PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Deyle ER, May RM, Munch SB, Sugihara G. 2016 Tracking and forecasting ecosystem interactions in real time. *Proc. R. Soc. B* **283**: 20152258. http://dx.doi.org/10.1098/rspb.2015.2258

Received: 5 November 2015
Accepted: 3 December 2015

Subject Areas:
ecology, theoretical biology, environmental science

Keywords:
nonlinear, empirical dynamics, changing interaction strength, S-map, state space reconstruction, community matrix

Author for correspondence:
George Sugihara
e-mail: gsugihara@ucsd.edu

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2015.2258 or via http://rspb.royalsocietypublishing.org.

THE ROYAL SOCIETY PUBLISHING

Tracking and forecasting ecosystem interactions in real time

Ethan R. Deyle[1], Robert M. May[2], Stephan B. Munch[3] and George Sugihara[1]

[1]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA
[2]Department of Zoology, University of Oxford, Oxford OX1 3PS, UK
[3]National Marine Fisheries Service, Southwest Fisheries Science Center, Santa Cruz, CA, USA

Evidence shows that species interactions are not constant but change as the ecosystem shifts to new states. Although controlled experiments and model investigations demonstrate how nonlinear interactions can arise in principle, empirical tools to track and predict them in nature are lacking. Here we present a practical method, using available time-series data, to measure and forecast changing interactions in real systems, and identify the underlying mechanisms. The method is illustrated with model data from a marine mesocosm experiment and limnologic field data from Sparkling Lake, WI, USA. From simple to complex, these examples demonstrate the feasibility of quantifying, predicting and understanding state-dependent, nonlinear interactions as they occur *in situ* and in real time—a requirement for managing resources in a nonlinear, non-equilibrium world.

1. Introduction

A particularly challenging aspect of ecological interactions is that they are not generally static. Rather, they are state-dependent (i.e. nonlinear) and change as ecosystem factors shift: e.g. fish populations show sensitivity to oceanographic conditions that increases when populations decline [1]; competition among small desert mammals varies with rainfall [2,3]; predation on insect herbivores changes with vegetation structure [4] and tadpole competitors suppress feeding in the presence of predators [5]. Although controlled experiments and model studies show how varying interactions can arise in principle from mechanistic state-dependence, empirical tools to track and predict them in the field are lacking. Here, we present a practical method that uses available time-series data to quantify, predict and understand changing ecosystem interactions as they occur in real time, as required for managing resources in a nonlinear non-equilibrium world.

Although much is known about nonlinear interactions in principle [6], heuristic understanding from models or controlled experiments may not accurately reflect what occurs in any specific natural setting. For example, consider two species that occupy the same trophic level. If their diets overlap, we might expect mutually negative competitive effects. However, if feeding responses are nonlinear, the strength or even the existence of competitive effects can depend on food limitation [7]. Moreover, if they share a common predator, the possible net outcome could be either positive or negative, depending on the details of their interactions [8] as well as the timescale of effects [9]. When predators exhibit prey switching, there can be even more complicated interactions [10]. Thus, in nature it is difficult to say which of the plausible expectations has arisen.

Indeed, a dearth of quantitative tools for investigating state-dependent interactions has limited our ability to measure them directly in the field. A few exceptional studies have tracked changes in food-webs through labour-intensive field methods (e.g. [11,12]). But while these gut content and other studies verify the complexity of interactions in principle, they are of little use for tracking or predicting 'continually changing species' interactions in nature. A more readily scalable alternative would be to estimate specific mutual interactions from time series of abundance. Currently available tools

like multivariate autoregression and its generalizations are specifically designed for systems with constant, fixed linear interactions and are sometimes 'fit' as dynamic linear models (DLM) to randomly drifting linear interactions [13,14]. But for nonlinear systems, such models are at best an *ad hoc* approximation without mechanism or the ability to predict.

Here, by clear contrast, we present an explicitly nonlinear approach based on empirical dynamic modelling (EDM as described below) [15–20] that uses readily available time-series data to measure and predict nonlinear interactions as they occur in the field. This approach does not require linear assumptions of equilibrium or constancy or of stochastic linear dynamics. We first introduce the logic of the method, and then demonstrate it by applying it to three systems: a model ecosystem where the interactions are exactly known; a mesocosm experiment from the Baltic Sea where the interactions are as expected; and data on zooplankton from Sparkling Lake, WI, USA.

2. The framework: empirical dynamic modelling

This method extends the classic community matrix idea—defined for systems in equilibrium—to dynamic systems that are not in equilibrium. The community matrix is commonly computed as the matrix of partial derivatives of the system evaluated at equilibrium (i.e. the Jacobian) or its *per capita* equivalent [21,22]. It is a linearization of the system—a theoretical expedient where pair-wise interactions are treated as fixed coefficients. However, real ecological systems are rarely, if ever, in static equilibrium. They typically exhibit nonlinear behaviour, where the strength and sign of interactions can vary with ecosystem state so that a single matrix of interactions will not suffice. Rather, to represent such systems it would be necessary to recalculate the interaction matrix anew for each successive ecosystem state. This may seem infeasible for a system being studied in the field, but—as we explain below—it is easily accomplished with S-maps [17], a standard EDM method for analysing nonlinear time series involving sequentially calculated Jacobians.

Briefly, EDM is an equation-free, mechanistic modelling approach based on the idea of reconstructing the underlying dynamical system from observed time series [15–20]. The process of building a manifold from time series is explained in a 1-min video animation (https://youtu.be/fevurdpiRYg). In EDM, the state of a dynamical system is a specific location in a multivariate coordinate space, or state space, whose coordinate axes are causally coupled ecosystem variables such as species abundance, temperature, resources, etc. The state of the system changes in time according to the rules/equations that describe the system dynamics, and this in turn traces out a trajectory. The collection of these time-series trajectories forms a geometric object called an attractor manifold, which describes empirically how variables relate to each other in time—hence EDM [15–20]. The basic idea of the S-map method is to recalculate an interaction matrix at each successive ecosystem state as the system travels along the attractor [17].

