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

Willem Bonnaffé[1,2], Ben Sheldon[1], & Tim Coulson[2]

1. Edward Grey Institute of Field Ornithology, Department of Zoology, Oxford University, Zoology Research and Administration Building, 11a Mansfield Road, Oxford OX1 3SZ

2. Ecological and Evolutionary Dynamics Lab, Department of Zoology, Oxford University, Zoology Research and Administration Building, 11a Mansfield Road, Oxford OX1 3SZ

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

Running title: Repeatable interactions and dynamics

Keywords: Artificial Neural Networks; Ecological Dynamics; Ecological interactions; Geber Method; Neural Ordinary Differential Equations; Ordinary Differential Equations; Prey-predator dynamics; Time series analysis; Rotifers; Microcosm;

Specifications: 140 words in abstract; 7071 words in text; 40 references; 5 figures; 1 table

Contact: Willem Bonnaffé, 61 St Giles, Pusey House, St Cross College, Oxford, OX1 3LZ, UK (w.bonnaffe@gmail.com)

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

Abstract

Generalisation of dynamical processes across natural systems is difficult because they are complex and hard to observe. The hope is that generalisation may be achieved by adequately modelling the complexity of systems, and observe them in sufficient detail. We investigate this by looking at the consistency of ecological interactions across three replicates of a three-species prey-predator system, well-observed in an artificial environment, using neural ordinary differential equations. We find that dominant interactions are consistent across the replicates, while weaker interactions are not, leading to different dynamical patterns across replicated systems. Our study hence suggests that generalisation of dynamical processes across systems may not be possible, even in simpler systems in ideal monitoring conditions. This is a problem because if we are not able to make generalisations in a simple artificial system, how can we make generalisation in the real world?

1 Introduction

The repeatability of ecological and evolutionary dynamics varies widely across systems and species. Sticklebacks from different lakes in Canada have independently evolved to a similar river morph phenotype (Thompson, Taylor, and Mcphail 1997). In guppies, four replicated populations located in different streams in Trinidad evolved the same low-predation phenotype (Reznick, Bryga, and Endler 1990). Multiple studies in experimental microcosms, particularly in rotifer populations, have shown that population dynamics were broadly repeatable (Yoshida et al. 2003; Yoshida et al. 2007; Becks et al. 2010; Becks et al. 2012; Hiltunen et al. 2013). Overall, this demonstrated that ecological and evolutionary dynamics may be repeatable across different instances of the same system, at least qualitatively. This was a fascinating finding given the complexity of the mechanisms involved and the subtle variations in environmental conditions across the different populations.

These systems hinted at the possibility for identifying global, generalisable, dynamical models. In practice, however, generalising dynamics and dynamical processes (i.e. functional representations describing which and how state variables affect each other and determine system dynamics) across natural systems has proven difficult (Lawton 1999). First, even if the dynamical patterns, and their outcomes, may appear to be conserved across similar systems, they may be underpinned by different processes. For instance, the evolution of the sticklebacks to highly similar river-adapted phenotypes has been shown to be underpinned by radically different genetic alterations (Raeymaekers et al. 2017). Second, it is often unclear whether quantitative differences across replicated systems

arise from pure stochasticity (Dallas et al. 2021), observation error (De Meester et al. 2019), or deterministic changes in the dynamical processes. Finally, the complexity of biological processes themselves (Adamson and Morozov 2013), differences in genetic and environmental contexts, may prevent the identification of a suitable dynamical model. For example, Becks and colleagues found that differences in the initial amount of genetic variation in otherwise identical rotifer populations led to subtle changes to the dynamics (Becks et al. 2010). Different access to seed supplies can modify the strength of the interaction between a plant and its herbivore, leading to either stable or oscillatory dynamics (Bonsall, Van Der Meijden, and Crawley 2003). Differences in temperature can alter the ecological interaction structure of entire ecosystems (Shurin et al. 2012; Bonnaffé et al. 2021). Because of this, vital rates are often found to be inconsistent in time (Gross, Ives, and Nordheim 2005; Adamson and Morozov 2013), and space (e.g. Gamelon et al. 2019). Overall, a growing body of evidence shows that generalisation of dynamical processes across similar natural systems often fails (Lawton 1999, e.g. Kendall et al. 2005; Demyanov, Wood, and Kedwards 2006; Ezard, Côté, and Pelletier 2009).

So how could repeatable dynamics arise across multiple instances of the same system? We would expect dynamics to be repeatable if the components of the system (e.g. species), as well as interactions between components, are conserved. For this, populations should have similar distributions for the traits that underpin these interactions, and should further share the same environmental conditions, across instances. While this is unreasonable to expect from a natural system, it may be achievable in an artificial setting. In such a setting, it is possible to understand the structure of

the system, to control the environment, and to reduce observation error. So if we fail to identify and generalise dynamical models in natural systems, perhaps we may be able to do so in artificial systems.

In spite of this there are few studies that have attempted to characterise the generalisability of dynamics across replicated systems in a laboratory setting. In such a setting, idiosyncrasies in population dynamics can arise from (1) variations in ecological interactions and individual processes, as a result of evolution (e.g. Yoshida et al. 2003), or stochasticity (Dallas et al. 2021), (2) variations in initial conditions due to the experimental setting (Yoshida et al. 2003; Becks et al. 2010; De Meester et al. 2019), and (3) the complexity of the system which can lead to large changes in system dynamics with small changes in the system state and structure (Adamson and Morozov 2013). Two studies, one in aphids and the other in rotifers, found substantial variation in vital rates across replicated populations, by fitting a stage-structured population ODE model to population dynamics time series data (Bruijning, Jongejans, and Turcotte 2019; Rosenbaum et al. 2019). These studies hint that generalisability of population dynamical processes may not be possible because of intrinsic population structure and evolution, even in virtually identical populations hosted in artifical environments.

We identified three gaps in the literature. First, this kind of evidence remains scarce, due in part to the fact that dynamical modelling approaches guided by empirical data are still not widespread (Pontarp, Brännström, and Petchey 2019). Second, most of these studies relied on parametric frameworks, which impose arbitrary pre-determined forms for the dynamical processes at play, so

that their model may not capture properly the complexity of the dynamics of these populations (Jost and Ellner 2000; Adamson and Morozov 2013; Bonnaffé, Sheldon, and Coulson 2021). Finally, most studies usually analyse dynamics in single-species systems, but not multi-species systems, such as those with intraguild predation, which are more biologically realistic scenarios (Hiltunen et al. 2013). Further studies are consequently required to investigate the consistency of dynamical processes in simple multi-species and well-observed systems, to conclude about the generalisability of population dynamics across systems.

Our aim in this study is to provide an assessment of the repeatability of dynamical processes across different instances of a realistic multi-species system hosted in a well-observed environment. We do this by quantifying the direction, strength, and consistency of interactions in time and across replicates of a three-species microcosm in an experimental setting. We hypothesise that if the system is (1) simple enough, (2) well-observed, (3) in a controlled environment, then dynamical effects/interactions should be broadly consistent in time and across replicates, hence allowing for generalisation of dynamics across systems. We consider three replicates of a three-species system, consisting in a prey (algae), intermediate-predator (flagellate), and top-predator (rotifer). The algae is consumed by the flagellate and rotifer, and the flagellate is consumed by the rotifer. We use three replicated system runs from a study by Hiltunen and colleagues which feature sequential oscillations of the density of the three species (Hiltunen et al. 2013). We analyse the time series with neural ordinary differential equations (Bonnaffé, Sheldon, and Coulson 2021), which allows us to approximate non-parametrically population growth rates, and quantify the direction, strength,

and consistency of inter- and intra-specific effects on the growth of each population. We find that the interaction between the rotifer and algae is consistent throughout time and across replicates, while the interaction between the flagellate and the two other species is not. Our study suggests that dynamical processes may sometimes not be consistent and generalisable across systems, even when they are as close to identical as experimentation permits. We discuss these results and hint at the underlying impact of evolution driving differences in these systems.

