Exploring the cause-effects of biodiversity and ecosystem functioning in natural lake ecosystems

Introduction

One of the most pressing concerns in current ecology and environmental science is accelerating global biodiversity loss in the biosphere (Sala *et al.* 2000). Biodiversity loss is a consequence of high global extinction rate which is two orders of magnitude higher than the background rate found in fossil record (Pimm *et al.* 1995). The loss of diversity potentially threatens the sustainability and ecological services of earth ecosystems (Hooper *et al.* 2012) because of the potential biodiversity effects on ecosystem functioning (BEF). In the past two decades, the importance of BEF was demonstrated by mounting studies which manipulated the biodiversity in experimental systems (Loreau *et al.* 2001; Hooper *et al.* 2005; Loreau 2010). Accumulating evidence of BEF has successfully invoked the attention of general public on biodiversity conservation (Rands *et al.* 2010). Although most of the BEF experiments have confirmed the effects of biodiversity on ecosystem functioning (Cardinale *et al.* 2006; Hooper *et al.* 2012), some concerns still wonder whether the BEF obtained from experimental systems can really reflect real scenarios of biodiversity loss in natural systems (Hooper *et al.* 2005; Naeem 2009). In fact, the causal relationship between biodiversity and ecosystem functioning has not been demonstrated in field because of the difficulties in methodologies. The existing methodologies using linear statistical analysis can only detect the correlation but not pinpoint the causation from field observations (Andow 1991). Critically, correlation does not imply causation (Sugihara *et al.* 2012).

To examine the causal relationship between biodiversity and ecosystem functioning in natural ecosystems, we employed an alternative approach, convergent cross map (CCM) (Sugihara *et al.* 2012), which can detect the causality from time series data based on empirical dynamic modeling, EDM. The dynamics theories proved by Takens states that two variables are causally linked if they belong to the same dynamic system (Takens 1981). Based on Takens' theory, CCM is developed as a novel algorithm to detect the causality between two variables using time series data. CCM outperforms conventional statistical analysis in two perspectives. First, CCM is able to distinguish the condition that two variables have no causal relationship but were affected by the third factor and thus showing correlations (i.e. Moran effects, Moran 1953). Second, CCM can detect the causality from a mirage correlation. Mirage correlation can be easily demonstrated by the following two-species difference equation:

$$X(t+1) = X(t)[r_x - r_x X(t) - \beta_{x,y} Y(t)]$$
$$Y(t+1) = Y(t)[r_y - r_y Y(t) - \beta_{y,x} X(t)]$$

In these equations, two species, *X* and *Y*, with chaotic dynamics (Fig. 1) may be positively coupled for a long period and then spontaneously become anticorrelated or uncorrelated for other periods (Mysterud *et al.* 2001). As a consequence, no overall correlation can be detected by linear statistical analysis (Fig. 1). Therefore, using statistical analysis based on correlation to infer causation is risky, especially when recognizing ubiquitous nonlinear behavior in natural systems. Comparing CCM and the other two approaches (stochastic statistical analysis and experimental manipulation) in examining the empirical causal relationship, we listed their advantages and disadvantages in Table 1. Although CCM cannot totally replace the importance of manipulated experiments, CCM might be a powerful tool to fill up the knowledge gap between the patterns in lab and the patterns in field. Through CCM, we aim to examine i) the causes-effect of biodiversity, ecosystem functioning, and abiotic environment in natural lake ecosystems (Table 2; ii) the cause-effect of biodiversity and ecosystem functioning across multiple trophic levels; iii) how BEF emergent from community inter-specific interactions, including inter-specific competition and trophic interaction.

Methods

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.

