 719 than the sum of squares (in the case of normal ordinary least square).

720 The issue with this formula is that the marginal posterior density is not finite when the parameters
 721 are 0, which leads to underfitting. In this paper we use a modified criterion, which corrects for that
 722 problem,

$$\log P(\theta|\mathcal{D}) \propto -\frac{I}{2} \log \left(1 + \sum_{i=1}^I e_i(\mathcal{D}, \theta)^2 \right) - \frac{J}{2} \log \left(1 + \sum_{j=1}^J \theta_j^2 \right) \quad (17)$$

723 where the marginal posterior density depends only on the residuals of the model when the parame-
 724 ters are equal to 0, and otherwise depends on both the parameters and the residuals. This construc-
 725 tion can be obtained simply by assuming a gamma prior for the parameters $p(\xi) \propto \frac{1}{\xi} \exp\{-\xi\}$,
 726 where ξ is the regularisation parameter, instead of the improper Jeffreys' prior that Cawley and
 727 Talbot used in their original study, namely $p(\xi) \propto \frac{1}{\xi}$. The details of the integration of the posterior
 728 distribution over ξ can be found in Cawley and Talbot's original paper.

B Complementary results case study 2 replicate A

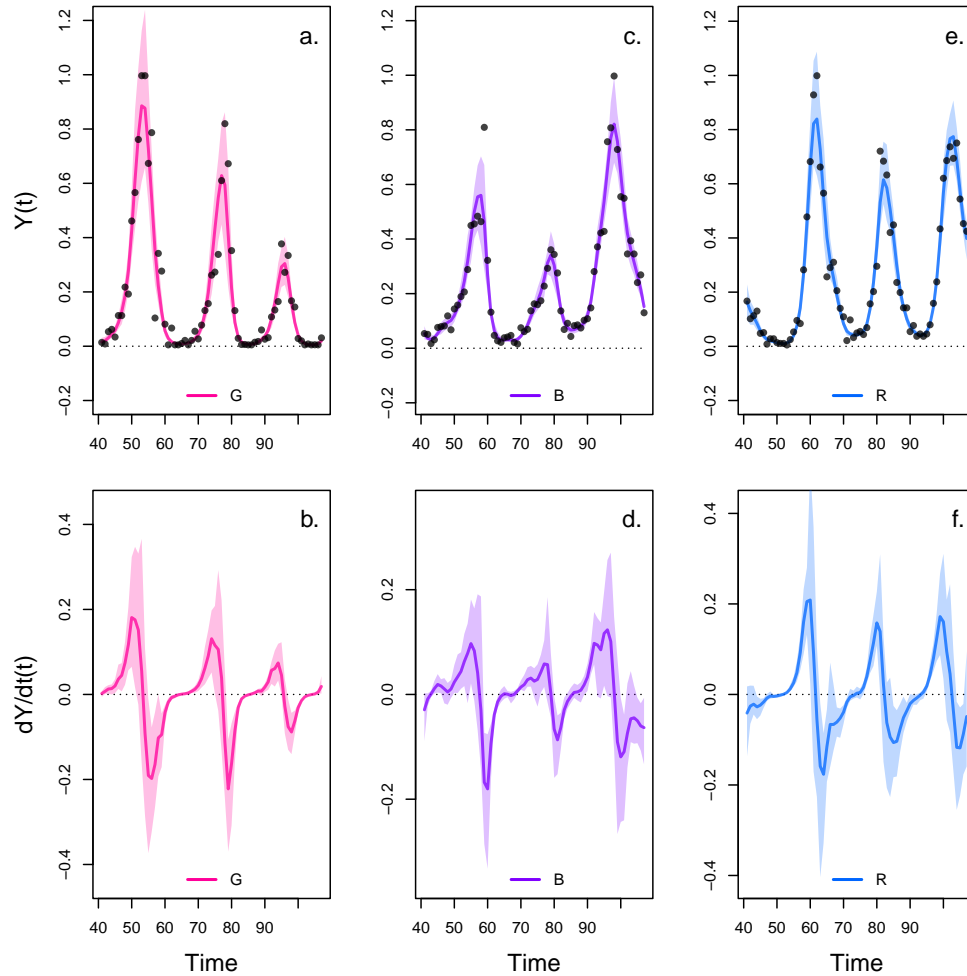


Figure S1: Interpolation of state and dynamics of algae, flagellate, and rotifer density in replicate A. Graph a., c., and e. display the neural interpolation of the population density of algae (G), flagellate (B), and rotifer (R), respectively (obtained with Eq. 7). Graph b., d., and f. show the corresponding interpolated dynamics, obtained by differentiating the interpolation of the states with respect to time (Eq. 5). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log marginal posterior distribution (Eq. 7) (Pearce et al. 2018). Time series are obtained from digitising the time series in Hiltunen et al. 2013.

