UNIVERSITY OF CALIFORNIA, SAN DIEGO

Multivariate Empirical Dynamic Approaches to State-Dependence in Ecological Dynamics and Management: A practical, mathematical investigation into sidestepping reductionism in the irreducible natural world

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Oceanography

by

Ethan Robert Deyle

Committee in charge:

George Sugihara, Chair
Lin Chao
Peter Franks
Arthur Miller
Stephan Munch
Stuart Sandin

2015

The dissertation of Ethan Robert Deyle is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2015

EPIGRAPH

Far away in the heavenly abode of the great god Indra, there is a wonderful net that stretches out indefinitely in all directions. At each node of the net is hung a single glittering jewel. If we select one of these jewels for inspection we will discover that in its polished surface are reflected all the other jewels in the net, infinite in number, with each reflected in this one jewel reflecting all the others, so that the process of reflection is infinite.

—-**Avatamsaka Sutra**

Fortney & Onellion Seeking Truth: the With Doubt

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to above all acknowledge Professor George Sugihara for his tireless support and guidance as the Chair of my dissertation committee and for the crucial role he played in introducing me to this most exciting field of science. I also would like to acknowledge the past and present lab-mates that shared their time, ideas, and code with me- Dr. Sarah Glaser, Dr. Charles Perretti, and Hao Ye. Finally, I would like to acknowledge the long list of co-authors and collaborators who made the various pieces of this unlikely and interdisciplinary dissertation possible. In no particular orderÑ Dr. Michael Fogarty, Dr. Alec MacCall, Prof. Les Kaufman, Dr. Andy Rosenberg, Dr. Emily Klein, Professor David Tillman, Adam Clark, Prof. Chih-hao Hsieh, Profssor Stephan B. Munch, Dr. Sue Hopkins, Dr. Gerald Pao, Prof. John McGowan, Dr. Melissa Carter, Dr. Mary Hilborn, Dr. Alain de Verneil, Dr. Michael DeFlorio, Prof. Art Miller, Kerri Seger, Dr. Eugene Ke, Prof. Inder Verma, and Prof. Robert May.

Chapter 2, in full, is a reprint of the material as it appears in Public Library of Science ONE 2011. Deyle, Ethan Robert; Sugihara. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in Proceedings of the National Academy of Science USA, 2012. Deyle, Ethan Robert; Fogarty, Michael; Hsieh, Chih-hao; Kaufman, Les; MacCall, Alec D.; Munch, Stephan B; Perretti, Charles T.; Ye, Hao; and Sugihara, George. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, has been submitted for publication of the material as it may appear in Proceedings of the Royal Society B, 2015. Deyle, Ethan Robert; Munch,

Stephan B; May, Robert M; Sugihara, George. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in part is currently being prepared for submission for publication of the material. Mayer, M. Cyrus; Hernandez, Ryan; Basu, Sanjay; Sugihara, George. The dissertation author was the primary investigator and author of this material.

VITA

2008	B. A. in Physics *magna cum laude*, Swarthmore College
2009	M. Sci. in Applied Mathematics, University of Cambridge
2009-2015	Graduate Researcher, University of California, San Diego
2012	M. S. in Marine Biology, University of California, San Diego
2013	Graduate Teaching Assistant, University of California, San Diego
2015	Ph. D. in Oceanography, University of California, San Diego

AWARDS

2010	EPA-STAR Fellowship
2011	NSF Graduate Research Fellowship

PUBLICATIONS

Ye H., Deyle E.R., Gilarranz L., and Sugihara G. Distinguishing time-delayed causal interactions using convergent cross mapping. Scientific Reports, *in press*.

Tsonis A.A., Deyle E., May R.M., Sugihara G., Swanson K., Verbeten J.D., and Wang G. Dynamical evidence for causality between galactic cosmic rays and interannual variation in global temperature. Proceedings of the National Academy of Sciences USA 112: 3253Ð3256 (2015).

van Nes E., Scheffer M., Brovkin V., Lenton T.M., Ye H., Deyle E., and Sugihara G. Causal feedbacks in climate change. Nature Climate Change 5: 445-448 (2015).

Deyle E., Fogarty M.J., Hsieh C.H., Kaufman L., MacCall A., Munch S., Perretti C., Ye H., and Sugihara G. Predicting climate effects on Pacific sardine. Proceedings of the National Academy of Sciences USA 110: 6430-6435 (2013).

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

Deyle, E., Sugihara G. Generalized Theorems for Nonlinear State Space Reconstruction. PLoS 6: e18295 (2011).

ABSTRACT OF THE DISSERTATION

Multivariate Empirical Dynamic Approaches to State-Dependence in Ecological Dynamics and Management: A practical, mathematical investigation into sidestepping reductionism in the irreducible natural world

by

Ethan Robert Deyle

Doctor of Philosophy in Oceanography

University of California, San Diego, 2015

Professor George Sugihara, Chair

In this dissertation, I investigate a series of seemingly disparate topicsÑthe theorem of a mathematician working on turbulence in fluid flows, the collapse of the great California sardine fishery in the mid 20th century, competition between zooplankton grazers in a marine mesocosm, and the occurrence of influenza in the tropics. All of the work, however, is motivated by the ongoing endeavor to develop ecosystem-based approaches to fisheries management. Ecosystem-based management remains an open problem in part because the fundamental complexity of natural systems does not readily

map onto the set of tools we've developed through our very successful accomplishments in engineering. In combination, the chapters of this thesis address this disparity by developing practical, empirical tools for studying and managing ecosystems that directly address the complex reality of nature.

Chapter 1

Introduction

Caveat

For the author, this introduction is both the first and last bit of substantial writing for quite a stretch of time that does not need to pass anonymous journal or grant review. As such, the author's long repressed personal writing idiom, now unchecked and having finally found outlet, has proceeded to run rampant through the text. The reader may well find it excessively pedantic, self-indulgent, and possibly even inflammatory. They may also find it repetitive, verbose, and possibly even redundant. I offer my sincere apologies (to my committee in particular), and ask for the reader for their patience in getting to the later chapters, where much greater restraint was exercised.

