Inconsistent species interactions across replicated systems may hinder generalisation of dynamical processes

Willem Bonnaffé[1,2] and 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; ben.sheldon@zoo.ox.ac.uk; tim.coulson@zoo.ox.ac.uk

Running title: Inconsistent dynamics prevent generalisation

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;

Article type: *Letters*

Specifications: 149 words in abstract; 4957 words in text; 29 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; Tim Coulson led investigations, provided input for the manuscript, commented on the manuscript.

Abstract

Generalisation of dynamical processes across natural systems is difficult because of their complexity and unobserved variables. The hope is that generalisation may be achieved if we model adequately the complexity of systems, and observe them in sufficient detail. Yet, there is still limited support for this claim. 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

Dynamical processes, such as ecological interactions, drive population dynamics and generate dynamical patterns. Generalising dynamical processes across biological systems is hard. This is due to biological contingencies, stemming from the complexity of biological systems (esp. ecological interactions), differences in environmental contexts (i.e. unobserved variables), and observation error (De Meester et al. 2019). These contingencies alter dynamical processes by, for instance, modifying the strength of species interactions, which ultimately results in variation in dynamics across systems with the same species composition and population structure. This prevents the identification of global dynamical models, and therefore hinders generalisation, repeatability, and transfer of knowledge across systems and studies (Lawton 1999).

This problem has been repeatedly identified in natural systems. For instance, 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). 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). Attempts at identifying a single population dynamics model in two mesocosms have led to partial fits, as the model could not accomodate the dynamics of the two different systems (Demyanov, Wood, and Kedwards 2006). A growing body of evidence showing that generalisation of dynamical processes across similar systems often fails (Lawton 1999, e.g. Kendall et al. 2005; Ezard, Côté, and Pelletier

2009).

A burning question is to what extent we would be able to generalise dynamics across systems if we were to properly account for contingencies, by appropriately modelling the complexity of the system, the structure of the environment, and reduce observation errors. This is difficult to assess in practice, especially in a natural setting (Brunner et al. 2019; De Meester et al. 2019). However, it may be possible in an artificial setting, where contingencies can be virtually eliminated. 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; 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 consistency of dynamical processes in a simple multi-species system hosted in a controlled environment. We do this by quantifying the direction, strength, and consistency of interactions in time and across replicates of a simple biological system 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.

2 Material and Methods

2.1 Method overview

The aim of the study is to determine the extent to which the strength and direction of interactions between three species in a tri-trophic prey-predator system are consistent in time and across replicates, as a way to assess the generalisibility of dynamics across simple and well-observed systems. To do this we approximate the dynamics of each species by fitting neural ordinary differential equations (NODEs, Bonnaffé, Sheldon, and Coulson 2021) to replicated time series data of changes in prey and predator densities. We then derive the interactions between species by looking at the sensitivity of the dynamics to a change in the density of each species, and assess their consistency in time and across replicates.

2.2 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, the top predator also consumes the intermediate predator. 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, Fig. 1). The aim of their experiment was to determine which type of population dynamics would arise in a system with two predators competing for the same resource (the algae), where one predator (the rotifer) would also be able to consume its competitor (the flagellate). According to their expectations, they found prey-predator oscillations, where the lag between the density peaks of each species reflected their position in the food web. Namely that the peak of algae preceded the flagellate peak, which itself preceded the rotifer peak.

Their microcosms are close to true replicates in that environmental conditions, namely temperature, salinity, and nutrient influx, where maintained constant, and initial conditions, that is the initial density of each species were shared across all replicates. In spite of that, they still found evidence for algae evolution in some parts of the time series, which resulted in a shift of the dynamics from fast prey-predator cycles to slower oscillations, similar to those documented in previous studies on similar systems (Yoshida et al. 2003), even in lineages where genetic variation in predator defense traits was eliminated at the start of the experiment. Consequently, the time series that they reported

are the ones that did not present evidence of evolution, and therefore displayed purely ecological dynamics.

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.

2.3 Model specifications

The aim of the modelling approach is to infer the drivers of the dynamics of each species from the time series data. More specifically, we want to quantify the effect of a change in the density of one species on the dynamics of the other species. In this way we can understand which, and to what extent, species interactions drive population dynamics. To do this we use neural ordinary differential equation (NODEs), which is a novel methodology allowing us to infer dynamical processes non-parametrically from time series data (Bonnaffé, Sheldon, and Coulson 2021). We choose this methodology over traditional approaches because it offers two advantages. The first lies in the fact that NODEs approximate the dynamics of populations non-parametrically, and are therefore not subject to incorrect model specifications (Jost and Ellner 2000; Adamson and Morozov 2013). This is important as it offers an unbiased estimation of the inter-dependences between state variables, and hence a reliable assessment of whether a species is contributing to the dynamics

of another. The second advantage is that it is a dynamical systems approach, which means that the effects are estimated in a dynamically consistent system of ODEs (Bonnaffé, Sheldon, and Coulson 2021). This is useful because it accounts for the dynamical nature of the system, so that it includes lag effects, not just direct correlations between variables.