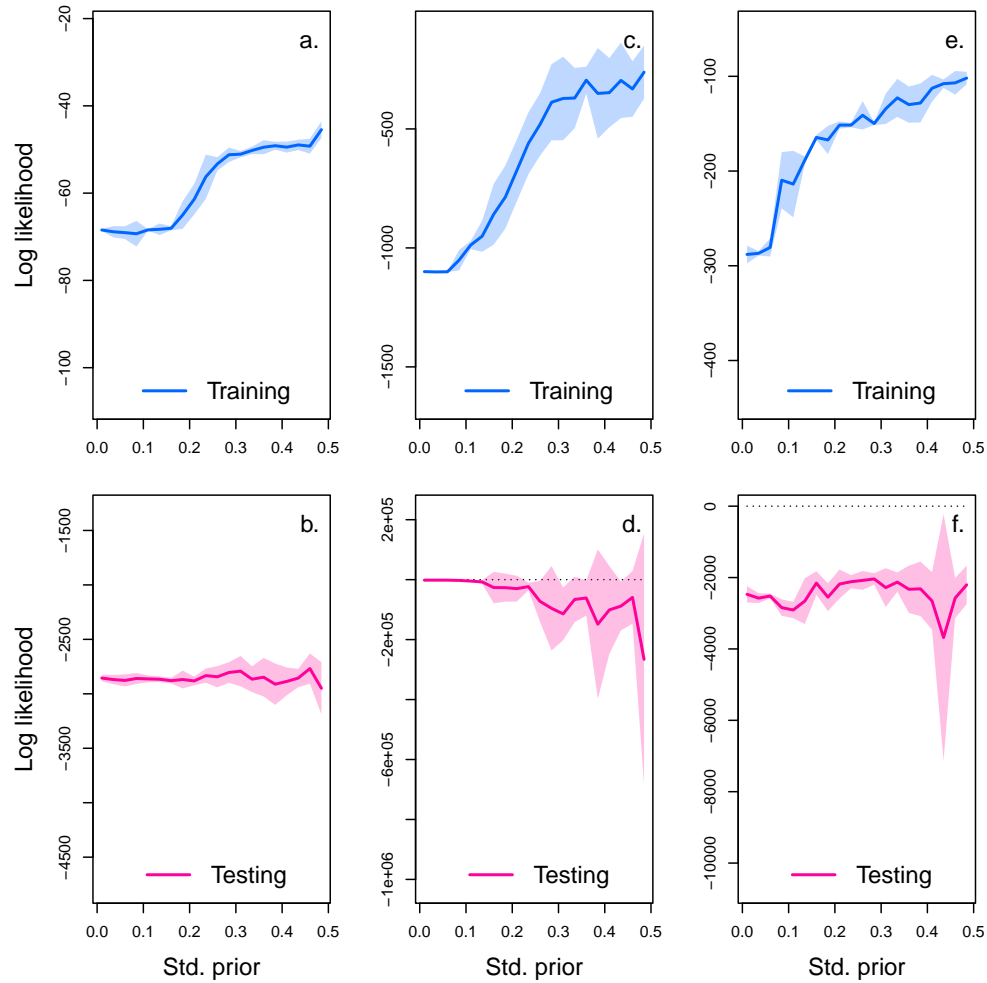


Figure S2: Cross-validation plot of the NODE analysis of replicate A. The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values (towards 0.5) correspond to a highly nonlinear model. Time series of algae, flagellate, and rotifer are split in half to create a train set and a test set. The model is fitted to the train set for increasing value of standard deviation (from 0.05 to 0.5 by 0.05 increments), and evaluated on the test set. Graph a., c., and e. show the log likelihood of the NODE system fitted by BNGM to the train set of algae, flagellate, and rotifer, respectively. Graph b., d., and f. show the log likelihood of the fitted NODE, evaluated on the corresponding test set. The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log marginal posterior distribution (Eq. 7) (Pearce et al. 2018).

