Partitioning time series to improve process-based models with machine learning

Victor Boussange[a,b,*], Pau Vilimelis Aceituno[c], Frank Schäfer[d], and Loïc Pellissier[a,b]

[a]Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

[b]Landscape Ecology, Institute of Terrestrial Ecosystems, Department of Environmental System Science, ETH Zürich, Zürich, Switzerland

[c]Institute of Neuroinformatics, ETH Zürich and University of Zürich, Zürich, Switzerland

[d]CSAIL, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[*]Corresponding author,

Email: vic.boussange@gmail.com

April 10, 2024

Abstract

Process-based community models are required to extrapolate beyond current dynamics and anticipate biodiversity response to global change, but their adoption has been limited in practice because of issues with model parameter estimation and model inaccuracies. While observation data could be used to directly estimate parameters and refine model structures, model nonlinearities, sparse and noisy datasets and the complexity of ecological dynamics complicate this process. Here, we propose an inverse modelling framework designed for addressing these issues, that relies on a segmentation method combined with automatic differentiation, deep learning gradient-based optimizers, minibatching, and a parameter normalization technique. By partitioning the data into time series with a short time horizon, we show that the segmentation method regularizes the inverse problem. Implemented in the Julia package `PiecewiseInference.jl`, the resulting framework enables the efficient calibration of realistic dynamical community models and supports the use of neural networks to represent complex ecological processes within the model. We evaluate the performance of the inverse modelling framework on simulated food-web dynamics of increasing complexity, including scenarios with partial observations. Our results demonstrate its performance in inferring model parameters and providing forecasts. We further showcase that the regularization induced by the framework, together with its efficiency, are instrumental for the online training of neural network-based parametrizations, alongside the other model parameters. Neural network-based parametrizations not only mitigate model inaccuracies but can also contribute to the refinement of ecological theory through their offline examination. The proposed inverse modelling framework is data-efficient, interpretable and capable of extrapolating beyond observed dynamics, showing promise in improving our ability to understand and forecast biodiversity dynamics.

Inverse modelling | Process-based modelling | Dynamical community models | Parameter inference | Scientific machine learning

Contents

1 Introduction **4**

2 Materials and methods **7**

2.1 Model parameter inference as a learning problem 7

2.1.1 Ecosystem models 7

2.1.2 Inverse modelling 7

2.1.3 Ill-conditioning of the inverse modelling problem 9

2.2 Inverse modelling framework . . . 11
2.2.1 Description of the segmentation method . . . 11
2.2.2 Numerical implementation of the inverse modelling framework . . . 12

3 Experiments **13**
3.1 Food-web ecosystems . . . 13
3.2 Effect of segment size on parameter, forecast error, and computational complexity . . . 15
3.3 Learning model parametrization . . . 18

4 Discussion **20**

5 Conclusion **23**

6 Acknowledgements **23**

7 Competing interests **23**

8 Code availability **23**

9 Contributions **24**

S1 Parameter values of community models **1**
S1.1 $\mathcal{M}_3$ . . . 1
S1.2 $\mathcal{M}_5$ . . . 1
S1.3 $\mathcal{M}_7$ . . . 2
S1.4 $\mathcal{M}_3^\dagger$ . . . 2

S2 Supplementary Figures **2**

1 Introduction

Ecosystems currently face intense disruption from global change [IPBES, 2019], calling for models that can extrapolate ecological dynamics beyond observations [Boyd, 2012]. A major challenge is that the processes driving ecological dynamics are nonlinear, resulting in complex responses and feedbacks [Scheffer et al., 2001]. Nonlinearity greatly affects the capacity of modelling approaches that do not incorporate specific biological knowledge to reliably project current trends into the future [Barnosky et al., 2012]. For instance, while methods based on statistical descriptions [Guisan and Zimmermann, 2000, Deneu et al., 2021] and nonparametric methods [Ye et al., 2015, Ye and Sugihara, 2016, Deyle et al., 2016] have interesting interpolation capabilities, they are ill-suited for extrapolating beyond observed trends [Barnosky et al., 2012, Urban et al., 2016]. In contrast, process-based community models integrate constraints on the expected dynamics by explicitly modelling interactions, feedback loops, and dependencies between ecosystem components [Geary et al., 2020]. While this should ensure a more robust forecast under large disruptions [Norberg et al., 2012], process-based ecosystems models often fall short in describing observational data with the same accuracy as more standard data-driven techniques [Fordham et al., 2018]. This discrepancy primarily stems from suboptimal parameter values, which are critical in determining the simulation outcomes, and from inaccuracies in the mathematical formulation of the processes [Schartau et al., 2017, Cabral et al., 2017, Briscoe et al., 2019, Pagel and Schurr, 2012]. Novel approaches that can combine the ecological constraints embedded in process-based models with observational data are needed to overcome these limitations and improve our predictive ability regarding ecosystem responses to global changes [Hartig et al., 2012].

The parametrization of community models may be indirectly learnt from observations by estimating the parameters from the data collectively using inference methods [Evans et al., 2016]. Parameter estimation methods proceed by maximizing the posterior probability of the parameters given the observations, but their success is subject to a number of issues, some of which specifically apply to community model properties. Among these issues, the exploration of the posterior landscape demands repeated model simulations, but process-based models are usually associated with a high computational cost that limits the number of possible runs [Fisher et al., 2018]. Additionally, the complexity of processes requires a large number of parameters which, due to the curse of dimensionality [Boyd, 2012], complicates the exploration of the posterior distribution and can further leave many parameters poorly constrained [Gutenkunst et al., 2007]. The limited availability of observations, which are usually composed of multiple partial short-term time series [Scheiter et al., 2013, Schartau et al., 2017], accentuates the lack of parameter constraints. Moreover, ecosystem dynamics can be strongly dependent

on the initial conditions (ICs) and show a chaotic behaviour [Hastings et al., 1993, Huisman and Weissing, 1999, Benincà et al., 2008], or can be associated with a large panel of dynamics depending on the parameter values. In such a case, small perturbations of the ICs or parameter values lead to large divergences in the model outcomes, causing numerical problems in finding the most probable parameters [DeAngelis and Yurek, 2015, Pagel and Schurr, 2012]. Last but not least, in contrast to fields such as climate and weather modelling, a precise quantitative formulation of the ecological processes regulating ecosystems is far from being established [Gehlen et al., 2015, Schartau et al., 2017, Purves et al., 2013]. Complex ecological processes such as functional responses (i.e., the process determining how intake changes with resource density) [Rosenbaum and Rall, 2018], the dependence of demographic parameters on species traits [Chalmandrier et al., 2022], or on environmental variables [Amarasekare and Johnson, 2017] are not adequately captured in process-based models and result in predicting spurious dynamics [Gentleman et al., 2003]. To summarize, the parametrization of community models requires inference methods that are robust despite the models' complexity, the limited observation data, and the inaccurate description of ecological processes.

A variety of data assimilation methods are increasingly being used to parametrize community models. Bayesian inference with Markov Chain Monte Carlo methods, used in Lignell et al. [2013], Higgins et al. [2010], Xu et al. [2006], Fiechter et al. [2013], Rosenbaum et al. [2019], Heiland et al. [2023], Rosenbaum and Fronhofer [2023], Pagel and Schurr [2012], offer the advantage of quantifying uncertainties by inferring the full posterior probability distribution of the unknown parameters. This is achieved by a global exploration of the parameter space, which makes Bayesian methods computationally expensive and particularly prone to the curse of dimensionality [Gosh et al., 2021]. Simulated annealing [Matear, 1995], genetic algorithms [Ward et al., 2010], and sequential methods such as extended Kalman filtering and ensemble Kalman methods [Bertino et al., 2003, Doron et al., 2013, Gharamti et al., 2017] have been used as alternatives, but are similarly subject to the curse of dimensionality and demand a large number of model evaluations. Variational methods typically rely on the model sensitivity to the parameters to explore more efficiently the parameter space, iteratively updating the parameter estimates using the gradient of the posterior landscape. Such methods therefore demand less evaluation [Schartau et al., 2017], which explains their wide adoption in the field of machine learning to train highly parametrized models (up to the order of 10^{10} parameters [Köpf et al., 2023]) and their use in calibrating marine community models (Fennel et al. [2001], Spitz et al. [1998], Xiao and Friedrichs [2014], Pelc et al. [2012], see Schartau et al. [2017] for a review) and terrestrial community models [Zhu and Zhuang, 2015, DeLong et al., 2014, Curtsdotter et al., 2019]. However, as the complex dynamics of community models tend to be associated with rugged posterior landscapes, variational methods are prone to converging to local minima,

making variational methods very sensitive to the choice of initial model parameters [Gábor and Banga, 2015, Schartau et al., 2017]. Ecosystem models are furthermore specified as differential equations that depend not only on parameters but also on ICs. The state-dependency of ecosystems means that neglecting the estimation of initial ICs might compromise the correct fitting of the parameters and the forecast skill [Lignell et al., 2013]. However, few of the aforementioned studies have addressed the problem of IC estimation (but see Pelc et al. [2012]). Finally, the numerical implementation of variational methods is also challenging, as the calculation of the gradient is difficult to implement manually and demands revisions each time the model is modified [Lawson et al., 1995, Pelc et al., 2012, Gharamti et al., 2017]. Novel methods for model parametrization are emerging, thanks to advances in the field of artificial intelligence [Willard et al., 2020, Kashinath et al., 2021, Alber et al., 2019, Peng et al., 2021, Rasp et al., 2018], providing new opportunities to better address these issues.

Here, we propose an inverse modelling framework relying on a multiple-shooting-like method inspired by Pisarenko and Sornette [2004] combined with automatic differentiation, deep learning gradient-based optimizers, minibatching, and a parameter normalization technique to efficiently estimate parameters and learn the parametrization of complex dynamical models from observation data. In Section 2.1, we outline the general inverse modelling problem and the issues arising from the characteristic features of process-based community models. In Section 2.2, we present the proposed framework together with its numerical implementation. In Section 3, we evaluate the performance of the inverse modelling framework through its application to simulated food webs with increasing complexity. We test the performance of our framework in recovering the true parameters and in providing forecasts based on noisy and incomplete observations, and assess the effects of the framework hyper-parameters. Furthermore, we investigate whether the inverse modelling framework can provide neural network-based parametrization of uncertain functional forms that characterize typical ecological processes. We discuss our findings in Section 4.

2 Materials and methods

2.1 Model parameter inference as a learning problem

2.1.1 Ecosystem models

Ecosystem models generally consist of a system of ordinary differential equations (ODEs) of the form

$$\begin{aligned} &\frac{d}{dt}x(t) = f(t, x(t), u(t), p) \\ &x(0) = x_0 \\ &y(t) = h(x(t)) \odot e^{\epsilon(t)} \end{aligned} \tag{1}$$

where $x(t) \in \mathbb{R}^m$ is a vector of state variables that might represent species abundance, resources availability or functional group biomass, $u(t)$ describes external forcings, $y(t) \in \mathbb{R}^d$ is a vector of observables that contains a subset or aggregates of the state variables, $p \in \mathbb{R}^q$ is the model parameter vector and $\odot$ denotes element-wise multiplication. The function h maps the state variables to the observables, and we assume that the observables are contaminated with a white noise ϵ of Gaussian type, with zero mean and covariance matrix Σ_y. Denoting by $\theta = (x_0, p)$ the tuple containing the ICs and the parameters, the model may be viewed as a map $\mathcal{M}$ parametrized by time t that takes the parameters θ to the state variables x

$$\mathcal{M}(t, \theta) = x(t) = \int_0^t f(s, x(s), u(s), p)ds + x_0. \tag{2}$$

2.1.2 Inverse modelling

Taking the expectation over the logarithm of the noise realizations yields $\mathbb{E}[\log y(t)] = \log h(\mathcal{M}(t, \theta))$, and it follows that the conditional likelihood of the logarithm of each observation $y_k \equiv y(t_k)$, given the parameters θ and the model $\mathcal{M}$, denoted by $p(\log y_k|\theta, \mathcal{M})$, follows the distribution of the residuals $\epsilon_k \equiv \epsilon(t_k) = \log y_k - \log h\left(\mathcal{M}(t_k, \theta)\right)$. By assumption, ϵ_k is thus normally distributed with $\mathcal{N}_{0,\Sigma_y}$. Following a Bayesian approach, inference of the model parameters can be performed on the basis of the parameter and model posterior probability $p(\theta, \mathcal{M}|\mathbf{y}_{0:K})$, i.e., the conditional probability density of the parameter values θ and the model $\mathcal{M}$ conditioned on the data, given by Bayes theorem as

$$p(\theta, \mathcal{M}|\mathbf{y}_{0:K}) \propto p(\mathbf{y}_{0:K}|\theta, \mathcal{M})p(\theta, \mathcal{M}) \tag{3}$$

where $\mathbf{y}_{0:K} = (y_0, \ldots, y_K)$ denotes the $K+1$ observations. Assuming independent observations, $p(\mathbf{y}_{0:K}|\theta, \mathcal{M})$ is the product of the conditional likelihood of each observation y_k

$$\begin{aligned} p(\mathbf{y}_{0:K}|\theta, \mathcal{M}) &= \prod_{k=0}^{K} p(y_k|\theta, \mathcal{M}) \\ &= \prod_{k=1}^{K} \frac{1}{\prod_{i=1}^{d} y_{k,i} \sqrt{(2\pi)^d \det \Sigma_y}} \exp\left(-\frac{1}{2}\|\epsilon_k\|^2_{\Sigma_y}\right) \end{aligned} \tag{4}$$

where we use the notation $\|y\|^2_\Sigma = y\Sigma^{-1}y^{\mathbf{T}}$ and inserted that the individual observations are log-normally distributed, and $p(\theta, \mathcal{M})$ is the prior distribution of the model and its associated parameter values. The model $\mathcal{M}$ is included in the probabilistic quantities to accommodate multiple candidate models.

