

Neural Ordinary Differential Equations for Ecological and Evolutionary Time Series Analysis

Willem Bonnaffé¹, Ben Sheldon¹, and Tim Coulson¹

¹University of Oxford Department of Zoology

May 11, 2020

Abstract

We present a novel method, Neural Ordinary Differential Equations, for learning ecological and evolutionary processes from time series data. The method consists in modelling dynamical systems with Ordinary Differential Equations and dynamic functions with Artificial Neural Networks, which upon successful training converge to functional shapes that best describe the processes. We tested NODEs by inferring per-capita growth rates of hare and lynx in simulated and real time series, which revealed that prey-predator oscillations were mainly driven by stronger predation at low hare and lynx density, as well as negative density-dependence in lynx, in line with the literature, thus demonstrating the validity and utility of NODEs. The approach is applicable to any system that can be modelled with differential equations, and particularly suitable for linking ecological, evolutionary, and environmental dynamics where parametric approaches are too challenging to implement, opening new avenues for theoretical and empirical investigations.

Introduction

Understanding and predicting ecosystem dynamics is at the heart of many fields in Ecology and Evolution, whether it is to forecast the outcome of evolutionary responses (Norberg et al. 2012), adverse consequences of climate change (Cotto et al. 2017), or to assess population viability (Pierson et al. 2015). Contrary to physical systems, the complexity of biological systems hinders our ability to anticipate how they change. For instance, even a seemingly simple system of two interacting species like the hare-lynx system can lead to intriguing and unpredictable outcomes, such as the hare apparently feeding on lynx (Deng 2018). While incorrectly inferring the drivers of historical hare and lynx population dynamics might have limited consequences, wrong predictions of environmental threats, such as effects of pests on crops, harvesting on natural resources, or epidemics (Ferguson et al. 2020), may have a greater societal impact. It is hence paramount that we continue improving our capacity to understand and forecast natural systems.

In practice, gaining both understanding and predictability over a given system can be achieved through mathematical modelling of time series data, which is broadly divided in into parametric and non-parametric approaches. Parametric models designed for this purpose consist of time difference equation systems in discrete time, such as auto-regressive models (e.g. Stenseth et al. 2002), and Ordinary Differential Equation systems in continuous time (e.g. Deng 2018). The general approach consists in translating hypotheses regarding the dynamics of an ecosystem into a mathematical model, the behaviour of which is confronted

with observations, and modified by trial and error until it reproduces satisfactorily the behaviour of the system. This approach is limited by our capacity to formulate assumptions and conceptualise the system, which makes it challenging to build a suitable model for a given ecosystem, and further problematic in cases where fast action is required (e.g. dealing with invasive species, algal blooms).

The counterpart of parametric models are non-parametric models, primarily consisting of recurrent Neural Networks (RNN) (Lek and Gue 1999), which allow for the identification of the impact and interdependences between different explanatory variables. In practice, RNNs are fitted to the time series by predicting iteratively the state of the system at the next time step given the current state and associated covariates. These models have been applied successfully to study ecological dynamics, for instance to forecast the growth of pine trees (Chon et al. 2000), changes in aquatic benthic communities (Chon et al. 2001), and spatial dynamics, for example to model the dispersal of ungulates (Dalziel et al. 2017). The main disadvantage of these approaches is that they are not embedded in a theoretical framework and thus provide limited insights into the mechanisms at play.

Neural Ordinary Differential Equations (NODEs), one of the latest breakthroughs in AI research (Chen et al. 2018, Rubanova et al. 2019), offer a promising alternative, as they combine the strengths of both parametric and non-parametric approaches to time series analysis. NODEs are defined by embedding Artificial Neural Networks (ANNs), which are universal function approximators (Chen and Chen 1993), in Ordinary Differential Equation systems (ODEs), which are powerful mathematical models for dynamical systems, allowing ANNs to approximate key dynamical processes (Chen et al. 2018, Rubanova et al. 2019). Then by training the NODE system to replicate time series of observations of the system state, the ANNs converge to the functional shapes that best describe the underlying dynamical processes (Wu et al. 2005). NODEs hence provide a novel way to infer hidden dynamical processes from time series data (Chen et al. 2018, Rubanova et al. 2019), as the theory known a priori can be specified with parametric ordinary differential equations, and missing theory can be approximated non-parametrically by the ANNs (Wu et al. 2005).

Hybrid approaches combining ANNs and ODEs have been applied with success in the past across a broad range of fields. For instance, Physicists have recently applied approaches combining Partial differential equations and ANNs to model fluid dynamics (Sinchev et al. 2018), and to discover governing physical equations from data (Berg and Nystrom 2019). Neuroscientists have modelled the activation dynamics of interacting populations of neurones in an Ape with a NODE system, faithfully reproducing the behaviour of motoneurons (Sussillo et al. 2015). Only one study in Ecology, to our knowledge, successfully combined ANNs with ODEs to learn functional responses in a prey-predator system, reproducing intriguing non-stationary oscillations (Wu et al. 2005). Altogether, the work demonstrates that combining ANNs and ODEs can produce realistic system dynamics and spearhead data-driven theoretical investigations and predictions.

The potential of NODEs to further our understanding of ecological and evolutionary ecosystem dynamics remains largely unexplored. We believe that this has to do with the fact that no study has formally introduced the framework in an ecological or evolutionary context and properly highlighted the potential of ANNs for approximating and analysing key dynamical processes, such as trophic interactions or fitness landscapes. Our aim in this paper is to introduce the NODE mathematical framework and demonstrate how it can be applied more broadly to study ecological and evolutionary dynamics, both spatial and temporal. To do this we define the mathematical pre-requisites and general steps to implement NODE systems for a given time series. Then, we demonstrate how the approach can be used in practice by analysing the drivers of prey-predator oscillations in the hare-lynx pelt count time series (Odum and Barrett 1972). In particular, we highlight how ANNs uncover non-linearity and phase-dependence in the intra and inter-specific interactions that govern hare and lynx dynamics, which fall in line with the literature. Finally, we discuss these results and high-

light further applications and developments of the framework in the context of ecological and evolutionary dynamics.

Material and Methods

From ODEs to NODEs

The dynamics of biological systems can be modelled with systems of Ordinary Differential Equations (ODEs, e.g. Van Velzen and Graedke 2017, Deng 2018)

$$\dot{X}^{(k)} = f^{(k)}(X, \theta^{(k)})$$

(1)

where each equation governs changes in the k^{th} state of the system in continuous time $\dot{X}^{(k)} = dX^{(k)}/dt$ as a function of the state vector $X = [X^{(1)}, X^{(2)}, \dots, X^{(J)}]^T$ and parameter vector $\theta^{(k)}$. Classically, we would formulate parametric equations for process functions $f^{(k)}$ based on biological assumptions regarding the dynamics of the states. For instance, the Lotka-Volterra model assumes a linear decrease in the per-capita growth rate of the prey $\dot{X}^{(1)} = (\alpha - \beta X^{(2)})X^{(1)}$ with predator density $X^{(2)}$.

