ICES Journal of Marine Science

ICES Journal of Marine Science (2020), 77(4), 1463–1479. doi:10.1093/icesjms/fsz209

Contribution to the Themed Section: '*Science in support of a nonlinear non-equilibrium world*'

Food for Thought

Frequently asked questions about nonlinear dynamics and empirical dynamic modelling

Stephan B. Munch[1]*, Antoine Brias [1], George Sugihara[2], and Tanya L. Rogers[1]

[1]*Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Santa Cruz, CA 95060, USA*

[2]*Scripps Institution of Oceanography, University of California, La Jolla, CA 92037, USA*

**Corresponding author: tel: 831-420-3909; e-mail: steve.munch@noaa.gov.*

Munch, S. B., Brias, A., Sugihara, G., and Rogers, T. L. Frequently asked questions about nonlinear dynamics and empirical dynamic modelling. – ICES Journal of Marine Science, 77: 1463–1479.

Received 14 December 2018; revised 24 September 2019; accepted 8 October 2019; advance access publication 26 November 2019.

Complex nonlinear dynamics are ubiquitous in marine ecology. Empirical dynamic modelling can be used to infer ecosystem dynamics and species interactions while making minimal assumptions. Although there is growing enthusiasm for applying these methods, the background required to understand them is not typically part of contemporary marine ecology curricula, leading to numerous questions and potential misunderstanding. In this study, we provide a brief overview of empirical dynamic modelling, followed by answers to the ten most frequently asked questions about nonlinear dynamics and nonlinear forecasting.

Keywords: empirical dynamic modelling, nonlinear dynamics, nonlinear forecasting, time series analysis

Introduction

Ecosystems are complex systems consisting of many species interacting with one another and the environment. Ecology, as a science, seeks a quantitative understanding of these relationships through observations, experiments, and theory, but many serious challenges inhibit this pursuit. Most basically, many of the relevant species or variables in an ecosystem may go unobserved because they are unknown and/or difficult to measure. Uncertainty about the relevant variables may arise in part because in complex nonlinear systems like ecosystems, causes and effects can appear decoupled (i.e. lack of correlation between variables does not imply lack of a causal relationship; Sugihara *et al.*, 2012). Finally, because experiments are not always feasible at the relevant spatial and temporal scales, our understanding of ecosystems is often limited to that which can be built from field observations, forcing us to confront the uncertainties mentioned earlier.

We typically try to identify relevant variables and understand how they relate to each other by looking at correlations or regressions between candidate pairs of variables. Under this scheme, the relationships between variables are assumed to be constant and independent of each other, e.g. competition between two species will manifest as a constant negative correlation that is independent of changes in underlying resource availability. The ubiquitous deviations from idealized lines and curves (the scatter around correlations, linear and otherwise) are typically regarded as noise and an unavoidable part of reality. While extremely useful for simple physical systems and for controlled experiments, regarding natural ecosystems in this way may actually inhibit our understanding if ecological dynamics are nonlinear (i.e. pairwise associations are not independent of each other) and not constant (i.e. the system is not in static equilibrium).

The "dynamical systems" perspective offers an alternative and more holistic view of ecosystems. It does not assume constancy or separability of ecosystems into independent components with fixed relationships. It begins by thinking of each species, nutrient, environmental driver, etc., as a state variable or coordinate that

defines a so-called "state space," which frames the system. Thus, the number of coordinates or dimensionality of the state space reflects how many species or environmental drivers there are. A point in the state space corresponds to the current state of the system and the location of this point changes through time according to the rules governing the system dynamics. This traces out a trajectory: a rendering of the system dynamics where, depending on the location in state space, pairwise relationships among coordinate variables may change through time. The "attractor" is the set of values in the state space towards which this trajectory tends to converge. Under this dynamical systems viewpoint, deviations from simple curves may not actually be "noise" but may represent deterministic dynamics, where the apparent noisiness results from variables that were simply not taken into account.

Viewing ecosystems in this way allows us to leverage some powerful mathematical concepts from dynamical systems theory. These concepts are particularly useful in cases where we lack observations on all of the relevant variables, we do not know (but would like to know) which of the observed variables are relevant, and/or we do not know how the relevant variables interrelate. These are, of course, incredibly common problems in ecology.

Nonlinear time series methods, such as empirical dynamic modelling (EDM, also referred to as methods for state-space reconstruction, attractor reconstruction, time-delay embedding, and nonlinear forecasting), provide a path to understand dynamics that can be used to gain insight into how ecosystems work as well as to make accurate out-of-sample forecasts about future ecosystem states (e.g. Fogarty *et al.*, 2016). Nonlinear time series methods were used early on by Schaffer and Kot (1986) to construct classical unimodal maps, but the idea to reconstruct attractors to recover hidden variables and make forecasts was introduced to ecology by Sugihara and May (1990). This followed seminal studies by May (1976) and others establishing the potential for deterministic chaos to arise in nonlinear models of ecological dynamics, offering an explanation for the complex dynamics observed in nature, and heralding searches for chaos in empirical data (Hastings *et al.*, 1993; Ellner and Turchin, 1995). Early applications of nonlinear forecasting in ecology found these techniques to be especially useful for understanding ecological dynamics, provided time series of sufficient length (Grenfell *et al.*, 1994). While several seminal studies have established that chaos occurs in both laboratory and natural populations (Costantino *et al.*, 1997; Becks *et al.*, 2005; Graham *et al.*, 2007; Benincà *et al.*, 2008, 2015), other authors have concluded that chaos is rare in ecology (Berryman and Millstein, 1989; Upadhyay *et al.*, 1998; Sibly *et al.*, 2007) and that ecological time series were too short, noisy, and non stationary for nonlinear time series methods to be of much use (Hsieh *et al.*, 2008). Although model-based inference and prediction remain the norm (Dietze, 2017), their lack of performance, especially in fisheries contexts (Glaser *et al.*, 2014), has increased interest in methods such as EDM. And while EDM has yet to become mainstream, recent work has demonstrated its utility in a wide range of ecological applications (Deyle and Sugihara, 2011; Sugihara *et al.*, 2012; Benincà *et al.*, 2015; Tajima *et al.*, 2015; Ye *et al.*, 2015; Deyle *et al.*, 2016a; Munch *et al.*, 2018).

There are, of course, many other approaches to ecosystem forecasting (Dietze, 2017). For example, dynamic linear models (West and Harrison, 1997; Stow *et al.*, 1998), extended or unscented Kalman filters (Wan and van der Merwe, 2001; Lillacci and Khammash, 2010), and more general hidden Markov models (Morales *et al.*, 2004; Fukaya and Royle, 2013) and data assimilation methods (Luo *et al.*, 2011; Niu *et al.*, 2014; Massoud *et al.*, 2018; Dietze, 2017) are widely used to understand and predict ecological dynamics. These methods are particularly powerful when the parametric model structure is a good approximation of the real dynamics. However, a complete description of these methods is beyond the scope of this article.

Although nonlinear time series methods are being used with increased frequency, we suspect that their adoption in ecology is hampered by the fact that the mathematical foundation required to understand them, which is rooted in dynamical systems theory and topology, can be difficult to penetrate. However, the resulting murk generally follows a few common channels, epitomized by a suite of frequently asked questions. The purpose of this Food for Thought essay is to dispel some of the mystery surrounding EDM by providing answers to these questions.

In this study, we begin with a brief, but novel, description of the modelling approach and then address ten of the most commonly asked questions about nonlinear dynamics (Questions 1–3) and nonlinear forecasting using EDM (Questions 4–10). These are the questions that are almost always asked following our talks on EDM. Although we cite relevant literature throughout, we have focused on providing answers rather than a comprehensive review [but see Chang *et al.* (2017) for an excellent overview]. The answers to each question are more or less self-contained; rather than reading from beginning to end, the reader is encouraged to skim the questions and decide which seem the most relevant. This article is intended for quantitative ecologists who, like the authors, are not theoreticians with formal training in dynamical systems theory but are learning it as they go. To this end, a glossary of the mathematical jargon is provided in the supplement and verbal arguments and simulations are used to illustrate our answers rather than formal proofs. To keep the background a manageable length, we defer to the supplement extended descriptions of attractors, Lyapunov exponents, and the connection between discrete- and continuous-time approaches. For further information, Alligood *et al.* (1996) provide an excellent introduction nonlinear dynamics and chaos. For specific information on code implementing EDM, the interested reader is encouraged to consult the rEDM package and its documentation (Ye *et al.*, 2018).

Background

Let us say we have an ecosystem consisting of M different "state variables." The state variables could represent the population density of M different species, or the concentrations of M different nutrients. They could also represent the density of just one species in M different locations. More likely, the state variables represent some combination of population densities, nutrients, and abiotic factors in several different locations. If we use $x_{1,t}$, $x_{2,t}$, …, $x_{M,t}$ to represent the value of all state variables at time t, then the vector $\mathbf{x}_t = \{x_{1,t}, x_{2,t}, \ldots, x_{M,t}\}$ represents the state of the system. As the system changes through time, the result is a trajectory through this state space. Apart from transients and random perturbations, the trajectory for most systems of interest will converge to an attractor, e.g. a point, a closed loop, or a more complex shape. More complex shapes arise when the dynamics are "chaotic." In this case, trajectories that are initially close together tend to diverge at a rate governed by the dominant Lyapunov exponent, though they ultimately remain on the attractor.

We can describe the dynamics of the system in discrete time by a set of coupled equations. Since there are M state variables, there are M different "maps," i.e. discrete-time models, each of which is a function of the current system state. That is, $x_{1,t+1} = F_1(\mathbf{x}_t)$,

$x_{2,t+1} = F_2(\mathbf{x}_t)$, etc. The notation $\mathbf{x}_{t+1} = F(\mathbf{x}_t)$ is shorthand for this collection of equations. In the supplement, we describe the connection between this discrete-time model and its analogue in continuous time.

If we have data on all of the state variables over a wide enough range of values, we can empirically estimate the functions $F_i(\mathbf{x}_t)$ from the data. As an example of this idea, Figure 1 shows abundance time series for a single species in a two-site metapopulation system. Plotting the future population size in each location as a function of the current population size at both locations reveals that there is smooth function governing the dynamics. This function can be estimated using any number of flexible, non-parametric regression approaches (e.g. splines, neural networks, or Gaussian processes). Use of this empirical, non-parametrically estimated map to make predictions and inferences about the dynamics is one of the core ideas of EDM (Sugihara and May, 1990; Sugihara *et al.*, 2012; Chang *et al.*, 2017).

Unfortunately, it is almost always the case that we only have measurements on a subset of the relevant state variables and that the true dimensionality of the system, M, is effectively unknown. To make this explicit, we split the state variables $\mathbf{x}_t$ into the following two subsets: $\mathbf{y}_t = \{x_{1,t}, x_{2,t}, \ldots, x_{O,t}\}$ representing the observed state variables and $\mathbf{z}_t = \{x_{O+1,t}, \ldots, x_{M,t}\}$ containing the remaining unobserved state variables. We rewrite the dynamics as follows:

$$\mathbf{y}_{t+1} = \mathrm{F}(\mathbf{y}_t, \mathbf{z}_t), \tag{1}$$

$$\mathbf{z}_{t+1} = \mathrm{G}(\mathbf{y}_t, \mathbf{z}_t), \tag{2}$$

where F represents the maps for the observed states (1 through O) and G represents the maps for the unobserved states ($O+1$ through M). There is more than one way to proceed in this situation, including (i) modelling only the observed states and treating the unobserved states as process noise, (ii) implicitly accounting for the unobserved states using time lags (Deyle *et al.*, 2013; Munch *et al.*, 2018), or (iii) modelling the complete dynamics and imputing the unobserved states using a hidden Markov approach e.g. Morales *et al.* (2004). As there are several good books on hidden Markov models for ecologists (Ruth King *et al.*, 2010; Dymarski, 2011; Newman *et al.*, 2014), we focus here on the first two approaches.

Process noise

The most commonly adopted approach is to focus solely on modelling the observed states, implicitly treating the unobserved state variables as noise. To do so, we replace our deterministic model with an approximation incorporating process uncertainty. The standard approach would be to use something like $y_{i,t+1} = \hat{F}_i(\mathbf{y}_t) + \varepsilon_{i,t}$ where $\varepsilon_{i,t}$ represents the process noise affecting the ith observed state variable.