Humans have generally found it is most intelligent to design devices and systems that behave predictably and reproducibly. As a mundane example, consider a bike pump. Simple action on one end (pumping) produces simple reaction on the other end (air-flow). This is what a bicycle pump doesÑnight or day, Tuesday or Sunday, summer or winter. When a bicycle pump ceases to do these things, it is no longer a bicycle pump. It is a broken bicycle pump.

What relevance does a metaphoric bicycle pump have to ecology? It should have

very little. Unlike a bike pump, the natural world is full of complexity. But what does "complexity" mean in this context? We might consider a watch complex. It has many small, intricate parts. However, like a bicycle pump, the watch does the same thing day after day; simple, unchanging cycle upon cycle. This is not the complexity that defines the natural world.

We might instead consider an automated automobile assembly line complex. It has many sophisticated parts- robotic arms with six degrees of freedom that grasp, weld, and drill. However, the whole system involves a linear chain of causative events that takes the same set of inputs (stamped sheet metal, glass panes, molded plastic bumpers, nuts, bolts, paint, etc. etc.) and produces a fixed product (a Kia Sorrento). This is also not the complexity of the natural world.

Rather, natural systems typically have many components that act with mutual interdependence. This then gives rise to a range of complex, dynamic behaviors including thresholds, feedbacks, vicious cycles, regime shifts, and catastrophic change[1]. Central to all this is state-dependence. When system components, e.g. two species, have a state-dependent interaction, the magnitude and even direction of the effect (whether it's small or large, positive or negative) can depend on the state of these variables or other interdependent variables (e.g. climate). State dependence has fundamental consequences for the way natural systems can be studied and managed, because systems with state-dependence cannot be understood solely as the sum of their pairwise parts.

1.1 State dependence in principle

State-dependence in nature is well known in principal from empirical studies and from theory. Indeed, empirical examples of state dependence can be found almost

[1] Recognizing here, of course, that some of these terms have murky, overlapping definitions.

anywhere. A common case of state-dependence is when the presence of a third species changes the behavior of two competitors (1, 2). For example, two co-occurring species of tadpole decrease foraging in the presence of a predator, and this leads to less intense competition between them (3).

This is just one very particular case. Work in semi-arid Peru found that predation and competition among small mammals and rises and falls with inter-annual variations in rainfall (4). In Atlantic-coast salt marshes, variations in vegetation structure fundamentally determining the impact of predators on insect herbivores (5). More broadly, fish populations at low abundance show greater sensitivity to physical variability than at high abundance (6, 7).

On the theoretical side, there are a host of general mechanisms that give rise to state-dependent interactions- prey switching, adaptive foraging behavior (3), and even alternating limiting resources. Perhaps most fundamental, nonlinear functional responses readily give rise to state-dependent interactions. For example, in the very basic case of two competitors with saturating feeding responses(8), competition between the two will be state-dependent. At high food limitation, competition will be intense, but if the relative prey abundance is large enough that consumers are food saturated, consumer populations will not be actively competing with each other.

1.2 The Linear Expedient

The ecologists quantitative toolbox is chocked full of methods that were inherited from the linear world of engineering and so are poorly suited to nonlinear, state-dependent systems. Methods like PCA and ANOVA implicitly assume that system of study is separable and thus contradict the fundamental idea of state-dependence. Does this make sense in ecology?

The common line of reasoning (when explicitly made) is to assume the system is in equilibrium. This is critical, because systems that have nonlinear functional coupling between components can still be treated with linear methods if they remain at or near equilibrium over time (9). Thus, one can acknowledge the (ubiquitously accepted) presence of nonlinear functional responses while still applying methods that do not address state-dependence.

Facing ecosystem-based management, the equilibrium assumption looms large. To my mind, there has been no convincing argument e.g. based on evolution or principles of community assembly that we should a priori assume that systems are in equilibrium. There is a sense by many that since equilibrium systems are easier to treat quantitatively, that equilibrium behavior is somehow more simple, more parsimonious than non-equilibrium behavior. By Occam's notoriously double-edged razor, the burden of proof then falls on demonstrating that a system is not in equilibrium. This is false reasoning!

Moreoever, evidence of nonlinear, non-equilibrium dynamics can be readily found when tested properly. To say "properly" is an important caveat. Chaos first came on the ecological scene through analysis of simple population models like the logistic equation, ala (10). To address the implications of these simple models to understanding ecology in the real world, May noted that the range of demographic parameters necessary for chaos in these models were in fact feasible for some species (e.g. insects). However, this led to later misunderstanding in subsequent works (e.g. (11)) that measuring single species demographic parameters could be a test of chaos or nonlinearity in the intrinsic dynamics of natural populations. This in turn led to conclusions that the growth rates and demographics of many (e.g. non-insect) populations dismissed the possibility of chaos in their dynamics. This overly interpreted exercise was in spite of the warning from May in the original paper that "there are no single species populations in the natural world" and

"replacing a population's interactions ... by passive parameters may do great violence to reality".

In fact, more practical and empirical tests of nonlinearity and instability in dynamics have arisen in the intervening decades. Most notable of these is S-maps, which explicitly compares the predictive skill of equivalent linear and nonlinear models of empirical time-series dynamics (12). Analysis across a wide range of systems and populations has shown time (13) and time (14) again that marine populations, particularly in the face of human exploitation (6), show nonlinear, non-equilibrium, state-dependent dynamics. Straw men and hard evidence notwithstanding, there is a very fundamental reason to be exceedingly wary of equilibrium assumptions in marine environments. The ocean is not in equilibrium. It is dynamic. Highly so. Even if populations are simply chasing the equilibrium under ever-changing environmental conditions (although evidence speaks to the contrary (13)) , state dependence is still unavoidable. Thus, even to call equilibrium dynamics an "assumption" may be too generous; in highly dynamic environments like the coastal marine realm, it might be more fairly called a "myth".

1.3 The challenges of Ecosystem-Based Management

Traditional "command and control" fisheries management sidesteps the well known complex, interdependence of ecosystems by treating fish stocks as if they were, well, bicycle pumps. The basic framework (written into Federal law!) relies on the principle of maximum sustainable yield, which involves a simple calculus between the reproductive potential of a stock and the mortality from fishing. The management target is the highest rate of fishing possible while still maintaining *fish out = fish in*.