We are interested in computing a Bayesian point estimate of θ given the observations $\mathbf{y}_{0:K}$. We use the maximum a posteriori (MAP) estimator [Bocquet et al., 2019], given by

$$\begin{aligned} \hat{\theta} &= \arg\max_{\theta} \log p(\theta, \mathcal{M}|\mathbf{y}_{0:K}) \\ &= \arg\max_{\theta} \log p(\mathbf{y}_{0:K}|\theta, \mathcal{M}) + \log p(\theta, \mathcal{M}) \end{aligned} \tag{5}$$

The MAP estimator $\hat{\theta}$ is similar to a maximum likelihood estimator (MLE), $\arg\max_\theta \log p(\mathbf{y}_{0:K}|\theta, \mathcal{M})$, but allows incorporating the prior distribution $p(\theta, \mathcal{M})$, which can generally be seen as a regularization to maximum likelihood estimation. Note that the MAP and the MLE estimators are identical under a uniform prior distribution of the parameters, i.e., when no prior information is used [Schartau et al., 2017]. Observing that maximizing $p(\theta, \mathcal{M}|\mathbf{y}_{0:K})$ is equivalent to minimizing $-\log p(\theta|\mathbf{y}_{0:K}, \mathcal{M})$ and assuming a normal prior distribution of the parameters $\mathcal{N}_{p_b, \Sigma_p}$, we can compute the MAP estimate via

$$\hat{\theta} = \arg\min_{\theta} L_{\mathcal{M}}(\theta) \tag{6}$$

where $L_{\mathcal{M}}$ is referred to as the loss function and is defined as [Schneider et al., 2017, Raue et al., 2009]

$$L_{\mathcal{M}}(\theta) = \sum_{k=0}^{K} \| \log y_k - \log h\left(\mathcal{M}(t_k, \theta)\right) \|^2_{\Sigma_y} + \|p - p_b\|^2_{\Sigma_p}. \tag{7}$$

Equation 7 is similar to a traditional least squares function commonly used in regression, where the second summand corresponds to an L2 regularization term.

Gradient-based optimizers can then be used to obtain $\hat{\theta}$ in Eq. (6), iteratively updating the parameter vector

θ_m at iteration m given the gradient of the loss function, denoted by $\nabla_\theta L_\mathcal{M}$, to navigate the surface defined by $L_\mathcal{M}$ with the aim to find the minimum where $\nabla_\theta L_\mathcal{M}(\hat{\theta}) = 0$. As an example, the plain vanilla gradient descent algorithm is given by

$$\theta_{m+1} = \theta_m - \gamma \nabla_\theta L_\mathcal{M}(\theta_m) \tag{8}$$

where γ is the learning rate. More sophisticated gradient-based algorithms, such as the Adam optimizer used in the section below, employ more advanced updating strategies to accelerate convergence.

2.1.3 Ill-conditioning of the inverse modelling problem

The use of gradient based methods for the the minimization of $L_\mathcal{M}$ as formulated in Eq. (7) is likely to fail, due to the dynamics of complex ecological community models and the characteristic of their associated loss function. Firstly, gradient-based methods following Eq. (8) are unconstrained optimization techniques in which parameter updates may result in proposing parameter and ICs values for which the ODE model is undefined. Secondly, the loss surface associated with community models characterized by complex dynamics is ill-conditioned for gradient-based optimization. As illustrated in Fig. 1**A**, it consists of multiple local minima, which cause problems of convergence in efforts to reach the global minimum (see the orange curve in Fig. 1**B**). Additionally, in a neighbourhood of the global minimum, the gradient of the loss function is very large, which may lead the optimizer to overshoot the true parameter values (see the green curve in Fig. 1**B**). These problematic features are due to the dynamical properties of community models: when the dynamics are very sensitive to small perturbations of parameters and ICs (as is the case for ecological dynamics that often exhibit chaos or limit cycle behaviour Bjørnstad and Grenfell [2001], Hastings et al. [1993], Huisman and Weissing [1999], Benincà et al. [2008]), the sensitivity of the dynamical trajectories to the model parameters and ICs grow exponentially over time (see Figs. S2 and S3). This large increase in model parameter sensitivity and ICs results in an unreasonably large gradient around the true parameter values, and results in many local minima in its vicinity. Systematic experiments detailed in Section 3.2 demonstrate more generally that surfaces associated with $L_\mathcal{M}$ in Eq. (7) cannot be navigated properly with gradient descent methods. Since the smoothness of the surface defined by $L_\mathcal{M}(\theta)$ depends on the time horizon, we reformulate Eq. (7) in the following section by partitioning the time series into segments with a short time horizon.

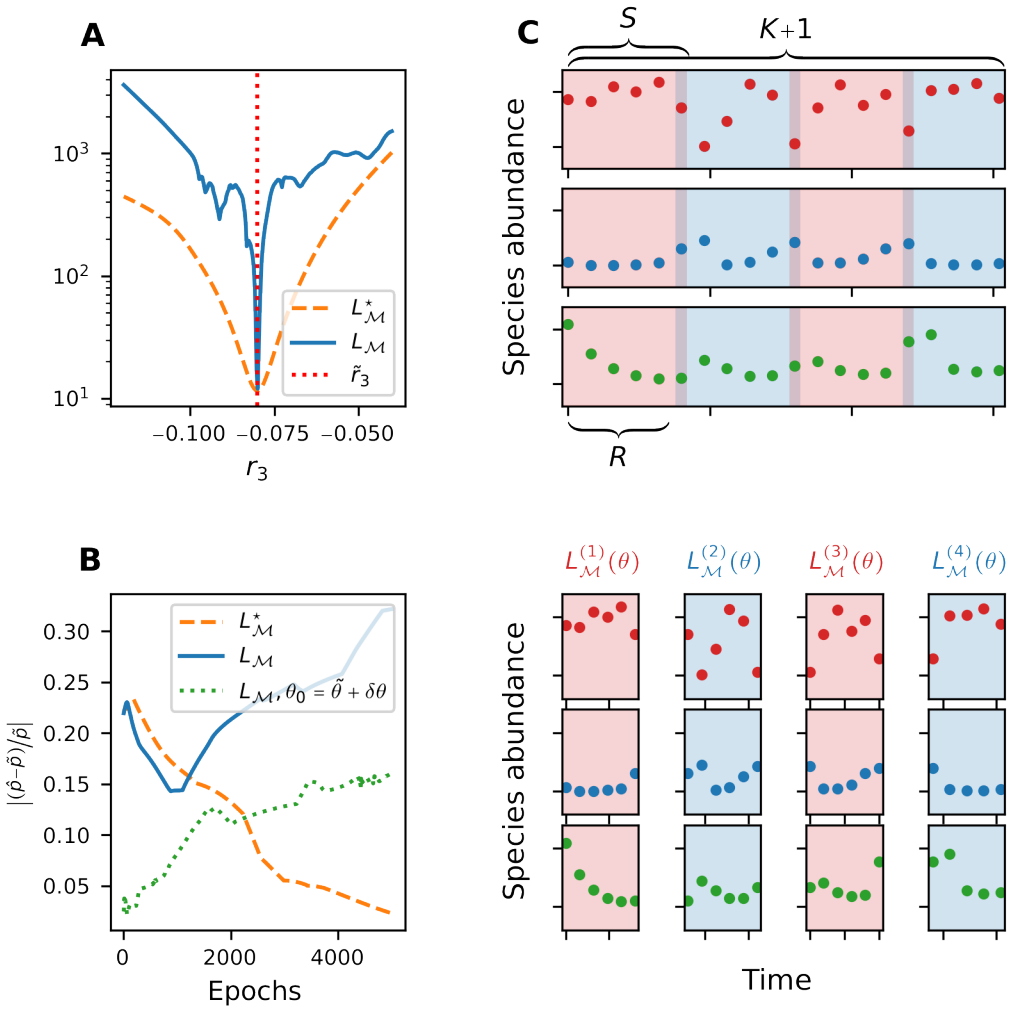

Figure 1: **Illustration of the segmentation method.** **A**. Cross-section of the surfaces defined by the naive loss function $L_{\mathcal{M}_3}(\theta)$ and the segmentation loss function $L^\star_{\mathcal{M}_3}$, for model $\mathcal{M}_3$ presented in Section 3 and in Fig. 2. The surface defined by $L_{\mathcal{M}_3}$ presents many local minima and a very large gradient in the neighbourhood of the true parameter, which renders its navigability difficult. On the contrary, the cross section of $L^\star_{\mathcal{M}_3}$ is smooth and shows a single minimum, illustrating the regularization induced by the segmentation method. **B**. Convergence of the Adam optimizer used to minimize $L_{\mathcal{M}_3}$ and $L^\star_{\mathcal{M}_3}$. The blue, green and orange lines correspond to values of $L_{\mathcal{M}_3}$, $L_{\mathcal{M}_3}$ where the initial guess of the parameter value is close to the true parameters, and $L^\star_{\mathcal{M}_3}$, respectively, evaluated against the number of epochs. The ill-conditioning of $L_{\mathcal{M}_3}$ leads the algorithm to converge to a local minimum, while $L^\star_{\mathcal{M}_3}$ is associated with a smooth convergence to the true parameters. **C**. Graphical representation of the proposed segmentation method. To improve the navigability of the loss surface, the algorithm partitions the time series into M segments with short time horizons, consisting of S data points (blue and red portions of the time series), with shift R. Since segments are treated independently, the method can naturally be extended to harness independent time series.

2.2 Inverse modelling framework

2.2.1 Description of the segmentation method

We use a segmentation method inspired by Pisarenko and Sornette [2004] that partitions the data into segments with a short time horizon. By matching the model simulations against segments of data with short time horizons, the high sensitivity of the dynamics to parameters and ICs, that cause the poor navigability of the landscape, are therefore mitigated. Additionally, we constrain the parameter values to ecologically feasible values using bijective maps, as explained below.

We assume the availability of a time series with $K+1$ observations $\mathbf{y}_{0:K}$. To improve the ill-conditioning of Eq. (7), we partition the time series into segments of length S, shifted by R data points (see Fig. 1**C**). Assuming for notational simplicity that $K+1$ is a multiple of $S-R$ (the overlap between each segment), this provides us with an augmented training dataset containing $M = \frac{K-S}{R} + 1$ segments, $\{\mathbf{y}_{mR:\, mR+S-1}\}_{m\in\{0,\dots,M-1\}}$. To further constrain both parameters and initial conditions to ecologically meaningful values, we use the bijective maps $\mathfrak{P}\colon p \to \mathfrak{p} \in \mathbb{R}^q$ and $\mathfrak{X}\colon x_0 \to \mathfrak{x}_0 \in \mathbb{R}^d$. Given a random minibatch $\mathcal{S} \in \{0,\dots,M-1\}$ (so that $\mathcal{S}$ selects $|\mathcal{S}| = b$ segments out of all segments), we reformulate the loss function in Eq. (7) as

$$
\begin{aligned}
L^{\star}_{\mathcal{M}}(\mathfrak{p}, \mathfrak{x}_0^{(1)}, \dots, \mathfrak{x}_0^{(M)}) &= \sum_{m\in\mathcal{S}} L^{(m)}_{\mathcal{M}}(\mathfrak{p}, \mathfrak{x}_0^{(m)}) \\
&= \sum_{m\in\mathcal{S}} \Big[\textstyle\sum_{k=0}^{S-1} \Big\| \log y_{(m-1)R+k} - \log h\Big(\mathcal{M}\Big(t_{(m-1)R+k}, (\mathfrak{P}^{-1}(\mathfrak{p}), \mathfrak{X}^{-1}(\mathfrak{x}_0^{(m)}))\Big)\Big)\Big\|_{\Sigma_y} \\
&\quad + \Big\| \log y_{(m-1)R} - \log h\Big(\mathfrak{X}^{-1}(\mathfrak{x}_0^{(m)})\Big)\Big\|_{\Sigma_{x_0}} + \|p_b - \mathfrak{P}^{-1}(\mathfrak{p})\|_{\Sigma_p} \Big],
\end{aligned}
\tag{9}
$$

where $(\mathfrak{p}, \mathfrak{x}_0^{(1)}, \mathfrak{x}_0^{(2)}, \dots, \mathfrak{x}_0^{(M)})$ consists of the augmented parameter vector containing M additional ICs transformed in the unconstrained domain and denoted by $x_0^{(m)} = \mathfrak{X}(x_0^{(m)})$, and p_b incorporates the prior knowledge of the parameter values. A minibatch is randomly sampled without replacement and used for updating the parameters and ICs according to the gradient-based optimizer rule. Once all segments have been sampled, marking the end of an epoch, the process repeats with all segments being considered again for the next epoch. In contrast to Eq. (7), the ICs for each segment are constrained by an extra term with weight $\Sigma_{x_0}^{-1}$, which is needed in practice for better convergence. We refer to Fig. 1**C** for a graphical representation of Eq. (9). By smoothing the ruggedness of the loss surface (orange curve in Fig. 1**A**), the proposed segmentation method yields improved navigability (orange curve in Fig. 1**B**). The segmentation method involves three hyper-parameters, namely the mini-batch size b, the segment size S and the segment shift R, which are associated with trade-offs in terms of

accuracy of inference, convergence and computational time. We investigate the effect of these hyper-parameters on inference performance in Section 3.