We can explicitly connect this model with process noise to (1) via a first-order approximation around the mean of the unobserved states:

$$y_{i,t+1} \approx F_i(\mathbf{y}_t, \bar{\mathbf{z}}) + \sum_{j=O+1}^{M} \frac{\partial F_i}{\partial z_j}(z_{j,t} - \bar{z}_j) \ , \tag{3}$$

from which we can see that the variance in the process noise is approximately,

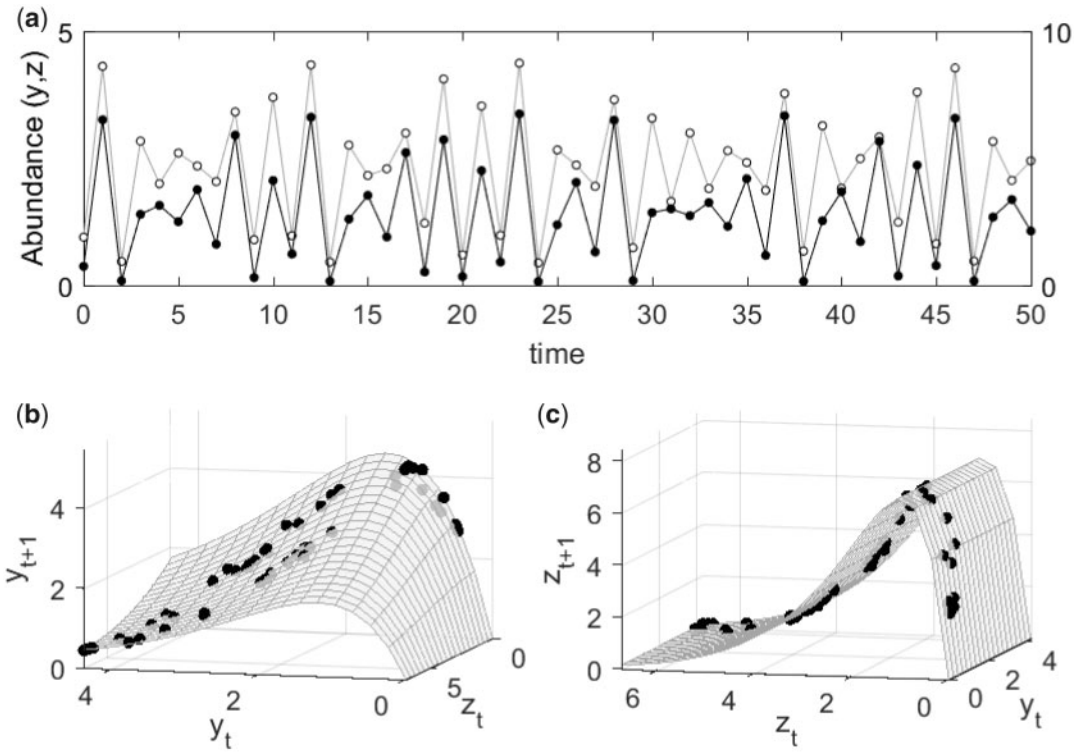

Figure 1. Empirical dynamics of a single-species, two-location population model when all state variables are observed. (a) Time series of simulated abundance for sites 1 (y) and 2 (z). (b) The next population size at site 1 as a function of current population sizes at sites 1 and 2. The points represent the simulated data as plotted in panel (a). The surface is estimated using a Gaussian process regression (Munch *et al.*, 2017). (c) Same as in (b) but for the next population size at site 2.

$$\mathrm{Var}(\varepsilon_{i,t}) \approx \sum_{j=O+1}^{M} \left(\frac{\partial F_i}{\partial z_j}\right)^2 \mathrm{Var}(z_j).$$

In some sense, this approximation is almost unavoidable in ecology, but it is something we rarely make explicit. Importantly, doing so reveals that the quality of the approximation depends on both the variance in the unobserved variables through time ($\mathrm{Var}(z_j)$) and the sensitivity of the observed state variables to changes in the unobserved ones $\left(\frac{\partial F_i}{\partial z_j}\right)$.

As a simple example of this approach, we return to the two-location system from Figure 1 where only the abundance in site 1, y, is measured. The complete dynamics are as follows:

$$y_{t+1} = (1-m_1)y_t e^{r_1 - y_t} + m_2 z_t e^{r_2 - z_t}$$
$$z_{t+1} = (1-m_2)z_t e^{r_2 - z_t} + m_1 y_t e^{r_1 - y_t}.$$

Since we only have data for site 1, we might fit a model of the form $y_{t+1} = y_t e^{\hat{r} - y_t + \varepsilon_t}$ where the apparent growth rate implicitly includes migration, i.e. $\hat{r} = r_1 + \ln(1-m_1)$ and the noise term is driven by immigration from site 2. Fitting this model to some data generated with (4) looks pretty good by ecological standards (Figure 2a). However, as we will see, it is possible to do much better in this case.

Delay embedding

Rather than treating the unobserved variables as noise, we could make use of Takens' theorem of time-delay embedding (Takens, 1981). Briefly, Takens' theorem says that if the trajectory of an autonomous deterministic system $\{y_1, y_2, \ldots, y_M\}$ converges to an attractor, and the dimensionality of the attractor (d) is less than the dimensionality of the system (M), then we only need data on the observable variables $\{y_1, y_2, \ldots, y_O\}$ to fully reconstruct the system dynamics. Specifically, we can faithfully reconstruct the attractor using time lags of the observed variables as a synthetic coordinate system, provided that the number of time lags (delay coordinates) used is $> 2d$. In its original form, Takens' theorem uses lags of only a single observable variable $\{y_{i,t}, y_{i,t-\tau}, y_{i,t-2\tau}, \ldots, y_{i,t-E\tau}\}$ to reconstruct the attractor, where τ is the "time delay" and E ($\geq 2d$) is the "embedding dimension." This result was later generalized to include externally driven (Stark, 1999) and stochastic systems (Stark *et al.*, 1997; Kantz and Schreiber, 2003). Mixtures of time delays can be used to account for multiple time scales (Judd and Mees, 1998) and (Munch *et al.*, 2017) used automatic relevance detection (Neal, 1997) to select relevant lags. Reconstructions from multiple variables (Deyle and Sugihara, 2011; Ye and Sugihara, 2016) can improve prediction and facilitate the inference of different mechanisms (Deyle *et al.*, 2016b). Takens' theorem is a deep mathematical result with far-reaching implications. Unfortunately, to really understand it, it requires a background in topology.

We can, however, build intuition for how lags implicitly account for unobserved state variables. To do so, shift the map for the unobserved states back by one time step and substitute this into the dynamics for the observed states. That is, plug $z_t = G(y_{t-1}, z_{t-1})$ into (1) to get $y_{t+1} = F[y_t, G(y_{t-1}, z_{t-1})]$. Now, if we are very lucky, we can solve (1) for the unobserved states and obtain an equation of the form $z_{t-1} = \Phi(y_t, y_{t-1})$. Substituting this in for z_{t-1}, we find

$$\begin{aligned} y_{t+1} &= F(y_t, z_t) = F(y_t, G[y_{t-1}, z_{t-1}]) \\ &= F(y_t, G[y_{t-1}, \Phi\{y_t, y_{t-1}\}]). \end{aligned} \quad (5)$$

We now have a new map for y_{t+1} that is a complete description of the dynamics that depends only on $\{y_t, y_{t-1}\}$; no information about z is required.

If it is not possible to solve for z_{t-1} using one lag, we can push z back in time another step, i.e. $z_{t-1} = G(y_{t-2}, z_{t-2})$ to get $y_{t+1} = F(y_t, G[y_{t-1}, G\{y_{t-2}, z_{t-2}\}])$. Obviously, we can continue in this way indefinitely and keep going until we have enough information to write $z_{t-E} \approx \Phi(y_t, y_{t-1}, \ldots, y_{t-E})$. We may not be able to solve for z_{t-E} exactly, but we can think about this as an approximation analogous to (3). In this case, the quality of the approximation depends on how sensitive y_{t+1} is to z_{t-E} and on the variance in z_{t-E} *conditional on* $\{y_t, \ldots, y_{t-E}\}$. If having more information about past values of y reduces the variance in z_{t-E}, we can expect including lags in the model to reduce the

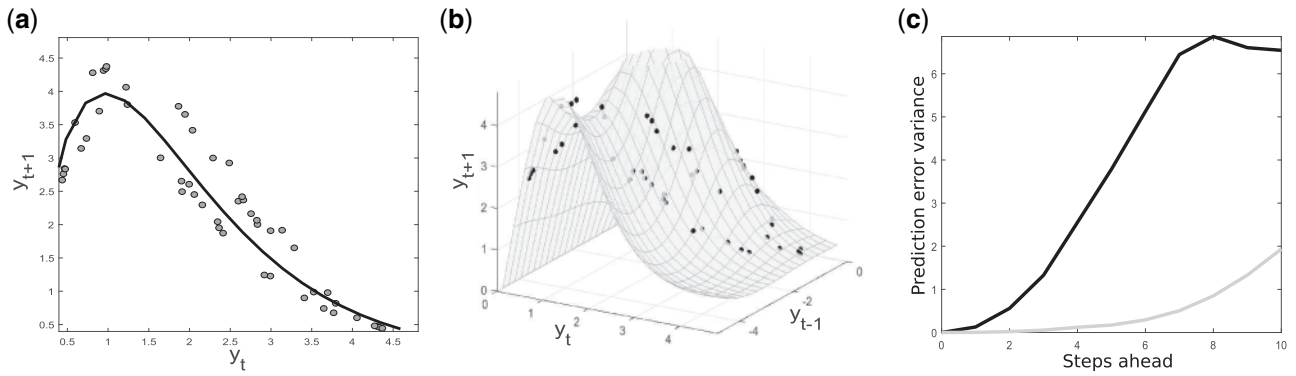

Figure 2. Empirical dynamics of a single-species, two-location population model when only one site is observed. (a) Approximate dynamics using a 1-d model (black line). The next population size at site 1 is plotted as a function of the current population size at site 1. The points are the simulated data. (b) Dynamics using a 2-d delay-coordinate model. The next population size at site 1 is plotted as a function of the current and previous population sizes at site 1. (c) Multi-step prediction using the 1-d approximation (black line) and the delay-coordinate model (grey line). The horizontal axis is the number of steps into the future we are trying to predict. The vertical axis is the mean squared prediction error, i.e. $\sum_t (\hat{y}_{t+\tau} - y_{t+\tau})^2/N$. The sum is obtained by making predictions τ steps ahead starting from each point in the time series shown in Figure 1.

prediction error for the observed variables. If the true system is deterministic and we have unlimited quantities of noise-free data, this conditional variance, and hence the prediction error, will go to zero. In real life, we can expect including lags to reduce the prediction error down to some finite limit.

We do not intend this as a proof of Takens' theorem or a complete statement about the implications of time-delay embedding. Nevertheless, this line of thinking should provide some intuition for why it is possible to reconstruct the dynamics of a system using delay coordinates. The practical upshot is that the same function approximation tools that we would use to estimate F in the case where we had a complete state vector can now be used to estimate the map in delay coordinates. Making use of time lags to implicitly account for unobserved state variables is the second core idea of EDM.

Returning to the two-site illustration, we can use the shift-and-substitute recipe to find the unobserved variable (z) as a function of the observed variable (y), i.e. $Z(y_t, y_{t-1}) = (1 - 1/m_2)y_t + (1 - (1 - m_1)/m_2)y_{t-1}e^{r_1 - y_{t-1}}$, and use this to rewrite (4) as $y_{t+1} = (1 - m_1)y_t e^{r_1 - y_t} + m_2 Z(y_t, y_{t-1})e^{r_2 - Z(y_t, y_{t-1})}$. Thus, we have an exact description of the two-site system expressed solely in terms of the abundance in site 1. If we knew these dynamics explicitly, fitting this model to the data from site 1 would require just three more parameters than fitting the one-dimensional (1-d) approximation. If we did not know enough about the system to write this parametric expression for the dynamics, we would use a non-parametric regression with y_t and y_{t-1} as inputs (Figure 2b).

Iterating the models shown in Figure 2a and b several steps into the future highlights the most salient difference between the time-delay embedding approach and treating the unobserved states as noise: predictions using delay embedding are substantially better than the 1-d noisy model up to eight steps into the future (Figure 2c). Of course, these results are specific to this example; the difference in forecast accuracy generally depends on both the sensitivity of the dynamics to the unobserved states and our ability to reconstruct the dynamics from the available data.

Because the shape of the model must be inferred from the data, time-delay embedding is most useful when the time series cover a broad range of states. This is more likely to happen in a nonlinear or chaotic system than it is for stable dynamics perturbed by noise. However, even when the data are not sufficient to completely recover the deterministic dynamics, including lags may still improve predictions.

We note that in actual application, we typically need to identify the relevant time delay, τ, in addition to the embedding dimension, E. The selection of both E and τ for a given data set is usually based on minimizing prediction error, cross-validation, or mutual information (see e.g. Chang *et al.*, 2017). The optimal values for E and τ for a given time series are not always obvious, though several methods have been developed to automate their identification (e.g. Garland *et al.*, 2016; Munch *et al.* 2017).