To illustrate, figure 1 presents a schematic for a hypothetical ecosystem consisting of two consumers, C_1 and C_2, competing for a shared resource, R. The empirical attractor manifold for this system is constructed from the time series (figure 1*a*) simply by taking the three time-series variables as Cartesian coordinates, $\underline{x}(t) = \{C_1(t),\ C_2(t),\ R(t)\}$ and plotting out the system trajectory (figure 1*b*). The dynamics of the first consumer, or how C_1 changes from one time point to another, is a function of the current ecosystem state $\underline{x}(t)$ and can be written as

$$C_1(t+1) = F(\underline{x}(t)).$$

Here F represents the three-dimensional dynamics on the attractor with respect to C_1. The two points on the hypothetical attractor, **p** and **q**, represent specific ecosystem states. Zooming into small neighbourhoods of these points, we see the interactions between C_1 and the other variables are nearly linear (figure 1*c*,*d*), and so F can be characterized by the appropriate row of the Jacobian matrix, which in discrete time is taken over the time interval t to $t+1$, and is simply

$$DF = \left\{\frac{\partial C_1(t+1)}{\partial C_1(t)}, \frac{\partial C_1(t+1)}{\partial C_2(t)}, \frac{\partial C_1(t+1)}{\partial R(t)}\right\}.$$

The Jacobian elements of this row define the interaction strengths or net local effect that each of the three variables C_1, C_2 and R has on the predicted variable C_1 [23]. Clearly, the interactions differ at **p** and **q**. The surface of F at **p** (figure 1*d*) has a steep positive slope in the R direction, indicating strong dependence on food abundance, and a steep negative slope along the C_2 direction, indicating strong competition. By contrast, at **q**, $C_1(t+1)$ is not sensitive to changes in R or C_2 and F is flat (figure 1*c*)—partial derivatives zero. Thus, the partial derivatives or Jacobian elements corresponding to these slopes define the interaction strengths for the system states **p** and **q**.

3. The method: measuring interactions with S-maps

The key to implementing these ideas is that the Jacobian elements (and thus the interaction strengths) can be recovered at any target point $\underline{x}(t)$ on the attractor using S-maps [16,17,19], where 'S' in Sugiharaś S-map denotes the 'sequentially' calculated Jacobians as the system moves along its attractor. Simply put, S-maps are a locally weighted multivariate linear regression scheme that approximates the best local linear model by giving greater weight to points on the attractor that are near the current ecosystem state. Because S-maps involve weighted linear regression, it is readily implemented in common statistical languages such as MATLAB and R. Example, marked-down R code is provided in the supplement, and the procedure is as follows.

Similar to other regression schemes, S-maps involves computing the linear model **C** that approximates the dynamics

$$\hat{x}_i(t^*+1) = C_0 + \sum_{j=1}^{E} C_{ij} x_j(t^*),$$

where E is the model order or embedding dimension (i.e. number of system variables)

However, in S-maps the linear approximation is done locally at each location $\underline{x}(t^*)$ on the attractor manifold. Stronger weight is given to vectors closer to the target point $\underline{x}(t^*)$ on the attractor manifold. Hence weighting is local to each state on the attractor. The exact weight given to observation k is given by

$$w_k = \exp\frac{-\theta \|\underline{x}(t_k) - \underline{x}(t^*)\|}{\bar{d}},$$

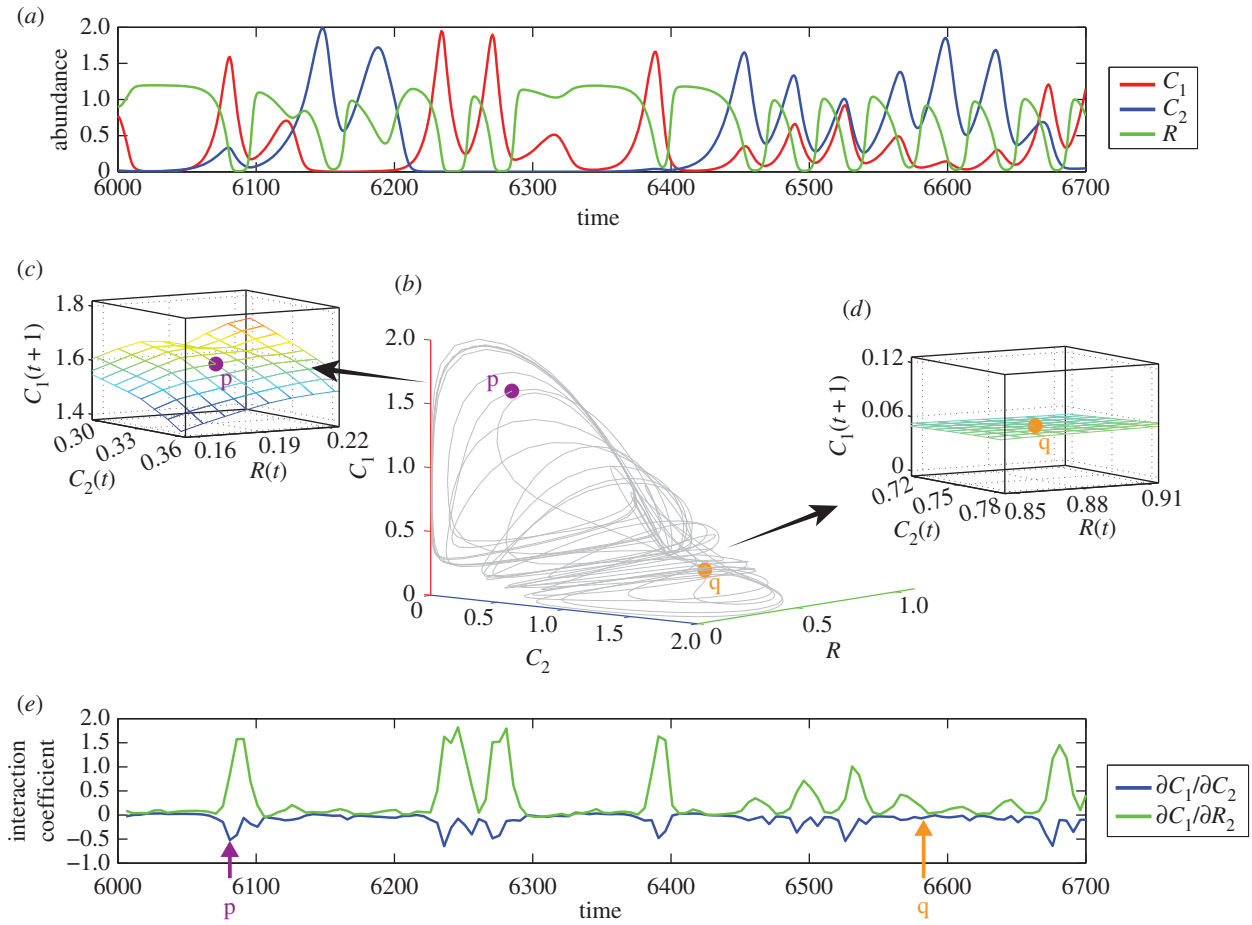

Figure 1. A schematic of measuring interactions in a hypothetical three-species ecosystem. The empirical attractor is constructed by re-plotting the time series of C_1, C_2 and R (*a*) simultaneously in three dimensions (*b*). The attractor displays the historical relationships between variables. The magnitudes of the interaction effects on C_1 are different at the two attractor states, **p** (purple) and **q** (orange). Panels (*c*) and (*d*) show the local effect of C_2 and R on C_1 at these states. The slopes of these local surfaces (i.e. the partial derivatives or Jacobian elements) define the interaction strengths calculated by the S-map coefficients (*e*). The surface at **p** is steep (*c*), thus the estimated interaction coefficients in (*e*) have large magnitude (purple arrow). Conversely, the surface at **q** is flat (*d*), so the interaction coefficients are near zero (orange arrow).

where $||\underline{x} - \underline{y}||$ denotes the Euclidian distance between two vectors and

$$\bar{d} = \frac{1}{n}\sum_{k=1}^{n} ||\underline{x}(t_k) - \underline{x}(t^*)||.$$

The parameter $\theta \geq 0$ tunes how strongly the regression is localized to the region of state space around each target. Note that if $\theta = 0$, the S-map model reduces to a VAR model. Thus, constant coefficient VAR models or closely related multivariate autoregressive (MAR) models arise as a special case of S-maps—the pathological case where location on the attractor manifold is irrelevant (no state dependence). For $\theta > 0$, the coefficients of **C** can vary with location on the attractor and with increasing θ they can vary more strongly as the system changes state. If θ is too small, the coefficients will underestimate the true variability in interaction strength. However, with larger θ the regression hinges on only the most proximal points on the manifold and will therefore be more sensitive to observation error. In practice, some intermediate value of θ will optimally balance bias and uncertainty (see discussion in the electronic supplementary material on choosing θ).

