

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

Willem Bonnaffé^{1,2} & Tim Coulson²

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

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

Emails: willem.bonnafe@stx.ox.ac.uk; tim.coulson@~~zo~~^{bio}.ox.ac.uk;

Running title:

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: xxx words in abstract; xxxx words in text; xx references; 6 figures; 2 tables

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

Statement of authorship:

Abstract

Inferring ecological interactions is hard because we often lack suitable parametric representations to portray them. NODEs provide a way to estimate interactions nonparametrically from time series data, with no assumptions required for the interactions. NODEs however are slow to fit, and the reliability of interactions ~~thereby~~ inferred had not been compared to ground truth. We provide a method, Bayesian neural gradient matching, that allows for ~~fitting~~ NODE systems within seconds. We ~~by~~ interpolating time series with neural networks, and fitting NODEs to the interpolated dynamics with Bayesian regularisation. We show that NODEs ~~thereby fitted~~ provide accurate estimates of ecological interactions between variables in an artificial system where ground truth is known. Finally, we infer interactions in ~~real~~ replicated time series of a tri trophic system, as well as in the hare lynx system. We find that dynamics are driven by a mixture of linear and nonlinear ecological interactions, of which only the strongest are consistent across replicates.

1 Introduction

Ecological interactions ~~correspond to interactive effects between different populations in an ecosystem.~~ ^{describe how the size + structure of one population influence the dynamics of other species} The direction and strength of interactions determine ~~the~~ ^{community} dynamics, ~~namely the temporal~~

~~changes in size and structure~~, of the populations. Much effort goes into characterising these interactions, as they are key to understanding and forecasting how ~~our~~ ^{are impacted by anthropogenic} ecosystems change.

Ecological interactions are inherently complex processes as they depend on the interactions between individuals within and across populations, displaying different phenotypes, living in a changing environment. As a result, interactions themselves are highly context dependent ~~traits~~ ^{is an interaction a trait?} so that they change in time with the state of the ecosystem and of the environment (Song et al. 2020).

Much work has provided evidence of this in nature and in the lab (e.g. Bonsall, Meijden, and Crawley 2003; Kendall et al. 2005; Bruijning, Jongejans, and Turcotte 2019; Rosenbaum et al. 2019; Bonnaffé et al. 2021). Negative ~~intra-specific~~ ^{density} dependence is perhaps the most well studied and ubiquitous type of ecological interaction (e.g. Lingjaerde et al. 2001; Moe et al. 2005; Brook and Bradshaw 2006). Even in ~~the most~~ simple ecological systems, featuring ~~an~~ algae and rotifer predator, studies have shown that the predation rate of algae by rotifer is dependent on the density of the predator (Jost and Ellner 2000; Yoshida et al. 2003). The metabolic theory of ecology also uncovered that many vital rates underpinning ecological interactions are temperature and size dependent (Brown et al. 2004). This makes it difficult to identify generally conserved interactions and hence make robust predictions regarding the fate of a given ecosystem (Lawton 1999).

Characterising precisely these interactions and how they vary with extraneous variables is ex-

This paragraph is a bit naive. I am not sure the key message.

→ grow some earlier sentences: Turchin, Rogana etc.

21 tremely difficult, but can be achieved through ~~mathematical~~ modelling of time series data (e.g.
22 Bruijning, Jongejans, and Turcotte 2019; Rosenbaum et al. 2019; Adams et al. 2020). Typically,
23 we would fit a dynamical model, such as an ODE system, to time series of the state variables, and
24 study the parameters that quantify the interactions between the different variables. ~~Paradoxically~~,
25 in order to do so we need suitable mathematical representations for the interactions, which we often
26 lack given the complexity of the processes (Jost and Ellner 2000; Wood 2001; Ellner, Seifu, and
27 Smith 2002; Wu, Fukuhara, and Takeda 2005). This hinders our capacity to infer ecological inter-
28 actions reliably, as any interaction quantified is ultimately contingent on the parametric function
29 chosen arbitrarily by the observer (Adamson and Morozov 2013).

normally
assumed
to be
linear.

30 Nonparametric modelling has proven to be a powerful alternative to solve that problem (e.g. Jost
31 and Ellner 2000; Wood 2001; Ellner, Seifu, and Smith 2002; Wu, Fukuhara, and Takeda 2005;
32 Pasquali and Soresina 2018)). Nonparametric models require minimal assumptions regarding the
33 mathematical nature of ecological interactions (Jost and Ellner 2000; Gross, Ives, and Nordheim
34 2005), and hence provide interaction estimates that are more robust to model structure. In previ-
35 ous work, we introduced a powerful framework, relying on neural ordinary differential equations
36 (NODEs, Chen et al. 2019) to approximate the dynamics of populations nonparametrically, from
37 which we derive ecological interactions (Bonnaffé, Sheldon, and Coulson 2021). More specif-
38 ically, the neural networks embedded in the ODEs learn nonparametrically the shape of the per
39 capita growth rate of the populations and its dependence on the state variables of the system (Bon-
40 naffé, Sheldon, and Coulson 2021). Combined with the Geber method (Hairston et al. 2005), we

not sure what this means.

41 are able to estimate the direction, strength, and degree of nonlinearity of interactions.

42 The main limitation of the approach lies in the computational cost of fitting the NODEs (Chen et al.

43 2019; Bonnaffé, Sheldon, and Coulson 2021). This is due to the fact that NODEs, just like ODEs,

44 need to be stimulated over the entire range of the time series in order to compute the likelihood

45 of the trajectories of the model. This can be avoided by using gradient matching, which requires

46 interpolating the time series, and fitting the ODEs directly to the interpolated dynamics (Jost and

47 Ellner 2000; Ellner, Seifu, and Smith 2002). Although a similar approach has been proposed (see

48 Treven et al. 2021), there are no implementations of this approach to fitting NODEs, in spite of its

49 great potential for cutting down computational costs.

do we
need to
elaborate
further?
what did
Treven do?

50 In addition, given the novelty of the framework, the accuracy and robustness of NODEs in estimat-

51 ing ecological interactions remain largely unexplored. Most of the work to date is concerned with

52 the accuracy of the fitted trajectories and of the forecasts (Mai, Shattuck, and O'Hern 2016; Treven

53 et al. 2021; Frank 2022), little attention is given to the shape of the processes that are producing

54 the dynamics approximated by NODEs (but see Hu et al. 2020). It is important to understand to

55 what extent the neural networks embedded within NODEs carry meaningful biological informa-

56 tion.

functional forms?

57 In this manuscript, we first introduce a novel fitting technique for NODEs, Bayesian neural gradient

58 matching. The method extends gradient matching by using neural networks to interpolate the

59 time series data instead of splines, and Bayesian regularisation to fit NODEs to the interpolated

60 dynamics (Cawley and Talbot 2007). This cuts down the fitting time of NODEs to only a few

seconds, compared to about 30 minutes in our previous work (Bonnaffé, Sheldon, and Coulson 2021), allowing for efficient cross validation, and uncertainty quantification. We then demonstrate that NODEs are highly accurate in recovering ecological interactions in an artificial three species prey predator system where ground truth is known. Finally, we conclude the work by characterising ecological interactions in three replicates of a real three species prey predator system with an algae, flagellate, and rotifer (Hiltunen et al. 2013), as well as in the classic ^alynx time series (Odum and Barrett 1972). We find that only main interactions, between the algae and the rotifer, are conserved across the three replicates, and not the interaction of the flagellate with the other species. We also find that in most cases linear interactions are sufficient to explain the dynamics apart from nonlinearity in the effect of the prey on the top predator in both the rotifer and lynx.