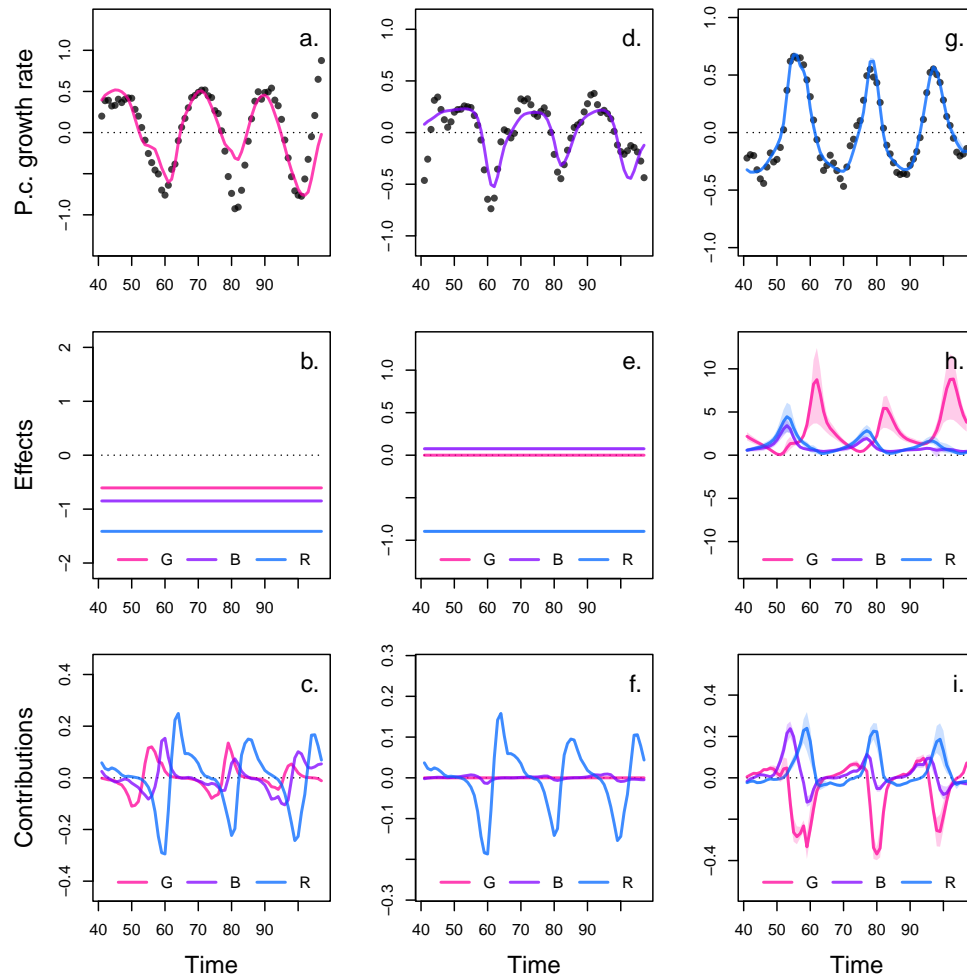


Figure S3: Drivers of dynamics of algae, flagellate, and rotifer in replicate A. This figure displays the NODE nonparametric approximations of the per-capita growth rate of algae (a., b., c.), flagellate (d., e., f.), and rotifer (g., h., i.). We obtain the NODE approximations (a., d., g., solid line) by fitting the interpolated per-capita growth rates (black dots) with ANNs that take population densities as input. We then estimate the direction of ecological interactions (effects, b., e., h.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, c., f., i.) by multiplying the interpolated dynamics of each population with its effects. The shaded area shows the 90% confidence interval, obtained by approximately sampling the posterior distributions. The replicated time series were obtained by digitising the time series in Hiltunen et al. (2013).

C Complementary results case study 2 replicate B

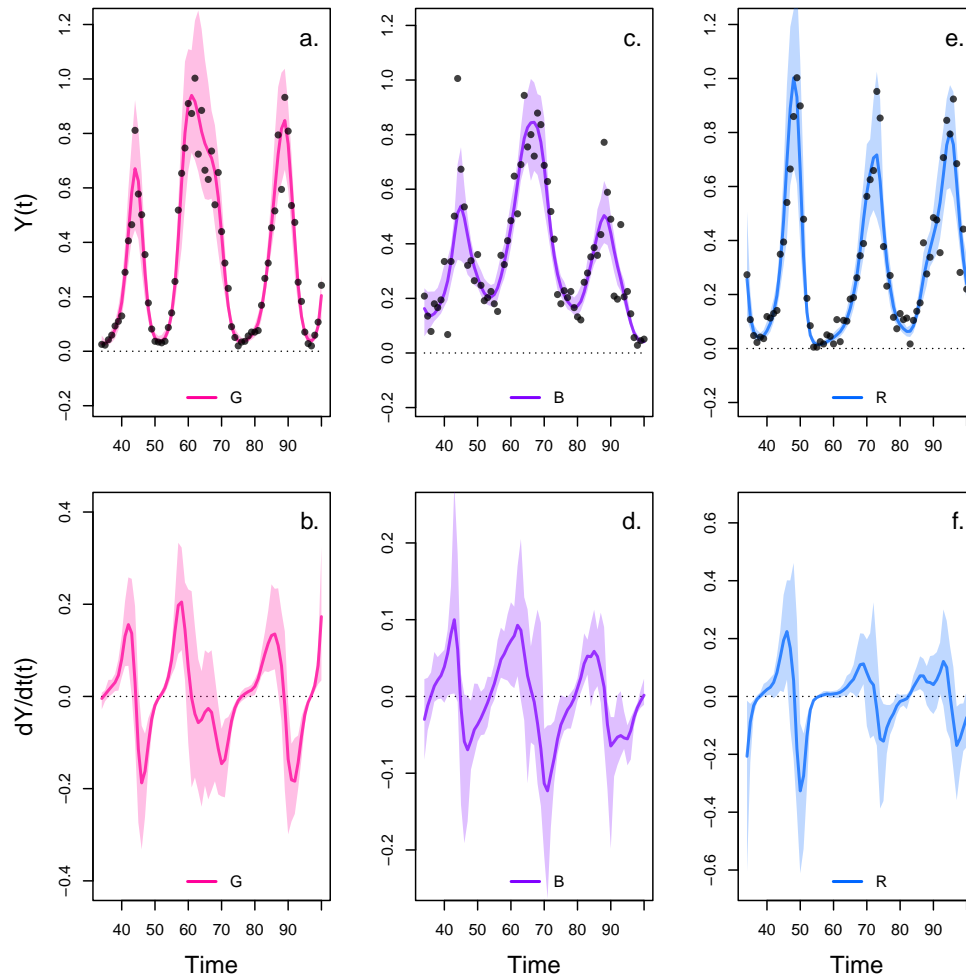


Figure S4: Interpolation of state and dynamics of algae, flagellate, and rotifer density in replicate B. Graph a., c., and e. display the neural interpolation of the population density of algae (G), flagellate (B), and rotifer (R), respectively (obtained with Eq. 7). Graph b., d., and f. show the corresponding interpolated dynamics, obtained by differentiating the interpolation of the states with respect to time (Eq. 5). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log marginal posterior distribution (Eq. 7) (Pearce et al. 2018). Time series are obtained from digitising the time series in Hiltunen et al. 2013.