Ecosystem-based management seeks to acknowledge that fish and fishing do not exist in a vacuum, but occur in the context of a complex, interdependent system. Fishing

on a large predator that might be sustainable if it occurred in isolation may quickly deplete the stock when fishing pressure is also applied to its prey. Fishing that may be sustainable for one stretch of time may be disastrous if continued when the environment shifts to a less favorable regime.

State dependence is central to these ecosystem behaviors. If you acknowledge the ecosystem, you must allow for state-dependence. Yet despite the long list of laboratory, field, and theoretical demonstrations of state-dependence in principle, acknowledging and addressing state-dependence in practice has been slow. While the misleading equilibrium assumption looms large, the biggest barrier to effectively treating state-dependence in ecosystem management may well be simply a lack of practical methodology. Major problems can arise in applying both traditional statistics and parametric model-based approaches to address nonlinear, state-dependent systems.

1.3.1 Correlation

Nevertheless, statistical methods based on a linear framework are still commonly applied to ecosystems, despite the potential mismatch between the implicit assumption of the methods and reality. Correlation is a great and foundational example. Despite the well-rehearsed mantra "correlation does not imply causation" that we have inherited from Bishop Berkeley, correlation and correlation-based approaches like principle components analysis still remain a powerful paradigm for studying interactions in ecosystems. And it can certainly bear fruit. For example, the Pacific Decadal Oscillation shows a strong correlation to zooplankton populations like the krill *Nyctiphanes simplex* in the Southern California Bight (15).

In other instances, however, correlation has produced little insight or outright confusion. Correlations between populations and the environment can be hard to find in marine ecosystems even where interaction is suspected (16), and even when they

are identified, they frequently disappear when retested later in time (17). This reflects the simple fact that systems with state-dependent dynamics are apt to produce mirage correlations. That is, the magnitude and even sign of interactions is liable to change through time (maybe even rapidly!), and so correlations can be positive, negative, or in-significant depending on when you look at the system. In this way, not only does "correlation not imply causation", but equally important *lack* of correlation does not imply *lack* of causation.

Of course, these are all problems that can happen in nonlinear systems. This also means that there are circumstances when correlation can potentially work. For example, synchronization between populations and the environment has been observed in numerous cases (18). This occurs when an external driver has a very strong influence, so that the dynamics of the response system (e.g. a population) become entrained to the driver. In this situation, the internal dynamics of the response no longer matter, and lead to strong, persistent correlations. Thus, correlation and PCA can be well suited e.g. to identify patterns due to strong environmental forcing. Witness that correlations between fish species and the environment appear most robust along the edge of species ranges (17), and indeed that in the previously mentioned case of *Nyctiphanes simplex*, the Southern California Bight is on the edge of its range. However, it is important to keep in mind that applying these methods will only identify a limited subset of possible ecological relationships.

1.3.2 Problems modeling systems with state-dependent interactions

The obvious alternative to traditional statistics is parametric model-based approaches. There are a lot of reasons to like models. Models let you predict. They let you understand interactions between two or more species, e.g. by taking partial derivatives. You can also explore "what if" scenarios- e.g. what happens to species x if change fishing

pressure on species y. Particularly appealing for management, they can let you explicitly evaluate trade-offs (e.g. like Ecosim). And most relevant to the discussion at hand, it is certainly possible to incorporate state-dependent behavior into ecological models.

Even so, using models to study ecosystems with nonlinear, state-dependent dynamics has some major roadblocks. Each particular model represents very specific hypotheses about the way different variables interact and the choices, e.g. of functional form, are often arbitrary. Importantly, even small differences in the functional form (e.g. of species interactions) can lead to dramatically different conclusions about management (19). Even if the correct model structure is somehow known a priori, fitting the coefficients of even simple population models (e.g. stock-recruitment relationships) can be unreliable and grossly misleading (20, 21).

1.4 EDM as a solution

In contrast to these conventional approaches that have explicit or implicit assumptions (e.g. stable equilibrium dynamics, linearly separable cause-and-effect relationships, simple parametric equations, etc.), Empirical Dynamic Modeling (EDM) relies on extracting system behavior directly from observed time series (12, 22-24). The essential idea of EDM is to view a dynamic system from a geometrical perspective. At any point in time the ecosystem occupies a single point in some n-dimensional state-space (e.g., Hutchinson's n-dimensional niche), where each axis is a state variable like species abundance or resource concentration. As the system changes over time, it occupies different points in the state-space. This forms trajectories in the n-dimensional space that comprise the geometric attractor (Figure 2). Time series of the state variables can then be viewed as sequential projections of the attractor onto the coordinate axes. Just like model equation, a dynamic system can also be completely described by these trajectories that it creates on

its attractor manifold in state space. Moreover, the dynamics can be reconstructed from the time-series simply by re-plotting them in multi-dimensional Cartesian space rather than separately in the time-domain. In practice, all of the often important ecosystem variables are not all measured together. This is where the embedding theorem of Takens becomes important.

Takens' Theorem states that instead of requiring observations of all n state variables to construct an attractor from data, we can substitute lags of a single time series for the unknown or unobserved variables. This allows the dynamics of a complex system to be recovered from just a single time series.

The most immediate way to practically implement these ideas is with forecasting. This single-variable forecasting with EDM can be an end unto itself (25), but can also validate hypotheses about nonlinear dynamics (12) or environmental forcing (13, 23). Recent work has specifically highlighted the power of forecasting with EDM (20, 21) over parametric-based forecasting. In the most generous circumstances, e.g. when the system equations are assumed to be exactly known, the empirical forecasts of EDM are equivalent to the performance of other methods. In many other cases, EDM does substantially better. One of the main objections to these methods based on Takens' theorem has been that they are overly phenomenological. Indeed, univariate EDM on its own does not address many of the needs of ecosystem-based management, such as identifying and characterizing interactions or exploring different scenarios of environment or management. Work by Dixon said. (23) showed how including multiple time-series in the reconstruction can begin to address these criticisms.