To summarize, ecosystems involve high-dimensional state spaces with complex dynamics. We rarely have data on all of the relevant state variables and our mechanistic understanding of the dynamics hardly ever complete. EDM uses time delays to account for unobserved variables and non-parametric modelling to flexibly infer the dynamics. These methods are being applied successfully in marine ecosystems to understand their dynamics (Deyle *et al.*, 2013, 2016b) and make better predictions (Ye *et al.*, 2015; Munch *et al.*, 2018; Pierre *et al.*, 2018). Nevertheless, there are often questions about the relevance of nonlinear or chaotic dynamics in ecology, the conditions under which EDM is expected to produce useful results, and what can be learned from EDM beyond making forecasts. The ten questions we have encountered most often are addressed below.

Question 1. Do nonlinear models always generate nonlinear dynamics?

Although statistical ecologists frequently fit nonlinear dynamical models to data, we need to distinguish between nonlinearity in the equations and nonlinearity in the dynamics that the fitted models generate.

If a dynamical model is linear, the system can be written as follows:

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t, \tag{6}$$

where $\mathbf{A}$ is a matrix of coefficients which does not depend on $\mathbf{x}$. Note that the trajectories $x_i(t)$ will not typically be linear with respect to time. A 1-d linear model leads to exponential growth/decay, but when the dimension of $\mathbf{x}$ is large enough, linear models can display other behaviours, including rather complicated-looking periodic cycles. Of course, almost no ecological models start out linear because this is inconsistent with ecological reality. More often, we say that a model is given by $\mathbf{x}_{t+1} = \mathbf{F}(\mathbf{x}_t)$ where $\mathbf{F}$ is a collection of nonlinear functions. However, for many nonlinear models, the dynamics near a stable equilibrium can be well approximated by a linear model. This occurs whenever there is an equilibrium point, x^*, and the Jacobian matrix, $J_{ij} = \frac{\partial F_i}{\partial x_j}$, evaluated at x^* has eigenvalues with modulus <1. Then, the dynamics following any small perturbation from x^* will be well approximated by (6) with $\mathbf{A} = \mathbf{J}$.

Importantly, nonlinear and chaotic dynamics can only be generated by nonlinear models; thus, chaotic dynamics and nonlinear dynamics are closely associated and the terms are sometimes used interchangeably in the literature. An unstable equilibrium is also a necessary, but not sufficient, condition for nonlinear (chaotic) dynamics in a deterministic setting. Thus, studies of nonlinear dynamics often include discussions of stability, as measured by eigenvalues or Lyapunov exponents.

As an example of a nonlinear model generating linear dynamics, take the classic Ricker model (Ricker, 1975) with noise, $x_{t+1} = x_t e^{r - x_t + \varepsilon_t}$, where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ and r controls the population growth rate. This *model* is clearly nonlinear. However, when $r < 2$, the *dynamics* are very nearly linear, which is shown in Figure 3a. That is, a straight line fit of x_{t+1} to x_t is pretty good—and about the best we can do with the information in the time series. Increasing the value of r destabilizes the equilibrium, leading to limit cycles or chaos (Figure 3b and c). In this case, a straight line fit of x_{t+1} to x_t does not adequately capture the dynamics.

In light of this, it can be useful to measure whether the dynamics for a given system are linear or nonlinear. If we have a sufficiently long time series, we can empirically evaluate whether the dynamics are well approximated by (6) by comparing the fit of (6) with the fit of a nonlinear alternative. Unfortunately, it is not always obvious what the nonlinear alternative model should be. Another approach that has been applied quite widely is to fit a model in which $\mathbf{A}$ is allowed to vary with $\mathbf{x}$ using a *local* linear regression and thereby test whether the dynamics depend on $\mathbf{x}$. This approach is often referred to as "s-map" (Sugihara, 1994).

Question 2. Do vital rates have to be physiologically or ecologically unreasonable to generate nonlinear dynamics?

There appears to be widespread belief that the growth rate (r) of a population at low population sizes needs to be unreasonably large

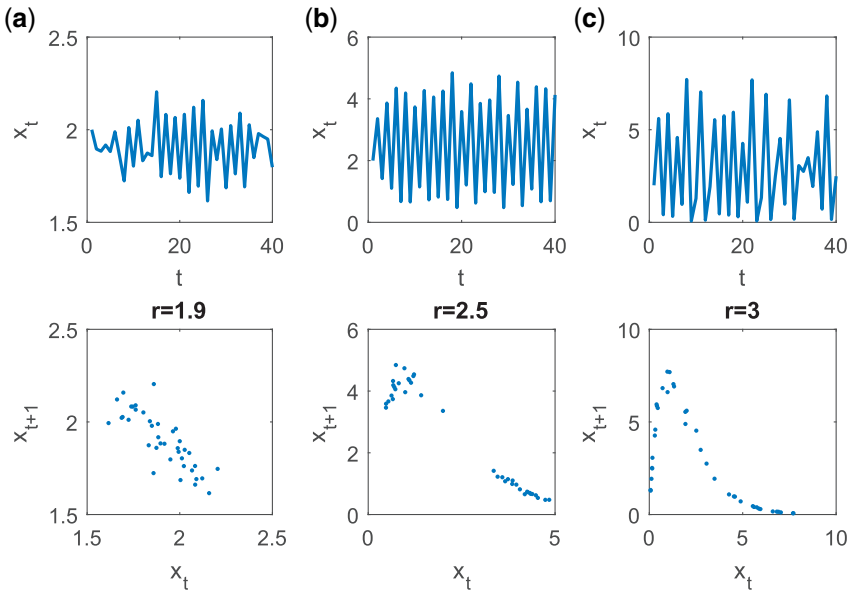

Figure 3. The Ricker model with noise ($\sigma = 0.05$) exhibits several different behaviours depending on the population growth rate: (a) nearly linear when $r = 1.9$, (b) a noisy two-cycle when $r = 2.5$, and (c) chaos when $r = 3$.

for nonlinear dynamics or chaos to occur (May, 1976). Although this is true for the simple, 1-d difference equation models that were used in early explorations of chaos, such as the logistic and Ricker maps, it is not true in general (see e.g. Gross *et al.*, 2005). The connection between population growth at low densities and chaos is highly model specific and inferences from low-dimensional models tell us little about real, high-dimensional, dynamics.

A necessary, but not sufficient, condition for nonlinear dynamics in a difference equation setting is that the model has an unstable equilibrium. To have an unstable equilibrium in a 1-d map, the slope of the return map at equilibrium must be >1 in absolute value. The apparent relationship between the onset of chaos and unreasonably high growth rates at low population densities is solely because—in early theoretical studies—there was a single parameter governing both the slope at the origin and the slope at the equilibrium.

It is straightforward to obtain chaotic dynamics with small population growth rates at low population sizes by decoupling the slope at the origin and the slope at the equilibrium. As an example, we can generalize the logistic map with one additional "shape" parameter to obtain the discrete theta-logistic model: $x_{t+1} = rx_t(1 - x_t^{\theta})$. As in the discrete logistic, the slope at the origin is given by r, but, the slope at equilibrium is $1 - \theta(r - 1)$, which can be a large negative number even when r is close to 1. Moreover, the value of r needed to generate chaotic dynamics (as indexed by a positive Lyapunov exponent) decreases with increasing θ (Figure 4).

Despite the allure of analytical tractability, 1-d biological systems exist only in chemostats and theory. The conditions for chaos to occur tend to be less stringent in larger systems. Even in classical Lotka–Volterra models, complex dynamics arise with the introduction of additional species (Vano *et al.*, 2006), reproductive delays (Zhao *et al.*, 2014), space (Wildenberg *et al.*, 2006), or contemporary evolution (Yu and Liu, 2016). As the number of species involved becomes large, random matrix theory can be used to show that instability is more likely (Stone, 2018), suggesting that nonlinear behaviour is easier to obtain as well. This is consistent with theory indicating that long food chains have general properties that make chaos likely (Gross *et al.*, 2005).

As an example, consider the multi-species Ricker model (Ackleh and Salceanu, 2015; Hartmann *et al.*, 2017). In this model, there are n species and the dynamics for the ith species are given by the following equation:

$$x_{i,t+1} = x_{i,t} e^{r_i - \sum_j A_{i,j} x_{j,t}} \tag{7}$$

The parameter r_i represents the growth of species i in the absence of interactions, and the $A_{i,j}$ terms represent the effect of species j on the growth of species i.

To see the effect of system size on the likelihood of chaos, we fixed r at 0.1 for all species, which is well below the value needed to generate chaos in the single-species Ricker model. We set the intraspecific interactions, $A_{i,i}$, to 0.1 as well so that in the absence of interspecific interactions, all species would converge to a stable equilibrium population size of 1. We then randomly assigned interspecific interaction strengths (i.e. the remaining $A_{i,j}$ terms). To randomly construct a predator–prey network, we drew interaction strengths, z, from a $N(0, \sigma^2/n)$ distribution. Since, by convention, we are subtracting A, if $z<0$, then species i preys on species j, and we set $A_{i,j} = fz$ and $A_{j,i} = -z$ where $f=0.1$ is the conversion efficiency. If $z>0$, then species j preys on species i, and we set $A_{i,j} = z$ and $A_{j,i} = -fz$. This algorithm constructs densely connected networks with many weak interactions, which is at least qualitatively consistent with diet studies in marine fishes (Link, 2002).

For each randomly constructed community of n species, we iterated the model 1000 steps to eliminate transients and then 1000 more steps to evaluate the dominant Lyapunov exponent. Since the Jacobian for the system is known at each time step, i.e. $J_{i,j} = -x_{i,t+1}A_{i,j}$ and $J_{i,i} = x_{i,t+1}/x_{i,t} - x_{i,t+1}A_{i,i}$, we can compute the complete set of Lyapunov exponents using the QR algorithm of

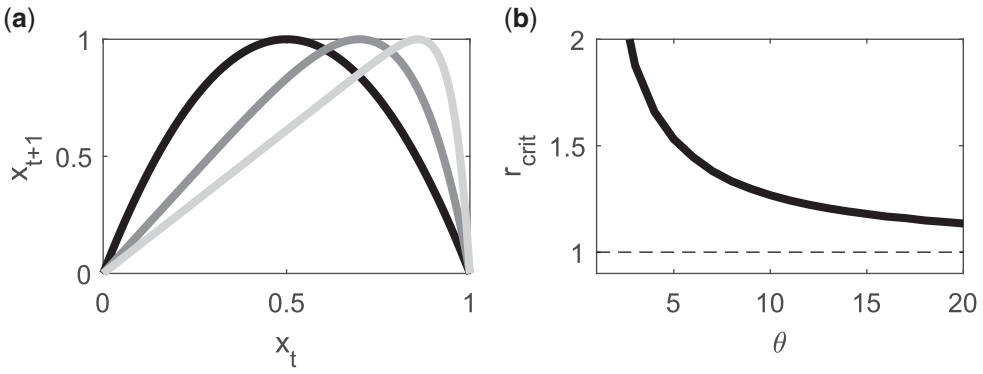

Figure 4. Chaos in the discrete theta-logistic model. (a) The theta-logistic for $\theta = 1$ (black), 5 (dark grey), and 20 (light grey). The horizontal and vertical axes are the population sizes at t and $t+1$ respectively. For each value of θ, r is chosen such that the maximum value of $x_{t+1} = 1$. Each of these examples exhibits chaotic dynamics. (b) The minimal population growth rate required for chaos. The vertical axis is r_{crit}, i.e. the smallest value of r for a given value of θ such that the dynamics are chaotic. For sufficiently large θ, chaos occurs with $r \approx 1$.

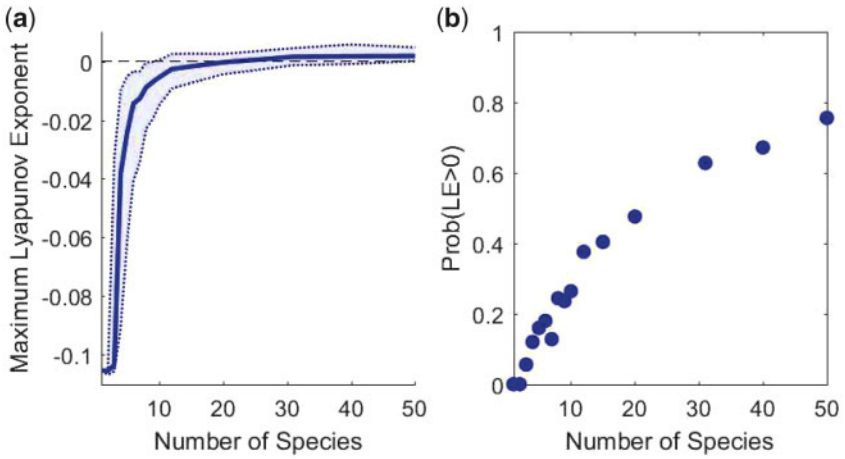

Figure 5. Dynamics of randomly constructed predator-prey networks. (a) The dominant Lyapunov exponent (median and interquartile range from 250 simulations) as a function of the number of species (n). (b) The frequency of dominant Lyapunov exponents that are positive, indicating chaotic dynamics. The distribution of Lyapunov exponents narrows with increasing n, and the dynamics are chaotic more than half the time for communities larger than about 20 species.

