

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

Willem Bonnaffé<sup>1,2</sup> & Tim Coulson<sup>2</sup>

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:** 324 words in abstract; 8268 words in text; 63 references; 8 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 compare these results against three existing approaches for estimating ecological interactions, standard NODEs, ODE models, and convergent cross mapping (CCM). We also infer interactions in experimentally replicated time series of a microcosm featuring an algae, flagellate, and rotifer population, in the hare and lynx system, and the Maizuru bay community featuring 11 species.
3. Our BNGM approach allows us to cut down the fitting time of NODE systems to only a few seconds and provides accurate estimates of ecological interactions in the artificial system, as true ecological interactions are estimated with minimal error. Our benchmark analysis reveals that our approach is both faster and more accurate than standard NODEs and parametric ODEs, while CCM was found to be faster but less accurate. The analysis of the replicated time series reveals that only strongest interactions are consistent across replicates, while the analysis of the Maizuru community shows the strong negative impact of the chameleon goby on most species of the community, and a potential indirect negative effect of temperature by favouring goby population growth.
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 concepts of population and community (i.e. groups of populations) are central in ecology (Berryman 2002). Ecologists have had a longstanding interest in finding laws that govern population and community dynamics, namely changes in the number of individuals in the populations present in a community (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

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

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

et al. 2021), and governed by environmental variables, such as temperature (Brown et al. 2004). Interactions also change following evolution of the traits that underpin them (Turchin et al. 2003; Yoshida et al. 2003). This makes it virtually impossible to model the full complexity of ecological interactions (Lawton 1999; Kendall et al. 1999).

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

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

2018). Nonparametric models require minimal assumptions regarding the mathematical nature of ecological interactions (Jost and Ellner 2000; Gross, Ives, and Nordheim 2005), and hence provide interaction estimates that are more robust to model structure (Wood 2001). In particular, artificial neural networks (ANNs) offer a promising, yet underused, nonparametric alternative to linear functional forms. In previous work, we introduced a powerful framework, relying on neural ordinary differential equations (NODEs, Chen et al. 2019) to approximate the dynamics of populations nonparametrically, from which we derive ecological interactions (Bonnaffé, Sheldon, and Coulson 2021). More specifically, the ANNs embedded in the ODEs learn nonparametrically the shape of the per capita growth rate of the populations and its dependence on the state variables of the system (Bonnaffé, Sheldon, and Coulson 2021). Combined with the Geber method (Hairston et al. 2005), we are able to estimate the direction, strength, and degree of nonlinearity of interactions.

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

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

87 In this manuscript, we first introduce a novel fitting technique for NODEs, Bayesian neural gradient  
88 matching (BNGM). The method extends gradient matching by using neural networks to interpolate  
89 the time series data instead of splines (Ellner, Seifu, and Smith 2002), and Bayesian regularisa-  
90 tion to fit NODEs to the interpolated dynamics (Cawley and Talbot 2007). This cuts down the  
91 fitting time of NODEs to only a few seconds, compared to about 30 minutes in our previous work  
92 (Bonnaffé, Sheldon, and Coulson 2021), allowing for efficient cross-validation, and uncertainty  
93 quantification. We then demonstrate that NODEs are highly accurate in recovering ecological in-  
94 teractions in an artificial three-species prey-predator system where truth is known. Finally, we  
95 conclude the work by characterising ecological interactions in three replicates of an experimental  
96 three-species prey-predator system with an algae, flagellate, and rotifer (Hiltunen et al. 2013), in  
97 the classic hare and lynx time series (Odum and Barrett 1972), as well as in the larger aquatic  
98 community of the Maizuru bay in Japan (Ushio et al. 2018). We find that only main interactions,  
99 between the algae and the rotifer, are conserved across the three replicates, and not the interactions  
100 of the flagellate with the other species. We also find that in most cases linear interactions are suffi-

cient to explain the dynamics apart from nonlinearity in the effect of the prey on the top predator in both the rotifer and lynx. Finally, we find that the dynamics of the aquatic species of the Maizuru bay community are largely determined by a single species, the chameleon goby, and by an indirect effect of temperature via its impact on goby population dynamics.

## 2 Material and Methods

### 2.1 Method overview

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

### 2.2 Neural ordinary differential equation

A NODE is a class of ordinary differential equation (ODE) that is partly or entirely defined as an artificial neural network (ANN) (Chen et al. 2019). They are useful to infer dynamical processes non-parametrically from time series data (Bonnaffé, Sheldon, and Coulson 2021). We choose NODEs over standard statistical approaches because they offer two advantages. The first is that NODEs



119 approximate the dynamics of populations nonparametrically. NODEs are therefore not subjected  
 120 to incorrect model specifications (Jost and Ellner 2000; Adamson and Morozov 2013). This pro-  
 121 vides a more objective estimation of the inter-dependences between state variables. The second  
 122 advantage is that it is a dynamical systems approach. So that the approach includes lag effects  
 123 through interacting state variables, not only direct effects between them.

124 We first consider a general NODE system,

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

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

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

131 which feature a single layer, containing  $J$  neurons, that maps the inputs, here the state variables  $y$ ,  
 132 to a single output, the dynamics of state variable  $i$ ,  $dy_i/dt$ . The parameter vector  $\theta_i$  contains the  
 133 weights  $\theta^{(l)}$  of the connections in the SLPs. SLPs can be viewed as weighted sums of activation

134 functions  $f_\sigma$ , which are usually chosen to be sigmoid functions  $f(x) = 1/(1 + \exp(-x))$ . The link  
 135 function  $f_\lambda$  allows to map the output of the network to a specific domain, for instance applying tanh  
 136 will constrain the dynamics between -1 and 1,  $dy_i/dt \in ]-1, 1[$ . Multi-layer networks can also  
 137 be used but are generally considered unnecessary since pioneering work established that a single  
 138 layer is sufficient to approximate any continuous function to a desired level of error (Funahashi and  
 139 Nakamura 1993).

140 This general form can be changed to represent biological constraints on the state variables. In  
 141 particular for population dynamics, the state variables are strictly positive population densities,  
 142  $y_i = N_i \in \mathcal{R}^+$ . We could hence re-write equation (1) as,  $dN_i/dt = f_p(N, \theta_i)N_i$ , where the SLPs  
 143 approximate the per-capita growth rate of the populations. More details regarding these models  
 144 can be found in our previous work (Bonnaiffé, Sheldon, and Coulson 2021).

### 145 **2.3 Fitting NODEs by Bayesian neural gradient matching**

146 In this section, we describe how to estimate the parameters  $\theta$  of the NODE system given a set of  
 147 time series. Fitting NODEs can be highly computationally intensive, which hinders uncertainty  
 148 quantification, cross-validation, and model selection (Bonnaiffé, Sheldon, and Coulson 2021). We  
 149 solve this issue by introducing BNGM, a computationally efficient approach to fit NODEs. The  
 150 approach involves two steps (Fig. 1). First, we interpolate the state variables and their dynamics  
 151 with neural networks (Fig. 1, red boxes). Second, we train each NODE to satisfy the interpolated  
 152 state and dynamics (Fig. 1, blue boxes). This bypasses the costly numerical integration of the

153 NODE system and provides a fully mathematically tractable expression for the posterior distribu-  
 154 tion of the parameter vector  $\theta$ , and hence analytical expressions for the gradients. We coin the  
 155 term BNGM to emphasise two important refinements of the standard gradient matching algorithm  
 156 (Ellner, Seifu, and Smith 2002). The first is that we use neural networks as interpolation functions,  
 157 and the second is that we use Bayesian regularisation to limit overfitting and estimate uncertainty  
 158 around parameters (Cawley and Talbot 2007).

### 159 **Interpolating the time series**

160 The first step is to interpolate the time series and differentiate it with respect to time in order to ap-  
 161 proximate the state and dynamics of the variables. We perform the interpolation via nonparametric  
 162 regression of the interpolating functions on the time series data,

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

163 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  
 164 interpolation function given the parameter vector  $\omega_i$ , and  $\varepsilon_{it}^{(o)}$  is the observation error between the  
 165 observation and prediction. The interpolation function is chosen to be a neural network,

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

166 where the parameter vector  $\omega_i$  contains the weights  $\omega^{(l)}$  of the network. We can further differentiate  
 167 this expression with respect to time to obtain an interpolation of the dynamics of the state variables

168 (Fig. 1, red boxes),

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

### 169 **Fitting NODEs to the interpolated time series**

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

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

173 where  $\varepsilon_{it}^{(p)}$  is the process error, namely the difference between the interpolated dynamics,  $\partial \tilde{y}_i / \partial t$   
 174 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  
 175 boxes).

### 176 **Bayesian regularisation**

177 In the context of standard gradient matching, defining the observation model (Eq. 3) and process  
 178 model (Eq. 6) would be sufficient to fit the NODE system (Eq. 1) to the time series via optimisation  
 179 (Jost and Ellner 2000; Ellner, Seifu, and Smith 2002; Wu, Fukuhara, and Takeda 2005). We could  
 180 find the parameter vector  $\omega_i$  and  $\theta_i$  that minimise the sum of squared observation and process errors,  
 181  $\varepsilon_{it}^{(o)}$  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,  $\epsilon_{ij}^{(o)} \sim \mathcal{N}(0, \sigma_i)$ , 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  $\sigma_i$  and  $\gamma_{ij}$ . Therefore, we use the approach developed by Cawley and Talbot to average out the value of the parameters  $\sigma_i$  and  $\gamma_{ij}$  in the full posterior distribution (Cawley and Talbot 2007), assuming hyperpriors  $p(\xi) \propto \frac{1}{\xi} \exp\{-\xi/2\}$  for both parameters. This yields the following expression for the log marginal posterior density of the parameters,

$$\log P(\omega_i | Y_i) \propto -\frac{N^{(o)}}{2} \log \left( 1 + \sum_{t=1}^{N^{(o)}} \left( \epsilon_{it}^{(o)} \right)^2 \right) - \frac{M^{(o)}}{2} \log \left( 1 + \sum_{j=1}^{M^{(o)}} \omega_{ij}^2 \right) \quad (7)$$

where  $P$  is the marginal posterior density,  $\omega_i = \{\omega_{i1}, \omega_{i2}, \dots, \omega_{iM^{(o)}}\}$  is the observation parameter vector controlling the interpolation function,  $Y_i = \{Y_{i1}, Y_{i2}, \dots, Y_{iN^{(o)}}\}$  corresponds to the sequence of observations of state variable  $i$  at time step  $t$ ,  $N^{(o)}$  is the total number of time steps in the time series,  $\epsilon_{it}^{(o)}$  is the observation error at time step  $t$  between the interpolated and observed value of variable  $i$ ,  $M^{(o)}$  is the total number of parameters. More details on how to derive this expression can be found in a supplementary file (Supplementary A).

Then, we define a simple Bayesian model to fit the NODEs to the interpolated dynamics, given the

interpolated states. We assume normal distributions for the observation error,  $\varepsilon_{it}^{(p)} \sim \mathcal{N}(0, \sigma_i)$ , and parameters,  $\theta_{ij} \sim \mathcal{N}(0, \delta_{ij})$ . This gives the following expression for the log posterior density of the parameters given the interpolations,

$$\log p(\theta_i \mid \omega) \propto -\frac{1}{2} \sum_{t=1}^{N^{(p)}} \left( \frac{\varepsilon_{it}^{(p)}}{\sigma_i} \right)^2 - \frac{1}{2} \sum_{j=1}^{M^{(p)}} \left( \frac{\theta_{ij}}{\delta_{ij}} \right)^2 \quad (8)$$

where  $\theta_i = \{\theta_{i1}, \theta_{i2}, \dots, \theta_{iM^{(p)}}\}$  are the NODE parameters of the  $i^{th}$  variable,  $\omega = \{\omega_1, \omega_2, \dots, \omega_I\}$  are the interpolation parameters of each state variable,  $\varepsilon_{it}^{(p)}$  is the process error of variable  $i$  at time step  $t$  between the interpolated dynamics and NODE prediction,  $\sigma_i$  is the standard deviation of the likelihood,  $M^{(p)}$  is the total number of parameters,  $\delta_{ij}$  is the standard deviation of the prior distribution of parameter  $\theta_{ij}$ .

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,  $\delta_{ij}$  (Cawley and Talbot 2007; Bonnaffé, Sheldon, and Coulson 2021). We could set small values of  $\delta$  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  $\delta_i$  on NODE parameters via cross-validation. We split the interpolated data into a train, validation, and test set, for instance, in three thirds. We fit the NODE model to the train set and predict the validation set. We repeat this process for increasing values of  $\delta_i$ , until we find the value that maximises the log likelihood of the validation data. We can perform multiple folds of validation by swapping the train and validation set, or by varying the size of the train/validation split. Ultimately, once we have

218 identified the appropriate value of the constraint parameters  $\delta_i$ , we fit the model to both the training  
219 and validation set, and assess the accuracy of the predictions on the test data, which is never seen  
220 by the model during training.

## 221 **2.4 Inference and uncertainty quantification**

222 Finally, we estimate uncertainty in parameter values by anchored ensembling, which produces ap-  
223 proximate Bayesian estimates of the posterior distribution of the parameters (Pearce et al. 2018).  
224 This involves sampling a parameter vector from the prior distributions,  $\theta_i \sim \mathcal{N}(0, \delta_i)$ , and then  
225 optimising the posterior distribution from this starting point,  $\theta_i^* = \underset{\theta_i}{\operatorname{argmax}} \log p(\theta_i | \omega)$ . By repeat-  
226 edly taking samples, the sampled distribution  $\theta^*$  approaches the posterior distribution and provides  
227 estimates and error around the quantities that can be derived from the models. The expectation and  
228 uncertainty around derived quantities can then be obtained by computing the mean and variance of  
229 the approximated posterior distributions. The strength of this approach is that it is unlikely to get  
230 stuck in local maxima hence providing a more thorough exploration of the parameter space.

## 231 **2.5 Analysing NODEs**

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

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

236 Knowing the direction, however, is not sufficient to determine the importance of a variable for the  
 237 dynamics of another. Given the same effects, a variable that fluctuates a lot will have a greater  
 238 impact on the dynamics of a focal variable, compared to a variable that remains quasi-constant.  
 239 For example, a predator can have a negative effect on the prey population, but its actual im-  
 240 pact/contribution to the dynamics of the prey population depends on its own dynamics, that is  
 241 if the predator population decreases, it has a positive contribution to the change in growth rate of  
 242 the prey population. We hence compute the strength of the interaction by multiplying the dynamics  
 243 of a variable  $y_j$  by its effect on the focal variable  $y_i$ , also known as the Geber method (Eq. 3 in  
 244 Hairston et al. 2005),

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

245 To summarise results across the entire time series we can compute the mean effects  $e_{ij}$  by aver-  
 246 aging  $e_{ijt}$  across all time steps,  $e_{ij} = 1/N^{(p)} \sum_t e_{ijt}$ , as well as the relative total contribution,  $c_{ij}$ ,  
 247 of a variable to the dynamics of another by computing the relative sum of square contributions,  
 248  $c_{ij} = \left( \sum_{ijt} c_{ijt}^2 \right)^{-1} \sum_t c_{ijt}^2$ . By computing the direction and strength of interactions between all the  
 249 variables in the system we can build dynamically informed ecological interaction networks (e.g.  
 250 Fig. 5). Other metrics can be computed by analysing the NODEs, such as equilibrium states, these



are discussed in our previous work (Bonnaillé, Sheldon, and Coulson 2021).

### 3 Case studies

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

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

##### System

We consider a tri-trophic ODE system consisting of a prey, an intermediate predator, and a top predator. The system is built on the real tri-trophic system featuring algae, flagellates, and rotifers, 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,  $\alpha$  is the prey intrinsic growth rate, limited by a carrying capacity  $\kappa$ ,  $\beta$  and  $\gamma$  are the predation rates by the intermediate and top predator,  $\delta$  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),  $\phi$  is the predation rate of the intermediate predator by the top predator,  $\mu$  and  $\nu$  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.

### **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 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 simple linear model to a nonlinear model by releasing the constraint on the parameters of the network during cross-validation. The number of units in the hidden layer  $J$  is chosen to be 10, as

281 this is a commonly used number for systems of that size (e.g. Wu, Fukuhara, and Takeda 2005;  
 282 Bonnaffé, Sheldon, and Coulson 2021).

### 283 **Time series interpolation**

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

### 295 **Fitting NODEs to the interpolated time series**

296 We fit the NODE system to the interpolated time series. In practice, we fit the NODE to the expect-  
 297 ation of the interpolated state and dynamics,  $E(\tilde{y}_i)$  and  $E(d\tilde{y}_i/dt)$ , by averaging over all sampled  
 298 interpolation parameters. An alternative approach could be to consider the interpolation that max-  
 299 imises the log marginal posterior density, but this may decrease repeatability due to the difficulty of  
 300 reliably identifying a global maximum. Averaging across multiple interpolations ensures an overall

301 smoother and robust interpolation. In addition, we standardise the response and explanatory vari-  
 302 ables with respect to their mean and standard deviation (i.e.  $Z = (Y - \mu)/\sigma$ ). This is to facilitate  
 303 the training of the NODE by equalizing the scale of the different parameters in the neural network.  
 304 Then, we identify the optimal regularisation parameter  $\delta$  (Eq. 8) by cross-validation. To do that,  
 305 we split the data in three thirds, train NODEs on the first third, and calculate the log likelihood of  
 306 the validation set for increasing values of  $\delta$ , from 0.1 (linear) to 1.0 (highly nonlinear), by incre-  
 307 ments of 0.1. This allows us to identify the maximum degree of nonlinearity,  $\delta$ , in the per-capita  
 308 growth rate that ensures generalisability throughout the time series. Then, we approximate the  
 309 posterior distribution of the NODE parameters by taking 30 samples from the posterior distribution  
 310 (Eq. 8). We ensure moderate temporal autocorrelation and normality by visualising the residuals  
 311 of the models. We also ensure results repeatability by running the entire fitting process a second  
 312 time.

### 313 **Computing ecological interactions**

314 Finally, we analyse the shape of the per-capita growth rates to recover the interaction between the  
 315 three species in the system. In particular, we look at the effect and contribution of each species  
 316 to the dynamics of the others. The effect is computed as the sensitivity (i.e. the gradient) of the  
 317 per-capita growth rate of a given species with respect to the density of the other species (Sugihara  
 318 et al. 2012; Bonnaiffé, Sheldon, and Coulson 2021). The contribution is computed following the  
 319 Geber method (Hairston et al. 2005), which consists in multiplying the dynamics of a variable by  
 320 its effects on the other variables. We further compute the importance of a species in driving the

321 dynamics of another by computing its relative total contribution compared to other species. More  
322 details on how to compute these quantities can be found in section 2.5 and in our previous study  
323 (Bonnaffé, Sheldon, and Coulson 2021).

## 324 **Benchmark**

325 In order to demonstrate the suitability of BNGM for fitting NODEs and inferring ecological in-  
326 teractions we compare our approach to three existing methods. For this purpose, we focus on the  
327 artificial time series, as this offers the possibility for comparing predictions to the truth, known  
328 from the equations that generated the time series.

329 We first consider a standard NODE model (Bonnaffé, Sheldon, and Coulson 2021), as our BNGM  
330 approach seeks to alleviate the computational cost of fitting NODEs. We define the per capita  
331 growth rate as an ANN with a single layer, 3 inputs, 10 hidden nodes, and exponential activa-  
332 tion functions. We use a Bayesian model, assuming log normal distributions for species den-  
333 sity  $Y_i \sim \log \mathcal{N}(y_i, \sigma_i)$ , and uniform uninformative prior distributions for the network parameters  
334  $\theta_i \sim \mathcal{U}(-10, 10)$ , initial densities  $y_{i0} \sim \mathcal{U}(0, 10)$ , and variance  $\sigma_i \sim \log \mathcal{N}(0.5, 0.5)$ . Our imple-  
335 mentation of standard NODEs differs from our BNGM approach in three ways. First, the standard  
336 NODE ANN has 3 outputs instead of one, as variables are fitted jointly. Second, computing the  
337 posterior density of the parameters requires to solve the NODE system with a numerical ODE  
338 solver (Runge-Kutta, package *deSolve*). Third, we do not constrain the parameters of the network  
339 given that the prohibitive fitting times prevent the tuning of the regularisation parameters.

340 We also consider a parametric ODE model, as this is the closest parametric alternative to NODEs

341 to infer ecological interactions. This model only differs from the standard NODE model in that  
342 the per-capita growth rate is approximated by second order polynomial functions  $r_i(y, \theta_i) = \theta_i^{(0)} +$   
343  $\sum_j \theta_{ij}^{(1)} y_j + \sum_j \sum_k \theta_{ijk}^{(2)} y_{ij} y_{ik}$ , instead of an ANN, which can handle simple nonlinearities.

344 To ensure the most meaningful comparison, we implemented the NODEBNGM, standard NODE,  
345 and parametric ODE models in base R, using BFGS for optimisation (function *optim*, R v4.2.0). We  
346 also followed a similar fitting procedure by independently training 30 models on the train/validation  
347 set (i.e. 2/3 of the time series) and predicting the test set (remaining third).

348 Finally, we implement convergent cross-mapping (CCM). This technique performs locally linear  
349 approximations of the state space of the system to estimate the sensitivity of the dynamics of a  
350 variable to a change in other variables (Sugihara et al. 2012). For this we use the package rEDM  
351 (v1.13.1, Sugihara et al. 2012), and adapt the example code provided for the three species system.  
352 We train the CCM model on the train set and predict outcomes on the test set. We then retrieve  
353 s-map coefficients (i.e. the interactions) and approximate the population dynamics and per-capita  
354 growth using finite differences, given that the standard implementation of CCM does not provide  
355 these estimates by default.

356 For all four methods, we compute the runtime as the average time required to train a single model.  
357 Using the best performing model on the train set, we then predict the population dynamics, growth  
358 rate, and ecological effects for the entire time series, including the test set. We compute the accu-  
359 racy of the predictions by computing the mean sum of squared error (MSE) of predictions versus  
360 the truth for both the per capita growth rate and ecological effects, on the train and test set. We also

361 build the corresponding dynamical interaction networks, using the inferred mean effects and total  
362 contributions, and compare them to the true network of interactions. Results are shown in Figure 4  
363 and presented in detail in supplementary figures (Fig. S1-5).

## 364 **3.2 Case study 2: real tri-trophic prey-predator oscillations**

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

### 368 **System**

369 We consider a three-species laboratory microcosm consisting of an algal prey (*Chlorella autroph-*  
370 *ica*), a flagellate intermediate predator (*Oxyrrhis marina*), and a rotifer top predator (*Brachionus*  
371 *plicatilis*). The algal prey is consumed by the intermediate and top predator, which also consumes  
372 the intermediate predator (Arndt 1993). The dynamics of this system, here the daily change in  
373 the density of each species, were recorded in three replicated time series experiments performed  
374 by Hiltunen and colleagues (Hiltunen et al. 2013). We use their time series because they describe  
375 a simple yet biologically realistic ecosystem, and because the quality of the replication of their  
376 microcosm reduces as much as possible observational and experimental error, and rules out envi-  
377 ronmental variation (Hiltunen et al. 2013). We digitised these time series by extracting by hand  
378 the coordinates of every points in the referential of the axis of the graph of the original study, and  
379 analysed them.