In previous work, we developed a simulation-based approach to fit NODE systems to time series data (Bonnaffé, Sheldon, and Coulson 2021). We would first simulate the NODE system over the entire time series. Then we would compute the error between the predictions of the NODE model and the observations. Finally, we would change the weights of the NODEs to minimise this error. There are two caveats with this approach. The first caveat is that the NODE system has to be simulated over the entire range of the data at every step of the optimisation, which is computationally expensive to perform. Second, the numerical integration prevents the computation of gradients of the posterior distribution of the model, which prevents the use of efficient gradient descent approaches.

Ellner and colleagues introduced a technique called *gradient matching* to fit ODEs (Ellner, Seifu, and Smith 2002). The approach involves two steps. First, they interpolate the time series of each state variable with cubic splines to obtain interpolated states and dynamics. Second, they train each ODE to satisfy the interpolated state and dynamics. The interpolation allows them to bypass the simulation of the ODE system, and further makes the error function mathematically tractable,

allowing for the computation of gradients.

2 Material and Methods

2.1 Method overview

We aim to 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.

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

We first consider a general NODE system,

$$\frac{dy_i}{dt} = f_p\left(y, \theta_i\right), \tag{1}$$

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

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

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

We would like to stress that this general form can be changed to represent biological constraints on the state variables. In particular for population dynamics, the state variables are strictly positive

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

2.3 Fitting NODEs by Bayesian neural gradient matching

In this section, we describe how to estimate the parameters θ of the NODE system given a set of time series. Fitting NODEs can be highly computationally intensive, which hinders uncertainty quantification, cross-validation, and model selection (Bonnaffé, Sheldon, and Coulson 2021). We solve this issue by introducing *Bayesian neural gradient matching* (BNGM), a computationally efficient approach to fit NODEs. The approach involves two steps (Fig. 1). First, we interpolate the state variables and their dynamics with neural networks. Second, we train each NODE to satisfy the interpolated state and dynamics. This bypasses the costly numerical integration of the NODE system and provides a fully mathematically tractable expression for the posterior distribution of the parameter vector θ. We coin the term BNGM to emphasise two important refinements of the standard gradient matching algorithm. The first is that we use neural networks as interpolation functions, and the second is that we use Bayesian regularisation to limit overfitting and estimate uncertainty around parameters.

Interpolating the time series

The first step is to interpolate the time series and differentiate it with respect to time in order to approximate the state and dynamics of the variables. We perform the interpolation via non-

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

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

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

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

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

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

Fitting NODEs to the interpolated time series

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

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

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

Bayesian regularisation

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

First, we define a simple Bayesian model to fit the interpolation functions (Eq. 3) to the time series data. We assume normal distributions for the observation error, $\varepsilon_{ij}^{(o)} \sim \mathcal{N}(0, \sigma_i)$, and for the parameters, $\omega_{ij} \sim \mathcal{N}(0, \gamma_{ij})$. Here, we are only interested in interpolating the time series accurately, irrespective of the value of σ_i and γ_{ij}. Therefore, we use the approach developed by Cawley and Talbot to average out the value of the parameters σ_i and γ_{ij} in the full posterior distribution (Cawley and Talbot 2007), assuming gamma hyperpriors $p(\xi) \propto \frac{1}{\xi} \exp\{-\xi\}$ for both parameters. This

yields the following expression for the log marginal posterior density of the parameters,

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

where P is the marginal posterior density, $\omega_i = \{\omega_{i1}, \omega_{i2}, ..., \omega_{iK}\}$ is the observation parameter vector controlling the interpolation function, $Y_i = \{Y_{i1}, Y_{i2}, ..., Y_{iJ}\}$ corresponds to the sequence of observations of state variable i at time step j, J is the total number of time steps in the time series, $\varepsilon_{ij}^{(o)}$ is the observation error at time step j between the interpolated and observed value of variable i, K 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, $\varepsilon_{ij}^{(p)} \sim \mathcal{N}(0, \sigma_i)$, and parameters, $\theta_{ik} \sim \mathcal{N}(0, \delta_{ik})$. This gives the following expression for the log posterior density of the parameters given the interpolations,

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

where $\theta_i = \{\theta_{i1}, \theta_{i2}, ..., \theta_{iK}\}$ are the NODE parameters of the i^{th} variable, $\omega = \{\omega_1, \omega_2, ..., \omega_I\}$ are the interpolation parameters of each state variable, $\varepsilon_{ij}^{(p)}$ is the process error of variable i at time step j between the interpolated dynamics and NODE prediction, σ_i is the standard deviation of the likelihood, K is the total number of parameters, δ_{ik} is the standard deviation of the prior distribution

of parameter θ_{ik}.

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

2.4 Inference and uncertainty quantification

Finally, we estimate uncertainty in parameter values by *anchored ensembling*, which produces approximate Bayesian estimates of the posterior distribution of the parameters (Pearce et al. 2018). The technique requires sampling a parameter vector from the prior distributions, $\theta_i \sim \mathcal{N}(0, \delta_i)$, and then optimising the posterior distribution from this starting point, $\theta_i^* = \underset{\theta_i}{argmax}\ \log p(\theta_i \mid \omega)$. By repeatedly taking samples, the sampled distribution θ^* approaches the posterior distribution and provides estimates and error around the quantities that can be derived from the models. The expectation and uncertainty around derived quantities can then be obtained by computing the mean and variance of the approximated posterior distributions. The great strength of this approach is that it is unlikely to get stuck in local maxima and provides a more robust optimisation of the

posterior.

2.5 Analysing NODEs

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

$$e_{ijt} = \frac{\partial}{\partial y_j} \frac{dy_i}{dt}. \tag{9}$$

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

$$c_{ijt} = \frac{dy_j}{dt} \frac{\partial}{\partial y_j} \frac{dy_i}{dt}. \tag{10}$$

To summarise results across the entire time series we can compute the mean effects e_{ij} by averaging e_{ijt} across all time steps, $e_{ij} = K^{-1} \sum_k e_{ijk}$, as well as the relative total contribution, c_{ij}, of a variable to the dynamics of another by computing the relative sum of square contributions, $c_{ij} = \left(\sum_{ijk} c_{ijk}^2 \right)^{-1} \sum_t c_{ijt}^2$. By computing the direction and strength of interactions between all the

variables in the system we can build dynamically informed ecological interaction networks (See case study below for examples). Other metrics can be computed by analysing the NODEs, such as equilibrium states, these are discussed in our previous work (Bonnaffé, Sheldon, and Coulson 2021).

3 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 ground 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.

3.1 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 3,

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

where G, B, and R, correspond to the prey, intermediate and top predator population densities,

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

We simulate a case of invasion, by introducing the top predator from rare, 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 rotifer predator is rare, and in the later section populations have attained a fixed equilibrium point. The resulting time series are presented in figure 2.

3.2 NODE model

In order to learn non-parametrically 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 nonparametric nonlinear model by releasing the constraint on the nonlinear/exponential section of the neural network during cross-validation. The number of units in the hidden layer J is chosen to be 10, as this is a commonly used number for systems of that size (e.g. Wu, Fukuhara, and Takeda 2005).

3.3 Time series interpolation

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

3.4 Fitting NODEs to the interpolated time series

We fit the NODE system to the interpolated time series. In practice, we fit the NODE to the expectation of the interpolated state and dynamics, $\mathrm{E}(\tilde{y}_i)$ and $\mathrm{E}(d\tilde{y}_i/dt)$, by averaging over all sampled interpolation parameters. An alternative approach could be to consider the interpolation that maximises the log marginal posterior distribution, but this may decrease repeatability due to the difficulty of reliably identifying a global maximum. Averaging across multiple interpolations ensures an overall smoother and robust interpolation. In addition, we standardise the response and explanatory variables with respect the their mean and standard deviation (i.e. $Z = (Y - \mu)/\sigma$). This is to facilitate the training of the NODE by equalizing the scale of the different parameters in the neural network. Then, we identify the optimal regularisation parameter δ (Eq. 8) by cross validation. To do that, we split the data in half and calculate the log likelihood of the test set for increasing values of δ, from 0.05 (linear) to 0.5 (highly non-linear), by increments of 0.05. This allows us to identify the maximum degree of non-linearity, δ, in the per-capita growth rate that ensures generalisability throughout the time series. Then, we approximate the posterior distribution of the NODE parameters by taking 100 samples from the posterior distribution (Eq. 8). Finally, we perform model selection by removing variables that do not result in a significant decrease in the log likelihood of the model (assessed by comparing log likelihood confidence intervals). We ensure moderate temporal autocorrelation and normality by visualising the residuals of the models. We also ensure results repeatability running the entire fitting process a second time.