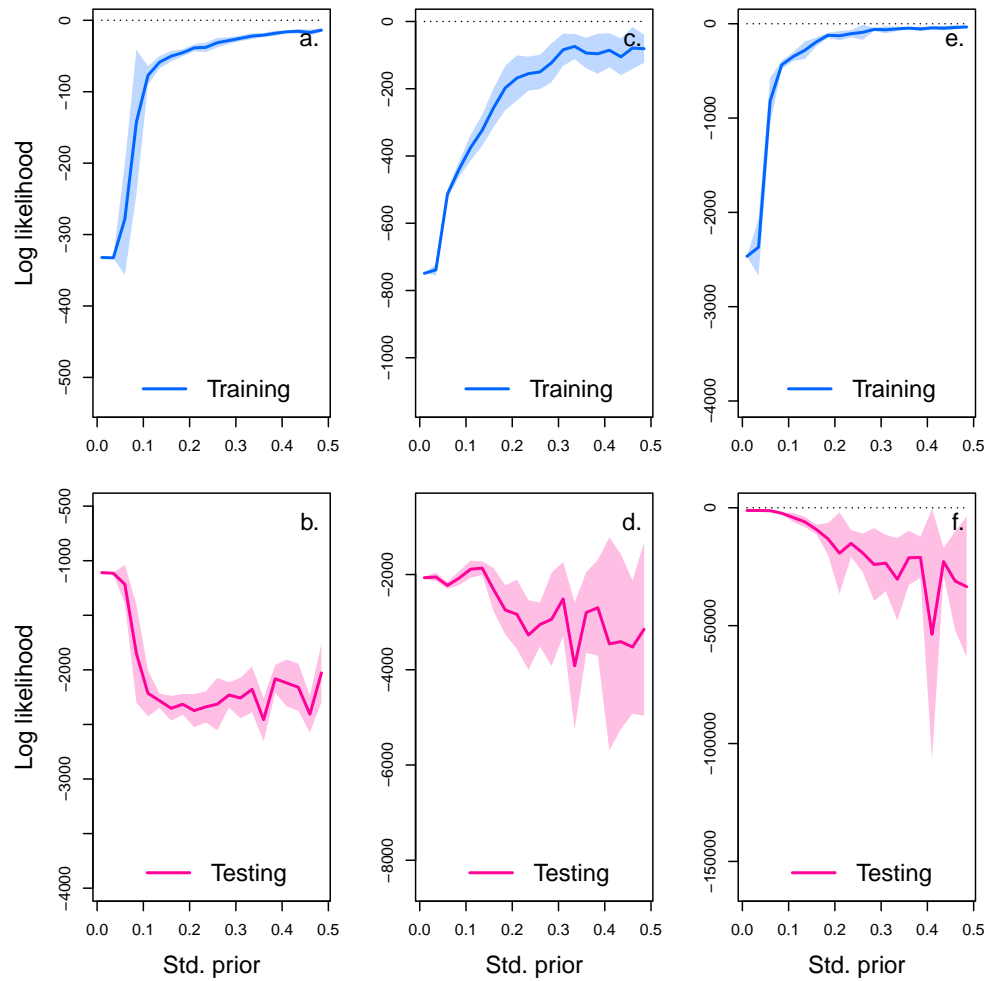


Figure S5: Cross-validation plot of the NODE analysis of replicate B. The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values (towards 0.5) correspond to a highly nonlinear model. Time series of algae, flagellate, and rotifer are split in half to create a train set and a test set. The model is fitted to the train set for increasing value of standard deviation (from 0.05 to 0.5 by 0.05 increments), and evaluated on the test set. Graph a., c., and e. show the log likelihood of the NODE system fitted by BNGM to the train set of algae, flagellate, and rotifer, respectively. Graph b., d., and f. show the log likelihood of the fitted NODE, evaluated on the corresponding test set. The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log marginal posterior distribution (Eq. 7) (Pearce et al. 2018).

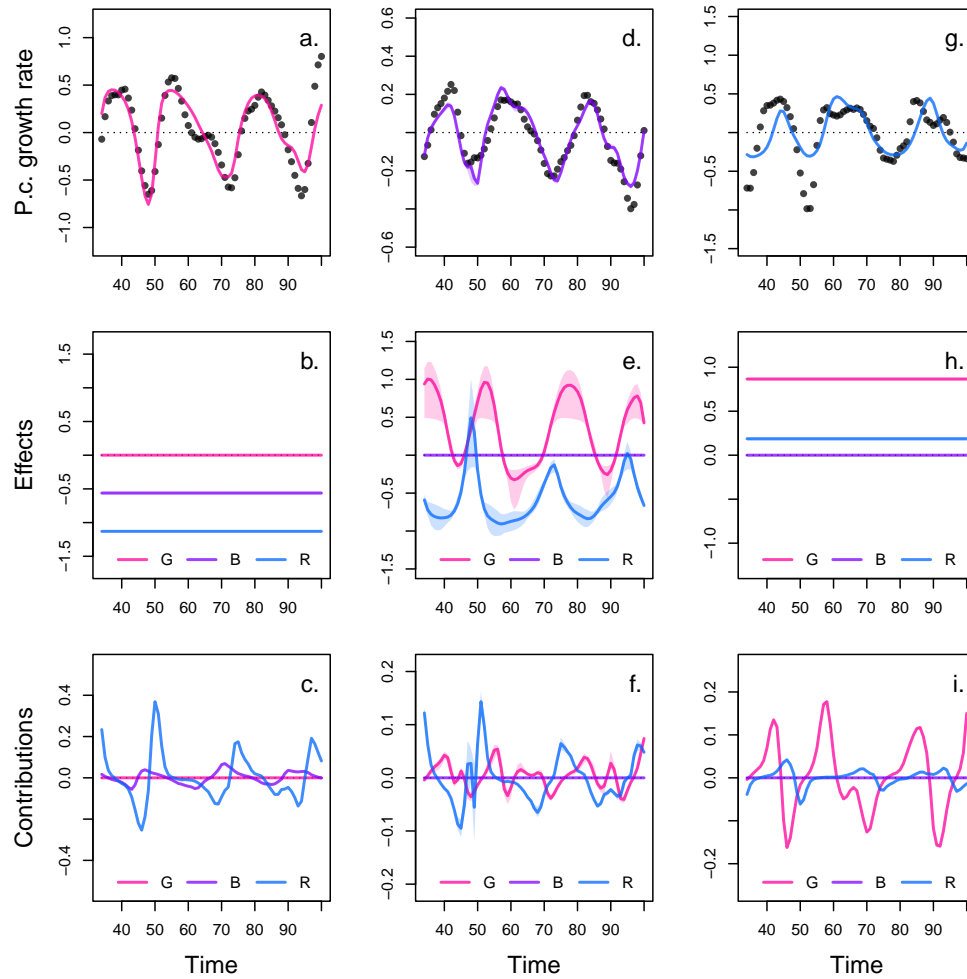


Figure S6: Drivers of dynamics of algae, flagellate, and rotifer in replicate B. This figure displays the NODE nonparametric approximations of the per-capita growth rate of algae (a., b., c.), flagellate (d., e., f.), and rotifer (g., h., i.). We obtain the NODE approximations (a., d., g., solid line) by fitting the interpolated per-capita growth rates (black dots) with ANNs that take population densities as input. We then estimate the direction of ecological interactions (effects, b., e., h.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, c., f., i.) by multiplying the interpolated dynamics of each population with its effects. The shaded area shows the 90% confidence interval, obtained by approximately sampling the posterior distributions. The replicated time series were obtained by digitising the time series in Hiltunen et al. (2013).