A Neural Ordinary Differential Equation (NODE) is a class of Ordinary Differential Equation (ODE) that relies instead on Artificial Neural Networks (ANNs) to approximate process functions that constitute differential equations (Funahashi et al. 1993, Chen et al. 2018), and thereby the processes (e.g. per-capita growth rate) that underpin the changes in the biological states (e.g. population size). ANNs are non-parametric equations that can approximate any continuous functions to a certain degree of precision (Chen and Chen 1993), giving NODE systems the potential to approximate any dynamical process, starting with net changes in biological states

$$\dot{X}^{(k)} = \tilde{f}^{(k)}(X, \Omega^{(k)})$$

(2)

where non-parametric functions $\tilde{f}^{(k)}$ are ANNs (thereafter denoted by the \sim) that return the net change in the k^{th} system state given the state vector X and weight vector $\Omega^{(k)}$ (see Fig. 1). This general form can be modified to model specific biological features and assumptions about the system by embedding ANNs in parametric or constrained equations (see case study, and Wu et al. 2005). External covariates, the dynamics of which is not modelled by a NODE, can be added as inputs to the ANNs as long as they have a fine temporal resolution.

In this study, we focus on ANNs with a single intermediate layer and one output value, namely Single-Layer Perceptrons (SLPs), as they are common and easy to use mathematically and computationally (Chen and Chen 1993, Jagdish et al. 1999, Wu et al. 2005),

$$\tilde{f}^{(k)}(X, \Omega^{(k)}) = \delta_0^{(k)} + \sum_{i=1}^I \omega_{i0}^{(k)} f_{\sigma} \left(\delta_i^{(k)} + \sum_{j=1}^J \omega_{ij}^{(k)} X^{(j)} \right)$$

(3)

where $X = [X^{(1)}, X^{(2)}, \dots, X^{(J)}]^T$ is the vector of J input variables, Ω is the vector of biases and weights, δ and ω , respectively. Biases $\delta_i^{(k)}$ control the baseline value of each node i in the absence of signal, and $\omega_{ij}^{(k)}$ is the weight of the input j on the intermediate node i , and $\omega_{i0}^{(k)}$ the weight of that intermediate node on the output node. In essence, a SLP is simply a weighted sum of non-linear basis functions f_{σ} (e.g. sigmoids, gaussians) taking as argument weighted sums of explanatory variables (Jagdish et al. 1999), introducing non-linear effects of covariates on the response

Implementation and training of NODE systems

Below we provide the general steps to implement and fit a NODE system to a given set of time series of K observed system states $x^{(k)}$. For each latent state $X^{(k)}$,

- Define $dX^{(k)}/dt$ as a combination of parametric and non-parametric functions $f_i^{(k)}$ and $\tilde{f}_j^{(k)}$ and states $X^{(l)}$
- Define parametric functions $f_i^{(k)}$ as combinations of parameters $\theta_{ip}^{(k)}$ and states $X^{(l)}$
- Define non-parametric functions $\tilde{f}_j^{(k)}$ as SLPs of weight vector $\Omega_j^{(k)}$ and states $X^{(l)}$
- Initiate all parameters and weights $\theta_{ip}^{(k)}$ and $\Omega_j^{(k)}$ (e.g. draws in uniform distributions)
- Set the initial state of the system $X_0^{(k)}$ at time t_0 (e.g. draw in uniform distribution)

Fitting a NODE model involves the following steps,

- Step 1: compute predicted states $\hat{X}_t^{(k)}$ for each time t in the time series, by integrating the K equations from t_0 to t , $\hat{X}_t^{(k)} = X_0^{(k)} + \int_{t_0}^t dX^{(k)}$ (function *ode*, package *deSolve*, R)
- Step 2: calculate total error e by summing the squared distances between observed and predicted states $x_t^{(k)}$ and $\hat{X}_t^{(k)}$ across all t and k (i.e. Ordinary Least Squares).
- Step 3: Reduce e by adjusting parameters and weights $\theta_{ip}^{(k)}$ and $\Omega_j^{(k)}$. We rely solely on the *optim* function (method *BFGS*) in R to fit all NODE systems, which we achieve with reasonable training times (< 5 mins). For harder problems, there exist more advanced algorithms (Wu et al. 2005, Chen et al. 2018, Rackauckas et al. 2019).

- Step 4: quantify uncertainty in NODE predictions via *ensembling* (Pearce et al. 2018). The ensemble is obtained by optimising multiple NODE systems with different initial conditions and retaining only a fraction of models that best fit the time series (e.g. 25%). The uncertainty is given by the mean and range of predictions and quantities derived from the NODE systems in the ensemble. The size of the ensemble depends on the problem (see case study).

We chose these techniques (*OLS*, *BFGS*, *ensembling*) as they facilitate the implementation and use of NODEs. We would like to stress the fact that our method is hence primarily a mathematical modelling approach, rather than statistical. Though this exceeds the aim of our study, casting NODEs in a Bayesian framework is an exciting prospect, and would benefit from recent work in this direction (Zhu et Zabarar 2018, Pearce et al. 2018).

Analysing NODEs

ANNs provide non-parametric approximations of the processes that drive the changes in the system (e.g. birth rates). Once the training of the NODE is complete, the ANNs will have converged to the functional shape that best fit the processes they approximate. To characterise the functional shape of the dynamical processes we can thus simply analyse the ANNs, by evaluating their output across different combinations of dependent states X (e.g. heatmap plots, see case study).

In addition, by providing non-parametric approximations of dynamical processes, ANNs automatically provide approximations for any (latent) quantities that can be derived from them. This includes the entire set of linear algebraic quantities aimed at analysing dynamical systems, such as the determinant, Jacobian and Hessian matrices. These quantities convey critical biological information, which depending on context, can be species interactions (Montoya et al. 2009), eco-evolutionary feedbacks (Hairston et al. 2005, Patel et al. 2018, Van Velzen and Graedke 2017), equilibrium points (Patel et al. 2018, Cortez and Patel 2017), and ultimately, the future state of the system.

