CrossMark

Chun-Wei Chang · Masayuki Ushio · Chih-hao Hsieh

Empirical dynamic modeling for beginners

Abstract Natural systems are often complex and dynamic (i.e. nonlinear), making them difficult to understand using linear statistical approaches. Linear approaches are fundamentally based on correlation. Thus, they are ill-posed for dynamical systems, where correlation can occur without causation, and causation may also occur in the absence of correlation. "Mirage correlation" (i.e., the sign and magnitude of the correlation change with time) is a hallmark of nonlinear systems that results from state dependency. State dependency means that the relationships among interacting variables change with different states of the system. In recent decades, nonlinear methods that acknowledge state dependence have been developed. These nonlinear statistical methods are rooted in state space reconstruction, i.e. lagged coordinate embedding of time series data. These methods do not assume any set of equations governing the system but recover the dynamics from time series data, thus called empirical dynamic modeling (EDM). EDM bears a variety of utilities to investigating dynamical systems. Here, we provide a step-by-step tutorial for EDM applications with rEDM, a free software package written in the R language. Using model examples, we aim to guide users through several basic applications of EDM, including (1) determining the complexity (dimensionality) of a system, (2) distinguishing nonlinear dynamical systems from linear stochastic systems, and quantifying the nonlinearity (i.e. state dependence), (3) determining causal variables, (4) forecasting, (5) tracking the strength and sign of interaction, and (6) exploring the scenario of external perturbation. These methods and applications can be used to provide a mechanistic understanding of dynamical systems.

C.-W. Chang (✉) · C. Hsieh (✉)
Taiwan International Graduate Program (TIGP)–Earth System
Science Program, Academia Sinica, Taipei 11529, Taiwan
E-mail: clark.ms37@yahoo.com.tw
E-mail: chsieh@ntu.edu.tw

C.-W. Chang
Taiwan International Graduate Program (TIGP)–Earth System
Science Program, National Central University, Taoyuan 32001,
Taiwan

M. Ushio (✉)
PRESTO, Japan Science and Technology Agency, Kawaguchi 332-0012, Japan
E-mail: ong8181@gmail.com

M. Ushio
Center for Ecological Research, Kyoto University, Otsu 520-2113,
Japan

C. Hsieh
Institute of Oceanography, National Taiwan University, Taipei
10617, Taiwan

C. Hsieh
Research Center for Environmental Changes, Academia Sinica,
Taipei 11529, Taiwan

C. Hsieh
Institute of Ecology and Evolutionary Biology, Department of Life
Science, National Taiwan University, Taipei 10617, Taiwan

C. Hsieh
National Center for Theoretical Sciences, Taipei 10617, Taiwan

Keywords Embedding · State space reconstruction · State dependence · Forecast · Interaction

Introduction

『月暈而風，礎潤而雨』~宋，蘇洵：辨奸論。 A famous old Chinese saying: "*A halo around the moon indicates the rising of wind; the damp on a plinth is a sign of approaching rain*" is believed to be written by Xun Su (approximated to appear in 1069AD in the Sung Dynasty of China). However, a halo does not cause wind; wind does not cause the halo, either! Two things appear
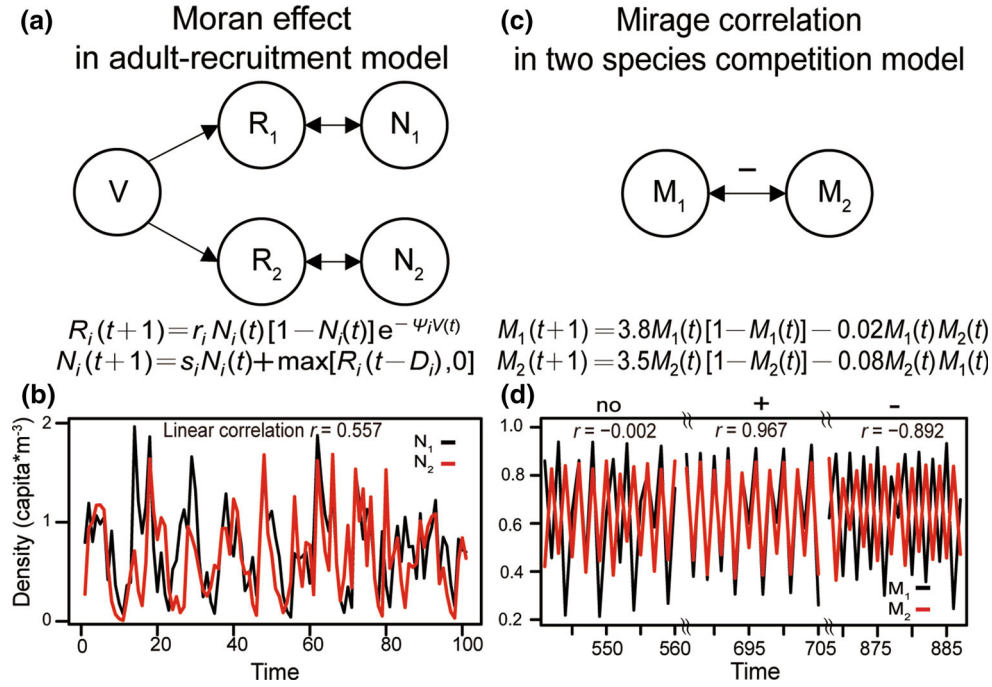
(a) Moran effect in adult-recruitment model

$$R_i(t+1) = r_i N_i(t)[1-N_i(t)]e^{-\Psi_i V(t)}$$
$$N_i(t+1) = s_i N_i(t) + \max[R_i(t-D_i), 0]$$

(b) Linear correlation r = 0.557

(c) Mirage correlation in two species competition model

$$M_1(t+1) = 3.8M_1(t)[1-M_1(t)] - 0.02M_1(t)M_2(t)$$
$$M_2(t+1) = 3.5M_2(t)[1-M_2(t)] - 0.08M_2(t)M_1(t)$$

(d) no $r = -0.002$ + $r = 0.967$ − $r = -0.892$

Fig. 1 Model examples demonstrating a confusing conclusion based on linear correlation analysis. The first model is a two-species adult (N)-recruitment (R) fishery model (**a**). Both species are driven by a shared environmental driver (V). Although no interaction exists between species N_1 and N_2, their dynamics show a strong positive correlation (**b**), driven by the shared environmental force. The second model is a two-species competition model (M_1 and M_2), demonstrating mirage correlation (**c**). Although M_1 and M_2 have a fixed negative interaction, the sign of correlation between their dynamics changes over time (**d**). In **a**, **c**, the *arrow* indicates causal interaction, with the cause pointing to the effect. The model examples are modified from Sugihara et al. (2012)

to be highly correlated, but there is no causal relationship between them. That is, correlation does not imply causation. This can be demonstrated using a simple model of two independent populations driven by the same external forcing (Fig. 1a, b). As can be seen from the model, two species show strong correlation, although they do not interact. This strong correlation is simply driven by a third, shared component (e.g., the environment). This is analogous to the well-known Moran effect (Moran 1953).

Even more counter-intuitively, a lack of correlation does not imply lack of causation. This can be demonstrated using a two-species competition model (Fig. 1c). The two species exhibit a mirage correlation (Fig. 1d): positive correlation for a period of time, negative correlation for another period of time, and then no correlation in yet another period of time. If one uses correlation to infer causality, one may erroneously conclude that the two species have no causal interaction. Such a "mirage correlation" is a hallmark of nonlinear dynamical systems (Sugihara et al. 2012).