Benettin *et al.* (1980). We repeated this procedure 250 times for each n over a range of ns from 1 to 50.

For predator–prey communities constructed in this way, the plausibility of chaotic dynamics increased sharply as the number of species increased (Figure 5). Under this setup, where all of the r_is are small, at least three species were required for chaos to occur. More than half of the randomly constructed networks exhibited chaos when the number of species exceeded 20.

This simple model illustrates that intuition from single-species models can be quite misleading; it is clearly possible to generate chaos with small population growth rates (here $r = 0.1$) and very modest interaction strengths; in this model, they are typically much <1, particularly when n is large.

Question 3. If a fitted model is stable, does this mean the system dynamics are stable?

To paraphrase E.T. Jaynes, this reflects a "model projection fallacy" in which the way a model describes the world is assumed to reflect the way the world really is. In fact, the stability of a fitted model can produce qualitatively incorrect inference about the stability of the system that generated the time series.

Stability is essentially a statement about the long-run behaviour of a system. If we have a model that accurately predicts out-of-sample data over an extended period of time into the future, then the stability of the model is likely to reflect reality. However, even seemingly trivial degrees of model mis-specification can lead to incorrect conclusions about the stability of a dynamical system. This is particularly problematic when discrepancies between model and data are attributed to "noise."

For example, consider the data in Figure 6. The fit of the Beverton–Holt model (i.e. $x_{t+1} = rx_t(a + x_t)^{-1}$) appears quite good, and we would be well-justified in selecting this model. However, the data were actually generated with a two-dimensional model in which the focal species is consumed by a generalist predator. So what looks like stable dynamics with noise when viewed in 1-d is really deterministic chaos in two-dimensions. However, we will never find any evidence of chaos by fitting a Beverton–Holt model: the dynamics under Beverton–Holt are always stable.

As a second example, let us say we generated data from the multi-species system from Question 2 (7) but then fitted it using a single-species Ricker model. To do this in the simplest way possible, we fitted $y = \ln(x_{t+1}/x_t) = \hat{r} + \hat{g}x_t + \epsilon$ via least squares for

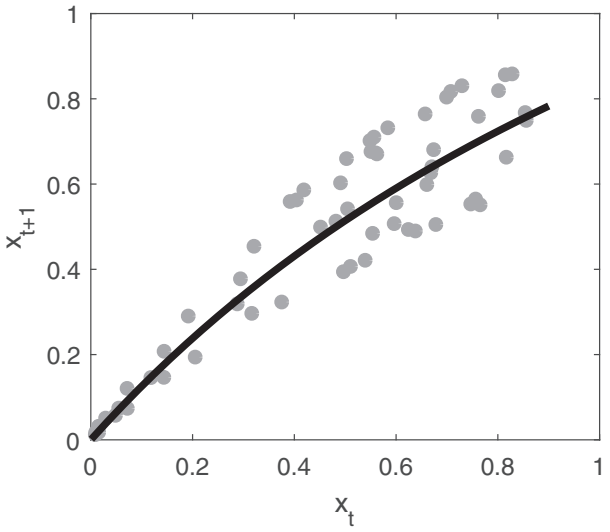

Figure 6. Seemingly stable population dynamics. The points are simulated data and the black line is fit of a Beverton–Holt model to them. The data, x_t were generated by a chaotic predator-prey system: $x_{t+1} = r_x x_t(1 - cy_t)(a + x_t(1 - cy_t))^{-1}$, $y_{t+1} = r_y * y_t(1 - y_t) + fx_t y_t$ with $r_x = 2$, $c = 1$, $r_y = 3.9$, and $f = 0.005$.

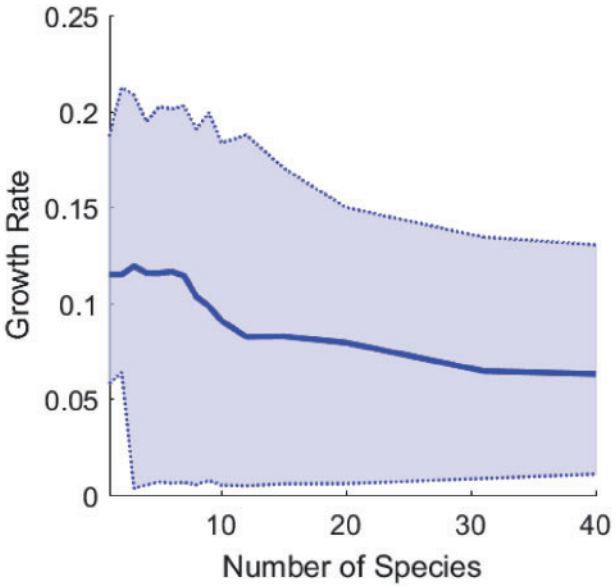

Figure 7. Single-species growth rates, $\hat{r}$, estimated from data generated by the randomly constructed predator-prey networks in Figure 5. For each randomly constructed community, we fit the Ricker model to the time series for each species separately and recorded the maximum estimate of $\hat{r}$ over all species, r_{max}. We did this 250 times for each community size, n, ranging from 1 to 50. The line indicates the median r_{max} and the bands are the interquartile range.

each simulated species and recorded the maximum estimate of $\hat{r}$ from each simulated community. Since the Ricker model is exactly the right shape for an approximation of the form (3), and chaos in the Ricker model emerges at high growth rates, we might expect the maximum growth rate to be a reasonably good indicator of whether or not the dynamics are chaotic. This is precisely the sort of thing that ecologists have done for decades (see Hassell and Comins, 1976; Sibly *et al.*, 2007; Shelton and Mangel, 2011).

How often would we conclude that the dynamics were chaotic? The answer, perhaps surprisingly, is never (Figure 7). Of the 4000 simulations with n ranging from 1 to 50, in which 1168 had positive Lyapunov exponents, none of the resulting regressions had a maximum $\hat{r}$ of >1. Since chaos in the Ricker model does not emerge until $r > 2.65$, we would be forced to conclude that chaos is rare in ecology.

As these examples illustrate, fitting a model to data does not necessarily tell us what we want to know about system stability. It might seem obvious that if you fit a wrong model you get a wrong answer, but many times, we focus on selecting models from a short list of *a priori* candidates by asking which one fits the data best. Although there will always be a best fitting model, this is not a guarantee that the selected model faithfully re-creates the dynamical properties of the system (see e.g. Boettiger *et al.*, 2015).

For complex systems, this issue is even more subtle—we could fit a correctly specified model to data and *still* get the wrong answer for stability. For a suite of ecological models in the chaotic domain, parameters estimated using traditional likelihood-based methods are biased towards stability (Perretti *et al.*, 2013). This problem is well known outside of ecology (see Abarbanel *et al.*, 1996; Judd, 2008) and arises because of extreme sensitivity to parameters and initial conditions. Synthetic likelihoods (Wood, 2010), shadowing (Judd, 2008), and synchronization (Abarbanel *et al.*, 1996) allow us to fit chaotic models to data without bias, but these methods are not routinely applied by quantitative ecologists.

So what should we do if we want a robust measure of the stability of a system? Farmer and Sidorowichl (1989) pioneered the idea of estimating Lyapunov exponents, and hence system stability, directly from data by measuring the divergence rate of trajectories that are initially close in state. There is now a considerable literature generalizing and applying this idea (see e.g. McCaffrey *et al.*, 1992; Benincà *et al.*, 2015; Ushio *et al.*, 2018). These methods allow us to characterize the stability of given system from data, rather than the stability of a model analogue.

Question 4. How long does the time series need to be to use EDM?

A generic but vague answer is that to empirically quantify system dynamics, the length of the time series needs to be several multiples of the characteristic return time for the system, i.e. the time it takes for the system state to return to a small neighbourhood of a given starting state. In addition, the maximum embedding dimension that we can recover with a given data set scales roughly as the square root of the time series length (Cheng and Tong, 1992). Thus, we can expect a fair amount of unexplained variation to remain when the time series is short and the attractor dimension is large.

In our experience, it is often the case that we see a significant reduction in prediction error when we have >30 years of data, particularly for short-lived species (Sugihara *et al.*, 2012; Ye *et al.*, 2015; Munch *et al.*, 2018). For example, Munch *et al.* (2018) found that when the length of the time series was ten times the mean age at maturation, that EDM could explain >50% of the variation in recruitment in several species.

However, time series length alone does not guarantee our ability to reconstruct dynamics. Dynamics that occur on time scales much shorter than the sampling interval will be difficult to reconstruct and lead to apparent indeterminism. To see this, imagine that the dynamics are governed by a logistic map: $x_{t+1} = rx_t(1 - x_t)$ with $r = 4$. If we have data at every time step, we can do a good job of reconstructing the dynamics with EDM. If we

have data every other time step, the map is now a fourth-order polynomial, i.e. $x_{t+2} = r^2 x_t(1 - x_t)(1 - rx_t(1 - x_t))$, and will require twice the data to resolve. If we only have data every five time steps, the map is now a 32nd-order polynomial that we have little hope of resolving. Unless we are able to sample more frequently, we are likely to conclude that the dynamics are stochastic.

Question 5. Can we use the embedding dimension to estimate the number of relevant species in the system?

The short answer is no. In Takens' theorem, the number of lags used must be greater than twice the dimension of the attractor, not the dimension of the system. This threshold guarantees that the delay-coordinate representation shares a one-to-one correspondence with the original attractor. However, the attractor dimension in a deterministic system is nearly always less than the dimension of the state space, sometimes very much so. For example, in any system that converges to a stable fixed point, the attractor dimension is 0. An arbitrarily complex system that exhibits a limit cycle may have an attractor dimension as low as 1. Attractor dimension and system dimension are definitely not the same.

Second, Takens' requirement that the embedding dimension be at least twice the attractor dimension is a generic condition intended for an arbitrary dynamical system. In practice, many systems can be reconstructed with fewer lags. For instance, the attractor for the Ricker map in the chaotic domain is 1-d, so a literal use of Takens' theorem would indicate that we needed $E = 2 \times 1 + 1 = 3$ lags, but really, we can do just fine in 2, i.e. $x_{t+1} = f(x_t)$. So if we are using some measure of goodness of fit to evaluate the embedding dimension, we often end up with an estimate that is less than $2d + 1$. Third, the maximum embedding dimension scales as $\sqrt{T}$ (Cheng and Tong, 1992). For an ecological time series of 50 years, we should expect the embedding dimension estimate to be <7, which seems consistent with the results that Glaser *et al.* (2014) found for abundance and landings time series. In >90% of the 135 time series they analysed, the estimated embedding dimension was <7.

Question 6. Does EDM work with stochastic dynamics or observation error?

Originally, Takens' theorem was restricted to deterministic systems but subsequent work extended these results to systems with noisy dynamics (Stark *et al.*, 1997; Kantz and Schreiber, 2003). As an illustration, consider the Lorenz attractor (Figure 8) perturbed by noise. Small amounts of process noise do not dramatically change the shape of the attractor, even in the delay-coordinate space. More extreme amounts of noise definitely distort the attractor, but this does not necessarily render delay-embedding useless.

To evaluate the utility of EDM for systems with stochastic dynamics, consider a fully stochastic two-stage population model representing juveniles, J_t, and adults, A_t. In this model, the number of offspring born to each female follows a Poisson distribution, $b_i \sim \text{Poi}(\lambda_i)$, and the expected birth rate λ_i varies among females and follows a gamma distribution, $\lambda_i \sim \Gamma(\alpha, \beta)$. The number of juveniles surviving each year follows a binomial distribution with survival probability $s_j e^{-\gamma A_t}$ representing agonistic interactions with adults. The number of surviving juveniles that mature follows a binomial with maturation probability μ. The number of adults that survive also follows a binomial distribution, with survival probability s_a. The resulting adult time series are illustrated in Figure 9 for several values of α. Models of this sort are currently used in the conservation literature (e.g. Schaub *et al.*, 2007; Fujiwara and Diaz-Lopez, 2017), and it is of interest to see how well EDM might recover these dynamics, using just the time series of adults.