2 Material and Methods

2.1 Method overview

We provide a non-parametric 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). We provide a novel method, *Bayesian neural gradient matching*, which results in over 300 times faster NODEs fitting.

79 2.2 Neural ordinary differential equation

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

89 We first consider a general NODE system,

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

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

*did we
define them ??*

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

96 which feature a single layer, containing J neurons, that maps the inputs, here the state variables y ,
 97 to a single output, the dynamics of state variable i , dy_i/dt . The parameter vector θ_i contains the
 98 weights $\theta^{(l)}$ of the connections in the SLPs. SLPs can be viewed as weighted sums of activation
 99 functions f_σ , which are usually chosen to be *sigmoid* functions $f(x) = 1/(1 + \exp(-x))$. The link
 100 function f_λ allows to map the output of the network to a specific domain, for instance applying
 101 *tanh* will constrain the dynamics between -1 and 1, $dy_i/dt \in]-1, 1[$.
 102 ~~We would like to stress that this~~ ^{The} general form can be changed to represent biological constraints
 103 on the state variables. In particular for population dynamics, the state variables are strictly positive
 104 population densities, $y_i = N_i \in \mathcal{R}^+$. We could hence re-write equation (1) as, $dN_i/dt = f_p(N, \theta_i)N_i$,
 105 where the SLPs approximate the per-capita growth rate of the populations. More details regarding
 106 these models can be found in our previous work (Bonnaiffé, Sheldon, and Coulson 2021).

107 2.3 Fitting NODEs by Bayesian neural gradient matching

108 In this section, we describe how to estimate the parameters θ of the NODE system given a set of
 109 time series. Fitting NODEs can be highly computationally intensive, which hinders uncertainty
 110 quantification, cross-validation, and model selection (Bonnaiffé, Sheldon, and Coulson 2021). We
 111 solve this issue by introducing *Bayesian neural gradient matching* (BNGM), a computationally
 112 efficient approach to fit NODEs. The approach involves two steps (Fig. 1). First, we interpolate

113 the state variables and their dynamics with neural networks (Fig. 1, red boxes). Second, we train
 114 each NODE to satisfy the interpolated state and dynamics (Fig. 1, blue boxes). This bypasses the
 115 costly numerical integration of the NODE system and provides a fully mathematically tractable
 116 expression for the posterior distribution of the parameter vector θ . We coin the term BNGM to
 117 emphasise two important refinements of the standard gradient matching algorithm (Ellner, Seifu,
 118 and Smith 2002). The first is that we use neural networks as interpolation functions, and the
 119 second is that we use Bayesian regularisation to limit overfitting and estimate uncertainty around
 120 parameters (Cawley and Talbot 2007). *↳ define ?*

*should be
in
abstract
too*

121 Interpolating the time series

122 The first step is to interpolate the time series and differentiate it with respect to time in order
 123 to approximate the state and dynamics of the variables. We perform the interpolation via non-
 124 parametric regression of the interpolating functions on the time series data,

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

125 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
 126 interpolation function given the parameter vector ω_i , and $\varepsilon_{it}^{(o)}$ is the observation error between the
 127 observation and prediction. The interpolation function is chosen to be a neural network,

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

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

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

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

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

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

135 where $\varepsilon_{it}^{(p)}$ is the process error, namely the difference between the interpolated dynamics, $\partial \tilde{y}_i / \partial t$
 136 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
 137 boxes).

138 **Bayesian regularisation**

139 In the context of standard gradient matching, defining the observation model (Eq. 3) and process
 140 model (Eq. 6) would be sufficient to fit the NODE system (Eq. 1) to the time series via optimisation
 141 (Jost and Ellner 2000; Ellner, Seifu, and Smith 2002; Wu, Fukuhara, and Takeda 2005). We could

find the parameter vector ω_i and θ_i that minimise the sum of squared observation and process errors, $\varepsilon_{it}^{(o)}$ and $\varepsilon_{it}^{(p)}$ (Eq. 3 and 6). However, this approach is prone to overfitting, and does not provide estimates of uncertainty around model predictions. To account for this, we introduce Bayesian regularisation, which allows us to control for overfitting by constraining parameters with prior distributions (Cawley and Talbot 2007), and to root our interpretation of uncertainty in a Bayesian framework.

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

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

where P is the marginal posterior density, $\omega_i = \{\omega_{i1}, \omega_{i2}, \dots, \omega_{iK}\}$ is the observation parameter vector controlling the interpolation function, $Y_i = \{Y_{i1}, Y_{i2}, \dots, Y_{iJ}\}$ 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

160 in a supplementary file (Supplementary A).

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

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

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

170 This approach allows us to limit overfitting by adjusting the constraint on the parameters, which
 171 is controlled by the standard deviation of the parameter prior distributions, δ_{ik} (Cawley and Talbot
 172 2007; Bonnaffé, Sheldon, and Coulson 2021). We could set small values of δ to limit the degree
 173 of non-linearity in the response, or to eliminate specific variables from the model by constraining
 174 their parameters to be close to zero. We identify the appropriate degree of constraint δ_i on NODE
 175 parameters via cross-validation. We train the NODE model on the first half of the interpolated data
 176 and predict the remaining half. We repeat this process for increasing values of δ_i , until we find the
 177 value that maximises the log likelihood of the test data.

178 2.4 Inference and uncertainty quantification

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

188 2.5 Analysing NODEs

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

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

193 Knowing the direction, however, is not sufficient to determine the importance of a variable for the
194 dynamics of another. Given the same effects, a variable that fluctuates a lot will have a greater

195 impact on the dynamics of a focal variable, compared to a variable that remains quasi-constant. We
 196 hence compute the strength of the interaction by multiplying the dynamics of a variable y_j by its
 197 effect on the focal variable y_i , also known as the Geber method (Hairston et al. 2005),

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

198 To summarise results across the entire time series we can compute the mean effects e_{ij} by av-
 199 eraging e_{ijt} across all time steps, $e_{ij} = K^{-1} \sum_k e_{ijk}$, as well as the relative total contribution, c_{ij} ,
 200 of a variable to the dynamics of another by computing the relative sum of square contributions,
 201 $c_{ij} = \left(\sum_{ijk} c_{ijk}^2 \right)^{-1} \sum_t c_{ijt}^2$. By computing the direction and strength of interactions between all the
 202 variables in the system we can build dynamically informed ecological interaction networks (e.g.
 203 fig. 5). Other metrics can be computed by analysing the NODEs, such as equilibrium states, these
 204 are discussed in our previous work (Bonnaffé, Sheldon, and Coulson 2021).