In this paper, we restrain ourselves to retrieving the Jacobian matrix of processes, which contains inter-dependences between the different states, as this is most useful to a mounting body of work linking ecology, evolution, and environmental change (Hairston et al. 2005, Ellner et al. 2011, Van Velzen and Graedke 2017, Pigeon et al. 2017). More specifically, we quantify direct effects of dependent states $X^{(i)}$ on a given dynamical process $\tilde{f}^{(k)}$ of state $X^{(k)}$ by analysing its Jacobian matrix, noted $J_{\tilde{f}}$, which contains the gradient of the process along each dependent state, namely $J_{\tilde{f}} = \left[\partial \tilde{f}^{(k)} / \partial X^{(i)} \right]_J$. We further compute explicitly how much each dependent state contributes to changes in the focal process through the chain rule (Hairston et al. 2005)

$$\frac{d\tilde{f}^{(k)}}{dt} = \sum_i \dot{X}^{(i)} \frac{\partial \tilde{f}^{(k)}}{\partial X^{(i)}}$$

(4)

where each contribution is computed as the product of the dynamics of the dependent state $\dot{X}^{(i)}$ and its direct effect on the dynamical process $\partial \tilde{f} / \partial X^{(i)}$ (i.e. the Jacobian element). Likewise, indirect and interactive effects and contributions can be obtained by applying the same steps to the Hessian matrix of the process.

Case study: prey-predator oscillations

In this section we provide an example of NODE implementation to study density-dependence in prey-predator cycles, using the hare-lynx system as a model system (Krebs et al. 2006), as it provides a simple and well-studied example, thus facilitating the assessment of the consistency of our results with the literature. Hare and lynx display clear 10-year long population cycles, which are thought to be driven by a combination of phase dependence in predation and negative density-dependence in lynx (Slough and Mowat 1996, Stenseth et al. 2002, Krebs et al. 2006). Our aim is to use a NODE system to learn the per-capita growth rate of the prey and predator from time series of lynx and hare density. Then, by recovering the Jacobian of the system, we quantify intra- and inter-specific density-dependence in growth rates, and through the chain rule, establish the respective contribution of intra- and inter-specific interactions to the demography of each population. The code, and additional explanations, can be found on a Github repository referenced in the data accessibility section.

The overarching goal of this case study is to establish that our approach returns (i) valid results when recovering growth rate from a simulated time series and (ii) biologically sensible results when applied to a real time series. Hence, we analyse first a simulated time series, generated from the classic Lotka-Volterra model for the dynamics of the prey $dN_1/dt = r_1 N_1$ and predator $dN_2/dt = r_2 N_2$ (Fig. 2, a. and b.), where the per-capita growth rate of the hare is a linear decreasing function of lynx density $r_1 = \alpha - \beta N_2$, while that of the lynx is a linear increasing function of hare density $r_2 = \gamma N_1 - \delta$. To generate the time series we let $N_1(0) = 1$ and $N_2(0) = 1$, set $\alpha = \delta = 1$ and $\beta = \gamma = 0.5$ to obtain a single 15 year long prey-predator oscillation, integrate the system over $t = [0, 15]$, and sample the model run every $\Delta t = 0.5$ time steps. Then we analyse a real time series of hare and lynx pelts (in tens of thousands), collected by trappers in the Hudson Bay area, from 1904 to 1934 (31 years, Fig. 4 a. and b.), which displays 3 characteristic oscillations, each lasting ten years (Odum and Barrett 1972). We chose these oscillations as they have been studied extensively, allowing for the most sensible comparisons between the results of our case study and the literature (Stenseth et al. 2002, Krebs et al. 2006).

We define a NODE system in which changes in hare and lynx density are governed by a differential equation and their per-capita growth rate is approximated by an ANN

$$\begin{aligned} \dot{N}_1 &= \tilde{r}_1(N_1, N_2, \Omega_1) N_1 \\ \dot{N}_2 &= \tilde{r}_2(N_1, N_2, \Omega_2) N_2 \end{aligned} \quad (5)$$

where $\dot{N}_k = dN_k/dt$ and \tilde{r}_k denote the total, and per-capita, population growth rate of the k^{th} population, respectively. The per-capita growth rates are approximated with SLP functions (equation (3))

$$\tilde{r}_k(N_1, N_2, \Omega_k) = \delta_0^{(k)} + \sum_{i=1}^I \omega_{i0}^{(k)} f_{\sigma} \left(\delta_i^{(k)} + \omega_{i1}^{(k)} N_1 + \omega_{i2}^{(k)} N_2 \right) \quad (6)$$

of the hare and lynx densities N_1 and N_2 , weight vector Ω_1 and Ω_2 , and where the number of intermediate neurones is $I = 10$ and sigmoid activation function has the form $f_\sigma(x) = 1/(1 + \exp(-x))$. The value of 10 hidden neurones was chosen by trial and error and served all purposes, it should be well over the dimension of the input (by at least two times, Chen and Chen 1993), but as low as possible (Wu et al. 2005). To account for non-stationarity in the real time series, namely that the per-capita growth rates are not constant in time, we introduce an explicit dependence on time in the NODE systems that we fit to the real time series, which is done by adding the year t as an input node in the SLPs.

We fit the NODE model following the optimisation approach described in section , by running 200 optimisations, with random initial weights and biases (normal distribution, mean=0, sd=0.001) and retaining 25% of the best models, which resulted in two ensembles of 50 elements, at which point adding elements would not change results. All quantities below are averaged over all elements in the respective ensembles. For each set of time series, (i) we visualise the dependence of per-capita growth rates on hare and lynx densities by evaluating the SLPs $\tilde{r}_k(N_1, N_2, \Omega_k)$, at every pairwise combinations of $N_1 \in \{0, 0.05, \dots, \max(N_1)\}$ and $N_2 \in \{0, 0.05, \dots, \max(N_2)\}$ (Fig. 3 and 5). Then, (ii) we quantify intra- and inter-specific direct effects on hare and lynx per-capita growth rates by computing the Jacobian matrix of the SLPs at every time step

$$\frac{\partial \tilde{r}_k}{\partial N_j} = \sum_i \omega_{i0}^{(k)} \frac{\partial}{\partial N_j} f_\sigma \left(\delta_i^{(k)} + \omega_{i1}^{(k)} N_1 + \omega_{i2}^{(k)} N_2 \right) \quad (7)$$

where $\partial \tilde{r}_k / \partial N_j$ is the change in the per-capita growth rate of population k with respect to a change in the density of population j , which captures the intra-specific effect when $j = k$, and the inter-specific effect otherwise. We visualise the changes in these effects throughout the time series by evaluating the Jacobian at every time step, namely for each N_1 and N_2 predicted by the NODE system (Fig. 2 and 4, c. and d.). Finally, (iii) we quantify the actual contribution of these effects to the temporal changes in the per-capita growth rates via the chain rule (a.k.a. the Geber method, Hairston et al. 2005)

$$\frac{d\tilde{r}_k}{dt} = \dot{N}_1 \frac{\partial \tilde{r}_k}{\partial N_1} + \dot{N}_2 \frac{\partial \tilde{r}_k}{\partial N_2} \quad (8)$$

where $\dot{N}_k \partial \tilde{r}_k / \partial N_k$ and $\dot{N}_i \partial \tilde{r}_k / \partial N_i$ captures the intra- and inter-specific contribution, respectively. We can also visualise the change in these contributions throughout the time series by evaluating them at every time step t , and N_1, N_2 predicted by the NODE system (Fig. 2 and 4, e. and f.).