Mirage correlations result from a fundamental property of nonlinear dynamical systems known as state dependency (Sugihara et al. 2012; Ye et al. 2015a). State dependency means that the relationships among interacting variables change with different states of the dynamical system (Ye et al. 2015a). For example, the sign of the correlation between two variables may

change with different system states, and therefore appears to change with time (i.e., mirage correlation; Fig. 1d). State dependent behavior is clearly demonstrated in the Lorenz butterfly attractor (Lorenz 1963), in which two variables exhibit opposite correlations when they are on different lobes of the butterfly attractor, i.e., different system states depending on the state of a third variable. (For an illustration, see animation: https://www.youtube.com/watch?v=8DikuwwPWsY). Importantly, for a dynamical system, variables are interdependent and cannot be analyzed separately (Sugihara et al. 2012). Such kind of state-dependent behavior cannot be studied by linear approaches (such as regression or structural equation modeling), because linear approaches are fundamentally based on correlation and assume that the systems are additive (Sugihara et al. 2012). Thus, from a methodological viewpoint, the study of nonlinear dynamical systems requires nonlinear methods that acknowledge state dependency, whereas linear methods should be applied for linear stochastic systems.

To study dynamical systems, nonlinear time series analytical methods have been developed over recent decades (e.g. Sugihara and May 1990; Anderson et al. 2008; Glaser et al. 2014; Ye et al. 2015a). These nonlinear statistical methods are rooted in state space reconstruction (SSR), i.e. lagged coordinate embedding of time series data (Takens 1981). The basic idea of SSR

is illustrated in the animation: http://deepeco.ucsd.edu/video-animations. These methods do not assume any set of equations governing the system but recover the dynamics from time series data, thus called empirical dynamic modeling. Essentially, dynamical systems can be described as the evolution of a set of states over time based on some rules governing the movement of states in a high dimensional state space (i.e. a manifold). Motion on the manifold can be projected onto a coordinate axis, forming a time series. More generally, any set of sequential observations of the system state (i.e. a function that maps the state onto the real number line) is a time series. For example, when we collect time series data, we actually design and apply an observation function. Conversely, time series (observations) can be plotted in a multidimensional state space to recover the dynamics, known as attractor reconstruction (Packard et al. 1980).

For example, if we know that the dynamics of zooplankton are affected by phytoplankton and fish, we can reconstruct the system by plotting time series of phytoplankton, zooplankton, and fish along the x, y, and z axis, respectively, in a state space, and view the evolution of the system over time. However, in practice, we may lack the phytoplankton and fish data needed to reconstruct the dynamics; or, in a more general situation, we may not even know all the critical variables for the system. To overcome these difficulties, Takens (1981) offered a solution by demonstrating that a shadow version of the attractor (motion vectors or phase space) governing the original process can be reconstructed from time series observations on a single variable in the process (for example, the time series of zooplankton abundance) using lagged coordinate embedding. To embed such a series of scalar measurements (with an equal sampling interval), vectors in the putative phase space are formed from time-delayed values of the scalar measurements, $\{x_t, x_{t-1\tau}, x_{t-2\tau} \ldots x_{t-(E-1)\tau}\}$, where E is the embedding dimension (i.e., the dimension or number of time-delayed coordinates required for the attractor reconstruction), and τ is the lag (see Sugihara and May 1990) for the choices of E and τ). Takens' theorem states that the shadow version of the dynamics reconstructed by such an embedding preserves the essential features of the true dynamics (so-called "topological invariance"). That is, if enough lags are taken, this form of reconstruction is generically a diffeomorphism and preserves essential mathematical properties of the original system. In other words, local neighborhoods (and their trajectories) in the reconstruction map to local neighborhoods (and their trajectories) of the original system. Thus, in our plankton example, even if the fish and phytoplankton abundances over time are not measured, we can still reconstruct a shadow that accounts for these missing variables by taking the E prior values from just the zooplankton time series as a coordinate in E-dimensional space. Based on the concept of attractor reconstruction, EDM can be used to study nonlinear dynamical systems.

As time series data accumulate, EDM is gaining significant attention. To a large extent, this is a result of the powerful free software package rEDM, which is written in the R language (Ye et al. 2016). However, for nonspecialists, there is often a steep learning curve toward the effective use of this package. Our objective here is not to explain the theory and algorithm of EDM, which requires a deep understanding of the theory of dynamical systems, but to guide EDM novices through several basic applications. Nevertheless, an introductory-level understanding of dynamical systems is required before using the methods; we recommend some textbooks (e.g. Nicolis and Prigogine 1989; Alligood et al. 1996). Here, we provide model examples for which the exact answers are known. We demonstrate the functions in the rEDM package to analyze the model time series data step by step, and then explain the output and statistics and provide the ecological interpretation of the results. All the example model data and R codes are included in the Electronic Supplementary Material (ESM), allowing readers to reproduce the results. We then briefly touch upon some technical issues concerning data requirements and processing. We conclude by pointing readers to useful references for more advanced applications of the EDM framework.

Applications of EDM

EDM bears a variety of utilities to investigating dynamical systems: (1) determining the complexity (dimensionality) of the system (Sugihara and May 1990; Hsieh et al. 2005), (2) distinguishing nonlinear dynamical systems from linear stochastic systems (Sugihara 1994) and quantifying the nonlinearity (i.e. state dependence) (Anderson et al. 2008; Sugihara et al. 2011), (3) determining causal variables (Sugihara et al. 2012), (4) forecasting (Sugihara and May 1990; Dixon et al. 1999; Ye et al. 2015a; Ye and Sugihara 2016), (5) tracking the strength and sign of interaction (Deyle et al. 2016b), and (6) exploring the scenario of external perturbation (Deyle et al. 2013). These methods and applications can be used to give a mechanistic understanding of dynamical systems and provide effective policy and management recommendations on ecosystem, climate, epidemiology, financial regulation, medical diagnosis, and much else. Below, we provide examples for some basic applications of EDM. In ESMs, we provide a step-by-step guideline for each analysis using the R language (ESM1).

Determining the complexity of system

The complexity of a system can be practically defined as the number of independent variables needed to reconstruct the attractor (i.e. dimensionality of the system). Based on Takens' Theorem, the dynamics of the system can be reconstructed from the time lags of a single time

series, e.g., $\{x_t, x_{t-1\tau}, x_{t-2\tau} \ldots x_{t-(E-1)\tau}\}$. For simplicity, throughout this manuscript we set the time lag $\tau = 1$ for demonstration. Here, E is the embedding dimension (note that the practical embedding dimension E is not necessarily equal to the true dimension of the system D). Moreover, E is not necessarily equal to the number of interacting components (e.g., the number of species or the number of coupled equations). Nevertheless, it is proved that $E < 2D + 1$ (Whitney 1936); that is, E has an upper bound. In most real-world cases, E is not known a priori, and needs to be estimated. Determining embedding dimension E is a fundamental first step in all EDM analyses.

The dimensionality of a dynamical system can be determined by simplex projection (Sugihara and May 1990; Hsieh et al. 2005). When using simplex projection, a time series is typically divided into two halves, where one half (X) is used as the library set for out-of-sample forecasting of the reserved other half, the prediction set (Y). Note that the prediction set is not used in the model construction, and thus the prediction is made out of sample. Simplex projection is a nonparametric analysis in state space. The forecast for a predictee $Y(t_k) = \{ Y(t_k), Y(t_k-1), \ldots, Y(t_k-E+1)\}$ is given by the projections of its neighbors in the state space in the library set, $\{X_{(1)}, X_{(2)}, \ldots, X_{(E+1)}\}$, where $||X_{(1)} - Y(t_k)|| = \min(||X - Y(t_k)||)$ for all $X \neq Y$, $X_{(2)}$ is the second-nearest neighbor, and so on. All $E + 1$ neighboring points from the library set form a minimal polygon (i.e., simplex) enclosing the predictee under embedding dimension E. The one-step forward prediction $\hat{Y}(t_k + 1)$ can then be determined by averaging the one-step forward projections of the neighbors $\{X_{(1)}(t_1 + 1), X_{(2)}(t_2 + 1), \ldots, X_{(E+1)}(t_{(E+1)} + 1)\}$. By carrying out simplex projection using different values of E, the optimal embedding dimension E can be determined according to the predictive skill. There are several ways to evaluate the predictive skill of simplex projection, such as the correlation coefficient (ρ) or the mean absolute error (MAE) between the observations and the forecast results (i.e., comparing $Y(t_k + 1)$ with $\hat{Y}(t_k + 1)$). Statistical issues concerning whether to use ρ or MAE with empirical data are discussed by Hsieh and Ohman (2006). Note that, in the case where the time series is rather short, leave-one-out cross-validation can be performed instead of dividing the time series into halves (Sugihara et al. 1996; Glaser et al. 2014).