205 **3 Case studies**

206 **3.1 Case study 1: artificial tri-trophic prey-predator oscillations**

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

211 3.1.1 System

212 We consider a tri-trophic ODE system consisting of a prey, an intermediate predator, and a top
 213 predator. The system is built on the real tri-trophic system featuring algae, flagellates, and rotifers,
 214 considered in case study 3 (Hiltunen et al. 2013),

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

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

221 We simulate a case of invasion, by introducing the top predator ~~from rare~~ *at a low density*, with a set of parameters
 222 that result in dampening prey-predator oscillations, namely $\alpha = 1$, $\beta = 2.5$, $\gamma = 1.5$, $\kappa = 3$, $\delta = \phi =$
 223 $\mu = \nu = 1$. We focus on the middle section of the time series, $t \in [20, 50]$, as in the initial section
 224 the ~~rotifer~~ *top* predator is rare, and in the later section populations have attained a fixed equilibrium
 225 point. The resulting time series are presented in figure 2.

226 3.1.2 NODE model

227 In order to learn non-parametrically the per-capita growth rate of each species, and to derive eco-
 228 logical interactions, we define a three-species NODE system,

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

229 where the per-capita growth rates r_R , r_G , and r_B are neural network functions of the density R , G , B
 230 of each species (function f_p , Eq. 2). We choose a combination of linear and exponential activation
 231 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
 232 simple linear model to a non-linear model by releasing the constraint on the exponential section of
 233 the neural network during cross-validation. The number of units in the hidden layer J is chosen to
 234 be 10, as this is a commonly used number for systems of that size (e.g. Wu, Fukuhara, and Takeda
 235 2005; Bonnaffé, Sheldon, and Coulson 2021).

236 3.1.3 Time series interpolation

237 We interpolate the time series using the neural network described in section 2.3 (Eq. 4). We set
 238 the number of neurons in the network to $J = 30$. We use sinusoid activation functions, $f_{\sigma}(x) =$
 239 $\sin(x)$, so that the weights $\omega_{ij}^{(1)}$, $\omega_{ij}^{(2)}$, and $\omega_{ij}^{(3)}$ control the amplitude, shift, and frequency of the
 240 oscillations in the time series, respectively. Given that the population densities are strictly positive

241 $R, G, B \in \mathcal{R}^+$, we use an exponential link function, $f_\lambda(x) = \exp(x)$. We then approximate the
 242 marginal posterior distribution of the interpolation parameters, and thereby of interpolated states
 243 and dynamics, by taking 100 samples from the log marginal posterior distribution (Eq. 7) via
 244 anchored ensembling. In practice, the high number of parameters in the neural network equation
 245 may impede the fit of the time series, especially for ~~small~~^{short} time series. We found that dividing
 246 the number of parameters K (Eq. 7) by the number of neurons in the network J (Eq. 2) yields
 247 consistent fitting results. Interpolated states and dynamics are presented in figure 2.

248 3.1.4 Fitting NODEs to the interpolated time series

249 We fit the NODE system to the interpolated time series. In practice, we fit the NODE to the expect-
 250 ation of the interpolated state and dynamics, $E(\tilde{y}_i)$ and $E(d\tilde{y}_i/dt)$, by averaging over all sampled
 251 interpolation parameters. An alternative approach could be to consider the interpolation that max-
 252 imises the log marginal posterior density, but this may decrease repeatability due to the difficulty of
 253 reliably identifying a global maximum. Averaging across multiple interpolations ensures an overall
 254 smoother and robust interpolation. In addition, we standardise the response and explanatory vari-
 255 ables with respect to their mean and standard deviation (i.e. $Z = (Y - \mu)/\sigma$). This is to facilitate
 256 the training of the NODE by equalizing the scale of the different parameters in the neural network.
 257 Then, we identify the optimal regularisation parameter δ (Eq. 8) by cross validation. To do that,
 258 we split the data in half and calculate the log likelihood of the test set for increasing values of δ ,
 259 from 0.05 (linear) to 0.5 (highly non-linear), by increments of 0.05. This allows us to identify
 260 the maximum degree of non-linearity, δ , in the per-capita growth rate that ensures generalisability

261 throughout the time series. Then, we approximate the posterior distribution of the NODE param-
262 eters by taking 30 samples from the posterior distribution (Eq. 8). Finally, we perform model
263 selection by removing variables that do not result in a significant decrease in the log likelihood
264 of the model (assessed by comparing log likelihood confidence intervals). We ensure moderate
265 temporal autocorrelation and normality by visualising the residuals of the models. We also ensure
266 results repeatability running the entire fitting process a second time.

267 **3.1.5 Computing ecological interactions**

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

277 **3.2 Case study 2: real tri-trophic prey-predator oscillations**

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

280 ecological interactions across virtually identical replicated time series.

281 **3.2.1 System**

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

293 **3.2.2 NODE analysis**

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

299 3.3 Case study 3: real di-trophic prey-predator oscillations

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

have
common
only one Ref -
doesn't suggest
common.

304 3.3.1 System

305 The system is described in details in our previous work (Bonnaffé, Sheldon, and Coulson 2021).
306 The data consist in a 90-year long time series of counts of hare and lynx pelts collected by trappers
307 in the Hudson bay area in Canada (Odum and Barrett 1972). The time series displays characteristic
308 10-year long prey-predator oscillations.

309 3.3.2 NODE analysis

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

312 4 Results

313 4.1 Model runtimes

314 We present a breakdown of the runtime of fitting NODEs by BNGM for each system in table
315 1. We find that it takes on average 5.35 minutes to fit NODEs by BNGM. This includes taking

316 390 samples, and thereby performing 390 full optimisations, of the posterior distribution of the
317 interpolation and NODE parameters. This amounts to about 5.37 second to sample each variable
318 of the NODE system once. This is a 335 fold improvement over our previous approach, which took
319 on average 30 minutes (Bonnaiffé, Sheldon, and Coulson 2021). *Very Impressive !*

320 4.2 Case study 1: artificial tri-trophic system

321 We present the results of fitting NODEs by BNGM to the artificial tri-trophic time series in figure
322 2 and 3. We find that both the interpolation of the state variables and dynamics are highly accurate
323 (Fig. 2), given that they closely match the ~~ground~~ truth, known from the equations of the ODE
324 model that we used to generate the time series (Eq. 11). Similarly, we find that the NODE approx-
325 imation of the per-capita growth rate of each species also closely matches the ground truth (Fig. 3,
326 a., d., g.). We find negative non-linear effects of the two predators on the growth rate of the algae
327 (Fig. 3, b., blue and purple lines). This non-linear pattern is mirrored by the effect of the algae
328 on the growth rate of the predators (Fig. 3, e. and h., red line). The linear interaction between
329 the two predators is also well-recovered (Fig. 3, e., blue line, and h., purple line). We find that
330 removing the intra-specific dependence in the growth rate of the predators did not affect the fit of
331 the model (Fig. e., purple line, and h., blue line). The BNGM approach ~~hence~~ recovers accurately
332 the dynamical characteristics of the artificial system.