D Complementary results case study 2 replicate C

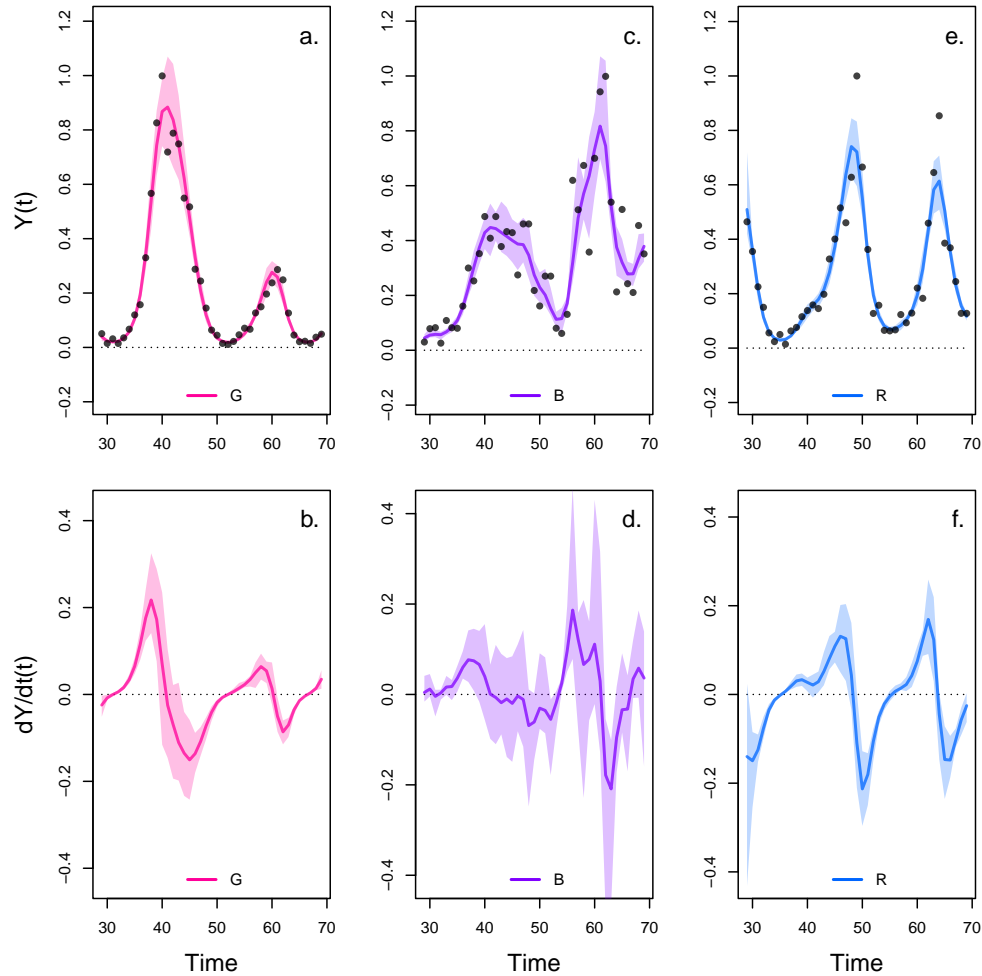


Figure S7: Interpolation of state and dynamics of algae, flagellate, and rotifer density in replicate B. Graph a., c., and e. display the neural interpolation of the population density of algae (G), flagellate (B), and rotifer (R), respectively (obtained with Eq. 7). Graph b., d., and f. show the corresponding interpolated dynamics, obtained by differentiating the interpolation of the states with respect to time (Eq. 5). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log marginal posterior distribution (Eq. 7) (Pearce et al. 2018). Time series are obtained from digitising the time series in Hiltunen et al. 2013.