Here, we demonstrate an example comparing two systems: linear stochastic red noise and a nonlinear logistic map (Fig. 2a, b; ESM2). By trial-and-error using different values of E for simplex projection, we determine that the best embedding dimension for red noise is $E = 7$ whereas that for the simple nonlinear logistic map is $E = 2$. In this example, the optimal E is selected based on the criterion that maximizes the predictive skill by evaluating the correlation coefficients (ρ) between the forecasts and observations (Fig. 2c, d). The results indicated that, although both time series show large fluctuations, the dimensionality (or com-plexity) of the logistic map is much smaller than that of the red noise.

Distinguishing nonlinear dynamical systems from linear stochastic systems and quantifying the nonlinearity

The ability to distinguishing nonlinear dynamical systems from linear stochastic systems is a critical concern, because "nonlinearity" is formally associated with the ideas of nonlinear amplification, multiple stable states, hysteresis and fold catastrophe (Scheffer et al. 2001; Hsieh et al. 2005). Moreover, if a system is nonlinear (i.e. driven mainly by low-dimensional, deterministic processes), then in principle it should be possible to construct a reasonable mechanistic model that captures this behavior with much better forecast skill (Sugihara 1994; Hsieh et al. 2008). In contrast, it is impossible to construct a mechanistic model for linear stochastic systems. One should also understand that it is impossible to distinguish high-dimensional nonlinear systems from linear stochastic systems given time series data (Sugihara 1994).

As mentioned above, nonlinearity is formally defined as the state dependency of a nonlinear dynamical system. In other words, the degree of state dependency reflects the nonlinearity of a dynamical system. State dependency (nonlinearity) can be quantified by S-map analysis (S-map stands for "sequential locally weighted global linear map" (Sugihara 1994)). Similar to simplex projection, S-map also provides forecasts in state space. However, instead of using only neighboring points surrounding the predictee, S-map makes forecasts using the whole library of points with certain weights (hence the name, locally weighted global linear map). In fact, S-map analysis is a locally weighted linear regression performed under the state space associated with a weighting function in the form of an exponential decay kernel, $w(d) = \exp(-\theta d/d_m)$. Here, d is the distance between the predictee and each library point, and d_m is the mean distance of all paired library points. The parameter θ controls the degree of state dependency. If $\theta = 0$, all library points have the same weight regardless of the local state of the predictee; mathematically, this model reduces to linear autoregressive model. In contrast, if $\theta > 0$, the forecast given by the S-map depends on the local state of the predictee, and thus produces locally different fittings. Therefore, by comparing the performance of equivalent linear ($\theta = 0$) and nonlinear ($\theta > 0$) S-map models, one can distinguish nonlinear dynamical systems from linear stochastic systems.

Moreover, state dependency (nonlinearity) can be examined using the improvement in forecasting skill of the nonlinear model over the linear model as $\Delta\rho = \max(\rho_\theta - \rho_{\theta=0})$: the maximum difference between the correlation ρ_θ at each θ to the correlation $\rho_{\theta=0}$ found for $\theta = 0$. If the $\Delta\rho$ is significantly different from the expectation of the null model, the system is deemed
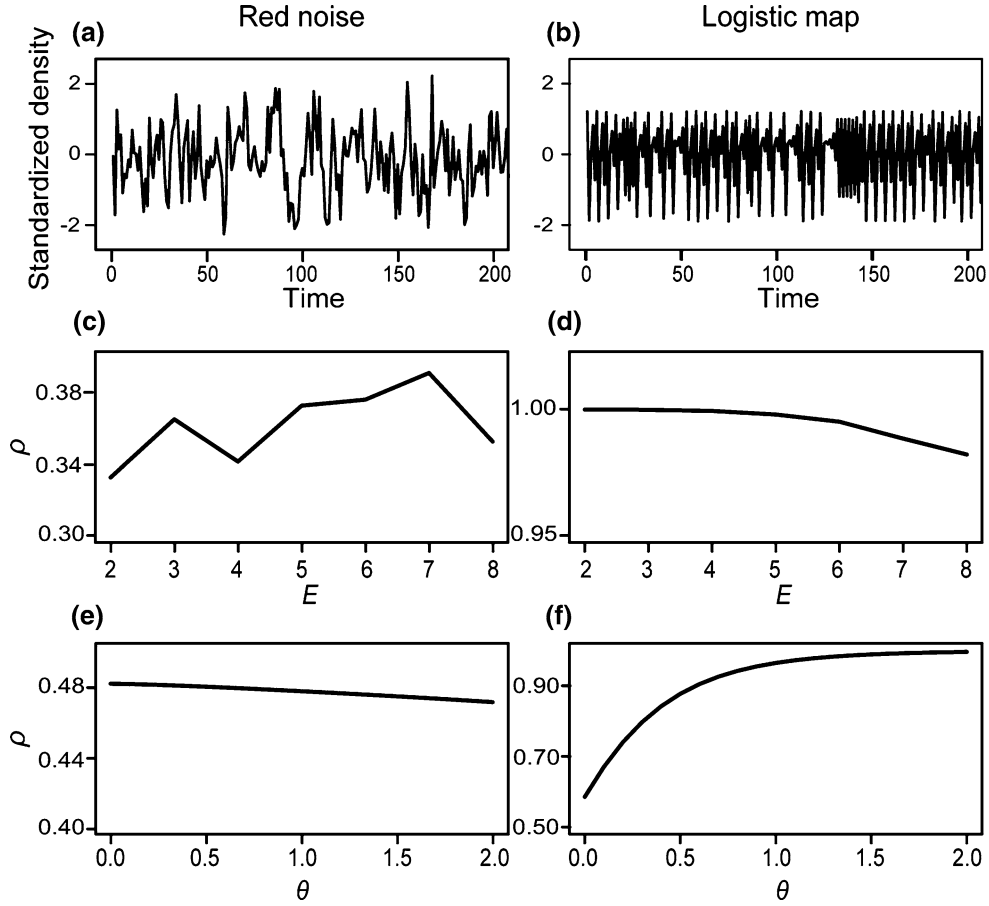
Fig. 2 Model examples comparing **a** linear stochastic red noise and **b** a nonlinear logistic map. By simplex projection, we determine the best embedding dimension (E) for **c** red noise and **d** the logistic map. In this study, we use the maximal predictive skill, ρ, as the criterion for selecting E. With the optimal E, we then use S-map to quantify the nonlinearity of red noise and the logistic map. For linear red noise (**e**), increasing the state dependency parameter θ does not improve the predictive skill (i.e. the optimal $\theta = 0$). In contrast, the predictive skill is maximized at $\theta = 2$ for the nonlinear logistic map (**f**). The improvement in forecasting skill of the nonlinear over the linear model, $\Delta\rho = \max(\rho_\theta - \rho_{\theta=0})$: the maximum difference between the forecasting skill ρ_θ at each θ to the skill $\rho_{\theta=0}$ found for $\theta = 0$, is used to quantify the nonlinearity (state dependency). If the $\Delta\rho$ is significantly different from that expected in the null model, the system is deemed nonlinear

nonlinear (see details in Hsieh et al. 2005; Deyle et al. 2013).