We define a simple NODE system for the three-species system described previously

$$\begin{aligned} \frac{dR}{dt} &= r_R(R,G,B,\beta_R)R \\ \frac{dG}{dt} &= r_G(R,G,B,\beta_G)G \\ \frac{dB}{dt} &= r_B(R,G,B,\beta_B)B \end{aligned} \tag{1}$$

where dR/dt, dG/dt, and dB/dt denote the change in rotifer (R), algae (G), and flagellate (B) density in continuous time. The per-capita growth rates r_R, r_G, and r_B are non-parametric functions of the density R, G, B of each species. The shapes of the non-parametric functions are controlled by the parameter vectors β_R, β_G, and β_B. Fitting the NODE system (1) amounts to finding the parameter vectors, and thereby the per-capita growth rates, that best describe the changes in density observed in the time series data.

Each non-parametric functions is an artificial neural network (ANN). ANNs are powerful mathematical objects that can be trained to approximate the shape of dynamical processes (Funahashi and Nakamura 1993). For the sake of simplicity, we consider the simplest form of an ANN which contains a single hidden layer, namely a single layer peceptron (SLP)

$$r_R = \sum_{i=1}^{N} \beta_i f_\sigma \left(\beta_{i0} + \beta_{i1} R + \beta_{i2} G + \beta_{i3} B\right) \tag{2}$$

which takes as input the density of each species R, G, and B, and output the corresponding per-capita growth rate. The parameter vector β_R, β_G, β_B, contain the weight of the connections in the ANNs. The SLP can be viewed as a weighted sum of basis functions f_σ of the state variables of the system. In this study we consider sigmoid basis functions, as they are commonly used and their capacity to approximate any continuous function is well established theoretically (Funahashi and Nakamura 1993). The number of units in the hidden layer N is chosen to be 10, as this is a commonly used number for systems of that size (e.g. Wu, Fukuhara, and Takeda 2005). More details regarding these models can be found in our previous work (Bonnaffé, Sheldon, and Coulson 2021).

2.4 Model fitting

This section describes how to recover the parameters β of the NODE system given the time series data at hand. In a previous study, we developed an approach to fit NODE systems to time series data (Bonnaffé, Sheldon, and Coulson 2021). The technique relied on simulating the NODE system over the times covered by the time series, and then computing the difference between the predictions of the NODE model, and the observations of the time series. The model is fitted to the time series by adjusting the parameter vectors until temporal dynamics of the state variables matched

the observations as closely as possible. There are two caveats with this approach that we solve in this study by opting for a different fitting approach. The first caveat is that the fitting process previously described is computationally expensive, because the NODE system has to be simulated over the entire range of the data at every step of the optimisation. Second, the simulation prevents the computation of gradients of the posterior distribution of the model, and thereby prevents the use of efficient gradient descent approaches. Overall, this makes it hard to avoid getting trapped in local maxima.

Instead, we propose an alternative fitting approach which relies on data interpolation to approximate populations state and dynamics. In this way we avoid the simulation step as the NODE system can then be directly compared to the interpolated dynamics. We proceed in two steps, (1) we interpolate the time series data to estimate the states and dynamics of each variable, and (2) we fit the NODE system directly to these estimated dynamics.

Interpolating the data

We interpolate the time series and differentiate it with respect to time in order to approximate the dynamics of the system. The interpolation is found via a non-parametric regression of the interpolating function on the time series data

$$Y(t_i) = \tilde{Y}(t_i, \Omega) + \varepsilon_i \tag{3}$$

where $Y(t_i)$ is the observed value of the variable at time t_i and $\tilde{Y}(t_i, \Omega)$ is the value predicted by

the interpolating function, up to an error $\varepsilon_i \sim \mathcal{N}(0, \sigma_1)$. In the present case, the variables are either the rotifer density R, algae density G, or flagellate density B, depending on which time series is interpolated. The interpolating function is chosen to be an SLP with sinusoid activation functions

$$\tilde{Y}(t, \Omega) = \sum_{i=1}^{N} \omega_{i0} sin\left(\pi(\omega_{i1} + \omega_{i2} t)\right) \tag{4}$$

where $\tilde{Y}(t, \Omega)$ is the interpolated state variable, either R, G, or B, and is determined to be a weighted sum of sinusoid functions of time. The interpolation parameter vector Ω contains the weights ω_{i0}, ω_{i1}, and ω_{i2} which control the amplitude, shift, and frequency of the oscillations in the time series, respectively. We found sinusoid activation functions to be most efficient for interpolating population dynamics compared to other functions (such as sigmoid, hyperbolic). Following this approach we obtain directly an approximation of the dynamics of the state variable by differentiating the SLP with respect to time

$$\frac{\partial}{\partial t}\tilde{Y}(t) = \sum_{i=1}^{N} \omega_{i0} \pi \omega_{i2} cos\left(\pi(\omega_{i1} + \omega_{i2} t)\right) \tag{5}$$

as well as an analytical expression of the interpolated per-capita growth rate of the populations, by combining equation (4) and (5)

$$\tilde{r}_Y = \frac{1}{\tilde{Y}} \frac{\partial \tilde{Y}}{\partial t} \tag{6}$$

Overall, interpolating the data amounts to finding the parameter vector Ω that minimises the error in equation (3).

Fitting NODEs to the interpolated data

The second step is to match the NODE system to the interpolated dynamics, given the interpolated state variables. Thanks to the interpolation of the data, this simply amounts to performing a regression of the non-parametric approximation of the per-capita growth rate (equation 2) on the interpolated per-capita growth rate (equation 6)

$$\tilde{r}_Y(t_i) = r_Y(t_i, \beta) + \eta_i \tag{7}$$

up to an error term $\eta_i \sim \mathcal{N}(0, \sigma_2)$. For instance, the per-capita growth rate of the rotifer writes as $\tilde{r}_R(t_i) = r_R\left(\tilde{R}(t_i), \tilde{G}(t_i), \tilde{B}(t_i), \beta_R\right) + \eta_i$. Fitting the NODE per-capita growth rate hence amounts to finding the parameter vector β that minimises the error in equation 7, given the interpolation.