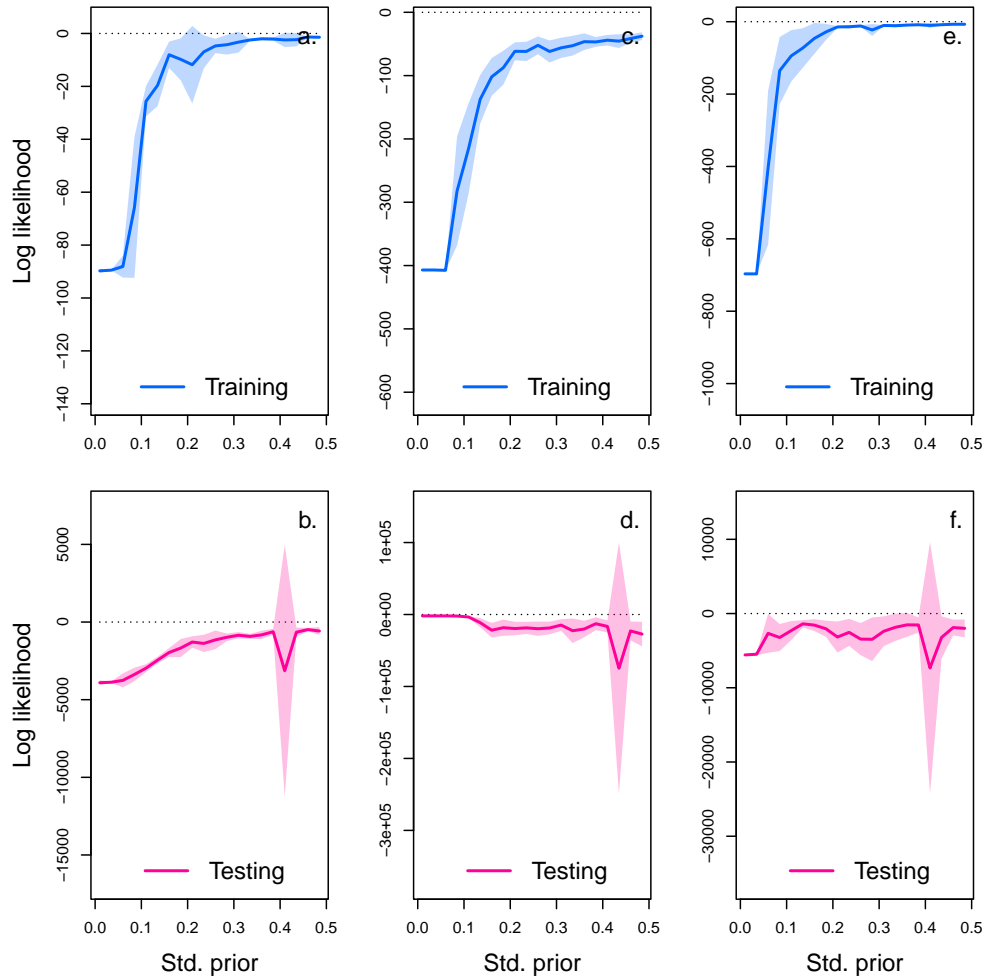


Figure S8: Cross-validation plot of the NODE analysis of replicate C. The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values (towards 0.5) correspond to a highly nonlinear model. Time series of algae, flagellate, and rotifer are split in half to create a train set and a test set. The model is fitted to the train set for increasing value of standard deviation (from 0.05 to 0.5 by 0.05 increments), and evaluated on the test set. Graph a., c., and e. show the log likelihood of the NODE system fitted by BNGM to the train set of algae, flagellate, and rotifer, respectively. Graph b., d., and f. show the log likelihood of the fitted NODE, evaluated on the corresponding test set. The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log marginal posterior distribution (Eq. 7) (Pearce et al. 2018).

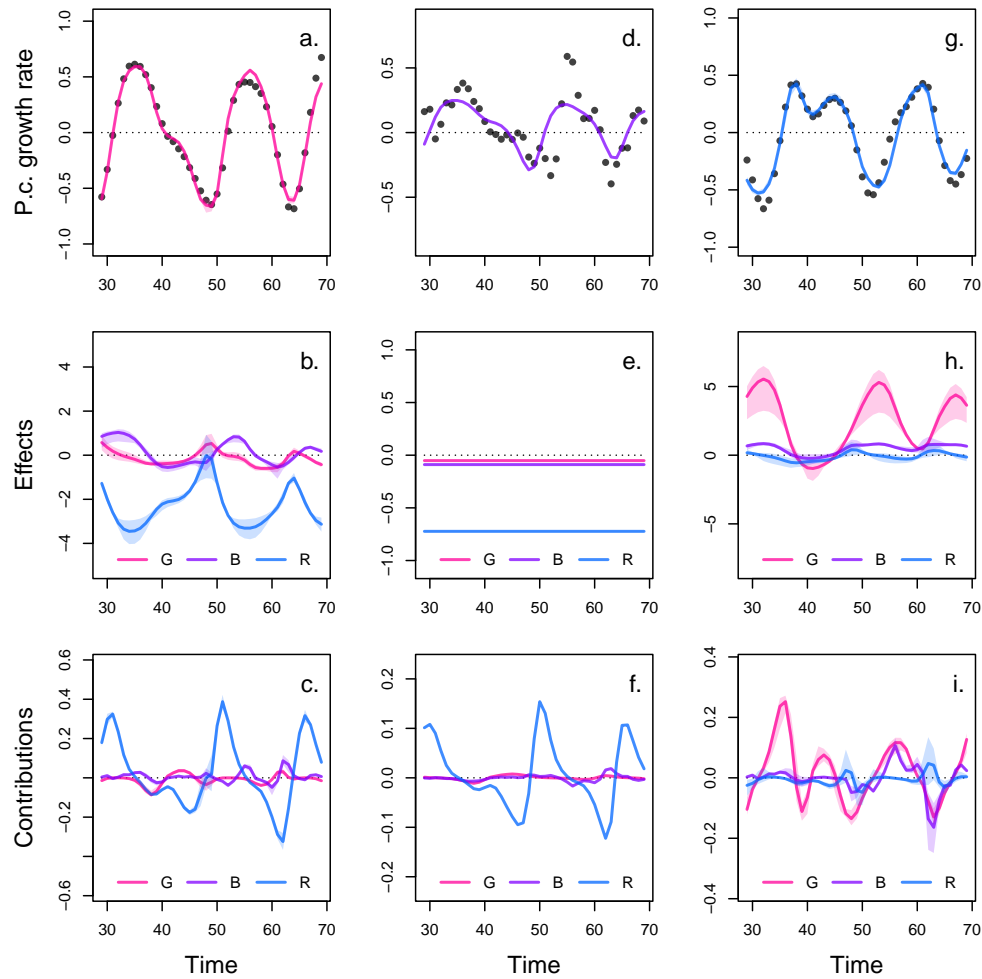


Figure S9: Drivers of dynamics of algae, flagellate, and rotifer in replicate C. This figure displays the NODE nonparametric approximations of the per-capita growth rate of algae (a., b., c.), flagellate (d., e., f.), and rotifer (g., h., i.). We obtain the NODE approximations (a., d., g., solid line) by fitting the interpolated per-capita growth rates (black dots) with ANNs that take population densities as input. We then estimate the direction of ecological interactions (effects, b., e., h.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, c., f., i.) by multiplying the interpolated dynamics of each population with its effects. The shaded area shows the 90% confidence interval, obtained by approximately sampling the posterior distributions. The replicated time series were obtained by digitising the time series in Hiltunen et al. (2013).