It is important to note that when the dynamics are stochastic, there are intrinsic limits to prediction that are not present in the deterministic case. Some error will always remain because of the stochasticity. One standard measure of model performance in this context is the mean square prediction error $V_{\text{pred}} = \text{Var}(A_{t+1} - \hat{A}_{t+1})/\text{Var}(A)$ where $\hat{A}_{t+1}$ is the predicted population size at the next time step. In this Markovian model, the conditional mean (i.e. setting $\hat{A}_{t+1} = \mathbb{E}(A_{t+1}|A_t, J_t)$) provides the lower bound on the mean square prediction error. Any other prediction will inflate the mean square error. Using the sample mean as the prediction for all time steps, i.e. setting $\hat{A}_{t+1} = \mathbb{E}(A)$ gives

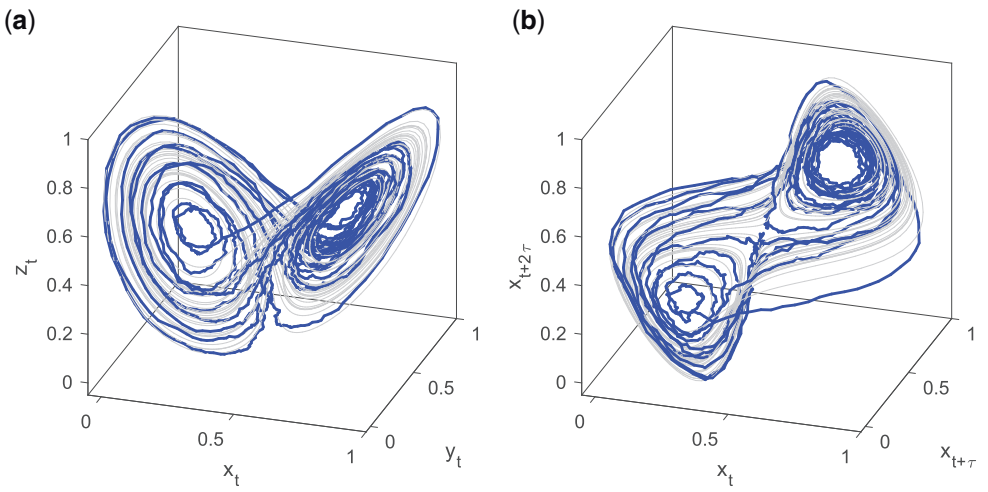

Figure 8. Lorenz attractor with process noise (dark lines) in (a) the native coordinate space and (b) the delay-coordinate space for x_t. The deterministic Lorenz attractor is shown in light grey. Low process noise will not strongly affect the nonlinear forecasts.

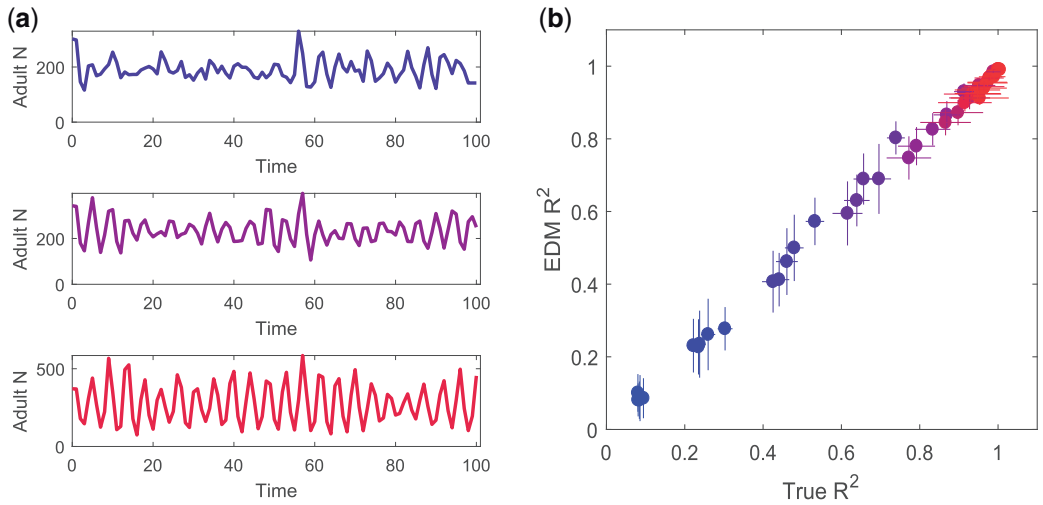

Figure 9. (a) Time series generated with a two-stage stochastic population model (parameters $\beta = 0.003$ and, from top to bottom, $\alpha = 0.13, 0.2, 0.27$). (b) Comparison of the true prediction R^2 (maximum explainable variance, see text for details) for the stochastic population model and the R^2 obtained using EDM. The colours correspond to different values of r: light gray indicates $r = 10$ and black indicates $r = 50$. Within each colour group, the different points indicate different values of β. When the dynamics are stable (light gray) most of the variation is due to noise and both the true and estimated R^2_{pred} are close to 0. As the dynamics become more nonlinear, the R^2_{pred} increases. Other parameters are $s_j = 0.5$, $s_a = 0.02$, $\gamma = 0.01$ and $\mu = 0.5$.

the sample variance, V_0. The scaled mean square error is defined as V_{pred}/V_0. As our measure of model performance, we subtract the scaled mean square error from 1 to obtain the "variance explained" by the model (where 0 is bad and 1 is best). By analogy with standard regression, we refer to this as R^2_{pred} (i.e. $R^2_{\text{pred}} = 1 - V_{\text{pred}}/V_0$). Of course, other metrics of model performance may be relevant, but this gives us an interpretable benchmark for determining how close our EDM predictions (based solely on the adult time series) can come to recovering the conditional mean.

When the dynamics are stable, most of the change in population size from one step to the next is driven by stochasticity. In this case, we expect the conditional mean to be close to the long-run mean and $R^2_{\text{pred}} \approx 0$. On the other hand, when the dynamics are cyclic or chaotic and the deterministic component of the dynamics dominates the stochastic component, the current state of the population will be useful for making predictions and we expect $R^2_{\text{pred}} \to 1$. Importantly, any model that fails to approximate the conditional mean will have a smaller value of R^2_{pred}.

To evaluate how EDM performs in this stochastic setting, we iterated the model 100 time steps and computed R^2_{pred} using the Gaussian process EDM (GP-EDM) approach of Munch *et al.* (2017) with a maximum embedding dimension of 4 applied to the simulated time series of adults. For comparison, we computed the R^2_{pred} from the original model, conditioning on the current numbers of adults and juveniles and refer to this as the true R^2. To explore a range of dynamics, we simulated the model with $\beta = 1/q$, $q = 0.01..300$, and $\alpha = rs^{-1}\beta$ with $r = 10..50$, and repeated the simulation 50 times for each parameter combination. This combination of parameters generates models that span nearly the full range of possible true R^2 values.

The results indicate a close correspondence between the true R^2 and R^2_{pred} using GP-EDM for this stochastic model (Figure 9). Our experience with other simulations is broadly similar, suggesting that stochastic dynamics are nearly as predictable with EDM as they are with a correctly specified model; the process noise increases the range of states sampled by the dynamics, which actually makes interpolation easier.

Although we have focused our answer on process stochasticity, there has also been considerable effort invested in "nonlinear noise reduction," i.e. models that explicitly deal with measurement error (Bröcker *et al.*, 2002). The most recent major advance is the Takens–Kalman filter, which uses delay-coordinate embedding in the context of an unscented Kalman filter to deal with both measurement and process uncertainties (Hamilton *et al.*, 2017).

Question 7. How can we include other factors, like temperature, in these forecasting models?

Takens' theorem was extended to driven and stochastic systems by Stark *et al.* (1997, 1999). From a practical viewpoint, this implies that the discrete-time map includes lags of the driving variable. To demonstrate how this might work and some of the potentially counter-intuitive results that can emerge, consider a two-species system, in which both species are affected by temperature, T_t.

$$\begin{aligned} x_{t+1} &= F(x_t, y_t, T_t) \\ y_{t+1} &= G(x_t, y_t, T_t). \end{aligned} \tag{8}$$

Pretending for the moment that we can invert F to solve for $y_{t-1} = F^{-1}(x_t, x_{t-1}, T_{t-1})$, we can rewrite the dynamics for x as follows:

$$x_{t+1} = F\Big(x_t, G[x_{t-1}, F^{-1}(x_t, x_{t-1}, T_{t-1}), T_{t-1}], T_t\Big). \tag{9}$$

So, the prediction for x depends not just on two lags of x but also on two lags of T. There are two things worth noting about this. The first is that the direction of the lagged temperature effect

now depends on the nature of the species interaction. If the growth of x is reduced by the presence of y, and the growth of y increases with T, then we should expect an increase in lagged T to decrease the growth of x. This is an important cautionary note for the interpretation of lagged effects more generally—the apparent effect of the lagged variable reflects both its direct effect and the cumulative impact of all the indirect effects on the species (and other state variables) that have been left out of the analysis.

The second issue is that we need to be careful about how the neighbourhoods are defined with such mixed inputs. In the S-map (local linear regression) framework (Sugihara, 1994), the Euclidean distance in the delay-coordinate space is used to weight the points to generate the locally linear map. To ensure that the distances between delay vectors are appropriate, the different variables need to be scaled correctly. There is, to our knowledge, no universal solution to this problem. However, we can make progress by introducing a scaling parameter for the environmental driver and estimating this during the course of model training (Munch *et al.*, 2017).

Question 8. Does EDM work if the environment is not stationary?

Since EDM constructs forecasts from past states of the system, nonstationarity tends to limit the time horizons over which we can make accurate predictions. However, there are several reasons why nonstationarity may not be as much of a problem as it might first seem. First, nonlinear systems often look nonstationary over a short interval. Second, even when the dynamics truly are nonstationary, reasonably good predictions may be possible if the system is not changing too fast.

We typically envision nonstationarity in ecology arising from temporal changes in vital rates, carrying capacity, species interactions, or other parameters driven by shifts in abiotic drivers, changes in community structure, or contemporary evolution. However, it is worth noting that some trajectories that appear nonstationary may actually result from nonlinear dynamics. Various nonlinear oscillators, such as the three-dimensionnal prey–predator model shown in Figure 10, can remain in distinct "regimes" for extended periods of time and then make a rapid shifts to a new regime without any changes in the parameters or external forcing (Guckenheimer and Holmes, 1983). This is not to say that nonstationarity is not a real issue for ecology but just to suggest that some apparent nonstationarity may instead reflect complexity. For a more in-depth discussion of the difficulties in identifying nonstationarity from finite time series, see Manuca and Savit (1996).

In the case where nonstationarity is driven by a variable that is changing over time, there are two possible approaches to take. If we know the driving variable, we can attempt to include the driver (e.g. temperature) as a predictor. In a mechanistic modelling framework, we would make one or more parameters a function of temperature to represent the influence of the environment on population growth. In the EDM framework, we can likewise include temperature as a predictor in the delay-coordinate system (see Question 7). Both approaches, of course, assume that the functional dependence on temperature continues to apply outside the current temperature regime (i.e. we can extrapolate). As long as the relationship between the environment and population dynamics is constant, EDM approaches can produce robust short-term predictions.

If, on the other hand, we do not know what the driving variable is, we may still be able to make short-term predictions. As long as the system behaviour does not change *qualitatively* in response to the driver, the recent past will be a good proxy for the near future. For instance, consider a two-species model where the growth rate of one species is linked to an unknown driver that is increasing through time. If we focus only on making short-term forecasts and the driver is changing slowly relative to the time scales of interest for prediction, this approach will not be too bad (Figure 11). Success depends on the rate at which the system is changing relative to the time scale over which we are trying to predict (Perretti *et al.*, 2013; Munch *et al.*, 2017).

If the exogenous driver is changing more rapidly, another approach is needed. For instance, we might allow F to change through time directly, i.e. by asserting that $F_{t+1} = F_t + \delta F$ and assigning a prior to δF (Munch *et al.*, 2017). Another possibility is to introduce a latent (i.e. unobserved) variable representing the

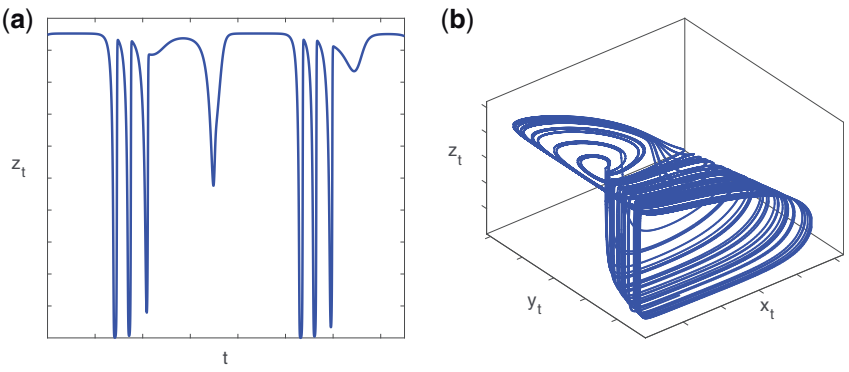

Figure 10. Deterministic three-dimensional prey–predator model. (a) Time series for prey species x. (b) System trajectory plotted in the native coordinate space. Apparent shifts between steady and fluctuating regimes are due to nonlinear dynamics. Equations for the prey density x, predator density y, and a prey trait z are $\frac{dx}{dt} = x\left(a_1 \frac{z}{1+b_1 z} - a_2 \frac{y}{1+b_2 x} - d_1\right)$, $\frac{dy}{dt} = y\left(y_a a_2 \frac{x}{1+b_2 x} - d_2\right)$ and $\frac{dz}{dt} = zV(2k_2 d_1 - 4k_4 d_1 z^2 - a_1 k_1 \frac{x}{1+b_1 z})$ from Gilpin and Feldman (2017) with $a_1 = 2.5$, $a_2 = 0.05$, $d_1 = 0.16$, $d_2 = 0.004$, $b_1 = 6$, $b_2 = \frac{4}{3}$, $k_1 = 6$, $k_2 = 9$, $k_4 = 9$, $y_a = 1$, and $V = \frac{1}{3}$.