In a practical sense, we quantify state dependence by analyzing time series data of the system. It is necessary to emphasize that we do not derive or fit equations using data for the system, because such equations are generally unknown and fitting equations is unreliable for mathematical reasons (Perretti et al. 2013). Moreover, even when the equations are known or can be hypothesized, one cannot determine the nonlinearity of a system simply by asking whether the underlying equations are linear or nonlinear. In fact, nonlinear equations do not necessarily always exhibit nonlinear dynamic properties (e.g. chaos). Depending on the parameters, nonlinear equations can actually exhibit simple linear behaviors, such as equilibria and periodic cycles. Failure to make this distinction often causes confusion in the literature concerning the definition of nonlinearity.

As a demonstration, we analyze the aforementioned linear (red noise) and nonlinear (logistic map) systems using S-map (Fig. 2e, f; ESM2). Nonlinearity can be evaluated by examining the relationship between predictive skill ρ and the state-dependency parameter θ (Fig. 2e, f). The linear stochastic red noise does not exhibit any state dependency, as the S-map performance is optimized at $\theta = 0$. In contrast, the nonlinear logistic map reaches the optimal predictive skill at some $\theta > 0$, indicating the improved S-map forecast ability accompanied with increasing state-dependency (i.e. nonlinearity).

Determining causal variables

EDM can be used to reveal causation between variables. Two variables are causally linked if they interact in the same dynamical system. Following Takens' theorem, the system manifold reconstructed from univariate embedding (SSR using a single variable) gives a 1-1 map to the original system, i.e., topological invariance. Because

all manifolds reconstructed from univariates give 1-1 maps to the original manifold, it is not surprising that all the reconstructed manifolds result in 1-1 mappings if they are causally linked. Based on this idea, Sugihara et al. (2012) developed a cross-mapping algorithm to test the causation between a pair of variables in dynamical systems. This algorithm predicts the current quantity of one variable M_1 using the time lags of another variable M_2 and vice versa. If M_1 and M_2 belong to the same dynamical system (i.e., they are causally linked), the cross-mapping between them shall be "convergent." Convergence means that the cross-mapping skill (ρ) improves with increasing library size. This is because more data in the library makes the reconstructed manifold denser, and the highly resolved attractor improves the accuracy of prediction based on neighboring points (i.e., simplex projection). Sugihara et al. (2012) stated that convergence is a practical criterion to test causation, and called this phenomenon convergent cross-mapping (CCM). To evaluate convergence in cross-mapping, the state space is reconstructed using different library lengths (L) subsampled randomly from time series. Here, L_i starts from the minimal library length, L_0, which is equal to the embedding dimension, to the maximal library length, L_{max}, which equal to the whole length of the time series. To test the convergence of CCM, two approaches are widely used. First, the convergence can be tested by investigating how the cross-mapping skill changes with respect to the library size (e.g., trend or increment). For example, one can consider the following two statistical criteria: (1) testing the existence of a significant monotonic increasing trend in $\rho(L)$ using Kendall's τ test, and (2) testing the significance of the improvement in $\rho(L)$ by Fisher's $\Delta\rho$ Z test, which checks whether the cross-mapping skill obtained under the maximal library length ($\rho(L_{max})$) is significantly higher than that obtained using the minimal library length ($\rho(L_0)$). The convergence of CCM is deemed significant when both Kendall's τ test and Fisher's $\Delta\rho$ Z test are significant. Second, the convergence and the significance of cross-mapping skill can be tested by comparison with the null model expectation generated using surrogate time series (van Nes et al. 2015). However, there is no consensus on the optimal approach or null model.

Note that, the direction of cross-mapping is opposite to the direction of cause-effect. That is, a convergent cross-mapping from $\mathbf{M_2}(t)$ to $\mathbf{M_1}(t)$ indicates that M_1 causes M_2. This is because M_1, as a causal variable driving M_2, has left its footprints on $\mathbf{M_2}(t)$. The footprints of M_1 are transcribed on the past history of M_2, and thus M_2 is able to predict the current value of M_1.

We revisit the two model examples of the Moran effect and mirage correlation (Fig. 1), and compare the results of CCM and linear correlation analysis at identifying causation. In the Moran effect model (ESM3), the cross-mapping between the two variables does not converge at all, even though their linear correlation is
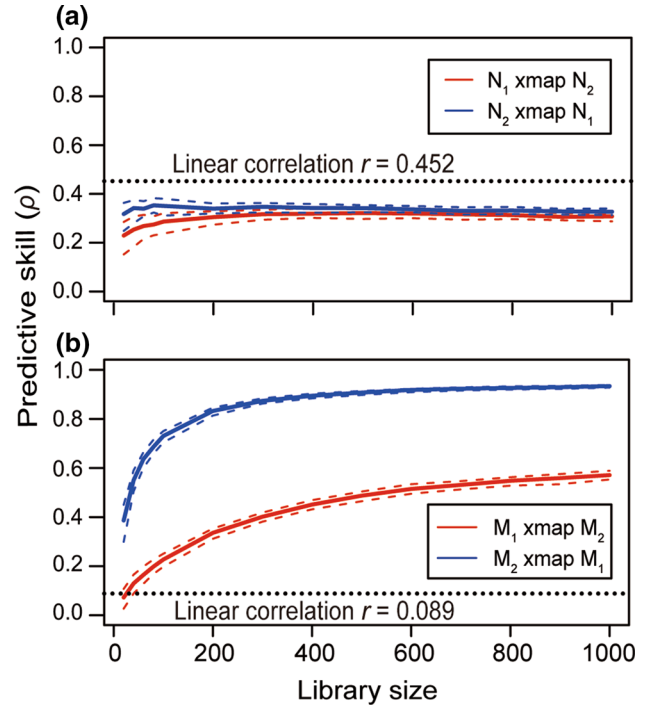

Fig. 3 Model examples demonstrating convergent cross mapping (CCM) to identify causality. In the Moran effect model (Fig. 1b), although the overall correlation is significant (*dotted line* $r = 0.452$), CCM does not exhibit any convergence with increasing library size, indicating no causation between the two populations (**a**). In contrast, although the overall correlation is very weak (*dotted line* $r = 0.089$) due to the mirage correlation (Fig. 1d), CCM exhibits strong convergence with increasing library size, indicating bidirectional causation between two competing populations (**b**). The *solid line* represents the median of predictive skill for each library size, and the *dash lines* represent the 1st and 3rd quantiles of the predictive skills from randomly subsampled library sets

significantly high (Fig. 3a). In contrast, the mirage correlation model (Fig. 3b) demonstrates clear convergence in CCM, although no significant correlation is found between the two populations. On the one hand, CCM avoids the wrong conclusion being drawn for the Moran effect (in contrast to the significant correlation concluded by the linear analysis) (Fig. 3a). On the other hand, CCM successfully detects the mutual causality in the competition model (ESM4) that is otherwise masked by the lack of significant correlation due to the mirage correlation (Fig. 3b) in nonlinear systems. A recent study indicates that CCM is generally robust even when the interaction coefficient is time-varying (BozorgMagham et al. 2015).

