Detecting Causality in Complex Ecosystems

George Sugihara,[1]* Robert May,[2] Hao Ye,[1] Chih-hao Hsieh,[3]* Ethan Deyle,[1] Michael Fogarty,[4] Stephan Munch[5]

Identifying causal networks is important for effective policy and management recommendations on climate, epidemiology, financial regulation, and much else. We introduce a method, based on nonlinear state space reconstruction, that can distinguish causality from correlation. It extends to nonseparable weakly connected dynamic systems (cases not covered by the current Granger causality paradigm). The approach is illustrated both by simple models (where, in contrast to the real world, we know the underlying equations/relations and so can check the validity of our method) and by application to real ecological systems, including the controversial sardine-anchovy-temperature problem.

Identifying causality (1) in complex systems can be difficult. Contradictions arise in many scientific contexts where variables are positively coupled at some times but at other times appear unrelated or even negatively coupled depending on system state (movie S1). Baltic Sea fisheries, for example, exhibit radically different dynamic control regimes (top-down versus bottom-up) depending on the threshold abundance of planktivores, causing the correlations between fish and zooplankton to change sign (2). Such state-dependent behavior is a defining hallmark of complex nonlinear systems (3, 4), and nonlinearity is ubiquitous in nature (3–11).

Ephemeral or "mirage" correlations are common in even the simplest nonlinear systems (7, 11–13), such as shown in Fig. 1 for two coupled difference equations that exhibit chaotic behavior (14):

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

When this happens, variables that may be positively coupled for long periods can spontaneously become anticorrelated or decoupled; this can create problems when fitting models to observational data (15).

Although correlation is neither necessary nor sufficient to establish causation, it remains deeply ingrained in our heuristic thinking (8, 13, 16, 17). One might conclude, for example, that the variables in Fig. 1 have no causal relation because they are uncorrelated. Obviously, lack of correlation does not imply lack of causation. Because of this and for reasons just given, the use of correlation to infer causation is risky, especially as we come to recognize that nonlinear dynamics are ubiquitous.

An alternative approach, Granger causality (GC) (18), provides a framework that uses predictability as opposed to correlation to identify causation between time-series variables. GC is recognized as the primary advance on the causation problem since Berkeley (1).

Variable X is said to "Granger cause" Y if the predictability of Y (in some idealized model) declines when X is removed from the universe of all possible causative variables, U (18). The key requirement of GC is separability, namely that information about a causative factor is independently unique to that variable (e.g., information about predator effects is not contained in time series for the prey) and can be removed by eliminating that variable from the model. Separability is characteristic of purely stochastic and linear systems, and GC can be useful for detecting interactions between strongly coupled (synchronized) variables in nonlinear systems. Separability reflects the view that systems can be understood a piece at a time rather than as a whole.

However, as Granger (18) realized early on, this approach may be problematic in deterministic settings, especially in dynamic systems with weak to moderate coupling. For example, GC gives ambiguous results for the system in Fig. 1 (see GC calculations S1). This is because separability is not satisfied in such systems, which, unlike the tradition in economics and single-species fisheries management, need to be considered as a whole. That is to say, in deterministic dynamic systems (even noisy ones), if X is a cause for Y, information about X will be redundantly present in Y itself and cannot formally be removed from U—a consequence of Takens' theorem (19, 20). To see this directly, we note simply that Eq. 1 can be rewritten as a model for $Y(t + 1)$ in terms of $Y(t)$ and $Y(t - 1)$ (see box S1 for a worked example). Therefore, information about $X(t)$ that is relevant to predicting Y is redundant in this system and cannot be removed simply by eliminating X as an explicit variable. When Granger's definition is violated, GC calculations are no longer valid, leaving the question of detecting causation in such systems unanswered.

In addition to nonseparability, ecosystems differ from the systems typically studied with Granger's approach in other important ways. First, in eco-system dynamics, weak to moderate coupling is the norm. McCann (21) and others have developed a strong case for the ubiquity of weak coupling in ecological food webs and have demonstrated their importance for system stability. Second, ecosystems are typically subject to forcing by external driving variables such as temperature, precipitation, and upwelling [e.g., (6, 22)]. Because many species share similar abiotic environments, this can lead to correlations and apparent synchrony among noninteracting species [e.g., the Moran effect (23)], complicating the task of sorting out the real interactions from spurious correlations. It is therefore important in ecology to have methods that (i) address nonseparable systems, (ii) identify weakly coupled variables, and (iii) distinguish interactions among species from the effects of shared driving variables.

Here, we examine an approach specifically aimed at identifying causation in ecological time series. We demonstrate the principles of our approach with simple model examples, showing that the method distinguishes species interactions from the effects of shared driving variables. Finally, we apply the method to ecological data from experimental and field studies, showing how it distinguishes top-down from bottom-up control in the classic *Paramecium-Didinium* experiment and clarifies the ongoing debate about the nature of interactions among sardine, anchovy, and sea surface temperature in the California Current ecosystem.