Convergent cross-mapping

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., topologically 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 M1 using the time lags of another variable M2 and vice versa. If M1 and M2 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 (q) 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, Li starts from the minimal library length, L0, which is equal to the embedding dimension, to the maximal library length, Lmax, 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 crossmapping 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 q(L) using Kendall’s s test, and (2) testing the significance of the improvement in q(L) by Fisher’s delta rho Z test, which checks whether the cross-mapping skill obtained under the maximal library length (q(Lmax)) is significantly higher than that obtained using the minimal library length (q(L0)). The convergence of CCM is deemed significant when both Kendall’s s 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 M2(t) to M1(t) indicates that M1 causes M2. This is because M1, as a causal variable driving M2, has left its footprints on M2(t). The footprints of M1 are transcribed on the past history of M2, and thus M2 is able to predict the current value of M1. 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, the cross-mapping between the two variables does not converge at all, even though their linear correlation is significantly high (Fig. 2a). In contrast, the mirage correlation model (Fig.2b) 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. 2a). 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. 2b) in nonlinear systems. A recent study indicates that CCM is generally robust even when the interaction coefficient is time-varying (BozorgMagham et al. 2015).

To test these hypotheses, specifically, we construct the networks of empirical causality links among biodiversity, ecosystem functioning and abiotic factors across different lake ecosystems by CCM. We quantify the interaction strength of causality link as the rate of convergence in CCM mappings (Sugihara et al. 2012). To examine how BEF response to environmental stresses, such as eutrophication, we compare the time series of biodiversity and ecosystem functioning before and after eutrophication.

Preliminary results

In our preliminary analysis (Fig. 4), we found three consistent patterns associated with biodiversity across three lake ecosystems which supports our hypotheses. First, the BEF links exists in all natural lake ecosystems. The strength of the causal link between biodiversity and ecosystem functioning including phytoplankton biomass (Chl*a*) and primary production (only in Lake Geneva) are as strong as that between abiotic factors and ecosystem functioning. This result suggests that biodiversity significantly affects ecosystem functioning in natural lakes, implying that the loss of biodiversity may cause substantial degradations in ecosystem services. The second pattern is that characteristic abiotic factors can simultaneously control biodiversity and ecosystem functioning. Such characteristic factors might be distinct for different types of lakes. In Lake Geneva, an oligotrophic lake, the characteristic factor is phosphate. That is, the limitation of phosphate in oligotrophic Lake Geneva might also be the key factor influencing species coexistence and productivity in phytoplankton community (Anneville *et al.* 2005). In Lake Kasumigaura, a eutrophic lake, the characteristic common drivers are total nitrogen and pH. Nitrogen has been demonstrated as the primary limiting factors in some eutrophic lakes, such as Lake Taihu (Xu *et al.* 2010), suggesting that the blooming event in eutrophic lakes is highly correlated with nitrogen cycling. In Lake Biwa, a mesotrophic lake, water temperature is the characteristic common factor, which is consistent with the previous study (Hsieh *et al.* 2011) that the plankton community of lake Biwa is under the stress of warming. These results implies that the relationship between biodiversity and productivity in empirical pattern might be easily confounded by other abiotic factors that may influence biodiversity and ecosystem functioning at the same time (MacDougall *et al.* 2013). Third, biodiversity could have indirect effect on ecosystem functioning through changing the cycling of essential nutrients including phosphate, total nitrogen and total phosphorous (i.e. the feedback). This result describes how the biodiversity effects on ecosystem functioning through altering other ecosystem components.