Results

Simulated time series

We obtain a perfect fit of the NODE system to the time series simulated from the Lotka-Volterra model (Fig. 2 a. and b.). The SLPs functions of the per-capita growth rates approximate almost exactly the shape and levels of the true per-capita growth rates (Fig. 3 a. and c.). The analysis of the Jacobian elements reveals a negative direct effect of lynx density on hare per-capita growth rate, and a positive direct effect of hare density on lynx per-capita growth rate, both of which remain consistent throughout the oscillation (Fig. 2 c. and d., red line). Intra-specific direct effects are not different from zero in both lynx and hare (Fig. 2 c. and d., black line). The analysis of the contributions further reveals that changes in the per-capita growth rates are exclusively driven by the inter-specific effects, and not by the intra-specific effects (Fig. 2 e. and f.). Overall, the ANN does well at recovering the Lotka-Volterra behaviour underlying the time series, in that population dynamics are driven solely by the prey-predator interaction.

Real time series

The NODE system also does well at reproducing the real oscillations, apart from minor departures (Fig. 4 a. and b.). The analysis of the Jacobian reveals a negative effect of lynx density on hare growth, strongest at low lynx density (Fig. 4 c. red line). In contrast, the effect of hare density on hare growth is small, though slightly positive in the second half of the hare cycle, when hare density is decreasing (Fig. 4 c. black line). The analysis of the contributions indicates that the inter-specific effect is the main driver of hare growth (Fig. 4 e. red line), as the intra-specific contribution is small, and impactful only during the second half of the hare cycle (Fig. 4 e. black line). As for the lynx population, the Jacobian analysis reveals a positive effect of hare density on lynx growth, only at low hare density (e.g. years 11 to 16, Fig. 4 d. red line), and a negative effect of lynx density on lynx growth, only at low lynx density (e.g. years 13 to 17, Fig. 4 d. black line). The analysis of the contributions shows that both effects impact substantially and sequentially the dynamics of the lynx population, as the inter-specific contribution increases in the lynx off-peak (Fig. 4 f. red line), immediately followed by negative intra-specific density-dependence (Fig. 4 f. black line). Overall, these results show that the hare population is driven mostly by its interaction with the lynx, and that the lynx is driven in equal measures by the hare population and negative intra-specific density-dependence. These results also demonstrate that the interaction between the two species is intermittent, in that the inter-specific effects only occur in the off-peaks, when both populations are at low density. We also note slight temporal trends, such as a tendency for effects to become stronger over time (e.g. Fig. 4 f. red line), and for lynx growth to improve overall (Fig. 5, right column).

Discussion

The aim of this paper is to introduce the novel mathematical framework of Neural Ordinary Differential Equations (NODEs) and to illustrate how it can be used to learn ecological and evolutionary dynamical processes from time series data. We present the general framework and use it to identify the drivers of prey-predator oscillations, by retrieving the Jacobian of the system and applying the Geber method. In particular, we infer per-capita growth rate in a simulated time series of lynx and hare counts generated from the Lotka-Volterra equations, and a real time series of hare and lynx pelt counts from the Hudson bay company records (Odum and Barrett 1972). We were able to quantify precisely the intra- and inter-specific

drivers of the dynamics of each population and how these changed throughout the oscillations, revealing that the strength of the interaction between the hare and lynx was magnified at low density of both species. These results demonstrate that the approach is reliable and provides valuable biological insights into the drivers of population dynamics, both in a controlled and a real setting. We discuss below the validity of the approach and confirm the biological relevance of our results in light of the literature. We also highlight further applications to study ecological and evolutionary time series and connections with existing empirical and theoretical frameworks. We conclude by comparing this novel approach to existing methods for time series analysis, and providing avenues for future developments.

Hare-lynx case study

The NODE system fitted the simulated lynx-hare time series and recovered the per-capita growth rates almost perfectly, which demonstrates the validity of the approach in a controlled setting, adding to the list of validations that the method received in other contexts (Wu et al. 2005, Mai et al. 2016, Chen et al. 2018, Rubanova et al. 2019). Furthermore, our analysis of the real time series is consistent with the literature. Our results reveal that the hare population is primarily driven by predation, rather than intra-specific density-dependence, which is consistent with empirical and experimental studies that ruled out the influence of adverse interactions with conspecifics on population growth, partly in line with low food constraints (Stenseth et al. 2002, Krebs et al. 2006). In contrast, we find evidence for negative-density dependence in the lynx population, especially at low lynx density, which in the literature is attributed to the fact that lynx tend to be more territorial before the population peak, after which territoriality breaks down, thus alleviating intra-specific competition in the second half of the cycle (Slough and Mowat 1996). Perhaps our most interesting finding is the intermittence in the interaction between hare and lynx, as the effect of predation is magnified when both species are at low density, and otherwise relatively weak. Interestingly, this is consistent with empirical work that demonstrated that the trophic interaction between lynx and hare was phase dependent, as lynx are more efficient hunters in the increasing part of the cycle, possibly due to higher food constraints at low hare density (Stenseth et al. 2002). Overall, this case study shows that our approach produces valid results when compared to ground truth, and consistent results with respect to the biology of the hare-lynx system, which is highly promising and encouraging for further applications of the approach.

Further applications to ecological and evolutionary dynamics

The NODE system in our case study can be used to identify the drivers of the demography of populations from time series (Fig. 6, 1A and 1B), and is readily applicable to a range of other ecological interactions, whether it is to quantify competitive interactions (Gamelon et al. 2019) or trophic interactions (Hiltunen et al. 2013). This can be achieved by modelling changes in population densities with differential equations (Fig. 6, 1C), and by approximating demographic processes, such as per-capita growth rates, birth or death rates, consumption rates, with ANNs (Fig. 6, 1D). Then by retrieving the Jacobian of the system we can establish networks of ecological interactions and their contribution to the dynamics of each species in the community. In doing so, NODEs also provide a bridge between theoretical ecological interactions frameworks, such as that of Press Perturbation (Montoya et al. 2009), or classic trophic ODE models (Hiltunen et al. 2013), and real systems.