Our approach is not in competition with the many effective methods that use GC (see supplementary text); rather, it is specifically aimed at a class of system not covered by GC. As verified in GC calculations S1 to S5 and box S1, GC does not apply to this class of system.

Dynamic causation and CCM. GC applies if the world is purely stochastic. However, to the extent that it is deterministic and dynamics are not entirely random, there will be an underlying manifold governing the dynamics (representing coherent trajectories as opposed to a random tangle).

In dynamical systems theory, time-series variables (say, X and Y) are causally linked if they are from the same dynamic system (4, 19, 20)—that is, they share a common attractor manifold M (movies S1 to S3 illustrate this idea). This means that each variable can identify the state of the other (3, 19, 20, 24, 25) (e.g., information about past prey populations can be recovered from the predator time series, and vice versa). Additionally, when one variable X is a stochastic environmental driver of a population variable Y, information about the states of X can be recovered from Y, but not vice versa. For example, fish time series can be used to estimate weather, but not conversely. This runs counter to Granger's intuitive scheme (see explanation in box S1).

Our alternative approach, convergent cross mapping (CCM), tests for causation by measuring the extent to which the historical record of Y values can reliably estimate states of X. This happens only if X is causally influencing Y. In more detail, CCM looks for the signature of X in Y's time series by seeing whether there is a correspondence between the

[1]Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA. [2]Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. [3]Institute of Oceanography and Institute of Ecology and Evolutionary Biology, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 106, Taiwan. [4]Northeast Fisheries Science Center, National Oceanic and Atmospheric Administration, 166 Water Street, Woods Hole, MA 02543, USA. [5]Southwest Fisheries Science Center, National Oceanic and Atmospheric Administration, 110 Shaffer Road, Santa Cruz, CA 95060, USA.

*To whom correspondence should be addressed. E-mail: gsugihara@ucsd.edu (G.S.); chsieh@ntu.edu.tw (C.H.)

"library" of points in the attractor manifold built from Y, M_Y, and points in the X manifold, M_X, where these two manifolds are constructed from lagged coordinates of the time-series variables Y and X, respectively (3, 19, 24) (movies S1 and S2).

Essentially, the idea is to see whether the time indices of nearby points on the Y manifold can be used to identify nearby points on M_X. If so, then one can use Y to estimate X and vice versa. This procedure is illustrated in Fig. 2 and movie S3, with full technical details including an algorithm in (26).

Note that CCM is related to the general notion of cross prediction (3, 25) but with important differences. First, CCM estimates "states" across variables and does not forecast how the system "evolves" on the manifold. This eliminates possible information loss from chaotic dynamics (Lyapunov divergence) and accommodates nondynamic (i.e., random) variables. More important, CCM involves convergence, a key property that distinguishes causation from simple correlation. Convergence means that cross-mapped estimates improve in estima-

tion skill with time-series length L (sample size used to construct a library) (Fig. 3A, fig. S2, and box S1). With more data, the trajectories defining the attractor fill in, resulting in closer nearest neighbors and declining estimation error (a higher correlation coefficient) as L increases (Fig. 2). Thus, CCM becomes a necessary condition for causation. Indeed, failure to account for convergence explains conflicting results reported in the literature with related methods (supplementary text and fig. S5).

In practical applications, where shadow manifolds are low-dimensional approximations of the true system, convergence will be limited by observational error, process noise, and time-series length L. Thus, with limited or noisy field data, CCM is demonstrated by predictability that increases with L (fig. S3). See (26) for a discussion of data requirements.

Framework for identifying causation, case (i) Bidirectional causality via functional coupling. Bidirectional causality is analogous to the concept of "feedback" between two time series described by Granger (18) and is the primary case covered by Takens (19). Simply put, if variables are mutually coupled (e.g., predator and prey), they will cross map in both directions (Fig. 3A and fig. S1A). Thus, each variable can be estimated from the other (predator histories can estimate prey states). Figure 3B gives examples of the general case i.

Notice that as the strength of coupling increases, information becomes more distinct in the affected variables. As a result, their manifolds will contain stronger historical signatures of the causes. In Fig. 1 (Eq. 1), for example, where $\beta_{y,x} \gg \beta_{x,y}$ the much stronger effect of species X on Y implies faster convergence for predicting X than for Y (Fig. 3A). Thus, all things equal, the relative skill of cross mapping can indicate the relative strength of causative effect (Fig. 3B).

Framework for identifying causation, case (ii) Unidirectional causality. Here, species X influences the dynamics of Y, but Y has no effect on X (Fig. 3C and fig. S1B). This describes an amensal or commensal relationship, or where X represents external environmental forcing.