E Complementary results case study 3

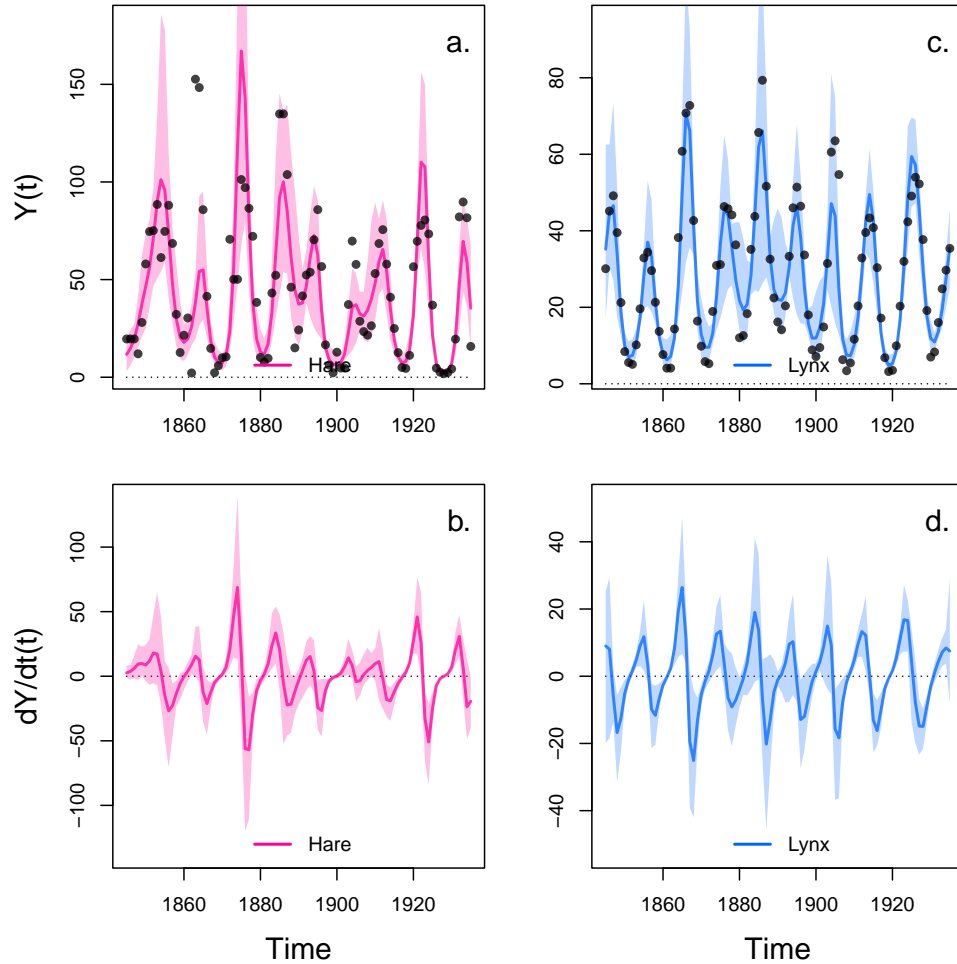


Figure S10: Interpolation of state and dynamics of hare and lynx. Graph a. and c. display the neural interpolation of the population density of hare and lynx respectively (obtained with Eq. 7). Graph b. and d. show the corresponding interpolated dynamics, obtained by differentiating the interpolation of the states with respect to time (Eq. 5). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log marginal posterior distribution (Eq. 7) (Pearce et al. 2018). Time series are obtained from Bonnaffé, Sheldon, and Coulson 2021, originally from Odum and Barrett 1972.