Forecasting: univariate, multivariate, and multiview embedding

Because vectors that are close in state space evolve similarly in time, the future value at one time point can be predicted based on the behavior of its nearest neighbors in the reconstructed state space. EDM uses
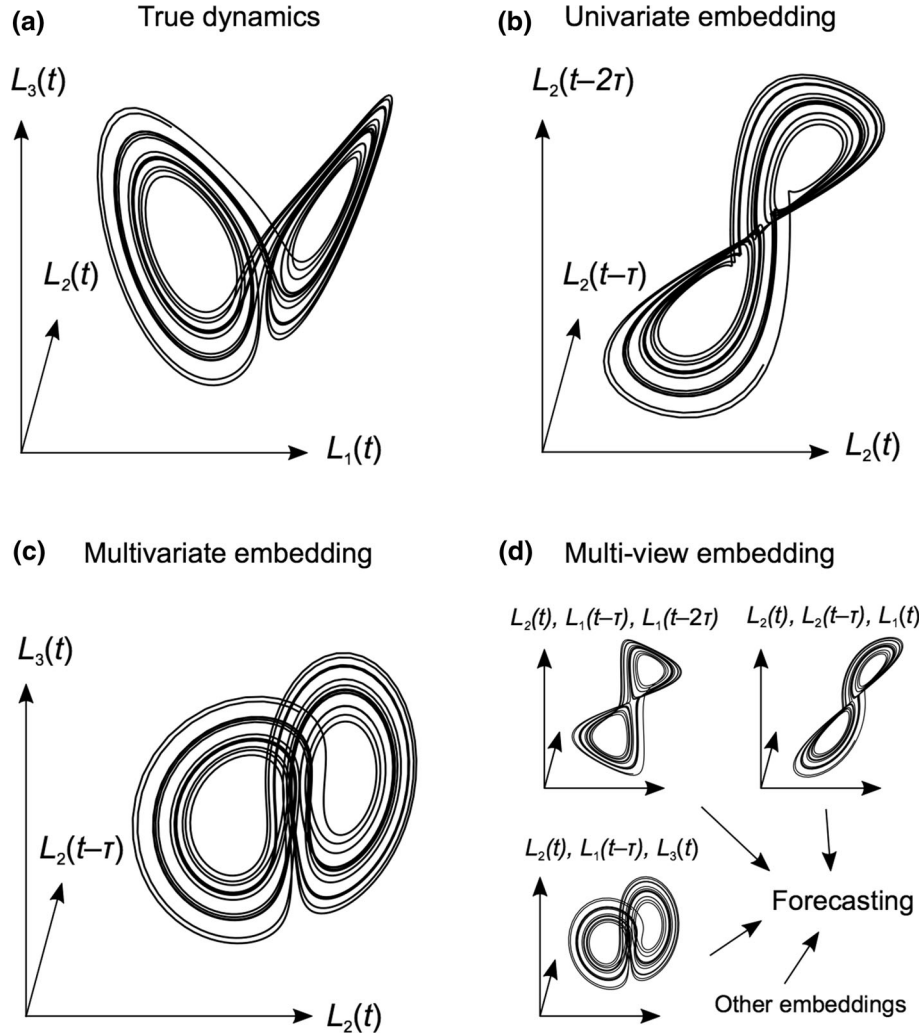
(a) True dynamics

$L_3(t)$

$L_2(t)$

$L_1(t)$

(b) Univariate embedding

$L_2(t−2\tau)$

$L_2(t−\tau)$

$L_2(t)$

(c) Multivariate embedding

$L_3(t)$

$L_2(t−\tau)$

$L_2(t)$

(d) Multi-view embedding

$L_2(t), L_1(t−\tau), L_1(t−2\tau)$ $L_2(t), L_2(t−\tau), L_1(t)$

$L_2(t), L_1(t−\tau), L_3(t)$

Forecasting

Other embeddings

Fig. 4 Conceptual explanation of univariate, multivariate, and multi-view embedding, using the Lorenz attractor ($L_1(t)$, $L_2(t)$, and $L_3(t)$). The original attractor (**a**) is approximated using state space reconstruction by univariate (**b**), multivariate (**c**), and multi-view embedding (**d**). Under a univariate embedding (**b**), only one variable (L_2 in this example) is used based on lagged time series ($L_2(t)$, $L_2(t − \tau)$, and $L_2(t − 2\tau)$). In the multivariate embedding (**c**), multiple variables (L_2 and L_3 in this example) are used based on a combination of more than one lagged time series ($L_2(t)$, $L_2(t − \tau)$, and $L_3(t)$). Moreover, the attractor can be reconstructed by multi-view embedding (**d**), that is, combinations of various multiple embeddings

the information on historical trajectories to forecast future values rather than specific equations that assume a mechanistic relationship between variables. Simplex projection (Sugihara and May 1990) and S-map (Sugihara 1994) (as explained in previous sections) enable forecasting for dynamical systems using information in the reconstructed state space.

As simplex projection and S-map are applied in a reconstructed state space, the method of reconstructing the state space is a critical issue for forecasting. In the framework of EDM, three different methods have been proposed so far (Fig. 4): (1) univariate embedding (Takens 1981; Sugihara and May 1990), (2) multivariate embedding (Dixon et al. 1999; Deyle and Sugihara 2011), and (3) multi-view embedding (Ye and Sugihara 2016). In this section, we demonstrate these three forecasting methods using time series generated from a re-

source–consumer–predator model (Fig. 5; ESM5). The details of this model are described by Deyle et al. (2016b).

The univariate embedding uses time-lagged values of a single variable to reconstruct the state space. Suppose we are interested in forecasting the population dynamics of Consumer 1 (C_1). We can use univariate embedding to reconstruct the state space using only information (history) encoded in C_1. The results of simplex projection indicate that the best embedding dimension of C_1 is $E = 3$, so the state space is reconstructed using {$C_1(t)$, $C_1(t − 1)$, $C_1(t − 2)$}. The forecasting skill (i.e., correlation coefficient between observed and predicted values) is 0.970 in this case (Fig. 6a).

Multivariate embedding uses multiple variables to reconstruct the state space. In the resource–consumer–predator model, Resource (R) and Predator 1 (P_1)
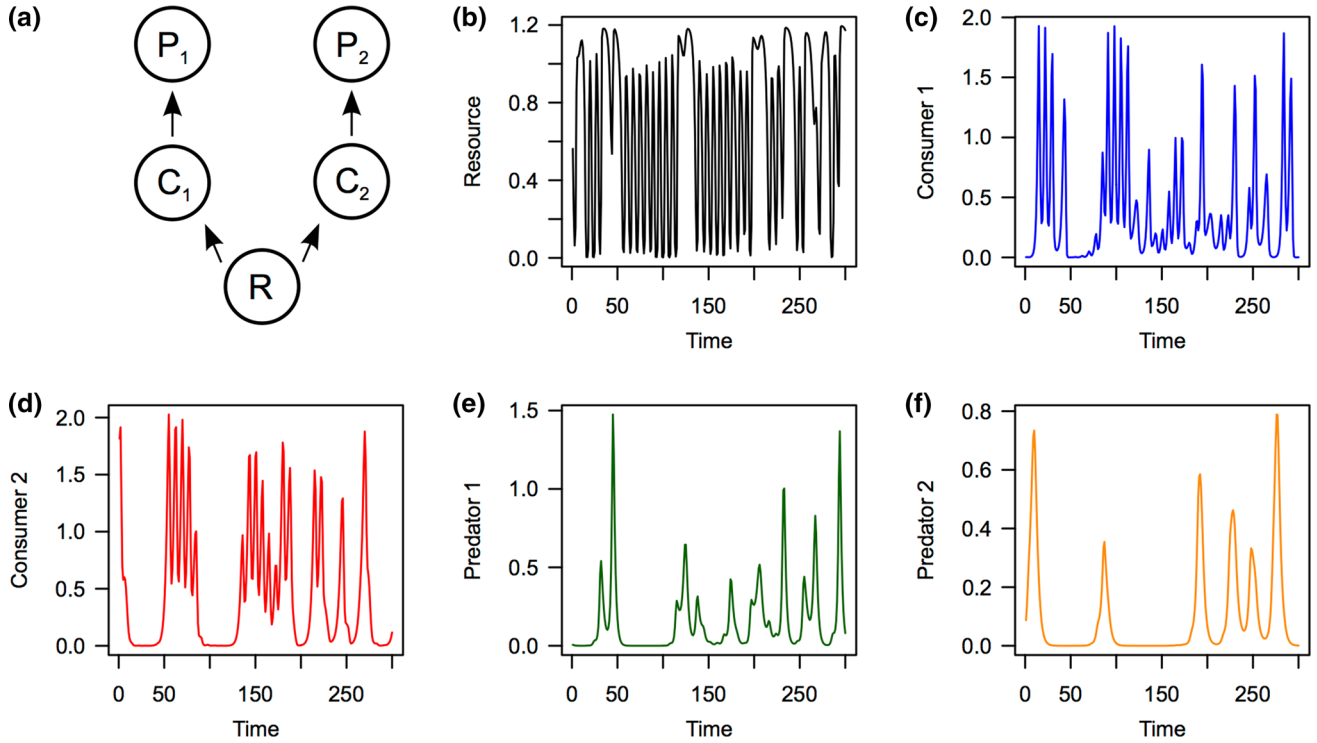
Fig. 5 Model example of five-species resource–consumer–predator interaction. Resource, Consumer 1, Consumer 2, Predator 1, and Predator 2 are represented by R, C_1, C_2, P_1, and P_2, respectively (**a**). In **a**, the *arrow* indicates energy flow. Example time series are shown for R (**b**), C_1 (**c**), C_2 (**d**), P_1 (**e**), and P_2 (**f**)