Figure 3C examines the system when $\beta_{x,y} = 0$. Notice that with moderately strong forcing from X (via $\beta_{y,x}$), even though Y exerts no effect, there may still be partial cross mapping of Y arising from the contemporaneous dependence of Y on X. However, this statistical effect is not convergent (shown by the asymptotic level curves with respect to L in Fig. 3E). With extremely strong forcing, the intrinsic dynamics of the forced variable become subordinate to the forcing variable, leading to the well-studied phenomenon of "synchrony" (27). The red plateau in Fig. 3E shows that bidirectional convergence can occur with strong forcing. Thus, strong forcing (synchrony) must be ruled out for CCM to unequivocally imply bidirectional coupling, although it still implies membership in a common dynamic system.

Transitivity. Notice that causation is transitive (e.g., if foxes prey on rabbits, and rabbits eat grass, then foxes and grass are causally linked). More formally, $X \Leftrightarrow Y \Leftrightarrow Z$ implies $X \Leftrightarrow Z$, whether or not X and Z interact directly. Similarly, for unidirectional

$$X(t+1) = X(t) [3.8 - 3.8\ X(t) - 0.02\ Y(t)]$$
$$Y(t+1) = Y(t) [3.5 - 3.5\ Y(t) - 0.1\ X(t)]$$
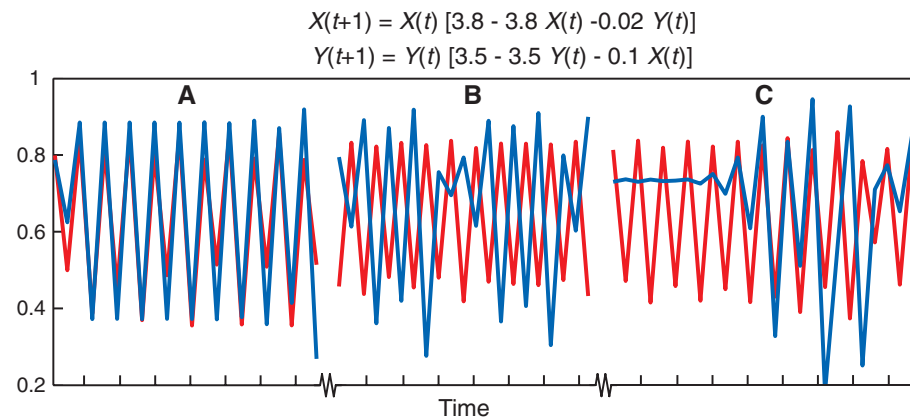
Fig. 1. Mirage correlations. (A to C) Three samples from a single run of a coupled two-species nonlinear logistic difference system with chaotic dynamics. Variables X (blue) and Y (red) appear correlated in the first time segment (A), anticorrelated in the second time segment (B), and lose all coherence 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 ($n = 1000$, $\rho = 0.0054$, $P = 0.864$).
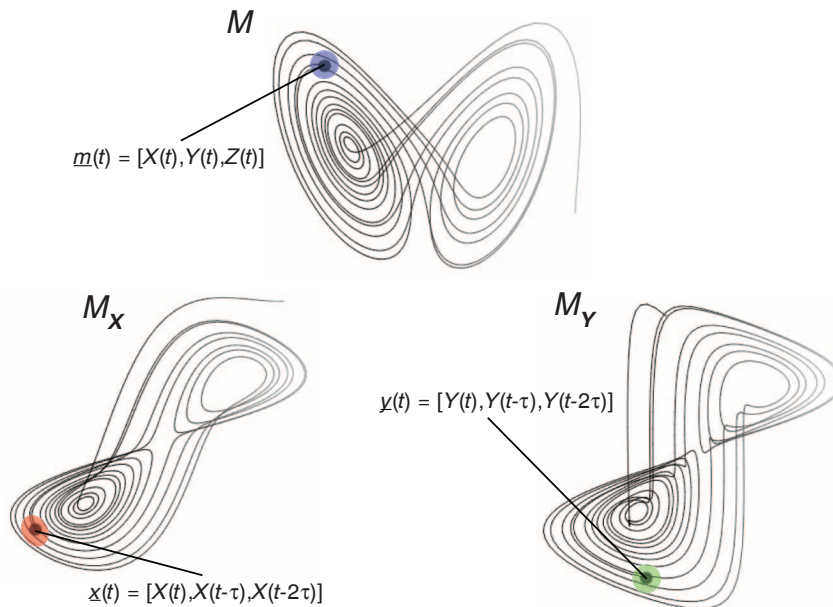
Fig. 2. Convergent cross mapping (CCM) tests for correspondence between shadow manifolds. This example based on the canonical Lorenz system (a coupled system in X, Y, and Z; eq. S7 without V) shows the attractor manifold for the original system (M) and two shadow manifolds, M_X and M_Y, constructed using lagged-coordinate embeddings of X and Y, respectively (lag = τ). Because X and Y are dynamically coupled, points that are nearby on M_X (e.g., within the red ellipse) will correspond temporally to points that are nearby on M_Y (e.g., within the green circle). That is, the