2.2.2 Numerical implementation of the inverse modelling framework

We implement the inverse modelling framework based on the reformulation of the learning problem in Eq. (9) in the Julia package `PiecewiseInference.jl`. Building upon the software ecosystem SciML [Rackauckas et al., 2020], `PiecewiseInference.jl` relies on `OrdinaryDiffEq.jl` for the forward integration of the process-based community models, as it provides highly efficient ODE solvers [Rackauckas and Nie, 2017]. `OrdinaryDiffEq.jl` is additionally compatible with automatic differentiation and includes an extensive set of sensitivity analysis methods through the `SciMLSensitivity.jl` package [Ma et al., 2021]. This automatic generation of the gradient greatly reduces the effort and potential errors associated with a manual implementation. In addition, it permits to control the numerical accuracy of the gradient computation, which may compromise the convergence of gradient-based optimizers [Gholaminejad et al., 2019]. In the following, we specifically use the reverse-mode automatic differentiation package `Zygote.jl` for our experiments, with custom rules from `SciMLSensitivity.jl`. In our experiments, the number of parameters p and states m is large enough, so that we find the `BacksolveAdjoint` adjoint sensitivity analysis method from `SciMLSensitivity.jl`, with theoretical scaling $\mathcal{O}(b(2m+q))$ [Ma et al., 2021], to perform best. For smaller numbers of parameters and/or states forward sensitivity analysis methods, with scaling $\mathcal{O}(bm(m+q))$, will eventually become more efficient [Ma et al., 2021].

The interoperability of the SciML ecosystem further makes it possible to benefit from the tooling of deep learning libraries such as `Flux.jl` [Innes, 2018] and various nonlinear optimization libraries, providing gradient-based optimizers which are well suited for highly parameterized models [Ruder, 2016]. In particular, `PiecewiseInference.jl` relies on `Optimization.jl` [zot, a], which provides a convenient common interface between an extensive set of optimization libraries. In the following, we specifically use the adaptive, momentum-based Adam optimizer [Kingma and Ba, 2014] implemented in `Optimisers.jl`. `ComponentArrays.jl` is used to conveniently handle model parameters, and `Bijectors.jl` is used in conjunction with `Distributions.jl` [zot, b] for constructing the bijective mappings $\mathfrak{P}$ and $\mathfrak{X}$. We evaluate the performance of the inverse modelling framework and its implementation in `PiecewiseInference.jl` in the next section.

3 Experiments

We evaluate the performance of the inverse modelling framework by considering three food-web ecosystems with increasing complexity. We use reference models to generate the observation data and first evaluate the performance of the inverse modelling framework in recovering the reference model parameters and measuring its forecast skill for varying segment size, mini-batch size, and noise levels. Second, we demonstrate how the proposed inverse modelling framework can leverage neural networks to parametrize ecological processes within a process-based model, allowing the inference of ecological laws from time series data.

3.1 Food-web ecosystems

We consider a generalized $\mathfrak{S}$-species food-web model inspired by [Hastings and Powell, 1991, McCann and Yodzis, 1994b,a, Klebanoff and Hastings, 1994, Post et al., 2000, Åkesson et al., 2021], where species interact through trophic and competitive relationships. Denoting the abundance of species i by N_i, and defining $N = (N_1, \ldots, N_{\mathfrak{S}})$, the generalized community model can be expressed as

$$\frac{1}{N_i}\frac{d}{dt}N_i = r_i(u(t)) - \sum_{j=1}^{\mathfrak{S}} N_j \left[\alpha_{i,j} + [F(N)]_{i,j} - [F(N)]_{j,i}\right]. \tag{10}$$

In Eq. (10), the first two terms capture intrinsic growth rate and intra- and interspecific competition. The last two terms capture growth and loss due to trophic interactions. The per capita growth rate r_i may depend on a time-dependent environmental forcing $u(t)$. The competition coefficient between species i and j is denoted by $\alpha_{i,j}$. The feeding rate $[F(N)]_{i,j}$ of species i on j follows a functional response of type II

$$[F(N)]_{i,j} = \frac{q_{i,j}W_{i,j}}{1 + q_{i,j}H_{i,j}\sum_{k=1}^{\mathfrak{S}} W_{i,k}N_k}, \tag{11}$$

where $q_{i,j}$ and $H_{i,j}$ represent the attack rate and handling time of species i when feeding on species j, respectively.

We specialize Eq. (10) into three community models with increasing complexity, where each model, denoted by $\mathcal{M}_3, \mathcal{M}_5, \mathcal{M}_7$ simulates the dynamics of 3, 5, and 7 species and consists of 9, 18 and 24 parameters, respectively. The food web of each ecosystem is illustrated in Fig. 2. These models have been used as benchmarks for proposed ecosystem forecasting methods [Perretti et al., 2013, Deyle et al., 2016, Ye and Sugihara, 2016] and produce fluctuations that resemble observed ecological time series [Bjørnstad and Grenfell, 2001] and are challenging to forecast [Post et al., 2000]. We set the parameters in the simulations to biologically realistic values proposed by [McCann and Yodzis, 1994a, McCann and Hastings, 1997], ensuring that the dynamics of

the system are chaotic or oscillatory (see Section S1 for details on the numerical parameter values). We scale time relative to the primary producer growth rate, adopting a biologically realistic increase for the primary producer biomass of 100% per day, whereby one unit of time equates to one day.

To generate observation data, we first simulate the dynamics of the ecosystem and then introduce noise by sampling from a log-normal distribution with zero mean and a covariance matrix $\Sigma = r^2 I$, where r represents the noise level. This error model is commonly used in population dynamics [Schartau et al., 2017]. Specifically, for a given time point t and species i, the observed value $y_i(t)$ is calculated as $N_i(t)\exp(\epsilon_i(t))$, where $N_i(t)$ is the true population size and $\epsilon_i(t)$ is the normally distributed noise term. We ensure that transient dynamics are not observed by sampling after a long burn-in time ($t > 500$). We assume that the parameter priors are uniformly distributed. For each scalar parameter p_i, we randomly draw initial parameter estimates from a uniform distribution with range $(0.5\,\tilde{p}_i, 1.5\,\tilde{p}_i)$, where $\tilde{p}_i$ denotes the true parameter value, we constrain the parameter value to the same range, and we constrain IC values to $(1 \cdot 10^{-3}, 5)$. We set $\Sigma_{x_0} = \frac{M^{(1)}}{K^{(1)}}\Sigma_y$, we use Adam with $\gamma = 5 \cdot 10^{-3}$ and $\beta_2 = 0.999$, and set the number of epochs to 500.

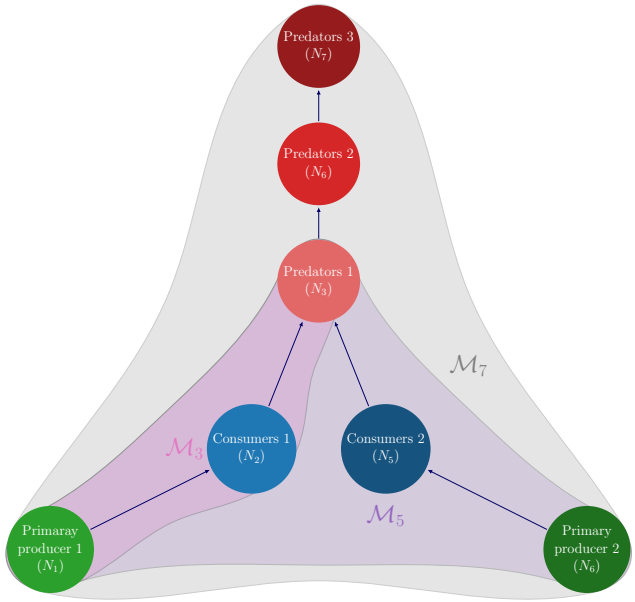

Figure 2: **Food web systems considered.** $\mathcal{M}_3$ corresponds to the 3-species chaotic model presented in [Hastings and Powell, 1991], where a predator group feeds upon a consumer group, itself feeding upon on a primary producer group with type II functional responses. $\mathcal{M}_5$ is an extension of $\mathcal{M}_3$ proposed in [Post et al., 2000], where another food chain is linked to the foodweb through the predator group. $\mathcal{M}_7$ extends $\mathcal{M}_5$ by further considering two additional top predator groups. In Section 3.3, alternative variants of $\mathcal{M}_3$ are also considered, where the primary producer group growth rate depends on an environmental forcing.

3.2 Effect of segment size on parameter, forecast error, and computational complexity

We generate time series from the reference food-web models $\mathcal{M}_3$, $\mathcal{M}_5$, and $\mathcal{M}_7$ (see Fig. 2) under varying noise levels, sampling from the model simulations every 4 days (see Fig. 1**C** for an illustration of the generated data), which provides us with time series with 111 points. The first 101 points of the generated time series are used as the training dataset, so that $K = 100$. For evaluating the performance of the inverse modelling framework, we use the median relative parameter error across the estimated parameters to assess the parameter error,

$$\text{med}\left(\|(\hat{p} - \tilde{p}) \oslash \tilde{p}\|\right), \tag{12}$$

where $\oslash$ denotes the element-wise division. Additionally, we assess the forecast skill, calculated as the squared logarithmic sum of the difference between the last 10 points of the generated time series and the predicted time series, obtained by using the estimated ICs from the last segment of the time series considered,

$$\sum_{k=102}^{111} \left\| \log y_k - \log \mathcal{M}(t_k, (\hat{p}, \hat{x_0}^{(M-1)})) \right\|. \tag{13}$$

These two metrics are complementary: as the food-web dynamics are chaotic (see Section S1), seemingly good performance in parameter estimation might not be sufficient to provide accurate forecasts. The performance of our proposed inverse modelling framework is assessed for varying segment sizes ranging from $S = 2$, corresponding to the minimal segment size, to $S = K$, the maximum segment size. The case where $S = K$ is identical, up to the use of the bijections, to using the naive $L_{\mathcal{M}}$ loss function as outlined in Eq. (7), that we consider as a reference point for our analysis. We use a constant minibatch size of $b = 10$, which we find convenient in terms of memory requirements for this exploration, and a shift $R = S - 1$ so that we have a segment overlap $S - R = 1$.

We consistently find for all models that partitioning the data outperforms the naive $L_{\mathcal{M}}$ loss (Fig. 3 top and middle panels). However, we find that there exists an optimal segment size which maximizes the inference results, both in terms of parameter error and forecast skill. This phenomenon arises as decreasing S tends to flatten the loss surface, which improves its navigability at the cost of decreasing its curvature. As a consequence, the segmentation strategy deteriorates the precision of the inferred parameters because the loss function takes similar values in an extended neighbourhood of the true parameters. Although the optimal segment size appears to be $S = 9$ for all three models and all levels of noise investigated, we expect this optimal size to depend on

the timescale of the system considered. In practice, determining the optimal S for a particular set of model and empirical data may be performed by splitting the dataset into a training and a testing dataset, and choosing S based on the assessed forecast skill (Fig. 3 middle panels). The inverse modelling framework handles increasing levels of observational noise well, and we report similarly good performance in Fig. S1 in a partial observation setting, where only the data on consumer and prey dynamics is used for training the model $\mathcal{M}_3$. In this case, the capacity of our framework to infer ICs is essential, since the IC for the primary producer state variable cannot be estimated from the data alone. We note that as the segment size decreases, the number of ICs increases, so that the augmented dataset becomes larger, which consequently increases the simulation time for one epoch (Fig. 3 bottom panels).

The choice of the minibatch size b is tightly linked to the choice of S. As the number of segments M increases, the memory pressure increases with $\mathcal{O}(M)$, which advocates for decreasing b to mitigate this effect. The consideration of large community models or the disposal of large time series datasets may also advocate for reduced b. Although a smaller minibatch size introduces randomness that can help diminish overfitting [Masters and Luschi, 2018] and may affect simulation time, we did not find a significant impact of minibatch size on either performance metrics or simulation time. Overall, we obtain promising results in estimating parameters and obtaining realistic forecasts from noisy and partial observations with large-scale non-linear community models.