Statistical modelling approach

The following section describes how to recover the parameters for the interpolation and NODE system that best describe the time series, while controlling for overfitting. 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

$$p(\theta|\mathscr{D}) \propto p(\mathscr{D}|\theta)p(\theta) \tag{8}$$

where θ is the parameter vector of the model, and $\mathscr{D}$ the evidence, namely the data that the model is fitted to. In the case of the interpolation, the evidence is the population densities, either $R(t)$, $G(t)$, or $B(t)$, and the parameters are the weights Ω in the sinusoid SLPs. In the case of fitting the NODE model to the interpolated data, the evidence is the interpolated per-capita growth rate of each population, either $\tilde{r}_R$, $\tilde{r}_G$, or $\tilde{r}_B$, and the parameters are the weights β_R, β_G, and β_B in the non-parametric per-capita growth rates r_R, r_G, and r_B.

Assuming a normal likelihood for the residuals given the evidence we get

$$p(\mathscr{D}|\theta) = \prod_{i=1}^{I} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{e_i(\mathscr{D},\theta)^2}{2\sigma^2}\right\} \tag{9}$$

where $e_i(\mathscr{D},\theta)$ are the residuals of the model given the parameters. In the case of the interpolation, the residuals are $Y(t_i) - \tilde{Y}(t_i)$, and in the case of the NODE model, $\tilde{r}_Y(t_i) - r_Y(t_i)$. The dispersion term σ in the likelihood is measured by the parameters σ_1 in the case of the interpolation, and σ_2 in the case of the NODE fitting. 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

$$p(\theta) = \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi\delta^2}} \exp\left\{-\frac{\theta^2}{2\delta^2}\right\} \tag{10}$$

where J is the number of parameters in the models. The parameter δ controls the dispersion of the priors, and thereby the complexity/level of constraint of the model. Low values of δ will lead to underfitting, as it constrains the model to be simple, while high values of δ will lead to overfitting, by allowing for more complex shapes.

There is no standard approach for choosing δ. To account for overfitting, we opt for a regularisation approach by optimising 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

$$\log P(\theta|\mathscr{D}) \propto \frac{I}{2}\log\left(1+\sum_{i=1}^{I} e_i(\mathscr{D},\theta)^2\right) + \frac{J}{2}\log\left(1+\sum_{j=1}^{J}\theta_j^2\right) \tag{11}$$

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).

Finally, we estimate uncertainity in parameter values through anchor sampling, which produces

approximate Bayesian estimates of the posterior distribution of the parameters (Pearce et al. 2018). The technique is simple in that it requires sampling a parameter vector from the prior distributions, and then optimising the posterior distribution from this starting point. By repeatedly taking samples, the sampled distribution approaches the posterior distribution and provides estimates and error around the quantities that can be derived from the models. The expectation of the quantities can then be approached by computing the mean 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 Model analysis

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 other. 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 comes down to multiplying the dynamics of a variable by its effects of the other variables. We further compute the importance of a species in driving the dynamics of another by computing its relative contribution compared to other species at each time step. More details on how to recover these quantities can be found in our previous study (Bonnaffé, Sheldon, and Coulson 2021).

3 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.

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 postive 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%, 43%, 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. 1, 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.

4 Discussion

Our ability to generalise dynamical processes and patterns across populations and communities is limited by the complexity of the dynamics, differences in environments, and incomplete and/or erroneous observations. Yet 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, thus featuring limited observation error. 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. Futhermore, 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.

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, 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 (Hiltunen et al. 2013). In their original paper the authors limited the initial genotypic 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 cannot 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. 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, other driving factors could be demographic stochasticity as is it often regarded as a driver of differences across systems (Dallas et al. 2021). Yet, we find that the dynamics of the three species are well-explained by relatively simple deterministic effects between the state variables, which means that that though dynamical processes differ across replicates, they are consistent in time within each system. This suggests that stochasticity plays a minor role in driving the system. 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). For instance, this study already provides a more accurate and reliable 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. 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 suggestions on early versions of the work. The work was supported by the Oxford-Oxitec scholarship and the NERC DTP.

Data accessibility

All data and code will be made fully available at https://github.com/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: *Proc. R. Soc. A Math. Phys. Eng. Sci.* 469.2149, pp. 1–19.

Bonnaffé, Willem, Ben C. Sheldon, and Tim Coulson (2021). "Neural ordinary differential equations for ecological and evolutionary time series analysis". In: *Methods Ecol. Evol.* 2, pp. 1–46.

Bonnaffé, Willem, Stéphane Legendre, Alain Danet, and Eric 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.

Bonsall, M B, E. Van Der Meijden, and M J Crawley (2003). "Contrasting dynamics in the same plant-herbivore interaction". In: *Proc. Natl. Acad. Sci. U. S. A.* 100.25, pp. 14932–14936.

Bruijning, Marjolein, Eelke Jongejans, and Martin M. Turcotte (2019). "Demographic responses underlying eco-evolutionary dynamics as revealed with inverse modelling". In: *J. Anim. Ecol.* 88.5, pp. 768–779.