points inside the red ellipse and green circle will have corresponding time indices (values for t). This enables us to estimate states across manifolds using Y to estimate the state of X and vice versa using nearest neighbors (3). With longer time series, the shadow manifolds become denser and the neighborhoods (ellipses of nearest neighbors) shrink, allowing more precise cross-map estimates (see movies S1 to S3).
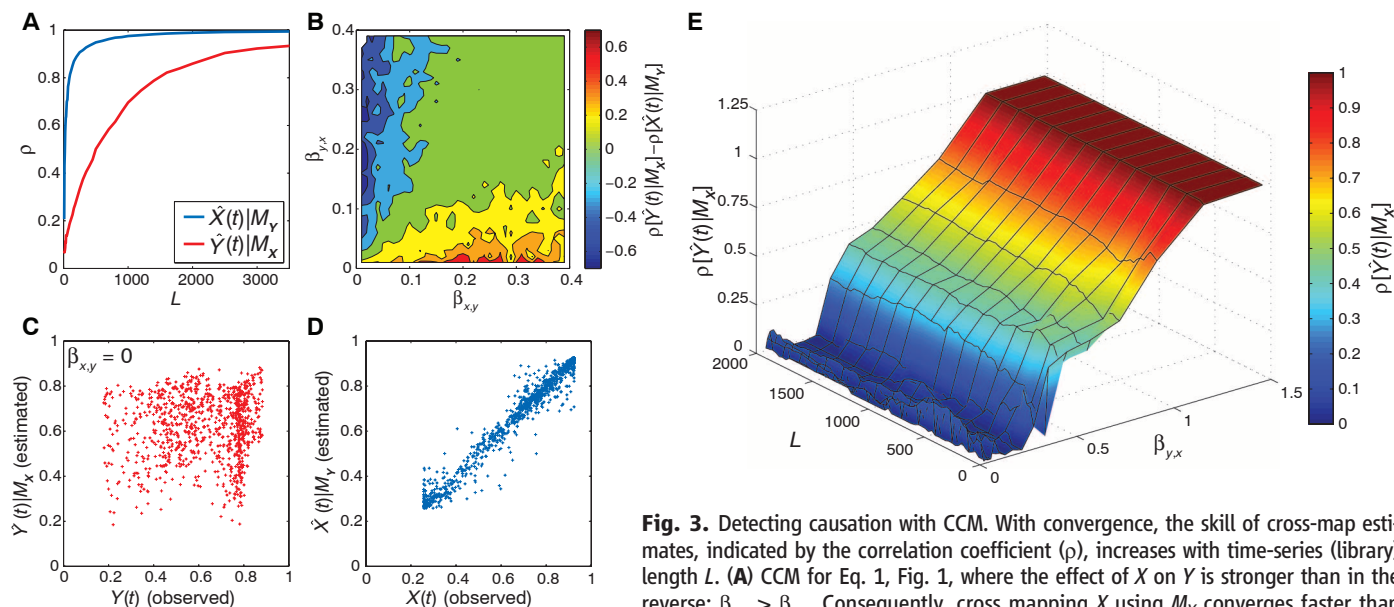
Fig. 3. Detecting causation with CCM. With convergence, the skill of cross-map estimates, indicated by the correlation coefficient (ρ), increases with time-series (library) length L. (**A**) CCM for Eq. 1, Fig. 1, where the effect of X on Y is stronger than in the reverse: $\beta_{y,x} > \beta_{x,y}$. Consequently, cross mapping X using M_Y converges faster than cross mapping Y using M_X. (**B**) Summary of this effect for Eq. 1, $L = 400$. (**C** to **E**) When Y (red) has no effect on X (blue) (i.e., $\beta_{x,y} = 0$), (C) shows that cross mapping of Y using M_X fails; however, the cross map of X succeeds (D) because the time series for Y contains information about the dynamics of X. (E) demonstrates nonconvergence of $\hat{Y}(t)$ as a function of forcing strength when $\beta_{x,y} = 0$. Convergence only occurs as a special case if strong forcing causes the system to collapse dimensionality (dark red plateau at high $\beta_{y,x}$), thus removing the dynamics of Y.

forcing, $X \Rightarrow Y$ and $Y \Rightarrow Z$ implies $X \Rightarrow Z$. Transitivity provides the basis for extending CCM to larger interaction networks, enabling us to distinguish variables that are coupled from those sharing a common driver. This is illustrated with two model examples below.

Complex model examples: External forcing of noncoupled variables. Consider the case where two species, X and Y, do not interact but are both driven by a common environmental variable Z (example 1 schematic in Fig. 4A). This occurs commonly in ecological systems [the Moran effect (*23*)] and remains problematic in studies of causation. Here we expect no cross mapping between species X and Y because there is no information flow between variables; however, information about the external forcing variable (Z) should still be recoverable from X and Y.