3.5 Computing ecological interactions

Finally, we analyse the shape of the per-capita growth rates to recover the interaction between the three species in the system. In particular, we look at the effect and contribution of each species to the dynamics of the others. The effect is computed as the sensitivity (i.e. the gradient) of the per-capita growth rate of a given species with respect to the density of the other species. The contribution is computed following the Geber method (Hairston et al. 2005), which consists in multiplying the dynamics of a variable by its effects on the other variables. We further compute the importance of a species in driving the dynamics of another by computing its relative total contribution compared to other species. More details on how to compute these quantities can be found in section 2.5 and in our previous study (Bonnaffé, Sheldon, and Coulson 2021).

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

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

4.1 System

We consider a three-species laboratory microcosm consisting of an algal prey (*Chlorella autrophica*), a flagellate intermediate predator (*Oxyrrhis marina*), and a rotifer top predator (*Brachionus plicatilis*). The algal prey is consumed by the intermediate and top predator, which also consumes

the intermediate predator (Arndt 1993). The dynamics of this system, here the daily change in the density of each species, were recorded in three replicated time series experiments performed by Hiltunen and colleagues (Hiltunen et al. 2013). We use their time series because they describe a simple yet biologically realistic ecosystem, and because the quality of the replication of their microcosm reduces as much as possible observational and experimental error, and rules out environmental variation (Hiltunen et al. 2013). We digitised these time series by extracting by hand the coordinates of every points in the referential of the axis of the graph of the original study, and analysed them.

4.2 NODE analysis

We apply the same analysis as performed on the artificial tri-trophic prey-predator oscillations. This allows us to recover a non-parametric approximation of the growth rate of each species, and then derive the direction and strength of the ecological interactions that underpin their dynamics. We present detailed results of the analysis of the first time series (Fig. 4), and a summary comparison of the three time series (Fig. 5).

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

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

and Coulson 2021).

5.1 System

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

5.2 NODE analysis

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

6 Results

We analyse sequentially the dynamics of each species, focussing on the amount of variation in per-capita growth rates explained by the NODE model, the overall direction, consistency, and importance of ecological interactions, and differences across replicates. Results are summarised in Table 1 and described in details for each species in the following section.

Drivers of top predator dynamics

Figure 2 presents the drivers of the dynamics of rotifer. The NODE approximation of the per-capita growth rate fits quite well the interpolated per-capita growth rate across all replicates (Fig. 2, A2

B2 and C2, $r^2 > 0.7$, Table 1). The analysis of effects reveals overall a positive effect of algae on rotifer growth in all replicates (Fig. 2, A3, B3, C3, green line). The intermediate predator has a positive effect on rotifer growth in replicates A and C only (Fig. 2, A3, B3, C3, blue line). We find positive intra-specific density-dependence in the first replicate only (Fig. 2, A3, red line). Overall, all effects are consistent throughout the time series. The algae is the dominant driver of rotifer dynamics as it accounts for 55%, 93%, and 74% of the change in per-capita growth rates across the three replicates (Table 1, Fig. 2, A5, B5, C5, green line).

Drivers of the prey dynamics

The per-capita growth rate of the algae is well explained by the NODE approximation (Fig. 3, A2, B2, C2, $r^2 > 0.8$, Table 1). Overall, rotifers have a negative impact on the growth of algae in all replicates (Fig. 3, A3, B3, C3, red line). We find evidence for negative density-dependence in replicate A and positive density-dependence in replicate B, but not in replicate C (Fig. 3, A3, B3, C3, green line). The intermediate predator has an overall negative effect on Algae only in replicate B (Fig. 3, B3, blue line). The main driver of algae dynamics is the rotifer population, which accounts for 58%, 44%, and 90% of the change in algae per-capita growth rate across the three replicates. Density-dependence, however, plays a role in replicate A and B, with 40% and 24% of total change in growth, respectively (Table 1). The intermediate predator contributes only to algae growth in replicate B, accounting for 32% change in growth (Table 1). Overall, effects are found to be consistent throughout the time series except in replicate B (Fig. 3, B3), where effects vary in complicated ways, leading to a period in the time series where the algae is mostly driven by

the intermediate predator and positive density-dependence, and less impacted by the top predator (Fig. 3, B5, from time 3 to 7.5).

Drivers of the intermediate predator dynamics

The per-capita growth rate of the intermediate predator is quite well captured by the NODE approximation (Fig. 4, A2, B2, C2, $r^2 > 0.7$, Table 1). The intermediate predator is mainly negatively affected by the rotifer population (Fig. 4, A3, B3, C3, red line). The algae has a negative effect on flagellate growth in replicate A, and a positive one in replicate B (Fig. 4, A3, B3, green line). The rotifer predator dynamics accounts for 78%, 62%, 91% of the change in the flagellate growth rate, and the algae 20% and 37% in replicate A and B, respectively (Table 1, Fig. 4, A5, B5, C5). Overall, effects are consistent throughout the time series.

7 Discussion

Our ability to generalise dynamical processes and patterns across populations and communities is limited by the complexity of the processes, differences in environments, and incomplete and/or erroneous observations. It remains unclear to what extent generalisation would be possible if we overcame these limitations. We tackle this question by looking at the consistency of dynamical patterns across three replicated runs of a simple three-species community, hosted in identical environmental conditions in the lab. We expected to find consistency in the drivers of population dynamics, both in time and across replicates, and thereby demonstrate that generalisation of dynamical processes may be possible if the system states were well-observed and environmental

conditions were known. To verify this expectation we (1) characterised the amount of variation in per-capita growth rates that is explainable deterministically, (2) quantified the direction, strength, and importance of ecological interactions for the growth of each population, and (3) described how these varied in time and across replicates. Our results are summarised in Figure 5. We find that only the effect of algae on rotifer ($G \to R$), and that of rotifer on algae ($R \to G$) and flagellate ($R \to B$) are conserved across the replicates. We find strong variation in the direction and importance of intra-specific density-dependence in rotifer ($R \to R$) and algae ($G \to G$) growth across the three replicates. The role played by the intermediate predator in the system was also different in all replicates, in that it only contributed substantially to the dynamics of the algae in replicate B ($B \to G$), and was either negatively, positively, or not affected by the algae ($G \to B$). Overall, this shows that the dominant interactions are conserved across replicates, but that minor interactions vary substantially in importance and effect. Furthermore, we find that these dynamical processes are more consistent in time within a system, than across replicates. Our results demonstrate that because of partially generalisable dynamical processes, dynamical patterns may not be generalisable across systems, even with limited observation error and when environmental conditions and community structure are conserved.

Overall, our results are consistent with the biology of the system. The rotifer top-predator is found to have a strong negative impact on the two other species, in spite of variation in prey preference across replicates. This is consistent with previous study which have established the importance of rotifers for top-down control of flagellate and algal populations (Arndt 1993; Hiltunen et al.

2013). What is more suprising is the positive intra-specific density-dependence in the growth rate of the rotifer population in replicate A. This implies that the population of rotifer grows more at high density. This might be explained by various biological mechanisms, such as cannibalism (Gilbert 1976), though evidence remains limited in the *Brachionus* genus, or higher mating success at high density (Snell and Garman 1986). Similarly, the algae shows signs of positive intra-specific density-dependence in replicate B, though this effect remains confined to a brief period in the time series. This may be due to a higher chance of evading predators at high-density. This shows that the NODEs approach used here recovers results consistent with existing knowledge, but also identify subtle, more intriguing dynamical processes.