Brunner, Franziska S., Jacques A. Deere, Martijn Egas, Christophe Eizaguirre, and Joost A.M. Raeymaekers (2019). "The diversity of eco-evolutionary dynamics: Comparing the feedbacks between ecology and evolution across scales". In: *Funct. Ecol.* 33.1, pp. 7–12.

Cawley, Gavin C. and Nicola L.C. Talbot (2007). "Preventing over-fitting during model selection via bayesian regularisation of the hyper-parameters". In: *J. Mach. Learn. Res.* 8, pp. 841–861.

Dallas, Tad, Brett A. Melbourne, Geoffrey Legault, and Alan Hastings (2021). “Initial abundance and stochasticity influence competitive outcome in communities”. In: *J. Anim. Ecol.* Pp. 1–26.

De Meester, Luc et al. (2019). “Analysing eco-evolutionary dynamicsThe challenging complexity of the real world”. In: *Funct. Ecol.* 33.1, pp. 43–59.

Demyanov, V., S. N. Wood, and T. J. Kedwards (2006). “Improving ecological impact assessment by statistical data synthesis using process-based models”. In: *J. R. Stat. Soc. Ser. C Appl. Stat.* 55.1, pp. 41–62.

Ellner, Stephen P., Yodit Seifu, and Robert H. Smith (2002). “Fitting Population Dynamic Models to Time-Series Data by Gradient Matching”. In: *Ecology* 83.8, p. 2256.

Ezard, Thomas H.G., Steeve D. Côté, and Fanie Pelletier (2009). “Eco-evolutionary dynamics: Disentangling phenotypic, environmental and population fluctuations”. In: *Philos. Trans. R. Soc. B Biol. Sci.* 364.1523, pp. 1491–1498.

Funahashi, Ken ichi and Yuichi Nakamura (1993). “Approximation of dynamical systems by continuous time recurrent neural networks”. In: *Neural Networks* 6.6, pp. 801–806.

Fussmann, Gregor F. and Bernd Blasius (2005). “Community response to enrichment is highly sensitive to model structure”. In: *Biol. Lett.* 1.1, pp. 9–12.

Gamelon, Marlène et al. (2019). “Accounting for interspecific competition and age structure in demographic analyses of density dependence improves predictions of fluctuations in population size”. In: *Ecol. Lett.* 22.5, pp. 797–806.

Gross, Kevin, Anthony R. Ives, and Erik V. Nordheim (2005). “Estimating fluctuating vital rates from time-series data: A case study of aphid biocontrol”. In: *Ecology* 86.3, pp. 740–752.

Hairston, Nelson G., Stephen P. Ellner, Monica A. Geber, Takehito Yoshida, and Jennifer A. Fox (2005). "Rapid evolution and the convergence of ecological and evolutionary time". In: *Ecol. Lett.* 8.10, pp. 1114–1127.

Hiltunen, Teppo, L. E. Jones, S. P. Ellner, and Jr G. Hairston Nelson G. (2013). "Temporal dynamics of a simple community with intraguild predation: an experimental test". In: *Ecology* 94.4, pp. 773–779.

Jost, C. and S. P. Ellner (2000). "Testing for predator dependence in predator-prey dynamics: A non-parametric approach". In: *Proc. R. Soc. B Biol. Sci.* 267.1453, pp. 1611–1620.

Kendall, Bruce E. et al. (2005). "Population cycles in the pine looper moth: Dynamical tests of mechanistic hypotheses". In: *Ecol. Monogr.* 75.2, pp. 259–276.

Lawton, John H (1999). "Nordic Society Oikos Are There General Laws in Ecology ?" In: *Oikos* 84.2, pp. 177–192.

Pearce, Tim, Felix Leibfried, Alexandra Brintrup, Mohamed Zaki, and Andy Neely (2018). "Uncertainty in Neural Networks: Approximately Bayesian Ensembling". In: *arXiv*, pp. 1–10.

Pontarp, Mikael, Åke Brännström, and Owen L. Petchey (2019). "Inferring community assembly processes from macroscopic patterns using dynamic eco-evolutionary models and Approximate Bayesian Computation (ABC)". In: *Methods Ecol. Evol.* 10.4, pp. 450–460.

Posey, Martin H., Troy D. Alphin, and Lawrence Cahoon (2006). "Benthic community responses to nutrient enrichment and predator exclusion: Influence of background nutrient concentrations and interactive effects". In: *J. Exp. Mar. Bio. Ecol.* 330.1, pp. 105–118.

Rosenbaum, Benjamin, Michael Raatz, Guntram Weithoff, Gregor F. Fussmann, and Ursula Gaedke (2019). “Estimating parameters from multiple time series of population dynamics using bayesian inference”. In: *Front. Ecol. Evol.* 6.234, pp. 1–14.

Shurin, Jonathan B., Jessica L. Clasen, Hamish S. Greig, Pavel Kratina, and Patrick L. Thompson (2012). “Warming shifts top-down and bottom-up control of pond food web structure and function.” In: *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 367.1605, pp. 3008–17.

Wu, Jun, Makoto Fukuhara, and Tatsuoki Takeda (2005). “Parameter estimation of an ecological system by a neural network with residual minimization training”. In: *Ecol. Modell.* 189.3-4, pp. 289–304.