In fisheries, for example, noninteracting populations with common peak recruitment years due to favorable environmental conditions may be correlated even though they do not interact. The simple fisheries model in Fig. 4B illustrates this situation (*26*), where although the significant cross-correlation between species suggests that they might be coupled, cross mapping shows no evidence of convergence, proving that they are not coupled. This shows that CCM can distinguish true interaction from a simple correlation produced by shared driving variables.

Figure 4C provides an interesting further illustration of the method with a more complex five-species model [schematic in Fig. 4A, model details in (*26*)]. In this example, species 1, 2, and 3 represent a mutually interacting guild that externally force species 4 and 5, whereas 4 and 5 do not influence any other species. Species 1, 2, and 3 are akin to Z in the discussion above, with 4 and 5 akin to the externally forced noncoupled pair X and Y. Figure 4C shows that CCM is able to deduce the correct network of
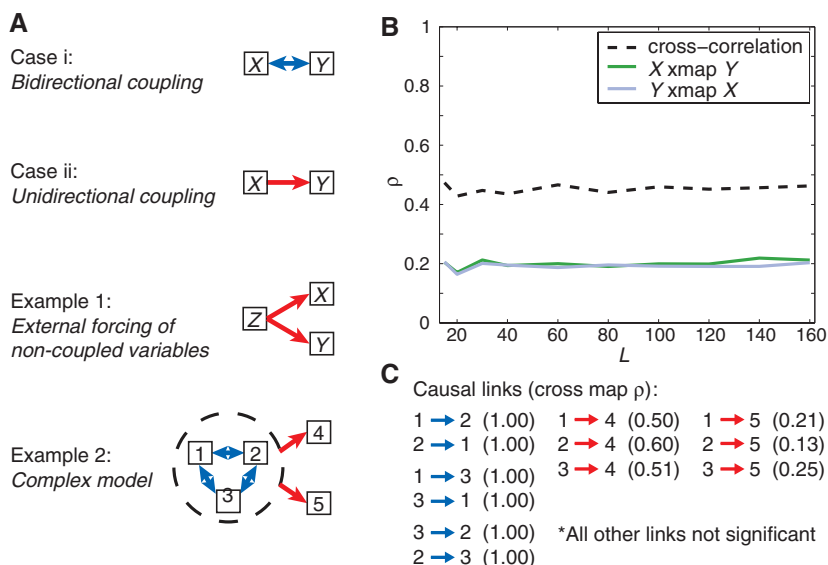
Fig. 4. Model causal networks. (**A**) Schematics of causal networks: two base cases and two model examples showing external forcing of noncoupled variables. (**B**) Cross-map results for example 1: external forcing of noncoupled variables. Cross-correlation erroneously suggests that X and Y are interacting, whereas cross mapping correctly shows that there is no interaction. (**C**) Cross-map results for the complex five-species model example. All significant ($P < 0.05$) mappings are given and indicate that species 1, 2, and 3 (the subsystem in the circle) all interact mutually (case i), but interact only asymmetrically as external forcing variables with respect to 4 and 5 (case ii), which do not interact directly themselves.

interactions getting all bidirectional and unidirectional links correct.

Real-world examples: Demonstration with ecological data. Keep in mind that attractors constructed from real data are approximations of dynamics occurring in higher dimensions. Thus, although observational error and process noise will limit the level of convergence attainable, low-dimensional approximations can still produce significant cross-map estimates of causal effects.

Bidirectional causation in an experimental predator-prey system. We apply the analysis to time series from the classic experimental predator-prey system, first studied in the 1920s by Gause and later improved by Veilleux (*28*), involving *Didinium* (predator) and *Paramecium* (prey) [methodological details in (*26*)].