333 4.3 Case study 2: real tri-trophic system

334 First, we present the in-depth analysis of the drivers of the dynamics of the algae, flagellate, and
335 rotifer population in replicate A (Fig. 4). Cross validation reveals that there is no support for non-
336 linear effects in the growth rate of the algae and flagellate for replicate A (Fig. 4, a. and b., d. and
337 e.). We find negative linear intra-specific density-dependence (Fig. 4, b., red line), and negative
338 linear inter-specific effects of the two predators (purple and blue line). We find that the growth rate
339 of the flagellate is virtually solely driven by predation by the rotifer (Fig. 4, e. and f., blue line).
340 The rotifer population itself is driven by a positive non-linear effect of both prey (Fig. h., red
341 and purple line). There is also evidence for positive non-linear intra-specific density-dependence
342 (Fig. h., blue line). Overall, comparing results across the three replicates reveals that the effect of
343 the rotifer population on the flagellate and algae, and the effect of the algae on the rotifer, are the
344 strongest and most consistent interactions (Fig. 5, table 2). The interactions of the flagellate with
345 the algae, and its effect on the rotifer population varies substantially (Fig. 5, table 2). Interestingly,
346 intra-specific density-dependence in rotifer and algae is also found to be inconsistent across the
347 three replicates.

prey-predator

348 4.4 Case study 3: real di-trophic system

349 Finally, we present the analysis of the drivers of the hare-lynx population dynamics in figure 6.
350 Cross-validation provides support for non-linear effects in the per-capita growth rate of the hare and
351 lynx. We find that the hare population growth rate is mostly determined by a non-linear negative

352 effect of the lynx population (Fig. 6, b. and c. blue line), and by weak non-linear positive density-
353 dependence (red line). The lynx growth rate is determined by a positive non-linear effect of the
354 hare (Fig. 6, e. and f., red line), and to a lesser extent by negative non-linear intra-specific density-
355 dependence (blue line).

356 5 Discussion

357 We provide a novel method for estimating ecological interactions nonparametrically, by using
358 neural ordinary differential equations (NODEs) fitted with Bayesian neural gradient matching
359 (BNGM). We demonstrate that ^{the}~~this~~ approach is fast, thanks to the gradient matching step, and
360 accurate, as NODEs approximate with minimal error the ecological interactions in artificial time
361 series. Finally, we conclude by estimating the strength, direction, importance, and nonlinearity of
362 ecological interactions in 4 real time series, showing variation in ecological interactions within and
363 across the time series.

364 Bayesian neural gradient matching

365 The Bayesian neural gradient matching (BNGM) approach that we ^{develop}~~propose here~~ extends standard
366 gradient matching, by using artificial neural networks as interpolating functions instead of splines
367 (Ellner, Seifu, and Smith 2002), and Bayesian regularisation to control the nonlinearity of the
368 processes (Cawley and Talbot 2007). This allows for accurately fitting NODEs within seconds,
369 making it potentially [?]~~one of the~~ most efficient current technique for fitting NODEs ^{available}~~(See also Treven~~
370 et al. 2021). The use of ANNs as interpolating functions sets it apart from the initial approach of

371 Ellner and colleagues, who used splines to interpolate the time series and approximate the ODEs
 372 (Ellner, Seifu, and Smith 2002). ANNs are more general and flexible than splines, as well as being
 373 easier to manipulate given that they are defined continuously on the state space and can easily
 374 handle multiple interactions between variables. Our approach is ^{related} ~~also similar~~ to that of Wu ~~and~~
 375 ~~et al.~~ ^{et al.} ~~colleagues~~, who use ANNs to approximate both the states and ODEs of prey-predator systems (Wu,
 376 Fukuhara, and Takeda 2005), as well as that of Treven and colleagues, who developed the Gaussian
 377 process equivalent (Treven et al. 2021). In both ^{approaches} ~~studies~~, they train the interpolation functions at
 378 the same time as the NODEs, as a way to constrain the interpolation to trajectories that can be
 379 achieved by the NODE system, and thereby introduce dynamical coupling between state variables.
 380 One of the ^{risks} ~~caveats~~ of this dynamical coupling is that ^{approaches} ~~misestimating~~ ^{ion of} one of the state variables of
 381 the model ^{ed} ~~will~~ bias the estimation of the states and dynamics of other variables. To avoid this, we
 382 ~~instead~~ fit each interpolation and NODE independently to each time series. In addition, this makes
 383 it possible to parallelise the code, resulting in potentially even faster computation. By making
 384 ⁻ ~~fitting~~ NODEs fast and robust, we ^{remove a} ~~break down~~ the main limitation of using NODEs, allowing
 385 for quick and extensive model comparison, cross-validation, and uncertainty quantification around
 386 estimates.

387 Accuracy of NODEs in estimating ecological interactions

388 Our approach relies on approximating population dynamics with NODEs and then computing their
 389 sensitivity to a change in the density of the different populations in the system. We demonstrate
 390 that NODEs ^{recovery} ~~recover~~ accurately the dynamics, strength, direction and nonlinearity of ecological

391 interactions in artificial tri-trophic prey-predator time series, where ^{the}~~ground~~ truth is known. In
392 particular, we find that the interactions between the algae and the two predators are nonlinear,
393 and thereby oscillate throughout the time series, which is consistent with the model, that features a
394 resistance to predation at high algal density. We also recover the linear interactions between the two
395 predators, which shows that the NODEs are sensitive enough to discriminate between linear and
396 nonlinear interactions within and across time series. To our knowledge, this is the first assessment
397 of the accuracy of NODEs in recovering interactions between variables from time series data, as
398 most of the work focuses on assessing the accuracy of the fitting and forecasting of time series (e.g.
399 Chen et al. 2019; Treven et al. 2021; Frank 2022).

400 **Ecological interactions in real prey-predator systems**

401 We further tested NODEs in a real setting, by inferring ecological interactions across three repli-
402 cated time series of a real tri-trophic system featuring an algae, flagellate, and rotifer population
403 (Hiltunen et al. 2013). Our approach reveals that only stronger interactions, namely the negative
404 effects of the rotifer top predator on the other species, and the positive effect of algae on the rotifer,
405 are conserved across the three replicated time series. We also find evidence for nonlinearity in
406 the dynamics of the rotifer, as the positive effect of the algae on rotifer growth oscillates through-
407 out the time series. This is consistent with the biology of the system, as the algae tends to form
408 anti-predation clumps at higher density, which would dampen the positive effect of algal density on
409 rotifer growth at high algal density (Yoshida et al. 2003; Hiltunen et al. 2013). We find it interesting
410 that the weaker interactions with the flagellate predator are not consistent across time series, given

411 that they are close to true replicates. This system is known to evolve rapidly, it is hence possible
412 that fast evolution of the different populations from the onset of the time series may have driven
413 the system onto different attractors (Yoshida et al. 2003; Yoshida et al. 2007; Hiltunen et al. 2013).
414 Additionally, stochasticity in population dynamics may have a similar effect (Dallas et al. 2021).
415 Disentangling these two sources of variation would require refining the modelling framework, for
416 instance by including explicitly evolution in the model (e.g. with the Price equation, Ellner, Geber,
417 and Hairston 2011), and by using neural stochastic differential equations (i.e. NSDEs, Rackauckas
418 et al. 2019) fitted with a particle filter. While these constitute interesting potential further devel-
419 opments, our method is a useful first step, ^{identifying} ~~pointing at~~ differences between the time series, and
420 demonstrating a reasonable amount of deterministic consistency in the dynamics, judging by the
421 cross-validation and fits.