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

Yoshida, Takehito, Stephen P. Ellner, Laura E. Jones, Brendan J.M. Bohannan, Richard E. Lenski, and Nelson G. Hairston (2007). “Cryptic population dynamics: Rapid evolution masks trophic interactions”. In: *PLoS Biol.* 5.9, pp. 1868–1879.

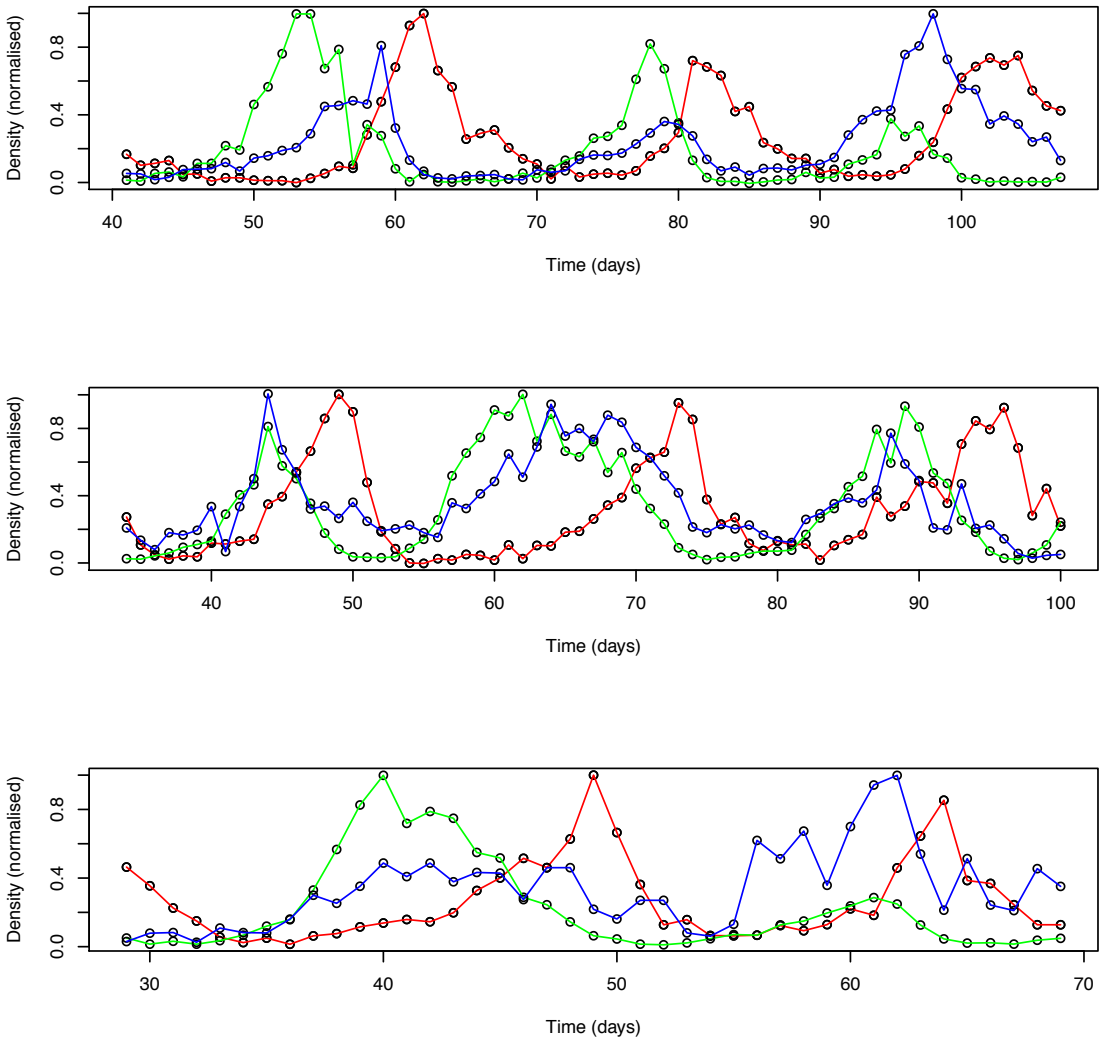

Figure 1: Time series of the prey algae (green), flagellate intermediate predator (blue), and top predator (red). Each time series correspond to one out of three replicates. Black open dots correspond to observed densities, scaled between 0 and 1. Time series were obtained from the study by Hiltunen and colleagues (Hiltunen et al. 2013).