Similarly, NODEs can be applied to identify the drivers of fast evolution in natural systems (Fig. 6, 2A) from time series of the traits, demography, and environmental variables (Fig. 6, 2B), such as those available for the

Darwin's Finches system (Grant and Grant 1993). This can be done by defining a set of differential equations describing changes in phenotypic, demographic, and environmental states (Fig. 6, 2C), for instance by merging demographic models with the Price equation (Ellner et al. 2011, Lion 2018), and by approximating the fitness landscape and fitness gradient with ANNs (Fig. 6, 2D). Following an application of the Geber method to both the fitness landscape and fitness gradient, we could derive the contribution of ecology and evolution to the dynamics of the phenotype and populations (Hairston et al. 2005). In this way NODE provide a novel non-parametric way to apply the Geber method, which would benefit a mounting body of work relying on this method (Hairston et al. 2005, Van Velzen and Graedke 2017, Pigeon et al. 2017), and provide an interesting way to test theoretical frameworks for eco-evolutionary feedbacks in real time series (e.g. Patel et al. 2018, Lion 2018).

NODEs can also be used to learn the drivers of spatial dynamics of individuals and populations (Fig. 6, 3A) from records of paths and trajectories (e.g. foraging, dispersal, migrations) (Fig. 6, 3B). This can be achieved by modelling the changes in the spatial position of individuals, or populations with differential equations (Fig. 6, 3C), and approximate the movement kernels, that is the probability of a given move depending on a set of cues, with ANNs (Fig. 6, 3D). We view this last application of NODEs as particularly useful given that we often lack a consistent mathematical framework to explain animal movement (Dalziel et al. 2017), and would lead to novel insights if combined with theoretical frameworks for shoaling and flocking behaviours, or flying predators (Olberg 2012, Corcoran et al. 2016, Brighton et al. 2017).

Comparison to parametric and non-parametric approaches

The two main categories of parametric approaches for inferring hidden processes in time series are ODE models in continuous time (e.g. Deng 2018), and auto-regressive models (or difference equation) in discrete time (e.g. Stenseth et al. 2002). In practice, parametric models require (i) a strong knowledge to formulate assumptions regarding the dynamical processes, (ii) advanced mathematical skills to translate these assumptions into parametric equations, and (iii) a considerable amount of model selection to find a suitable model, which becomes impractical with increasing process complexity, in such a way that models are often only compared visually to time series (e.g. Becks et al. 2010, Hiltunen et al. 2013). NODEs address each of these points, as ANNs (i) require a limited amount of assumptions, which merely consists in the specification of inputs, (ii) emulate infinitely many different processes shapes, (iii) browse through them continuously, and that regardless of the complexity of the process. The main benefit of NODEs is hence that they permit to infer dynamical processes where complexity and lack of prior knowledge makes a classical parametric approach impractical, and otherwise provide first-hand information to guide the design and selection on parametric equations. In essence, ANNs in NODEs handle the complexity of the processes at the micro-scale, allowing us to focus on an integrative view of the system, which is useful for the study of the interplay between demography, phenotypic, and environmental change, as this forces us to embrace new levels of complexity.

The main framework for non-parametric time series analysis are recurrent neural networks (RNNs) (Chon et al. 2001, Obach et al. 2001, Matsunaga et al. 2013), which rely on single ANNs to approximate the dynamics of the entire system. In doing so, RNNs are often criticised for their black box nature greatly reduces interpretability of the model and its applicability as an explanatory tool, making them more suitable for forecasting. NODEs circumvent this problem by embedding ANNs in a consistent theoretical framework outlined by the parametric ODE equations, making ANNs more interpretable and meaningful (Aarts et al. 2001). For instance in our case study, we could split the SLP functions approximating per-capita growth rates of the lynx and hare into birth and death rate functions to identify via which vital function (birth or death) these two species influence each other the most. By doing so, we progressively dissect the ANNs into

possible mechanisms, thus whitening the black box (Aarts et al. 2001).

The novelty of NODEs also comes from their hybrid nature, which by embedding ANNs in ODEs provides a novel and unique way to infer non-parametrically key linear algebraic quantities, such as the determinant, Jacobian, and Hessian of dynamical systems. Accessing these quantities offers insights into equilibrium points, stability, resilience of the ecosystem, but also interaction between its components (e.g. phenotype and environment). These quantities are at the heart of theoretical and empirical work (e.g. Hairston et al. 2005, Montoya et al. 2009, Pigeon et al. 2017, Van Velzen and Graedke 2017), that relies on parameterisation to obtain them indirectly, thus not providing an objective inference of their actual value (Wu et al. 2005). NODEs would provide instead the possibility to infer these quantities directly, thus creating a novel field of investigation.

Conclusion

NODEs combine the explanatory power of dynamical mathematical models and the learning capacity of ANNs, giving them the potential to approximate the dynamics of any biological system based on simple time series data. This framework has not been applied yet, its potential to further our understanding of ecosystem dynamics anywhere theory is missing, as well as making more reliable predictions, is unexplored. We showed that NODEs can be applied successfully to analyse time series in ecology, using the hare-lynx prey-predator system as a case-study. We further provided several directions for addressing key questions in ecological, spatial, and evolutionary dynamics, with NODEs, highlighting its generalisability and its potential to generate new insights in ecology and evolution. Successfully applying and developing this framework will also result in concrete applications, such as controlling and restoring systems under anthropogenic perturbations. In particular, we could design NODE systems to control ecosystems in real time, whether it is to maintain them at a maximum sustainable yield (e.g. in fisheries, agro-ecological systems), or to restore and protect them more efficiently against perturbations (e.g. invasive species, climate change). In doing so, the main hope with using NODEs is that we will not be impeded by our understanding of the systems in taking action.

Acknowledgments

We thank warmly the Ecological and Evolutionary Dynamics Lab and Sheldon Lab Group at the department of Zoology for their feedback and support, Michael Bonsall and José Lourenco for their feedback on early versions of the methods. The work was supported by the NERC DTP and Oxford Oxitec Scholarship.

Data accessibility

All data and code will be made fully available at <https://github.com/WillemBonnafe/NODEs>.

References

- Aarts, Lucie P. and Peter Van Der Veer (2001). Neural network method for solving partial differential equations. *Neural Process. Lett.* , 14.3, pp. 261-271.
- Becks, Lutz et al. (2010). Reduction of adaptive genetic diversity radically alters eco-evolutionary community dynamics. *Ecol. Lett.* , 13.8, pp. 989-997.
- Berg, Jens and Kaj Nystrom (2019). Data-driven discovery of PDEs in complex datasets. *J. Comput. Phys.* , 384, pp. 239-252.
- Brighton, Caroline H. et al. (2017). Terminal attack trajectories of peregrine falcons are described by the proportional navigation guidance law of missiles. *Proc. Natl. Acad. Sci. U. S. A.* , 114.51, pp. 13495-13500.