## 380 **NODE analysis**

381 We apply the same analysis as performed on the artificial tri-trophic prey-predator oscillations.  
382 This allows us to recover a nonparametric approximation of the growth rate of each species, and  
383 then derive the direction and strength of the ecological interactions that underpin their dynamics.  
384 We present detailed results of the analysis of the second time series (Fig. 5), and a summary  
385 comparison of the three time series (Fig. 6). Complementary results, including cross-validation  
386 plots, and detailed results for the other two replicates can be found in the supplementary material  
387 (Supplementary C-E).

## 388 **3.3 Case study 3: real di-trophic prey-predator oscillations**

389 We infer ecological interactions by NODE BNGM in the hare-lynx system (Odum and Barrett  
390 1972). This is to provide an example of a longer time series, and to offer a point of comparison  
391 with previous and future implementations of NODEs, which commonly use this time series (e.g.  
392 Bonnaffé, Sheldon, and Coulson 2021; Frank 2022).

## 393 **System**

394 The system is described in details in our previous work (Bonnaffé, Sheldon, and Coulson 2021).  
395 The data consist in a 90-year long time series of counts of hare and lynx pelts collected by trappers  
396 in the Hudson bay area in Canada (Odum and Barrett 1972). The time series displays characteristic  
397 10-year long prey-predator oscillations.

## 398 **NODE analysis**



399 We apply the same analysis as previously described, to the exception that the NODE system only  
400 features two variables,  $H$  and  $L$ , instead of 3. Results are presented in Figure 7.

### 401 **3.4 Case study 4: Maizuru bay aquatic community**

402 Finally, we demonstrate the capacity of NODEs to analyse the drivers of the dynamics of a larger  
403 community by analysing the time series of the Maizuru bay community (Ushio et al. 2018).

#### 404 **System**

405 The dataset for this system consists of 12-year long time series of fortnight abundance estimates of  
406 the 15 dominant species in the Maizuru bay, Japan. The data was collected every two weeks along  
407 three 200m long and 2m wide transects by underwater visual census conducted along the coast of  
408 the Maizuru fishery research station of Kyoto University from 2002 to 2014 (for more details see  
409 Ushio et al. 2018). Bottom sea temperature (at 10m depth) was also recorded on each census. The  
410 dataset contains 14 dominant species of fish and 1 genus of jellyfish. Only species with more than  
411 1000 sightings were included in the final dataset.

412 We focussed our analysis on the species with the least sparse records. We discarded the follow-  
413 ing species from our analysis *Engraulis japonicus*, *Plotosus lineatus*, *Chaenogobius gulosus*, and  
414 *Siganus fuscescens*. We also excluded periods which presented jellyfish blooms, as these were  
415 isolated events which could cause numerical errors in the estimation of the dynamics of species  
416 abundance. In total, we considered a time period of a hundred time steps from June 2004 to  
417 August 2008, and 11 species out of 15, namely *Aurelia* sp, *Sebastes inermis*, *Trachurus japoni-*

418 *cus*, *Girella punctata*, *Pseudolabrus sieboldi*, *Halichoeres poecilopterus*, *Halichoeres tenuispinnis*,  
419 *Pterogobius zonoleucus*, *Tridentiger trigonocephalus*, *Sphyræna pinguis*, and *Rudarius ercodes*.  
420 We included the sea bottom temperature (in degrees celsius) as an additional environmental vari-  
421 able.

## 422 **NODE analysis**

423 We then analysed this dataset following the approach described in the method section. We split  
424 the data into three thirds to create a training, validation, testing set (final third), and we followed  
425 the same procedure as described before to tune the regularisation parameters. The NODE system  
426 consisted of 11 NODEs, where the per-capita growth rate is determined by a single-layer ANN with  
427 12 input nodes, 10 hidden nodes, exponential activation functions, and 1 output node. Due to the  
428 high dimensionality of the effects obtained (11 by 12), we only present mean effects and relative  
429 total contributions, obtained by taking the mean, and the relative mean squares, respectively, of  
430 the effects and contributions across the entire time series (Fig. 8). The time series of effects  
431 and contributions are presented in greater details in the supplementary material (Supplementary  
432 G).

## 433 **4 Results**

### 434 **4.1 Case study 1: artificial tri-trophic system**

435 We present the results of fitting NODEs by BNGM to the artificial tri-trophic time series in Figure  
436 2 and 3. We find that both the interpolation of the state variables and dynamics are highly accurate  
437 (Fig. 2), given that they closely match the ground truth, known from the equations of the ODE  
438 model that we used to generate the time series (Eq. 11). Similarly, we find that the NODE approx-  
439 imation of the per-capita growth rate of each species also closely matches the ground truth (Fig. 3,  
440 a., d., g.). We find negative nonlinear effects of the two predators on the growth rate of the algae  
441 (Fig. 3, b., blue and purple lines). This nonlinear pattern is mirrored by the effect of the algae on  
442 the growth rate of the predators (Fig. 3, e. and h., red line). The interaction between the two preda-  
443 tors is also well-recovered (Fig. 3, e., blue line, and h., purple line), in spite of a slight tendency  
444 for overestimating the degree of nonlinearity of effects. The BNGM approach hence accurately  
445 recovers the dynamical characteristics of the artificial system.

### 446 **4.2 Benchmark**

447 Figure 4 shows the performance of fitting NODEs by BNGM, compared to standard NODEs, para-  
448 metric ODEs, and convergent cross mapping models (CCM). We find that fitting NODEs by BNGM  
449 provides the highest estimation accuracy for growth rates and ecological effects, both on the train-  
450 ing and test set, as well as competitive runtimes. Standard NODEs provide similar estimation  
451 accuracies, but take over 15 minutes to train. CCM is the fastest technique, as results are obtained

452 in under a second, but is relatively less accurate. Parametric ODEs are found to be both slow and  
453 less accurate.

454 We present a detailed breakdown of the runtime of fitting NODEs by BNGM for each system in  
455 Table 1. We find that it takes on average 5.35 minutes to fit NODEs by BNGM on the smaller  
456 systems with 3 or less species, and about 23 minutes to fit the larger system from the Maizuru  
457 bay, which features 12 variables. This includes performing  $100 \times I$  and  $30 \times I$  full optimisations  
458 of the posterior distribution of the interpolation and NODE parameters, respectively. This amounts  
459 to about 5.37 second to sample each variable of the NODE system once in the smaller systems,  
460 and about 28 seconds for the Maizuru bay community. This is over a 100 fold improvement over  
461 standard NODE models, which take on average 20 minutes (Fig. 4).

### 462 **4.3 Case study 2: real tri-trophic prey-predator oscillations**

463 We present an in-depth analysis of the drivers of the dynamics of the algae, flagellate, and rotifer  
464 population in replicate B (Fig. 5). We find slightly positive nonlinear intra-specific density depen-  
465 dence in algal growth (Fig. 5, b., red line), and negative nonlinear inter-specific effects of the two  
466 predators (purple and blue line). We find that the growth rate of the flagellate is driven by a positive  
467 effect of algal density, a negative effect of predation by the rotifer and intra-specific density de-  
468 pendence (Fig. 5, e. and f.). The rotifer population is almost solely driven by a positive nonlinear  
469 effect of algal density (Fig. 5, h., red line). Overall, comparing results across the three replicates  
470 reveals that the effect of the rotifer population on the flagellate and algae, and the effect of the algae

on the rotifer, are the strongest and most consistent interactions (Fig. 6, Table 2). The interactions of the flagellate with the algae, and its effect on the rotifer population varies substantially across replicates (Fig. 6, Table 2).

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

We present the analysis of the drivers of the hare-lynx population dynamics in Figure 7. Cross-validation provides weak support for nonlinear effects in the per-capita growth rate of the hare and lynx. We find that the hare population growth rate is mostly determined by a nonlinear negative effect of the lynx population (Fig. 7, b. and c. blue line), and by weak nonlinear positive density dependence (red line). The lynx growth rate is determined by a positive nonlinear effect of the hare (Fig. 7, e. and f., red line), and to a lesser extent by negative nonlinear intra-specific density dependence (blue line).

#### 4.5 Case study 4: drivers of the Maizuru bay community dynamics

We show the results of the NODE analysis of the drivers of the dynamics Maizuru bay community in Figure 8. Our main finding is that the chameleon goby (*T. trigonocephalus*) has a strong negative effect on 8 of the 11 dominant species of the community. We find that *E. ercodes* also has a strong negative impact on other species in the community, although relatively smaller than that of the chameleon goby. We find a positive effect of sea bottom temperature on the growth rate of the chameleon goby. Other effects are found to be mostly positive and have a relatively smaller impact on community dynamics.

## 5 Discussion

Characterising ecological interactions from time series data is challenging. This is due to the fact that interactions can be highly context-dependent processes (Song and Saavedra 2021), making it difficult to identify parametric models that encapsulate their complexity (Wood 2001). Interactions estimated with parametric models are contingent on the parameterisation arbitrarily chosen by the observer, and hence risk being biased (Wood 2001; Adamson and Morozov 2013). We provide a novel method for estimating ecological interactions nonparametrically, by using neural ordinary differential equations (NODEs) fitted with Bayesian neural gradient matching (BNGM). First, we remove the cost of fitting NODEs by introducing BNGM, which allows for NODE fitting in only a few seconds. The method involves interpolating time series and dynamics with neural networks, and then fitting NODEs to interpolated dynamics with Bayesian regularisation. We further demonstrate that this approach is accurate, as NODEs approximate with minimal error the ecological interactions in artificial time series, where real interactions are known, performing better than three existing methods. Finally, we estimate the strength, direction, importance, and nonlinearity of ecological interactions in 3 natural and experimental systems, showing variation in ecological interactions within and across the time series.

### Performance of NODEs fitted by BNGM compared to existing methods

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 (Bonnaiffé, Sheldon, and Coulson 2021). We demonstrate that NODEs accurately recover the dynamics, strength,

510 direction, and nonlinearity of ecological interactions in artificial tri-trophic prey-predator time se-  
511 ries, where truth is known. In particular, we find that the interactions between the prey and the  
512 two predators are nonlinear, and thereby oscillate throughout the time series, which is consistent  
513 with the model, that features a resistance to predation at high prey density. We also recover the  
514 interactions between the two predators, in spite of a slight tendency to overestimate the degree of  
515 nonlinearity. To our knowledge, this is the first assessment of the accuracy of NODEs in recovering  
516 interactions between variables from time series data, as most of the work focuses on assessing the  
517 accuracy of the fitting and forecasting of time series (e.g. Mai, Shattuck, and O’Hern 2016; Chen  
518 et al. 2019; Treven et al. 2021; Frank 2022).

519 We find that fitting NODEs by BNGM provides higher estimation accuracies of ecological inter-  
520 actions, and reduces substantially fitting times compared to standard NODEs (Bonnaiffé, Sheldon,  
521 and Coulson 2021), and parametric ODEs (Rosenbaum et al. 2019). This difference is attributable  
522 to three factors. First, BNGM alleviates the need for solving numerically the NODE system, which  
523 makes it faster to evaluate the posterior distribution. Second, it allows for the calculation of ana-  
524 lytical gradients of the posterior distribution, which greatly improve the speed and efficiency of the  
525 gradient descent optimisation algorithm. Finally, it makes it possible to fit each variables indepen-  
526 dently on each other, which results in a simpler optimisation problem.

527 CCM remains faster than our approach in recovering estimates of ecological interactions (Sugihara  
528 et al. 2012), however its accuracy is lower. A possible explanation for this comes from the fact  
529 that CCM computes the sensitivity of the total population growth rate, rather than the per-capita

530 growth rate, which can change estimated effects. Additionally, CCM relies on piecewise linear  
531 reconstructions of the state space (Deyle et al. 2015), whereas NODEs computes a global nonlinear  
532 approximation of the per-capita growth rate on the entire range covered by the data. We view the  
533 former as potentially more sensitive to local noise in the state space (Cenci, Sugihara, and Saavedra  
534 2019), compared to the latter, which uses all evidence available to inform local inference.

535 Our BNGM approach extends standard gradient matching, by using artificial neural networks  
536 (ANNs) as interpolating functions, and Bayesian regularisation to control the nonlinearity of the  
537 processes (Cawley and Talbot 2007). The use of ANNs as interpolating functions sets it apart  
538 from the initial approach of Ellner et al., who use splines to interpolate the time series before ap-  
539 proximating the ODEs (Ellner, Seifu, and Smith 2002). ANNs are more general and flexible than  
540 splines, as well as being easier to manipulate given that they are defined continuously on the state  
541 space, which is especially useful when handling multiple interactions between variables. Our ap-  
542 proach is related to that of Wu et al., who use ANNs to approximate both the states and ODEs of  
543 prey-predator systems (Wu, Fukuhara, and Takeda 2005), as well as that of Treven and colleagues,  
544 who developed the Gaussian process equivalent (Treven et al. 2021). In both approaches, they  
545 train the interpolation functions at the same time as the NODEs, in order to constrain the interpola-  
546 tion of trajectories such that they can be achieved by the NODE system, which thereby introduces  
547 dynamical coupling between state variables. One of the downsides of this approach is that mis-  
548 timating one of the state variables of the model biases the estimation of the states and dynamics  
549 of other variables. To avoid this, we fit each interpolation and NODE independently to each time



550 series. In addition, this makes it possible to parallelise the code, resulting in potentially even faster  
551 computation.

552 Our approach opens new possibilities for nonparametric inference of ecological interactions from  
553 time series data. The lower fitting times makes it possible to tackle larger systems, quick and  
554 extensive model comparison, cross-validation, and apply more thorough statistical treatments of the  
555 uncertainty of these models, for instance by implementing Markov-chain Monte-Carlo (MCMC)  
556 sampling.

### 557 **Ecological interactions in real prey-predator systems**

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

570 the system onto different attractors (Yoshida et al. 2003; Yoshida et al. 2007; Hiltunen et al. 2013).  
571 Additionally, stochasticity in population dynamics may have a similar effect (Dallas et al. 2021).  
572 Disentangling these two sources of variation would require refining the modelling framework, for  
573 instance by explicitly including evolution in the model (e.g. with the Price equation, Ellner, Geber,  
574 and Hairston 2011), and by using neural stochastic differential equations (i.e. NSDEs, Rackauckas  
575 et al. 2019) fitted with a particle filter. While these would constitute interesting developments, our  
576 method is still a useful first step, identifying differences between the time series, and demonstrating  
577 a reasonable amount of deterministic consistency in the dynamics, judging by the cross-validation  
578 and fits.

579 We also analysed the hare-lynx time series (Odum and Barrett 1972), as it is a common benchmark  
580 in the field of time series analysis, and provides a comparison point with our previous implemen-  
581 tation of NODEs (Bonnaffé, Sheldon, and Coulson 2021). As in our previous study, we found a  
582 predatory inter-specific interaction between lynx and hare, and negative intra-specific density de-  
583 pendence in the lynx. Evidence for positive density dependence in the hare was more limited than  
584 previously found. We also found stronger evidence for nonlinearity, as intra- and inter-specific ef-  
585 fects oscillated throughout the time series, as a result of density dependence. This difference with  
586 our previous study is due to the fact that our previous implementation of NODEs was based on sim-  
587 ulating the full NODE system, and hence imposed dynamical coupling between the variables. This  
588 dynamical coupling comes at a cost, if one variable is not explained well by the model, it will bias  
589 the interactions and dynamics of other variables. Here, the time series of lynx and hare are analysed

590 independently, each state variable is interpolated as closely as desired, its effects on the dynamics  
591 of other variables are hence even more robust to model misspecification than before.

592 We applied our approach to analyse the drivers of the dynamics of 11 species in the Maizuru bay  
593 (Ushio et al. 2018). We inferred  $11 \times 11$  ecological interactions, and 11 dependencies on water  
594 temperature. We found that the chameleon goby had a strong negative impact on the other species  
595 of the system, showing a strong competitive potential. This species is viewed as an aggressive  
596 competitor (Ushio et al. 2018), and is considered an invasive species in places where it has been  
597 introduced (Goren, Gayer, and Lazarus 2009). We also find a positive effect of temperature on  
598 the growth rate of the chameleon goby, which suggests that warming could have indirect negative  
599 effects on many species on Maizuru bay by favouring the reproduction of the goby.

600 Surprisingly, our results differ substantially from those obtained by Ushio et al. in their original  
601 analysis of the system with convergent cross mapping (Ushio et al. 2018). This may be due to a  
602 several factors. First, we considered a different set of species, as some of the time series that Ushio  
603 and colleagues used were too sparse to be suitable for our analysis. Additionally, this difference  
604 may be explained by the disparity in the estimation accuracies revealed by our benchmark anal-  
605 ysis, relating to fundamental mathematical differences between the two approaches, as discussed  
606 previously. If nothing else, our analysis of the Maizuru community dynamics demonstrates the  
607 usefulness of our BNGM method for fitting NODEs to a larger, more realistic system.

608 Overall, our approach provides a novel and powerful way of estimating interactions nonparametri-  
609 cally from time series data. The benefit of using NODEs is that they make no assumptions about

610 the nature of the ecological interactions that drive the dynamics of the species (Chen et al. 2019;  
611 Bonnaiffé, Sheldon, and Coulson 2021). Hence, we have a better chance at estimating the actual  
612 value of the interactions, knowing that it is not subjected to potential incorrect model specifications  
613 (Jost and Ellner 2000; Ellner, Seifu, and Smith 2002; Wu, Fukuhara, and Takeda 2005; Kendall  
614 et al. 2005; Adamson and Morozov 2013).

### 615 **Limits and prospects**

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

630 One further issue is that some interactions may depend on variables that are not observed. For  
631 instance, some population dynamics are strongly determined by their demographic state (Lande et  
632 al. 2002; Coulson et al. 2004), which would call for time series of the relevant demographic stages.  
633 In the system considered here, the dynamics of algae in the rotifer system are most likely coupled  
634 with that of nitrogen, for which no time series was available (Hiltunen et al. 2013). Our method  
635 only accounts for observed variables, so that time series for all important variables are required,  
636 though unaccounted variables are captured to some extent by nonlinear fluctuations in interactions.  
637 One interesting prospect would hence be to incorporate unobserved/latent state variables into the  
638 NODE system (Dupont, Doucet, and Teh 2019; Zhang et al. 2019; Frank 2022). Careful thought  
639 has to be given here as whether to use an ODE or NODE for the latent states given that they are not  
640 constrained by observations.

641 A further question is whether we could use similar approaches to analyse systems larger than the  
642 ones considered in this study. In particular, microbial communities feature thousands of species,  
643 and hence potentially millions of interactions, which poses a real problem for inference with dy-  
644 namical models (Ridenhour et al. 2017). Even the simplest linear ODE model would contain  
645 millions of parameters, hence entering deep learning territory. We believe that our success there is  
646 more readily limited by the availability of time series sufficiently long to identify this many inter-  
647 actions, rather than by our models. A possible next step could be to address the capacity of more  
648 complex NODE and ODE models to estimate interactions in large artificial communities (e.g. Co-  
649 enen et al. 2020), which could inform us on the relationship between model complexity and data

650 requirements in terms of time series length and sampling frequency.

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

## 658 **Conclusion**

659 We provide a method, BNGM, which allows for NODE fitting in a matter of seconds. This is a  
660 crucial step for efficient model selection and uncertainty quantification in NODEs. We also demon-  
661 strate that NODEs allow for faster more accurate estimation of the direction, strength, and nonlin-  
662 earity of ecological interactions than existing approaches, in a system where truth is known. Finally,  
663 we estimate ecological interactions in real prey predator systems, showing that only stronger inter-  
664 actions seem to be consistent across replicated time series, and that a single species can account for  
665 a large part of the changes in community dynamics. Our study allows for efficient NODE fitting,  
666 and confirms the power of NODEs in identifying dynamical coupling between populations.

## 667 **Acknowledgments**

668 We thank warmly the Ecological and Evolutionary Dynamics Lab and Sheldon Lab Group at the

669 department of Zoology for their feedback and support. We thank Ben Sheldon for insightful sug-  
670 gestions on early versions of the work. The work was supported by the Oxford-Oxitec scholarship  
671 and the NERC DTP.

## 672 **Data accessibility**

673 All data and code is available on GitHub at <https://github.com/WillemBonnafe/NODEBNGM>,  
674 as well as on Zenodo at <https://zenodo.org/record/7825866#.ZDgXh-zMIUE> (DOI: 10.5281/zen-  
675 odo.7825866).

## 676 **Statement of authorship**

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

## 679 **References**

- 680 Aarts, L. P. and P. V. D. Veer (2001). “Neural network method for solving partial differential equa-  
681 tions”. In: *Neural Processing Letters* 14 (3), pp. 261–271.
- 682 Adams, M. P. et al. (Apr. 2020). “Informing management decisions for ecological networks, using  
683 dynamic models calibrated to noisy time-series data”. In: *Ecology Letters* 23 (4), pp. 607–619.
- 684 Adamson, M. W. and A. Y. Morozov (2013). “When can we trust our model predictions? Un-  
685 earthing structural sensitivity in biological systems”. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 469 (2149), pp. 1–19.

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

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

691 Berryman, A. A. (2002). “Population: a central concept for ecology?” In: *Oikos* 97 (3), pp. 439–  
692 442.

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

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

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

701 Bonnaiffé, W., B. C. Sheldon, and T. Coulson (2021). “Neural ordinary differential equations for  
702 ecological and evolutionary time series analysis”. In: *Methods in Ecology and Evolution* 2, pp. 1–  
703 46.

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



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

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

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

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

717 Cenci, S., G. Sugihara, and S. Saavedra (May 2019). “Regularized S-map for inference and fore-  
718 casting with noisy ecological time series”. In: *Methods in Ecology and Evolution* 10 (5), pp. 650–  
719 660.

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

722 Coenen, A. R., S. K. Hu, E. Luo, D. Muratore, and J. S. Weitz (Apr. 2020). “A Primer for Micro-  
723 biome Time-Series Analysis”. In: *Frontiers in Genetics* 11.

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

726 Dallas, T., B. A. Melbourne, G. Legault, and A. Hastings (2021). “Initial abundance and stochas-  
727 ticity influence competitive outcome in communities”. In: *Journal of Animal Ecology*, pp. 1–  
728 26.

729 Deyle, E. R., R. M. May, S. B. Munch, and G. Sugihara (Jan. 2015). “Tracking and forecasting  
730 ecosystem interactions in real time”. In: *Proceedings of the Royal Society B: Biological Sciences*  
731 283, pp. 1–9.

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

733 Ellner, S. P., M. A. Geber, and N. G. J. Hairston (2011). “Does rapid evolution matter? Measuring  
734 the rate of contemporary evolution and its impacts on ecological dynamics”. In: *Ecology Letters*  
735 14 (6), pp. 603–614.

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

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

740 Funahashi, K. ichi and Y. Nakamura (1993). “Approximation of dynamical systems by continuous  
741 time recurrent neural networks”. In: *Neural Networks* 6 (6), pp. 801–806.

742 Goren, M., K. Gayer, and N. Lazarus (2009). “First record of the far East Chameleon goby tri-  
743 dentiger trigonocephalus (Gill, 1859) in the Mediterranean Sea”. In: *Aquatic Invasions* 4 (2),  
744 pp. 413–415.

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

747 Hairston, N. G. J., S. P. Ellner, M. A. Geber, T. Yoshida, and J. A. Fox (2005). “Rapid evolution  
748 and the convergence of ecological and evolutionary time”. In: *Ecology Letters* 8 (10), pp. 1114–  
749 1127.