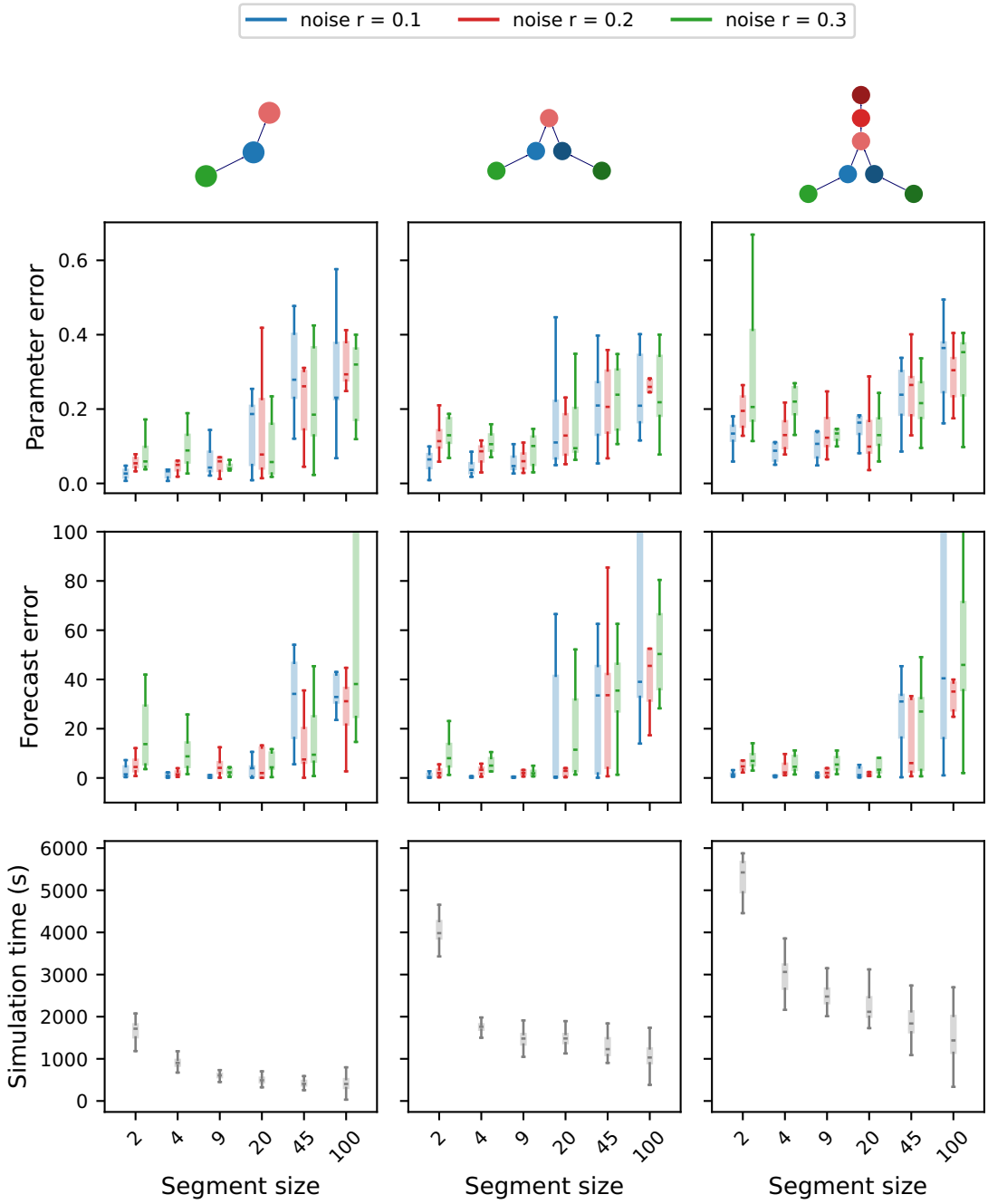

Figure 3: **Effect of segment length S on parameter error (see Eq. (12)), forecast error (see Eq. (13)) and simulation time, for varying noise level and model complexity.** The first column corresponds to results obtained with model $\mathcal{M}_3$, consisting of 3 state variables and 9 parameters. The second column corresponds to results obtained with model $\mathcal{M}_5$, consisting of 5 state variables and 18 parameters. The last column corresponds to results obtained with model $\mathcal{M}_5$, consisting of 7 state variables and 24 parameters. Results for the segment length $S = K = 100$ are similar to using the naive loss $L_{\mathcal{M}}$ in Eq. (7), i.e., without partitioning the data. The top and bottom panels consistently demonstrate that the inverse modelling framework performs better when $S < K$. In particular, there exists an optimal segment length S which maximizes both parameter inference and forecast performance. The bottom panels show the additional computational cost induced by partitioning the data, which is limited for large enough S.

3.3 Learning model parametrization

Neural networks are well suited to represent nonlinear, high-dimensional functions, making them ideal for parametrizing complex ecological processes such as functional responses [Rosenbaum and Rall, 2018], or the dependence of demographic parameters on species traits [Chalmandrier et al., 2022] or on environmental variables [Amarasekare and Johnson, 2017]. Here, we use a neural network to parametrize the growth rate of a primary producer species. Based on a single time series dataset, we use the proposed inverse modelling framework to train the resulting model, whose parameters consist of both the neural network weights together with the additional model parameters, and investigate whether the trained neural network has appropriately recovered the growth rate functional form.

To generate the observation data, we consider a variant of model $\mathcal{M}_3$ that we denote $\mathcal{M}_3^\dagger$, where the primary producer growth rate r_1 depends on a time-varying forcing u, which could e.g. correspond to temperature (see Amarasekare and Johnson [2017]):

$$r_1(u) = \mathfrak{r} e^{-\frac{1}{2}\left(\frac{u}{s}\right)^2} \qquad \text{and} \qquad u(t) = \sin(\omega t) \tag{14}$$

where $\mathfrak{r}$ is the maximum primary producer growth rate, reached when $u = 0$, and $1/s$ is the growth rate sensitivity to the forcing.

We generate multiple observation datasets for different values of $1/s$ with $\mathcal{M}_3^\dagger$, that we contaminate with noise with level $r = 0.1$. We then use the inverse modelling framework to train a model $\mathcal{M}_3^{\text{NN}}$ on the observation data, where the primary producer growth rate is parametrized by a feed-forward neural network NN

$$r_1(u) = \text{NN}(p, u). \tag{15}$$

NN consists of 1 input layer, 1 output layer, and 3 hidden layers with 5 neurons, and Gaussian radial basis functions as activation functions (see Fig. 4 for an illustration of the architecture). Following, NN involves 76 parameters defining the layer weights and biases, resulting in $\mathcal{M}_3^\dagger$ consisting of a total of 85 parameters. Similarly to Section 3.2 and Eq. (13), we split the observation data into a training and a test dataset, and use the ICs inferred on the last segment to generate an out-of-sample forecast. We then evaluate the statistics of the forecast error obtained for $\mathcal{M}_3^{\text{NN}}$ over 10 inference runs with the test dataset, which we compare to that obtained for $\mathcal{M}_3$, the model with constant growth rate, taken as a null model. We set the hyperparameters as $b = 10$, $S = 11$, and $R = S - 1 = 10$. We expect the forecast skill associated with $\mathcal{M}_3^{\text{NN}}$ to be significantly

improved compared to $\mathcal{M}_3$ for high $1/s$.

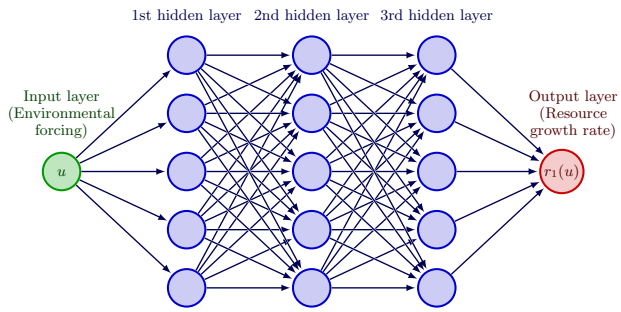

Figure 4: **Architecture of the neural network** NN **in Eq. (15), used in $\mathcal{M}_3^{\mathrm{NN}}$ for parametrizing the primary producer growth rate and its dependence on the environmental forcing** u.

We find that the forecast skill associated with $\mathcal{M}_3^{\mathrm{NN}}$ significantly exceeds that of $\mathcal{M}_3$ for $1/s > 0.4$ (Fig. 4**B**). In such case, the signature of the time varying forcing u on the primary producer growth rate is strong enough so that the more complex model with neural network-based parametrization $\mathcal{M}_3^{\mathrm{NN}}$ performs better, and we expect that the neural network has learned the correct parametrization between u and the primary producer growth rate. A comparison of the growth rate used in the reference model $\mathcal{M}_3^{\dagger}$ against the neural network predictions (Fig. 4**C**) shows that the neural network has adequately recovered the gaussian shape of the relationship, together with its right mean and standard deviation.

Overall, the inverse modelling framework enables the construction of data-driven community models where intricate processes with uncertain mathematical formulations like functional responses are parametrized by neural networks.

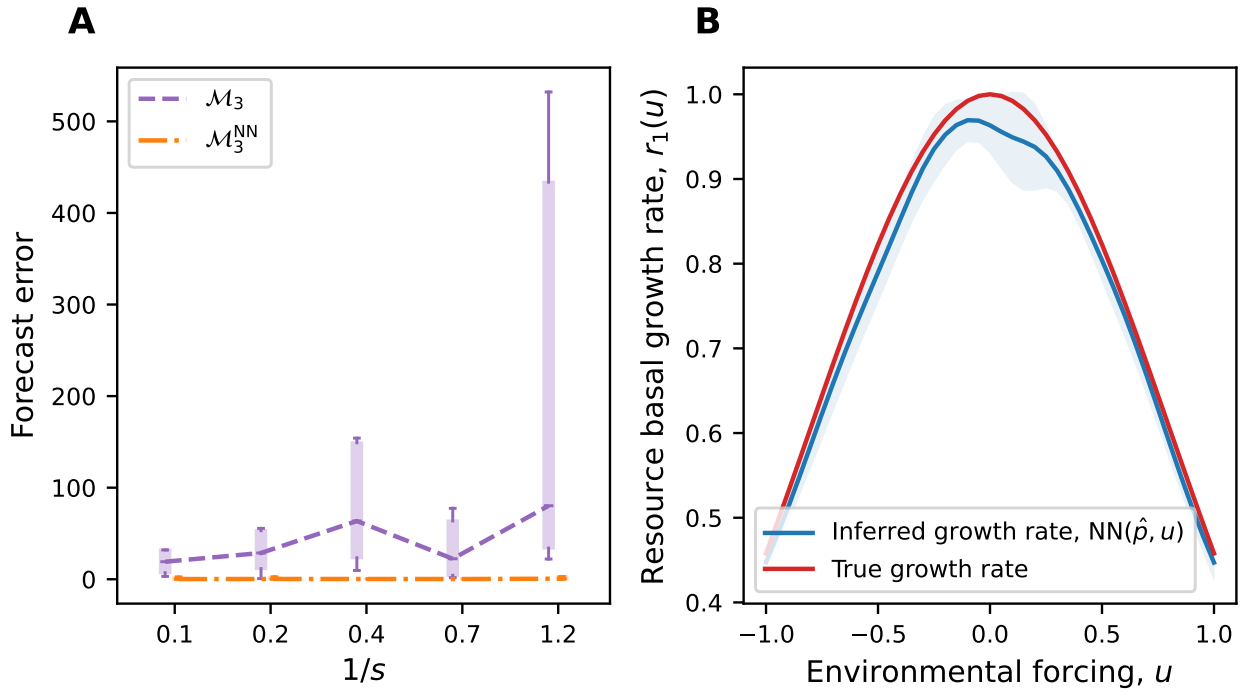

Figure 5: **Neural network-based model parametrization** $\mathcal{M}_3^{\mathrm{NN}}$. **A**. Forecast error for models $\mathcal{M}_3$, which assumes a constant primary producer growth rate, and the neural network-augmented model $\mathcal{M}_3^{\mathrm{NN}}$. Both models are fitted to the time series data generated by the reference model $\mathcal{M}_3^{\dagger}$, for varying strength of environmental forcing sensitivity $1/s$. When $1/s$, which characterises the dependency between environmental forcing u and primary producer growth rate r_1, is sufficiently strong, $\mathcal{M}_3^{\mathrm{NN}}$'s forecast skill significantly outperforms that of $\mathcal{M}_3$. **B**. Offline comparison of the post-training neural network output $\mathrm{NN}(\hat{p}, u)$ against the true primary producer growth rate in Eq. (14) for $1/s = 1.2$. **B** demonstrates that the neural network parametrization NN successfully recovers the dependence between r_1 and u. The blue line corresponds to the average value of $\mathrm{NN}(\hat{p}, u)$ obtained for 10 simulation runs weighted by the loss value, with the shaded blue area indicating one standard deviation. Results displayed were obtained for noise level $r = 0.1$.

4 Discussion

We propose an inverse modelling framework combining a segmentation method [Pisarenko and Sornette, 2004] with automatic differentiation [Rackauckas et al., 2020], deep-learning gradient-based optimizers [Kingma and Ba, 2014], and a parameter normalization technique to efficiently and accurately infer parameters of complex dynamical models. We show that partitioning the data into segments with a short time horizon regularizes the ill-conditioning of the loss function arising from the strong nonlinearities of community models (Fig. 1). This reformulation, together with the use of bijective maps to constrain parameter values to ensure that model parameters and initial conditions (ICs) remain within bounds that are valid both from an ecological and a numerical perspective, allows the use of gradient-based optimization methods to solve the inverse modelling

problem. Gradient-based optimizers enable the calibration of highly parametrized models, including process-based models parametrized with neural networks (Figs. 1, 3 and 5). By automatically generating the gradient of the loss function required by gradient-based optimizers, automatic differentiation streamlines the implementation of the method, simplifying the comparison of multiple models' goodness-of-fits. The segmentation method is relevant beyond variational methods and applies to any method navigating a loss surface, such as evolutionary algorithms [Wilke et al., 2001, Rodriguez-Fernandez et al., 2006] or Markov Chain Monte Carlo methods [Lignell et al., 2013, Higgins et al., 2010, Xu et al., 2006, Fiechter et al., 2013, Rosenbaum et al., 2019]. We observe that our segmentation method resembles state-space modelling approaches (see [Auger-Méthé et al., 2021] and references therein). As such, it should not only regularise the inverse problem and the problem's memory requirements, but also reduce bias arising from observation and process errors [Rosenbaum and Fronhofer, 2023]. Overall, the inverse modelling framework successfully blends ML methods with process-based models to learn from ecological time series. It could therefore improve our quantitative understanding of ecosystem dynamics and help to anticipate their responses to global changes [Urban et al., 2016].

Our work contributes to the ongoing effort to better assimilate observational data into process-based models [Schartau et al., 2017, Raissi et al., 2019, Kashinath et al., 2021, Auger-Méthé et al., 2021, Rosenbaum and Fronhofer, 2023, Paredes et al., 2023, Bolibar et al., 2023], with a specific focus on the parametrization of differential equation-based models with strong nonlinearities. Recently, Yazdani et al. [2020] proposed an alternative framework dubbed "systems biology informed deep learning", where a neural network is fitted to the data and the additional process-based model constraints are integrated. This alternative framework extends previous colocation methods [Ramsay et al., 2007, Cao et al., 2008] and has the advantage of being able to parametrize stochastic models. As it requires the design of a non-trivial neural network architecture to emulate the full system's dynamics and a "goodness of fit" parameter, it nevertheless imposes an additional layer of complexity, which might negatively affect the model parametrization [Yazdani et al., 2020]. In contrast, the inverse modelling framework proposed here trains the model directly against data, using automatic differentiation and sensitivity analysis in order to apply gradient-based optimizers directly to the model simulations. This makes it possible to bypass the use of neural networks for approximating the full dynamics, rendering the parametrization process simpler and more amenable to model selection [Ramsay et al., 2007].