- Chen, Ricky T. Q. et al. (2019). Neural Ordinary Differential Equations. *arXiv* , pp. 1-19.
- Chen, Tianping and Hong Chen (1993). Approximations of Continuous Functionals by Neural Networks with Application to Dynamic Systems. *IEEE Trans. Neural Networks* , 4.6, pp. 910-918.
- Chon, Tae-Soo et al. (2000). Use of an Artificial Neural Network to Predict Population Dynamics of the ForestPest Pine Needle Gall Midge (Diptera:Cecidomyiida). *Environ. Entomol.* , 29.6, pp. 1208-1215.
- Chon, Tae Soo et al. (2001). Patterning and short-term predictions of benthic macroinvertebrate community dynamics by using a recurrent artificial neural network. *Ecol. Modell.* , 146.1-3, pp. 181-193.
- Corcoran, Aaron J. and William E. Conner (2016). How moths escape bats:Predicting outcomes of predator-prey interactions. *J. Exp. Biol.* , 219.17, pp. 2704-2715.
- Cortez, Michael H. and Swati Patel (2017). The Effects of Predator Evolution and Genetic Variation on PredatorPrey Population-Level Dynamics. *Bull. Math. Biol.* , 79.7, pp. 1510-1538.
- Cotto, Olivier et al. (2017). A dynamic eco-evolutionary model predicts slow response of alpine plants to climate warming. *Nat. Commun.* , 8.May, pp. 1-9.
- Dalziel, Benjamin D., Juan M. Morales, and John M. Fryxell (2008). Fitting Probability Distributions to Animal Movement Trajectories: Using Artificial Neural Networks to Link Distance, Resources, and Memory. *Am. Nat.* , 172.2, pp. 248-258.
- Deng, Bo (2018). An Inverse Problem: Trappers Drove Hares to Eat Lynx. *Acta Biotheor.* , 66.3, pp. 213-242.
- Ellner, Stephen P., Monica A. Geber, and Nelson G. Hairston (2011). Does rapid evolution matter? Measuring the rate of contemporary evolution and its impacts on ecological dynamics. *Ecol. Lett.* , 14.6, pp. 603-614.
- Ferguson, Neil M et al. (2020). Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Prepr. Spiral* , March, pp. 3-20.
- Funahashi, Ken ichi and Yuichi Nakamura (1993). Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks* , 6.6, pp. 801-806.
- Gamelon, Marlene et al. (2019). Accounting for interspecific competition and age structure in demographic analyses of density dependence improves predictions of fluctuations in population size. *Ecol. Lett.* , 22.5, pp. 797-806.
- Grant, B. R. and P. R. Grant (1993). Evolution of Darwin's finches caused by a rare climatic event. *Proc. R. Soc. B Biol. Sci.* , 251.1331, pp. 111-117.
- Hairston, Nelson G. et al. (2005). Rapid evolution and the convergence of ecological and evolutionary time. *Ecol. Lett.* , 8.10, pp. 1114-1127.
- Hiltunen, Teppo et al. (2013). Temporal dynamics of a simple community with intraguild predation: an experimental test. *Ecology* , 94.4, pp. 773-779.
- Jagdish, C. et al. (1999). Identification of Nonlinear Dynamic Systems Using Neural Networks. *IEEE Trans. Systems, Man, Cybern.* , 29.2, pp. 254-262.
- Krebs, Charles J. et al. (2006). What Drives the 10-year Cycle of Snowshoe Hares? *Bioscience* , 51.1, pp. 1-35.

- Lek, Sovan and J F Gue (1999). Artificial neural networks as a tool in ecological modelling, an introduction. *Ecol. Modell.* , 120, pp. 1-9.
- Lion, Sebastien (2018). Theoretical approaches in evolutionary ecology: Environmental feedback as a unifying perspective. *Am. Nat.* , 191.1, pp. 21-44.
- Mai, Manuel, Mark D. Shattuck, and Corey S. O'Hern (2016). Reconstruction of Ordinary Differential Equations From Time Series Data. *arXiv* , pp. 1-15.
- Matsunaga, Fabio, Miroslava Rakocevic, and Jacques Duilio Brancher (2013). Artificial neural networks in modeling of environmental time series for yerbamate growth dynamics. *Proc. 7th Int. Conf. Funct. Plant Model.* , May 2016, pp. 2005-2007.
- Montoya, Jose M. et al. (2009). Press perturbations and indirect effects in real food webs. *Ecology* , 90.9, pp. 2426-2433.
- Norberg, Jon et al. (2012). Eco-evolutionary responses of biodiversity to climate change. *Nat. Clim. Chang.* , 2.10, pp. 747-751.
- Obach, Michael et al. (2001). Modelling population dynamics of aquatic insects with artificial neural networks. *Ecol. Modell.* , 146.1-3, pp. 207-217.
- Odum, Eugene P. and Gary W. Barrett (1972). Fundamentals of Ecology. *J. Wildl. Manage.* , 36.4, p. 1372.
- Olberg, Robert M. (2012). Visual control of prey-capture flight in dragonflies. *Curr. Opin. Neurobiol.* , 22.2, pp. 267-271.
- Patel, Swati, Michael H. Cortez, and Sebastian J. Schreiber (2018). Partitioning the Effects of Eco-Evolutionary Feedbacks on Community Stability. *Am. Nat.* , 191.3, pp. 381-394.
- Pearce, Tim et al. (2018). Uncertainty in Neural Networks: Approximately Bayesian Ensembling. *arXiv* , pp. 1-10.
- Pierson, Jennifer C. et al. (2015). Incorporating evolutionary processes into population viability models. *Conserv. Biol.* , 29.3, pp. 755-764.
- Pigeon, Gabriel et al. (2017). Fluctuating effects of genetic and plastic changes in body mass on population dynamics in a large herbivore. *Ecology* , 98.9, pp. 2456-2467.
- Rackauckas, Chris et al. (2019). DiffEqFlux.jl - A Julia Library for Neural Differential Equations. *arXiv* , pp. 1-17.
- Rubanova, Yulia, Ricky T. Q. Chen, and David Duvenaud (2019). Latent ODEs for Irregularly-Sampled Time Series. *arXiv* , pp. 1-11.
- Sinchev, Bakhtgerey et al. (2018). Some methods of training radial basis neural networks in solving the Navier-Stokes equations. *Int. J. Numer. Methods Fluids* , 86.10, pp. 625-636.
- Slough, Brian G. and Garth Mowat (1996). Lynx Population Dynamics in an Untrapped Refugium. *J. Wildl. Manage.* , 60.4, pp. 946-961.
- Stenseth, N. C. et al. (2002). From patterns to processes: Phase and density dependencies in the Canadian lynx cycle. *Proc. Natl. Acad. Sci.* , 95.26, pp. 15430-15435.