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

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

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

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

760 Jost, C. and S. P. Ellner (2000). “Testing for predator dependence in predator-prey dynamics: A  
761 non-parametric approach”. In: *Proceedings of the Royal Society B: Biological Sciences* 267  
762 (1453), pp. 1611–1620.

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

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

- 767 Lande, R, S Engen, B.-E Saether, F Filli, E Matthysen, and H Weimerskirch (2002). “Estimating  
768 Density Dependence from Population Time Series Using Demographic Theory and Life-History  
769 Data”. In: *American Naturalist* 159, pp. 321–337.
- 770 Lawton, J. H. (1999). “Are There General Laws in Ecology ?” In: *Oikos* 84 (2), pp. 177–192.
- 771 Lingjaerde, O. C. et al. (2001). “Exploring the density-dependent structure of blowfly populations  
772 by nonparametric additive modeling”. In: *Ecology* 82 (9), pp. 2645–2658.
- 773 Mai, M., M. D. Shattuck, and C. S. O’Hern (2016). “Reconstruction of Ordinary Differential Equa-  
774 tions From Time Series Data”. In: *arXiv*, pp. 1–15.
- 775 Moe, S. J., A. B. Kristoffersen, R. H. Smith, and N. C. Stenseth (2005). “From patterns to pro-  
776 cesses and back: Analysing density-dependent responses to an abiotic stressor by statistical and  
777 mechanistic modelling”. In: *Proceedings of the Royal Society B: Biological Sciences* 272 (1577),  
778 pp. 2133–2142.
- 779 Mysterud, A., T. Coulson, and N. C. Stenseth (2002). “The role of males in the dynamics of ung-  
780 late populations”. In: *Journal of Animal Ecology* 71, pp. 907–915.
- 781 Novak, M. and D. B. Stouffer (Nov. 2021). “Geometric Complexity and the Information-Theoretic  
782 Comparison of Functional-Response Models”. In: *Frontiers in Ecology and Evolution* 9.
- 783 Odum, E. P. and G. W. Barrett (1972). “Fundamentals of Ecology”. In: *The Journal of Wildlife*  
784 *Management* 36 (4), p. 1372.
- 785 Pasquali, S. and C. Soresina (2018). “Estimation of the mortality rate functions from time series  
786 field data in a stage-structured demographic model for *Lobesia botrana*”. In: *arXiv*, pp. 1–15.

787 Pearce, T., F. Leibfried, A. Brintrup, M. Zaki, and A. Neely (2018). “Uncertainty in Neural Net-  
788 works: Approximately Bayesian Ensembling”. In: *arXiv*, pp. 1–10.

789 Rackauckas, C., M. Innes, Y. Ma, J. Bettencourt, L. White, and V. Dixit (Feb. 2019). “DiffEqFlux.jl  
790 - A Julia Library for Neural Differential Equations”. In: *arXiv*, pp. 1–17.

791 Ridenhour, B. J. et al. (Nov. 2017). “Modeling time-series data from microbial communities”. In:  
792 *ISME Journal* 11 (11), pp. 2526–2537.

793 Rosenbaum, B., M. Raatz, G. Weithoff, G. F. Fussmann, and U. Gaedke (2019). “Estimating param-  
794 eters from multiple time series of population dynamics using bayesian inference”. In: *Frontiers*  
795 *in Ecology and Evolution* 6 (234), pp. 1–14.

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

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

801 Song, C. and S. Saavedra (July 2021). “Bridging parametric and nonparametric measures of species  
802 interactions unveils new insights of non-equilibrium dynamics”. In: *Oikos* 130 (7), pp. 1027–  
803 1034.

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

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

808 Turchin, P. (1999). “Population Regulation: A Synthetic View”. In: *Oikos* 84 (1), pp. 153–159.

809 – (2001). “Does population ecology have general laws?” In: *Oikos* 94, pp. 17–26.

810 Turchin, P. et al. (2003). “Dynamical effects of plant quality and parasitism on population cycles  
811 of larch budmoth”. In: *Ecology* 84 (5), pp. 1207–1214.

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

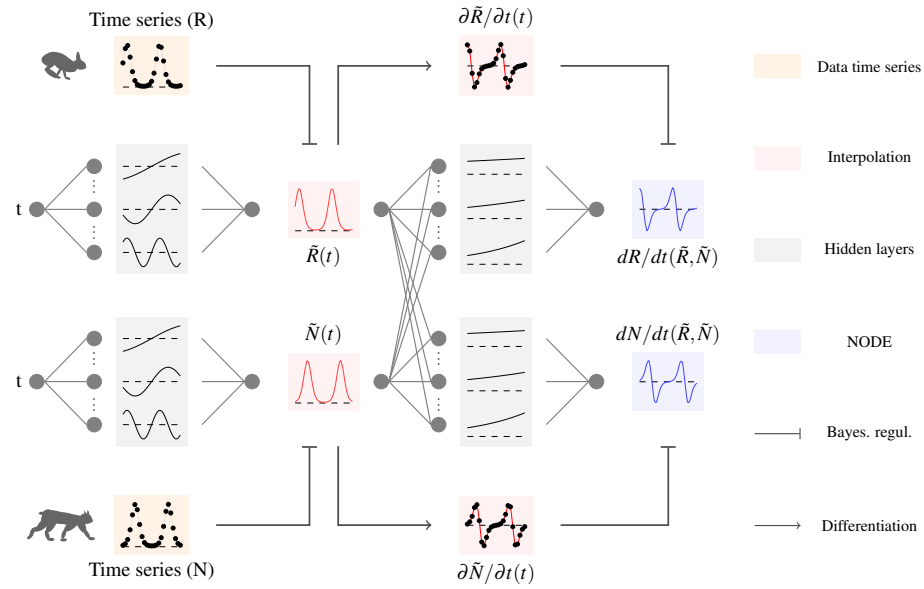
814 Wood, S. N. (2001). “Partially specified ecological models”. In: *Ecological Monographs* 71 (1),  
815 pp. 1–25.

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

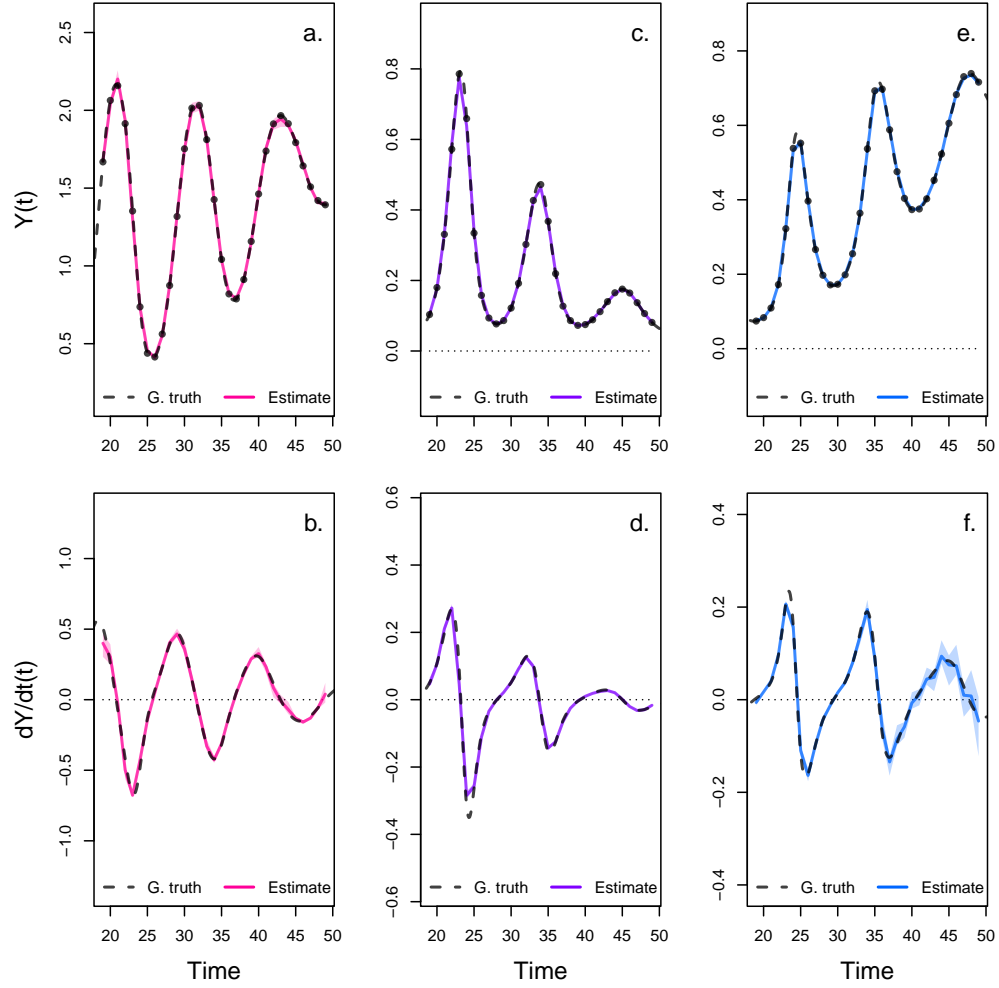
819 Yoshida, T., S. P. Ellner, L. E. Jones, B. J. M. Bohannan, R. E. Lenski, and N. G. J. Hairston (2007).  
820 “Cryptic population dynamics: Rapid evolution masks trophic interactions”. In: *PLoS Biology* 5  
821 (9), pp. 1868–1879.

822 Yoshida, T., L. E. Jones, S. P. Ellner, G. F. Fussmann, and N. G. J. Hairston (2003). “Rapid evo-  
823 lution drives ecological dynamics in a predator – prey system”. In: *Nature* 424 (July), pp. 303–  
824 306.

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

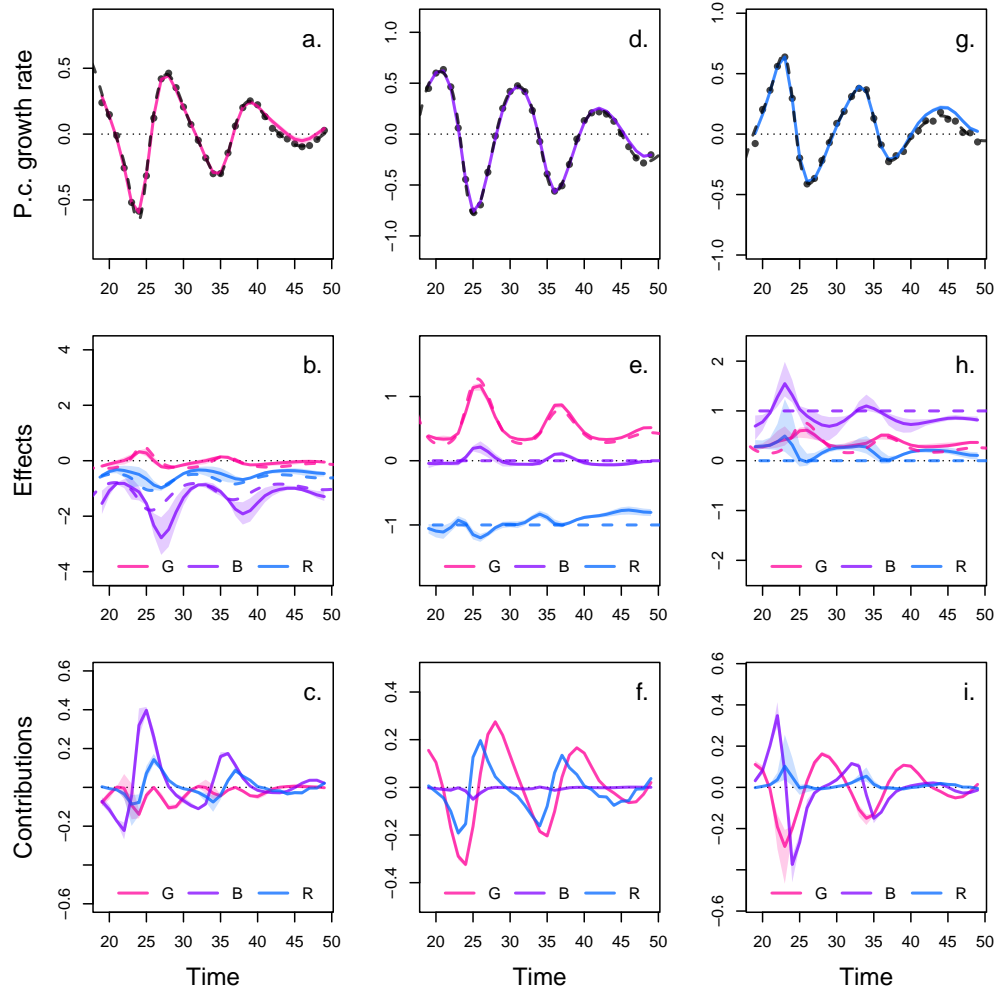


**Figure 1: Overview of fitting neural ordinary differential equations (NODE) by Bayesian neural gradient matching (BNGM).** In a first step we compute a continuous time approximation (interpolation) of each state variables, here the prey  $\tilde{R}(t)$  and predator density  $\tilde{N}(t)$  (red boxes). 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.



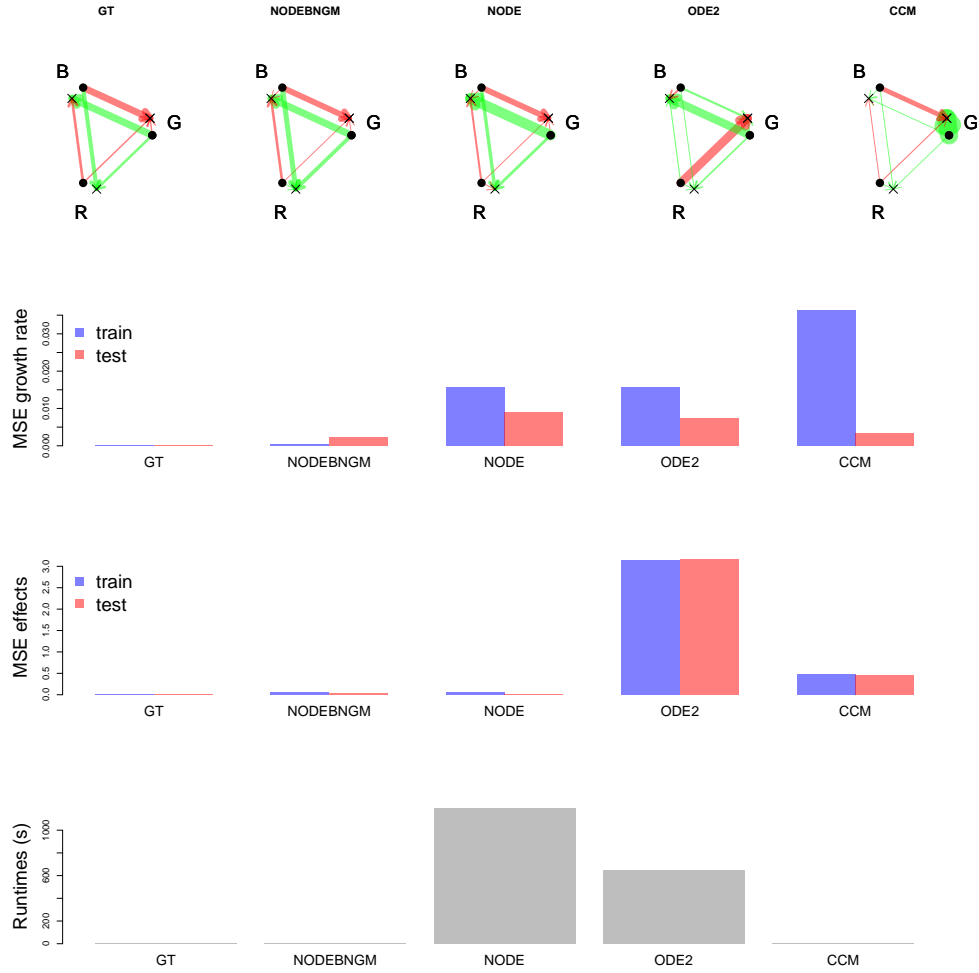
**Figure 2: Interpolated density and dynamics of prey, intermediate, and top predators in the artificial system.** This figure corresponds to the first step in the overview figure (Fig. 1). It shows the accuracy of the interpolated densities of prey (a.), intermediate (c.), and top predators (e.). We 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.



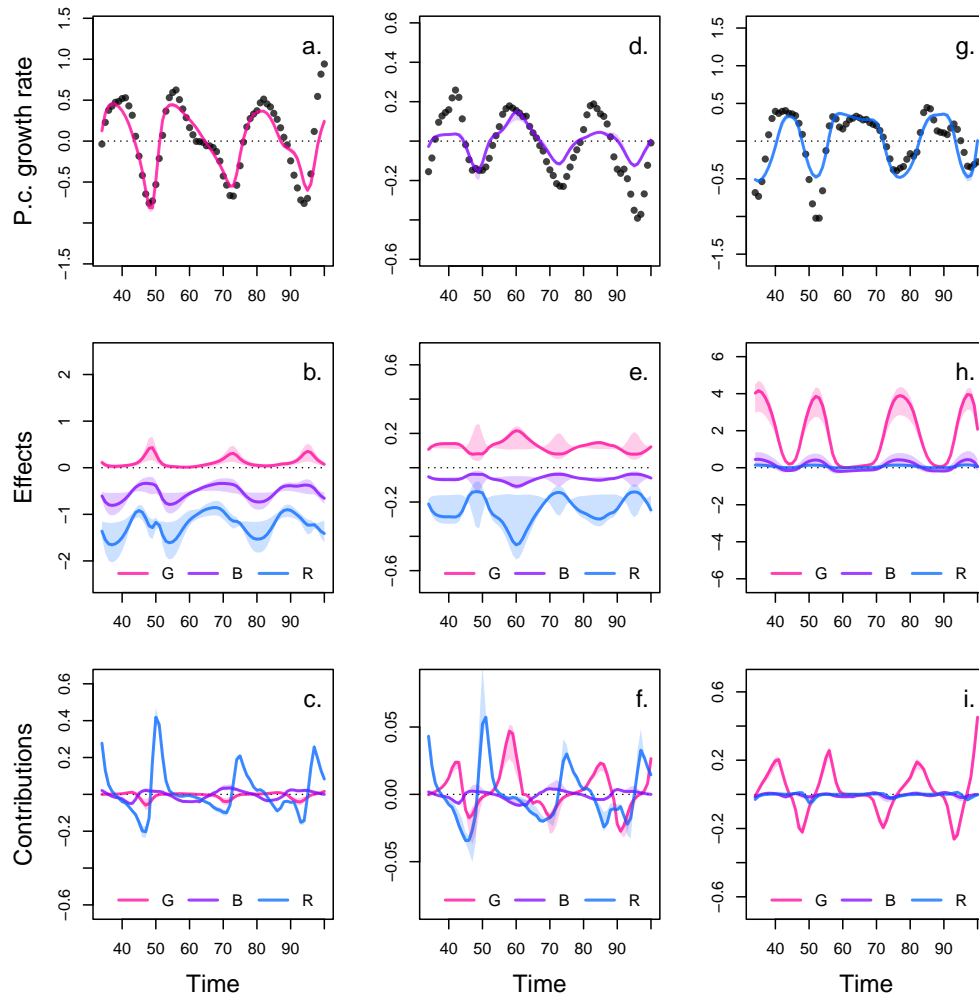


**Figure 3: Drivers of dynamics of prey, intermediate, and top predator in the artificial system.**

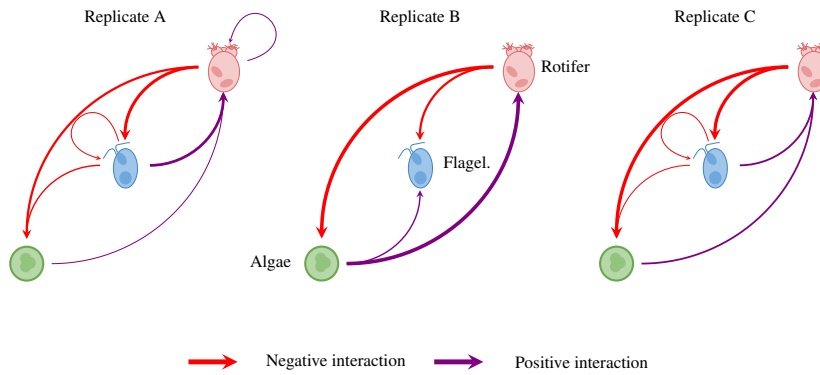
This figure corresponds to the second step in the overview figure (Fig. 1). It displays the NODE nonparametric approximations of the per-capita growth rate of prey (a., b., c.), intermediate (d., e., f.), and top predators (g., h., i.). We obtain the NODE approximations (a., d., g., solid line) by fitting 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.



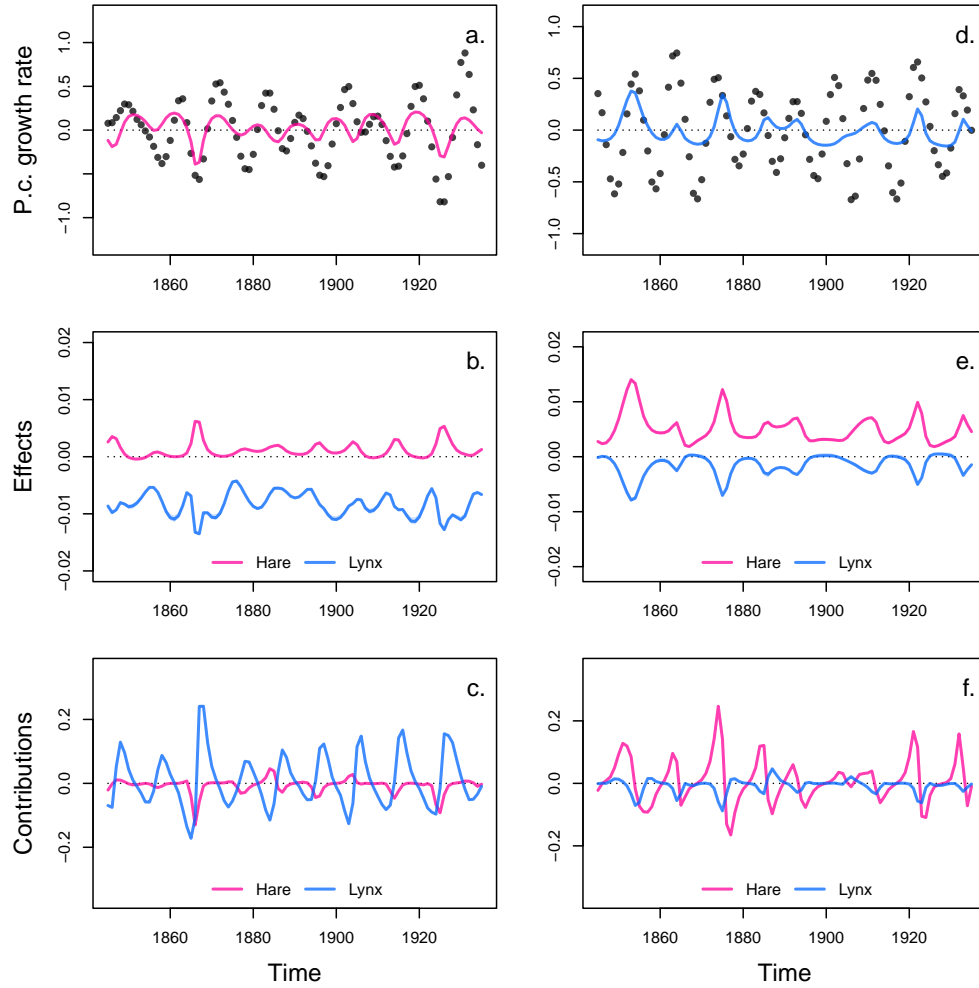
**Figure 4: Runtimes and accuracy of NODEs fitted by BNGM compared to standard NODEs, ODEs, and CCM.** The NODEBNGM method (nonparametric) involves fitting a NODE system by Bayesian neural gradient matching (BNGM). The NODE method (nonparametric) involves fitting a NODE system with an ODE solver. The ODE2 method (parametric) involves fitting an ODE system with quadratic functions of species densities with an ODE solver. The CCM method (non-parametric) involves computing locally linear approximations of the state space. For each method, we trained 30 models on the two first thirds of the artificial time series where ground truth is known (Fig. 2). We computed runtimes as the mean time (in seconds) required to train a single model. Using the best identified model, we predicted the growth rate and effects on the train and test set. We computed accuracies as the mean squared error of predictions vs ground truth (known from the equations that generated the data) (see Fig. S1-5 for more details). At the top, we show the dynamical interaction network of the system predicted by the best models, where G, B, R correspond to the prey, intermediate and top predator, respectively. Green and red colours correspond to positive and negative interactions, the width of arrows is proportional to relative total contribution to the growth rate of the population computed following the Geber method (Hairston et al. 2005; Bonnaiffé, Sheldon, and Coulson 2021).