What might be the drivers of differences in the dynamical processes across these three replicates? One of the main source of variation in dynamics may be differences in the intrinsic structure of populations, such as variation in traits influencing intra- and inter-specific interactions which may lead to different dynamics (Yoshida et al. 2003; Yoshida et al. 2007; De Meester et al. 2019; Bruijning, Jongejans, and Turcotte 2019). Differences in the phenotypic structure may be due to unaccounted variation in initial conditions (Becks et al. 2010), or variation that developed throughout time as a result of evolution (e.g. Yoshida et al. 2003; Yoshida et al. 2007). In particular, the algae in this system is prone to evolve a predator defence behaviour, by forming clumps, which reduce predation risk (Yoshida et al. 2003; Hiltunen et al. 2013). In their original paper, the authors limited the initial genetic diversity in the algae and focussed on replicates which did not display evidence of evolution, in an attempt to limit the impact of initial variation in phenotypic structure,

and of evolution, on the dynamics (Hiltunen et al. 2013). In spite of that, evolution may not be eliminated completely, thus variation in traits governing the interactions between the species in the system may still have developed during the experiment, and led to changes in the dynamical processes across replicates. This would further be consistent with results from Yoshida and colleagues, who showed that evolution of prey defense could lead to ecological dynamics inconsistent with the known trophic interactions (Yoshida et al. 2007). Becks and colleagues also showed that small changes in the initial genotypic diversity could lead to drastically different eco-evolutionary dynamics (Becks et al. 2010). Our study hence reinforces the idea that rapid evolution may prevent generalisation of dynamical processes (Ezard, Côté, and Pelletier 2009; De Meester et al. 2019), and further suggests that this may also be the case in simple systems with limited environmental variation and opportunity for evolution.

Alternatively, stochasticity may be a major driver of differences across systems (Dallas et al. 2021). First, stochasticity in initial conditions, arising from the sampling of the communities of each replicate, could introduce differences in the interactions between the three populations. Second, stochasticity in the population dynamics themselves may result in different changes in densitiy levels in communities that are otherwise identical. Because our modelling approach is deterministic, it does not directly provide an estimate of the total variation explained by stochasticity. Our modelling approach decomposes the variation in the data into observation and process error (Calder et al. 2003). First, the interpolation step introduces residual observation error, namely variation that is not captured by the interpolation. Second, the fitting of the NODE to the interpolation introduces

residual process error, which is variation in the observation model that is not explained by the process modelled by the NODE. Stochasticity in the dynamics could explain the observation and process residual error (Calder et al. 2003), while stochasticity in initial conditions can only influence differences across replicates. Yet, we find relatively small process and observation error ($> 70\%$ of variance explained). So that, the dynamics of the three species are well explained by relatively simple linear deterministic effects between the state variables, which means that though dynamical processes differ across replicates they are reasonably consistent in time within each system. This suggests that stochasticity in dynamics plays a minor role in driving differences in dynamics across replicates, compared to stochasticity in initial conditions. In order to quantify this, we would need to estimate the influence of stochasticity directly. This can be done by modelling explicitly the random distribution of model parameters that underpin the dynamics of populations, which would then inform us about the importance of stochasticity driven by variation at the individual-level (Fox and Kendall 2002). Additionally, we could model stochasticity explicitly in the model with neural stochastic differential equations, which would allow us to separate the amount of change explainable by the deterministic part of the model, from demographic stochasticity, at each time step (Jia and Benson 2019).

Finally, we cannot exclude the potential contribution of unobserved variables that were not monitored during the experiment, such as variation in nutrient levels in the chemostat, and which may also lead to differences in the predation and intra-specific interactions across systems (e.g. Bonsall, Van Der Meijden, and Crawley 2003; Fussmann and Blasius 2005; Posey, Alphin, and Cahoon

2006).

Should we expect limited generalisability of dynamics across systems, even if the complexity of the process is properly captured, environmental conditions known, and the system well-observed? A similar study, that inferred dynamical processes consistency from replicated time series of a simple rotifer system, found substantial variation in vital rates across replicates (Rosenbaum et al. 2019), also pointing at a low generalisability of dynamical processes. Yet, the level of replication of the time series of their studies was not as stringent as that of the time series we considered, which leaves room for variability in dynamics to be caused by differences in experimental setup, population history, initial densities. Bruijning and colleagues also found substantial variation in vital rates across clones in a replicated system of aphids, showing that slight phenotypic variations can change the population dynamics, all else being equal (Bruijning, Jongejans, and Turcotte 2019). This phenomenon is likely to be even more important in more complicated systems and in a natural setting where most variables are unobserved, which poses a problem for the generalisation of results across studies and systems (De Meester et al. 2019). How can we expect to generalise dynamics across real systems if we are not able to do so in artificial systems? Overall, our study reinforces the view that general inferences should not be drawn from a single system, and that more efforts are required to distinguish dynamical patterns that are conserved across systems from idiosyncratic ones.

Can we trust our models then if they are doomed to provide partly idiosyncratic answers? Our study demonstrates that processes can vary substantially across replicates, so that there may hence

not be a single suitable functional form and parametrisation to model them (Lawton 1999). Yet, most of the work to date has involved fitting parametric models to time series data (e.g. Bruijning, Jongejans, and Turcotte 2019; Pontarp, Brännström, and Petchey 2019; Rosenbaum et al. 2019), which provide a very narrow view of the range of possible functions to describe the biological processes at play (Jost and Ellner 2000; Adamson and Morozov 2013). These models are subjective by nature (Jost and Ellner 2000; Adamson and Morozov 2013), and hence not generalisable, so that they greatly reduce our chance at identifying dynamical processes that are idiosyncratic, and those that are transferable.

What alternatives do we have then? We propose that NODEs are a suitable framework to study dynamical processes, as they produce inferences that are free of model assumption and facilitate comparison across studies and systems (Bonnaffé, Sheldon, and Coulson 2021). In this sense, our study already provides a potentially more objective depiction of dynamical processes than previous work with parametric models. Furthermore, in this paper we overcame the practical challenges of implementing NODEs by providing a computationally efficient fitting procedure, relying on time series interpolation, and developed a model selection criterion robust to overfitting. Similar approaches have been proposed in the past, for instance Ellner and colleagues developed a method called gradient matching where they interpolated the data with cubic splines to which they fitted the differential equations (Jost and Ellner 2000; Ellner, Seifu, and Smith 2002). Wu and colleagues also relied on data interpolation of the data with ANNs to fit non-parametric approximations of population vital rates (Wu, Fukuhara, and Takeda 2005). But the approaches were too challenging

and cumbersome to be implemented routinely, and were not used to tackle ecological interactions. Overall, our work demonstrates the usefulness of NODEs for inferring ecological interactions from count time series, which could readily be applied to a substantial pool of time series data.

Conclusion

Generalising dynamics across biological systems is hard because of the complexity of the dynamical processes (e.g. ecological interactions), differences in environmental context, and monitoring limitations. It remains unclear whether we could generalise dynamics if we properly modelled complexity, controlled for environmental effects, and observed systems precisely. We addressed this question by looking at the generalisability of dynamical processes across three replicated time series of a three-species system, using the novel framework of NODEs. We found that only the dominant interactions were conserved across the three time series, namely that between the algae and the rotifer, while the role of the intermediate predator varied substantially. Our results hence suggest that generalisation may not seem possible, even in simple system with no environmental variation. Given previous work in this system, the main cause of differences across replicates may be evolution in prey defence traits. We conclude that more work is required, using NODEs, to identify dynamical patterns that are conserved and those that are idiosyncratic across a wider range of systems.

Acknowledgments

We thank warmly the Ecological and Evolutionary Dynamics Lab and Sheldon Lab Group at the department of Zoology for their feedback and support. We thank Ben Sheldon for insightful sug-

gestions on early versions of the work. The work was supported by the Oxford-Oxitec scholarship and the NERC DTP.