The inverse modelling framework integrates the practical constraints of ecological datasets, accommodating incomplete, noisy, and sparse observation data [Dornelas et al., 2018]. The partial-observation scenario in Fig. S1 is interesting for real-case applications where data may only be available for a subset of the whole pool of species or groups considered. Moreover, since segments are treated independently, the method could allow

us to process and combine information from independent datasets. This ability to harness independent time series embraces the practical constraints imposed by ecological datasets, since local ecosystem surveys, such as marine trawling surveys or local terrestrial surveys (see Pinsky et al. [2013], Dornelas et al. [2018], Burrows et al. [2019] and references therein), are generally shallow in time but composed of many replicates [Hsieh et al., 2008, Clark et al., 2015, Ye and Sugihara, 2016]. Overall, the proposed inverse modelling framework accommodates the characteristics of the current ecological time series and can improve the assimilation of ecological data into process-based community models.

Our work can help elucidate mechanistic pathways by enabling the parametrization of complex models incorporating key ecological and adaptive mechanisms [Urban et al., 2016, Nadeau and Urban, 2019]. These mechanisms are generally associated with strong nonlinearities [Bjørnstad and Grenfell, 2001, Hastings et al., 1993, Huisman and Weissing, 1999, Benincà et al., 2008, Govaert et al., 2019] and must be refined in models to improve predictive accuracy [Urban et al., 2016]. Furthermore, the inverse modelling framework enables the development of process-based models incorporating data-driven components such as neural network-based parametrizations, whose weights can be trained alongside other model parameters. This possibility effectively blends the flexibility of machine learning with process-based models, enhancing the explanatory power of process-based models over observational data. Consequently, the inverse modelling framework, along with the user-friendly Julia package PiecewiseInference.jl, promise to improve the representation of key ecological processes [Schartau et al., 2017], and will allow researchers to test ecological hypotheses by confronting process-based models with empirical data [Curtsdotter et al., 2019, Becks et al., 2010].

The proposed approach still presents a set of limitations, which might hamper its success under specific situations. First, while the use of segments smooths the loss surface and ensures better convergence, it also deteriorates the precision of the inferred parameter values. To circumvent this issue, iterative training can be performed, where the learning is initiated by a short segment length K to identify the region with the most probable parameters, and in subsequent iterations the segment length is increased while the learning rate is decreased, to improve the precision of the inference. Second, the data might not provide enough constraints for a correct parametrization. Pre-experimental analyses with simulated synthetic data might therefore be required to ensure an adequate sampling effort [Banks et al., 2017, Laubmeier et al., 2018]. Alternatively, one could decide on reducing the model complexity to adapt the available information content of the time series. Third, while in Eq. (9) it is assumed that the parameter values are the same across the time series, strong geographic or temporal variability might also be observed among the spatially replicated data, causing parameter values to vary across the replicates. The knowledge of this variability could motivate partial pooling [Beaumont,

2010] or parametrization of the parameters in terms of environmental conditions or space [Pahlow et al., 2008, Bolibar et al., 2023]. Finally, as with any local optimization approach, it cannot be guaranteed that the inferred parameters correspond to the true MAP. Performing multiple runs with varying initial parameter estimates can help ensure that the parameter estimates are reliable.

5 Conclusion

We proposed a scientific machine learning framework designed for realistic community models characterized by highly non-linear dynamics. This framework not only addresses the unique challenges posed by ecological datasets but also mitigates inaccuracies in current ecological models by incorporating data-driven parameterization through neural networks. Together with the increasing number of ecological datasets following the development of monitoring technologies such as environmental DNA [Ruppert et al., 2019], remote sensing [Jetz et al., 2019], bioacoustics [Aide et al., 2013], and citizen observations [GBIF: The Global Biodiversity Information Facility, 2022], the inverse modelling framework and its implementation in the Julia package `PiecewiseInference.jl` open up new avenues for leveraging observational data to improve our quantitative understanding of ecosystem processes [Curtsdotter et al., 2019] and predictions of future ecosystem trajectories [Urban et al., 2016].

6 Acknowledgements

L.P. and V.B. were supported by the SNF grant 310030E_205556. P.V.A. was supported by an ETH postdoctoral fellowship.

7 Competing interests

The authors declare no competing interests.

8 Code availability

The inverse modelling framework is implemented in the Julia package `PiecewiseInference.jl` available at https://github.com/vboussange/PiecewiseInference.jl, and the simulation code is available at https://github.com/vboussange/partitioning-time-series.

9 Contributions

Conceptualization V.B. and P.V.

Formal Analysis V.B., P.V. and F.S.

Funding Acquisition P.V. and L.P.

Investigation V.B. and P.V.

Methodology V.B. and F.S.

Software V.B. and F.S.

Validation V.B.

Visualization V.B.

Writing – Original Draft Preparation V.B., P.V., F.S., and L.P.

References

Optimization.jl: A Unified Optimization Package. a. doi: 10.5281/zenodo.7738525.

JuliaStats/Distributions.jl: V0.25.103. b. doi: 10.5281/zenodo.10070405.

T. Mitchell Aide, Carlos Corrada-Bravo, Marconi Campos-Cerqueira, Carlos Milan, Giovany Vega, and Rafael Alvarez. Real-time bioacoustics monitoring and automated species identification. *PeerJ*, 1(1):e103, July 2013. ISSN 2167-8359. doi: 10.7717/peerj.103.

Anna Åkesson, Alva Curtsdotter, Anna Eklöf, Bo Ebenman, Jon Norberg, and György Barabás. The importance of species interactions in eco-evolutionary community dynamics under climate change. *Nature Communications*, 12(1):4759, December 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-24977-x.

Mark Alber, Adrian Buganza Tepole, William R. Cannon, Suvranu De, Salvador Dura-Bernal, Krishna Garikipati, George Karniadakis, William W. Lytton, Paris Perdikaris, Linda Petzold, and Ellen Kuhl. Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *npj Digital Medicine*, 2(1):115, December 2019. ISSN 2398-6352. doi: 10.1038/s41746-019-0193-y.

Priyanga Amarasekare and Christopher Johnson. Evolution of Thermal Reaction Norms in Seasonally Varying Environments. *The American Naturalist*, 189(3):E31–E45, March 2017. ISSN 0003-0147, 1537-5323. doi: 10.1086/690293.

Marie Auger-Méthé, Ken Newman, Diana Cole, Fanny Empacher, Rowenna Gryba, Aaron A. King, Vianey Leos-Barajas, Joanna Mills Flemming, Anders Nielsen, Giovanni Petris, and Len Thomas. A guide to state–space modeling of ecological time series. *Ecological Monographs*, 91(4):e01470, 2021. ISSN 1557-7015. doi: 10.1002/ecm.1470.

H.T. Banks, J.E. Banks, R. Bommarco, A. Curtsdotter, T. Jonsson, and A.N. Laubmeier. PARAMETER ESTIMATION FOR AN ALLOMETRIC FOOD WEB MODEL. *International Journal of Pure and Apllied Mathematics*, 114(1):143–160, April 2017. ISSN 1311-8080. doi: 10.12732/ijpam.v114i1.12.

Anthony D. Barnosky, Elizabeth A. Hadly, Jordi Bascompte, Eric L. Berlow, James H. Brown, Mikael Fortelius, Wayne M. Getz, John Harte, Alan Hastings, Pablo A. Marquet, Neo D. Martinez, Arne Mooers, Peter Roopnarine, Geerat Vermeij, John W. Williams, Rosemary Gillespie, Justin Kitzes, Charles Marshall, Nicholas Matzke, David P. Mindell, Eloy Revilla, and Adam B. Smith. Approaching a state shift in Earth's biosphere. *Nature*, 486(7401):52–58, 2012. ISSN 00280836. doi: 10.1038/nature11018.

Mark A. Beaumont. Approximate Bayesian Computation in Evolution and Ecology. *Annual Review of Ecology, Evolution, and Systematics*, 41(1):379–406, 2010. ISSN 1543-592X. doi: 10.1146/annurev-ecolsys-102209-144621.

Lutz Becks, Stephen P. Ellner, Laura E. Jones, and Jr G. Hairston Nelson G. Reduction of adaptive genetic diversity radically alters eco-evolutionary community dynamics. *Ecology Letters*, 13(8):989–997, May 2010. ISSN 14610248. doi: 10.1111/j.1461-0248.2010.01490.x.

Elisa Benincà, Jef Huisman, Reinhard Heerkloss, Klaus D. Jöhnk, Pedro Branco, Egbert H. Van Nes, Marten Scheffer, and Stephen P. Ellner. Chaos in a long-term experiment with a plankton community. *Nature*, 451 (7180):822–825, February 2008. ISSN 0028-0836. doi: 10.1038/nature06512.

Laurent Bertino, Geir Evensen, and Hans Wackernagel. Sequential Data Assimilation Techniques in Oceanography. *International Statistical Review*, 71(2):223–241, August 2003. ISSN 0306-7734. doi: 10.1111/j.1751-5823.2003.tb00194.x.

Ottar N. Bjørnstad and Bryan T. Grenfell. Noisy Clockwork: Time Series Analysis of Population Fluctuations in Animals. *Science*, 293(5530):638–643, July 2001. ISSN 0036-8075. doi: 10.1126/science.1062226.

Marc Bocquet, Julien Brajard, Alberto Carrassi, and Laurent Bertino. Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models. *Nonlinear Processes in Geophysics*, 26 (3):143–162, July 2019. ISSN 1607-7946. doi: 10.5194/npg-26-143-2019.

Jordi Bolibar, Facundo Sapienza, Fabien Maussion, Redouane Lguensat, Bert Wouters, and Fernando Pérez. Universal differential equations for glacier ice flow modelling. *Geoscientific Model Development*, 16(22):6671–6687, November 2023. ISSN 1991-959X. doi: 10.5194/gmd-16-6671-2023.

Ian L. Boyd. The Art of Ecological Modeling. *Science*, 337(6092):306–307, July 2012. ISSN 0036-8075. doi: 10.1126/science.1225049.

Natalie J. Briscoe, Jane Elith, Roberto Salguero-Gómez, José J. Lahoz-Monfort, James S. Camac, Katherine M. Giljohann, Matthew H. Holden, Bronwyn A. Hradsky, Michael R. Kearney, Sean M. McMahon, Ben L. Phillips, Tracey J. Regan, Jonathan R. Rhodes, Peter A. Vesk, Brendan A. Wintle, Jian D.L. Yen, and Gurutzeta Guillera-Arroita. Forecasting species range dynamics with process-explicit models: Matching methods to applications. *Ecology Letters*, 22(11):1940–1956, 2019. ISSN 1461-0248. doi: 10.1111/ele.13348.

Michael T. Burrows, Amanda E. Bates, Mark J. Costello, Martin Edwards, Graham J. Edgar, Clive J. Fox, Benjamin S. Halpern, Jan G. Hiddink, Malin L. Pinsky, Ryan D. Batt, Jorge García Molinos, Benjamin L. Payne, David S. Schoeman, Rick D. Stuart-Smith, and Elvira S. Poloczanska. Ocean community warming responses explained by thermal affinities and temperature gradients. *Nature Climate Change*, 9(12):959–963, December 2019. ISSN 1758-678X. doi: 10.1038/s41558-019-0631-5.

Juliano Sarmento Cabral, Luis Valente, and Florian Hartig. Mechanistic simulation models in macroecology and biogeography: State-of-art and prospects. *Ecography*, 40(2):267–280, 2017. ISSN 1600-0587. doi: 10.1111/ecog.02480.

Jiguo Cao, Gregor F. Fussmann, and James O. Ramsay. Estimating a Predator-Prey Dynamical Model with the Parameter Cascades Method. *Biometrics*, 64(3):959–967, September 2008. ISSN 0006-341X. doi: 10.1111/j.1541-0420.2007.00942.x.

Loïc Chalmandrier, Daniel B. Stouffer, Adam S. T. Purcell, William G. Lee, Andrew J. Tanentzap, and Daniel C.

Laughlin. Predictions of biodiversity are improved by integrating trait-based competition with abiotic filtering. *Ecology Letters*, 25(5):1277–1289, 2022. ISSN 1461-0248. doi: 10.1111/ele.13980.

Adam Thomas Clark, Hao Ye, Forest Isbell, Ethan R. Deyle, Jane Cowles, G. David Tilman, George Sugihara, and B. D. Inouye. Spatial convergent cross mapping to detect causal relationships from short time series. *Ecology*, 96(5):1174–1181, 2015. ISSN 00129658. doi: 10.1890/14-1479.1.

Alva Curtsdotter, H. Thomas Banks, John E. Banks, Mattias Jonsson, Tomas Jonsson, Amanda N. Laubmeier, Michael Traugott, and Riccardo Bommarco. Ecosystem function in predator–prey food webs—confronting dynamic models with empirical data. *Journal of Animal Ecology*, 88(2):196–210, February 2019. ISSN 0021-8790. doi: 10.1111/1365-2656.12892.

Donald L. DeAngelis and Simeon Yurek. Equation-free modeling unravels the behavior of complex ecological systems. *Proceedings of the National Academy of Sciences*, 112(13):3856–3857, March 2015. ISSN 0027-8424. doi: 10.1073/pnas.1503154112.

John P. DeLong, Torrance C. Hanley, and David A. Vasseur. Predator-prey dynamics and the plasticity of predator body size. *Functional Ecology*, 28(2):487–493, April 2014. ISSN 02698463. doi: 10.1111/1365-2435.12199.

Benjamin Deneu, Maximilien Servajean, Pierre Bonnet, Christophe Botella, François Munoz, and Alexis Joly. Convolutional neural networks improve species distribution modelling by capturing the spatial structure of the environment. *PLOS Computational Biology*, 17(4):e1008856, April 2021. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1008856.

Ethan R. Deyle, Robert M. May, Stephan B. Munch, and George Sugihara. Tracking and forecasting ecosystem interactions in real time. *Proceedings of the Royal Society B: Biological Sciences*, 283(1822):20152258, January 2016. ISSN 0962-8452. doi: 10.1098/rspb.2015.2258.