interact directly with C_1. Thus, information in R and P_1 is useful for forecasting the population dynamics of C_1. In this case, the state space is reconstructed using $\{R(t), P_1(t), C_1(t)\}$ (i.e., native multivariate embedding without using lagged values). The forecasting skill is 0.987 (Fig. 6b). Note that, in this case, $\{R(t), P_1(t), C_1(t)\}$ is sufficient to recover the dynamics of $C_1(t)$, because the best embedding dimension of $C_1(t)$ is $E = 3$. However, if the best embedding dimension of $C_1(t)$ were $E \geq 4$, additional time-lagged values (e.g., $C_1(t - 1)$) may be added to sufficiently recover the dynamics of $C_1(t)$.

Multi-view embedding leverages information by combining many possible embeddings (Ye and Sugihara 2016). According to embedding theory (Takens 1981; Deyle and Sugihara 2011), many valid embeddings are possible even if there are only a few variables in a system. Given l lags for each of n variables, the number of E-dimensional variable combination is $m = \binom{nl}{E} - \binom{n(l - 1)}{E}$. Although all variable combinations are valid embeddings, the system dynamics may not be resolved equally well with limited data. Therefore, only the top-k reconstructions, as ranked by in-sample forecasting skill, are used in the multi-view embedding, with the heuristic value of $k = \sqrt{m}$ applied in the original paper (Ye and Sugihara 2016). The values predicted from the top-k reconstructions are then averaged, and a single predicted value is calculated. For example, the Lorenz attractor contains three variables,

$L_1(t)$, $L_2(t)$, and $L_3(t)$. If we allow a time-lag of up to two steps, then the number of possible three-dimensional combinations (three-dimensional embeddings), m, is $\binom{3 \times 2}{3} - \binom{3 \times (2 - 1)}{3} = 19$ (Fig. 4d). The top-k reconstructions, that is $\sqrt{19}$ (in practice, the top-4 or -5 reconstructions), are used to make predictions. We applied the multi-view embedding to the resource–consumer–predator model, and forecast C_1. The forecasting skill of multi-view embedding is 0.989 for C_1 (Fig. 6c).

In general, the forecasting skill runs in the following order: multi-view embedding > multivariate embedding > univariate embedding (Fig. 6d). That is, given limited-length time series, richer information results in better forecasting skill. In some applications, one may wish to estimate the uncertainty of a forecast. One potential solution is to consider error propagation (Ye et al. 2015a). However, other approaches may apply, and this topic remains an open question.

Tracking strength and sign of interactions

Interspecific interactions are of particular interest among ecologists, because they are thought to drive the dynamics (e.g., local stability) of an ecological community (e.g. May 1972; Mougi and Kondoh 2012). The S-map method enables partial derivatives to be calculated in a multivariate state space at each time point, and
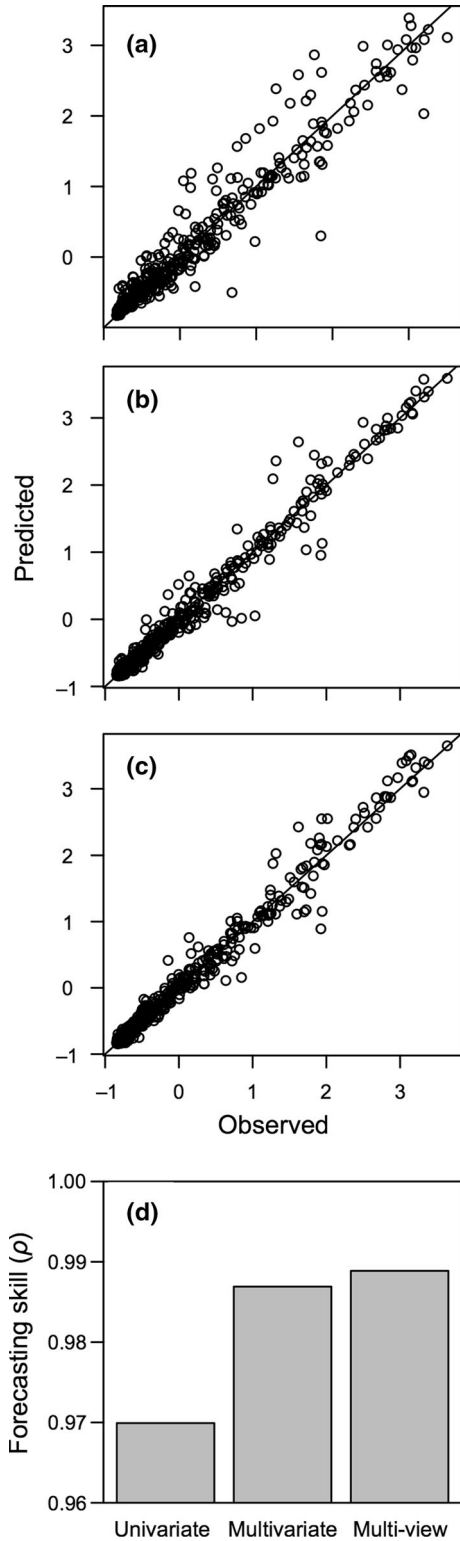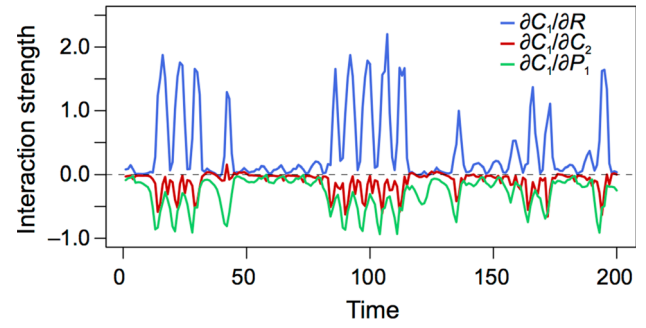
Fig. 7 Time-varying interspecific interaction strengths estimated by the S-map method as partial derivatives (i.e. S-map coefficients). The *blue*, *red*, and *green lines* represent $\partial C_1/\partial R$, $\partial C_1/\partial C_2$, and $\partial C_1/\partial P_1$, respectively. In this example, $\partial C_1/\partial R$, $\partial C_1/\partial C_2$, and $\partial C_1/\partial P_1$ can be regarded as bottom-up, competition, and top-down effects, respectively (color figure online)

2016b). For example, $\partial C_1/\partial R$ represents the influence of R on C_1 in the resource–consumer–predator model. Note that it is important to distinguish time-varying (realized) interaction strengths from interaction coefficients (which are often constant in differential or difference equations) (Hernandez 2009; Deyle et al. 2016b).