Data accessibility

All data and code will be made fully available at https://github.com/WillemBonnaffe/NODER/rotifer.

Statement of authorship

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

References

Adamson, M. W. and A. Y. Morozov (2013). "When can we trust our model predictions? Unearthing structural sensitivity in biological systems". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 469.2149, pp. 1–19.

Arndt, H. (1993). "Rotifers as predators on components of the microbial web (bacteria, heterotrophic flagellates, ciliates) - a review". In: *Hydrobiologia* 255-256.1, pp. 231–246.

Becks, L., S. P. Ellner, L. E. Jones, and N. G. J. Hairston (2010). "Reduction of adaptive genetic diversity radically alters eco-evolutionary community dynamics". In: *Ecology Letters* 13.8, pp. 989–997.

Becks, L., S. P. Ellner, L. E. Jones, and N. G. J. Hairston (2012). “The functional genomics of an eco-evolutionary feedback loop: Linking gene expression, trait evolution, and community dynamics”. In: *Ecology Letters* 15.5, pp. 492–501.

Bonnaffé, W., A. Danet, S. Legendre, and E. Edeline (2021). “Comparison of size-structured and species-level trophic networks reveals antagonistic effects of temperature on vertical trophic diversity at the population and species level”. In: *Oikos* 130.8, pp. 1297–1309.

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

Bonsall, M. B., E. Van Der Meijden, and M. J. Crawley (2003). “Contrasting dynamics in the same plant-herbivore interaction”. In: *Proceedings of the National Academy of Sciences of the United States of America* 100.25, pp. 14932–14936.

Bruijning, M., E. Jongejans, and M. M. Turcotte (2019). “Demographic responses underlying eco-evolutionary dynamics as revealed with inverse modelling”. In: *Journal of Animal Ecology* 88.5, pp. 768–779.

Calder, C., M. Lavine, P. Müller, and J. S. Clark (2003). “Incorporating multiple sources of stochasticity into dynamic population models”. In: *Ecology* 84.6, pp. 1395–1402.

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

563 Dallas, T., B. A. Melbourne, G. Legault, and A. Hastings (2021). “Initial abundance and stochas-
564 ticity influence competitive outcome in communities”. In: *Journal of Animal Ecology*, pp. 1–
565 26.
566 De Meester, L. et al. (2019). “Analysing eco-evolutionary dynamics—The challenging complexity
567 of the real world”. In: *Functional Ecology* 33.1, pp. 43–59.
568 Demyanov, V., S. N. Wood, and T. J. Kedwards (2006). “Improving ecological impact assessment
569 by statistical data synthesis using process-based models”. In: *Journal of the Royal Statistical*
570 *Society. Series C: Applied Statistics* 55.1, pp. 41–62.
571 Ellner, S. P., Y. Seifu, and R. H. Smith (2002). “Fitting Population Dynamic Models to Time-Series
572 Data by Gradient Matching”. In: *Ecology* 83.8, p. 2256.
573 Ezard, T. H. G., S. D. Côté, and F. Pelletier (2009). “Eco-evolutionary dynamics: Disentangling
574 phenotypic, environmental and population fluctuations”. In: *Philosophical Transactions of the*
575 *Royal Society B: Biological Sciences* 364.1523, pp. 1491–1498.
576 Fox, G. A. and B. E. Kendall (2002). “Demographic stochasticity and the variance reduction ef-
577 fect”. In: *Ecology* 83.7, pp. 1928–1934.
578 Fussmann, G. F. and B. Blasius (2005). “Community response to enrichment is highly sensitive to
579 model structure”. In: *Biology Letters* 1.1, pp. 9–12.
580 Gamelon, M. et al. (2019). “Accounting for interspecific competition and age structure in demo-
581 graphic analyses of density dependence improves predictions of fluctuations in population size”.
582 In: *Ecology Letters* 22.5, pp. 797–806.

583 Gilbert, J. J. (1976). “Selective cannibalism in the rotifer Asplanchna sieboldi: Contact recognition
584 of morphotype and clone”. In: *Proceedings of the National Academy of Sciences* 73.9, pp. 3233–
585 3237.
586 Gross, K., A. R. Ives, and E. V. Nordheim (2005). “Estimating fluctuating vital rates from time-
587 series data: A case study of aphid biocontrol”. In: *Ecology* 86.3, pp. 740–752.
588 Hairston, N. G. J., S. P. Ellner, M. A. Geber, T. Yoshida, and J. A. Fox (2005). “Rapid evolution and
589 the convergence of ecological and evolutionary time”. In: *Ecology Letters* 8.10, pp. 1114–1127.
590 Hiltunen, T., L. E. Jones, S. P. Ellner, and N. G. J. Hairston (2013). “Temporal dynamics of a simple
591 community with intraguild predation: an experimental test”. In: *Ecology* 94.4, pp. 773–779.
592 Jia, J. and A. R. Benson (2019). “Neural jump stochastic differential equations”. In: *Advances in
593 Neural Information Processing Systems* 32.NeurIPS.
594 Jost, C. and S. P. Ellner (2000). “Testing for predator dependence in predator-prey dynamics: A
595 non-parametric approach”. In: *Proceedings of the Royal Society B: Biological Sciences* 267.1453,
596 pp. 1611–1620.
597 Kendall, B. E. et al. (2005). “Population cycles in the pine looper moth: Dynamical tests of mech-
598 anistic hypotheses”. In: *Ecological Monographs* 75.2, pp. 259–276.
599 Lawton, J. H. (1999). “Are There General Laws in Ecology ?” In: *Oikos* 84.2, pp. 177–192.
600 Odum, E. P. and G. W. Barrett (1972). “Fundamentals of Ecology”. In: *The Journal of Wildlife
601 Management* 36.4, p. 1372.
602 Pearce, T., F. Leibfried, A. Brintrup, M. Zaki, and A. Neely (2018). “Uncertainty in Neural Net-
603 works: Approximately Bayesian Ensembling”. In: *arXiv*, pp. 1–10.

Pontarp, M., Å. Brännström, and O. L. Petchey (2019). “Inferring community assembly processes from macroscopic patterns using dynamic eco-evolutionary models and Approximate Bayesian Computation (ABC)”. In: *Methods in Ecology and Evolution* 10.4, pp. 450–460.

Posey, M. H., T. D. Alphin, and L. Cahoon (2006). “Benthic community responses to nutrient enrichment and predator exclusion: Influence of background nutrient concentrations and interactive effects”. In: *Journal of Experimental Marine Biology and Ecology* 330.1, pp. 105–118.

Raeymaekers, J. A. M. et al. (2017). “Adaptive and non-adaptive divergence in a common landscape”. In: *Nature Communications* 8.1, pp. 1–8.

Reznick, D. N., H. Bryga, and J. A. Endler (1990). “Experimentally induced life-history evolution in a natural population”. In: *Nature* 346.6282, pp. 357–359.

Rosenbaum, B., M. Raatz, G. Weithoff, G. F. Fussmann, and U. Gaedke (2019). “Estimating parameters from multiple time series of population dynamics using bayesian inference”. In: *Frontiers in Ecology and Evolution* 6.234, pp. 1–14.

Shurin, J. B., J. L. Clasen, H. S. Greig, P. Kratina, and P. L. Thompson (2012). “Warming shifts top-down and bottom-up control of pond food web structure and function.” In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 367.1605, pp. 3008–17.

Snell, T. W. and B. L. Garman (1986). “Encounter probabilities between male and female rotifers”. In: *Journal of Experimental Marine Biology and Ecology* 97.3, pp. 221–230.