With this local weighting scheme, the S-map model is simply the SVD (singular value decomposition) solution for **C** to the linear equation

$$\mathbf{B} = \mathbf{A} \cdot \mathbf{C},$$

where **A** is the $n \times E$ dimensional matrix of weighted state-space vectors given by

$$A_{kj} = w_k x_j(t_k)$$

and **B** is the n-element vector of the predicted variable, i.e. the future values of the target variable x_i given by

$$B_k = w_k x_i(t_k + 1).$$

For linear regression, SVD is equivalent to a least-squares fitting that minimizes the Euclidean distance between the true and estimated attractor points. In some cases, other distance metrics may be more appropriate. For example, for small population sizes, an error model that is appropriate for count data near zero would be appropriate (e.g. [13]). Finally, it is common practice with this sequential regression procedure to employ leave-one-out cross-validation to avoid in-sample fitting.

S-maps have been used both as a simple test for nonlinear dynamics [17] and as a non-parametric tool for ecosystem forecasting [19]. Here we note simply that, in multivariate embeddings (i.e. native embeddings using causal variables [19] rather than lags of a single variable), the S-map coefficients

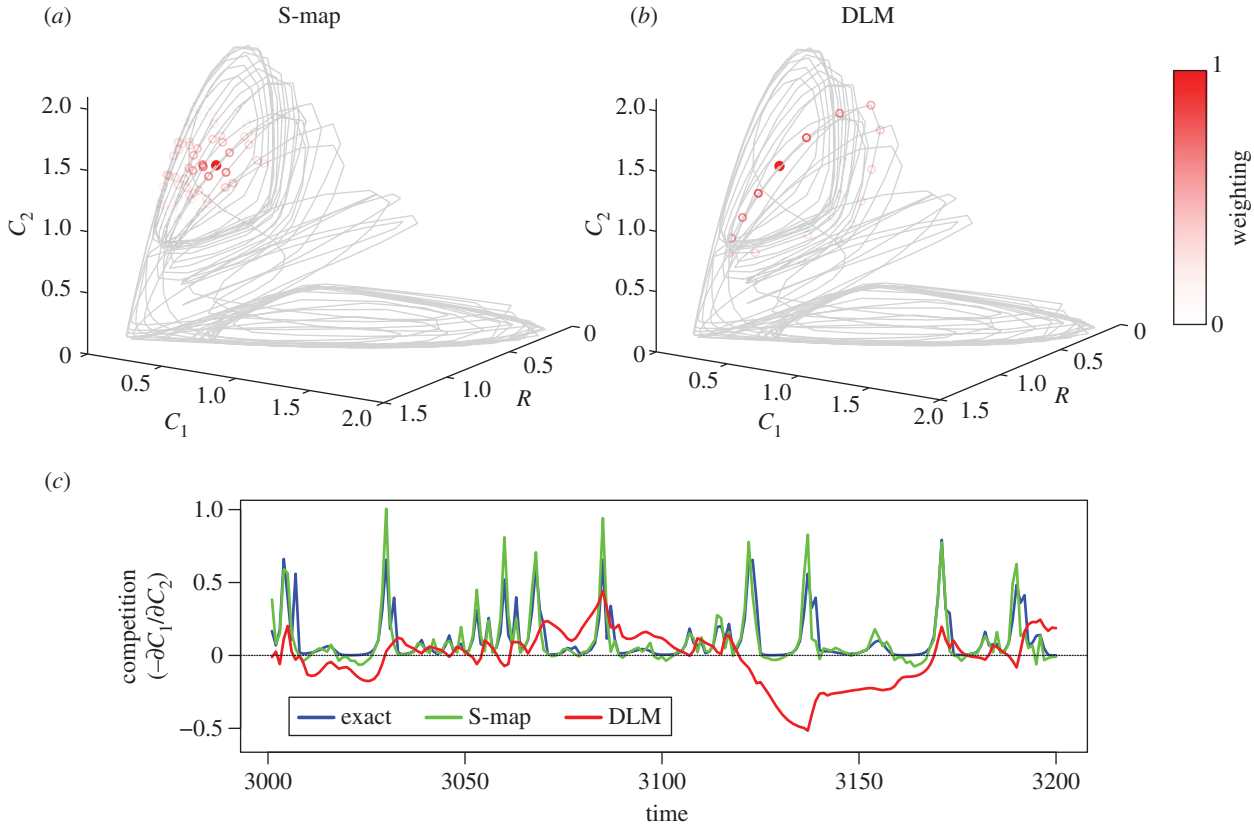

Figure 2. S-map versus dynamic linear model (DLM). S-map and DLM are both applied to measuring competition in the five-species food chain model described in the *test cases* below and supplement (target point = solid red circle). The DLM model is a vector autoregressive (VAR) model where the linear interaction coefficients are allowed to drift in time as a random walk. In (*a*), as with S-maps, weighting is determined by closeness in state space—position on the attractor—and is predictive, whereas in (*b*) as with DLM, it is determined by distance in time and it is 'fit' *post hoc* using forward information as a linear stochastic process. Note in (*b*) that weighting is given to several points where the ecosystem was in a substantially different state (e.g. much lower food abundance) than the target point (solid red). (*c*) Compares estimates of competition based on S-map forecasts, DLM *post hoc* fits and exact calculations from the model.

approximate the Jacobian or interaction elements at successive points along the attractor. That is, S-maps generate the relevant Jacobian elements that define the interaction strengths, and as required do so sequentially (S = 'Sequential Jacobian') as the system travels along its attractor. Moreover, in real time, when the time-series data for C_1 at time $t + 1$ are not available for fitting, the interaction strengths computed at that instant are actually *forecasts* of the influence of each variable on C_1.