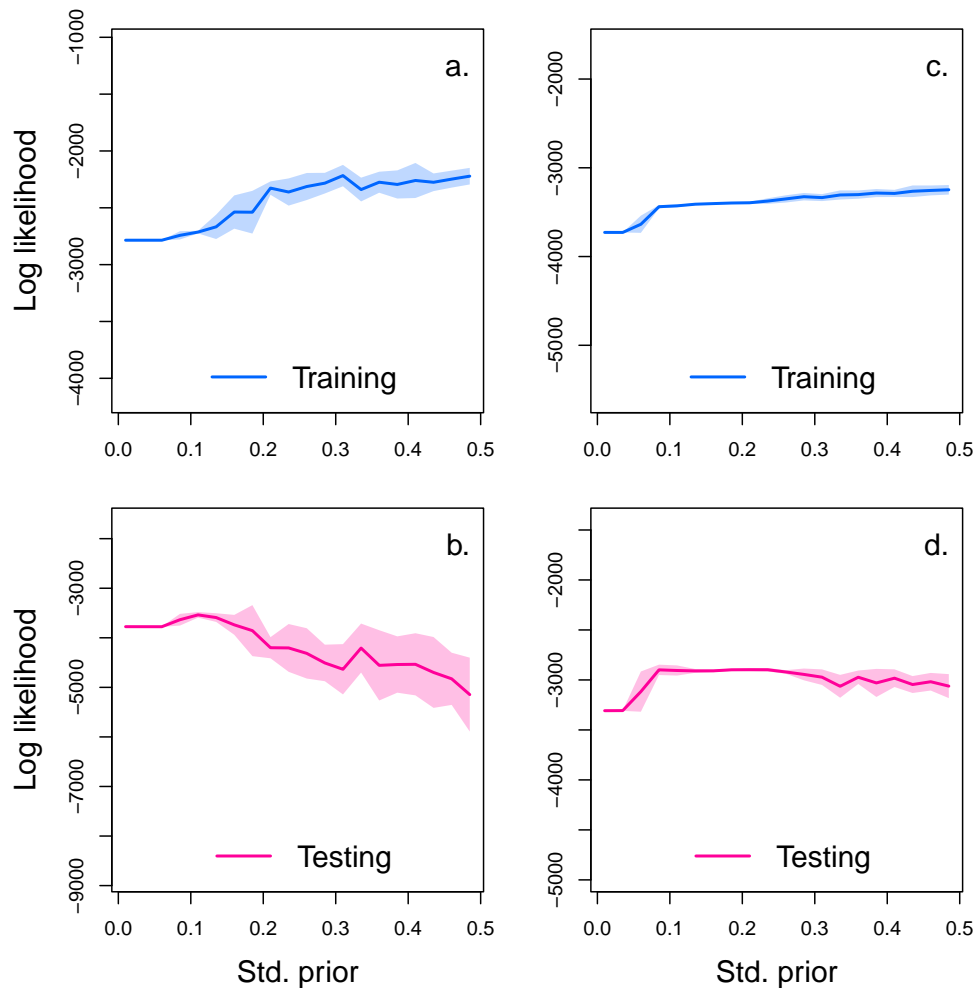


Figure S11: Cross-validation plot of the NODE analysis of the hare-lynx system The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values (towards 0.5) correspond to a highly nonlinear model. Time series of algae, flagellate, and rotifer are split in half to create a train set and a test set. The model is fitted to the train set for increasing value of standard deviation (from 0.05 to 0.5 by 0.05 increments), and evaluated on the test set. Graph a., c., and e. show the log likelihood of the NODE system fitted by BNGM to the train set of algae, flagellate, and rotifer, respectively. Graph b., d., and f. show the log likelihood of the fitted NODE, evaluated on the corresponding test set. The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log marginal posterior distribution (Eq. 7) (Pearce et al. 2018).

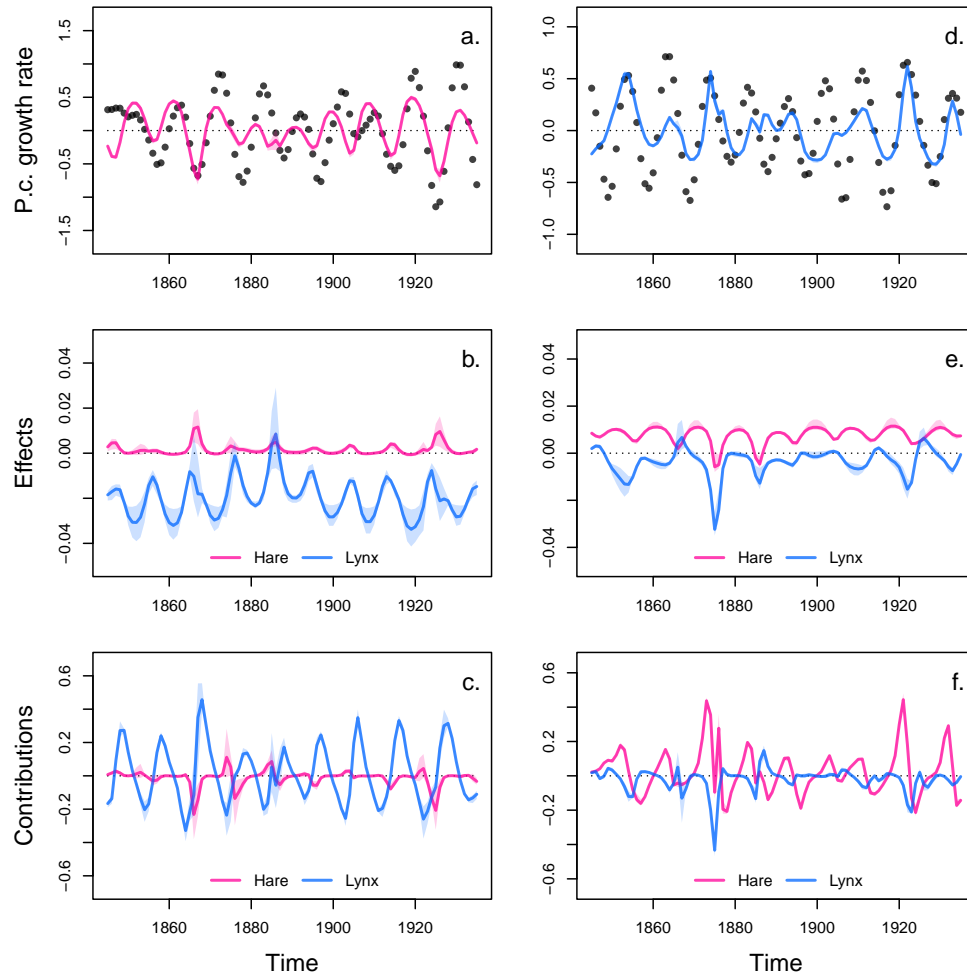


Figure S12: Drivers of dynamics of hare and lynx in the Odum and Barrett pelt count time series. This figure displays the NODE nonparametric approximations of the per-capita growth rate of hare (a., b., c.), and lynx (d., e., f.). We obtain the NODE approximations (a., d., solid line) by fitting the interpolated per-capita growth rates (black dots) with ANNs that take population densities as input. We then estimate the direction of ecological interactions (effects, b., e.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, c., f.) by multiplying the interpolated dynamics of each population with its effects. The shaded area shows the 90% confidence interval, obtained by approximately sampling the posterior distributions.