623 Thompson, C. E., E. B. Taylor, and J. D. Mcphail (1997). “Parallel Evolution of Lake-Stream
624 Pairs of Threespine Sticklebacks (Gasterosteus) Inferred from Mitochondrial DNA Variation”.
625 In: *Evolution* 51.6, pp. 1955–1965.
626 Wu, J., M. Fukuhara, and T. Takeda (2005). “Parameter estimation of an ecological system by a
627 neural network with residual minimization training”. In: *Ecological Modelling* 189.3-4, pp. 289–
628 304.
629 Yoshida, T., S. P. Ellner, L. E. Jones, B. J. M. Bohannan, R. E. Lenski, and N. G. J. Hairston (2007).
630 “Cryptic population dynamics: Rapid evolution masks trophic interactions”. In: *PLoS Biology*
631 5.9, pp. 1868–1879.
632 Yoshida, T., L. E. Jones, S. P. Ellner, G. F. Fussmann, and N. G. J. Hairston (2003). “Rapid evo-
633 lution drives ecological dynamics in a predator – prey system”. In: *Nature* 424.July, pp. 303–
634 306.

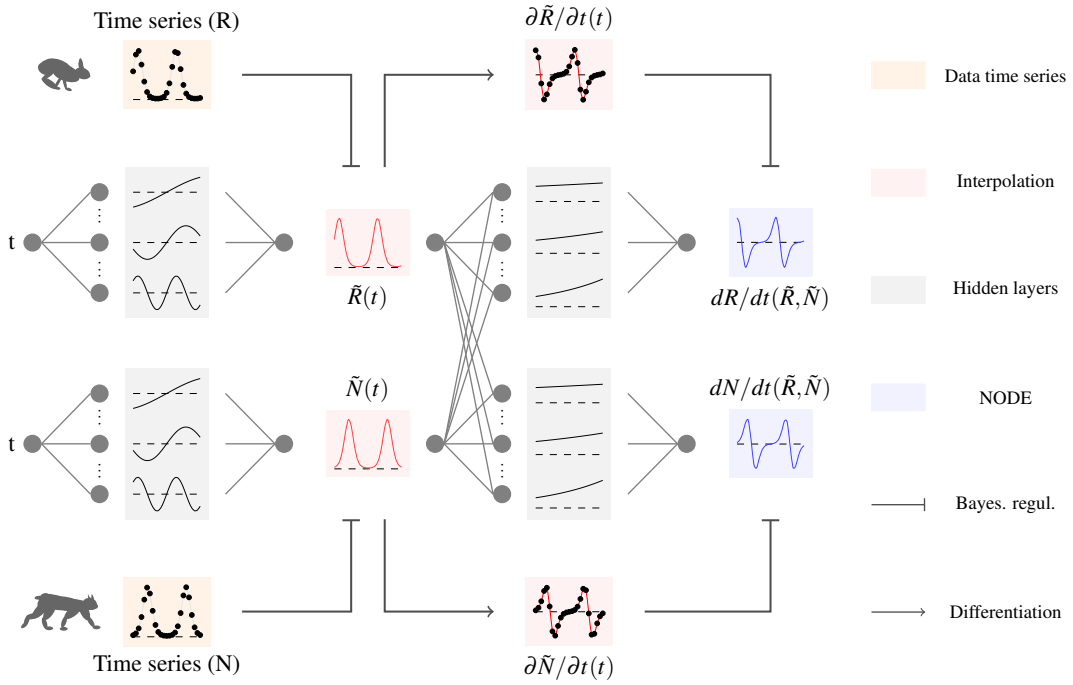

Figure 1: Overview of fitting neural ordinary differential equations (NODE) by Bayesian neural gradient matching (BNGM). In a first step we compute a continuous time approximation (interpolation) of each state variables, here the prey $\tilde{R}(t)$ and predator density $\tilde{N}(t)$. To do that we 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, 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. It takes on average 2.6 seconds to fit NODEs by BNGM, compared to 30 mins in a previous study (Bonnaffé, Sheldon, and Coulson 2021), which corresponds to a 735 fold increase in speed.

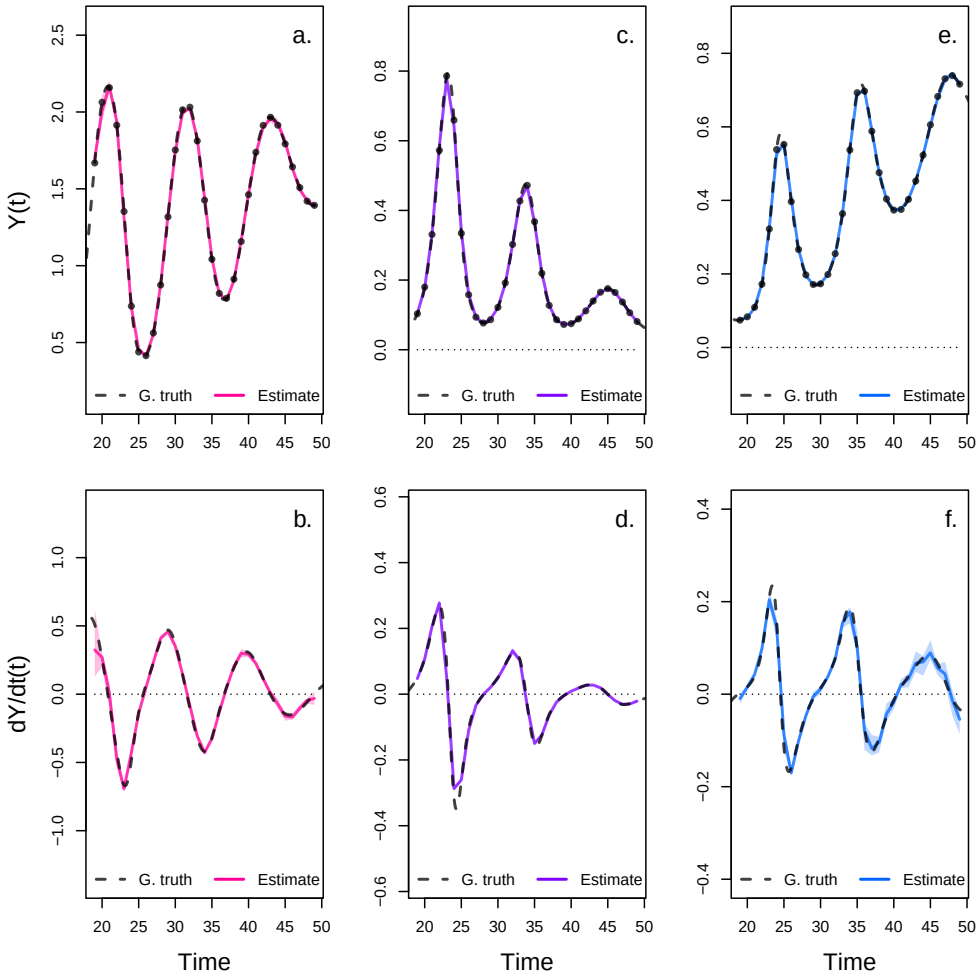

Figure 2: Interpolated density and dynamics of algae, flagellate, and rotifer in the artificial system. This figure corresponds to the first step in the overview figure. It shows the accuracy of the interpolated densities of algae (a.), flagellate (c.), and rotifer (e.). We obtain interpolated densities 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.