In the damsel fish *Pomacentrus amboinensis*, spawning is known to coincide with the lunar cycle (like many other reef fish). While the lunar cycle is simple and linear, univariate EDM analysis shows that the larval supply is strongly nonlinear. But by including additional factors in multivariate empirical dynamic models (EDM), Dixon

et al. were able to go a step further. They show not only that there are wind and tidal conditions for larval supply, but that these conditions must be met simultaneously to create large larval supply.

In Chapter 2 of this dissertation, I revisit the idea of multivariate EDM explored heuristically by Dixon et al. I show that multivariate EDM indeed on firm mathematical group by explicitly expanding the mathematical proof of Takens's theorem to include multivariate time series data. In the remaining chapters, I build upon this foundation to expand the multivariate EDM approach and demonstrate its emerging role for ecosystem-based management.

1.4.1 Sardine: the most basic problem of physical-biological coupling

One aspect of ecosystem-based management referred to in the Magnusson-Stevens Reauthorization Act is "to account for effects of environmental variation on fish stocks and fisheries." Even this one facet presents scientific questions both theoretical and applied in nature. On a basic level, we need to identify predict the effect that different environmental scenarios will have on population abundance and resilience. Critically, even in this seemingly straightforward question of ecosystem-based management strains conventional statistical and modeling approaches. Intuitively, it is clear that the effect of the environment has to be nonlinear- when the environment goes far enough to either extreme, it is bad for the species. But more importantly, there is comprehensive evidence that the effect of the environment can be state-dependent (6, 7).

Debate over the management of Pacific sardine is a perfect example. Jacobson and MacCall (26) found statistical correlation between log reproductive success (one way of quantifying recruitment) and a three-year average of the Scripps Institution of Oceanography (SIO) pier temperature (SST). They used a general additive model to develop a management strategy for sardine that incorporated the influence of sea

temperature. Based on this finding, the Pacific Fishery Management Council modified the sardine management plan to afford extra protection when the SIO pier temperature is unfavorable (27).

McClatchie et al. (28) repeated the correlation analysis of Jacobson and MacCall with an additional 17 years of new data. Surprisingly they found that the statistical relationship between recruitment and SST is no longer significant when more recent data are included in the analysis. McClatchie et al. argue that the SIO pier SST is a poor variable to use for sardine management since it is a broad-scale, synoptic measure of conditions in the California Current (29) that has no direct mechanistic link to sardine population dynamics. They conclude that a more mechanistically motivated environmental index must be developed that incorporates physical variables, larval predators, and prey. This temporarily led to the suspension of the environmental control rule. However, subsequent findings by Lindegren and Checkley (30) that used generalized additive models rather than correlation led to conclusion that there is still a quantifiable effect of sea surface temperature on Sardine in the recent decades, and led to revitalized discussions of a temperature control rule for Pacific sardine.

As discussed above, using correlation as a measure of physical-biological interaction might be unwise, since it ignores the potential for state-dependent dynamics. So despite the results of McClatchie et al., environmental forces reflected in large-scale, synoptic variables like the SIO pier SST may in fact influence the Pacific sardine, but in a state-dependent way. However, model based approaches also suffer difficulties. Models represent very specific hypotheses. The so-called “environmental Ricker”, which motivates the GAM analyses both of Jacobson & MacCall and of Lindegren & Checkeley, is a great example. The traditional Ricker model describes the relationship between recruitment and spawning stock biomass.

$$R = S\exp\left(r(1-S/k)\right)$$

From the classic work by Hjort (31), it has been understood for nearly a century that recruitment is influenced by more than just stock size or number of eggs spawned. Environmental conditions that affect larval retention or food supply during the critical feeding period. To test hypotheses about environmental forcing, it has been common practice to incorporate the environment (e.g. temperature, T) into the Ricker model as follows:

$$R = S\exp\left(r(1-S/k) + \psi T\right)$$

This form has the convenience that it appears log-linear. That is, if we rewrite the equations for the logarithm of recruitment, we have

$$\ln(R) = [\ln(S) + r(1-S/k)] + \psi T.$$

This form has the convenience that the effects of S and T are additive, and so easily separable using GAM analysis. However, just because it is convenient doesn't mean it is correct. There are other perfectly reasonable ways to think of the environment entering into the Ricker calculus. For example, growth can be a function of temperature,

$$R_t = S_{t-D}\exp\left(r(T_{t-D})(1-S_{t-D}/k)\right)$$

or the carrying capacity

$$R_t = S_{t-D}\exp\left(r(1-S_{t-D}/k(T_{t-D}))\right).$$

Neither of these rather plausible mechanisms leads to separable effects amenable

to the assumptions of GAMs.

This is not just talk and hand waving. While the analysis presented by Lindegren and Checkeley appears to suggest that the GAM approach is robust across time (unlike correlation), the different choices of L&C versus the original J&M in fact lead to different conclusions over different periods of time (Table 1.1 below). If the original SIO pier SST is used, the conventional F-test concludes there is no significant effect over the recent period (1981-2010) studied by Lindegren and Checkeley, or the complete span studied by McClatchie et al. (1935-63, 1981-2010).

In Chapter 3 of this dissertation, I show how re-examining environmental effects on Pacific sardine using Empirical Dynamic Modeling tools suited to systems with state dependence leads to a much clearer picture. We find that the original SIO pier SST does have a significant effect on Pacific sardine, but that the effect is state-dependent. Moreover, we show how multivariate EDM can be used to explore the effects of different plausible environmental scenarios on sardine dynamics.

1.4.2 Interactions between trophically similar species

There are plenty more complicated questions to consider for ecosystem-based management, particularly in trying to discern the multispecies effects of fishing. *In principle*, there are a number of possible relationships that can exist between trophically similar species (e.g. two planktivorous fish). The most well known, of course, is that two species that share prey may have mutual indirect negative effects on each other due to competition. Coupling by common predators can also lead to indirect effects, but depend on the functional forms (e.g. the feeding response of the predator) (32). Some functional forms can lead to apparent mutualism (mutual net positive effect between the two prey), while others lead to apparent competition (mutual net negative effect). If prey preference is introduced, the indirect effects can even be positive in one direct, but negative in the

other. More importantly, in a system with state-dependent dynamics, the strength of these various effects can wax and wane, meaning that the magnitude and sign of interactions are apt to shift e.g. as the system transitions from bottom-up dominated to top-down.