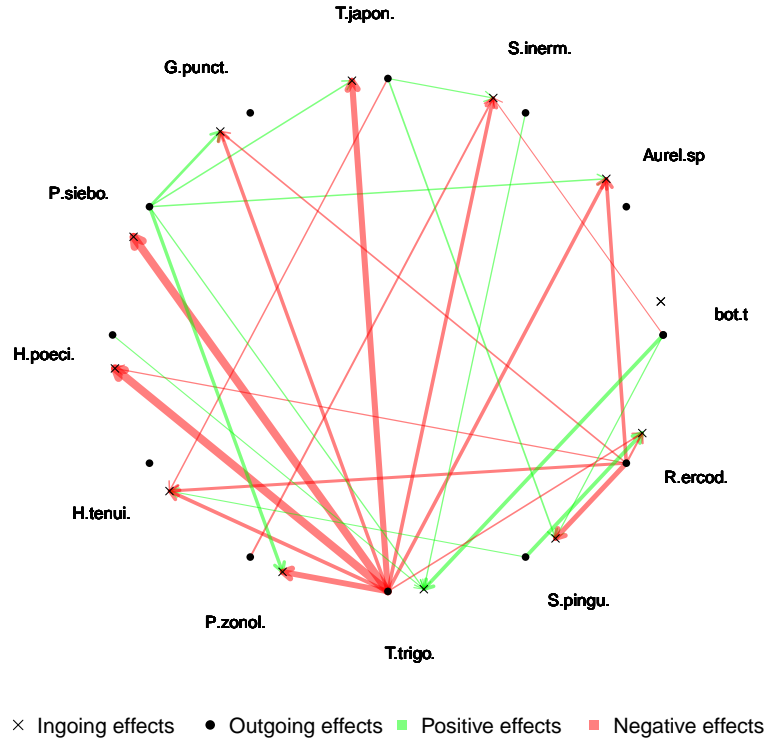
**Figure 5: Drivers of dynamics of algae, flagellate, and rotifer in replicate B.** This figure displays the NODE nonparametric approximations of the per-capita growth rate of algae (a., b., c.), flagellate (d., e., f.), and rotifer (g., h., i.). We obtain the NODE approximations (a., d., g., solid line) 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).



**Figure 6: 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 A and C were analysed in the same way as replicate B (see Fig. 5 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. 5 c., and comparing it to the sum of square of all contributions (Fig. 5 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 7: 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.



**Figure 8: Dynamical interaction network of the Maizuru bay community.** This figure summarises the results of the NODEBNGM analysis of the Maizuru bay community time series (Fig. S18). Bot.t corresponds to the temperature at sea bottom. Species are referred to by their shortened acronym. The main species are *Aurelia* sp., *Sebastes inermis*, *Trachurus japonicus*, *Girella punctata*, *Pseudolabrus sieboldi*, *Halichoeres poecilopterus*, *Halichoeres tenuispinnis*, *Pterogobius zonoleucus*, *Tridentiger trigonocephalus*, *Sphyræna pinguis*, and *Rudarius ercodes*. Red and green arrows correspond to negative or positive mean effects, obtained by averaging the sensitivity of the growth rate of a species to the density of other species across the time series. The width of the arrows is proportional to the relative total contribution (in %) of species density to the growth rate of other species, obtained by computing the relative sum the square of contributions across the time series. For the sake of clarity, we only display interactions which contribute to more than 10% of a change in the population growth rates.

**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 Ushio system corresponds to the 100 time step long time series dataset of the 11 dominant species in the Maizuru bay community (Ushio et al. 2018). 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). 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 t. steps	Interpolation		NODE fit		total	total p. fit
			N fits	time (s)	N fits	time (s)		
Replicate A	3	66	100	234.45	30	80.41	314.86	5.02
Replicate B	3	66	100	238.01	30	81.13	319.14	5.08
Replicate C	3	40	100	136.51	30	49.94	186.45	3.03
Hare-lynx	2	90	100	341.74	30	21.55	363.29	4.14
Ushio	12	100	100	806.12	30	604.42	1410.54	28.21

**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
<b>Replicate A</b>	$r^2$	0.11	0.37	0.47
<b>Mean effects</b>	on G	-0.08	-1.14	-1.13
	on B	0.28	-0.21	-0.66
	on R	0.60	1.09	0.32
<b>% of total contributions</b>	to G	0.01	0.34	0.65
	to B	0.02	0.04	0.93
	to R	0.26	0.66	0.08
<b>Replicate B</b>	$r^2$	0.52	0.4	0.65
<b>Mean effects</b>	on G	0.12	-0.53	-1.23
	on B	0.12	-0.06	-0.25
	on R	1.83	0.10	0.07
<b>% of total contributions</b>	to G	0.02	0.03	0.95
	to B	0.38	0.02	0.61
	to R	0.99	0.00	0.01
<b>Replicate C</b>	$r^2$	0.59	0.32	0.73
<b>Mean effects</b>	on G	0.09	-0.49	-1.96
	on B	0.04	-0.19	-0.65
	on R	1.07	0.79	-0.01
<b>% of total contributions</b>	to G	0.00	0.08	0.91
	to B	0.01	0.10	0.88
	to R	0.49	0.51	0.00



## 827 **6 Supplementary**

### 828 **A Bayesian regularisation**

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

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

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

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

833 where  $e_i(\mathcal{D}, \omega)$  are the residuals of the model given the parameters, and the evidence. In the case  
834 of the interpolation, the residuals correspond to the observation error  $\varepsilon^{(o)}$  (Eq. 3). In the case of  
835 the NODE approximation, they correspond to the process error  $\varepsilon^{(p)}$  (Eq. 7).  $N$  is the number of  
836 data points, either observations in the case of the interpolation,  $N^{(o)}$ , or interpolated points in the  
837 case of the NODE fitting,  $N^{(p)}$ .

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

$$p(\omega) = \prod_{j=1}^M \frac{1}{\sqrt{2\pi\delta_j^2}} \exp \left\{ -\frac{\omega_j^2}{2\delta_j^2} \right\}, \quad (15)$$

where  $M$  is the number of parameters in the models. The parameter  $\delta_j$  controls the dispersion of the priors, and thereby the complexity/level of constraint of the model.

Bayesian regularisation consists in constraining the values of the parameters in the model to be close to a desired value. Usually, parameters are constrained by choosing normal priors centered about 0. In this case, the standard deviation of the normal priors,  $\delta$ , governs the range of values that the parameters can take, and hence constrains more or less strongly the behaviour of the model (Cawley and Talbot 2007). Low values of dispersion may increase constraint on parameters too drastically, which would lead to underfitting, and result in a reduction of the variance of parameter estimates and bias mean estimates towards 0. In contrast, too high values of dispersion may lead to overfitting, by allowing for more complex shapes. To account for this, we optimise models on the second-level of inference. This means that we are finding the optimal value of  $\delta$ , in addition to optimising the model parameters.

In practice, choosing the level of constraint is difficult, Cawley and Talbot hence developed a criterion to perform model selection on the second level of inference. They proposed to optimise the marginal posterior distribution by averaging out the dispersion of the priors. With an improper Jeffrey's hyperprior,  $p(\xi) \propto \frac{1}{\xi}$ , on the dispersion parameters, the dispersion can be integrated out, leaving us with a formula for the posterior that only depends on the parameters of the model,

$$\log P(\omega|\mathcal{D}) \propto -\frac{N}{2} \log \left( \sum_{i=1}^N e_i(\mathcal{D}, \omega)^2 \right) - \frac{M}{2} \log \left( \sum_{j=1}^M \omega_j^2 \right), \quad (16)$$

where  $P(\omega|\mathcal{D})$  denotes the marginal posterior density,  $\mathcal{D}$  denotes the evidence,  $N$  and  $M$  denote

the number of data points and parameters, respectively,  $e_i$  denote the residuals, and  $\omega$  denote the parameters of the model. The construction is elegant because it is not sensitive to the choice of prior hyperparameters, and simple as it amounts to optimising the log of the sum of squares, rather than the sum of squares (in the case of normal ordinary least square).

One of the drawbacks of this construction is that the marginal posterior density is not finite when the parameters are 0, which can lead to underfitting. In this paper we use a modified criterion, which corrects for that problem,

$$\log P(\omega|\mathcal{D}) \propto -\frac{N}{2} \log \left( 1 + \sum_{i=1}^N e_i(\mathcal{D}, \omega)^2 \right) - \frac{M}{2} \log \left( 1 + \sum_{j=1}^M \omega_j^2 \right), \quad (17)$$

where the marginal posterior density depends only on the residuals of the model when the parameters are equal to 0, and otherwise depends on both the parameters and the residuals. This construction can be obtained by adding an exponential correction to the improper Jeffrey's hyperprior that Cawley and Talbot used in their original study, namely  $p(\xi) \propto \frac{1}{\xi} \exp\{-\xi/2\}$ , where  $\xi$  is the regularisation parameter.

We provide below details of the calculation of this modified criterion, starting by averaging out the regularisation parameter. We let  $\xi := 1/\delta^2$ , assuming that all parameters are controlled by the same regularisation parameter, which gives the following equation for the prior distribution,

$$p(\omega|\xi) = \prod_{j=1}^M \frac{1}{\sqrt{2\pi/\xi}} \exp\left\{-\frac{\xi}{2} \omega_j^2\right\}. \quad (18)$$

872 To follow the proof given by Cawley and Talbot we further use the following notation,  $d := M$ ,

873  $Z_\Omega := (2\pi/\xi)^{d/2}$ , and  $\Omega := \sum_j \omega_j^2$ , which yields

$$p(\omega|\xi) = Z_\Omega(\xi)^{-1} \exp\left\{-\frac{\xi}{2}\Omega(\omega)\right\}. \quad (19)$$

874 We can then integrate out the regularisation parameter by computing the marginal prior distribu-  
875 tion,

$$P(\omega) = \int p(\omega|\xi)p(\xi)d\xi. \quad (20)$$

876 This expression can be solved analytically with the right choice of prior. Cawley and Talbot use

877  $p(\xi) \propto 1/\xi$ , but instead we choose  $p(\xi) \propto 1/\xi \exp\{-\xi/2\}$ . By assuming positive support for  $\xi$

878 and expanding out the marginal prior distribution we get,

$$\begin{aligned} P(\omega) &= \int_0^\infty \left(\frac{2\pi}{\xi}\right)^{-d/2} \exp\left\{-\frac{\xi}{2}\Omega(\omega)\right\} \frac{1}{\xi} \exp\left\{-\frac{\xi}{2}\right\} d\xi \\ &= (2\pi)^{-d/2} \int_0^\infty \xi^{d/2-1} \exp\left\{-\frac{\xi}{2}(\Omega(\omega)+1)\right\} d\xi \end{aligned} \quad (21)$$

879 We use the gamma integral,  $\int_0^\infty x^{\nu-1} \exp\{-\mu x\} = \Gamma(\nu)/\mu^\nu$  (Cawley and Talbot 2007), to ob-

880 tain,

$$P(\omega) = (2\pi)^{-d/2} \frac{\Gamma(d/2)}{(1/2)^{d/2} (\Omega(\omega)+1)^{d/2}} \quad (22)$$

881 Finally, by applying the logarithm and discarding constants, we obtain the final expression for the

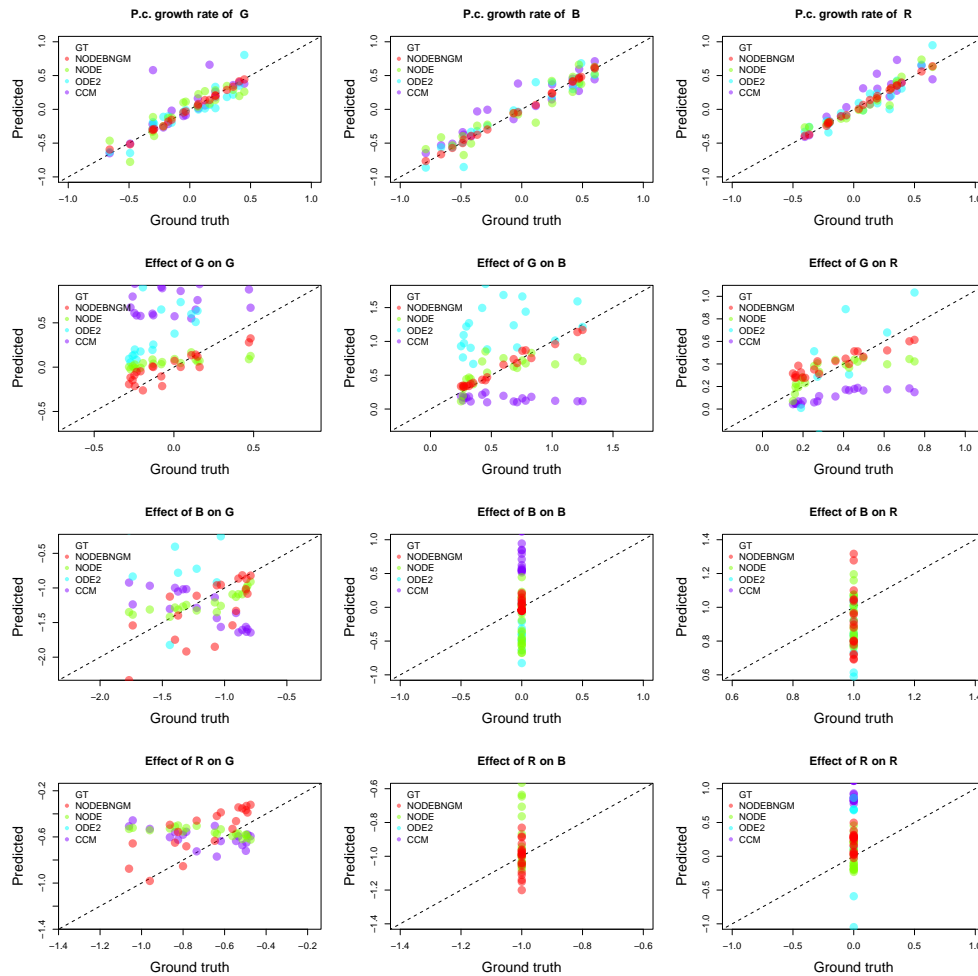
882 marginal prior distribution,

$$\begin{aligned}\log P(\omega) &\propto -\frac{d}{2} \log (\Omega(\omega) + 1) \\ &\propto -\frac{M}{2} \log \left( 1 + \sum_j^M \omega_j^2 \right)\end{aligned}\tag{23}$$

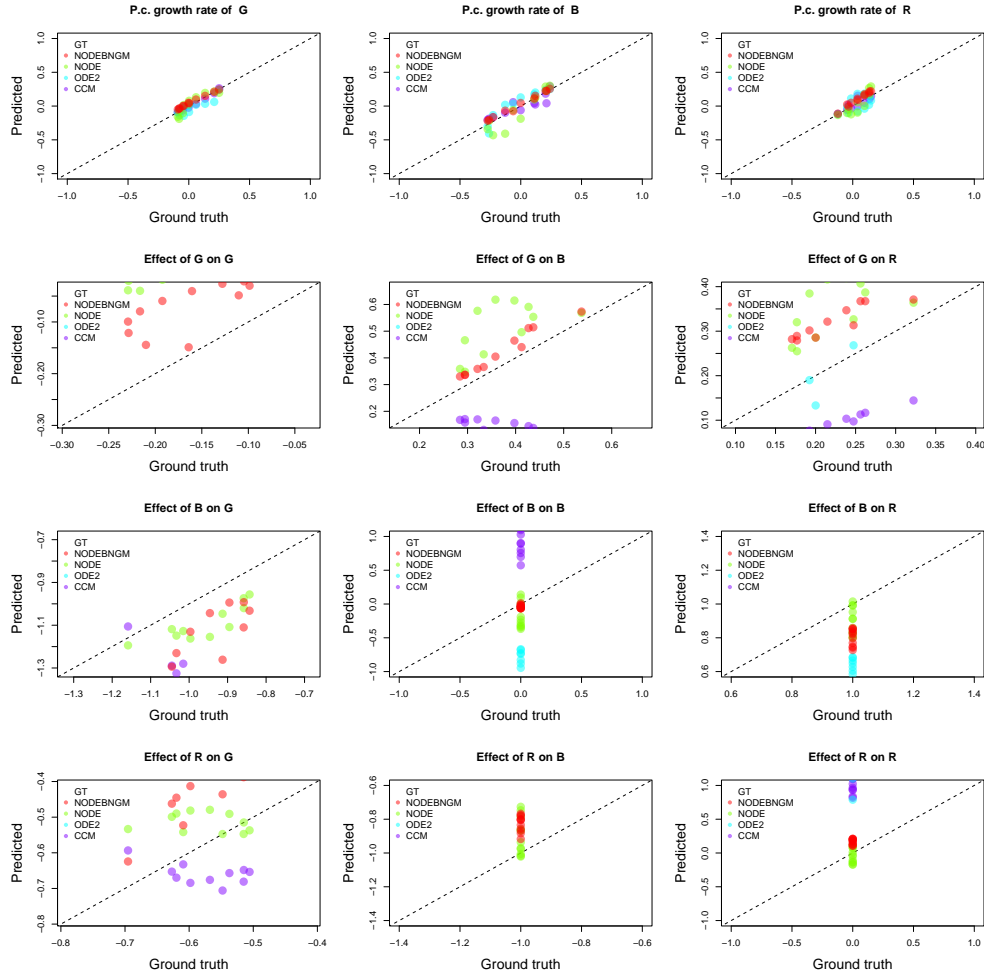
883 The same procedure can be applied to derive the marginal likelihood, by noting  $d := N$ ,  $\xi := 1/\sigma^2$ ,

884  $Z_\Omega := (2\pi/\xi)^{d/2}$ ,  $\Omega(\omega) := \sum_i e_i^2$ .

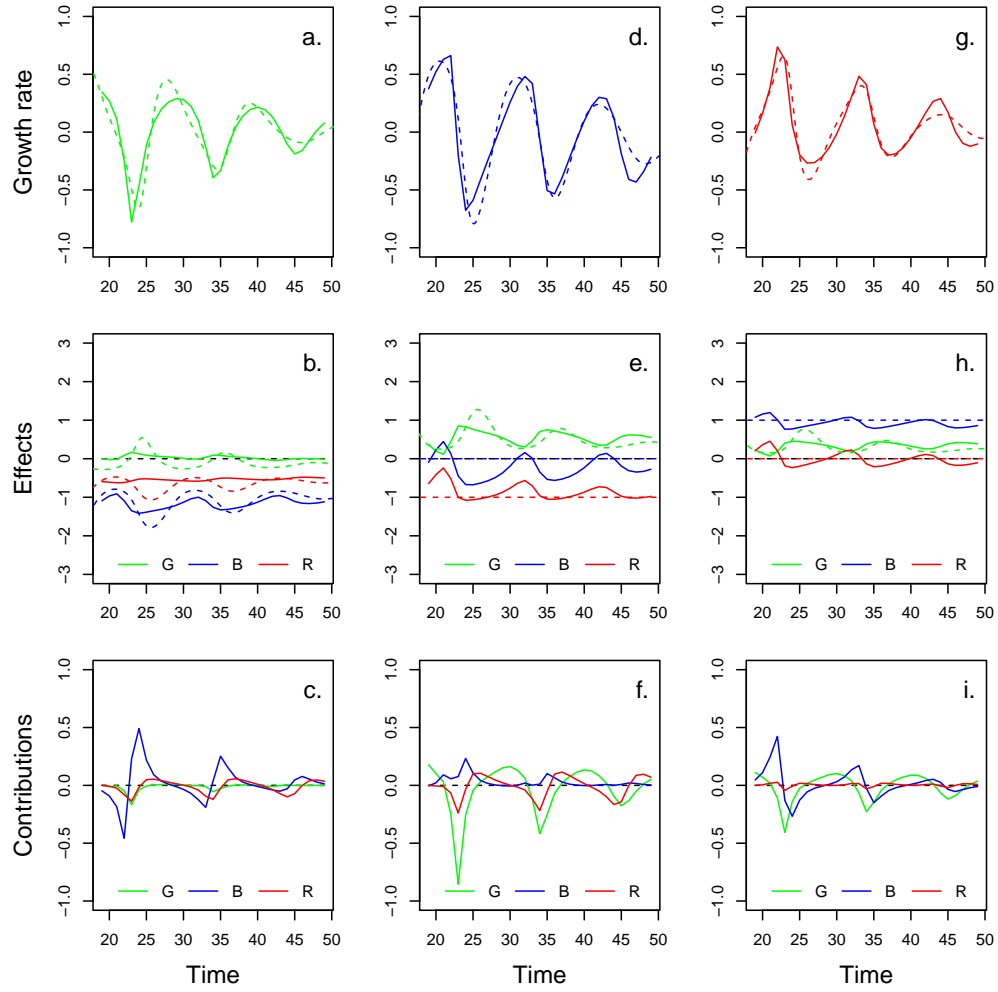
## B Complementary results benchmark analysis



**Figure S1: Train set accuracy of predicted per-capita growth rates and effects estimated by NODEBNGM, standard NODEs, ODE2, and CCM.** The NODEBNGM method (nonparametric) involves fitting a NODE system by Bayesian neural gradient matching (BNGM). The NODE method (nonparametric) involves fitting a NODE system with an ODE solver. The ODE2 method (parametric) involves fitting an ODE system with quadratic functions of species densities with an ODE solver. The CCM method (nonparametric) involves computing locally linear approximations of the state space. For each method, we trained 30 models on the two first thirds of the artificial time series where ground truth is known (Fig. 2). For each plot, the x-axis corresponds to the ground truth, known from the equations that generated the artificial time series, and the y-axis corresponds to the prediction of the best model. Effects are computed as the sensitivity (i.e. derivative) of the per-capita growth rate with respect to each species density  $G$ ,  $B$ , and  $R$ , namely the prey, intermediate and top predator.