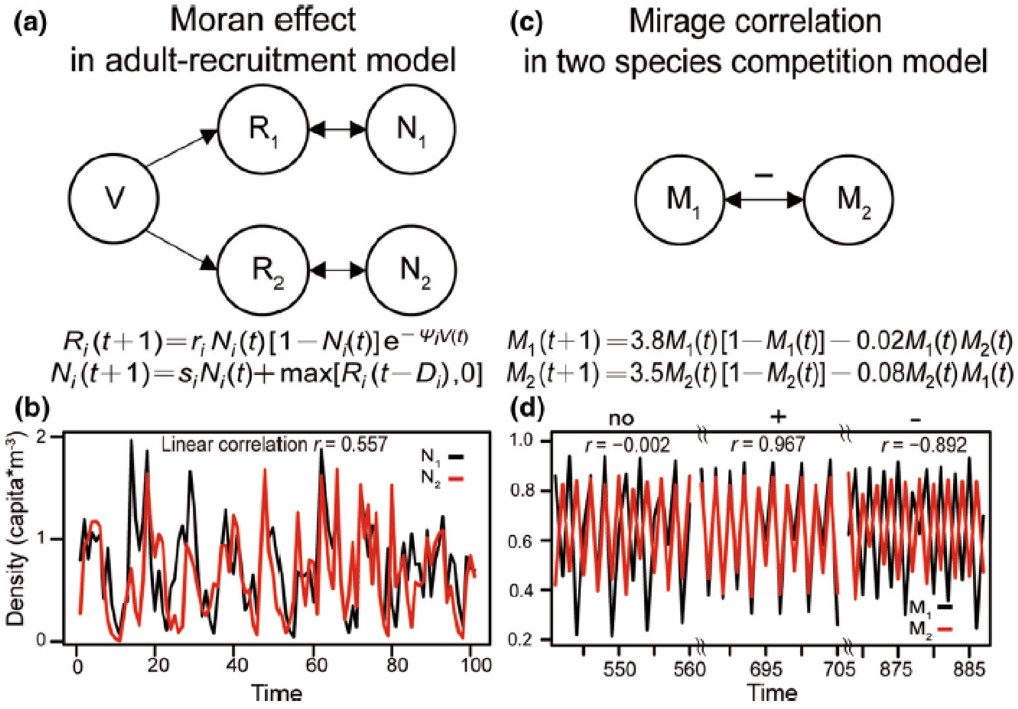

Fig. 1 Model examples demonstrating a confusing conclusion based on linear correlation analysis. The first model is a twospecies adult (N)-recruitment (R) fishery model (a). Both species are driven by a shared environmental driver (V). Although no interaction exists between species N1 and N2, their dynamics show a strong positive correlation (b), driven by the shared environmental force. The second model is a two-species competition model (M1 and M2), demonstrating mirage correlation (c). Although M1 and M2 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).

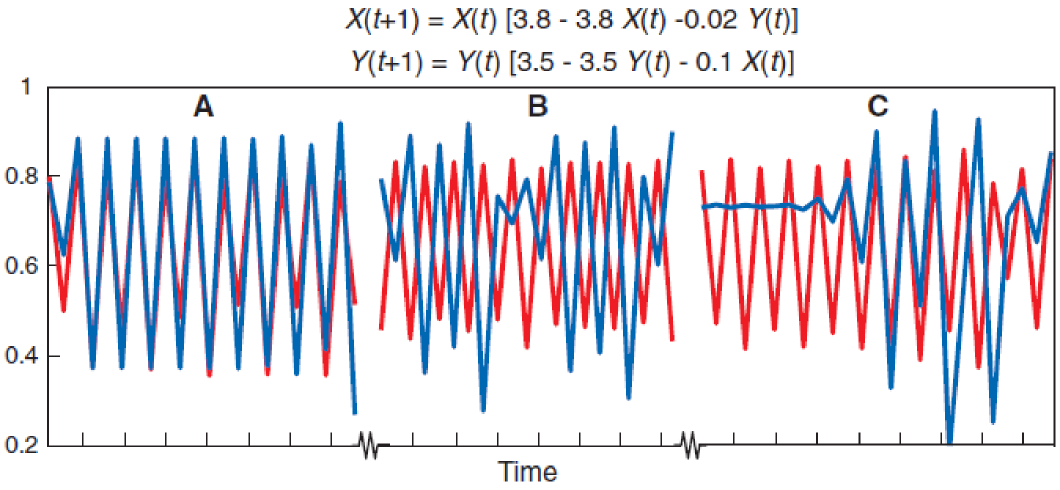

Fig. 2. Mirage correlations (A to C) adopting from Sugihara *et al.* (2012). Three samples from a single run of a coupled two-species nonlinear logistic difference system with chaotic dynamics. Variables X (blue) and Y (red) appear correlated in the first time segment (A), anticorrelated in the second time segment (B), and lose all coherence in the third time segment (C) with alternating interspersed periods of positive, negative, and zero correlation. Although the system is deterministic and dynamically coupled, there is no long-term correlation (n = 1000, r = 0.0054, P = 0.864).