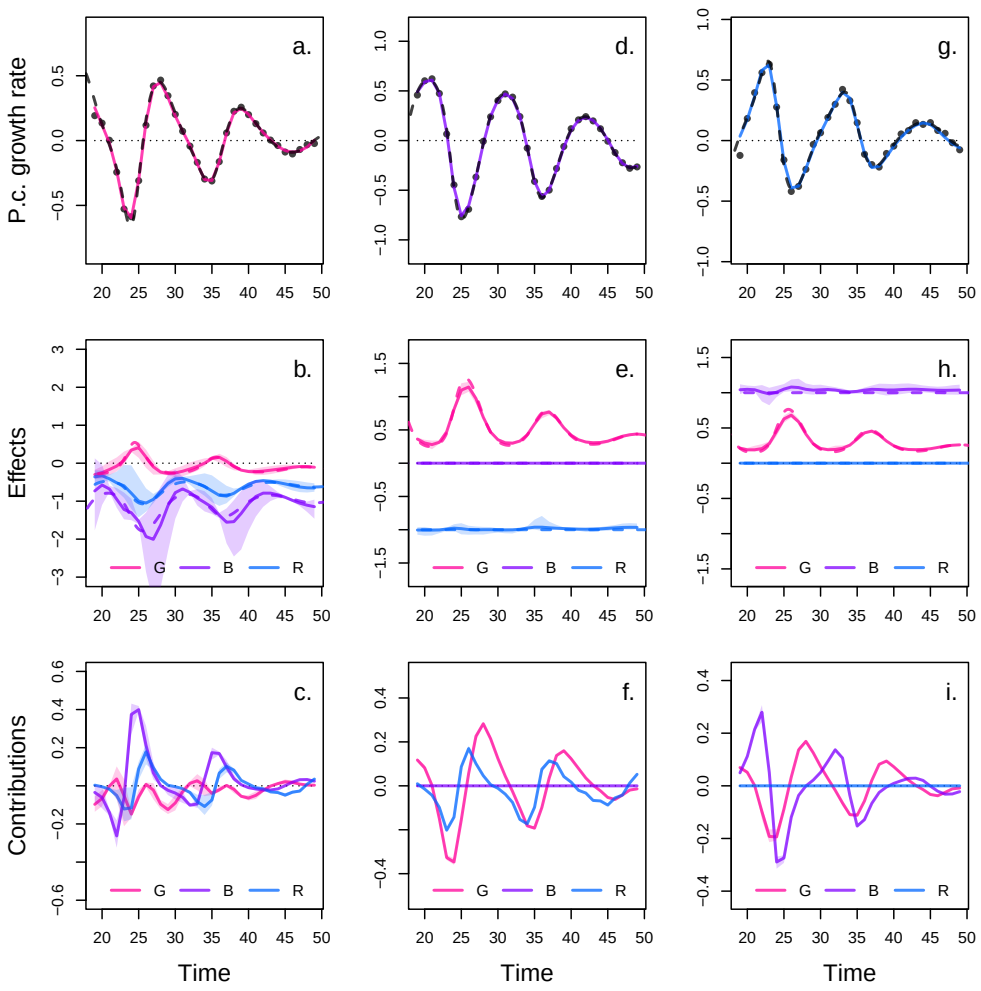

Figure 3: Drivers of dynamics of algae, flagellate, and rotifer in the artificial system. This figure corresponds to the second step in the overview figure. It displays the NODE non-parametric approximations of the per-capita growth rate of algae (a., b., c.), flagellate (d., e., f.), and rotifer (g., 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 (fig. 1, b., d., f.) with its effects. Dashed lines correspond to ground truth, obtained from the original trajectories of the tri-trophic ODE model. The shaded area shows the 90% confidence interval, obtained by approximately sampling the posterior distributions.

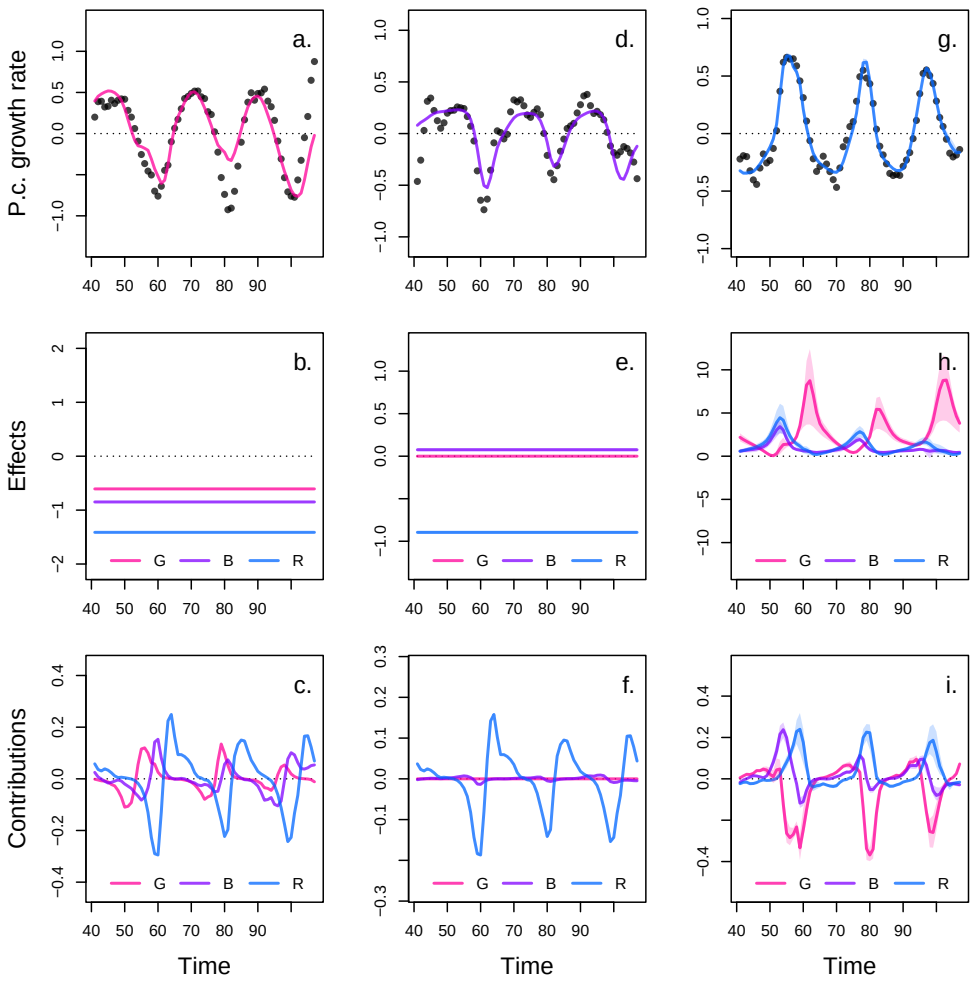

Figure 4: Drivers of dynamics of algae, flagellate, and rotifer in replicate A. This figure displays the NODE non-parametric approximations of the per-capita growth rate of algae (a., b., c.), flagellate (d., e., f.), and rotifer (g., h., i.). We obtain the NODE approximations (a., d., g., solid line)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

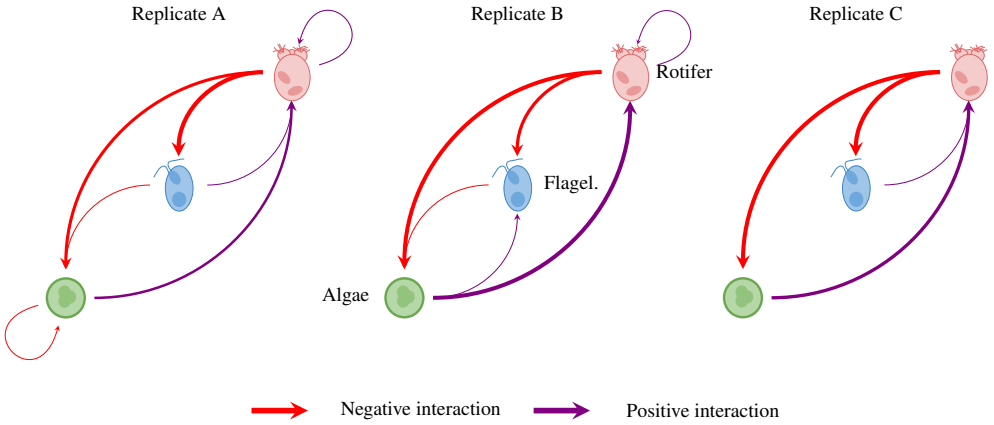

Figure 5: Interaction networks inferred from 3 replicated time series of algae, flagellate, and rotifers. This figure shows the direction and strength of ecological interactions inferred from 3 replicated sets of time series of algae, flagellate, and rotifer, using NODEs fitted by gradient matching. The replicates B and C were analysed in the same way as replicate A (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 red line in fig. 5 f., and computing the % of total contributions that it accounts for. We provide the value of the mean effects and relative strengths in Table 1. The replicated time series were obtained by digitising the time series in Hiltunen et al. (2013).