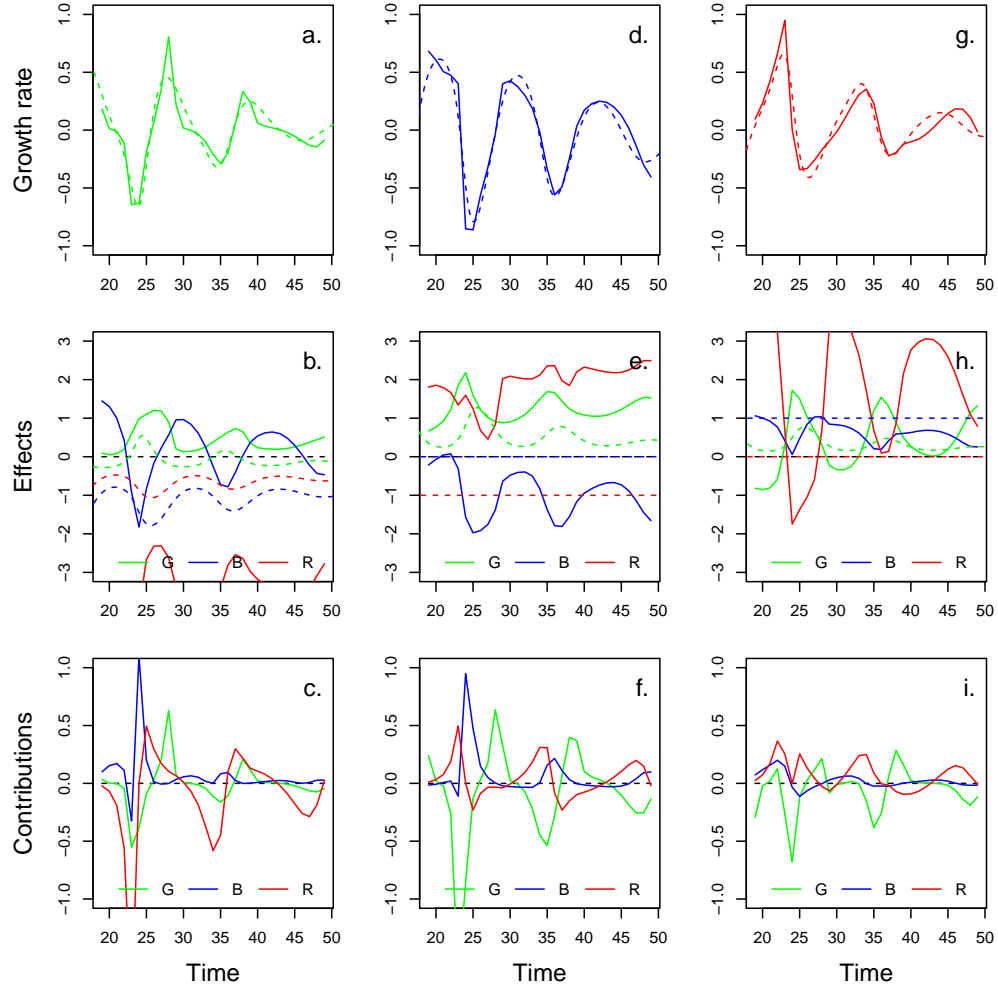


**Figure S2: Test set accuracy of predicted per-capita growth rates and effects estimated by NODEBNGM, standard NODEs, ODE2, and CCM.** The NODEBNGM method (nonparametric) involves fitting a NODE system by Bayesian neural gradient matching (BNGM). The NODE method (nonparametric) involves fitting a NODE system with an ODE solver. The ODE2 method (parametric) involves fitting an ODE system with quadratic functions of species densities with an ODE solver. The CCM method (nonparametric) involves computing locally linear approximations of the state space. For each method, we trained 30 models on the two first thirds of the artificial time series where ground truth is known (Fig. 2). For each plot, the x-axis corresponds to the ground truth, known from the equations that generated the artificial time series, and the y-axis corresponds to the prediction of the best model. Effects are computed as the sensitivity (i.e. derivative) of the per-capita growth rate with respect to each species density G, B, and R, namely the prey, intermediate and top predator.

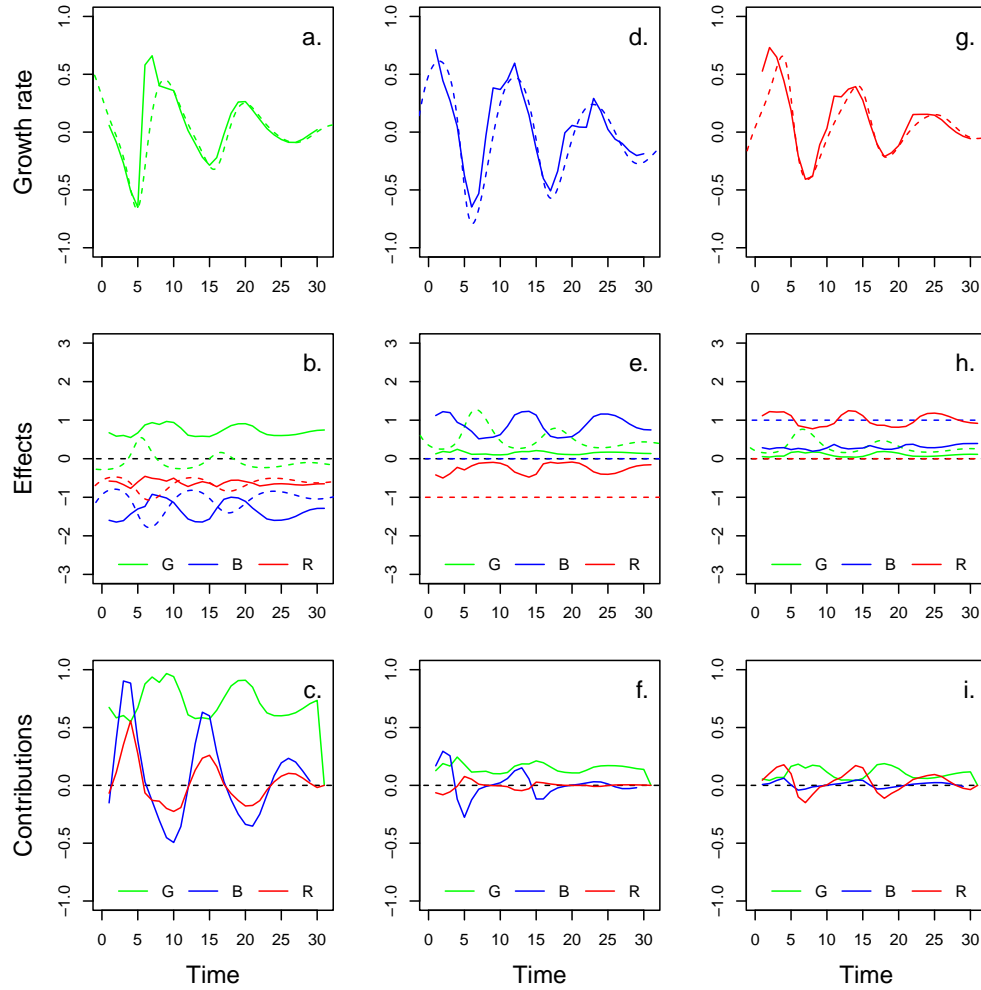


**Figure S3: Effects and contributions inferred by standard NODEs.** The standard NODE model nonparametrically approximates the per-capita growth rate of the 3 species with an ANN featuring a single layer, 3 input nodes, 10 hidden nodes, 3 outputs. 30 models are fitted to the two first third of the time series using BFGS and a Runge-Kutta ODE solver. The graphs present the predictions obtained for the model that maximises posterior density of the network parameters given the time series. We estimate the direction of ecological interactions (effects) by computing the derivative of the per-capita growth rate approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions) by multiplying the interpolated dynamics of each population with its effects.



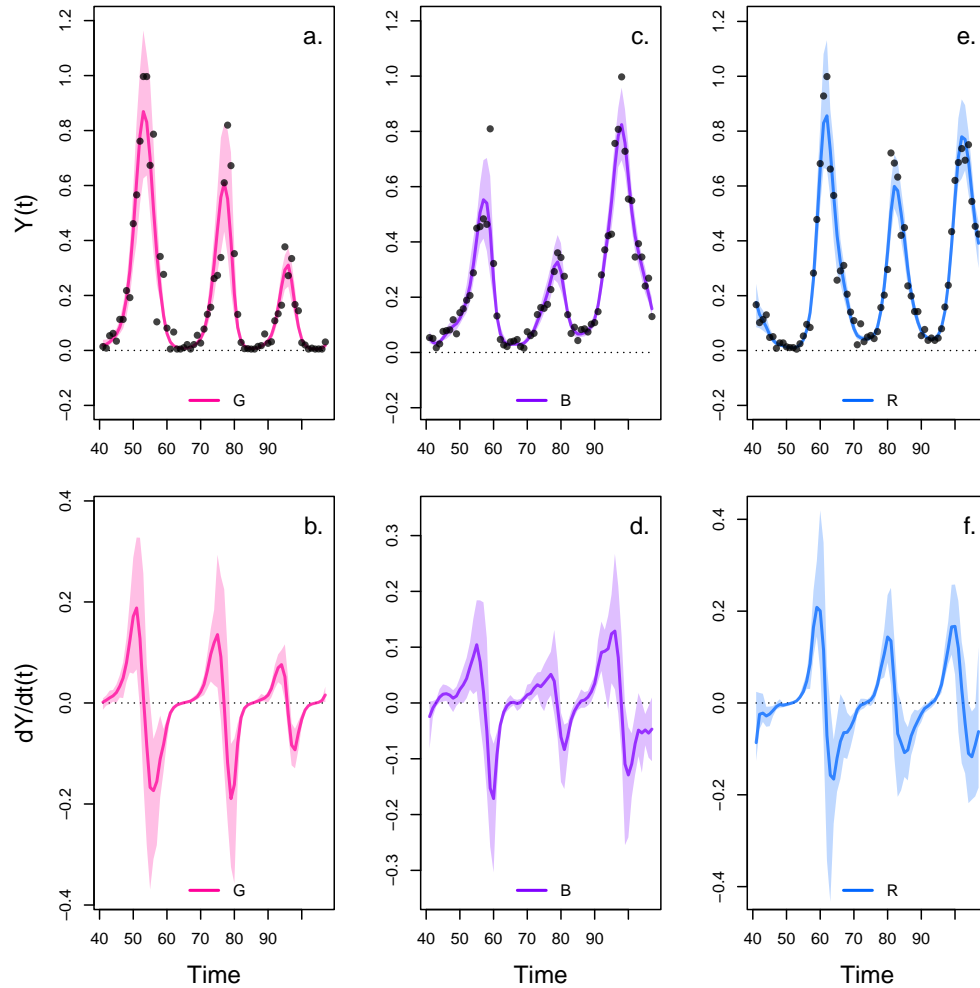


**Figure S4: Effects and contributions inferred by parametric ODE.** The ODE2 model parametrically approximates the per-capita growth rate of the 3 species with second order polynomial functions. 30 models are fitted to the two first third of the time series using BFGS and a Runge-Kutta ODE solver. The graphs present the predictions obtained for the model that maximises posterior density of the network parameters given the time series. We estimate the direction of ecological interactions (effects) by computing the derivative of the per-capita growth rate approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions) by multiplying the interpolated dynamics of each population with its effects.

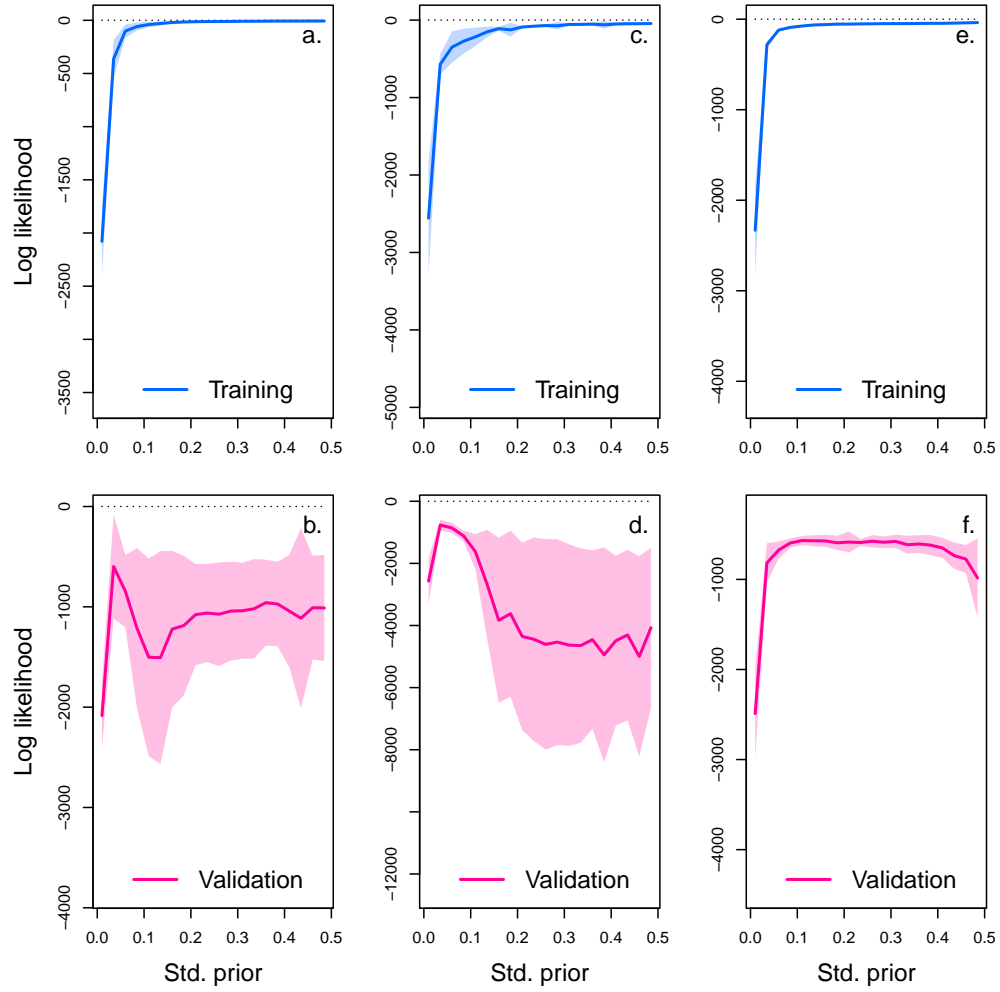


**Figure S5: Effects and contributions inferred by CCM.** The CCM method nonparametrically approximates the state space from which it derives the sensitivity of population dynamics to a change in the density of the each species. We use the rEDM implementation and derived our code from the three species example provided in the package tutorial (v1.13.1, Sugihara et al. 2012). We calculated the dynamics and per capita growth rate using finite differences, as the standard library does not readily provide estimates. The effects correspond to the s-map coefficients. Finally, we compute the strength of ecological interactions (contributions) by multiplying the interpolated dynamics of each population with its effects.

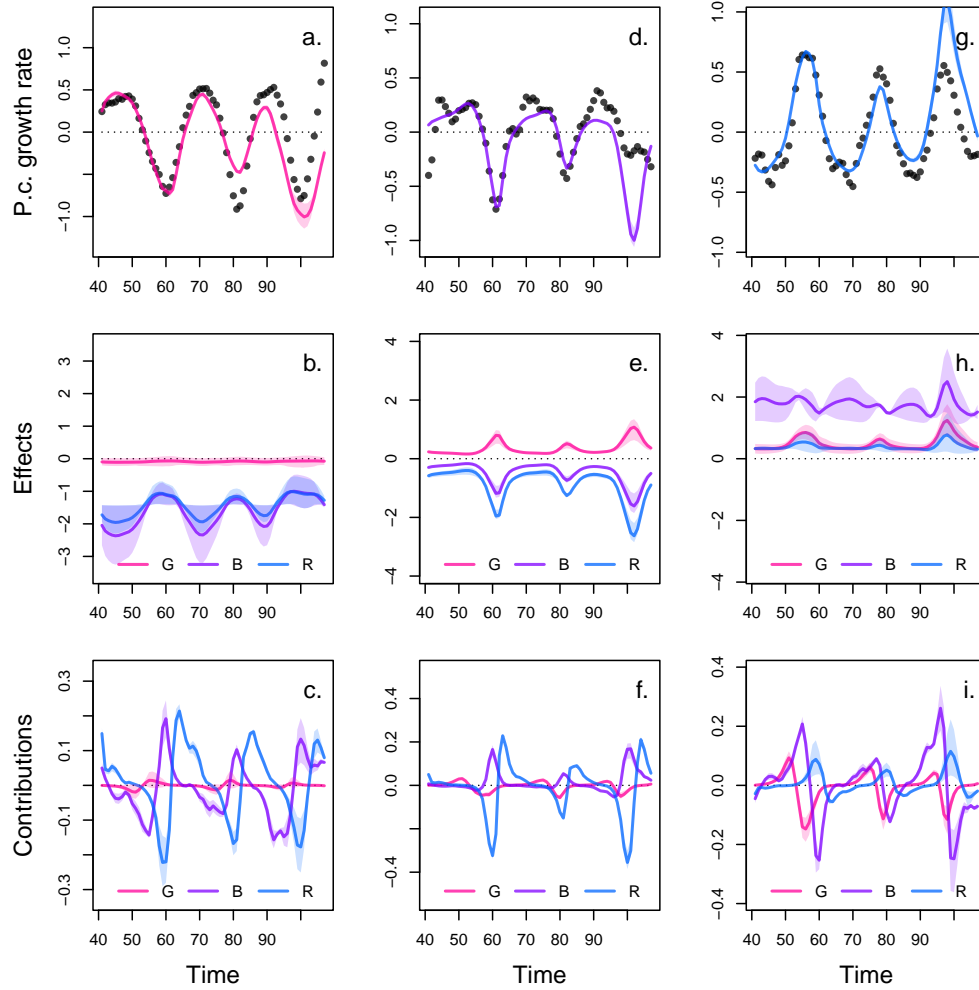
## C Complementary results case study 2 replicate A



**Figure S6: 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.

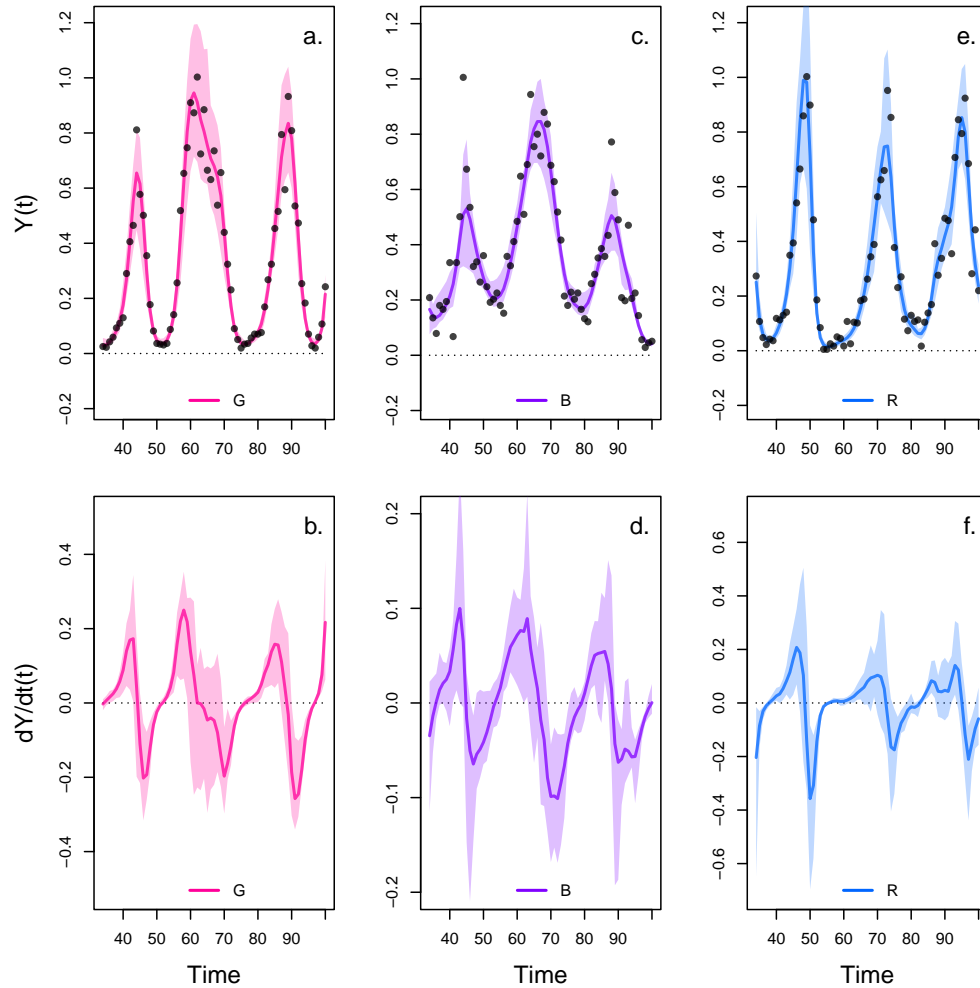


**Figure S7: Cross-validation plot of the NODE analysis of replicate A.** The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values (towards 0.5) correspond to a highly nonlinear model. Time series are split in three thirds to create a train, validation, and 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 validation 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 validation set. The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log posterior distribution (Eq. 8) (Pearce et al. 2018).

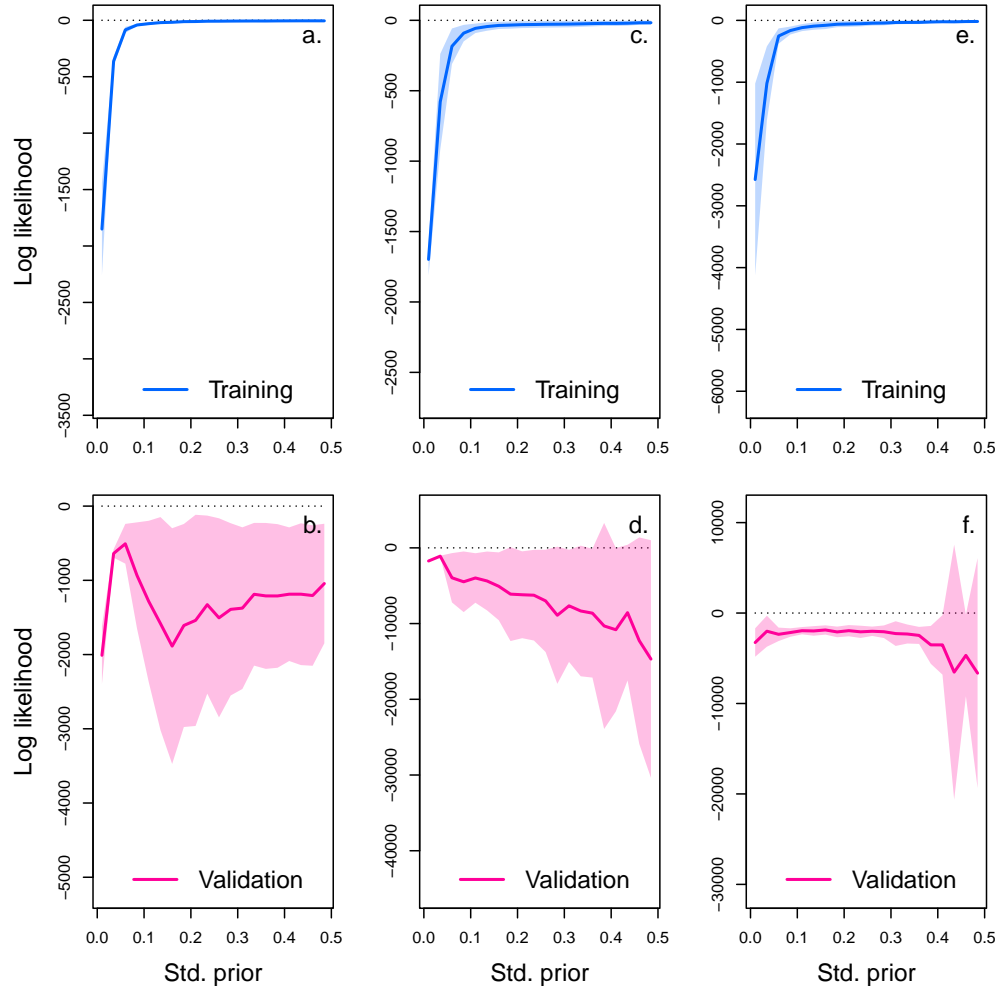


**Figure S8: Drivers of dynamics of algae, flagellate, and rotifer in replicate A.** This figure displays the NODE nonparametric approximations of the per-capita growth rate of algae (a., b., c.), flagellate (d., e., f.), and rotifer (g., h., i.). We obtain the NODE approximations (a., d., g., solid 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).