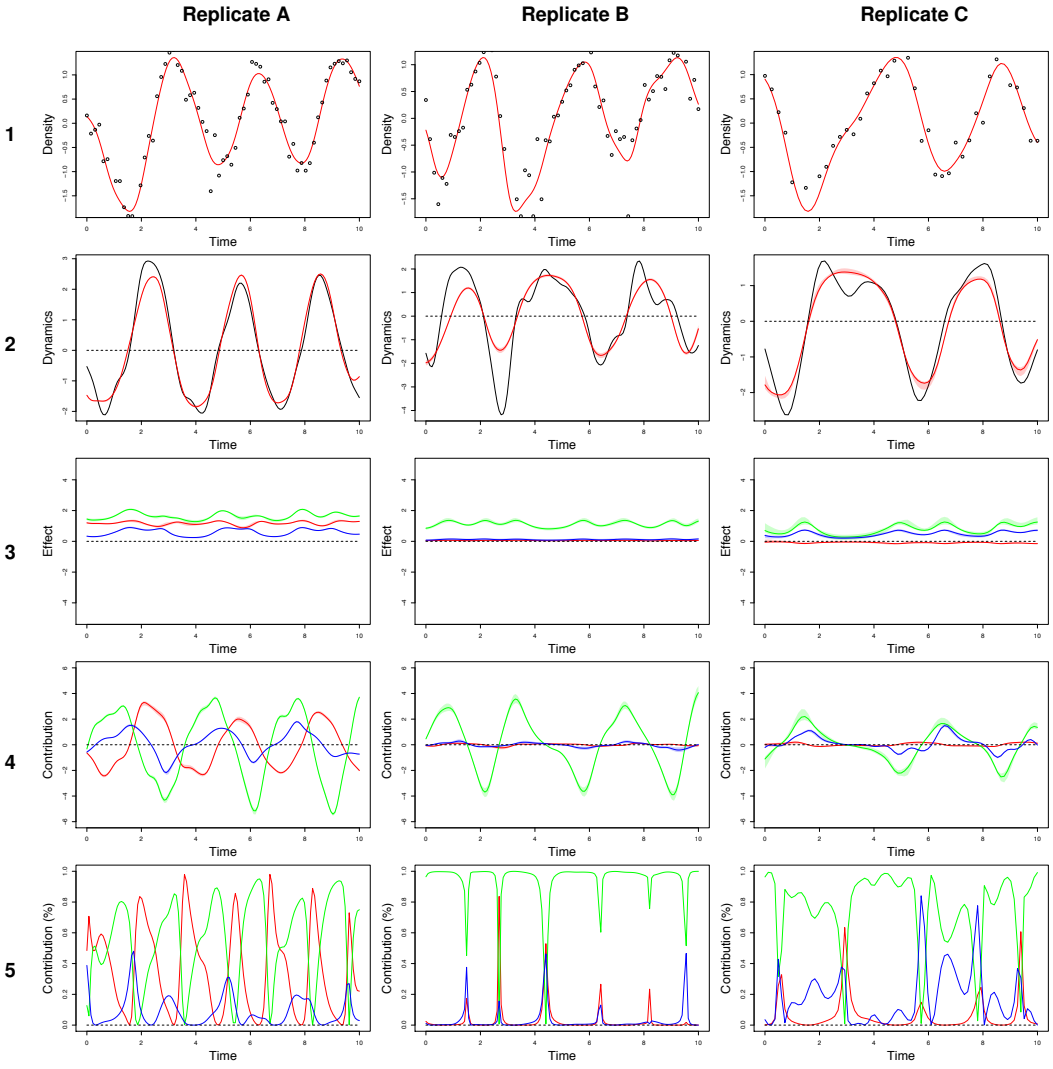

Figure 2: Drivers of rotifer dynamics. Row (1) presents the log density of rotifers in time, the dots are the observed densities while the red line is the interpolated data. Row (2) presents the change in log density at each time step, where the black line is the per-capita growth rate derived from the interpolated data and the red line is the NODE approximation. Row (3) presents the effect of each species on the dynamics/growth of rotifers, obtained by computing the sensitivity of the NODE approximation of the rotifer per-capita growth rate with respect to each species density. Row (4) presents the contribution of each species to the growth of the rotifer, obtained by multiplying its effect by its net change (i.e. row 2 and 3). Row (5) is the same as fourth but expressed as % of total change at each time step explained by each species. For all figures, green, blue, and red refer respectively to algae, flagellate, and rotifer.