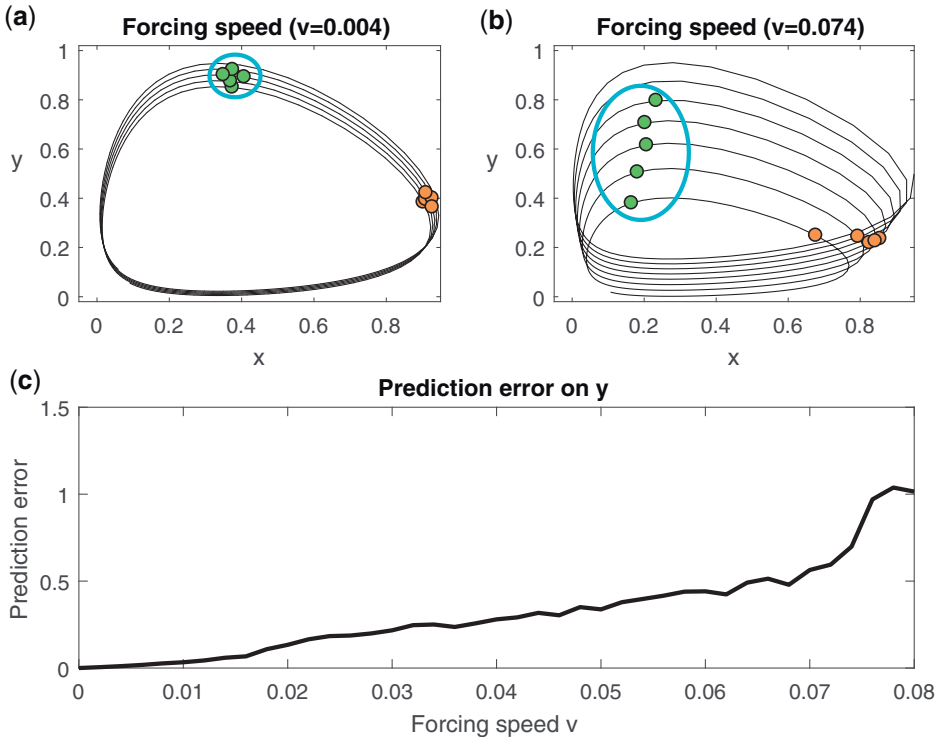

Figure 11. Prediction error in a nonstationary Lotka–Volterra predator-prey system, where the prey growth rate increases linearly over time at forcing speed v. Equations are $\frac{dx}{dt} = \frac{2}{3}\left(1 + r(t)\right)x - \frac{4}{3}xy$ and $\frac{dy}{dt} = -y + xy$ with $r(t) = vt$. (a, b) System trajectories for two different values of v, plotted in the state space. In a constant environment, the attractor would be a limit cycle with period 8; however, due to the changing growth rate, the cycle period and amplitude change over time. The white points indicate the current state in a small neighbourhood and the dark gray points are their corresponding future values (at $t + 2$). The future points are closer together when the forcing speed is low (a) than when the forcing speed is high (b). Consequently, the uncertainty (gray ellipse) of predictions increases with increased forcing speed. (c) The vertical axis is the scaled mean squared error using EDM on the time series for y from $t = 0$ to $t = 40$ (on this scale 0 means no error, 1 means the prediction error is equal to the total variance in the time series for y). Clearly, prediction error increases with v. Nevertheless, even if $v = 0.05$, such that the growth rate increases from 1 to 3 over the duration of the time series, the predictions are still useful (i.e. prediction error < 0.5).

environment into the delay-coordinate map (Verdes *et al.*, 2006). That is, $x_t = F(x_{t-1}, \ldots, x_{t-E}, u_t)$, where u_t represents environmental effects on x. We then use the time series to estimate both F and $u_1, \ldots, u_T$. Verdes *et al.* (2006) constrain u_t using a penalty analogous to the random walk priors used in many ecological models with time-varying coefficients (Congdon, 2007; Ives and Dakos, 2012). Both of these approaches can be thought of as generalizations of the time-delay embedding method that can explicitly be used to test for nonstationarity in the dynamics on the time scales of interest. Several other ideas for dealing with nonstationarity, collectively called over-embedding, can be found in the physics literature (Kantz and Schreiber, 2003).

Question 9. Can we really learn anything about biology this way?

This question probably arises from the fact that these methods have historically been framed as forecasting tools rather than general tools for ecological inference, but many useful insights can be obtained using EDM.

First, EDM can be used to evaluate whether the dynamics of a system are nonlinear (Sugihara, 1994; Sugihara *et al.*, 1999). In addition, Lyapunov exponents can be estimated from the reconstructed map to infer the dynamic stability of the system (see Abarbanel *et al.*, 1992; McCaffrey *et al.*, 1992; Benincà *et al.*, 2015; Ushio *et al.*, 2018). It seems plausible to us that analogous methods could also be used as the basis of robust tests for stationarity, although we are not aware that anyone has done so.

Second, EDM can be used to test for causal coupling between variables in complex systems (Sugihara *et al.*, 2012). For example, we can test whether two species are interacting directly (Sugihara *et al.*, 2012) and whether temperature is an important driver of population dynamics (Deyle *et al.*, 2013). The fact that inferences about these mechanisms of population dynamics are not filtered through parametric models is a great strength of EDM.

Third, since the coefficients in a local linear mapping approximate the Jacobian matrix, Deyle *et al.* (2016b) showed that they are a direct estimate of the net effect of species interactions at each time point. In this way, we can investigate how species interactions change with through time and in response to other state variables.

Fourth, by incorporating an hypothesized "mechanistic model" into the non-parametric structure (e.g. Sugihara *et al.*, 1999; Thorson *et al.*, 2014), these methods can also be used to identify—and correct for—model mis-specification. By explicitly determining how much of the residual variation around the mechanistic model is predictable using EDM, we can evaluate the

adequacy of a given model structure. Models with predictable residuals either are the wrong shape or are missing some important state variables. In contrast, residuals from a well-specified model structure should be unpredictable.

Question 10. When does EDM not work?

This is an important issue that, in our view, has not received enough attention in the ecological literature. There are, of course, several studies that conclude the methods do not work (Grenfell *et al.*, 1994; Ward *et al.*, 2014; Cobey and Baskerville, 2016), but to our knowledge, there has not been a concise summary of when we expect these methods to fail or to be outperformed by alternative approaches.

The delay embedding approach may fail either because the system in question does not meet the assumptions of the theorem or because the available data are insufficient to fully resolve the dynamics. For instance, we expect that the stationarity assumption is likely to be violated in marine ecosystems over long time scales (but see Question 8). Along these lines, EDM is not well-suited to analysing time series dominated by a monotonic trend, though some progress may be made by differencing (see Wu *et al.*, 2007). Cobey and Baskerville (2016) also note that causal inference using EDM is inhibited in the presence of strong external forcing driven by seasonality or highly correlated process noise.

Time series length and large observation errors are likely to be more problematic when applying EDM to ecological data. Since the recoverable embedding dimension scales roughly as the square root of the time series length [Cheng and Tong (1992), Question 4], this sets a practical upper bound on the dynamics that can be resolved. Moreover, since EDM requires time series that are several times longer than the characteristic return time of the system, success is more likely in systems with rapid turnover. We expect EDM to have difficulties when trying to make predictions for species whose lifespans exceed the length of the available time series (e.g. rockfishes, see Munch *et al.*, 2018). In addition, observation noise presents some practical problems, particularly when the time series is short or the observation variance is commensurate with the size of the attractor.

Linear time series models (e.g. Ives and Dakos, 2012) and dynamic linear modelling approaches (e.g. West and Harrison, 1997; Carpenter and Brock, 2006) are likely to outperform EDM when the system is strongly nonstationary or the observation noise is fairly large. The additional information provided by the parametric structure of these models helps compensate for these deficiencies in the data. More specific mechanistic models can be expected to outperform EDM when the model structure is a good approximation of the true dynamics. That said, mechanistic models with missing state variables or incorrect structure can also be quite misleading. *Post hoc* checks for the predictability of model errors may be helpful in identifying model deficiencies.

Conclusions

Although EDM has been described as non-mechanistic (Jabot, 2015; Lagergren *et al.*, 2018), this viewpoint essentially confounds the action of specific mechanisms with the existence of a set of equations describing them. In our view, EDM represents an alternative perspective in mechanistic modelling, one that regards the observed attractor as the fundamental description of the dynamics, rather than a prescribed set of equations.

Ecology is currently experiencing rapid growth as a quantitative discipline. Incredibly complex models can now be fit to data with relatively little effort. Some intuition for nonlinear dynamics is indispensable in formulating and evaluating the performance of these models. Although useful in its own right, EDM can also help in constructing and validating parametric models. In the early stages of model development, CCM (Sugihara *et al.*, 2012) and S-map estimates of interactions strengths (Deyle *et al.*, 2016b) can be useful tools for identifying the important components of multi-species models. Following model development, residual delay maps (Sugihara *et al.*, 1999) are useful for identifying unexplained structure and Gaussian process (GP) regression (Thorson *et al.*, 2014) is useful for identifying and compensating for model mis-specification. Rather than opposing viewpoints, we see EDM and parametric mechanistic modelling as highly complementary.

Although training in ecology has become increasingly quantitative, EDM is still something that ecologists typically have to learn on their own. Each of the topics we have addressed represent questions that we wrestled with when we began learning EDM. We hope that the answers we have provided are useful to other quantitative ecologists and facilitate future applications.

Supplement

Attractors

The trajectories of most dynamical systems of interest eventually converge to an attractor, which may be a point, a cycle, or a more complex shape in the state space. The attractor typically does not fill the entire M-dimensional state space (i.e. the attractor has a lower dimensionality than the space in which it exists). This has several implications. First, because we will likely only have observations $\mathbf{x}_t$ near the attractor, we will not be able to infer $F_i(\mathbf{x}_t)$ over all possible values of $\mathbf{x}_t$ due to the lack of data off the attractor. However, the states near the attractor are likely the ones that are the most relevant, as these are the neighbourhood the system is likely to visit again in the future. Modelling dynamics in this vicinity also allows us to make efficient use of the available data.

On the other hand, the absence of data far away from the attractor limits any statistical attempt to infer ecological dynamics from the time series, parametric or otherwise. For example, if the attractor is a stable fixed point, there is not much we can learn about the dynamics, regardless of the tools we use. Near a stable fixed point, the dynamics will be well described by a linear approximation; any explicit model capable of producing a fixed point with a similar Jacobian matrix (see the answer to Question 1) will appear to fit the data. On the other hand, when the attractor is a more complex object, e.g. a limit cycle or strange attractor, then there is more hope of learning something deeper from the observed fluctuations in the state variables through time.

Lyapunov exponents

The stability of a system along a trajectory may be characterized by the convergence of nearby trajectories. In the general case, this is determined by the collection of Lyapunov exponents for the system, which can be thought of as a generalization of eigenvalues, which are used to characterize stability in linear systems. Specifically, the distance ϵ between two points on the attractor that are initially close grows (or shrinks) approximately as

$$\epsilon_t \approx e^{\lambda t}\epsilon_0, \tag{10}$$

where λ is the dominant Lyapunov exponent. The distance grows (and hence the system is unstable or chaotic) if $\lambda > 0$. When the Lyapunov exponent for a system is positive, small errors in the

initial state estimate grow exponentially such that over sufficiently long time scales (roughly $1/\lambda$), the system is effectively unpredictable. This sensitivity to initial conditions is the hallmark of chaos. This exponential growth approximation is only good over short time intervals for nonlinear systems; since both initial points are on the attractor, there is a finite limit to the distance between them set by the size of the attractor. This is in contrast to an unstable linear system in which ϵ will grow without bound. Below, we present an informal derivation of Lyapunov exponents and their connection to eigenvalues, which we assume are more familiar. For clarity, we begin in 1-d and then generalize.

Let us say a discrete-time system is governed by the map $x_{t+1} = F(x_t)$. We start with the pair of nearby points x_0 and $y_0 = x_0 + \epsilon_0$, where $|\epsilon_0|$ is the initial distance between them. One step into the future, the distance is $|\epsilon_1| = |F(y_0) - F(x_0)|$. Using a first-order approximation of F at x_0, the distance is approximately $\epsilon_1 \approx F'(x_0)\epsilon_0$ where $F'(x)$ is the derivative of F at the point x. If $|F'(x_0)|$ is >1, then $|\epsilon_1|$ will be larger than $|\epsilon_0|$.