422 We also analysed the hare-lynx time series (Odum and Barrett 1972), as it is a common benchmark
423 in the field of time series analysis, and provides a comparison point with our previous implemen-
424 tation of NODEs (Bonnaffé, Sheldon, and Coulson 2021). As in our previous study, we found
425 a predatory inter-specific interaction between lynx and hare, and negative intra-specific density-
426 dependence in the lynx. Evidence for positive density-dependence in the hare was more limited
427 than previously found. We also found stronger evidence for nonlinearity, as intra- and inter-specific
428 effects oscillated throughout the time series, as a result of density-dependence in the ecological
429 interactions. This difference with our previous study is due to the fact that our previous implemen-
430 tation of NODEs was based on simulating the full NODE system, and hence imposed dynamical

431 coupling between the variables. This dynamical coupling comes at a cost, if one variable is not
432 explained well by the model, it will bias the interactions and dynamics of other variables. Here,
433 the time series of lynx and hare are analysed independently, each state variable is interpolated
434 as closely as desired, its effects on the dynamics of others are hence even more robust to model
435 misspecification than before.

436 Overall, our approach provides a novel and powerful way of estimating interactions nonparamet-
437 rically from time series data. The benefit of using NODEs is that we ~~are not making~~ ^{make no} assumptions
438 about the nature of the ecological interactions that drive the dynamics of the species (Chen et al.
439 2019; Bonnaiffé, Sheldon, and Coulson 2021). Hence, we have a better chance at estimating the
440 actual value of the interaction, knowing that it is not subjected to potential incorrect model spec-
441 ifications (Jost and Ellner 2000; Ellner, Seifu, and Smith 2002; Wu, Fukuhara, and Takeda 2005;
442 Kendall et al. 2005; Adamson and Morozov 2013). This approach is similar to Sugihara’s maps
443 (S-maps, Sugihara et al. 2012), which estimate interactions in time series by approximating the
444 Jacobian matrix nonparametrically via locally linear approximations of the state space (Deyle et al.
445 2015). However, because S-maps are locally linear, they do not assume the existence of a latent
446 trajectory generated from an overarching model. This creates two caveats, the first being that they
447 are more sensitive to noise in the time series (Cenci and Saavedra 2018), the second being that
448 they have no theoretical grounding given that they are at heart linear functions defined piecewise
449 on the state space. NODEs remain in essence deterministic ODE models, assuming an overarching
450 model driving the populations through the entire state-space, which can hence incorporate para-

metric assumptions regarding the driving processes (Bonnaffé, Sheldon, and Coulson 2021). For instance, we model the per capita growth rate of populations explicitly in NODEs, while S-maps approximate the population-level growth. Overall, this makes NODEs more suitable than S-maps when it comes to fitting noisy data or exploring theory by testing specific assumptions.

Limits and prospects

Our approach allows for the quantification of nonlinearity in interactions, but it does not directly indicate what it is attributable to. For instance, we identify nonlinear density-dependence in the effect of the algae on the rotifer, but we do not know whether this is due to a change in the effect with algae density or rotifer density, or both. In order to disentangle these higher order effects we could compute the Hessian of the system, namely the second order derivative of the dynamics with respect to the different state variables. Though this procedure is simple mathematically, it would ~~produce~~ ^{result} in 27 second order effects to analyse for the simple 3 species system considered here. This type of analysis would get rapidly out of hand for larger systems. Further work should hence consider how to derive and present these higher order effects.

Our method also only accounts for observed variables, so that time series for all important variables are required. In practice, the dynamics of variables are often dependent on unobserved quantities. For instance, the dynamics of algae in the rotifer system are most likely coupled with that of nitrogen, for which no time series is available. ^(Ref?) One interesting prospect would ~~hence~~ be to incorporate unobserved/latent state variables into the NODE system (Dupont, Doucet, and Teh 2019; Zhang et al. 2019; Frank 2022). Careful thought has to be given there as ~~to~~ whether to use an ODE or

471 NODE for the latent state given that it is not constrained by observations.

472 ~~In addition, as stated before,~~ we fit NODEs independently to each time series. This means that
473 a dynamical system simulated with the fitted NODEs may not reproduce the observed trajec-
474 ries for the state variables. Though this is not a problem if we are only interested in understand-
475 ing the drivers of each time series, this could become problematic when it comes to forecasting.
476 ~~Hence,~~ other studies have taken the approach of fitting the interpolations and NODEs together
477 (Wu, Fukuhara, and Takeda 2005; Treven et al. 2021). However, to our knowledge, no studies have
478 provided a Bayesian framework to perform ^{simultaneously} ~~explicitly~~ the interpolation and fitting step together.
479 As a further development ~~to the present work,~~ we ~~hence~~ propose building a Bayesian framework
480 combining both the neural interpolation and NODE fitting, in order to generate interpolations that
481 are consistent with the dynamical model.

482 ~~Ultimately,~~ we consider ^{only} ~~here~~ NODEs, which are defined along the ^{time} ~~temporal~~ dimension ~~only~~. The
483 framework could easily be extended to any other dimension by considering partial differential
484 equations instead (Rackauckas et al. 2019). For instance, in ~~a~~ spatial ecology ~~context~~ we could
485 model the dynamics of populations along two additional spatial dimensions. In an evolutionary
486 context, we could model ~~explicitly~~ the dynamics of populations in the ^e ~~phenotypic~~ space, by adding
487 the phenotype of the population as an additional dimension. The BNGM method ^c ~~would~~ prove
488 instrumental in fitting these models, which are notoriously expensive to stimulate. ^(ref?)

489 Conclusion

490 We provide ~~here~~ a method, BNGM, which allows for fitting NODEs within a matter of seconds.

491 This is a crucial step for efficient model selection and uncertainty quantification in NODEs. We also
492 demonstrate that NODEs allow for accurate estimation of the direction, strength, and nonlinearity
493 of ecological interactions, in a system where ~~ground~~ truth is known. Finally we estimate ecological
494 interactions in real prey predator systems, showing that system dynamics are driven by a mixture
495 of linear and nonlinear interactions, of which only strong ones seem to be generalisable across time
496 series.

497 **Acknowledgments**

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

502 **Data accessibility**

503 All data and code will be made fully available at <https://github.com/WillemBonnaffe/xxx/xxx>.

504 **Statement of authorship**

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

References

- Adams, M. P. et al. (Apr. 2020). “Informing management decisions for ecological networks, using dynamic models calibrated to noisy time-series data”. In: *Ecology Letters* 23 (4), pp. 607–619.
- 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.
- Bonnaffé, W., S. Legendre, A. Danet, 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*, pp. 1–14.
- 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. V. D. 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.
- Brook, B. W. and C. J. A. Bradshaw (2006). “Strength of evidence for density dependence in abundance time series of 1198 species”. In: *Ecology* 87 (6), pp. 1445–1451.

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

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

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

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

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

539 Dallas, T., B. A. Melbourne, G. Legault, and A. Hastings (2021). “Initial abundance and stochas-
540 ticity influence competitive outcome in communities”. In: *Journal of Animal Ecology*, pp. 1–
541 26.

542 Deyle, E. R., R. M. May, S. B. Munch, and G. Sugihara (Jan. 2015). “Tracking and forecasting
543 ecosystem interactions in real time”. In: *Proceedings of the Royal Society B: Biological Sciences*
544 283, pp. 1–9.

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

546 Ellner, S. P., M. A. Geber, and N. G. J. Hairston (2011). “Does rapid evolution matter? Measuring
547 the rate of contemporary evolution and its impacts on ecological dynamics”. In: *Ecology Letters*
548 14 (6), pp. 603–614.