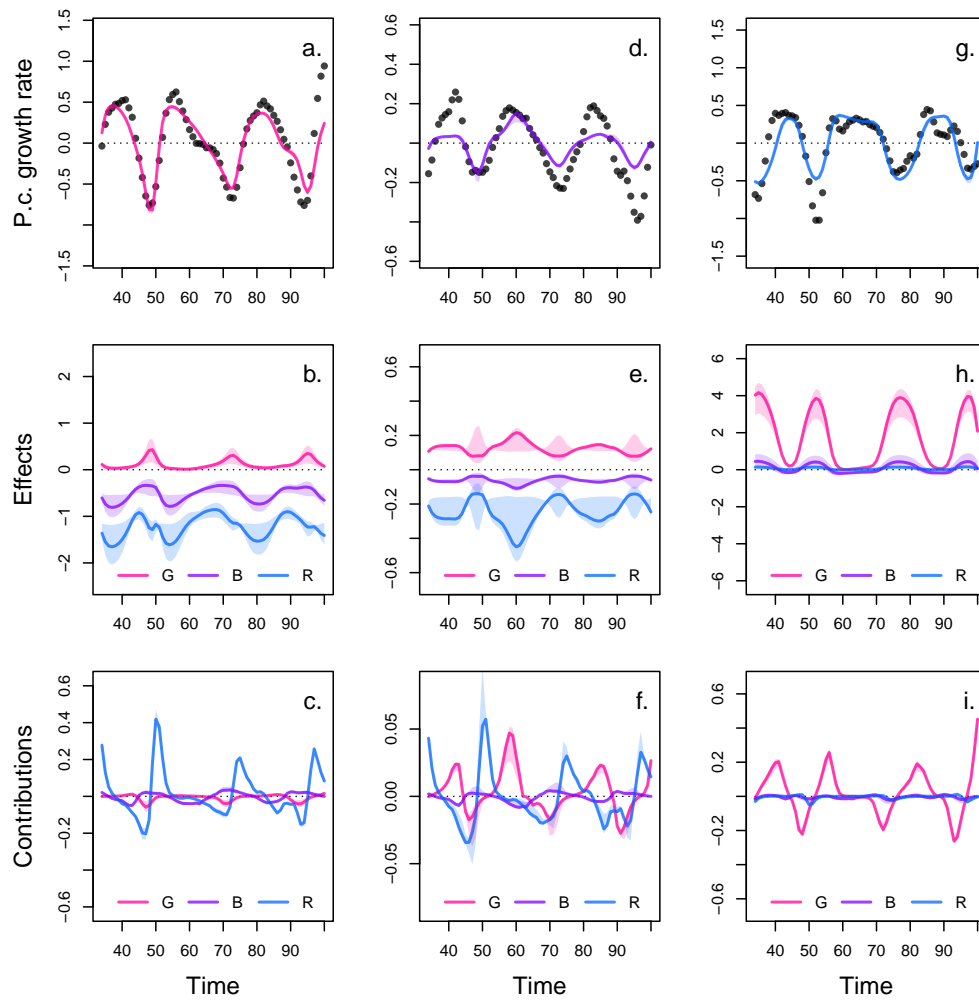
## D Complementary results case study 2 replicate B



**Figure S9: 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.



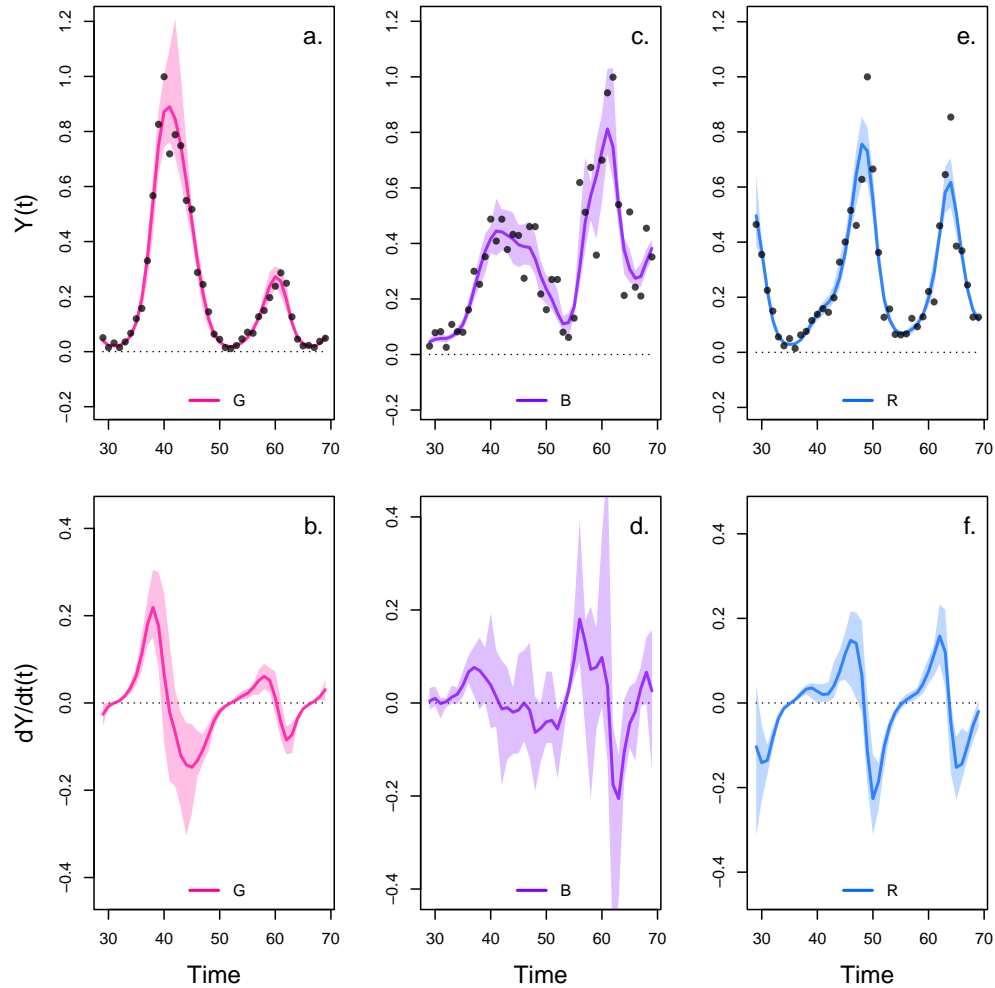
**Figure S10: Cross-validation plot of the NODE analysis of replicate B.** The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values (towards 0.5) correspond to a highly nonlinear model. Time series are split in three thirds to create a train, validation, and 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 validation 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 validation set. The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log posterior distribution (Eq. 8) (Pearce et al. 2018).



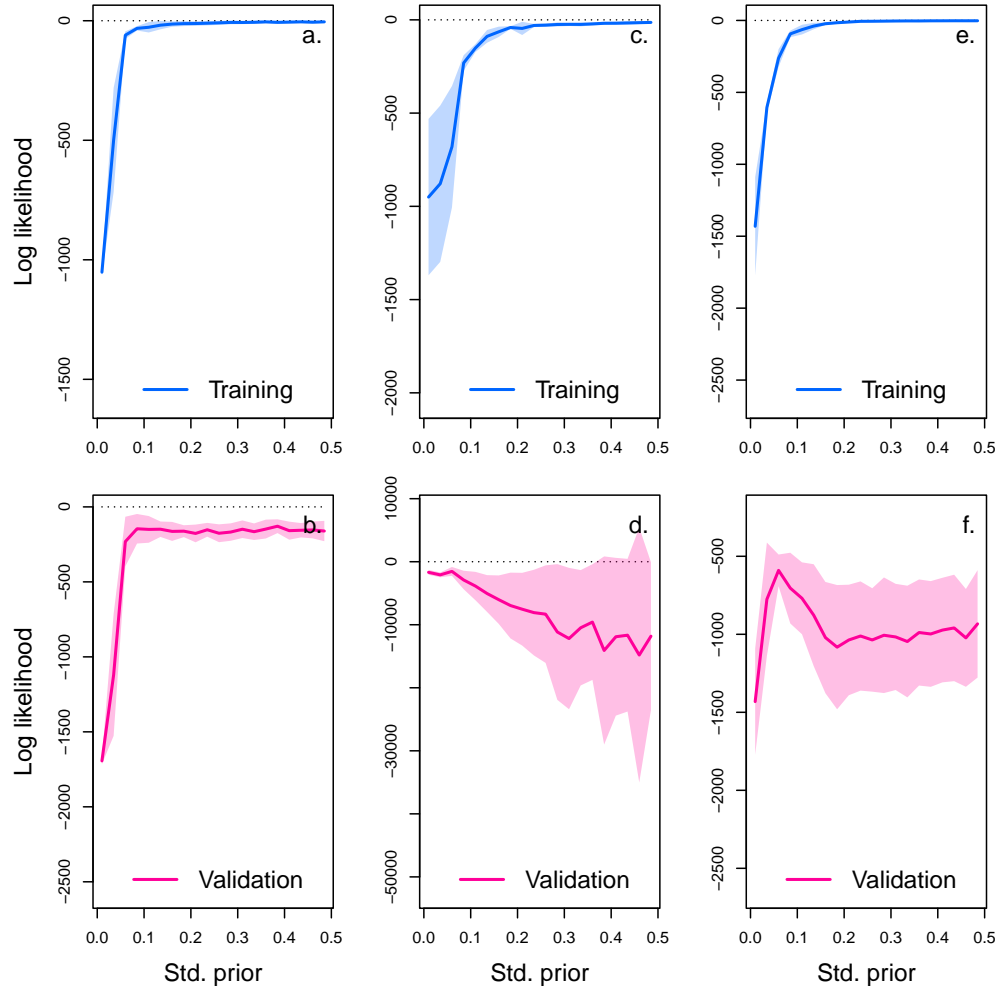
**Figure S11: Drivers of dynamics of algae, flagellate, and rotifer in replicate B.** This figure displays the NODE nonparametric approximations of the per-capita growth rate of algae (a., b., c.), flagellate (d., e., f.), and rotifer (g., h., i.). We obtain the NODE approximations (a., d., g., solid 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).



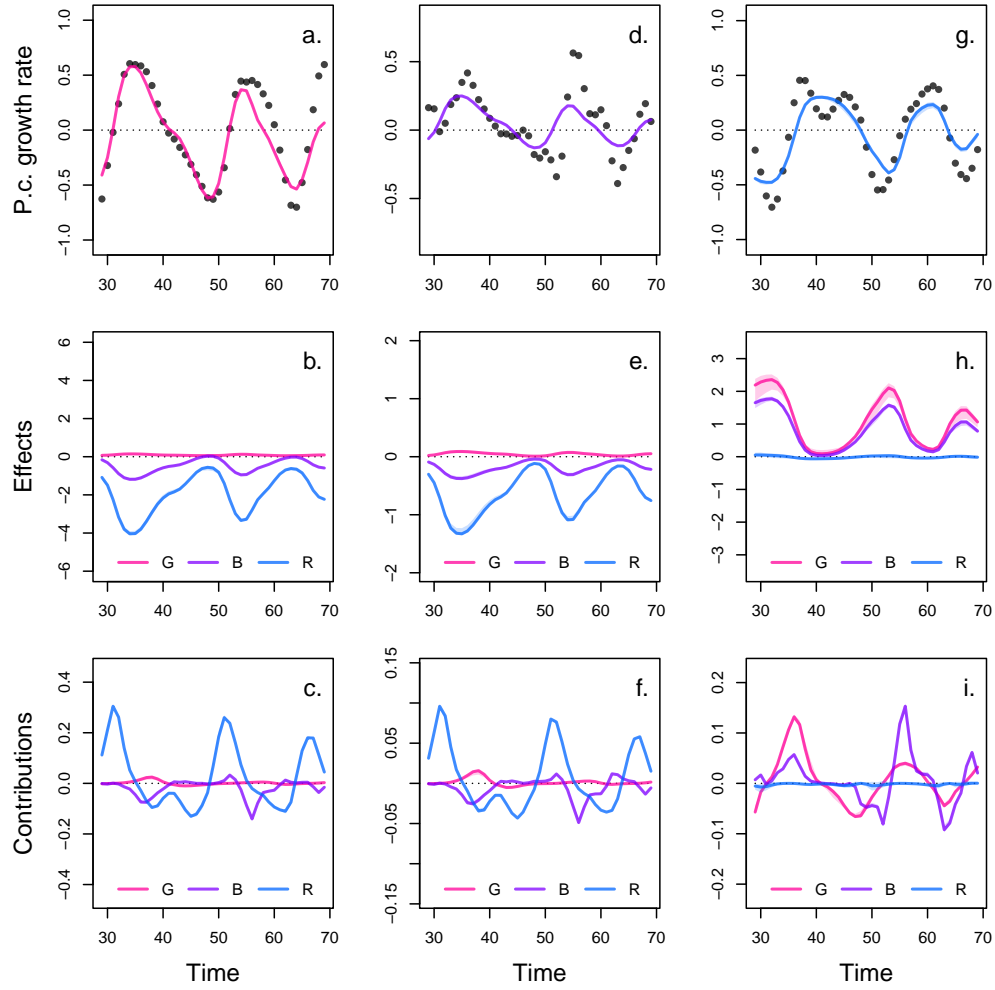
## E Complementary results case study 2 replicate C



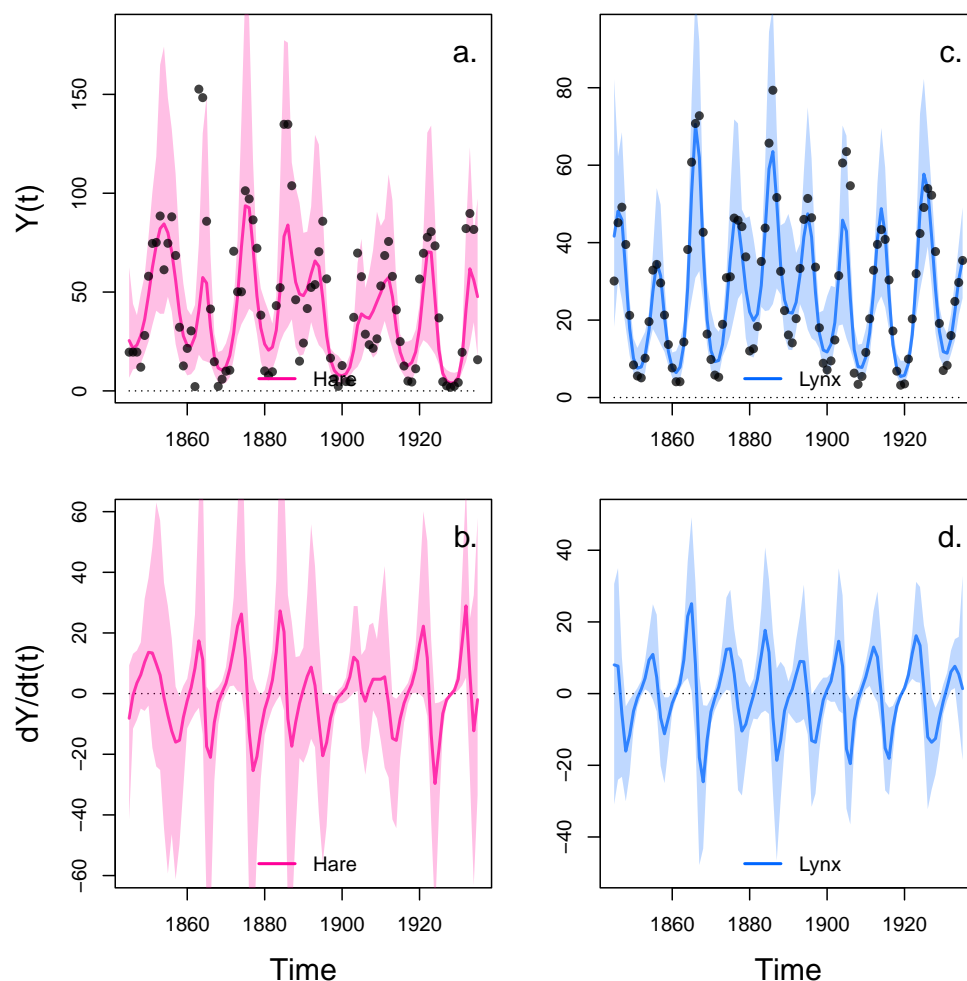
**Figure S12: Interpolation of state and dynamics of algae, flagellate, and rotifer density in replicate B.** Graph a., c., and e. display the neural interpolation of the population density of algae (G), flagellate (B), and rotifer (R), respectively (obtained with Eq. 7). Graph b., d., and f. show the 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.



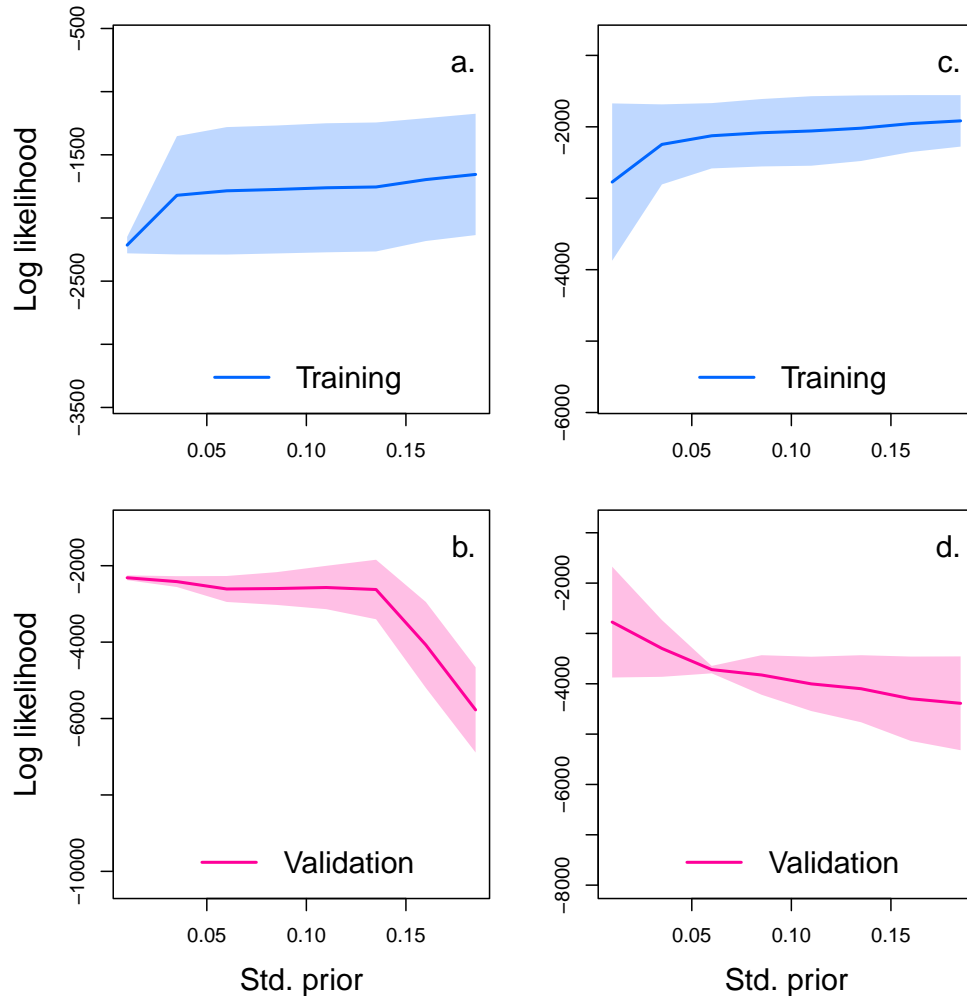
**Figure S13: Cross-validation plot of the NODE analysis of replicate C.** The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values (towards 0.5) correspond to a highly nonlinear model. Time series are split in three thirds to create a train, validation, and 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 validation 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 validation set. The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log posterior distribution (Eq. 8) (Pearce et al. 2018).



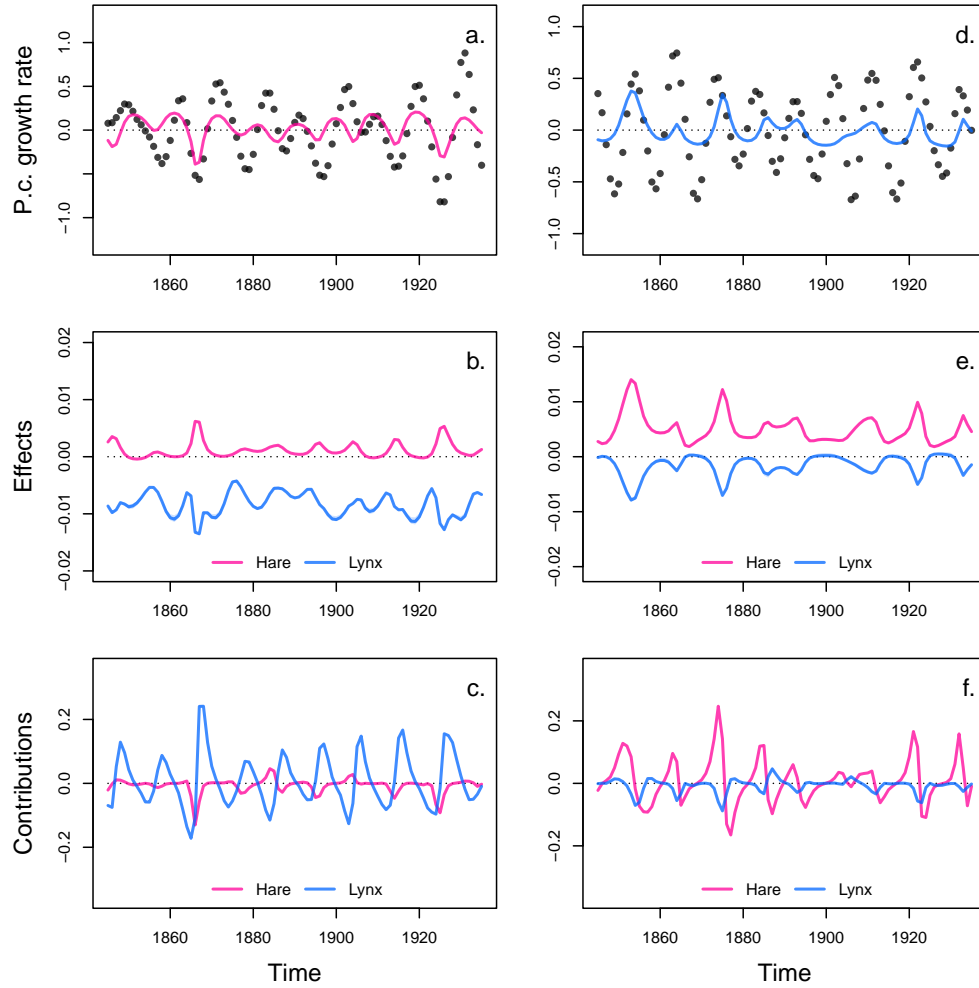
**Figure S14: Drivers of dynamics of algae, flagellate, and rotifer in replicate C.** This figure displays the NODE nonparametric approximations of the per-capita growth rate of algae (a., b., c.), flagellate (d., e., f.), and rotifer (g., h., i.). We obtain the NODE approximations (a., d., g., solid 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).