Using time series data from the resource–consumer–predator model (ESM5), we can calculate the interaction strengths from R, C_2, and P_1 to C_1. First, the state space is reconstructed using $C_1(t)$, $R(t)$, $C_2(t)$, and $P_1(t)$. Second, the best weighting parameter (θ) used in the S-map is determined by trial-and-error (see the previous section). Third, the partial derivatives at each time point are calculated using the multivariate S-map method. The partial derivatives $\partial C_1/\partial R$, $\partial C_1/\partial C_2$, and $\partial C_1/\partial P_1$ can be regarded as the bottom-up, competition, and top-down effect, respectively. The results indicate that the interaction strengths do indeed fluctuate in the model system, and that the bottom-up effects are larger than the competition and top-down effects (Fig. 7).

Scenario exploration of external perturbation

Ecological systems are often affected by an external force, e.g., temperature, and predicting what may happen in the system if the external force increases or decreases is a pressing concern. EDM facilitates the forecasting of the potential outcome of changes in an external force in the dynamics of a system (scenario exploration). For example, Deyle et al. (2013) used multivariate simplex projection to predict responses in the Pacific sardine population under a scenario of climate (temperature) changes.

In scenario exploration, we first need to determine which variables are included in SSR. In the demonstration, we again use the resource–consumer–predator model (Fig. 5; ESM5). In the model time series, we focus on predicting the influences of changes in R on C_1. First, we reconstruct the state space using univariate embedding. As the best embedding dimension of C_1 is 3, the

Fig. 6 Observed versus predicted values of univariate (**a**), multivariate (**b**), and multi-view (**c**) embedding. Forecasting skill (**d**) runs in the order: multi-view embedding > multivariate embedding > univariate embedding

the partial derivatives give a good approximation of interspecific interactions, capturing the time-varying dynamics of the interaction strengths (Deyle et al.
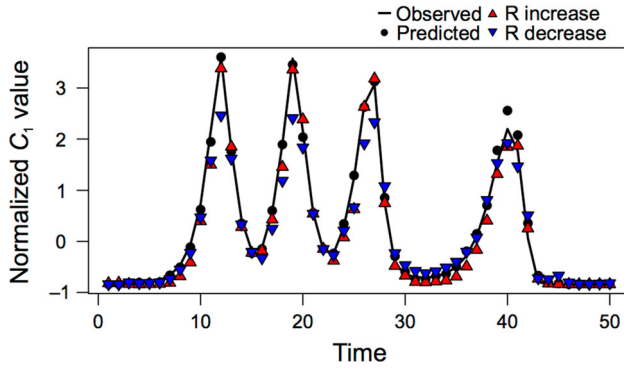
Fig. 8 Scenario exploration focusing on Consumer 1 (C_1) under increased/decreased Resource (R) situations. The *solid line*, *black filled-circles*, *red triangles*, and *blue triangles* represent the observed C_1, predicted C_1 with original data, predicted C_1 under increased R situation, and predicted C_1 under decreased R situation, respectively. Note that an increase in R does not always result in increased C_1. This is because the effect of R on C_1 also depends on the condition of other species, a phenomenon known as state-dependent behavior in dynamical systems (color figure online)

state space is reconstructed as $\{C_1(t), C_1(t - 1), C_1(t - 2)\}$. To predict the consequences of changes in R, we add $R(t)$ as an additional coordinate in the reconstructed state space. Thus, the final version of reconstructed state space is $\{C_1(t), C_1(t - 1), C_1(t - 2), R(t)\}$. In the state space, 50% of the standard deviation (σ) of R is added/subtracted to/from a target vector, and the future behavior of the modified vector ($\langle C_1(t), C_1(t - 1), C_1(t - 2), R(t) + 0.5\sigma \rangle$ or $\langle C_1(t), C_1(t - 1), C_1(t - 2), R(t) - 0.5\sigma \rangle$) is predicted by simplex projection. This scenario exploration suggests that changes in R result in changes in C_1 (Fig. 8). Note that the influence of R is not constant; that is, an increase in R results in increased C_1 at some time points, but decreased C_1 at other points (Fig. 8). Because the resource–consumer–predator model is a nonlinear dynamical system, the system behavior is state-dependent, and interactions between variables fluctuate over time. Thus, the effect of perturbation in R on C_1 changes depending on the state of the system. Again, this example demonstrates that EDM acknowledges state-dependence, and is therefore a powerful tool for analyzing and predicting nonlinear dynamical systems.

Data issues

One should bear in mind that EDM can only be applied to time series data of fixed, equal sampling intervals. Given a long, high-frequency time series, one can certainly bin the data into different time scales (i.e., dividing the sampling frequency). Nevertheless, analyses at different scales (e.g. daily or annual) reveal different dynamics, because the behavior of a dynamical system is scale dependent (Hsieh and Ohman 2006). In addition, the time series data needs to be stationary, as required by all time series analyses (Box et al. 1994).

As in any statistical method, errors can undermine the efficacy of EDM. Two types of error are often encountered: measurement (observational) and process error. Measurement error arises because of uncertainties in the measurements or observations; process error results from some processes that are not observed with the observation function (Sugihara 1994). For example, in an ecosystem, we cannot model all species; similarly, in a simple logistic growth model, the growth rate parameter is not fixed but randomly perturbed by environmental variation. The un-modeled part is considered the process error. An interesting phenomenon is that process error can drive a deterministic system from equilibrium to stochastic chaos (for mathematical account on this topic, see Sugihara 1994). For example, in a system of differential equations, even though the mean value(s) of the parameter(s) indicates that the system should reach equilibrium or stable cycle, the increasing variance(s) of the parameter(s) can drive the system to exhibit nonlinear behavior. This phenomenon is not well appreciated, but should actually be expected to appear very often in nature (Anderson et al. 2008; Sugihara et al. 2011) and warrants further study. EDM has been shown to be robust against moderate levels of measurement or process error (Hsieh et al. 2008; BozorgMagham et al. 2015); however, this is likely to be system-specific.

One critical concern is the number of data points needed for EDM to be applicable. Given the error in empirical systems, there is no theoretical justification for the minimal time series length. Generally, the required length of the time series increases with increasing complexity (embedding dimension) of the system. Sugihara et al. (2012) suggested that 35–40 data points are required for EDM. Nevertheless, data leveraging approaches have been developed to combine dynamically similar replicates (i.e., dynamic equivalence class) in cases where each individual time series is too short. For example, time series from different species that have the same dynamics can be concatenated to form a longer time series known as dewdrop regression (Hsieh et al. 2008). Spatial data from dynamic equivalence classes can be combined for analyses (Clark et al. 2015), and different combinations of time series data from interacting components can form a multivariate embedding, i.e., multi-view embedding (Ye and Sugihara 2016).

Another difficult issue is that many time series have missing data. Missing data (coded as NA in R) are automatically ignored in rEDM. Note that, as embedding is a necessary step in SSR, any vector (embedding) involved missing data is also omitted during computation. Therefore, missing data impart an unavoidably negative influence on the performance of EDM.

Data processing

Finally, we make a few suggestions on data processing prior to using EDM. First, the time series of variables

should always be normalized to zero mean and unit variance to ensure all variables have the same level of magnitude for comparison and to avoid constructing a distorted state space. Second, linear trends should be removed, either by simple regression or taking the first difference, to make the time series stationary. Third, unless there is strong mechanistic reason, we recommend that the time series data are not passed through a linear filter (e.g., smoothing or moving average), because smoothers can destroy the dynamics and make the signal linear. Finally, we caution that strong cyclic behavior or seasonality may mask the efficacy of EDM; data standardization methods (Ye et al. 2015a) or surrogate data tests to account for seasonality (Deyle et al. 2016a) have been developed to overcome these problems, although further methodological development is still underway.