The results in Fig. 5A suggest bidirectional coupling (case i), which accords with what is known. Moreover, the higher level of skill in
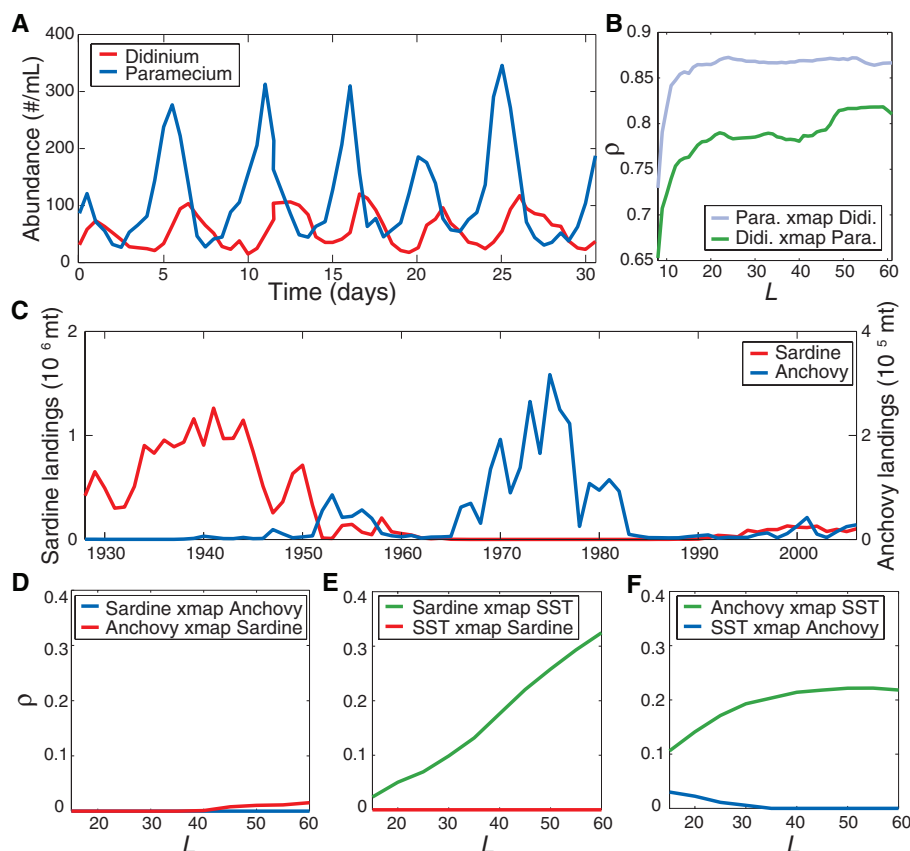
Fig. 5. Detecting causation in real time series. (**A**) Abundance time series of *Paramecium aurelia* and *Didinium nasutum* as reported in (*28*). (**B**) CCM of *Paramecium* and *Didinium* with increasing time-series length *L*. The pattern suggests top-down predator control. (**C**) California landings of Pacific sardine (*Sardinos sagax*) and northern anchovy (*Engraulis mordax*). (**D** to **F**) CCM (or lack thereof) of sardine versus anchovy, sardine versus SST (Scripps Pier), and anchovy versus SST (Newport Pier), respectively. This shows that sardines and anchovies do not interact with each other and that both are forced by temperature.

cross mapping *Didinium* from the *Paramecium* time series than the reverse (Fig. 5B) suggests that top-down control by the predator, *Didinium*, is stronger than bottom-up control by the prey, *Paramecium*. This finding is consistent with the experimental protocol and illustrates asymmetrical bidirectional coupling (case i).

Complex causation in the sardine-anchovy system. Here, we examine the relationship among Pacific sardine (*Sardinops sagax*) landings, northern anchovy (*Engraulis mordax*) landings, and sea surface temperature (SST) measured at Scripps Pier and Newport Pier, California (Fig. 5C).

Competing hypotheses have been advanced to explain the pattern of alternating dominance of sardine and anchovy across global fisheries on multidecadal time scales. Although the observed reciprocal abundance levels (Fig. 5A) resonates with ecological competition as an underlying mechanism, global synchrony in sardine and anchovy stanzas (*29*) suggests the operation of large-scale environmental forcing coupled with species-specific differences in optimal temperature levels. Recent evidence of regime-like behavior in these systems suggests the operation of nonlinear processes (*10*).

Similar to the global pattern, in California, 20th-century landings of Pacific sardine and northern

anchovy show one population peaking when the other is low. Whereas some (*30*) have hypothesized that the species act in direct competition, others (*31*) have argued that the species react differently to common large-scale environmental forcing. Moreover, paleoecological time series based on fish scales preserved in the anoxic sediments of the Santa Barbara basin revealed that the negative cross-correlation witnessed in the 20th century disappears in these longer time series (*32*). Correlation with environmental factors has also been elusive. Jacobson and MacCall (*33*) used two approaches, a generalized additive model and a linearized Ricker stock-recruitment model with environmental terms, and detected correlation between 3-year running averages of the Scripps Pier SST versus sardine recruitment and spawning stock size. However, when the analysis was expanded to include recent stock assessments from 1992 to 2009, the relationships vanished (*34*). Although there are many possible explanations, such behavior is consistent with nonlinear dynamics and mirage correlation.

We address this controversy using the same analytical protocol used for the *Didinium-Paramecium* example (*26*). The results in Fig. 5D show no significant cross-map signal between sardine and anchovy landings, indicating that sardines and anchovies

do not interact. In addition, as expected, there is no detectable signature from either sardine or anchovy in the temperature manifold; obviously, neither sardines nor anchovies affect SST. However, there is clear asymmetric CCM between sardines and SST as well as between anchovies and SST (Fig. 5, E and F), meaning that temperature information is encoded in both fishery time series. The recoverable temperature signature reveals a weak coupling of temperature to sardines and anchovies. Thus, although sardines and anchovies are not actually interacting, they are weakly forced by a common environmental driver, for which temperature is at least a viable proxy. Note that because of transitivity, temperature may be a proxy for a group of driving variables (i.e., temperature may not be the most proximate environmental driver). Our finding that SST influences sardine and anchovy population size (Fig. 5, E and F) is consistent with earlier findings of Jacobson and MacCall (*33*). Supporting evidence with other fishery-independent data are provided in the supplementary text (figs. S3 and S4).