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

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

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

555 Hairston, N. G. J., S. P. Ellner, M. A. Geber, T. Yoshida, and J. A. Fox (2005). “Rapid evolution
556 and the convergence of ecological and evolutionary time”. In: *Ecology Letters* 8 (10), pp. 1114–
557 1127.

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

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

562 Jost, C. and S. P. Ellner (2000). “Testing for predator dependence in predator-prey dynamics: A
563 non-parametric approach”. In: *Proceedings of the Royal Society B: Biological Sciences* 267
564 (1453), pp. 1611–1620.

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

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

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

570 Mai, M., M. D. Shattuck, and C. S. O’Hern (2016). “Reconstruction of Ordinary Differential Equations
571 From Time Series Data”. In: *arXiv*, pp. 1–15.

572 Moe, S. J., A. B. Kristoffersen, R. H. Smith, and N. C. Stenseth (2005). “From patterns to processes
573 and back: Analysing density-dependent responses to an abiotic stressor by statistical and
574 mechanistic modelling”. In: *Proceedings of the Royal Society B: Biological Sciences* 272 (1577),
575 pp. 2133–2142.

576 Odum, E. P. and G. W. Barrett (1972). “Fundamentals of Ecology”. In: *The Journal of Wildlife
577 Management* 36 (4), p. 1372.

578 Pasquali, S. and C. Soresina (2018). “Estimation of the mortality rate functions from time series
579 field data in a stage-structured demographic model for *Lobesia botrana*”. In: *arXiv*, pp. 1–15.

580 Pearce, T., F. Leibfried, A. Brintrup, M. Zaki, and A. Neely (2018). “Uncertainty in Neural Networks:
581 Approximately Bayesian Ensembling”. In: *arXiv*, pp. 1–10.

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

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

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

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

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

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

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

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

602 Yoshida, T., L. E. Jones, S. P. Ellner, G. F. Fussmann, and N. G. J. Hairston (2003). “Rapid evo-
603 lution drives ecological dynamics in a predator – prey system”. In: *Nature* 424 (July), pp. 303–
604 306.

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

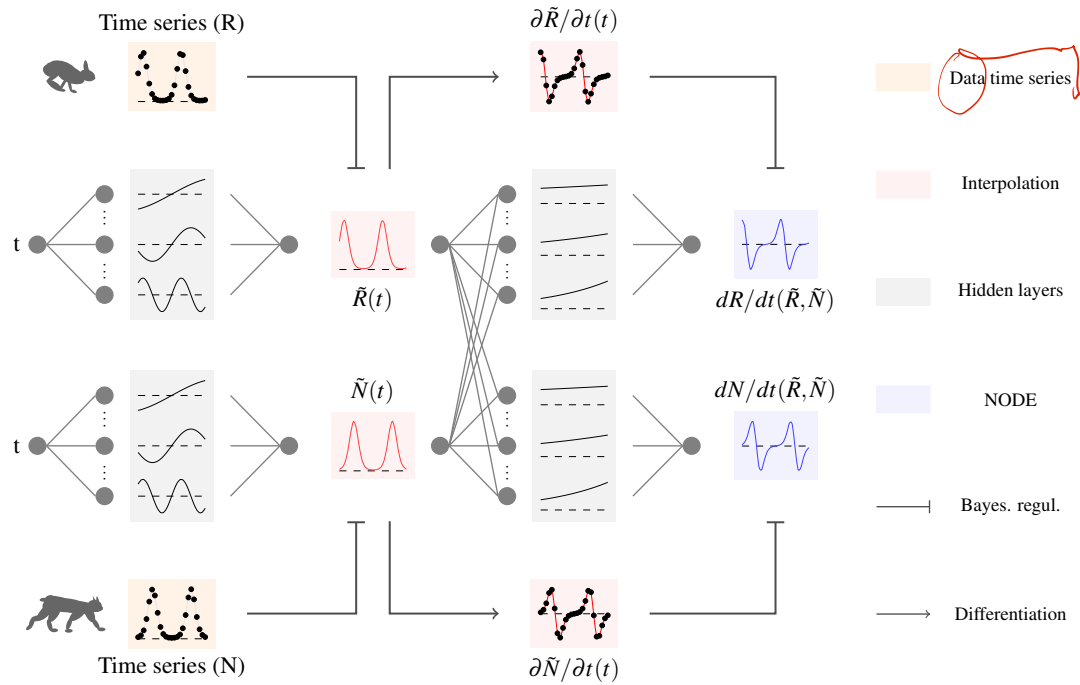
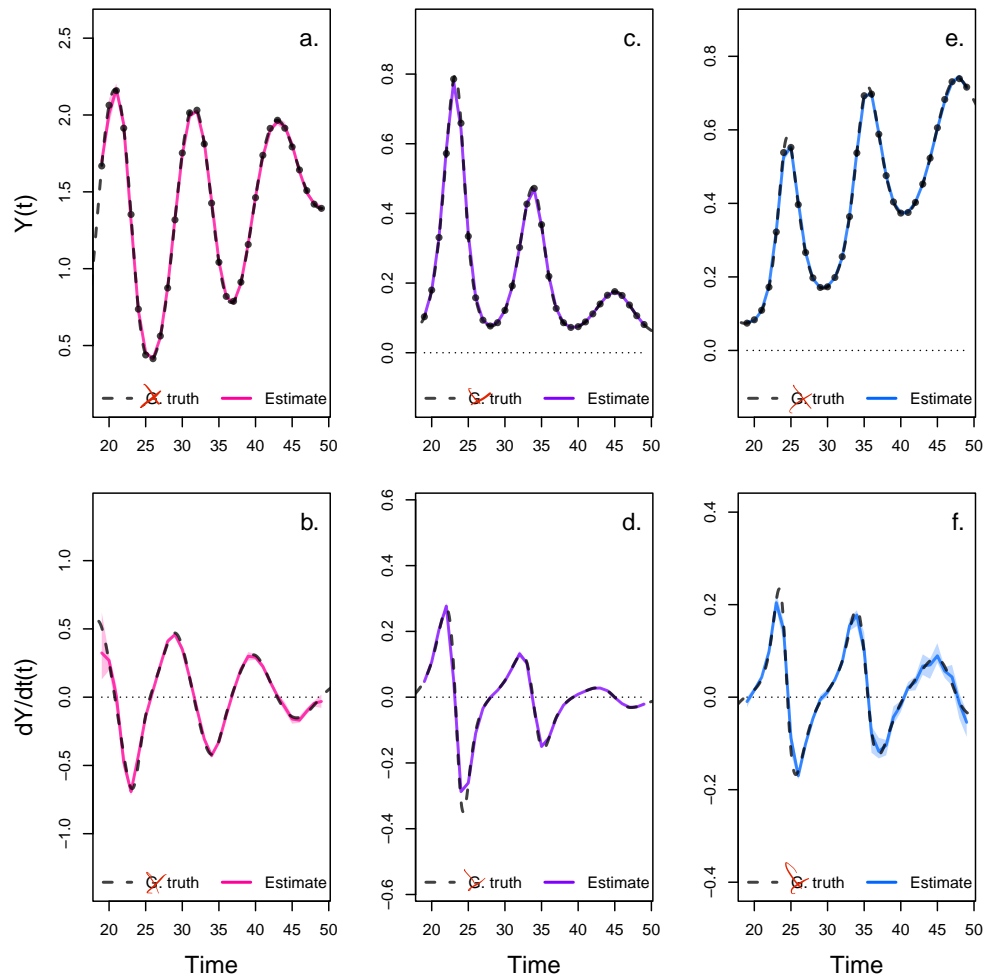