At this point, it is worthwhile to compare S-maps to other methods for reconstructing interaction strengths from time series. Ives *et al.* [14] provide a clearly written justification for VAR/MAR models to approximate systems in the vicinity of an equilibrium, where it is appropriate to treat interactions with a *constant* matrix. Relaxing the equilibrium assumption, Lamon *et al.* [24] apply DLM to ecosystem interactions, where the coefficients are allowed to drift stochastically through time. These models have been used to indicate impending regime shifts [25] and in some nearly linear cases they can be used *retrospectively* to estimate past changes in system dynamics [26]. DLM is a linear method where the system matrix is modelled as a random walk. These models and their state-space extensions are nicely described by Holmes *et al.* [27] and implemented in their MARSS package.

Because the interaction strengths are allowed to change, the DLM may appear to be an analogue to the S-map. However, there is a major difference between the two: DLM methods are intended for linear stochastic systems and do not explicitly address state dependence; as such they are not mechanistically predictive of changes driven by nonlinear dynamics. Specifically, DLM determine coefficients by weighting ecosystem states that are nearby in 'time,' rather than ecosystem states that are actually most similar (closest in the state space, figure 2*a*,*b*). If the system changes slowly relative to the sampling rate, to produce a nearly linear case, DLM will give results similar to S-maps. However, states can change quite rapidly in ecosystems (e.g. outbreaks of spruce budworms or fishery collapses), meaning states nearby in time may be very dissimilar and have different interaction strengths. Under these conditions, the DLM approach will fail to measure interactions correctly (figure 2*c*). DLM is a linear method and produces *linear* forecasts with uncertainty bounds that grow very rapidly in time—they are not intended for nonlinear systems.

By contrast, because S-maps are specifically designed for nonlinear systems, they are able to portray the mechanistic conditions (system state) governing system dynamics. This is a key point that differentiates the EDM approach from non-mechanistic DLM methods that treat time variation phenomenologically without providing a mechanistic basis for understanding *why* interactions are changing, and typically require forward information ($\underline{x}(t + 1)$) to fit the drifting coefficients. Importantly, S-maps have been shown to be robust to observational noise [28] (see the electronic supplementary material) with process noise handled by the regression procedure [17].

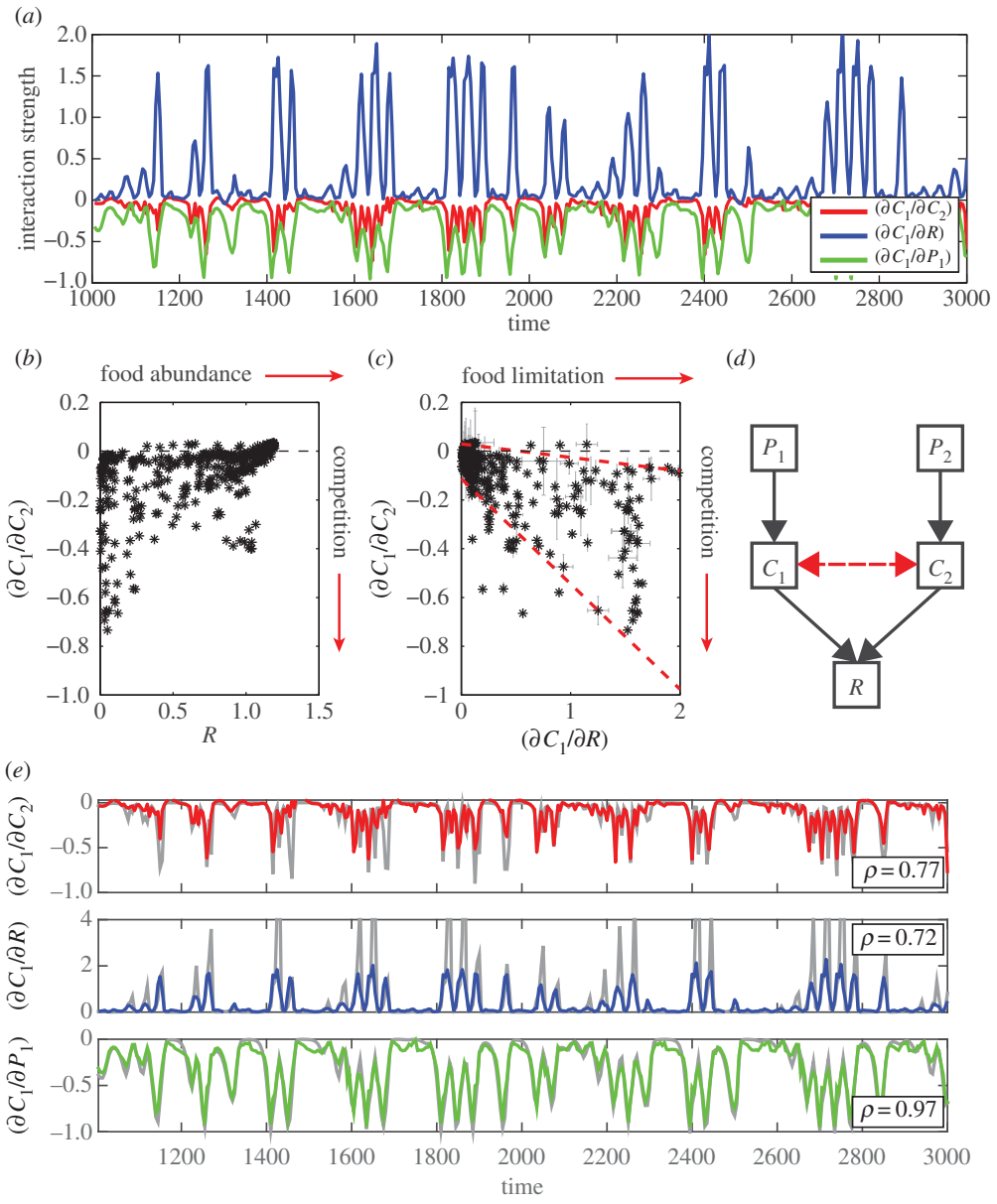

Figure 3. Dynamic interactions measured from the output of a five-species model food web (*d*). (*a*) The S-map estimated interaction coefficients over 1000 model years for the effects on consumer 1 (C_1) of the other variables. (*b*) and (*c*) The dependence of competition ($-\partial C_1/\partial C_2$) on food abundance ($R$) and food limitation ($\partial C_1/\partial R$), respectively. Red dashed line indicates 0.05 quantile regression (slope is significantly different from 0 ($p < 0.01$)). Grey error bars indicate 95% confidence limits on S-map coefficients. (*e*) Comparison of the estimated S-map forecasts of interaction strength with those computed directly from the model equations (correlation coefficients given in figure). Axes are in normalized units.

4. Test cases

To demonstrate the concept and the method we will apply it to three test cases: a model, an experimental mesocosm and a natural lake ecosystem. The model (figure 3*d*) is a classic food web [29,30] consisting of two consumers (C_1, C_2), their predators (P_1, P_2) and a single resource (R). The trophic interactions are governed by saturating Holling Type II feeding responses, and this gives rise to state-dependent competition [7,8]. The model is reckoned to be a transparent example of state-dependent interactions.