Finally, it is important to note that the measurable nonlinear coupling of temperature to sardine stocks means that the effect of temperature varies with system state. Therefore, contrary to the current regulatory framework for sardines, a fixed temperature index will not suffice for sound management decisions. Rather, a dynamic (state-dependent) rule involving temperature is required.

Final remarks on nonseparability. One of the fundamental ideas in this work is that when causation is unilateral, $X \Rightarrow Y$ ("*X* drives *Y*," as in case ii), then it is possible to estimate *X* from *Y*, but not *Y* from *X*. This runs counter to intuition (and GC), and suggests that if the weather drives fish populations, for example, we can use fish to estimate the weather but not conversely.

To further clarify how this works, consider the two-species logistic model described earlier (Eq. 1). We can recover the cross-map dynamics algebraically by rearranging Eq. 1 to give expressions for $Y(t)$ and $X(t)$, substituting these back into Eq. 1, and solving for $X(t)$ in terms of $Y(t)$ and $Y(t-1)$ (and conversely; see box S1).

In these expressions, the parameter $\beta_{x,y}$ governs the sensitivity of *X* to changes in *Y*. As $\beta_{x,y}$ approaches 0, *X* drives *Y* unilaterally (case ii) and the cross-map estimate of *X* remains well-behaved. But the cross-map model for *Y* has a singularity when $\beta_{x,y} = 0$, meaning that cross mapping allows the driver to be reconstructed from the driven variable, but not the other way around (fish reflect weather states, but not conversely).

Finally, because Eq. 1 (parameterized as in Fig. 1) can be algebraically rearranged as a model for $X(t+1)$ purely in terms of $X(t)$ and $X(t-1)$, the information from *Y* becomes redundant and can be removed without affecting our ability to predict $X(t+1)$. Thus, GC would conclude (incorrectly!) that *Y* does not cause *X* (GC calculation S1).

Summary. Despite the fundamental problems raised in Berkeley's 1710 *A Treatise on Principles of Human Knowledge* (*1*), correlation remains the analytical standard of modern science. This has be-

come more difficult to justify with increasing recognition that nonlinear dynamics are ubiquitous. Apparent relationships among variables can switch spontaneously in nonlinear systems as a result of mirage correlations or a threshold change in regime, and correlation can lead to incorrect and contradictory hypotheses. Growing recognition of the prevalence and importance of nonlinear behavior calls for a better criterion for evaluating causation where experimental manipulation is not possible.

Granger causality addresses Berkeley's issues with prediction rather than correlation as the criterion for causation in time series. This idea assumes that causes can be separated from effects, so that a variable is identified as causative if prediction skill declines when that variable is removed. This is possible in a purely stochastic world and is a powerful idea for systems that can be studied as independent pieces; however, it is not defined for all systems, and in particular not for deterministic dynamic systems (even noisy ones) where Takens' theorem applies (19, 20). To address this, we examine an approach that exploits nonseparability by using CCM to test for membership to a common dynamical system. CCM is not a method competing with GC, but deals with interdependence often found in ecological study where GC is simply not applicable. Thus it is not surprising that as a further check, the GC calculations for all the model and real data examples considered in this work were largely unsuccessful (table S2 and GC calculations S1 to S5).

Although many empirical measures of species interactions exist (e.g., inferring interaction proxies from diet matrices), we suggest that causation inferred from time-series information provides a "bottom-line" picture of interactions that is more direct than those possible with proxies. The ability to resolve causal networks from their dynamical behavior has implications for system identification and ecosystem-based management, particularly where it is important to know which species interact as a group and need to be considered together. In resource management, as elsewhere, accurate knowledge of the causal network can be essential for avoiding unforeseen consequences of regulatory actions.