Given that all these are possible in principle, how are we to ascertain which in fact is relevant to a given ecosystem in practice? This question is difficult to answer with linear statistics, as the interactions change in time, but are also difficult to answer with models, as the conclusions depend greatly are difficult to evaluate choices of functional forms and correct parameterization. Chapter 4, shows how EDM gives a straightforward approach to answering these questions, and gives an equation-free method for tracking and predicting species interactions for ecosystem study and management.

1.4.3 Related applied problems: the seasonality of influenza

Some of the same questions arising in EBM also arise in epidemiologyÑnamely, how we understand the effect of the environment on ecology. It is universally appreciated by residents of temperate countries like the U.S. that influenza is seasonal. You get the flu in the winter. However, the environmental causes (and hence possibility for productive public health policy) have continued to remain in debate. Part of this reason is that there are many different potential mechanisms that coincide with winter in temperate countries.

In the winter, air temperature is colder. When air temperature is colder, people spend more time indoors, and hence have increased contact rates. Winter also coincides with weaker solar irradiance, which suppressed the production of Vitamin D, and could in theory lead to decreased immune response. But correlation, as we know, does not imply causation. Recent lab analysis has suggested another route, which is that dry air in the winter increases the transmission rate of the virus due to longer residence time of droplets in the air (33) or greater survival when airborne (34). As with our discussions of ecosystem interactions above, there are multiple plausible mechanisms here that are

valid *in principle*, and hence the cause cannot be determined without examining the real system.

Human population scale analyses have focused on identifying correlative relationships. Hence, they have not addressed state-dependence in the environmental drivers of influenza and have not satisfactorily resolved the conundrum. In temperate countries, influenza outbreaks strongly correlate to the seasonal changes in temperature and absolute humidity, and so correlation does not distinguish alternate hypotheses. In tropical countries, annual climate cycles are much weaker and influenza seasonality is much harder to find. Naive interpretation (neglecting the nonlinear reality of epidemics) suggests that there must be other factors at work in the tropics.

In 5 chapter of this dissertation, I use multivariate EDM methods to directly examine global drivers of influenza outbreaks from country-level time series. By identifying causal drivers rather than correlations, I show that despite the apparent differences between temperate and tropical countries, absolute humidity and to a lesser extent temperature drive influenza outbreaks globally. Using multivariate EDM, I also corroborate a U-shaped relationship between absolute humidity and influenza at the population level that has been suggested in principle by experiment.

1.5 Summary

Together, these chapters address develop multivariate EDM as a practical, empirical tool for studying and managing ecosystems that directly address the complex reality of nature. There remain many potential ways to refine and improve EDM methods. The weighting schemes in simplex projection or S-maps can be modified to explicitly considering non-Gaussian sampling error. Quasi-parametric forms can be included in generalizations of S-maps to incorporate qualitative knowledge about constraints on

species dynamics. Nevertheless, the work herein demonstrates that even simple implementations of multivariate EDM can lead to powerful and novel insights by removing the disparity between the methods we use and the fundamental reality of state dependence in nature.

Table 1.1: Results of sequential F-test on Pacific sardine to detect effect of SIO SST. Over an initial 30-year period, there appears to be a highly significant effect of temperature on (log recruitment). However, when examining a more recent period of time or the entire span, the effect disappears. Using 3rd order thin-plate smoothing instead of LOESS, as in (30), F-test finds no significant temperature effect over any of these time periods.

	Resid. Df	Resid. Dev	Df	Deviance	F	Pr(>F)
1935-63, 1981-90 (Data Range of Jacobson & MacCall)						
$\mathbf{m_1}$	29.138	26.164				
$\mathbf{m_2}$	28.138	21.338	1	4.8256	6.3633	0.012
1981-2010 (Data Range of Lindegren & Checkley)						
$\mathbf{m_1}$	23.43	15.055				
$\mathbf{m_2}$	22.43	14.984	1	0.0706	0.1056	0.75
1935-63, 1981-2010 (Data Range of McClatchie et al.)						
$\mathbf{m_1}$	46.641	34.516				
$\mathbf{m_2}$	45.641	34.194	1	0.3219	0.4296	0.52

$\mathbf{m_1}$	$\ln R \sim 1 + lo(SSB_{t-2})$
$\mathbf{m_2}$	$\ln R \sim 1 + lo(SSB_{t-2}) + SST_{t-2}$

1.6 References

1. Wissinger S, McGrady J (1993) Intraguild Predation and Competition Between Larval Dragonflies: Direct and Indirect Effects on Shared Prey. Ecology 74:207-218.

2. Trussell GC, Ewanchuk PJ, Bertness MD (2002) Field evidence of trait-mediated indirect interactions in a rocky intertidal food web. Ecol Lett 5:241-245.

3. Werner EE (1992) Individual behavior and higher-order species interactions. Am Nat 140:S5-S32.

4. 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.

5. Gratton C, Denno RF Seasonal shift from bottom-up to top-down impact in phytophagous insect populations. Oecologia 134:487-495.

6. 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.

7. Brander KM (2005) Cod recruitment is strongly affected by climate when stock biomass is low. ICES J Mar Sci 62:339-343.

8. Abrams PA (1980) Consumer functional response and competition in consumer-resource systems. Theor Popul Biol 17:80-102.

9. MacArthur R (1970) Species packing and competitive equilibrium for many species. Theor Popul Biol 1:1-11.

10. May RM (1976) Simple mathematical models with very complicated dynamics. Nature 261:459-467.

11. Shelton AO, Mangel M (2011) Fluctuations of fish populations and the magnifying effects of fishing. Proc Natl Acad Sci USA 108:7075-7080.

12. Sugihara G (1994) Nonlinear forecasting for the classification of natural time series. Philos T Roy Soc A 348:477-495.

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

14. Glaser SM, Fogarty MJ, Liu H, Altman I, Hsieh C-H, Kaufman L, MacCall AD, Rosenberg AA, Ye H, and Sugihara G (2013) Complex dynamics may limit prediction in marine fisheries. Fish and Fisheries.