Maria Dornelas, Laura H. Antão, Faye Moyes, Amanda E. Bates, Anne E. Magurran, Dušan Adam, Asem A. Akhmetzhanova, Ward Appeltans, José Manuel Arcos, Haley Arnold, Narayanan Ayyappan, Gal Badihi, Andrew H. Baird, Miguel Barbosa, Tiago Egydio Barreto, Claus Bässler, Alecia Bellgrove, Jonathan Belmaker, Lisandro Benedetti-Cecchi, Brian J. Bett, Anne D. Bjorkman, Magdalena Błażewicz, Shane A. Blowes, Christopher P. Bloch, Timothy C. Bonebrake, Susan Boyd, Matt Bradford, Andrew J. Brooks, James H. Brown, Helge Bruelheide, Phaedra Budy, Fernando Carvalho, Edward Castañeda-Moya, Chaolun Allen Chen, John F.

Chamblee, Tory J. Chase, Laura Siegwart Collier, Sharon K. Collinge, Richard Condit, Elisabeth J. Cooper, J. Hans C. Cornelissen, Unai Cotano, Shannan Kyle Crow, Gabriella Damasceno, Claire H. Davies, Robert A. Davis, Frank P. Day, Steven Degraer, Tim S. Doherty, Timothy E. Dunn, Giselda Durigan, J. Emmett Duffy, Dor Edelist, Graham J. Edgar, Robin Elahi, Sarah C. Elmendorf, Anders Enemar, S. K. Morgan Ernest, Rubén Escribano, Marc Estiarte, Brian S. Evans, Tung-Yung Fan, Fabiano Turini Farah, Luiz Loureiro Fernandes, Fábio Z. Farneda, Alessandra Fidelis, Robert Fitt, Anna Maria Fosaa, Geraldo Antonio Daher Correa Franco, Grace E. Frank, William R. Fraser, Hernando García, Roberto Cazzolla Gatti, Or Givan, Elizabeth Gorgone-Barbosa, William A. Gould, Corinna Gries, Gary D. Grossman, Julio R. Gutierréz, Stephen Hale, Mark E. Harmon, John Harte, Gary Haskins, Donald L. Henshaw, Luise Hermanutz, Pamela Hidalgo, Pedro Higuchi, Andrew Hoey, Gert Van Hoey, Annika Hofgaard, Kristen Holeck, Robert D. Hollister, Richard Holmes, Mia Hoogenboom, Chih-hao Hsieh, Stephen P. Hubbell, Falk Huettmann, Christine L. Huffard, Allen H. Hurlbert, Natália Macedo Ivanauskas, David Janík, Ute Jandt, Anna Jażdżewska, Tore Johannessen, Jill Johnstone, Julia Jones, Faith A. M. Jones, Jungwon Kang, Tasrif Kartawijaya, Erin C. Keeley, Douglas A. Kelt, Rebecca Kinnear, Kari Klanderud, Halvor Knutsen, Christopher C. Koenig, Alessandra R. Kortz, Kamil Král, Linda A. Kuhnz, Chao-Yang Kuo, David J. Kushner, Claire Laguionie-Marchais, Lesley T. Lancaster, Cheol Min Lee, Jonathan S. Lefcheck, Esther Lévesque, David Lightfoot, Francisco Lloret, John D. Lloyd, Adrià López-Baucells, Maite Louzao, Joshua S. Madin, Borgþór Magnússon, Shahar Malamud, Iain Matthews, Kent P. McFarland, Brian McGill, Diane McKnight, William O. McLarney, Jason Meador, Peter L. Meserve, Daniel J. Metcalfe, Christoph F. J. Meyer, Anders Michelsen, Nataliya Milchakova, Tom Moens, Even Moland, Jon Moore, Carolina Mathias Moreira, Jörg Müller, Grace Murphy, Isla H. Myers-Smith, Randall W. Myster, Andrew Naumov, Francis Neat, James A. Nelson, Michael Paul Nelson, Stephen F. Newton, Natalia Norden, Jeffrey C. Oliver, Esben M. Olsen, Vladimir G. Onipchenko, Krzysztof Pabis, Robert J. Pabst, Alain Paquette, Sinta Pardede, David M. Paterson, Raphaël Pélissier, Josep Peñuelas, Alejandro Pérez-Matus, Oscar Pizarro, Francesco Pomati, Eric Post, Herbert H. T. Prins, John C. Priscu, Pieter Provoost, Kathleen L. Prudic, Erkki Pulliainen, B. R. Ramesh, Olivia Mendivil Ramos, Andrew Rassweiler, Jose Eduardo Rebelo, Daniel C. Reed, Peter B. Reich, Suzanne M. Remillard, Anthony J. Richardson, J. Paul Richardson, Itai van Rijn, Ricardo Rocha, Victor H. Rivera-Monroy, Christian Rixen, Kevin P. Robinson, Ricardo Ribeiro Rodrigues, Denise de Cerqueira Rossa-Feres, Lars Rudstam, Henry Ruhl, Catalina S. Ruz, Erica M. Sampaio, Nancy Rybicki, Andrew Rypel, Sofia Sal, Beatriz Salgado, Flavio A. M. Santos, Ana Paula Savassi-Coutinho, Sara Scanga, Jochen Schmidt, Robert Schooley, Fakhrizal Setiawan, Kwang-Tsao Shao, Gaius R. Shaver, Sally Sherman, Thomas W. Sherry, Jacek Siciński, Caya Sievers, Ana Carolina da Silva, Fernando Rodrigues da Silva, Fabio L. Silveira, Jasper Slingsby,

Tracey Smart, Sara J. Snell, Nadejda A. Soudzilovskaia, Gabriel B. G. Souza, Flaviana Maluf Souza, Vinícius Castro Souza, Christopher D. Stallings, Rowan Stanforth, Emily H. Stanley, José Mauro Sterza, Maarten Stevens, Rick Stuart-Smith, Yzel Rondon Suarez, Sarah Supp, Jorge Yoshio Tamashiro, Sukmaraharja Tarigan, Gary P. Thiede, Simon Thorn, Anne Tolvanen, Maria Teresa Zugliani Toniato, Ørjan Totland, Robert R. Twilley, Gediminas Vaitkus, Nelson Valdivia, Martha Isabel Vallejo, Thomas J. Valone, Carl Van Colen, Jan Vanaverbeke, Fabio Venturoli, Hans M. Verheye, Marcelo Vianna, Rui P. Vieira, Tomáš Vrška, Con Quang Vu, Lien Van Vu, Robert B. Waide, Conor Waldock, Dave Watts, Sara Webb, Tomasz Wesołowski, Ethan P. White, Claire E. Widdicombe, Dustin Wilgers, Richard Williams, Stefan B. Williams, Mark Williamson, Michael R. Willig, Trevor J. Willis, Sonja Wipf, Kerry D. Woods, Eric J. Woehler, Kyle Zawada, Michael L. Zettler, and Thomas Hickler. BioTIME: A database of biodiversity time series for the Anthropocene. *Global Ecology and Biogeography*, 27(7):760–786, July 2018. ISSN 1466-822X. doi: 10.1111/geb.12729.

Maéva Doron, Pierre Brasseur, Jean Michel Brankart, Svetlana N. Losa, and Angélique Melet. Stochastic estimation of biogeochemical parameters from Globcolour ocean colour satellite data in a North Atlantic 3D ocean coupled physical-biogeochemical model. *Journal of Marine Systems*, 117–118:81–95, 2013. ISSN 09247963. doi: 10.1016/j.jmarsys.2013.02.007.

Margaret E. K. Evans, Cory Merow, Sydne Record, Sean M. McMahon, and Brian J. Enquist. Towards Process-based Range Modeling of Many Species. *Trends in Ecology & Evolution*, 31(11):860–871, November 2016. ISSN 0169-5347. doi: 10.1016/j.tree.2016.08.005.

Katja Fennel, Martin Losch, Jens Schröter, and Manfred Wenzel. Testing a marine ecosystem model: Sensitivity analysis and parameter optimization. *Journal of Marine Systems*, 28(1-2):45–63, February 2001. ISSN 09247963. doi: 10.1016/S0924-7963(00)00083-X.

J. Fiechter, R. Herbei, W. Leeds, J. Brown, R. Milliff, C. Wikle, A. Moore, and T. Powell. A Bayesian parameter estimation method applied to a marine ecosystem model for the coastal Gulf of Alaska. *Ecological Modelling*, 258:122–133, June 2013. ISSN 03043800. doi: 10.1016/j.ecolmodel.2013.03.003.

Rosie A. Fisher, Charles D. Koven, William R.L. Anderegg, Bradley O. Christoffersen, Michael C. Dietze, Caroline E. Farrior, Jennifer A. Holm, George C. Hurtt, Ryan G. Knox, Peter J. Lawrence, Jeremy W. Lichstein, Marcos Longo, Ashley M. Matheny, David Medvigy, Helene C. Muller-Landau, Thomas L. Powell, Shawn P. Serbin, Hisashi Sato, Jacquelyn K. Shuman, Benjamin Smith, Anna T. Trugman, Toni Viskari, Hans Verbeeck, Ensheng Weng, Chonggang Xu, Xiangtao Xu, Tao Zhang, and Paul R. Moorcroft. Vegetation demographics

in Earth System Models: A review of progress and priorities. *Global Change Biology*, 24(1):35–54, 2018. ISSN 13652486. doi: 10.1111/gcb.13910.

Damien A. Fordham, Cleo Bertelsmeier, Barry W. Brook, Regan Early, Dora Neto, Stuart C. Brown, Sébastien Ollier, and Miguel B. Araújo. How complex should models be? Comparing correlative and mechanistic range dynamics models. *Global Change Biology*, 24(3):1357–1370, 2018. ISSN 1365-2486. doi: 10.1111/gcb.13935.

Attila Gábor and Julio R. Banga. Robust and efficient parameter estimation in dynamic models of biological systems. *BMC Systems Biology*, 9(1):74, December 2015. ISSN 1752-0509. doi: 10.1186/s12918-015-0219-2.

GBIF: The Global Biodiversity Information Facility. What is GBIF? 2022.

William L. Geary, Michael Bode, Tim S. Doherty, Elizabeth A. Fulton, Dale G. Nimmo, Ayesha I. T. Tulloch, Vivitskaia J. D. Tulloch, and Euan G. Ritchie. A guide to ecosystem models and their environmental applications. *Nature Ecology & Evolution*, 4(11):1459–1471, November 2020. ISSN 2397-334X. doi: 10.1038/s41559-020-01298-8.

M. Gehlen, R. Barciela, L. Bertino, P. Brasseur, M. Butenschön, F. Chai, A. Crise, Y. Drillet, D. Ford, D. Lavoie, P. Lehodey, C. Perruche, A. Samuelsen, and E. Simon. Building the capacity for forecasting marine biogeochemistry and ecosystems: Recent advances and future developments. *Journal of Operational Oceanography*, 8 (sup1):s168–s187, April 2015. ISSN 1755-876X. doi: 10.1080/1755876X.2015.1022350.

Wendy Gentleman, Andrew Leising, Bruce Frost, Suzanne Strom, and James Murray. Functional responses for zooplankton feeding on multiple resources: A review of assumptions and biological dynamics. *Deep Sea Research Part II: Topical Studies in Oceanography*, 50(22-26):2847–2875, November 2003. ISSN 09670645. doi: 10.1016/j.dsr2.2003.07.001.

M.E. Gharamti, J. Tjiputra, I. Bethke, A. Samuelsen, I. Skjelvan, M. Bentsen, and L. Bertino. Ensemble data assimilation for ocean biogeochemical state and parameter estimation at different sites. *Ocean Modelling*, 112: 65–89, April 2017. ISSN 14635003. doi: 10.1016/j.ocemod.2017.02.006.

Amir Gholaminejad, Kurt Keutzer, and George Biros. ANODE: Unconditionally Accurate Memory-Efficient Gradients for Neural ODEs. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, volume 2019-Augus, pages 730–736, California, August 2019. International Joint Conferences on Artificial Intelligence Organization. ISBN 978-0-9992411-4-1. doi: 10.24963/ijcai.2019/103.

Sanmitra Gosh, Paul Birrell, and Daniela De Angelis. Variational inference for nonlinear ordinary differential equations. *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, 130(29): 2719–2727, 2021.

Lynn Govaert, Emanuel A. Fronhofer, Sébastien Lion, Christophe Eizaguirre, Dries Bonte, Martijn Egas, Andrew P. Hendry, Ayana De Brito Martins, Carlos J. Melián, Joost A.M. Raeymaekers, Irja I. Ratikainen, Bernt Erik Saether, Jennifer A. Schweitzer, and Blake Matthews. Eco-evolutionary feedbacks—Theoretical models and perspectives. *Functional Ecology*, 33(1):13–30, 2019. ISSN 13652435. doi: 10.1111/1365-2435.13241.

Antoine Guisan and Niklaus E. Zimmermann. Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2):147–186, December 2000. ISSN 0304-3800. doi: 10.1016/S0304-3800(00)00354-9.

Ryan N. Gutenkunst, Joshua J. Waterfall, Fergal P. Casey, Kevin S. Brown, Christopher R. Myers, and James P. Sethna. Universally Sloppy Parameter Sensitivities in Systems Biology Models. *PLoS Computational Biology*, 3(10):e189, October 2007. ISSN 1553-7358. doi: 10.1371/journal.pcbi.0030189.

Florian Hartig, James Dyke, Thomas Hickler, Steven I. Higgins, Robert B. O'Hara, Simon Scheiter, and Andreas Huth. Connecting dynamic vegetation models to data - an inverse perspective: Dynamic vegetation models - an inverse perspective. *Journal of Biogeography*, 39(12):2240–2252, December 2012. ISSN 03050270. doi: 10.1111/j.1365-2699.2012.02745.x.

Alan Hastings and Thomas Powell. Chaos in a Three-Species Food Chain. *Ecology*, 72(3):896–903, June 1991. ISSN 00129658. doi: 10.2307/1940591.