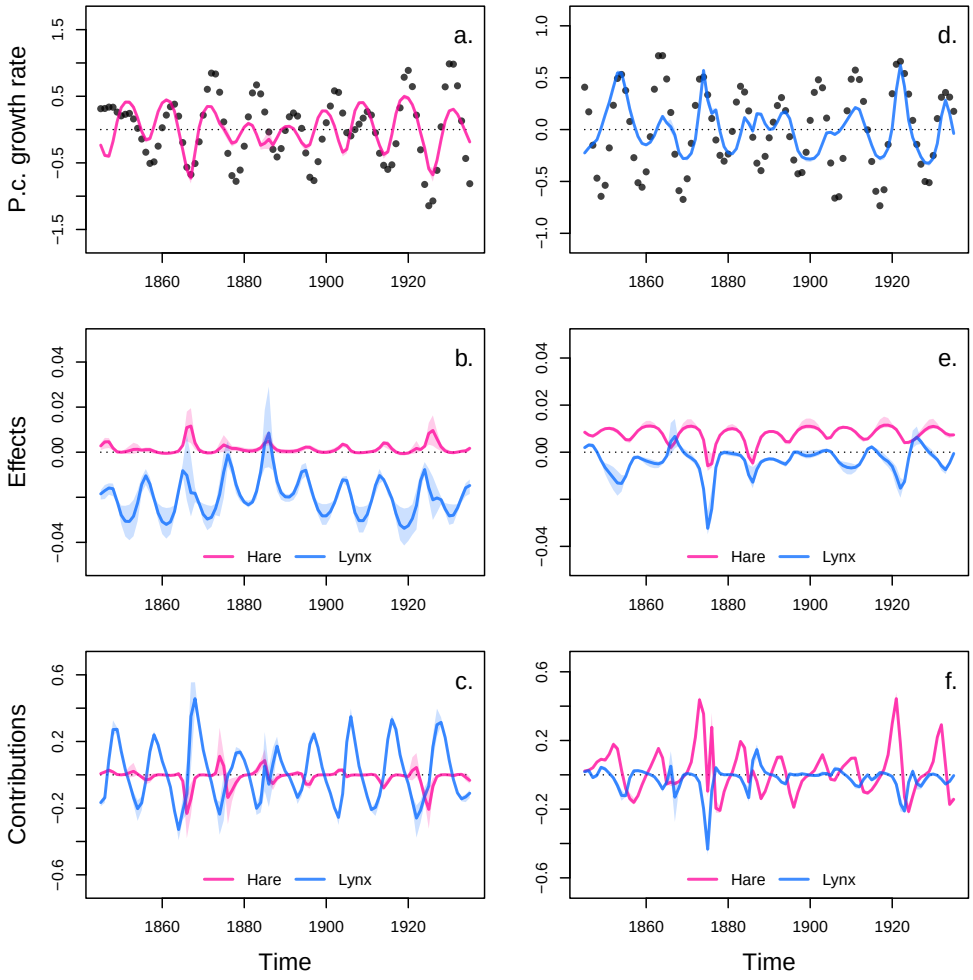

Figure 6: Drivers of dynamics of hare and lynx in the Odum and Barrett pelt count time series. This figure displays the NODE non-parametric approximations of the per-capita growth rate of hare (a., b., c.), and lynx (d., e., f.). We obtain the NODE approximations (a., d., solid line) by fitting the interpolated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

Table 1: Comparison of the direction and strentgh 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 non-parametric approximation of the pre-capita growth rate of the three species.

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

Table 2: Summary of model runtimes. We measured the time required to perform 100 interpolations and 30 NODE fits to each variable in the systems. Replicate A, B, and C correspond to each replicated time series of the aglae, flagellate, and rotifer tri-trophic system (Hiltunen et al. 2013). The Hare-Lynx system correspond to the 90 years long time series of hare and lynx pelt counts (Odum and Barrett 1972). The number of time steps (N steps) is given for each time series. The total time per fit is obtain by dividing the total time in seconds by the number of fits (i.e. 130). It takes on average 5.35 minutes for the 130 NODE fits NODE, which amounts to 2.47 seconds per sample taken. This is 735 times faster than the 30 minutes fitting times obtained in a previous study (Bonnaffé, Sheldon, and Coulson 2021).

			Interpolation		NODE fit			
System	N var.	N steps	N fits	time (s)	N fits	time (s)	total	total p. fit
Replicate A	3	66	100	239.47	30	129.41	368.88	2.84
Replicate B	3	66	100	233.59	30	133.13	366.72	2.82
Replicate C	3	40	100	136.51	30	74.01	210.52	1.62
Hare-lynx	2	90	100	303.64	30	33.56	337.20	2.59

8 Supplementary

A Bayesian regularisation

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

$$p(\theta|\mathscr{D}) \propto p(\mathscr{D}|\theta)p(\theta) \tag{13}$$

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

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

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

The prior probability density functions for the parameters are given by

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

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

There is no standard approach for choosing δ. 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 the models on the second-level of inference. This means that we are finding the optimal value of δ, in addition to optimising the model parameters. We do this by optimising the marginal posterior density of the parameters, obtained by averaging out δ following a modification of the approach developped by Cawley and Talbot (Cawley and Talbot 2007). This yields the following expression for the marginal log posterior density of the parameters

$$\log P(\Omega|\mathscr{D}) \propto -\frac{I}{2}\log\left(1+\sum_{i=1}^{I}\left(\varepsilon_i^{(o)}\right)^2\right)-\frac{J}{2}\log\left(1+\sum_{j=1}^{J}\Omega_j^2\right) \tag{16}$$

$$\log p(\beta|\Omega) \propto -\frac{1}{2}\sum_{i=1}^{I}\left(\frac{\varepsilon_i^{(p)}}{\sigma}\right)^2-\frac{1}{2}\sum_{j=1}^{J}\left(\frac{\beta_j}{\delta_j}\right)^2 \tag{17}$$

which amounts to optimising the log of the sum of squared residuals rather than the sum of squared residuals. $P(\theta|\mathscr{D})$ designates the marginal posterior distribution. More details on how to derive this expression from equation (8) can be found in a supplementary file (See supplementary A).

In this section we describe how to derive the modified model selection critieria developed by Caw-

ley and Talbot (Cawley and Talbot 2007). Bayesian regularisation simply amounts to 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 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). Performing inference on the second level means that we are trying to find the appropriate value of the dispersion of the priors, in other words, the appropriate level of constraint on the model. 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 appropriate choice of prior, the dispersion can be integrated out, leaving us with a formula for the posterior that only depends on the parameters of the model,

$$\log P(\theta|\mathscr{D}) \propto -\frac{I}{2}\log\left(\sum_{i=1}^{I} e_i(\mathscr{D},\theta)^2\right) - \frac{J}{2}\log\left(\sum_{j=1}^{J}\theta_j^2\right) \tag{18}$$

where $P(\theta|\mathscr{D})$ denotes the marginal posterior density, $\mathscr{D}$ denotes the evidence, I and J denote the number of data points and parameters, respectively, e_i denote the residuals of the model, and θ 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).

The issue with this formula is that the marginal posterior density is infinity when the parameters are 0, which leads to underfitting. In this paper we use a modified criterion, which corrects for that problem

$$\log P(\theta|\mathscr{D}) \propto -\frac{I}{2}\log\left(1+\sum_{i=1}^{I} e_i(\mathscr{D},\theta)^2\right) - \frac{J}{2}\log\left(1+\sum_{j=1}^{J}\theta_j^2\right) \tag{19}$$

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 residudals. This construction can be obtained simply by assuming a gamma prior for the parameters $p(\xi) \propto \frac{1}{\xi}\exp\{-\xi\}$, where ξ is the regularisation parameter, instead of the improper Jeffreys' prior that Cawley and Talbot used in their original study, namely $p(\xi) \propto \frac{1}{\xi}$. The details of the integration of the posterior distribution over ξ can be found in Cawley and Talbot's orginal paper.