**Figure S15: 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.

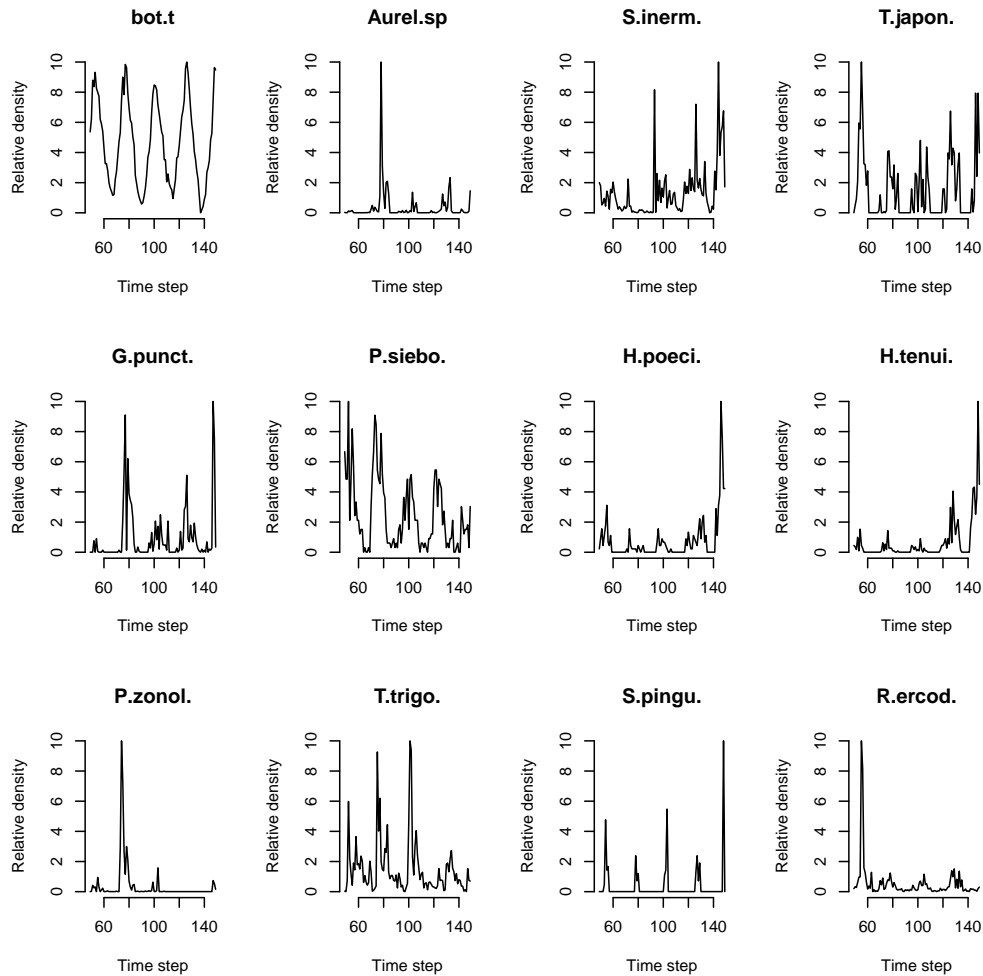


**Figure S16: Cross-validation plot of the NODE analysis of the hare-lynx system.** The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values (towards 0.2) correspond to a highly nonlinear model. Time series are split in three thirds to create a train, validation, and test set. The model is fitted to the train set for increasing value of standard deviation (from 0.0 to 0.2 by 0.025 increments), and evaluated on the validation set. Graph a. and c. show the log likelihood of the NODE system fitted by BNGM to the train set of hare and lynx, respectively. Graph b. and d. show the log likelihood of the fitted NODE, evaluated on the corresponding validation set. The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log posterior distribution (Eq. 8) (Pearce et al. 2018).

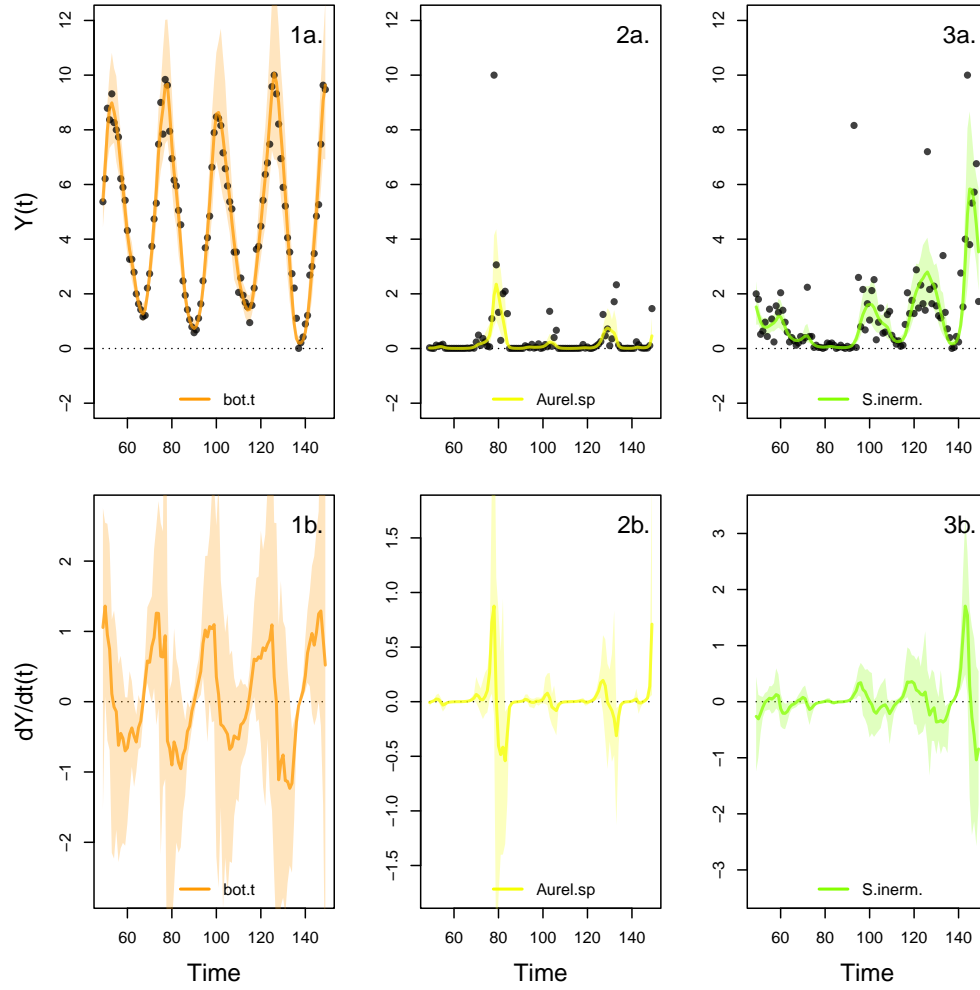


**Figure S17: Drivers of dynamics of hare and lynx in the Odum and Barrett pelt count time series.** This figure displays the NODE nonparametric approximations of the per-capita growth rate of hare (a., b., c.), and lynx (d., e., f.). We obtain the NODE approximations (a., d., solid line) by fitting 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.

## G Complementary results case study 4

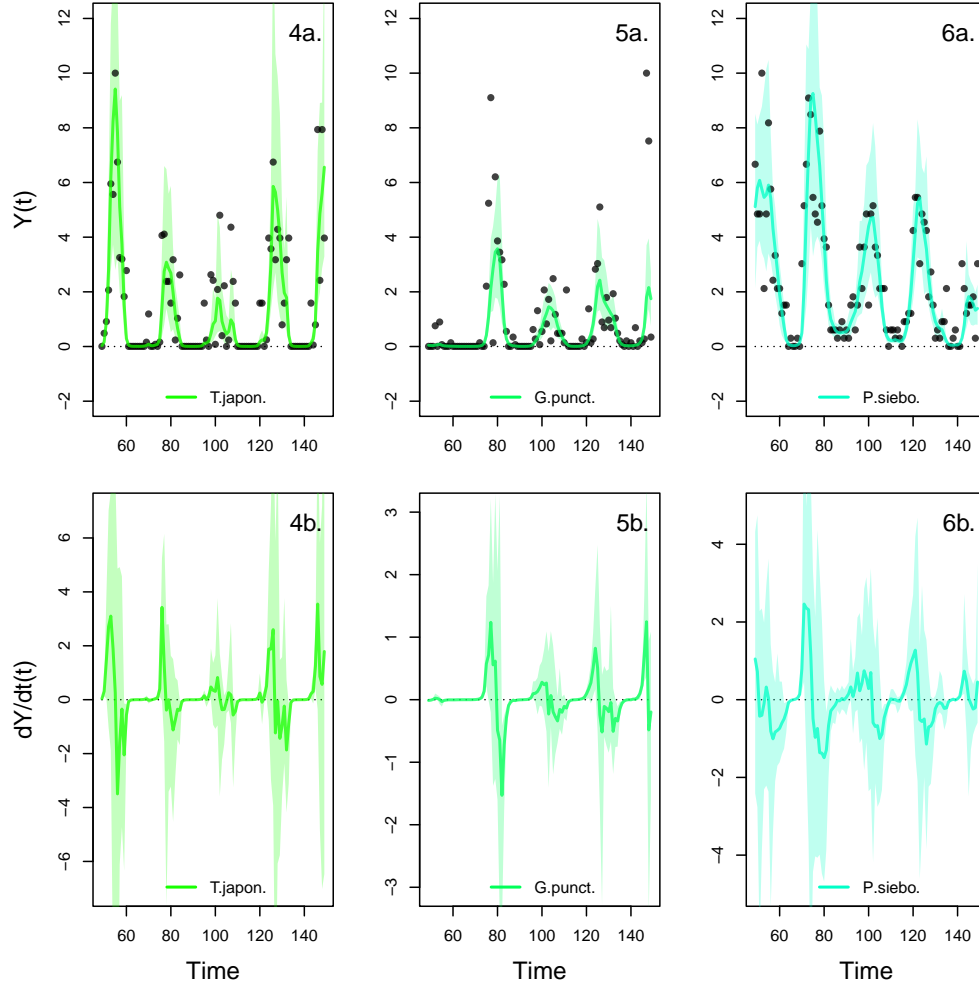


**Figure S18: Time series of the aquatic community in Ushio et al. 2018.** The time series was collected for 12-years on a fortnightly basis, for 15 dominant species in the Maizuru bay in Japan. We focus on the 11 species and the 100 time steps with the least sparse abundance records. Bot.t corresponds to water temperature near the bottom. The main species are *Aurelia sp.*, *Sebastes inermis*, *Trachurus japonicus*, *Girella punctata*, *Pseudolabrus sieboldi*, *Halichoeres poecilopterus*, *Halichoeres tenuispinnis*, *Pterogobius zonoleucus*, *Tridentiger trigonocephalus*, *Sphyrna pinguis*, and *Rudarius ercodes*. Variables are normalised with respect to minimum and maximum values.

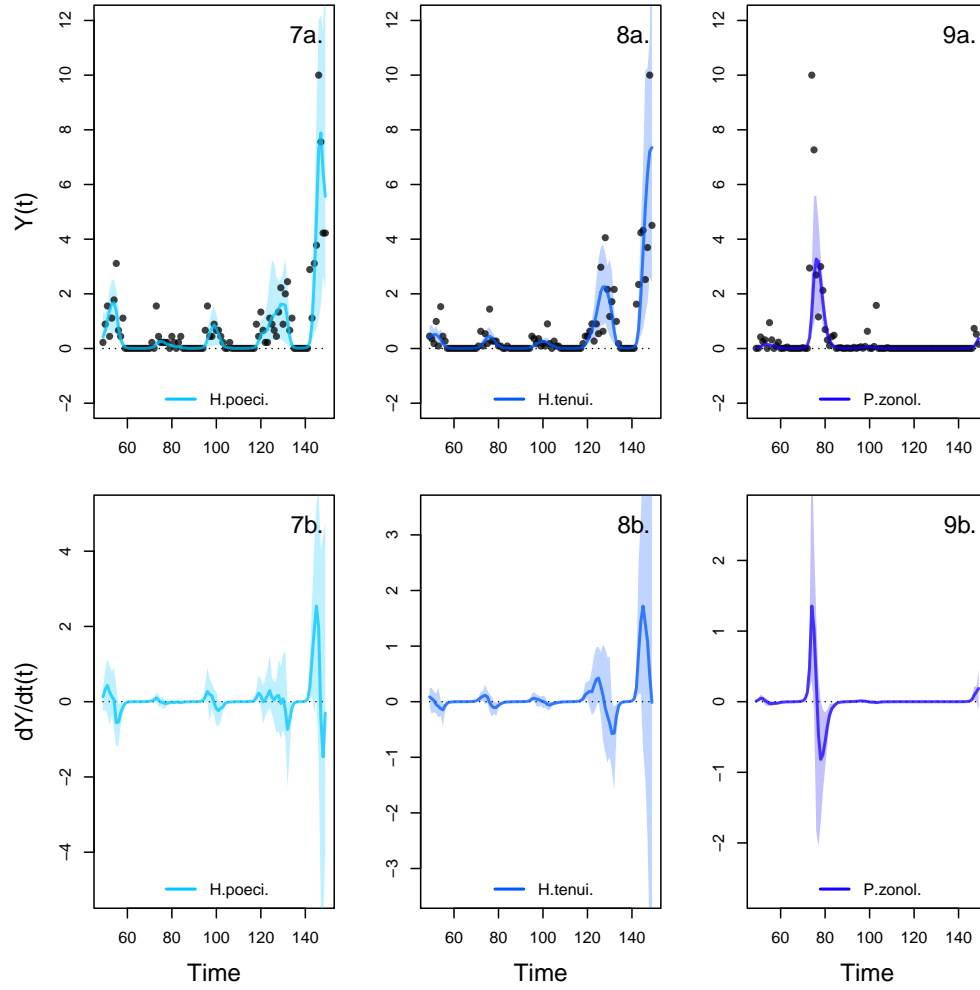


**Figure S19: Interpolation of state and dynamics of temperature and species abundance in the Maizuru bay community.** Graphs a. display the neural interpolations of the population density (obtained with Eq. 7), apart from graph 1a. which corresponds to sea bottom temperature. Graphs b. show the corresponding 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. 8).

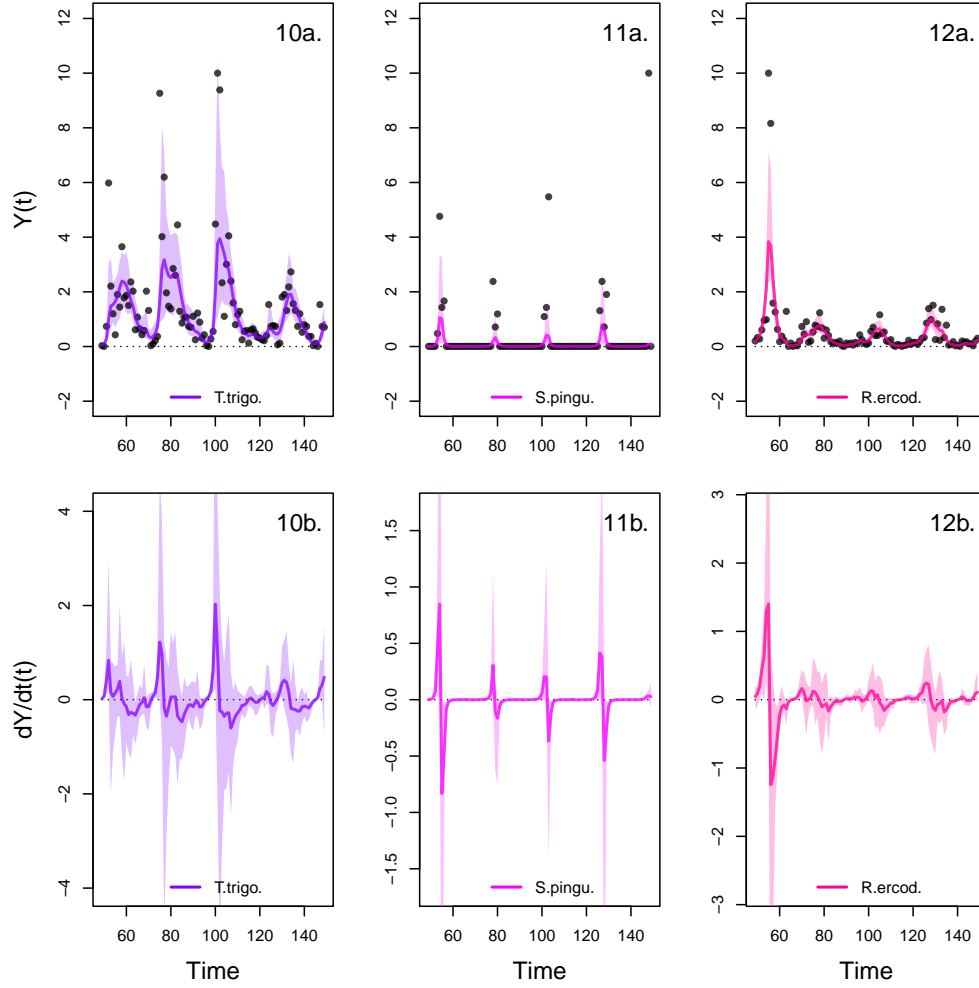




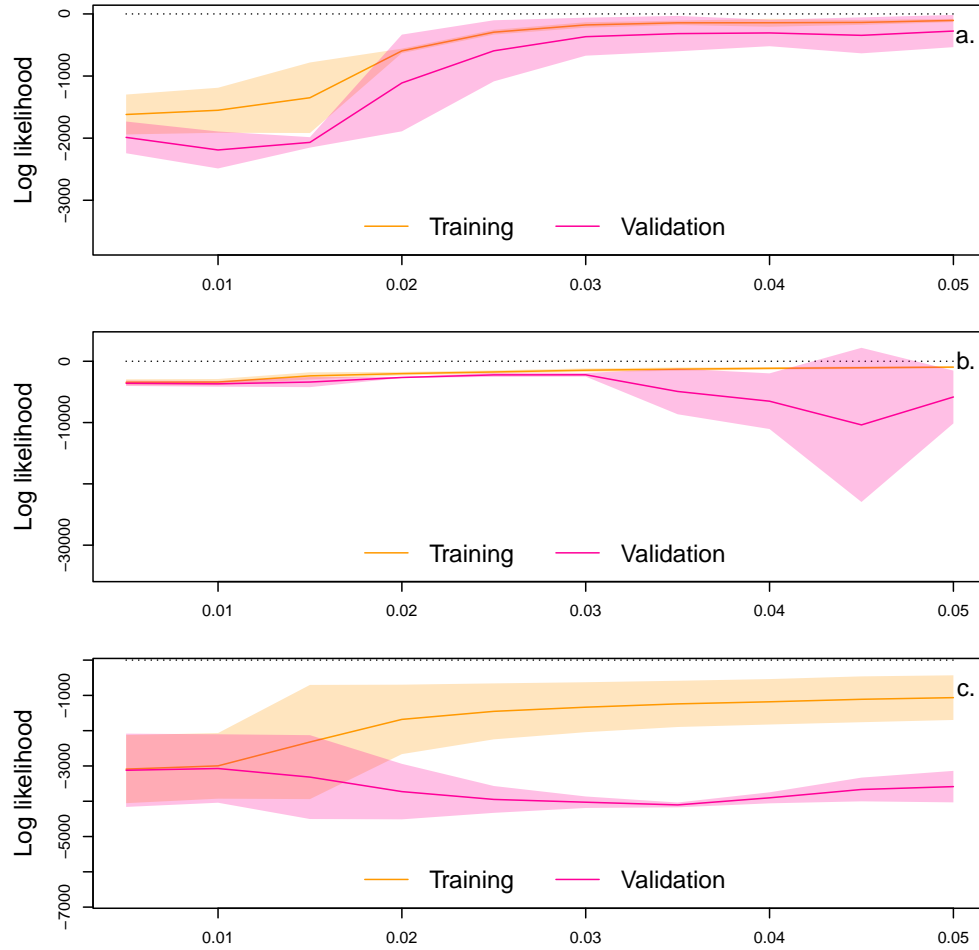
**Figure S20: Interpolation of state and dynamics of species abundance in the Maizuru bay community.** Graphs a. display the neural interpolations of the population density (obtained with Eq. 7). Graphs b. show the corresponding interpolated dynamics, obtained by differentiating the interpolation of 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. 8).