Figure 3 shows how the EDM approach using S-maps uncovers the mechanisms that cause the interaction strength to vary between the consumer C_1 and the other ecosystem components. For example, in figure 3*b* competition between C_1 and C_2 only occurs at low to moderate food levels, whereas at high food concentrations competition tends to zero. Moreover, figure 3*c* shows that $\partial C_1/\partial R$ (a direct measure of food limitation) sets a maximum on the strength of competition. This is explicitly demonstrated by the 0.05 quantile regression [31] of $\partial C_1/\partial C_2$ on $\partial C_1/\partial R$ (dashed red line). The effect is consistent with the underlying structure of the model, thus validating the approach. Finally, figure 3*e* shows that the interaction strengths forecasted by agrees well with the Jacobian elements computed directly from the system equations. Importantly, these predicted interaction coefficients are robust to realistic amounts of observational noise (see the electronic supplementary material S1 and figure S8).

Next, we apply the method to the interaction between calanoid copepods and rotifers in a freely evolving marine mesocosm isolated from the Baltic sea [32,33]. We focus on calanoids, rotifers and their two main prey items, nanoflagellates and picocyanobacteria (figure 4*d*). Figure 4*a* shows the S-map estimated interactions on calanoids, for the effects of rotifers, nanoflagellates and picocyanobacteria. As expected, interactions with the chief prey item (nanoflagellates) are always positive,

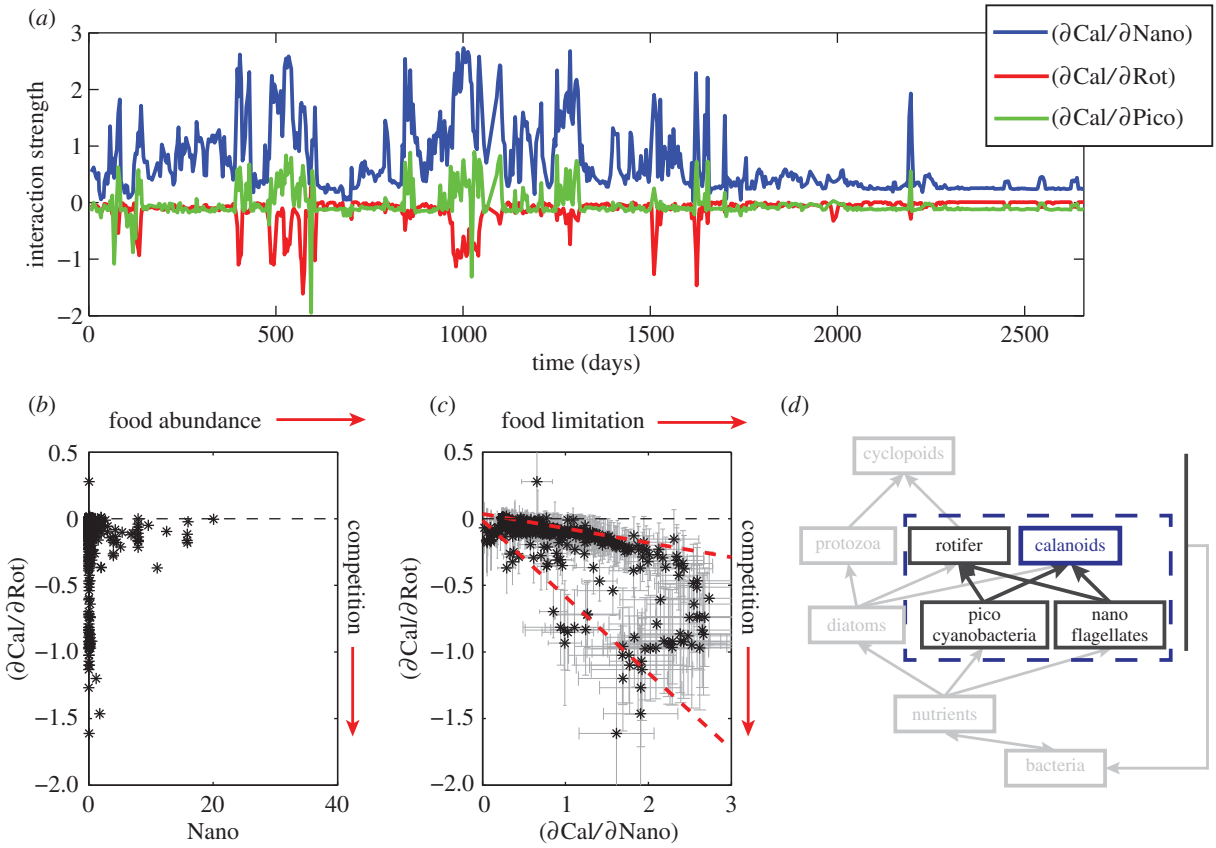

Figure 4. Dynamic interactions in the Baltic Sea mesocosm. (*a*) The S-map estimated interaction coefficients for calanoid copepods (Cal) with respect to the main prey item, nanoflagellates (Nano); secondary prey item, picocyanobacteria (Pico); and main competitor, rotifers (Rot) through the duration of the experiment. The effect of rotifers ∂Cal/∂Rot is shown as a function of (*b*) food abundance (Nano) and (*c*) food limitation (∂Cal/∂Nano), with the grey error bars indicating 95% confidence limits on S-map coefficients. The 0.05 quantile regression (red dashed line) has a significant slope ($p < 0.01$), and demonstrates the state-dependent nature of competition (indicated by −∂Cal/∂Rot) between rotifers and calanoids. (*d*) The focal species within a summary interaction network for the mesocosm. Axes are in normalized units.

and interactions with the other grazer (rotifers) are always negative. However, the intensity of these interactions changes through time. Competition (−∂Cal/∂Rot) is strong only when the prey (nanoflagellate) concentration is near-zero (figure 4*b*) with the maximum strength of competition set by food limitation (figure 4*c*), demonstrated by the 0.05 quantile regression line (dashed red). Analogous results have been obtained from experimental evidence of saturating feeding responses [6], however, here they are recovered noninvasively and directly in the freely evolving mesocosm, by analysis of the abundance time series. Although there is no direct way to validate the specific estimates of interaction strength (as we could for the model), our results align with ecological expectations—that competition depends on food limitation—and this validates the approach. Moreover, the convenience of the approach suggests its utility in cases where experimental manipulations are logistically infeasible, such as in large marine ecosystems.

As a final example, we consider the ecological interactions in Sparkling Lake, WI, USA, focusing again on copepod grazers (calanoids and cyclopoids). Figure 5*a* shows the S-map estimates for calanoids of the time-varying effects of cyclopoids, temperature and planktivorous fish. Note that the effect of cyclopoids on calanoids, ∂Cal/∂Cyc, is only negative in certain periods, indicating that there is only intermittent competition. Much of the time, the interaction is positive. In theory, a positive interaction can arise from apparent mutualism between trophically similar species who share common predators [8]. If so, competition should occur only when predation pressure is low. Thus, plotting ∂Cal/∂Cyc against total predator biomass, Fish (figure 5*b*), we see that competition, indeed, only occurs in periods with low planktivorous fish abundance. Similarly, plotting ∂Cal/∂Cyc against the predator : prey ratio (figure 5*c*) shows that the positive effect occurs most strongly at the highest ratios, whereas competition occurs only at the lowest predator : prey ratios. Here, the predator : prey ratio is Fish/(1 + Cal + Cyc), where 1 is added to accommodate times when both Cal and Cyc are measured as 0.