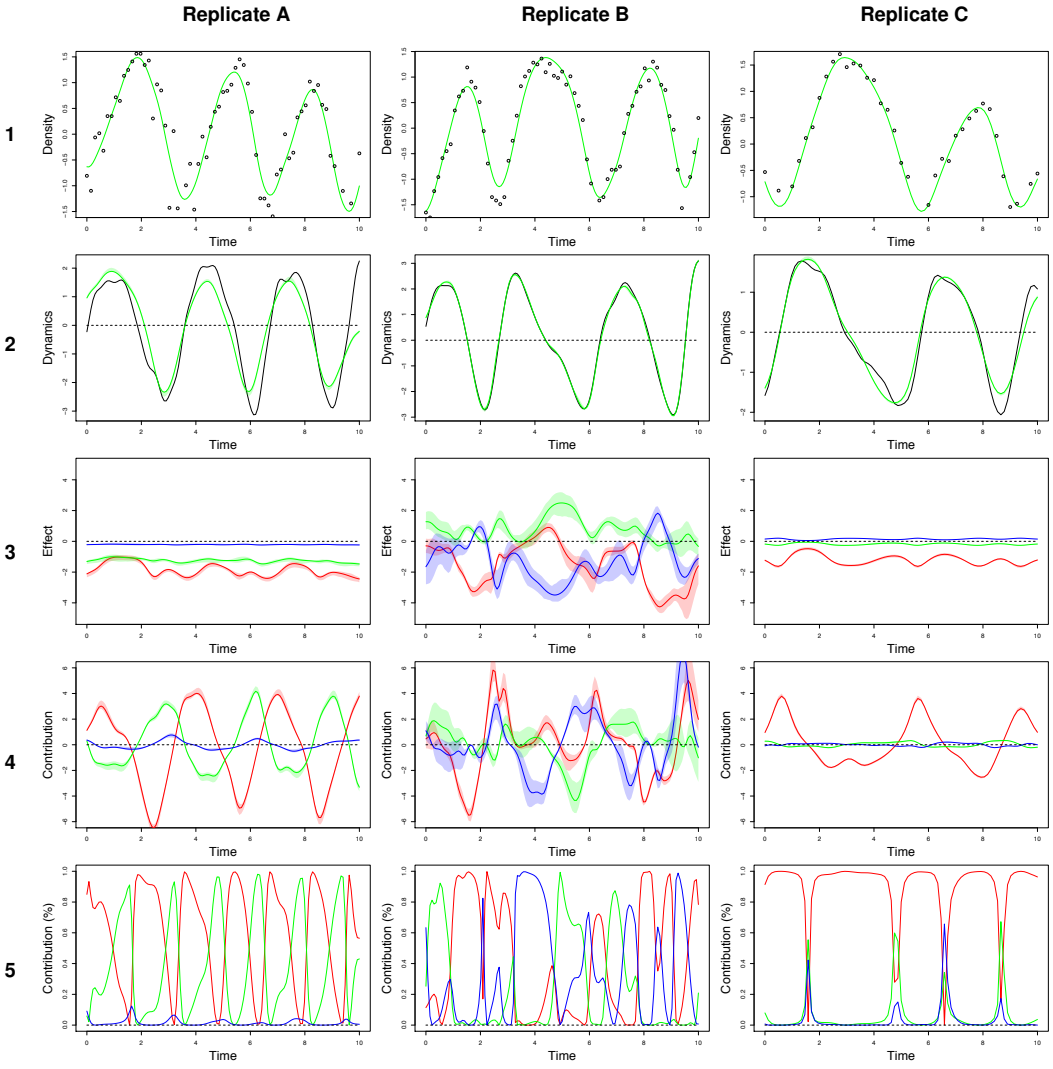

Figure 3: Drivers of algae dynamics. Row (1) presents the log density of algae in time, the dots are the observed densities while the green line is the interpolated data. Row (2) presents the change in log density at each time step, where the black line is the per-capita growth rate derived from the interpolated data and the green line is the NODE approximation. Row (3) presents the effect of each species on the dynamics/growth of algae, obtained by computing the sensitivity of the NODE approximation of the algae per-capita growth rate with respect to each species density. Row (4) presents the contribution of each species to the growth of the algae, obtained by multiplying its effect by its net change (i.e. row 2 and 3). Row (5) is the same as fourth but expressed as % of total change at each time step explained by each species. For all figures, green, blue, and red refer respectively to algae, flagellate, and rotifer.