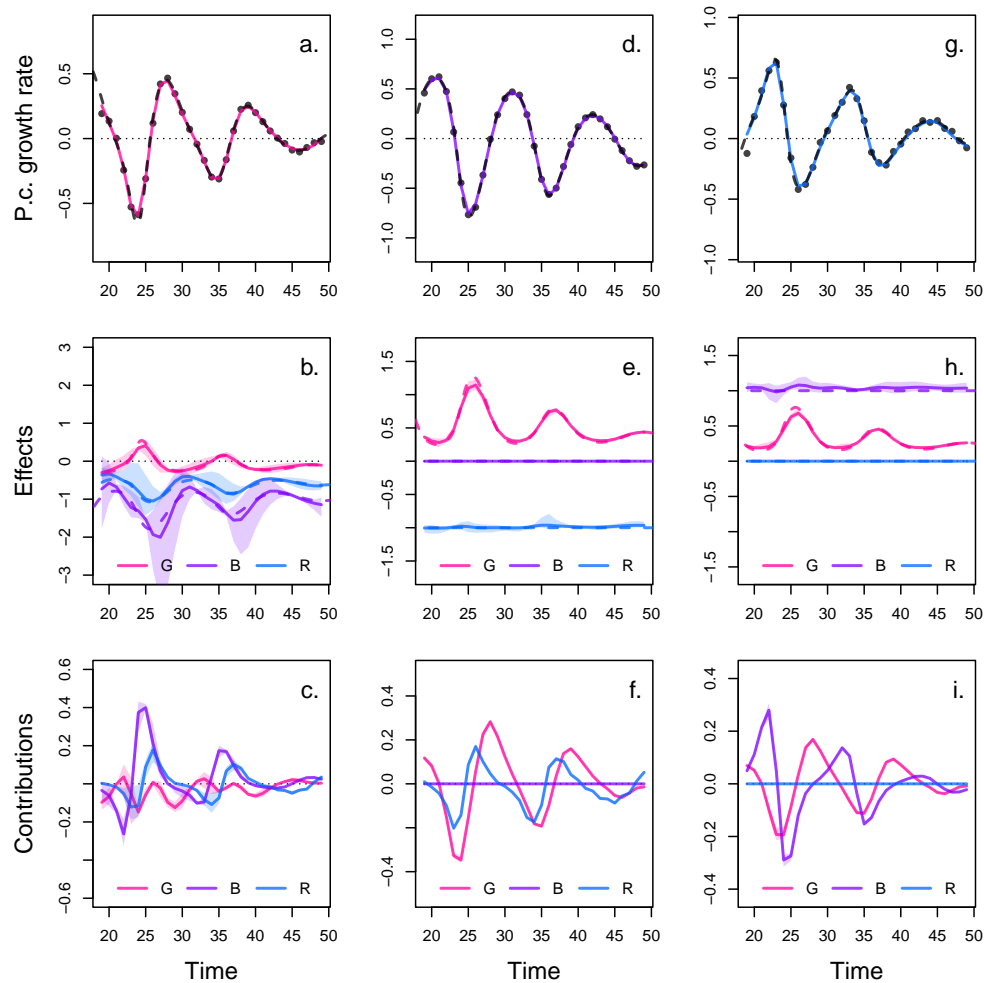
Figure 1: Overview of fitting neural ordinary differential equations (NODE) by Bayesian neural gradient matching (BNGM). In a first step we compute a continuous time approximation (interpolation) of each state variables, here the prey $\tilde{R}(t)$ and predator density $\tilde{N}(t)$. To do that we 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 5.37 seconds to fit NODEs by BNGM, compared to 30 mins in a previous study (Bonnaffé, Sheldon, and Coulson 2021), which corresponds to a 335 fold increase in speed.

Saved in results.



pred, intermediate for predators

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.



primary, intermediate and top predators.

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~~ 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.

Put similar
figures for
Reps B and
C in the SI?

cause for
concern?

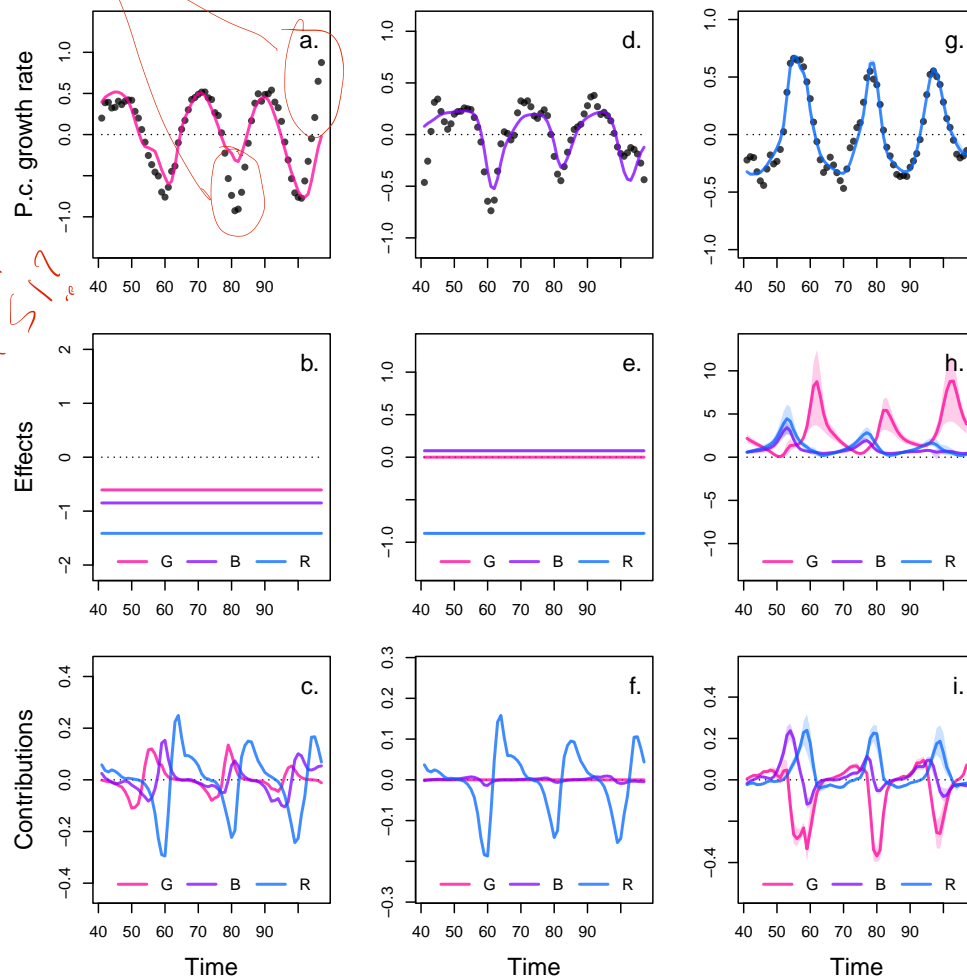


Figure 4: Drivers of dynamics of algae, flagellate, and rotifer in replicate A. This figure displays the NODE non-parametric approximations of the per-capita growth rate of algae (a., b., c.), flagellate (d., e., f.), and rotifer (g., h., i.). We obtain the NODE approximations (a., d., g., solid line) 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).