This evidence is consistent with predator-mediated mutualism and resonates with previous work showing that Sparkling Lake has been dominated by top-down forcing [34]. However, consistency is not proof. Indeed, because food supply will have a positive effect on both calanoids and cyclopoids, increases in one species could correlate with increases in the other for this reason alone. Unfortunately, this effect could not be examined in the Sparkling Lake study because there was no effective measure of food supply to drive the analysis (in terms of adequate temporal and/or taxonomic resolution, as discussed in the methods). Thus, although not conclusive, the weight of evidence in figure 5*b*,*c* points to apparent mutualism (or at least commensalism) mediated by common predators.

5. Concluding remarks

These three demonstrations illustrate how S-maps can be used to quantify changing species interactions and to

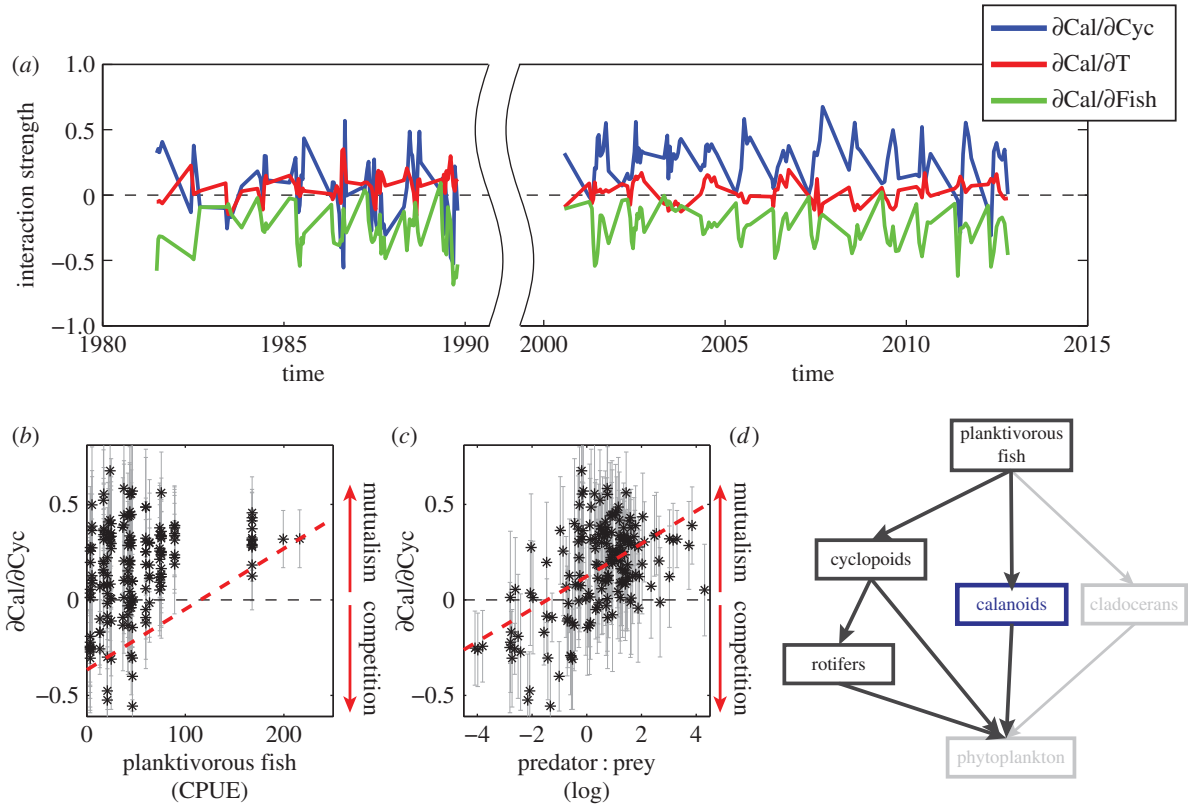

Figure 5. Dynamic interactions in Sparkling Lake. (*a*) S-map estimates of interaction coefficients quantify the changing effects of cyclopoids (Cyc), temperature (*T*) and planktivorous fish abundance (Fish) on calanoid copepods (Cal). The interaction of cyclopoids on calanoids ∂Cal/∂Cyc is shown as a function of (*b*) planktivorous fish abundance (CPUE) and (*c*) the log of the predator : prey ratio, with grey bars indicating 95% confidence on the S-map coefficients. The red dashed line represents the 0.05 quantile regression in (*c*) and regression on the mean in (*d*). Here, the predator : prey ratio is defined as Fish/(1 + Cal + Cyc).

identify the underlying mechanisms. In the model system, we are able to recover the known interactions directly from the time series. In the mesocosm, we find competition that intensifies as food becomes limiting (figure 4*b*,*c*)—as expected. Conversely, in Sparkling Lake, we find competition only when predator abundance is low (figure 5*b*) and a net positive interaction that intensifies as the predator : prey ratio increases (figure 5*c*)—suggestive of apparent mutualism.

The S-map procedure, extracts dynamics directly from time-series, and does not involve correlational evidence to construct a heuristic mode [35]. As noted elsewhere [15], such correlations can be inappropriate in nonlinear dynamic systems, which tend to produce 'mirage correlations', i.e. ephemeral associations among variables that appear then disappear, or even change sign. As a case in point, in Sparkling Lake where the sign of ∂Cal/∂Cyc clearly flips through time, the presence of an interaction may be missed with a linear time-averaged analysis. Indeed, this can explain why previous linear analysis of this system using vector autoregression did not find a significant linear-constant effect of cyclopoids on calanoids [34], as the positive and negative episodes would have cancelled out.

EDM applies if (i) there is a deterministic component to the ecosystem dynamics and (ii) there are sufficient time-series data to uncover an embedded attractor—sufficient data to generate an unfolded manifold where trajectories do not cross. Importantly, these two core assumptions can be validated by nearest-neighbour prediction (e.g. simplex projection [18]). Predictability indicates that there are deterministic dynamics and that there are few places where the future of the system is undetermined—where trajectories cross [36]. Testing these assumptions is a prerequisite for applying the method.

Conversely, EDM should not generally apply if the system cannot be properly embedded. This could occur in a purely stochastic system with no discernable dynamics or when observational noise dominates to the extent that no predictive nonlinear manifold can be uncovered [1,15–20,35,36]. While we have shown that S-maps accommodate reasonable amounts of observational noise (up to 30%) (electronic supplementary material, figure S8), there are other generalizations of vector autoregression that focus specifically on dealing with noise. Most notably, MARSS combines Kalman filters with the basic MAR framework to better identify fixed interactions in the presence of noise [13]. Although not trivial, it should be possible to combine noise-modelling methods like MARSS with S-maps—a possible avenue for future research.