Alan Hastings, Carole L. Hom, Stephen Ellner, Peter Turchin, and H. Charles J. Godfray. Chaos in Ecology: Is Mother Nature a Strange Attractor? *Annual Review of Ecology and Systematics*, 24(1):1–33, November 1993. ISSN 0066-4162. doi: 10.1146/annurev.es.24.110193.000245.

Lukas Heiland, Georges Kunstler, Vladimír Šebeň, and Lisa Hülsmann. Which demographic processes control competitive equilibria? Bayesian calibration of a size-structured forest population model. *Ecology and Evolution*, 13(7):e10232, 2023. ISSN 2045-7758. doi: 10.1002/ece3.10232.

Steven I. Higgins, Simon Scheiter, and Mahesh Sankaran. The stability of African savannas: Insights from the indirect estimation of the parameters of a dynamic model. *Ecology*, 91(6):1682–1692, June 2010. ISSN 0012-9658. doi: 10.1890/08-1368.1.

Chih-hao Hsieh, Christian Anderson, and George Sugihara. Extending Nonlinear Analysis to Short Ecological Time Series. *The American Naturalist*, 171(1):71–80, January 2008. ISSN 0003-0147. doi: 10.1086/524202.

Jef Huisman and Franz J. Weissing. Biodiversity of plankton by species oscillations and chaos. *Nature*, 402 (6760):407–410, November 1999. ISSN 0028-0836. doi: 10.1038/46540.

Mike Innes. Flux: Elegant machine learning with Julia. *Journal of Open Source Software*, 3(25):602, May 2018. ISSN 2475-9066. doi: 10.21105/joss.00602.

IPBES. Summary for policymakers of the global assessment report on biodiversity and ecosystem services. Technical report, Zenodo, November 2019.

Walter Jetz, Melodie A. McGeoch, Robert Guralnick, Simon Ferrier, Jan Beck, Mark J. Costello, Miguel Fernandez, Gary N. Geller, Petr Keil, Cory Merow, Carsten Meyer, Frank E. Muller-Karger, Henrique M. Pereira, Eugenie C. Regan, Dirk S. Schmeller, and Eren Turak. Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology and Evolution*, 3(4):539–551, 2019. ISSN 2397334X. doi: 10.1038/s41559-019-0826-1.

K. Kashinath, M. Mustafa, A. Albert, J. L. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, R. Wang, A. Chattopadhyay, A. Singh, A. Manepalli, D. Chirila, R. Yu, R. Walters, B. White, H. Xiao, H. A. Tchelepi, P. Marcus, A. Anandkumar, P. Hassanzadeh, and Prabhat. Physics-informed machine learning: Case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194), 2021. ISSN 1364503X. doi: 10.1098/rsta.2020.0093.

Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. pages 1–15, December 2014.

Aaron Klebanoff and Alan Hastings. Chaos in three species food chains. *Journal of Mathematical Biology*, 32 (5):427–451, May 1994. ISSN 0303-6812. doi: 10.1007/BF00160167.

Andreas Köpf, Yannic Kilcher, Dimitri von Rütte, Sotiris Anagnostidis, Zhi-Rui Tam, Keith Stevens, Abdullah Barhoum, Nguyen Minh Duc, Oliver Stanley, Richárd Nagyfi, Shahul ES, Sameer Suri, David Glushkov, Arnav Dantuluri, Andrew Maguire, Christoph Schuhmann, Huu Nguyen, and Alexander Mattick. OpenAssistant Conversations – Democratizing Large Language Model Alignment, October 2023.

A. N. Laubmeier, Kate Wootton, J. E. Banks, Riccardo Bommarco, Alva Curtsdotter, Tomas Jonsson, Tomas

Roslin, and H. T. Banks. From theory to experimental design—Quantifying a trait-based theory of predator-prey dynamics. *PLOS ONE*, 13(4):e0195919, April 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0195919.

Linda M. Lawson, Yvette H. Spitz, Eileen E. Hofmann, and Robert Bryan Long. A data assimilation technique applied to a predator-prey model. *Bulletin of Mathematical Biology*, 57(4):593–617, July 1995. ISSN 0092-8240. doi: 10.1007/BF02460785.

Risto Lignell, Heikki Haario, Marko Laine, and T. Frede Thingstad. Getting the "right" parameter values for models of the pelagic microbial food web. *Limnology and Oceanography*, 58(1):301–313, January 2013. ISSN 00243590. doi: 10.4319/lo.2013.58.1.0301.

Yingbo Ma, Vaibhav Dixit, Michael J. Innes, Xingjian Guo, and Chris Rackauckas. A Comparison of Automatic Differentiation and Continuous Sensitivity Analysis for Derivatives of Differential Equation Solutions. In *2021 IEEE High Performance Extreme Computing Conference (HPEC)*, number 2, pages 1–9. IEEE, September 2021. ISBN 978-1-66542-369-4. doi: 10.1109/HPEC49654.2021.9622796.

Dominic Masters and Carlo Luschi. Revisiting Small Batch Training for Deep Neural Networks, April 2018.

Richard J. Matear. Parameter optimization and analysis of ecosystem models using simulated annealing: A case study at Station P. *Journal of Marine Research*, 53(4):571–607, July 1995. ISSN 00222402. doi: 10.1357/0022240953213098.

Kevin McCann and Alan Hastings. Re–evaluating the omnivory–stability relationship in food webs. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1385):1249–1254, August 1997. ISSN 0962-8452. doi: 10.1098/rspb.1997.0172.

Kevin McCann and Peter Yodzis. Biological Conditions for Chaos in a Three-Species Food Chain. *Ecology*, 75 (2):561–564, March 1994a. ISSN 00129658. doi: 10.2307/1939558.

Kevin McCann and Peter Yodzis. Nonlinear Dynamics and Population Disappearances. *The American Naturalist*, 144(5):873–879, November 1994b. ISSN 0003-0147. doi: 10.1086/285714.

Christopher P. Nadeau and Mark C. Urban. Eco-evolution on the edge during climate change. *Ecography*, 42 (7):1280–1297, 2019. ISSN 1600-0587. doi: 10.1111/ecog.04404.

Jon Norberg, Mark C. Urban, Mark Vellend, Christopher A. Klausmeier, and Nicolas Loeuille. Eco-evolutionary responses of biodiversity to climate change. *Nature Climate Change*, 2(10):747–751, October 2012. ISSN 1758-678X. doi: 10.1038/nclimate1588.

Jörn Pagel and Frank M. Schurr. Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecology and Biogeography*, 21(2):293–304, 2012. ISSN 1466822X. doi: 10.1111/j.1466-8238.2011.00663.x.

Markus Pahlow, Alain F. Vézina, Benoit Casault, Heidi Maass, Louise Malloch, Daniel G. Wright, and Youyu Lu. Adaptive model of plankton dynamics for the North Atlantic. *Progress in Oceanography*, 76(2):151–191, 2008. ISSN 00796611. doi: 10.1016/j.pocean.2007.11.001.

Josefa Arán Paredes, Koen Hufkens, and Benjamin D. Stocker. Rsofun: A model-data integration framework for simulating ecosystem processes, November 2023.

Joanna S. Pelc, Ehouarn Simon, Laurent Bertino, Ghada El Serafy, and Arnold W. Heemink. Application of model reduced 4D-Var to a 1D ecosystem model. *Ocean Modelling*, 57–58:43–58, November 2012. ISSN 14635003. doi: 10.1016/j.ocemod.2012.09.003.

Grace C. Y. Peng, Mark Alber, Adrian Buganza Tepole, William R. Cannon, Suvranu De, Savador Dura-Bernal, Krishna Garikipati, George Karniadakis, William W. Lytton, Paris Perdikaris, Linda Petzold, and Ellen Kuhl. Multiscale Modeling Meets Machine Learning: What Can We Learn? *Archives of Computational Methods in Engineering*, 28(3):1017–1037, May 2021. ISSN 1134-3060. doi: 10.1007/s11831-020-09405-5.

Charles T. Perretti, George Sugihara, and Stephan B. Munch. Nonparametric forecasting outperforms parametric methods for a simulated multispecies system. *Ecology*, 94(4):794–800, April 2013. ISSN 0012-9658. doi: 10.1890/12-0904.1.

M. L. Pinsky, B. Worm, M. J. Fogarty, J. L. Sarmiento, and S. A. Levin. Marine Taxa Track Local Climate Velocities. *Science*, 341(6151):1239–1242, September 2013. ISSN 0036-8075. doi: 10.1126/science.1239352.

V. F. Pisarenko and D. Sornette. Statistical methods of parameter estimation for deterministically chaotic time series. *Physical Review E*, 69(3):036122, March 2004. ISSN 1539-3755. doi: 10.1103/PhysRevE.69.036122.

David M. Post, M. Elizabeth Conners, and Debra S. Goldberg. Prey preference by a top predator and the stability of linked food chains. *Ecology*, 81(1):8–14, 2000. ISSN 00129658. doi: 10.1890/0012-9658(2000)081[0008:PPBATP]2.0.CO;2.

Drew Purves, Jörn P. W. Scharlemann, Mike Harfoot, Tim Newbold, Derek P. Tittensor, Jon Hutton, and Stephen Emmott. Time to model all life on Earth. *Nature*, 493(7432):295–297, January 2013. ISSN 0028-0836. doi: 10.1038/493295a.

Christopher Rackauckas and Qing Nie. DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. *Journal of Open Research Software*, 5, 2017. ISSN 2049-9647. doi: 10.5334/jors.151.

Christopher Rackauckas, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, Ali Ramadhan, and Alan Edelman. Universal Differential Equations for Scientific Machine Learning. January 2020.

M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, February 2019. ISSN 00219991. doi: 10.1016/j.jcp.2018.10.045.

J. O. Ramsay, G. Hooker, D. Campbell, and J. Cao. Parameter estimation for differential equations: A generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5): 741–796, November 2007. ISSN 13697412. doi: 10.1111/j.1467-9868.2007.00610.x.

Stephan Rasp, Michael S. Pritchard, and Pierre Gentine. Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, 115(39):9684–9689, September 2018. doi: 10.1073/pnas.1810286115.

Andreas Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, August 2009. ISSN 1460-2059. doi: 10.1093/bioinformatics/btp358.

Maria Rodriguez-Fernandez, Pedro Mendes, and Julio R. Banga. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *Biosystems*, 83(2-3):248–265, February 2006. ISSN 03032647. doi: 10.1016/j.biosystems.2005.06.016.

Benjamin Rosenbaum and Emanuel A. Fronhofer. Confronting population models with experimental microcosm data: From trajectory matching to state-space models. *Ecosphere*, 14(4):e4503, April 2023. ISSN 2150-8925, 2150-8925. doi: 10.1002/ecs2.4503.

Benjamin Rosenbaum and Björn C. Rall. Fitting functional responses: Direct parameter estimation by simulating differential equations. *Methods in Ecology and Evolution*, 9(10):2076–2090, October 2018. ISSN 2041-210X. doi: 10.1111/2041-210X.13039.

Benjamin Rosenbaum, Michael Raatz, Guntram Weithoff, Gregor F. Fussmann, and Ursula Gaedke. Estimating Parameters From Multiple Time Series of Population Dynamics Using Bayesian Inference. *Frontiers in Ecology and Evolution*, 6(JAN), January 2019. ISSN 2296-701X. doi: 10.3389/fevo.2018.00234.

Sebastian Ruder. An overview of gradient descent optimization algorithms. pages 1–14, 2016.

Krista M. Ruppert, Richard J. Kline, and Md Saydur Rahman. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation*, 17:e00547, January 2019. ISSN 23519894. doi: 10.1016/j.gecco.2019.e00547.

Markus Schartau, Philip Wallhead, John Hemmings, Ulrike Löptien, Iris Kriest, Shubham Krishna, Ben A. Ward, Thomas Slawig, and Andreas Oschlies. Reviews and syntheses: Parameter identification in marine planktonic ecosystem modelling. *Biogeosciences*, 14(6):1647–1701, March 2017. ISSN 1726-4189. doi: 10.5194/bg-14-1647-2017.

Marten Scheffer, Steve Carpenter, Jonathan A Foley, Carl Folke, and Brian Walker. Catastrophic shifts in ecosystems. *Nature*, 413(6856):591–596, October 2001. ISSN 0028-0836. doi: 10.1038/35098000.

Simon Scheiter, Liam Langan, and Steven I. Higgins. Next-generation dynamic global vegetation models: Learning from community ecology. *New Phytologist*, 198(3):957–969, May 2013. ISSN 0028-646X. doi: 10.1111/nph.12210.

Tapio Schneider, Shiwei Lan, Andrew Stuart, and João Teixeira. Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations. *Geophysical Research Letters*, 44(24):12,396–12,417, December 2017. ISSN 0094-8276. doi: 10.1002/2017GL076101.

Y.H. Spitz, J.R. Moisan, M.R. Abbott, and J.G. Richman. Data assimilation and a pelagic ecosystem model: Parameterization using time series observations. *Journal of Marine Systems*, 16(1-2):51–68, September 1998. ISSN 09247963. doi: 10.1016/S0924-7963(97)00099-7.

M. C. Urban, G. Bocedi, A. P. Hendry, J.-B. Mihoub, G. Pe'er, A. Singer, J. R. Bridle, L. G. Crozier, L. De Meester, W. Godsoe, A. Gonzalez, J. J. Hellmann, R. D. Holt, A. Huth, K. Johst, C. B. Krug, P. W. Leadley, S. C. F. Palmer, J. H. Pantel, A. Schmitz, P. A. Zollner, and J. M. J. Travis. Improving the forecast for biodiversity under climate change. *Science*, 353(6304), September 2016. ISSN 0036-8075. doi: 10.1126/science.aad8466.

Ben A. Ward, Marjorie A.M. Friedrichs, Thomas R. Anderson, and Andreas Oschlies. Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models. *Journal of Marine Systems*, 81(1-2):34–43, April 2010. ISSN 09247963. doi: 10.1016/j.jmarsys.2009.12.005.