References and Notes

1. G. Berkeley, *A Treatise on Principles of Human Knowledge* (1710).
2. M. Casini *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 197 (2009).
3. G. Sugihara, R. M. May, *Nature* **344**, 734 (1990).
4. P. A. Dixon, M. J. Milicich, G. Sugihara, *Science* **283**, 1528 (1999).
5. W. A. Brock, C. L. Sayers, *J. Monet. Econ.* **22**, 71 (1988).
6. B. T. Grenfell *et al.*, *Nature* **394**, 674 (1998).
7. A. Mysterud *et al.*, *Nature* **410**, 1096 (2001).
8. C. H. Hsieh, S. M. Glaser, A. J. Lucas, G. Sugihara, *Nature* **435**, 336 (2005).
9. G. I. Bischi, *et al.*, Eds., *Nonlinear Dynamics in Economics, Finance and Social Sciences* (Springer, Berlin, 2010).
10. D. A. S. Patil *et al.*, *Mon. Weather Rev.* **129**, 2116 (2001).
11. T. T. Lo, H. H. Hsu, *Atmos. Sci. Lett.* **11**, 210 (2010).
12. X. Rodo, M. Pascual, G. Fuchs, A. S. G. Faruque, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12901 (2002).
13. B. K. Wagner *et al.*, *Nat. Biotechnol.* **26**, 343 (2008).
14. A. L. Lloyd, *J. Theor. Biol.* **173**, 217 (1995).
15. Committee on Major U.S. Oceanographic Research Programs, National Research Council, *Global Ocean Science: Toward an Integrated Approach* (National Academies Press, Washington, DC, 1999).
16. Y. Chen *et al.*, *J. Neurosci. Methods* **150**, 228 (2006).
17. R. M. May, S. A. Levin, G. Sugihara, *Nature* **451**, 893 (2008).
18. C. W. J. Granger, *Econometrica* **37**, 424 (1969).
19. F. Takens, in *Dynamical Systems and Turbulence*, D. A. Rand, L. S. Young, Eds. (Springer-Verlag, New York, 1981), pp. 366–381.
20. E. R. Deyle, G. Sugihara, *PLoS ONE* **6**, e18295 (2011).
21. K. McCann, A. Hastings, G. R. Huxel, *Nature* **395**, 794 (1998).
22. C. H. Hsieh *et al.*, *Fish. Oceanogr.* **18**, 102 (2009).
23. P. A. P. Moran, *Aust. J. Zool.* **1**, 291 (1953).
24. N. H. Packard *et al.*, *Phys. Rev. Lett.* **45**, 712 (1980).
25. F. A. Ascioti *et al.*, *J. Plankton Res.* **15**, 603 (1993).
26. See supplementary materials on *Science* Online.
27. K. Josic, *Nonlinearity* **13**, 1321 (2000).
28. B. G. Veilleux, thesis, University of Alberta (1976).
29. R. A. Schwartzlose *et al.*, *S. Afr. J. Mar. Sci.* **21**, 289 (1999).
30. G. I. Murphy, J. D. Isaacs, Species replacement in marine ecosystems with reference to the California current. *Minutes of Meeting Marine Research Committee* **7**, 1 (1964).
31. R. Lasker, A. MacCall, in *Proceedings of the Joint Oceanographic Assembly, Halifax, August 1982: General Symposia* (Department of Fisheries and Oceans, Ontario, 1983), pp. 110–120.
32. T. R. Baumgartner *et al.*, *CCOFI Rep.* **33**, 24 (1992).
33. L. D. Jacobson, A. D. MacCall, *Can. J. Fish. Aquat. Sci.* **52**, 566 (1995).
34. S. McClatchie *et al.*, *Can. J. Fish. Aquat. Sci.* **67**, 1782 (2010).

REPORTS

Biotinylated Rh(III) Complexes in Engineered Streptavidin for Accelerated Asymmetric C–H Activation

Todd K. Hyster,[1,2] Livia Knörr,[2] Thomas R. Ward,[2]* Tomislav Rovis[1]*

Enzymes provide an exquisitely tailored chiral environment to foster high catalytic activities and selectivities, but their native structures are optimized for very specific biochemical transformations. Designing a protein to accommodate a non-native transition metal complex can broaden the scope of enzymatic transformations while raising the activity and selectivity of small-molecule catalysis. Here, we report the creation of a bifunctional artificial metalloenzyme in which a glutamic acid or aspartic acid residue engineered into streptavidin acts in concert with a docked biotinylated rhodium(III) complex to enable catalytic asymmetric carbon-hydrogen (C–H) activation. The coupling of benzamides and alkenes to access dihydroisoquinolones proceeds with up to nearly a 100-fold rate acceleration compared with the activity of the isolated rhodium complex and enantiomeric ratios as high as 93:7.

Thanks to the advent of genetic engineering, enzymes are attracting increasing attention as versatile synthetic tools, even displacing established organometal-catalyzed industrial processes (1). However, creating an enzyme for an abiotic reaction from a noncatalytic scaffold remains a major challenge (2–5). One class of strategies has relied on the incorporation of non-natural metal cofactors within a protein scaffold to afford artificial metalloenzymes (6–9). The main focus in the area has been improving the selectivity of the hybrid catalysts, rather than reaction rates, which are, by and large, dictated by the first coordination sphere interactions around the metal (10–12). Among the various cofactor localization strategies (13, 14), the biotin-(strept)avidin technology has proven versatile: The geometry of the biotin-binding pocket is ideally suited to accommodate organometallic moieties, leaving enough room for substrate binding and activation (15–19).

[1]Department of Chemistry, Colorado State University, Fort Collins, CO 80523, USA. [2]Department of Chemistry, University of Basel, Basel CH-4056, Switzerland.

*To whom correspondence should be addressed. E-mail: rovis@lamar.colostate.edu (T.R.); thomas.ward@unibas.ch (T.R.W.)