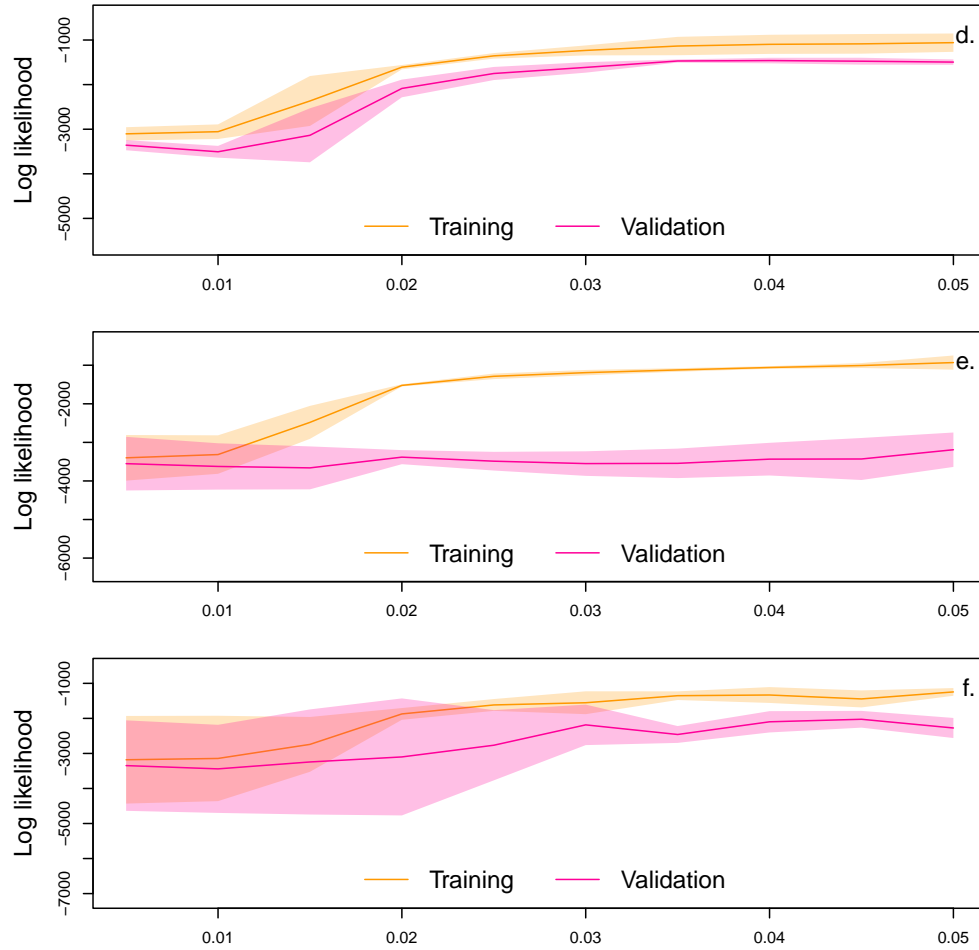
**Figure S21: Interpolation of state and dynamics of species abundance in the Maizuru bay community.** Graphs a. display the neural interpolations of the population density (obtained with Eq. 7). Graphs b. show the corresponding interpolated dynamics, obtained by differentiating the interpolation of 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. 8).



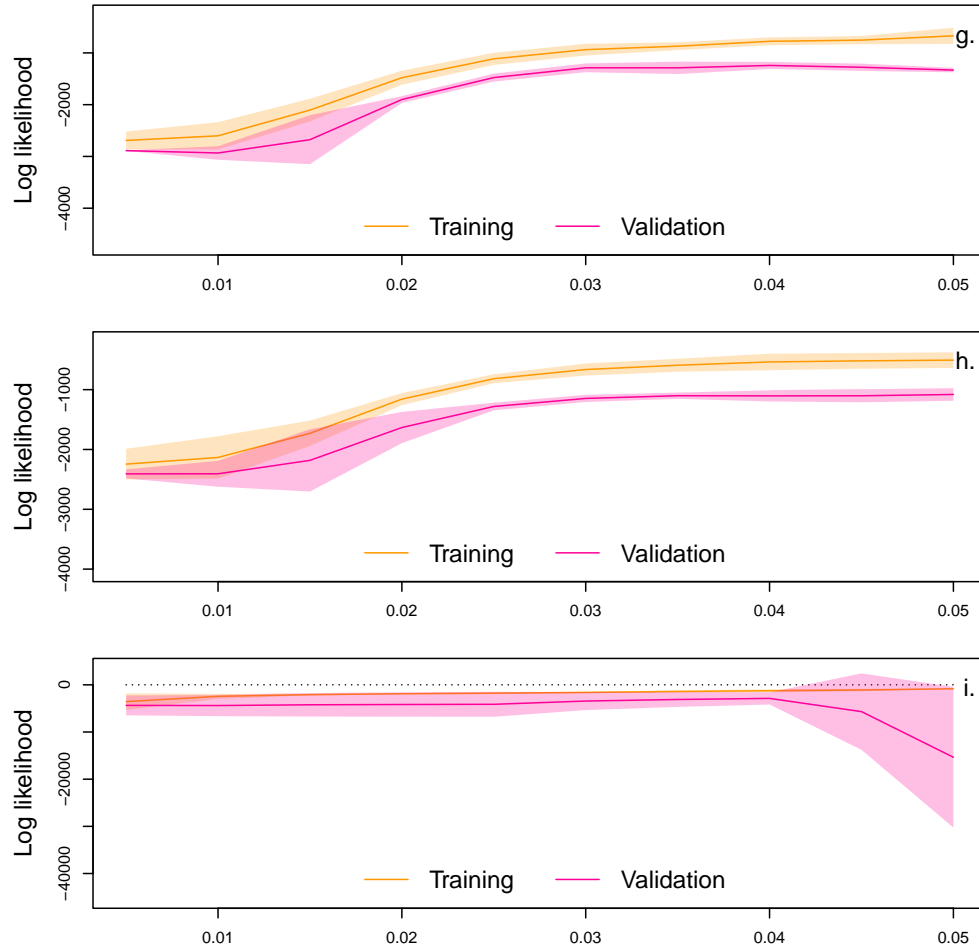
**Figure S22: Interpolation of state and dynamics of species abundance in the Maizuru bay community.** Graphs a. display the neural interpolations of the population density (obtained with Eq. 7). Graphs b. show the corresponding interpolated dynamics, obtained by differentiating the interpolation of 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. 8).



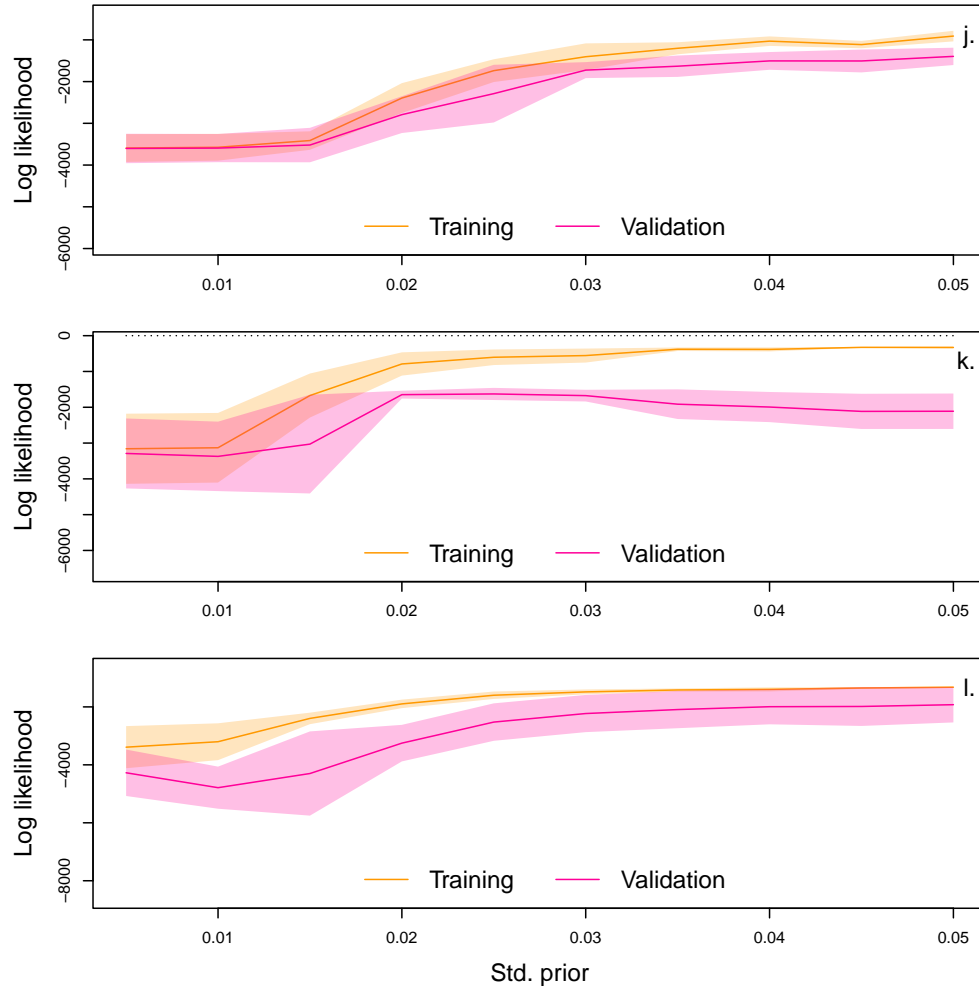
**Figure S23: Cross-validation plot of the NODE analysis of the Maizuru bay community.** The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values correspond to a highly nonlinear model. Time series are split in three thirds to create a train, validation, and test set. The model is fitted to the train set (i.e. first third) for increasing value of standard deviation (from 0.005 to 0.05 by 0.005 increments), and evaluated on the validation set. The operation is repeated by swapping the training and validation set. The graphs show the log likelihood of the NODE system fitted by BNGM to the train set (in orange), and evaluated on the corresponding validation set (in red). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log posterior distribution (Eq. 8) (Pearce et al. 2018).



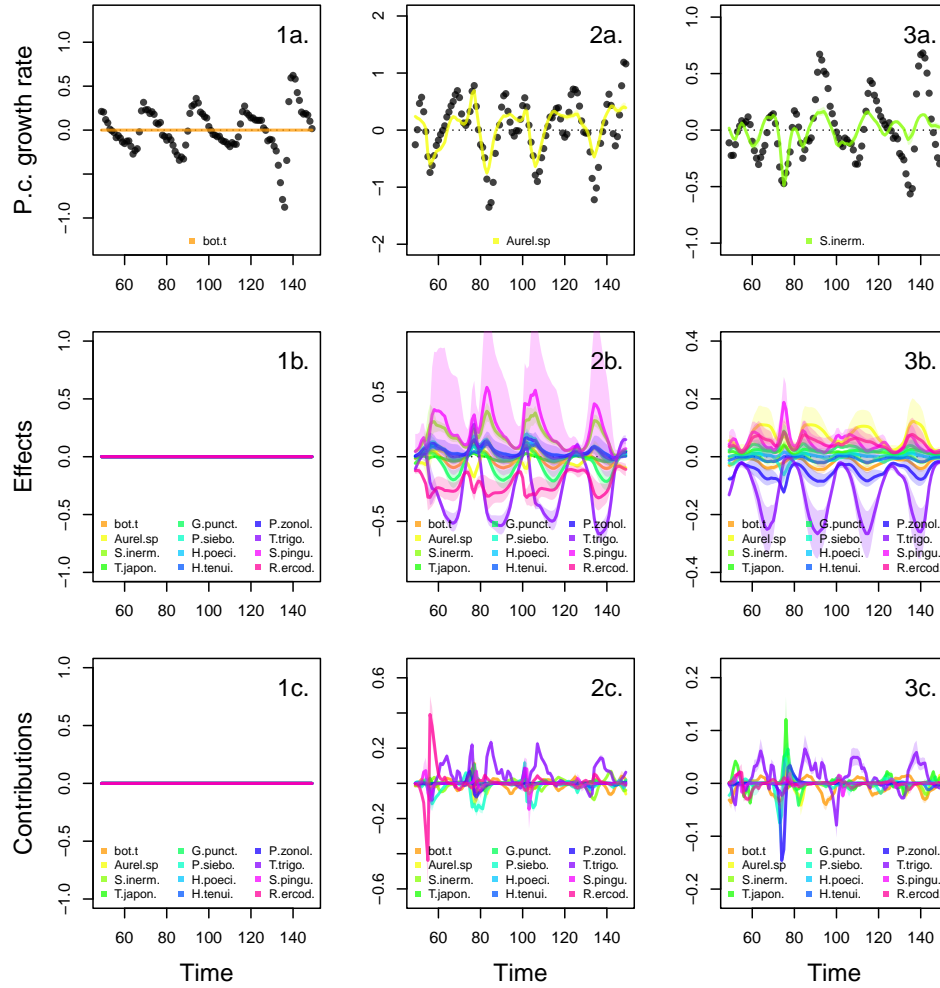
**Figure S24: Cross-validation plot of the NODE analysis of the Maizuru bay community.** The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values correspond to a highly nonlinear model. Time series are split in three thirds to create a train, validation, and test set. The model is fitted to the train set (i.e. first third) for increasing value of standard deviation (from 0.005 to 0.05 by 0.005 increments), and evaluated on the validation set. The operation is repeated by swapping the training and validation set. The graphs show the log likelihood of the NODE system fitted by BNGM to the train set (in orange), and evaluated on the corresponding validation set (in red). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log posterior distribution (Eq. 8) (Pearce et al. 2018).



**Figure S25: Cross-validation plot of the NODE analysis of the Maizuru bay community.** 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 correspond to a highly nonlinear model. Time series are split in three thirds to create a train, validation, and test set. The model is fitted to the train set (i.e. first third) for increasing value of standard deviation (from 0.005 to 0.05 by 0.005 increments), and evaluated on the validation set. The operation is repeated by swapping the training and validation set. The graphs show the log likelihood of the NODE system fitted by BNGM to the train set (in orange), and evaluated on the corresponding validation set (in red). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log posterior distribution (Eq. 8) (Pearce et al. 2018).

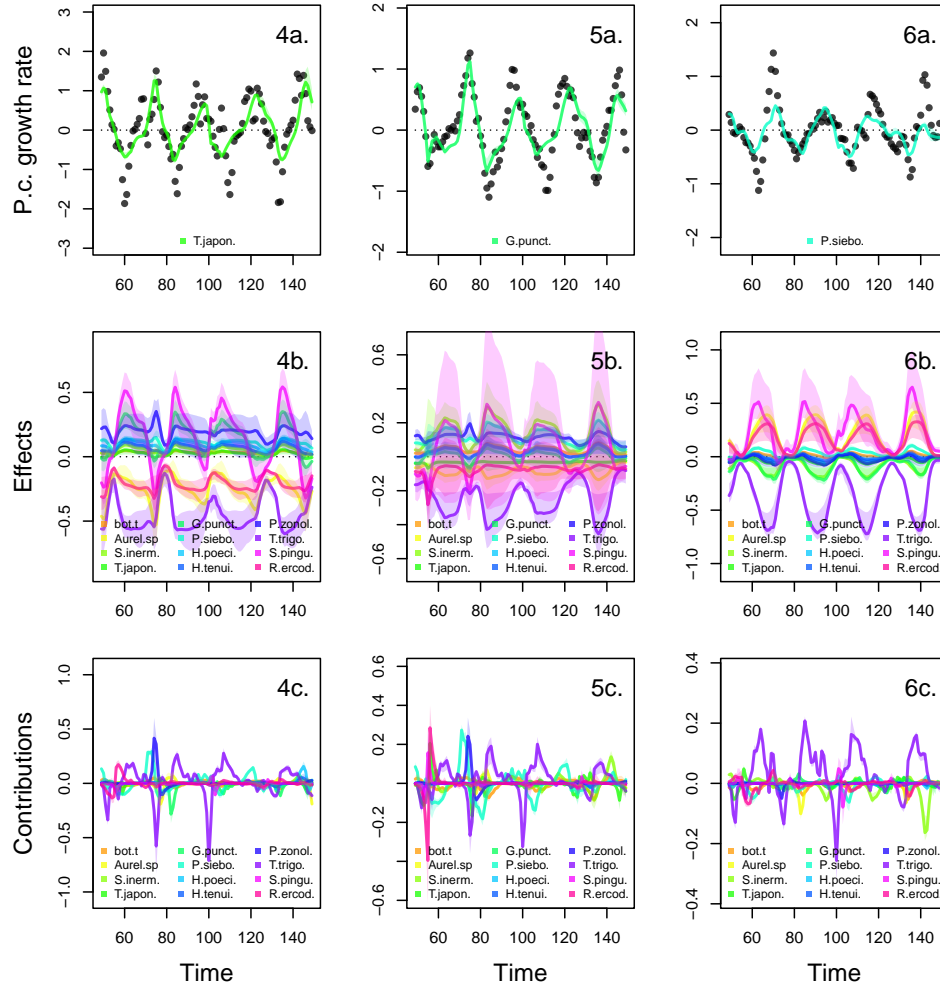


**Figure S26: Cross-validation plot of the NODE analysis of the Maizuru bay community.** The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values correspond to a highly nonlinear model. Time series are split in three thirds to create a train, validation, and test set. The model is fitted to the train set (i.e. first third) for increasing value of standard deviation (from 0.005 to 0.05 by 0.005 increments), and evaluated on the validation set. The operation is repeated by swapping the training and validation set. The graphs show the log likelihood of the NODE system fitted by BNGM to the train set (in orange), and evaluated on the corresponding validation set (in red). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log posterior distribution (Eq. 8) (Pearce et al. 2018).

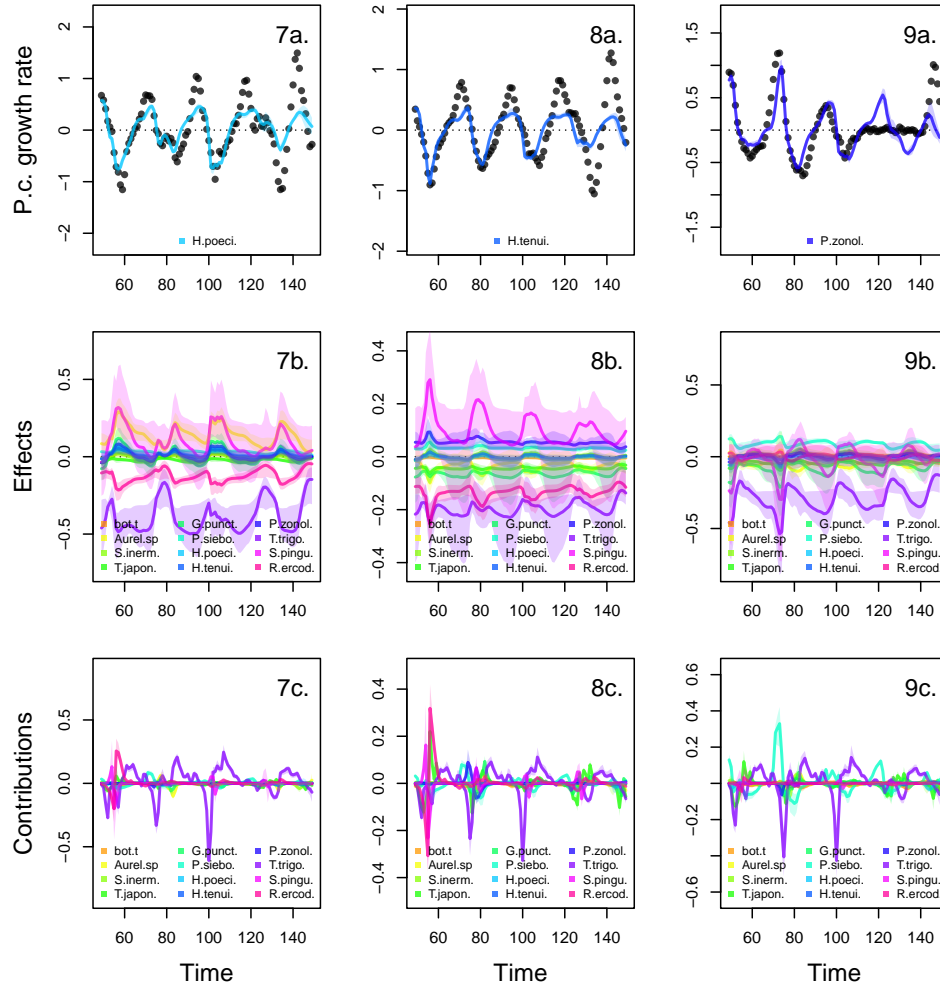


**Figure S27: Drivers of dynamics of species abundance in the Maizuru bay community.** This figure displays the NODE nonparametric approximations of the per-capita growth rates (2-3a.). We obtain the NODE approximations (2-3a., solid line) by fitting the interpolated per-capita growth rates (black dots) with ANNs that take population densities as input. We then estimate the direction of ecological interactions (effects, 2-3b.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, 2-3c.) 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.

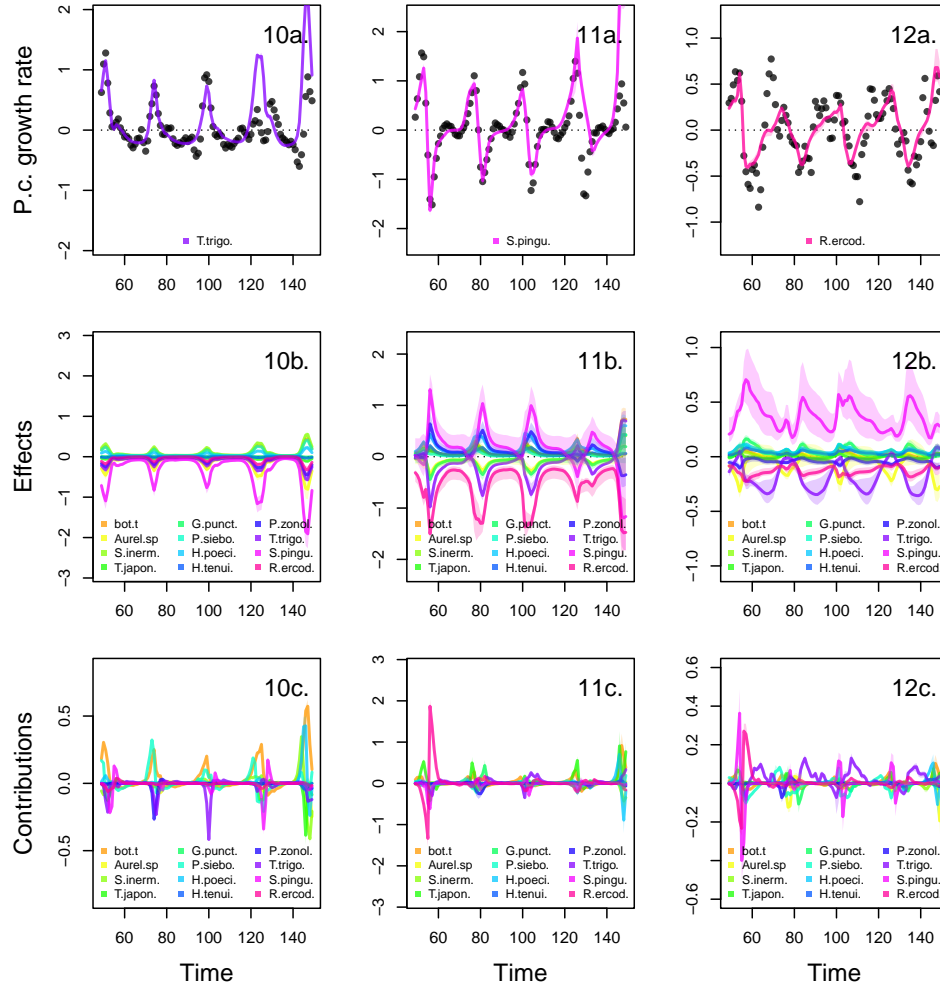




**Figure S28: Drivers of dynamics of species abundance in the Maizuru bay community.** This figure displays the NODE nonparametric approximations of the per-capita growth rates (4-6a.). We obtain the NODE approximations (4-6a., solid line) by fitting the interpolated per-capita growth rates (black dots) with ANNs that take population densities as input. We then estimate the direction of ecological interactions (effects, 2-6b.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, 2-6c.) 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.



**Figure S29: Drivers of dynamics of species abundance in the Maizuru bay community.** This figure displays the NODE nonparametric approximations of the per-capita growth rates (7-9a.). We obtain the NODE approximations (7-9a., 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, 7-9b.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, 7-9c.) 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.



**Figure S30: Drivers of dynamics of species abundance in the Maizuru bay community.** This figure displays the NODE nonparametric approximations of the per-capita growth rates (10-12a.). We obtain the NODE approximations (10-12a., 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, 10-12b.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, 10-12c.) 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.