Advanced applications

In addition to the examples given in this introductory paper, EDM has a wide variety of applications. For example, time-delayed causal interactions estimated from CCM may be used to infer direct versus indirect interaction (Ye et al. 2015b). Elevated nonlinearity, as quantified by S-map, is a useful early warning signal for anticipating critical transitions in dynamical systems (Dakos et al. 2017). EDM has been used to investigate scale-dependent system behavior (Hsieh and Ohman 2006; Jian et al. 2016). The prediction horizon (i.e. how quickly the predictive skill decays with time steps into future) of EDM can provide a guideline for fisheries management (Glaser et al. 2014). EDM has also been used to classify systems; this is because systems belonging to the same dynamic behavior can predict each other (Hsieh et al. 2008; Liu et al. 2012).

Final remark

This casual review is by no means comprehensive. To apply EDM, some basic knowledge of statistics and dynamical systems is essential to prevent the misuse of the software or the misinterpretation of results. EDM is a rapidly developing field and is a powerful tool for understanding nature. However, EDM tools can only be applied when sufficient time series data are available. Thus, we encourage long-term monitoring programs to be established and maintained, and recommend that time series data are shared.

References

Alligood KT, Sauer TD, Yorke JA (1996) Chaos: an intruduction to dynamical systems. Springer, New York

Anderson CNK, Hsieh CH, Sandin SA, Hewitt R, Hollowed A, Beddington J, May RM, Sugihara G (2008) Why fishing magnifies fluctuations in fish abundance. Nature 452:835–839

Box GEP, Jenkins GM, Reinsel GC (1994) Time series analysis: forecasting and control, 3rd edn. Prentice-Hall Inc., Englewood Cliffs

BozorgMagham AE, Motesharrei S, Penny SG, Kalnay E (2015) Causality analysis: identifying the leading element in a coupled dynamical system. PLoS ONE 10:e0131226. doi: 10.1371/journal.pone.0131226

Clark AT, Ye H, Isbell F, Deyle ER, Cowles J, Tilman GD, Sugihara G (2015) Spatial convergent cross mapping to detect causal relationships from short time series. Ecology 96:1174–1181

Dakos V, Glaser SM, Hsieh CH, Sugihara G (2017) Elevated nonlinearity as an indicator of shifts in the dynamics of populations under stress. J R Soc Interface 14:20160845

Deyle E, Sugihara G (2011) Generalized theorems for nonlinear state space reconstruction. PLoS ONE 6:e18295

Deyle ER, Fogarty M, Hsieh CH, Kaufman L, MacCall AD, Munch SB, Perretti CT, Ye H, Sugihara G (2013) Predicting climate effects on Pacific sardine. Proc Natl Acad Sci USA 110:6430–6435. doi:10.1073/pnas.1215506110

Deyle ER, Maher MC, Hernandez RD, Basu S, Sugihara G (2016a) Global environmental drivers of influenza. Proc Natl Acad Sci USA 113:13081–13086. doi:10.1073/pnas.1607747113

Deyle ER, May RM, Munch SB, Sugihara G (2016b) Tracking and forecasting ecosystem interactions in real time. Proc R Soc Lond B 283:20152258

Dixon PA, Milicich MJ, Sugihara G (1999) Episodic fluctuations in larval supply. Science 283:1528–1530

Glaser SM, Fogarty MJ, Liu H, Altman I, Hsieh C-H, Kaufman L, MacCall AD, Rosenberg AA, Ye H, Sugihara G (2014) Complex dynamics may limit prediction in marine fisheries. Fish Fish 15:616–633. doi:10.1111/faf.12037

Hernandez M-J (2009) Disentangling nature, strength and stability issues in the characterization of population interactions. J Theor Biol 261:107–119. doi:10.1016/j.jtbi.2009.07.001

Hsieh CH, Ohman MD (2006) Biological responses to environmental forcing: the linear tracking window hypothesis. Ecology 87:1932–1938

Hsieh CH, Glaser SM, Lucas AJ, Sugihara G (2005) Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. Nature 435:336–340

Hsieh CH, Anderson C, Sugihara G (2008) Extending nonlinear analysis to short ecological time series. Am Nat 171:71–80

Jian Y, Silvestri S, Brown J, Hickman R, Marani M (2016) The predictability of mosquito abundance from daily to monthly timescales. Ecol Appl 26:2609–2620. doi:10.1002/eap.1405

Liu H, Fogarty MJ, Glaser SM, Altman I, Hsieh C, Kaufman L, Rosenberg AA, Sugihara G (2012) Nonlinear dynamic features and co-predictability of the Georges Bank fish community. Mar Ecol Prog Ser 464:195–207

Lorenz EN (1963) Deterministic nonperiodic flow. J Atmos Sci 20:130–141. doi: 10.1175/1520-0469(1963)020 < 0130:dnf > 2.0.co;2

May RM (1972) Will a large complex system be stable? Nature 238:413–414

Moran PAP (1953) The statistical analysis of the Canadian lynx cycle. II. Synchronization and meteorology. Aust J Zool 1:291–298

Mougi A, Kondoh M (2012) Diversity of interaction types and ecological community stability. Science 337:349–351. doi:10.1126/science.1220529

Nicolis G, Prigogine I (1989) Exploring complexity: an introduction. W. H. Freeman and Company, New York

Packard NH, Crutchfield JP, Farmer JD, Shaw RS (1980) Geometry from a time series. Phys Rev Lett 45:712–716

Perretti CT, Munch SB, Sugihara G (2013) Model-free forecasting outperforms the correct mechanistic model for simulated and experimental data. Proc Natl Acad Sci USA 110:5253–5257. doi:10.1073/pnas.1216076110

Scheffer M, Carpenter S, Foley JA, Folkes C, Walker B (2001) Catastrophic shifts in ecosystems. Nature 413:591–596

Sugihara G (1994) Nonlinear forecasting for the classification of natural time series. Philos Trans R Soc A 348:477–495

Sugihara G, May RM (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in a data series. Nature 344:734–741

Sugihara G, Allan W, Sobel D, Allan KD (1996) Nonlinear control of heart rate variability in human infants. Proc Natl Acad Sci USA 93:2608–2613

Sugihara G, Beddington J, Hsieh CH, Deyle E, Fogarty M, Glaser SM, Hewitt R, Hollowed A, May RM, Munch SB, Perretti C, Rosenberg AA, Sandin S, Ye H (2011) Are exploited fish populations stable? Proc Natl Acad Sci USA 108:E1224–E1225. doi:10.1073/pnas.1112033108

Sugihara G, May R, Ye H, Hsieh CH, Deyle E, Fogarty M, Munch S (2012) Detecting causality in complex ecosystems. Science 338:496–500. doi:10.1126/science.1227079

Takens F (1981) Detecting strange attractors in turbulence. In: Rand DA, Young LS (eds) Dynamic systems and turbulence. Springer, New York, pp 366–381

van Nes EH, Scheffer M, Brovkin V, Lenton TM, Ye H, Deyle E, Sugihara G (2015) Causal feedbacks in climate change. Nat Clim Change 5:445–448. doi:10.1038/nclimate2568

Whitney H (1936) Differentiable manifolds. Ann Math 37:645–680

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

Ye H, Beamish RJ, Glaser SM, Grant SCH, Hsieh CH, Richards LJ, Schnute JT, Sugihara G (2015a) Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. Proc Natl Acad Sci USA 112:E1569–E1576. doi:10.1073/pnas.1417063112

Ye H, Deyle ER, Gilarranz LJ, Sugihara G (2015b) Distinguishing time-delayed causal interactions using convergent cross mapping. Sci Rep 5:14750. doi:10.1038/srep14750

Ye H, Clark A, Deyle E, Sugihara G (2016) rEDM: an R package for empirical dynamic modeling and convergent cross-mapping