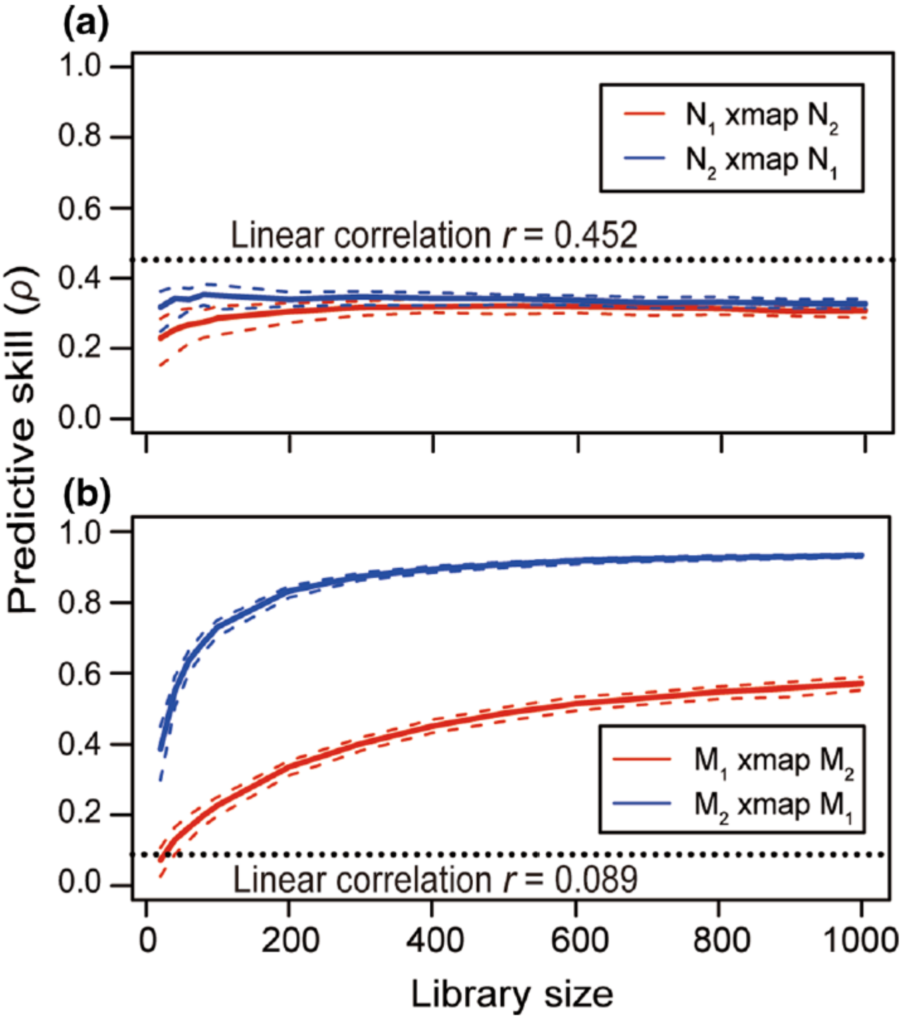

Fig. 3 Model examples demonstrating convergent cross mapping (CCM) to identify causality. In the Moran effect model (Fig. 1b), although the overall correlation is significant (dotted line r = 0.452), CCM does not exhibit any convergence with increasing library size, indicating no causation between the 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.