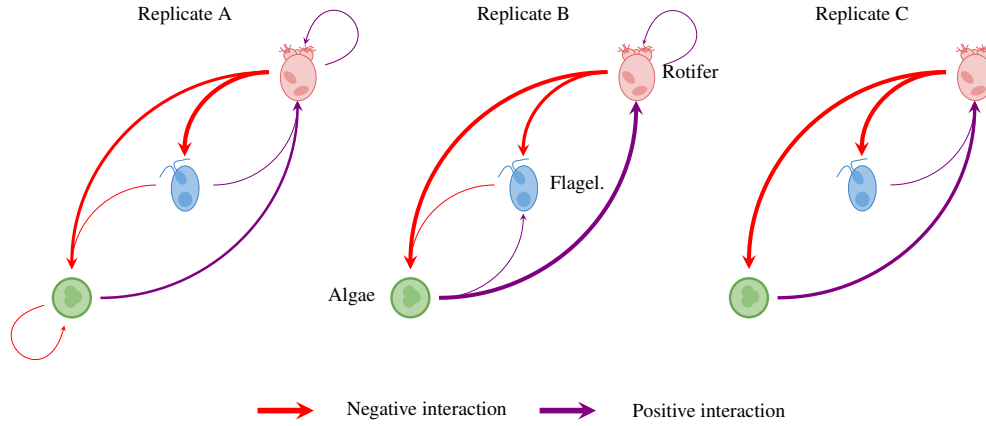


Figure 5: Interaction networks inferred from 3 replicated time series of algae, flagellate, and rotifers. This figure shows the direction and strength of ecological interactions inferred from 3 replicated sets of time series of algae, flagellate, and rotifer, using NODEs fitted by gradient matching. 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 2. The replicated time series were obtained by digitising the time series in Hiltunen et al. (2013).

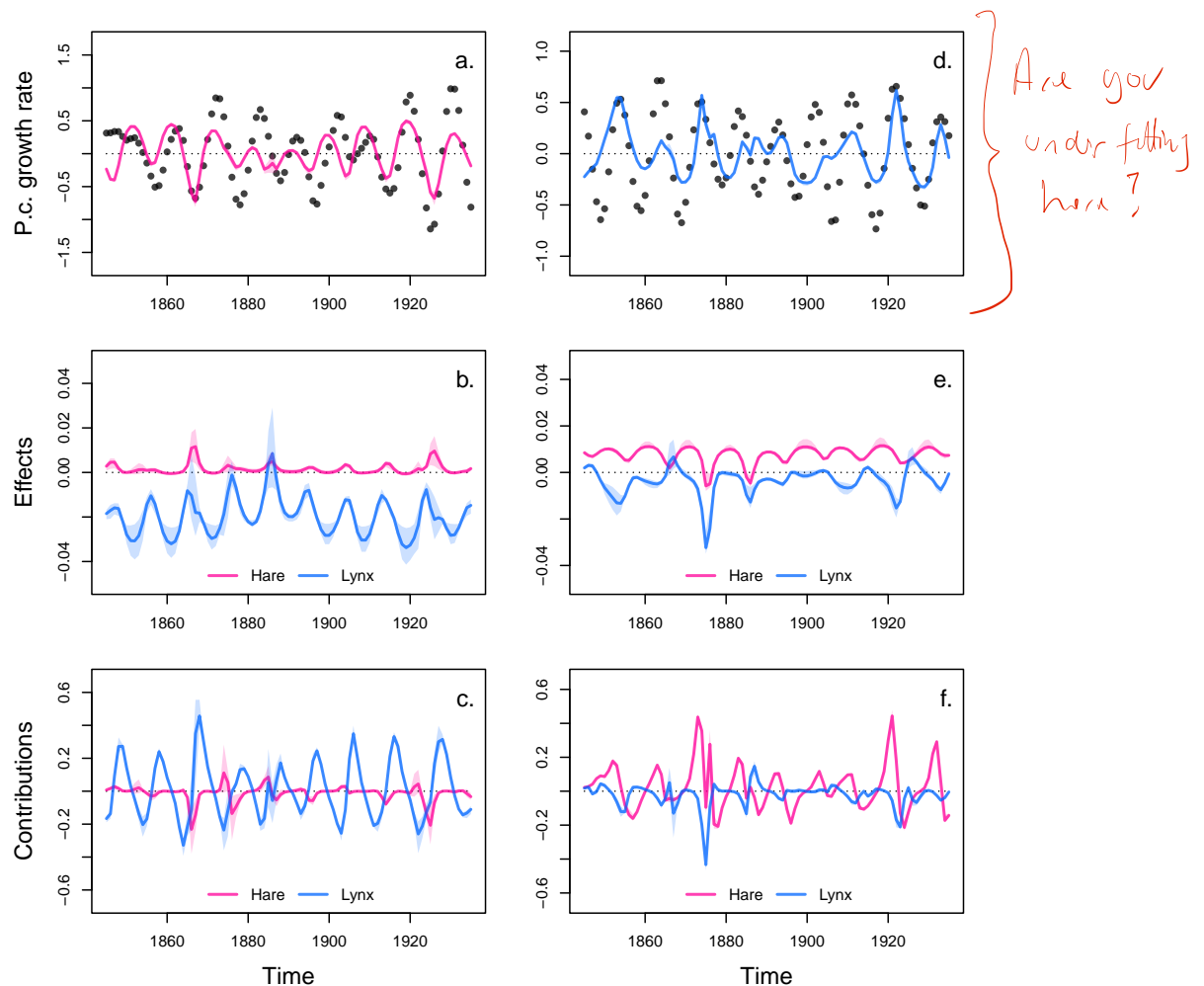


Figure 6: Drivers of dynamics of hare and lynx in the Odum and Barrett pelt count time series. This figure displays the NODE non-parametric approximations of the per-capita growth rate of hare (a., b., c.), and lynx (d., e., f.). We obtain the NODE approximations (a., d., solid line) by fitting the interpolated 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: Summary of model runtimes. We measured the time required to perform 100 interpolations and 30 NODE fits to each variable in the systems. Replicate A, B, and C correspond to each replicated time series of the aglae, flagellate, and rotifer tri-trophic system (Hiltunen et al. 2013). The Hare-Lynx system correspond to the 90 years long time series of hare and lynx pelt counts (Odum and Barrett 1972). The number of time steps (N steps) is given for each time series. The total time per fit is obtain by dividing the total time in seconds by the number of fits (i.e. 130). It takes on average 5.35 minutes for the 130 NODE fits NODE, which amounts to 5.37 seconds per sample taken. This is 335 times faster than the 30 minutes fitting times obtained in a previous study (Bonnaffé, Sheldon, and Coulson 2021). These results were obtained on a macbook pro M1 MAX 2022, in base R, with non-optimised code.

↓ version 7.

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

Could go in S.I.

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 non-parametric approximation of the pre-capita growth rate of the three species.

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

607 **6 Supplementary**

608 **A Bayesian regularisation**

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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