Sussillo, David et al. (2015). A neural network that finds a naturalistic solution for the production of muscle activity. *Nat. Neurosci.* , 18.7, pp. 1025-1033.

Van Velzen, Ellen and Ursula Gaedke (2017). Disentangling eco-evolutionary dynamics of predator-prey coevolution: The case of antiphase cycles. *Sci. Rep.* , 7.1, pp. 1-11.

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

Zhu, Yin hao and Nicholas Zabaras (2018). Bayesian deep convolutional encoderdecoder networks for surrogate modeling and uncertainty quantification. *J. Comput. Phys.* , 366, pp. 415-447.

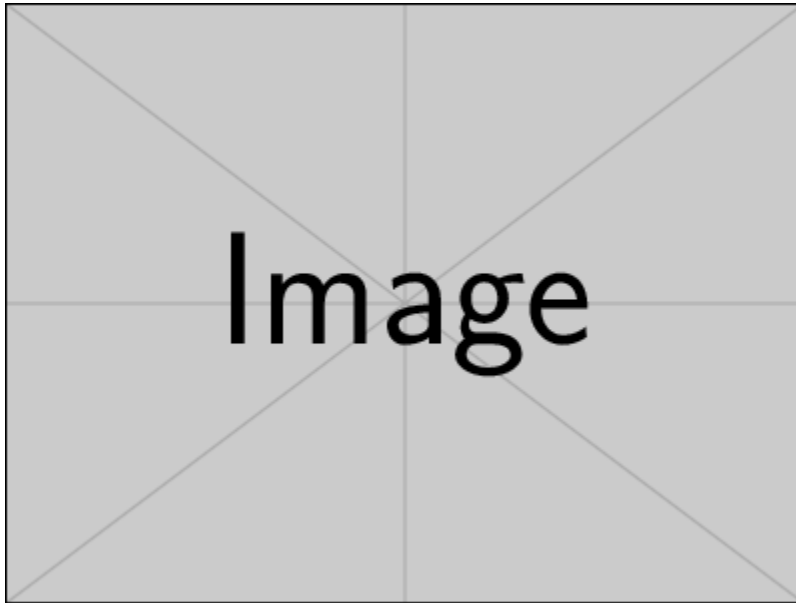


Figure 1: Schematic view of a 2-dimensional NODE system. The SLPs determine the net changes \tilde{f}_1 and \tilde{f}_2 in each system state, given the current system state X_1 and X_2 . The total change in a given state over a given time period t_0 to t_f is obtained by integrating the net changes in the different states with an ODE solver (*ode* function, *deSolve* package, R).

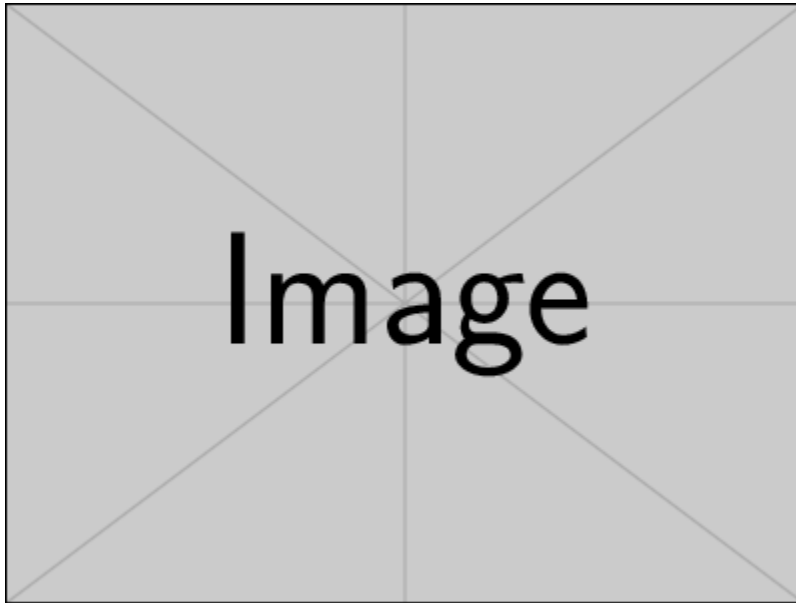


Figure 2: Analysis of the NODE model fitted to the simulated time series. The graph a. and b. represent the simulated time series of hare and lynx, respectively (black dots). The red line correspond to mean density predicted by the NODE system. The graph c. and d. present the changes in the inter-specific (red line) and intra-specific effects (black line) on hare and lynx per-capita growth rates, respectively, given by the Jacobian elements of the SLPs. The graph e. and f. correspond to the contribution of the effects to changes in hare and lynx per-capita growth rates, respectively. In all graphs, the dashed lines correspond to the ground truth, derived from the Lotka-Volterra equations. The shaded areas give the error around mean predictions, given by the range of predictions in the ensemble (grey: 5-95%, dark grey: 25-75%, 50 elements).

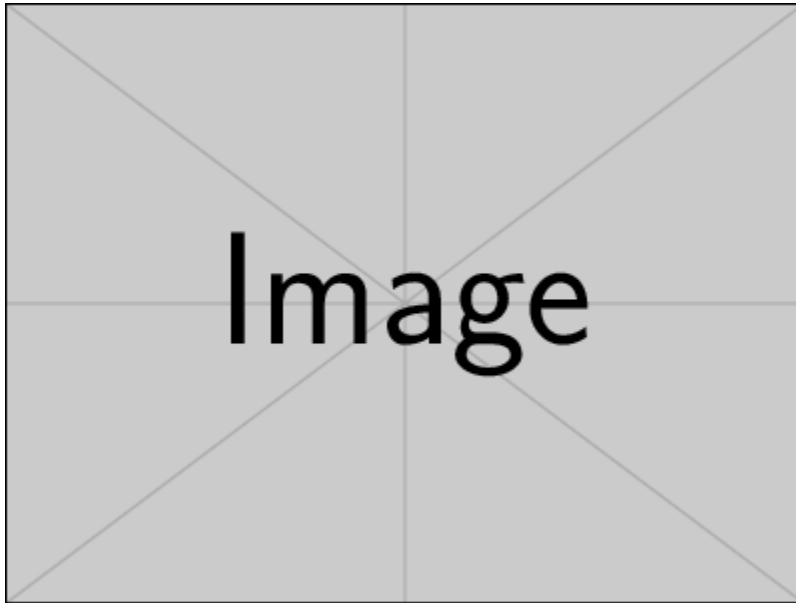


Figure 3: ANN approximation of the per-capita growth rates. The graph a. and c. correspond to the mean ANN approximation of the per-capita growth rates of the hare and lynx populations, \tilde{r}_1 and \tilde{r}_2 , respectively, as a function of hare and lynx density. The graphs b. and d. correspond to the ground truth, that is the per-capita growth rate in the Lotka-Volterra equations, namely $r_1 = 1 - 0.5N_2$ and $r_2 = 0.5N_1 - 1$. The black dots correspond to the data points in the time series.