Two steps into the future, $\epsilon_2 = F(F(y_0)) - F(F(x_0))$. Again, using a linear approximation around x, the distance is $|\epsilon_2| \approx |F'(F(x_0))F'(x_0)\epsilon_0|$, where $F'(F(x_0)) = F'(x_1)$ is the derivative evaluated at the next point on the trajectory starting from x_0. So it is the product $|F'(x_1)F'(x0)|$ that determines whether ϵ_2 grows or shrinks compared with ϵ_0. Typically, the per-time-step expansion or contraction is used, which is $|F'(x_1)F'(x_0)|^{1/2}$.

Note that if x_0 is a fixed point (i.e. $F(x_0) = x_0$), then the derivative does not change through time, so the per-time-step expansion is determined entirely by $|F'(x_0)|$. If x_0 is on a limit cycle with period τ, such that $x_\tau = x_0$, then the stability of the *cycle* is given by $[\prod_{i=0}^{\tau-1} |F'(x_i)|]^{1/\tau}$. Extending this argument for more complicated attractors, stability is characterized using the long-run limit of this product, i.e. $\lim_{t\to\infty} \prod_{i=0}^{t-1} |F'(x_i)|^{1/t}$. The natural log of this defines the Lyapunov exponent, which is given by

$$\lambda = \lim_{t\to\infty} \frac{1}{t} \sum_{i=0}^{t-1} \ln|F'(x_i)|. \tag{11}$$

This expression is identical to taking the average of $\ln|F'(x)|$ over the stationary distribution or "invariant measure" for x, which provides straightforward recipe for calculating the Lyapunov exponent when we have a 1-d model: at each iteration, evaluate $\ln|F'(x_t)|$ and take the average over a long enough interval to obtain convergence. In the main text, we do this for multiple, randomly selected starting values to avoid accidentally starting on unstable fixed points or limit cycles.

For systems with n state variables, there are n Lyapunov exponents. The largest of these is the "dominant Lyapunov exponent," which determines stability. To find the Lyapunov exponents, we again start with pair of nearby points $\mathbf{x}_0$ and $\mathbf{y}_0 = \mathbf{x}_0 + \epsilon_0$ and think about how the distance between these points grows over time. By analogy with the scalar case, this is determined by $[\prod_{i=0}^{t-1} \mathbf{J}(\mathbf{x_i})]^{1/t}$ where $\mathbf{J}(\mathbf{x_i})$ is the Jacobian matrix evaluated at the state $\mathbf{x_i}$. As $t \to \infty$, the eigenvalues of this long-run product determine the directions in which $\boldsymbol{\epsilon}$ grows and shrinks. As in the scalar case, if $\mathbf{x}_0$ is a fixed point, the eigenvalues of $\mathbf{J}(\mathbf{x}_0)$ completely characterize stability. For a limit cycle, stability is determined by the product of Jacobians over the sequence of states visited.

For a chaotic system, although it is possible to evaluate the Jacobian of a model at each point on a trajectory, calculating the long-run product directly and then taking its eigenvalues is not numerically stable. So, to avoid numerical artefacts, a QR algorithm is typically used (Eckmann and Ruelle, 1985). This is what was done in the main text to estimate the Lyapunov exponents for systems with two or more state variables. To avoid transients or unstable cycles, the system was randomly initialized and run for 1000 steps before computing the Lyapunov exponents.

Continuous time

For simplicity, the main text has focused on dynamics in discrete time. Nevertheless, this material applies to continuous-time systems as well. In this study, we describe the connection between continuous- and discrete-time systems.

Imagine we have an n-dimensional autonomous system in which the derivative of the ith state variable with respect to time is given by

$$\frac{dx_i}{dt} = f_i(\mathbf{x}_t). \tag{12}$$

If we want to know the value of each state variable x_i at the future time $t + \tau$, we could solve (12) by integrating. Doing so generates a discrete-time "map" over the time step τ, i.e. $x_{i,t+\tau} = F_i(\mathbf{x}_t, \tau)$. This map is, in general, a function of τ.

In the main text, we mention the Jacobian for a discrete-time system in several places. When observing a continuous-time system at discrete intervals, it may be useful to connect the Jacobian obtained over a discrete-time step and the Jacobian for the continuous-time dynamics. Let $J_{i,j} = \frac{\partial F_i}{\partial x_j}$ represent the Jacobian matrix for the discrete-time dynamics observed over time step τ and let $A_{i,j} = \frac{\partial f_i}{\partial x_j}$ be the Jacobian for the continuous-time system. Then

$$\boldsymbol{J} \approx e^{\boldsymbol{A}\tau}. \tag{13}$$

By analogy, the calculation of the Lyapunov exponents in a continuous-time setting involves, at least theoretically, the eigenvalues of $\boldsymbol{J}$ in the limit as $\tau \to \infty$. That is, if $\Lambda(\tau)$ are the eigenvalues of $\boldsymbol{J}(\tau)$, then the Lyapunov exponents are defined as $\lambda_i = \lim_{\tau\to\infty} \frac{1}{\tau} \ln(\Lambda_i(\tau))$.

Glossary

Attractor	A set of values towards which a dynamical system tends to converge. In continuous time, this could be a single point corresponding to a stable equilibrium, a closed loop corresponding to a limit cycle, or a more complex set corresponding to chaotic dynamics. In discrete time, limit cycles are finite collections of points, which are repeated indefinitely
Autonomous system	A system whose behaviour is not influenced by external forcing
Chaos	Deterministic dynamical systems that exhibit sensitive dependence to initial conditions but produce bounded

Continued

	trajectories are chaotic. These systems typically have at least one positive Lyapunov exponent and a complex or "strange" attractor. Because sensitivity to initial conditions effectively limits the time horizon over which a useful prediction can be made, chaotic dynamics often appear random over sufficiently long time scales
Dynamical system	Any system that changes through time according to some underlying governing equations
Embedding dimension	The number of lags of a single variable required to construct a 1:1 and invertible projection of the original attractor in the EDM framework
Empirical dynamic modelling	A type of nonlinear forecasting that uses non-parametric methods to reconstruct the dynamics of a system directly from time series data. Frequently, time lags are used as predictors to account for unobserved state variables and also known as state-space reconstruction, attractor reconstruction, and time-delay embedding
Jacobian matrix	The matrix of all first partial derivatives of a vector-valued function, that is $J_{i,j} = \partial F_i(\mathbf{x})/\partial x_j$. Just as the derivative determines a line tangent to a function at a point in 1-d, the Jacobian specifies the "tangent plane" approximation in more than 1-d. It is used to construct a linear approximation to the dynamics of a nonlinear system, either near a fixed point or along a trajectory. As in linear dynamical systems, the eigenvalues of the Jacobian matrix provide information about stability
Map	In discrete time, the "map" is the function or set of functions that convert the current state of the system into the next state
Non-parametric	Non-parametric approaches to model fitting make minimal assumptions about the shape of the function being approximated. Splines, Gaussian processes, neural networks, local linear models, kernel smoothers, and basis function expansions are considered non-parametric regression approaches, even though they involve estimating some number of parameters
Nonlinear dynamics	A system displays nonlinear dynamics if the state variables evolve through time in a way that is not directly (linearly) proportional to the values of those variables
Nonlinear forecasting	Predicting the future state of a system using either a mechanistic nonlinear model or a model-free approach based on the observed dynamics
Prediction error	The deviation between the future state of the model and observed data. Usually this is summarized as the variance of the difference (i.e. observed value−model value) and is often referred to as the mean square error. The mean square error ranges from 0 to infinity. To provide a more intuitive scaling for the prediction error, we divide by the total variance in the observed values and subtract from 1, so that a perfect model has a score of 1, a poor model has values closer to 0, and model that is worse than just using the mean will have negative values (which can occur when the population and model are oscillating but out of phase)
Stability	The tendency of a system to converge to an equilibrium point or other attractor. An equilibrium is locally stable if the system will return to it following a small perturbation
State space	The set of all possible configurations of a system. The state of the system can be represented as a point within this space where the axes are the state variables

Continued

Acknowledgements

We thank Sarah Glaser, Ethan Deyle, and Hao Ye for useful discussions early in the development of this manuscript. Andrew Hein, Ben Martin, Marm Kilpatrick, Hao Ye, Celia Symons, Franck Jabot, Jean-Denis Mathias, Guillaume Deffuant, and two anonymous reviewers provided useful feedback that helped make this manuscript more readable. Financial support for this work was generously provided by the Lenfest Oceans Program.

References

Abarbanel, H. D., Brown, R., and Kennel, M. B. 1992. Local Lyapunov exponents computed from observed data. Journal of Nonlinear Science, 2: 343.

Abarbanel, H. D., Rulkov, N. F., and Sushchik, M. M. 1996. Generalized synchronization of chaos: the auxiliary system approach. Physical Review E Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics, 53: 4528–4535.

Ackleh, A. S., and Salceanu, P. L. 2015. Competitive exclusion and coexistence in an n-species Ricker model. Journal of Biological Dynamics, 9: 321–331.

Alligood, K. T., Sauer, T. D., and Springer, J. A. Y. 1996. CHAOS: An Introduction to Dynamical Systems. Springer, New York.

Becks, L., Hilker, F. M., Malchow, H., Jürgens, K., and Arndt, H. 2005. Experimental demonstration of chaos in a microbial food web. Nature, 435: 1226–1229.

Benettin, G., Galgani, L., Giorgilli, A., and Strelcyn, J. M. 1980. Lyapunov Characteristic Exponents for smooth dynamical systems and for Hamiltonian systems; A method for computing all of them. Part 2: numerical application. Meccanica, 15: 21–30.

Benincà, E., Ballantine, B., Ellner, S. P., and Huisman, J. 2015. Species fluctuations sustained by a cyclic succession at the edge of chaos. Proceedings of the National Academy of Sciences of the United States of America, 112: 6389–6394.

Benincà, E., Huisman, J., Heerkloss, R., Jöhnk, K. D., Branco, P., Van Nes, E. H., Scheffer, M. *et al.* 2008. Chaos in a long-term experiment with a plankton community. Nature, 451: 822–825.

Berryman, A. A., and Millstein, J. A. 1989. Are ecological systems chaotic - And if not, why not? Trends in Ecology and Evolution, 4: 26–28.

Boettiger, C., Mangel, M., and Munch, S. 2015. Avoiding tipping points in fisheries management through Gaussian process dynamic programming. Proceedings of the Royal Society B, 282: 8–11.

Bröcker, J., Parlitz, U., and OgorzaŁek, M. 2002. Nonlinear noise reduction. Proceedings of the IEEE, 90: 898–918.

Carpenter, S. R., and Brock, W. A. 2006. Rising variance: a leading indicator of ecological transition. Ecology Letters, 9: 311–318.

Chang, C. W., Ushio, M., and Hsieh, C. H. 2017. Empirical dynamic modeling for beginners. Ecological Research, 32: 785–796.

Cheng, B., and Tong, H. 1992. On consistent nonparametric order determination and chaos. Journal of the Royal Statistical Society. Series B (Methodological), 54: 427–449.

Cobey, S., and Baskerville, E. B. 2016. Limits to causal inference with state-space reconstruction for infectious disease. PLoS One, 11: e0169050-22.

Congdon, P. 2007. Bayesian Statistical Modelling, 2nd edn. John Wiley & Sons, Chichester, UK.

Costantino, R. F., Desharnais, R. A., Cushing, J. M., and Dennis, B. 1997. Chaotic dynamics in an insect population. Science, 275: 389–391.

Deyle, E. R., Fogarty, M., Hsieh, C.-H., Kaufman, L., MacCall, A. D., Munch, S. B., Perretti, C. T. *et al.* 2013. Predicting climate effects on Pacific sardine. Proceedings of the National Academy of Sciences of the United States of America, 110: 6430–6435.

Deyle, E. R., Maher, M. C., Hernandez, R. D., Basu, S., and Sugihara, G. 2016a. Global environmental drivers of influenza. Proceedings of the National Academy of Sciences of the United States of America, 113: 13081–13086.

Deyle, E. R., May, R. M., Munch, S. B., and Sugihara, G. 2016b. Tracking and forecasting ecosystem interactions in real time. Proceedings of the Royal Society B: Biological Sciences, 283: 20152258.

Deyle, E. R., and Sugihara, G. 2011. Generalized theorems for nonlinear state space reconstruction. PLoS One, 6: e18295.

Dietze, M. C. 2017. Ecological Forecasting. Princeton University Press, Princeton, USA.

Dymarski, P. 2011. Hidden Markov Models. Intech Open, Rijeka, Croatia.

Eckmann, J. P., and Ruelle, D. 1985. Ergodic theory of chaos and strange attractors. Reviews of Modern Physics, 57: 617.