15. Brinton E, Townsend A (2003) Decadal variability in abundances of the dominant euphausiid species in southern sectors of the California Current. CalCOFI: A Half Century of Physical, Chemical and Biological Research in the California Current System 50:2449-2472.

16. Hsieh C-H, Reiss CS, Watson W, Allen MJ, Hunter JR, Lea RN, Rosenblatt RH, (2005) A comparison of long-term trends and variability in populations of larvae of exploited and unexploited fishes in the Southern California region: a community approach. Progress In Oceanography 67:160-185.

17. Myers RA (1998) When do environment-recruitment correlations work? Reviews in Fish Biology and Fisheries 8:285-305.

18. Liebhold A, Koenig WD, Bjornstad ON (2004) Spatial Synchrony in Population Dynamics*. Annu Rev Ecol Evol Syst 35:467-490.

19. Wood SN, Thomas MB (1999) Super-sensitivity to structure in biological models. Proceedings of the Royal Society B: Biological Sciences 266:565-570.

20. Perretti CT, Sugihara G, Munch SB (2012) Nonparametric forecasting outperforms parametric methods for a simulated multispecies system. Ecology 94:794-800.

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

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

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

24. Sugihara G, May R, Ye H, Hsieh CH, Deyle E, Fogarty M, Munch S (2012) Detecting causality in complex ecosystems. Science 338:496-500.

25. Glaser SM, Ye H, Sugihara G (2013) A nonlinear, low data requirement model for producing spatially explicit fishery forecasts. Fisheries Oceanography 23:45-53.

26. Jacobson LD, MacCall AD (1995) Stock-recruitment models for Pacific sardine (Sardinops sagax). Can J Fish Aquat Sci 52:566-577.

27. Hill KT, Yaremko M, Jacobson LD, Lo NCH, Hanan DA (1998) Stock assessment and management recommendations for Pacific sardine (Sardinops sagax) in 1997.

28. McClatchie S, Goericke R, Auad G, Hill K (2010) Re-assessment of the stock-recruit and temperature-recruit relationships for Pacific sardine (Sardinops sagax). Can J Fish Aquat Sci 67:1782-1790.

29. McGowan JA, Cayan D, Dorman L (1998) Climate-Ocean Variability and Ecosystem Response in the Northeast Pacific. Science 281:210-216.

30. Lindegren M, Checkley DM Jr (2013) Temperature dependence of Pacific sardine (Sardinops sagax) recruitment in the California Current Ecosystem revisited and revised. Can J Fish Aquat Sci 70:245-252.

31. Hjort J (1926) Fluctuations in the year classes of important food fishes. Journal du Conseil 1:5-38.

32. Abrams PA, Holt RD, Roth JD (1998) Apparent competition or apparent mutualism? shared predation when populations cycle. Ecology 79:201-212.

33. Lowen AC, Mubareka S, Steel J, Palese P (2007) Influenza virus transmission is dependent on relative humidity and temperature. PLoS Pathogens 3:1470-1476.

34. Tellier R (2009) Aerosol transmission of influenza A virus: a review of new studies. J R Soc Interface 6 Suppl 6:S783-90.

Chapter 2

Generalized theorems for nonlinear state space reconstruction

Abstract

Takens' theorem (1981) shows how lagged variables of a single time series can be used as proxy variables to reconstruct an attractor for an underlying dynamic process. State space reconstruction (SSR) from single time series has been a powerful approach for the analysis of the complex, non-linear systems that appear ubiquitous in the natural and human world. The main shortcoming of these methods is the phenomenological nature of attractor reconstructions. Moreover, applied studies show that these single time series reconstructions can often be improved *ad hoc* by including multiple dynamically coupled time series in the reconstructions, to provide a more mechanistic model. Here we provide three analytical proofs that add to the growing literature to generalize Takens' work and that demonstrate how multiple time series can be used in attractor reconstructions. These expanded results (Takens' theorem is a special case) apply to a wide variety of natural systems having parallel time series observations for variables believed to be related to

the same dynamic manifold. The potential information leverage provided by multiple embeddings created from different combinations of variables (and their lags) can pave the way for new applied techniques to exploit the time-limited, but parallel observations of natural systems, such as coupled ecological systems, geophysical systems, and financial systems. This paper aims to justify and help open this potential growth area for SSR applications in the natural sciences.

2.1 Introduction

A growing realization in many natural sciences is that simple idealized notions of linearly decomposable, fixed equilibrium systems often do not accord with reality. Rather, empirical measurements on ecosystems, metabolic systems, financial networks, and the like suggest a more complex, but potentially more information-rich paradigm at work [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. Despite a long history of linear methods development in the engineering sciences, natural systems are generally not well described as sums of independent frequencies that can be sensibly decomposed, analyzed as non-interacting, and reassembled (e.g. Fourier or spectral analysis) in the style of traditional reductionism [15, 16]. Rather, quantitative measurements show many systems to be fundamentally non-equilibrium and unstable, in a manner more consistent with nonlinear (state dependent) dynamics occurring on a strange attractor manifold M, where relationships between state variables cannot be studied independently of the overall system state [17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27]. This emergent comprehensive view may help explain why many natural systems, such a those mentioned above, appear so difficult to understand and predict. Mirage correlations are commonplace in nonlinear systems where the manifold may contain trajectories that can temporarily exhibit positive correlations between variables for surprisingly long time periods (and in some regions

of the state space) and can subsequently and rapidly exhibit negative correlations or no relationship in other time periods (and other regions of M). This transient property of apparent non-stationarity in correlations is one of the confounding phenomena faced by traditional linear models that require continual refitting and exhibit little or no predictive power.