In addition, and because the multivariate S-map method described here can be sensitive to the specific embedding coordinates, care must be taken to examine a comprehensive set of time-series variables that can be verified with a causation test (e.g. convergent cross mapping, CCM [15]) and a multivariate prediction test [17,19,28], as shown for the case studies here (see Material and methods and the electronic supplementary material). This is essential for applying the method sensibly.

Previous work has shown how EDM can be used for population forecasting [17,19,28] for exploring alternative environmental scenarios [20], and for detecting causal linkages [15]. Here, we apply the S-map approach and its intermediate output (the sequential jacobians) to track and forecast the changing interactions in ecosystems. While models and field

experiments can identify species interactions in the abstract, in the field these interactions are embedded in an evolving network of factors. Therefore, by allowing the study of interactions as they are realized in nature, EDM offers a path for studying biological systems as a dynamically changing and interconnected whole [15]. Moreover, insofar as the framework involves data that can be feasibly collected close to real time (e.g. as occurs at many LTER sites, fisheries systems and other monitoring programmes around the world) and can actually *forecast* expected interactions, we believe it could become a practical tool for ecosystem control and management. It is a conceptual framework that speaks to the critical importance of ongoing and long-term data collection, where additional data increases the accuracy of our forecasts, and our ability to handle novel scenarios.

6. Material and methods

(a) Mesocosm data

Data for the Baltic Sea mesocosm were obtained from the supplemental materials of Benincà *et al*. [32,33]. Because the data were sampled irregularly, they were processed to give time series with approximately one week (± 1 day) between successive observations. Only data containing periods with at least 15 successive (one week apart) observations were considered. All time series were normalized to have a mean of 0 and standard deviation of 1.

This system is an ideal test case because of its intermediate complexity, and previously noted nonlinear dynamics [32,33]. Major interactions are summarized in figure 5*d*. Prior analysis of interspecies relationships in the mesocosm [33] found evidence of coherent oscillations between predators and prey through certain periods, but did not reveal much about competition. We focus on the two main grazers, calanoids and rotifers, along with their two principle prey items, nanoflagellates and picocyanobacteria [33]. Each group was dominated by a single species or genus [32,33]. Cross-mapping [15] confirms that these four species interact closely (see electronic supplementary material, table S1 and figure S1), and the manifold with these four species gives excellent predictability of calanoid dynamics (see electronic supplementary material, figure S2). Thus, even though the mesocosm has many (greater than or equal to 10) potentially important state variables, the dynamics of the grazers can be well embedded in the lower, four-dimensional space. The results shown in figure 4 are robust to including additional mesocosm variables in the state space (see electronic supplementary material, figures S6 and S7).

(b) Sparkling lake data

Data for Sparkling Lake between 4 June 1981 and 13 November 2013 were from the Northern Lakes LTER online portal (http://lter.limnology.wisc.edu) and included the following datasets: Zooplankton—Trout Lake Area; Chlorophyll—Trout Lake Area; Chemical Limnology of Primary Study Lakes: Nutrients, pH and Carbon; and Fish Abundance. Following Beisner *et al*. [34], the zooplankton data were resolved to broad taxa—calanoids, cyclopoids, cladocerans and rotifers. The temperature time series, averaged measurements made at 1-m intervals from 1 to 15 m depth. For chlorophyll, we integrated across the regularly sampled depths {0 m, 3 m, 5 m, 8 m 10 m} as well as using surface measurements only. For the fish data, we aggregated catch per unit effort (CPUE) across the two main planktivores, cisco (*Coregonus artedii* LeSueur) and smelt (*Osmerus mordax* Mitchill). Beisner *et al*.'s [34] per cent smelt index was a less useful predictor of month-to-month changes in calanoids.

The zooplankton, chlorophyll and temperature data were processed to give time series with approximately 1 month (± 4 days) between successive observations. The same fish abundance was used across the whole calendar year. All time series were normalized to have a mean of 0 and standard deviation of 1.

Because of data limitations, CCM was not used to determine the best set of state variables. CCM requires taking consecutive time lags of observed variables, and the lake is not sampled regularly in winter months. Instead, as in [19], we used a system of sequential elimination. We began with an embedding containing all candidate variables: calanoids (Cal), cyclopoids (Cyc), cladocerans (Cld), rotifers (Rot), chlorophyll-*a* (Chl), temperature (T) and planktivorous fish abundance (Fish). We then evaluated which candidate (if any) most improved EDM predictability when excluded from the embedding [19]. This variable was eliminated, and the procedure was repeated sequentially until all remaining candidate variables were helpful predictors. This left us with the five-dimensional embedding $<$Cal(t), Cyc(t), Rot(t), T(t), Fish(t)$>$. Importantly, S-map analysis of this embedding shows evidence of nonlinear dynamics (see 'Weighting Parameter' in the electronic supplementary material, S1 and electronic supplementary material, figure S5). Thus, while previous research used linear methods [34], the dynamics can be more fully understood as nonlinear. Significantly, including either depth-integrated or surface chlorophyll-*a* degrades predictability. So while we expect food abundance to be an important variable, the monthly chlorophyll-*a* data reported in [34] does not appear to be informative. Whether this is an issue of temporal or taxonomic resolution is unclear.

Authors' contributions. The S-map framework of sequential Jacobians was described by G.S. and developed with examples by E.R.D. E.R.D. performed all the calculations, analyses and test cases. E.R.D. and G.S. wrote much of the text with help from R.M.M. and S.M.

Competing interests. We declare we have no competing interests.

Funding. This work was supported by National Science Foundation (NSF) grant no. DEB1020372, NSF-National Oceanic and Atmospheric Administration Comparative Analysis of Marine Ecosystem Organization (CAMEO) programme grant no. NA08OAR4320894/CAMEO, DoD-Strategic Environmental Research and Development Program 15 RC-2509; Lenfest Foundation Award 00028335, the Sugihara Family Trust, the Deutsche Bank-Jameson Complexity Studies Fund, The McQuown Fund and the McQuown Chair in Natural Sciences, University of California, San Diego.

Acknowledgement. We are grateful to P. Sugihara and H. Ye for the video animation, and we thank H. Ye, C. Hsieh, J. Melack, I. Altman, J. Thorson, N. Gallo for comments.