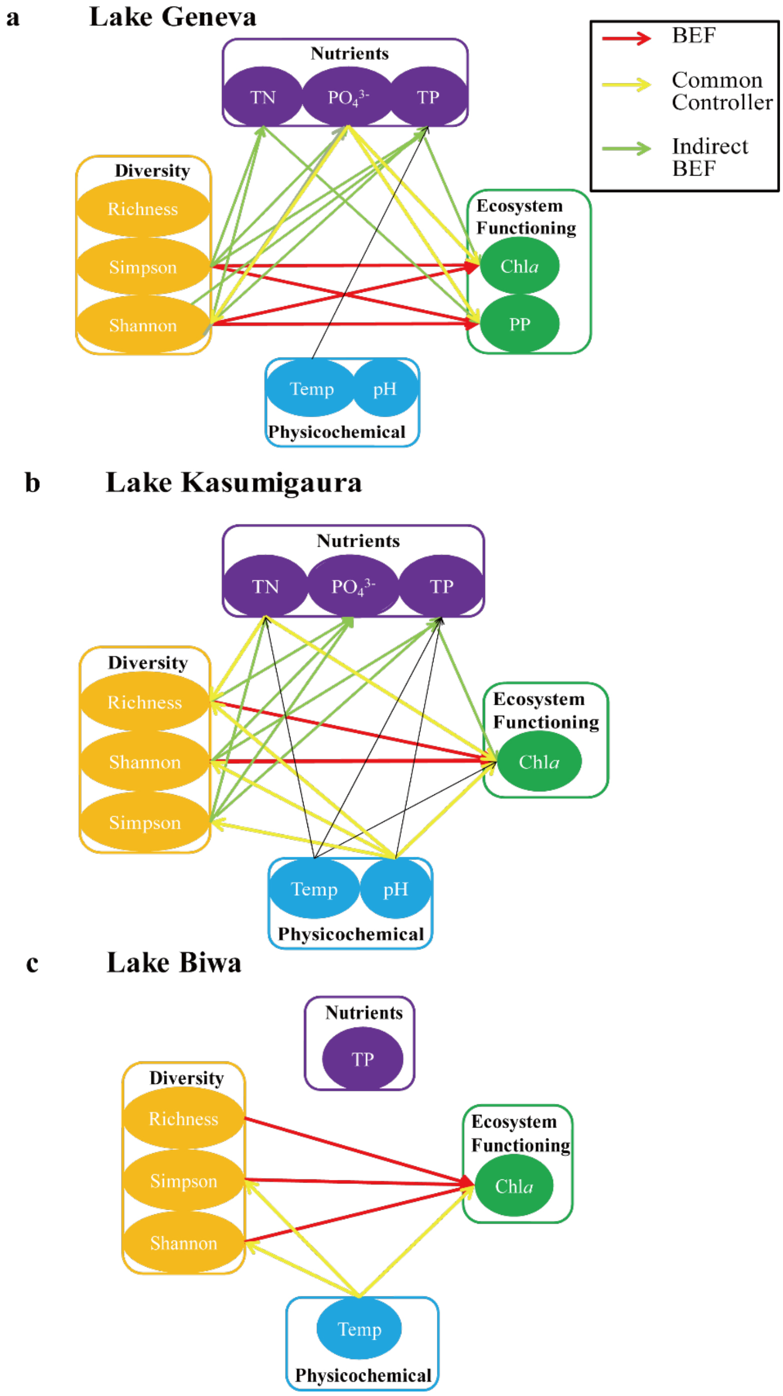

Fig. 4. The causality network among phytoplankton diversity (orange), ecosystem functioning (green), nutrients (purple) and physicochemical factors (blue) in a) Lake Geneva; b) Lake Kasumigaura; c) Lake Biwa. The red arrows are typical BEF links. The yellow arrow represent the effects of the common characteristic factors on biodiversity and ecosystem functioning. The green links represent the indirect BEF through the diversity effect on nutrient cycling.