In this paper, we present two general theorems that addresses the problem of characterizing the coupled dynamics of nonlinear systems using time series observations on a manifold M. A special case of this theorem, attributed originally to Takens [12], provided the first sketch of a mathematical proof for reconstructing a diffeomorphic shadow manifold M' using lags of a single time series as coordinate axes. The basic idea, that was earlier demonstrated by Packard, Crutchfield, Farmer, and Shaw [28] and Crutchfield [2], is that under generic conditions, a shadow manifold M' can be created using time-lagged observations of M based on a single observation function (Cartesian coordinate variable) that is a smooth and smoothly invertible $1:1$ mapping with M. Subsequently, Sauer, Yorke, and Casdagli [29] provided a definitive proof and an explicit extension of Takens' theorem to fractal sets; their theorems are also more powerful than the original theorem, as they show embeddings are not just generic in the sense of being open and dense in the set of all mappings, but in fact almost every mapping in the sense of prevalence [30] is an embedding (see [30] for an in-depth explanation of the advantages of "prevalence" over "generic"). The theorem was also extended by Stark, Broomhead, Davies, and Huke [31, 32] and Stark [33] to the certain classes of stochastic systems. Practical methods for reconstruction have also been explored, particularly to address the presence of noise in real data (e.g. [29, 34]). Casdagli et al. [35] give a thorough treatment of such techniques based on transformations of univariate maps, showing how optimal noise reduction can be achieved. These very important prior results all focused on reconstruction from a single time series; however, as proven below, they can be extended

to the more practically significant case where multiple observation functions are used to generate M'.

Here we prove the more general case of multivariate embeddings (embeddings using multiple time series and lags thereof), and show how time series information can be leveraged if multiple time series and their lags are used to construct embeddings of M'. These theorems pave the way for more extensive use of state space reconstruction methods in practical applications where long time series may not be available, so that multiple diffeomorphic embeddings may be created in factorial fashion to more fully exploit the coupled non-redundant information that can be extracted from multiple time series (multiple observation functions of dynamics on a manifold) to create predictive shadow manifolds [36]. The use of multiple time series allows the possibility of noise reduction that exceeds the limitations of univariate reconstructions in the presence of noise [35].

The possibility of extending Takens' theorem to allow lags of multiple observation functions was mentioned in Remark 2.9 from [29], but was not explicitly proven. The remark was also restricted to mappings strictly formed from consecutive lags, which is not the only possibility that needs to be considered in the multivariate case. Given the potential importance of multivariate reconstructions, we believe a full proof is required—in particular, one that extends the generalization to non-consecutive lags. We show how Takens' theorem is a special case of our more general Theorem 2 (below) and by following the structure of Takens' original proof we clarify the logic and highlight the restrictions and special cases (non-generic cases) that can arise in its application to real world systems. We then give explicit proof of a stronger version of Remark 2.9 from Sauer et al. [29] that allows non-consecutive lags. This third theorem is stronger than the first two in the sense that it shows embeddings are prevalent and not just generic. For those less familiar, we begin with a brief overview of some basic terms and concepts

used in our proofs.

2.1.1 Some Basic Concepts of Embedding Theory

Consider the classic Lorenz attractor [37] shown in Figure 2.1a, consisting of trajectories in three-dimensional space that together define a butterfly shaped surface or *manifold*. For simplicity, a manifold can be thought of as a generalized, n-dimensional surface embedded in some higher dimensional space, where the dimension of the manifold may be fractal (as is the case for the Lorenz attractor). More generally, an *embedding* is a multivariate transformation of a manifold that resolves all trajectories on the original manifold without crossings. That is, an embedding is globally $1:1$ in that it resolves all *singularities* in trajectories that define the manifold (singularities are points on the manifold where trajectories cross so that future paths are not uniquely determined).

An *immersion* is a local embedding that may not preserve the global topology of a manifold. Rather an immersion preserves the topology of every local neighborhood of the original manifold, so that each point of the tangent space of the immersed manifold has the same dimensionality as the true manifold. Thus, an immersion is a mapping that is $1:1$ between any given "piece" of the true manifold and the immersed manifold. However, this condition does not guarantee that the global topology is preserved. This is illustrated in Figure 2.1c, where two different pieces of the original manifold are mapped to the same piece of the immersed manifold, producing an immersion that is not an embedding. Immersions are nonetheless a useful conceptual stepping stone for constructing proofs about embeddings, since all embeddings are necessarily immersions.

The Lorenz attractor, Figure 2.1a, provides an excellent example to illustrate both of these concepts. Consider two different multivariate functions that transform the original manifold, $\Phi_y = (y(t), y(t-\tau), y(t-2\tau))$ and $\Phi_z = (z(t), z(t-\tau), z(t-2\tau))$ where τ is a small time lag as in Takens' theorem. Both of these functions map points

on the true manifold to points on a shadow manifold, shown in Figures 2.1b and 2.1c. Examining these shadow manifolds, it is evident that both are immersions of the Lorenz attractor, because zooming in on a particular piece of either will reveal that the tangent spaces have the same dimensionality as the original. However, only Figure 2.1b is an embedding that successfully reproduces the two lobes of the butterfly. The reconstruction in Figure 2.1c, based on lags of the z-coordinate, fails to do so, because the two fixed points of the original attractor have the same z-coordinate; they are mapped to the same point on the shadow manifold, so the two lobes are stacked on top of each other. This singularity is a consequence of a special, non-generic symmetry in the Lorenz system that violates an assumption of Takens' theorem. Figure 2.1d shows an embedding based on lags of both y- and z-coordinates and is an example of the generalized mappings addressed in this paper.

2.2 Results

2.2.1 Two Theorems in the Style of Takens: The Generic Case

Let M be a compact manifold of dimension m. A dynamical system is a diffeophorism ϕ defining the trajectories or "flow" on M for discrete time or a vector field X on M for continuous time. Takens [12] proved generically that given ϕ and M, a smooth observation function $y : M \to \mathbb{R}$ can be used to construct an embedding of M in $2m+1$ dimensions under the transformation $\Phi_{(\phi,y)} : M \to \mathbb{R}^{2m+1}$ where $\Phi_{(\phi,y)}(\mathbf{x}) = \langle y(\mathbf{x}), y(\phi(\mathbf{x})), y(\phi^2(\mathbf{x})), \ldots, y(\phi^{2m}(\mathbf{x}))\rangle$. Here the components

$$\langle y(\mathbf{x}), y(\phi(\mathbf{x})), y(\phi^2(\mathbf{x})), \ldots, y(\phi^{2m}(\mathbf{x}))\rangle$$

correspond to time-lagged observations of the dynamics on M defined by ϕ.