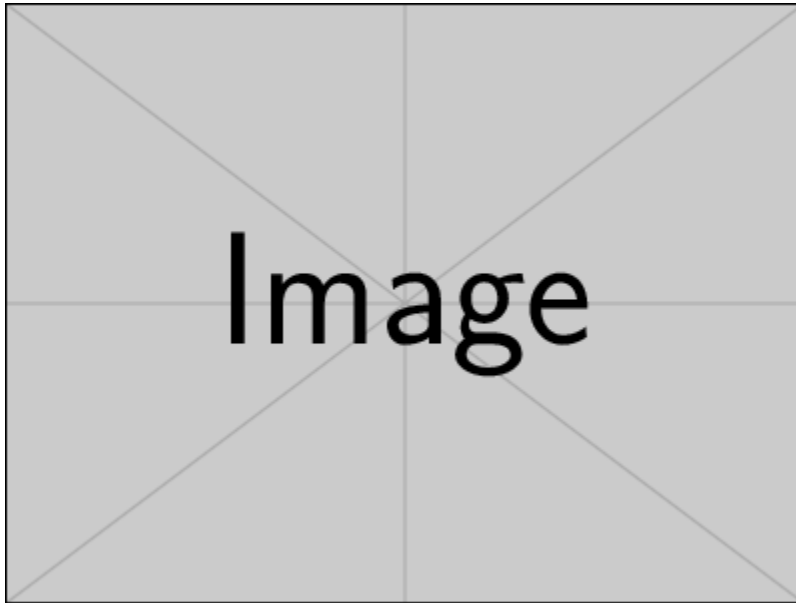


Figure 4: Analysis of the NODE system fitted to the real time series. The graph a. and b. represent the real time series of hare and lynx, respectively (black dots). The red line correspond to mean density predicted by the NODE system. The graph c. and d. present the changes in the inter-specific (red line) and intra-specific effects (black line) on hare and lynx per-capita growth rates, respectively. The effects correspond to the Jacobian elements of the SLPs. The graph e. and f. correspond to the contribution of these effects to changes in hare and lynx per-capita growth rates, respectively. In all graphs, the dashed lines indicate zero, and the shaded areas give the error around mean predictions, given by the range of predictions in the ensemble (grey: 5-95%, dark grey: 25-75%, 50 elements).

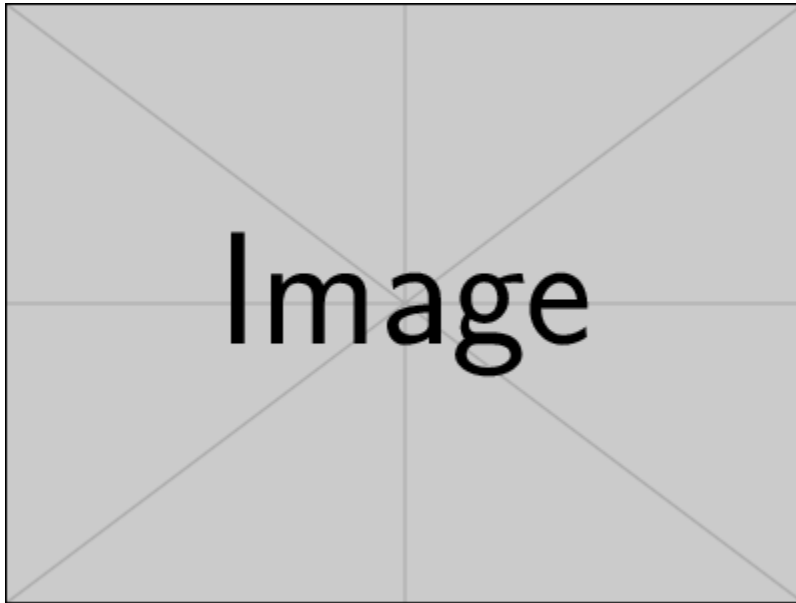


Figure 5: ANN approximation of the per-capita growth rates in the real time series. The graph correspond to the mean ANN approximation of the per-capita growth rates of the hare and lynx populations (left and right column, respectively), as a function of hare and lynx density (x , y axes, respectively), at time $t = 0, 10, 20, 30$ years. The dots correspond to the data points in the time series.

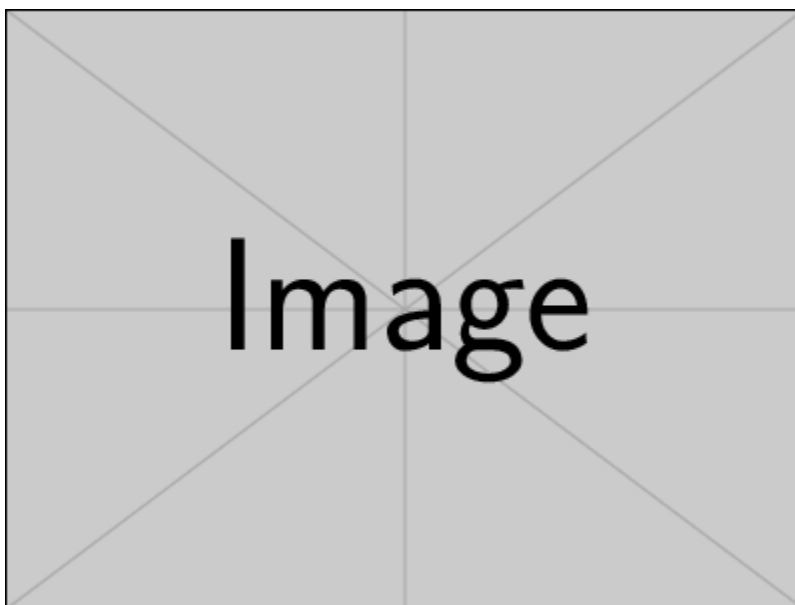


Figure 6: Applications of NODEs to ecological, evolutionary, and spatial dynamics. The biological functions of interest (e.g. p.c. growth rates, fitness gradients) are modelled with ANNs (D). The ANNs feed into the differential equation systems, which are iterated forward to generate predictions (C). The predictions are compared to the observed time series and an error is deduced from it (B). In a final step, the ANNs are adjusted in order to reduce the error (back to D). The process is repeated until the predictions match the observations, at which point the functions of interest should have converged to the most adequate shape given the time series at hand.