References

1. Anderson CNK, Hsieh C-H, Sandin SA, Hewitt R, Hollowed AB, Beddington J, May RM, Sugihara G. 2008 Why fishing magnifies fluctuations in fish abundance. *Nature* **452**, 835–839. (doi:10.1038/nature06851)
2. Lima M, Stenseth NC, Jaksic FM. 2002 Food web structure and climate effects on the dynamics of small mammals and owls in semi-arid Chile. *Ecol. Lett.* **5**, 273–284. (doi:10.1046/j.1461-0248.2002.00312.x)
3. Lima M, Morgan Ernest SK, Brown JH, Belgrano A, Stenseth NC. 2008 Chihuahuan Desert kangaroo rats: nonlinear effects of population dynamics, competition, and rainfall. *Ecology* **89**, 2594–2603. (doi:10.1890/07-1246.1)

4. Gratton C, Denno RF. 2003 Seasonal shift from bottom-up to top-down impact in phytophagous insect populations. *Oecologia* **134**, 487–495. (doi:10.1007/s00442-002-1137-8)
5. Werner EE. 1992 Individual behavior and higher-order species interactions. *Am. Nat.* **140**, S5–S32. (doi:10.1086/285395)
6. Jeschke JM, Kopp M, Tollrian R. 2004 Consumer-food systems: why type I functional responses are exclusive to filter feeders. *Biol. Rev.* **79**, 337–349. (doi:10.1017/S1464793103006286)
7. Abrams PA. 1980 Consumer functional response and competition in consumer-resource systems. *Theor. Popul. Biol.* **17**, 80–102. (doi:10.1016/0040-5809(80)90016-7)
8. Abrams PA, Holt RD, Roth JD. 1998 Apparent competition or apparent mutualism? Shared predation when populations cycle. *Ecology* **79**, 201–212. (doi:10.1890/0012-9658(1998)079[0201:ACOAMS]2.0.CO;2)
9. Chase JM *et al.* 2002 The interaction between predation and competition: a review and synthesis. *Ecol. Lett.* **5**, 302–315. (doi:10.1046/j.1461-0248.2002.00315.x)
10. Holt RD, Lawton JH. 1994 The ecological consequences of shared natural enemies. *Annu. Rev. Ecol. Syst.* **25**, 495–520. (doi:10.1146/annurev.es.25.110194.002431)
11. Schoenly K, Cohen JE. 1991 Temporal variation in food web structure: 16 empirical cases. *Ecol. Monogr.* **61**, 267–298. (doi:10.2307/2937109)
12. Tavares-Cromar AF, Williams DD. 1996 The importance of temporal resolution in food web analysis: evidence from a detritus-based stream. *Ecol. Monogr.* **66**, 91–113. (doi:10.2307/2963482)
13. Hampton SE, Holmes EE, Scheef LP, Scheuerell MD, Katz SL, Pendleton DE, Pendleton DE, Ward EJ. 2013 Quantifying effects of abiotic and biotic drivers on community dynamics with multivariate autoregressive (MAR) models. *Ecology* **94**, 2663–2669. (doi:10.1890/13-0996.1)
14. Ives A, Dennis B, Cottingham K, Carpenter S. 2003 Estimating community stability and ecological interactions from time-series data. *Ecol. Monogr.* **73**, 301–330. (doi:10.1890/0012-9615(2003)073[0301:ECSAEI]2.0.CO;2)
15. 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)
16. 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. (doi:10.1073/pnas.93.6.2608)
17. Sugihara G. 1994 Nonlinear forecasting for the classification of natural time series. *Phil. Trans. R. Soc. Lond. A* **348**, 477–495. (doi:10.1098/rsta.1994.0106)
18. Sugihara G, May RM. 1990 Nonlinear forecasting as a way of distinguishing chaos from measurement error in time-series. *Nature* **344**, 734–741. (doi:10.1038/344734a0)
19. Dixon PA, Milicich MJ, Sugihara G. 1999 Episodic fluctuations in larval supply. *Science* **283**, 1528–1530. (doi:10.1126/science.283.5407.1528)
20. 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)
21. Levins R, Pressick ML, Heatwole H. 1973 Coexistence patterns in insular ants. *Am. Sci.* **61**, 463–472.
22. MacArthur R. 1970 Species packing and competitive equilibrium for many species. *Theor. Popul. Biol.* **1**, 1–11. (doi:10.1016/0040-5809(70)90039-0)
23. 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)
24. Lamon EC, Carpenter SR, Stow CA. 1998 Forecasting pcb concentrations in lake michigan salmonids: a dynamic linear model approach. *Ecol. Appl.* **8**, 659–668. (doi:10.1890/1051-0761(1998)008[0659:FPC ILM]2.0.CO;2)
25. Carpenter SR, Brock WA. 2006 Rising variance: a leading indicator of ecological transition. *Ecol. Lett.* **9**, 311–318. (doi:10.1111/j.1461-0248.2005.00877.x)
26. Schindler DE, Rogers DE, Scheuerell MD, Abrey CA. 2005 Effects of changing climate on zooplankton and juvenile sockeye salmon growth in Southwestern Alaska. *Ecology* **86**, 198–209. (doi:10.1890/03-0408)
27. Holmes EE, Ward EJ, Wills K. 2012 MARSS: multivariate autoregressive state-space models for analyzing time-series data. *R J.* **4**, 11–19.
28. Perretti CT, Sugihara G, Munch SB. 2012 Nonparametric forecasting outperforms parametric methods for a simulated multispecies system. *Ecology* **94**, 794–800. (doi:10.1890/12-0904.1)
29. Hastings A, Powell TM. 1991 Chaos in a three-species food chain. *Ecology* **72**, 896–903. (doi:10.2307/1940591)
30. Post DM, Conners ME, Goldberg DS. 2000 Prey preference by a top predator and the stability of linked food chains. *Ecology* **81**, 8–14. (doi:10.1890/0012-9658(2000)081[0008:PPBATP]2.0.CO;2)
31. Cade BS, Noon BR. 2003 A gentle introduction to quantile regression for ecologists. *Front. Ecol. Environ.* **1**, 412–420. (doi:10.1890/1540-9295(2003)001[0412:AGITQR]2.0.CO;2)
32. Benincà E, Huisman J, Heerkloss R, Jöhnk KD, Branco P, van Nes EH, Scheffer M, Ellner SP. 2008 Chaos in a long-term experiment with a plankton community. *Nature* **451**, 822–825. (doi:10.1038/nature06512)
33. Benincà E, Jöhnk KD, Heerkloss R, Huisman J. 2009 Coupled predator–prey oscillations in a chaotic food web. *Ecol. Lett.* **12**, 1367–1378. (doi:10.1111/j.1461-0248.2009.01391.x)
34. Beisner BE, Ives AR, Carpenter SR. 2003 The effects of an exotic fish invasion on the prey communities of two lakes. *J. Anim. Ecol.* **72**, 331–342. (doi:10.1046/j.1365-2656.2003.00699.x)
35. Ye H, Beamish RJ, Glaser SM, Grant SCH, Hsieh C-H, Richards LJ, Schnute JT, Sugihara G. 2015 Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. *Proc. Natl Acad. Sci. USA* **112**, E1569–E1576. (doi:10.1073/pnas.1417063112)
36. Sugihara G, Grenfell B, May RM. 1990 Distinguishing error from chaos in ecological time series. *Phil. Trans. R. Soc. Lond. B* **330**, 235–251. (doi:10.1098/rstb.1990.0195)