Ellner, S., and Turchin, P. 1995. Chaos in a noisy world: new methods and evidence from time-series analysis. The American Naturalist, 145: 343–375.

Farmer, J. D., and Sidorowich, J. J. 1989. Exploiting chaos to predict the future and reduce noise. *In* Evolution, Learning and Cognition, pp. 277–330. Ed. by Y. C. Lee. World Scientific, London.

Fogarty, M. J., Gamble, R., and Perretti, C. T. 2016. Dynamic complexity in exploited marine ecosystems. Frontiers in Ecology and Evolution, 4: 1–20.

Fujiwara, M., and Diaz-Lopez, J. 2017. Constructing stage-structured matrix population models from life tables: comparison of methods. PeerJ, 5: e3971.

Fukaya, K., and Royle, J. A. 2013. Markov models for community dynamics allowing for observation error. Ecology, 94: 2670.

Garland, J., James, R. G., and Bradley, E. 2016. Leveraging information storage to select forecast-optimal parameters for delay-coordinate reconstructions. Physical Review E, 93: 022221.

Gilpin, W., and Feldman, M. W. 2017. A phase transition induces chaos in a predator-prey ecosystem with a dynamic fitness landscape. PLoS Computational Biology, 13: e1005644-20.

Glaser, S. M., Fogarty, M. J., Liu, H., Altman, I., Hsieh, C. H., Kaufman, L., Maccall, A. D. *et al.* 2014. Complex dynamics may limit prediction in marine fisheries. Fish and Fisheries, 15: 616–633.

Graham, D. W., Knapp, C. W., Van Vleck, E. S., Bloor, K., Lane, T. B., and Graham, C. E. 2007. Experimental demonstration of chaotic instability in biological nitrification. ISME Journal, 1: 385–393.

Grenfell, B. T., Kleczkowski, A., Ellner, S. P., and Bolker, B. M. 1994. Measles as a case study in nonlinear forecasting and chaos. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 348: 515–530.

Gross, T., Ebenhöh, W., and Feudel, U. 2005. Long food chains are in general chaotic. Oikos, 109: 135–144.

Guckenheimer, J., and Holmes, P. 1983. Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields. Springer, New York, USA.

Hamilton, F., Berry, T., and Sauer, T. 2017. Kalman-Takens filtering in the presence of dynamical noise. European Physical Journal: Special Topics, 226: 3239–3250.

Hartmann, M., Hosack, G. R., Hillary, R. M., and Vanhatalo, J. 2017. Gaussian process framework for temporal dependence and discrepancy functions in Ricker-type population growth models. Annals of Applied Statistics, 11: 1375–1402.

Hassell, M. P., and Comins, H. N. 1976. Discrete time models for two-species competition. Theoretical Population Biology, 9: 202.

Hastings, A., Hom, C. L., Ellner, S., Turchin, P., and Godfray, H. C. J. 1993. Chaos in ecology: is mother nature a strange attractor? Annual Review of Ecology and Systematics, 24: 1–33.

Hsieh, C., Anderson, C., and Sugihara, G. 2008. Extending nonlinear analysis to short ecological time series. The American Naturalist, 171: 71–80.

Ives, A. R., and Dakos, V. 2012. Detecting dynamical changes in nonlinear time series using locally linear state-space models. Ecosphere, 3: art58.

Jabot, F. 2015. Why preferring parametric forecasting to nonparametric methods? Journal of Theoretical Biology, 372: 205–210.

Judd, K. 2008. Shadowing pseudo-orbits and gradient descent noise reduction. Journal of Nonlinear Science, 18: 57–74.

Judd, K., and Mees, A. 1998. Embedding as a modeling problem. Physica D: Nonlinear Phenomena, 120: 273–286.

Kantz, H., and Schreiber, T. 2003. Nonlinear Time Series Analysis. Cambridge University Press, Cambridge, UK.

Lagergren, J., Reeder, A., Hamilton, F., Smith, R. C., and Flores, K. B. 2018. Forecasting and uncertainty quantification using a hybrid of mechanistic and non-mechanistic models for an age-structured population model. Bulletin of Mathematical Biology, 80: 1578–1595.

Lillacci, G., and Khammash, M. 2010. Parameter estimation and model selection in computational biology. PLoS Computational Biology, 6: e1000696.

Link, J. 2002. Does food web theory work for marine ecosystems? Marine Ecology Progress Series, 230: 1–9.

Luo, Y., Ogle, K., Tucker, C., Fei, S., Gao, C., LaDeau, S., Clark, J. S. *et al.* 2011. Ecological forecasting and data assimilation in a data-rich era. Ecological Applications, 21: 1429–1442.

Manuca, R., and Savit, R. 1996. Stationarity and nonstationarity in time series analysis. Physica D: Nonlinear Phenomena, 99: 134–161.

Massoud, E. C., Huisman, J., Benincà, E., Dietze, M. C., Bouten, W., and Vrugt, J. A. 2018. Probing the limits of predictability: data assimilation of chaotic dynamics in complex food webs. Ecology Letters, 21: 93–103.

May, R. M. 1976. Simple mathematical models with very complicated dynamics. Nature, 261: 459.

McCaffrey, D. F., Ellner, S., Gallant, A. R., and Nychka, D. W. 1992. Estimating the Lyapunov exponent of a chaotic system with nonparametric regression. Journal of the American Statistical Association, 87: 682–695.

Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., and Fryxell, J. M. 2004. Extracting more out of relocation data: building

movement models as mixtures of random walks. Ecology, 85: 2436.

Munch, S. B., Giron-Nava, A., and Sugihara, G. 2018. Nonlinear dynamics and noise in fisheries recruitment: a global meta-analysis. Fish and Fisheries, 19: 964–973.

Munch, S. B., Poynor, V., and Arriaza, J. L. 2017. Circumventing structural uncertainty: a Bayesian perspective on nonlinear forecasting for ecology. Ecological Complexity, 32: 134.

Neal, R. M. 1997. Monte Carlo Implementation of Gaussian Process Models for Bayesian Regression and Classification. Technical Report 9702, Department of Statistics, University of Toronto.

Newman, K., Buckland, S., Morgan, B., King, R., Borchers, D., Cole, D., Besbeas, P. *et al.* 2014. Modelling Population Dynamics: Model Formulation, Fitting and Assessment Using State-Space Methods. Springer, New York, USA.

Niu, S., Luo, Y., Dietze, M. C., Keenan, T. F., Shi, Z., Li, J., and Chapin, F. S. 2014. The role of data assimilation in predictive ecology. Ecosphere, 5: 1–16.

Perretti, C. T., Munch, S. B., and Sugihara, G. 2013. Model-free forecasting outperforms the correct mechanistic model for simulated and experimental data. Proceedings of the National Academy of Sciences of the United States of America, 110: 5253–5257.

Pierre, M., Rouyer, T., Bonhommeau, S., and Fromentin, J. M. 2018. Assessing causal links in fish stock-recruitment relationships. ICES Journal of Marine Science, 75: 903–911.

Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Bulletin of the Fisheries Research Board of Canada, Number 191. 382 pp.

King Ruth, Morgan, B. J. T., Gimenez, O., and Brooks, S. P. 2010. Bayesian Analysis for Population Ecology. Chapman & Hall/CRC.

Schaffer, W. M., and Kot, M. 1986. Chaos in ecological systems: the coals that Newcastle forgot. Trends in Ecology and Evolution, 1: 58–63.

Schaub, M., Gimenez, O., Sierro, A., and Arlettaz, R. 2007. Use of integrated modeling to enhance estimates of population dynamics obtained from limited data. Conservation Biology, 21: 945–955.

Shelton, A. O., and Mangel, M. 2011. Fluctuations of fish populations and the magnifying effects of fishing. Proceedings of the National Academy of Sciences of the United States of America, 108: 7075–7080.

Sibly, R. M., Barker, D., Hone, J., and Pagel, M. 2007. On the stability of populations of mammals, birds, fish and insects. Ecology Letters, 10: 970–976.

Stark, J. 1999. Delay embeddings for forced systems. I. Deterministic forcing. Journal of Nonlinear Science, 9: 255–332.

Stark, J., Broomhead, D. S., Davies, M. E., and Huke, J. 1997. Takens embedding theorems for forced and stochastic systems. Nonlinear Analysis, Theory, Methods and Applications, 30: 5303–5314.

Stone, L. 2018. The feasibility and stability of large complex biological networks: a random matrix approach. Scientific Reports, 8: 1–12.

Stow, C. A., Lamon, E. C., and Carpenter, S. R. 1998. Forecasting Pcb concentrations in lake michigan salmonids: a dynamic linear model approach. Ecological Applications, 8: 659–668.

Sugihara, G. 1994. Nonlinear forecasting for the classification of natural time series. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 348.

Sugihara, G., Casdagli, M., Habjan, E., Hess, D., Dixon, P., and Holland, G. 1999. Residual delay maps unveil global patterns of atmospheric nonlinearity and produce improved local forecasts. Proceedings of the National Academy of Sciences of the United States of America, 96: 14210–14215.

Sugihara, G., and May, R. M. 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature, 344: 734.

Sugihara, G., May, R., Ye, H., Hsieh, C. H., Deyle, E., Fogarty, M., and Munch, S. 2012. Detecting causality in complex ecosystems. Science, 338: 496–500.

Tajima, S., Yanagawa, T., Fujii, N., and Toyoizumi, T. 2015. Untangling brain-wide dynamics in consciousness by cross-embedding. PLoS Computational Biology, 11: e1004537-28.

Takens, F. 1981. Detecting strange attractors in turbulence. *In* Dynamical Systems and Turbulence, pp. 366–381. Ed. by D. Rand and L. S. Young. Springer, New York.

Thorson, J. T., Ono, K., and Munch, S. B. 2014. A Bayesian approach to identifying and compensating for model misspecification in population models. Ecology, 95: 329–341.

Upadhyay, R., Iyengar, S., and Rai, V. 1998. Chaos: an ecological reality? International Journal of Bifurcation and Chaos, 8: 1325–1323.

Ushio, M., Hsieh, C. H., Masuda, R., Deyle, E. R., Ye, H., Chang, C. W., Sugihara, G. *et al.* 2018. Fluctuating interaction network and time-varying stability of a natural fish community. Nature, 554: 360–363.

Vano, J. A., Wildenberg, J. C., Anderson, M. B., Noel, J. K., and Sprott, J. C. 2006. Chaos in low-dimensional Lotka-Volterra models of competition. Nonlinearity, 19: 2391.

Verdes, P. F., Granitto, P. M., and Ceccatto, H. A. 2006. Overembedding method for modeling nonstationary systems. Physical Review Letters, 96: 1–4.

Wan, E., and van der Merwe, R. 2001. The unscented Kalman filter. *In* Kalman Filtering and Neural Networks, Chapter 7, pp. 221–280. Ed. by S. Haykin. Wiley, New York.

Ward, E. J., Holmes, E. E., Thorson, J. T., and Collen, B. 2014. Complexity is costly: a meta-analysis of parametric and non-parametric methods for short-term population forecasting. Oikos, 123: 652–661.

West, M., and Harrison, J. 1997. Bayesian Forecasting and Dynamic Models. Springer, New York, USA.

Wildenberg, J. C., Vano, J. A., and Sprott, J. C. 2006. Complex spatiotemporal dynamics in Lotka-Volterra ring systems. Ecological Complexity, 3: 140.

Wood, S. N. 2010. Statistical inference for noisy nonlinear ecological dynamic systems. Nature, 466: 1102–1104.

Wu, Z., Huang, N. E., Long, S. R., and Peng, C.-K. 2007. On the trend, detrending, and variability of nonlinear and nonstationary time series. Proceedings of the National Academy of Sciences of the United States of America, 104: 14889–14894.

Ye, H., Beamish, R. J., Glaser, S. M., Grant, S. C. H., Hsieh, C-H., Richards, L. J., Schnute, J. T. *et al.* 2015. Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. Proceedings of the National Academy of Sciences of the United States of America, 112: E1569–E1576.

Ye, H., Clark, A., Deyle, E., and Munch, S. 2019. rEDM: Applications of Empirical Dynamic Modeling from Time Series. https://ha0ye.github.io/rEDM, https://github.com/ha0ye/rEDM.

Ye, H., and Sugihara, G. 2016. Information leverage in interconnected ecosystems: overcoming the curse of dimensionality. Science, 353: 922–925.

Yu, X., and Liu, L. 2016. The stationary distribution and extinction of the n-dimensional stochastic Lotka-Volterra predator-prey system. In Proceedings of 6th International Conference on Intelligent Control and Information Processing, ICICIP 2015, 2015: 57–61.

Zhao, H., Sun, Y., and Wang, Z. 2014. Control of Hopf bifurcation and Chaos in a delayed Lotka-Volterra predator-prey system with time-delayed feedbacks. Abstract and Applied Analysis, 2014: 1.

Handling editor: Gary Griffith