Claus O. Wilke, Jia Lan Wang, Charles Ofria, Richard E. Lenski, and Christoph Adami. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*, 412(6844):331–333, July 2001. ISSN 0028-0836. doi: 10.1038/35085569.

Jared Willard, Xiaowei Jia, Shaoming Xu, Michael Steinbach, and Vipin Kumar. Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems. 1(1):1–35, 2020. ISSN 2331-8422.

Yongjin Xiao and Marjorie A. M. Friedrichs. The assimilation of satellite-derived data into a one-dimensional lower trophic level marine ecosystem model. *Journal of Geophysical Research: Oceans*, 119(4):2691–2712, April 2014. ISSN 2169-9275. doi: 10.1002/2013JC009433.

Tao Xu, Luther White, Dafeng Hui, and Yiqi Luo. Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in parameter estimation and model prediction. *Global Biogeochemical Cycles*, 20(2): n/a–n/a, June 2006. ISSN 08866236. doi: 10.1029/2005GB002468.

Alireza Yazdani, Lu Lu, Maziar Raissi, and George Em Karniadakis. Systems biology informed deep learning for inferring parameters and hidden dynamics. *PLOS Computational Biology*, 16(11):e1007575, November 2020. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1007575.

Hao Ye and George Sugihara. Information leverage in interconnected ecosystems: Overcoming the curse of dimensionality. *Science*, 353(6302):922–925, August 2016. ISSN 0036-8075. doi: 10.1126/science.aag0863.

Hao Ye, Richard J. Beamish, Sarah M. Glaser, Sue C. H. Grant, Chih-hao Hsieh, Laura J. Richards, Jon T. Schnute, and George Sugihara. Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. *Proceedings of the National Academy of Sciences*, 112(13):E1569–E1576, March 2015. ISSN 0027-8424. doi: 10.1073/pnas.1417063112.

Qing Zhu and Qianlai Zhuang. Ecosystem biogeochemistry model parameterization: Do more flux data result in a better model in predicting carbon flux? *Ecosphere*, 6(12):art283, December 2015. ISSN 2150-8925. doi: 10.1890/ES15-00259.1.

S1 Parameter values of community models

S1.1 $\mathcal{M}_3$

We used the biologically realistic parameter values $r = \begin{pmatrix} 1 \\ -0.4 \\ -0.08 \end{pmatrix}$, $\alpha = \begin{pmatrix} 1.0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $W = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$, $H = \begin{pmatrix} 0 & 0 & 0 \\ 1.24 & 0 & 0 \\ 0 & 2.5 & 0 \end{pmatrix}$, $q = \begin{pmatrix} 0 & 0 & 0 \\ 4.98 & 0 & 0 \\ 0 & 0.8 & 0 \end{pmatrix}$, with initial conditions $x_0 = \begin{pmatrix} 0.77 \\ 0.06 \\ 0.945 \end{pmatrix}$. The dynamics of the system are chaotic for this set of parameter values [McCann and Yodzis, 1994a]. All non-zero coefficients were set as free parameters to be inferred, excluding the coefficients of the adjacency matrix W.

S1.2 $\mathcal{M}_5$

We used the biologically realistic parameter values $r = \begin{pmatrix} 1 \\ -0.15 \\ -0.08 \\ 1.0 \\ -0.15 \end{pmatrix}$, $\alpha = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$, $W = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & \omega & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-\omega & 0 \end{pmatrix}$ with $\omega = 0.2$, $H = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 2.89855 & 0 & 0 & 0 & 0 \\ 0 & 7.35294 & 0 & 0 & 7.35294 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.89855 & 0 \end{pmatrix}$, $q = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1.38 & 0 & 0 & 0 & 0 \\ 0 & 0.272 & 0 & 0 & 0.272 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.38 & 0 \end{pmatrix}$, with initial conditions $x_0 = \begin{pmatrix} 0.77 \\ 0.06 \\ 0.945 \\ 0.467 \\ 0.18 \end{pmatrix}$. The dynamics of the system are chaotic for this set of parameter values [Post et al., 2000]. All non-zero coefficients were set as free parameters, excluding the coefficients of the adjacency matrix W, where only ω was considered as a free parameter.

S1.3 $\mathcal{M}_7$

We used the biologically realistic parameter values $r = \begin{pmatrix} 1 \\ -0.15 \\ -0.08 \\ 1.0 \\ -0.15 \\ -0.01 \\ -0.005 \end{pmatrix}$, $\alpha = \begin{pmatrix} 1.0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$, $W = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \omega & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-\omega & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$ with $\omega = 0.2$, $H = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2.89855 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 7.35294 & 0 & 0 & 7.35294 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.89855 & 0 & 0 & 0 \\ 0 & 0 & 8.0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 12.0 & 0 \end{pmatrix}$, $q = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1.38 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.272 & 0 & 0 & 0.272 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.38 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.05 & 0 \end{pmatrix}$, with initial conditions $x_0 = \begin{pmatrix} 0.77 \\ 0.06 \\ 0.945 \\ 0.467 \\ 0.18 \\ 0.14 \\ 0.18 \end{pmatrix}$. The dynamics of the system are chaotic for this set of parameter values. All non-zero coefficients were set as free parameters, excluding the coefficients of the adjacency matrix W, where only ω was considered as a free parameter.

S1.4 $\mathcal{M}_3^\dagger$

We used the same parameter values as for $\mathcal{M}_3$, in addition to setting $\omega = 2\pi/3000$.

S2 Supplementary Figures

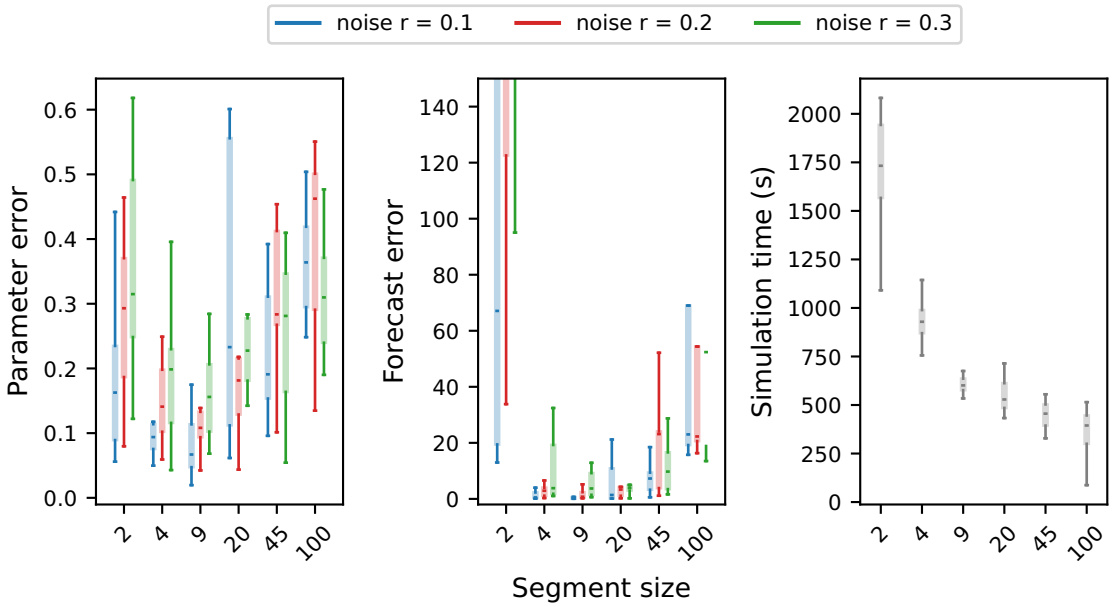

Figure S1: **Effect of segment length S on parameter error, forecast error and simulation time, for varying noise level, with model $\mathcal{M}_3$ under a partial observation setting.**

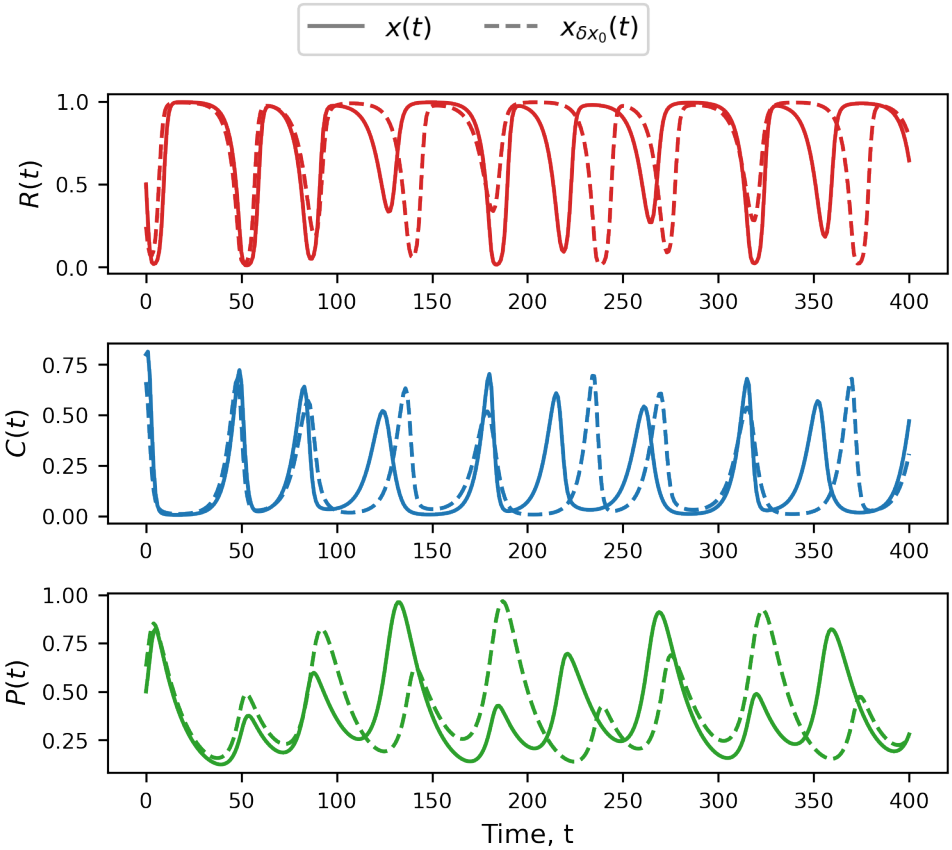

Figure S2: **Divergence between the trajectory $\mathcal{M}_3(t,(x_0,p))$ and a perturbed trajectory $\mathcal{M}_3(t,(x_0+\delta x_0,p))$.** For $t \lesssim 100$, $\mathcal{M}_3(t,(x_0,p))$ and $\mathcal{M}_3(t,(x_0+\delta x_0,p))$ are correlated and the divergence regime is informative, but for $t \gtrsim 100$ the trajectories become uncorrelated, illustrating the high sensibility of the model to ICs.

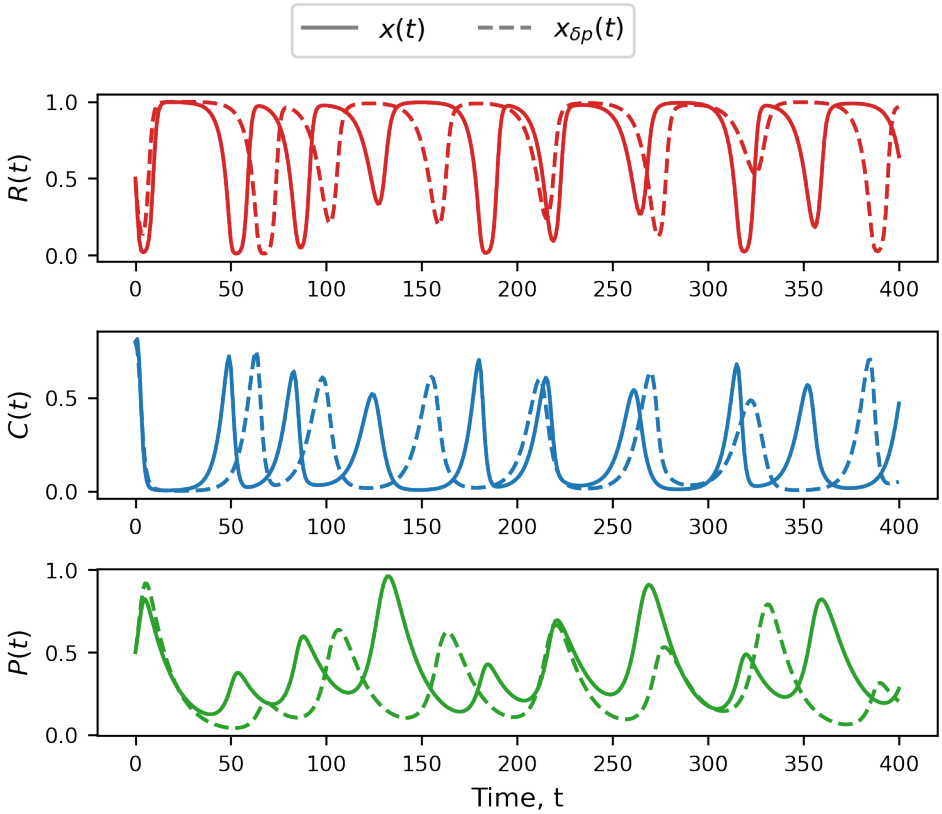

Figure S3: **Divergence between the trajectory $\mathcal{M}_3(t,(x_0,p))$ and a perturbed trajectory $\mathcal{M}_3(t,(x_0,p+\delta p))$.** For $t \lesssim 40$, $\mathcal{M}_3(t,(x_0,p))$ and $\mathcal{M}_3(t,(x_0,pp+\delta p))$ are correlated and the divergence regime is informative, but for $t \gtrsim 40$ the trajectories become uncorrelated, illustrating the high sensibility of the model to parameters.