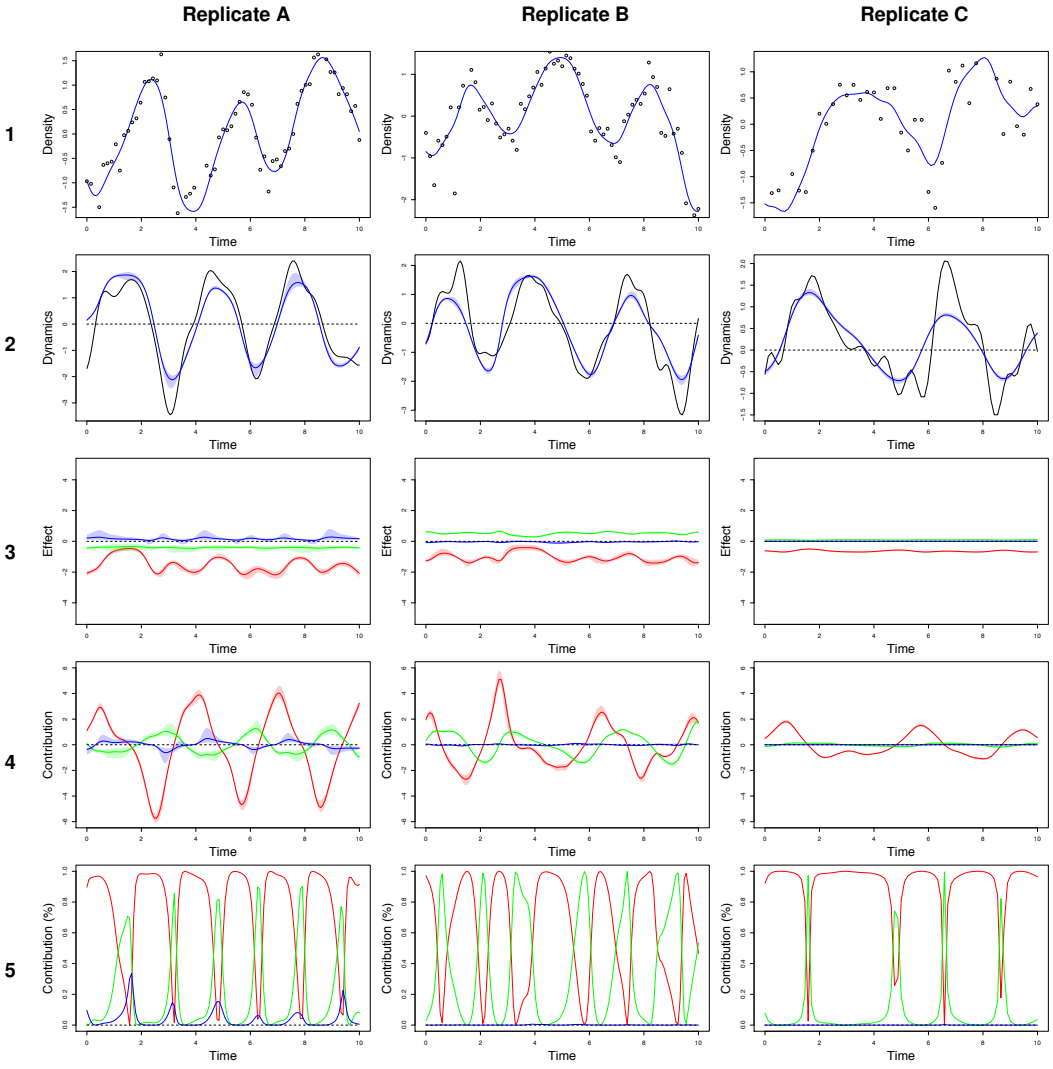

Figure 4: Drivers of flagellate dynamics. Row (1) presents the log density of flagellate in time, the dots are the observed densities while the blue line is the interpolated data. Row (2) presents the change in log density at each time step, where the black line is the per-capita growth rate derived from the interpolated data and the blue line is the NODE approximation. Row (3) presents the effect of each species on the dynamics/growth of flagellate, obtained by computing the sensitivity of the NODE approximation of the algae per-capita growth rate with respect to each species density. Row (4) presents the contribution of each species to the growth of the flagellate, obtained by multiplying its effect by its net change (i.e. row 2 and 3). Row (5) is the same as fourth but expressed as % of total change at each time step explained by each species. For all figures, green, blue, and red refer respectively to algae, flagellate, and rotifer.

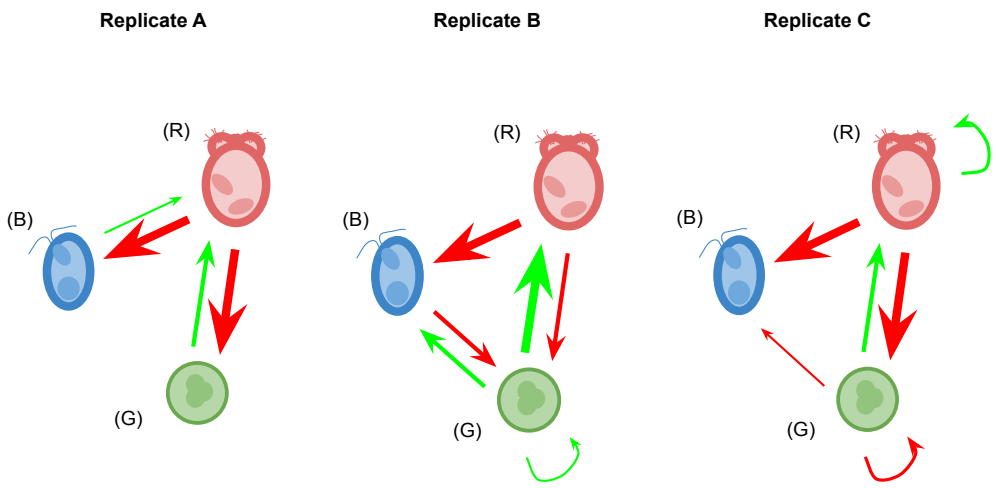

Figure 5: Interaction network inferred from the NODE analysis of the time series. The nodes G, R, B correspond to the algae, rotifer, and flagellate, respectively. Red and green arrows correspond to negative or positive mean effects, estimated by averaging the gradient of the per-capita growth rate of each species with respect to the density of the others across the time series. The width of the arrow is proportional to the % of total contributions of a variable on another, 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.

Table 1: Summary analysis. r^2 corresponds to the r squared of the NODE non-parametric approximation of the pre-capita growth rate compared to the interpolated per-capita growth rate for each of the three species. 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.

			R	G	B
Replicate A	r^2		0.968	0.839	0.831
	Mean effects	on R	1.180	1.723	0.632
		on G	-1.888	-1.286	-0.208
		on B	-1.476	-0.371	0.131
	% of total contributions	to R	0.356	0.552	0.091
		to G	0.583	0.399	0.016
		to B	0.785	0.194	0.020
Replicate B	r^2		0.744	0.997	0.765
	Mean effects	on R	0.058	1.146	0.113
		on G	-1.505	0.707	-1.287
		on B	-1.000	0.505	-0.046
	% of total contributions	to R	0.035	0.934	0.029
		to G	0.439	0.242	0.317
		to B	0.624	0.373	0.001
Replicate C	r^2		0.923	0.962	0.726
	Mean effects	on R	-0.106	0.7814	0.498
		on G	-1.234	-0.1840	0.146
		on B	-0.659	0.0912	0.014
	% of total contributions	to R	0.080	0.743	0.175
		to G	0.900	0.068	0.030
		to B	0.913	0.084	0.001

5 Supplementary

A Bayesian regularisation

In this section we describe how to derive the modified model selection critieria developed by Cawley 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,

$$\log P(\theta|\mathscr{D}) \propto \frac{I}{2}\log\left(\sum_{i=1}^{I} e_i(\mathscr{D},\theta)^2\right) + \frac{J}{2}\log\left(\sum_{j=1}^{J}\theta_j^2\right) \tag{12}$$

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 a 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

$$\log P(\theta|\mathscr{D}) \propto \frac{I}{2}\log\left(1+\sum_{i=1}^{I} e_i(\mathscr{D},\theta)^2\right)+\frac{J}{2}\log\left(1+\sum_{j=1}^{J}\theta_j^2\right) \tag{13}$$

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.