Notice that such mappings involve a single distinct observation function (i.e. a single time series), and represent a small subset in the larger set $\mathscr{Y}^{2m+1}$of all possible mappings $M \to \mathbb{R}^{2m+1}$ that could, for example, involve multiple time series and their lags.

Takens explicitly refers only to the unlagged y as an observation function, but in its most general sense an observation function is any $y : M \to \mathbb{R}$. Thus, the functions $y(\phi(\mathbf{x})), y(\phi^2(\mathbf{x})), \ldots$, corresponding to the lags of the time series are technically observation functions as well. This bears mention, because in the more general case of mappings $\Phi : M \to \mathbb{R}^{2m+1}$, the observation functions making up the components of Φ are not all derived from a single time series, but can be various lags of multiple time-series. To treat these cases, it is necessary to acknowledge that these are all observation functions, and we will refer to distinct time series as "unlagged" observation functions.

For a mapping Φ in the larger set $\mathscr{Y}^{2m+1}$of all mappings $M \to \mathbb{R}^{2m+1}$, consider the case with $2m+1$ component functions $y_k : M \to \mathbb{R}$ which are multiple unlagged observation functions of M (i.e. multiple time series). Again, an observation function is any function $M \to \mathbb{R}$ that assigns a real number to each point on the manifold M. For a mapping $\Phi : M \to \mathbb{R}^{2m+1}$, we can think of Φ in terms of its $2m+1$ component functions, which correspond to the coordinates in $\mathbb{R}^{2m+1}$. These component functions may all be lags of a single distinct observation function tracking a dynamical system, as in the case of Takens, or they may be multiple observation functions, as in the case of Whitney, or they may be lags of multiple observation functions, as in Theorems 2 and 7 below.

The question arises whether general multivariate mappings

$$\Phi(\mathbf{x}) = (y_1(\mathbf{x}), y_2(\mathbf{x}), \ldots, y_{2m+1}(\mathbf{x}))$$

form legitimate embeddings. Here we present two theorems: one that demonstrates that maps created from $2m+1$ distinct observation functions are generically embeddings and another that shows that maps created from lags of multiple observation

functions are also generically embeddings. Both of these theorems generalize Takens' theorem for which the component functions only involve a single observation function.

It follows from Whitney [38] that generically $\Phi \in \mathscr{Y}^{2m+1}$ is an embedding. Note, however, that Whitney's work does not apply to the specific subsets of $\mathscr{Y}^{2m+1}$ involving fixed lagged relationships as discussed by Takens for reconstructing attractor manifolds M for dynamic systems. That is, Whitney's theorem is generic and does not address these specific subsets of $\mathscr{Y}^{2m+1}$ which have "measure zero" (e.g. in the sense of "shy" defined in [30]). To tackle this problem, we look to the proof of Takens and see that it can be readily generalized to the other subsets of $\mathscr{Y}^{2m+1}$, including the case of generic $\Phi \in \mathscr{Y}^{2m+1}$.

Recall that, for a compact manifold, a mapping that is an immersion and injective is also necessarily an embedding. Thus, Takens' general approach was to first show that (i) immersions are dense in the set of mappings $\{\Phi_{(\phi,y)}\}$, then that (ii) there is a dense set of $1:1$ mappings within this set of immersions. Since the set of embeddings is open in the set of all possible mappings, Takens concludes that mappings in $\{\Phi_{(\phi,y)}\}$ are generically embeddings. The critical word here is "generically," meaning there can be exceptions (and as explained in [30], the set of such exceptions doesn't necessarily have zero measure).

To demonstrate both (i) and (ii), Takens argues that even when the property of interest (e.g. the $1:1$ property) does not hold for some particular mapping, by making an arbitrarily small perturbation, it is possible to find a nearby mapping for which that property holds. The key to the theorem and also to adapting it to other sets of mappings is finding how to make these perturbations. The proof is most straightforward for the general case involving $2m+1$ distinct observation functions (each a distinct time series) because it is possible to perturb the component functions of $\Phi_{\langle y_k \rangle}$ independently. Thus we begin with this proof to add clarity to the more powerful main theorem 2 involving

lags of multiple observation functions.

Theorem 1. *Consider a compact, m-dimensional manifold M and a set of $2m+1$ observation functions $\langle y_1, \ldots, y_{2m+1} \rangle$, where $y_k : M \to \mathbb{R}$ smoothly; by "smooth" we mean at least $\mathbb{C}^2$. Then it is a generic property of all possible $\langle y_k \rangle$ that the mapping $\Phi_{\langle y_k \rangle} : M \to \mathbb{R}^{2m+1}$ defined as*

$$\Phi_{\langle y_k \rangle} = (y_1(\mathbf{x}), y_2(\mathbf{x}), \ldots, y_{2m+1}(\mathbf{x}))$$

is an embedding.

Proof. Consider an arbitrary set of $2m+1$ observation functions $\langle \bar{y}_k \rangle$ on M. We define a corresponding mapping $\Phi_{\langle \bar{y}_k \rangle} \in \mathscr{Y}^{2m+1}$ by letting each of these $2m+1$ observation functions be one of the component functions of $\Phi_{\langle \bar{y}_k \rangle}$. Now, recall that an immersion is a map with a derivative that is globally injective, i.e. $1:1$. We denote the total derivative of a function f as Df. If the derivative is evaluated at a particular point $\mathbf{x}$ in the domain of f, we will write $(Df)_{\mathbf{x}}$, and if Df is a matrix, then we denote the derivative at a particular point and along a particular tangent vector $\mathbf{v}$ as $(Df)_{\mathbf{x}}(\mathbf{v})$.

For any point $\mathbf{x} \in M$, we can perturb the co-vectors $(D\bar{y}_k)_{\mathbf{x}} \in \mathscr{T}^*(M)$ independently by perturbing individual $\bar{y}_k$. By making infinitesimal perturbations at points $\mathbf{x} \in M$ for which $\operatorname{rank}(D\Phi_{\langle \bar{y}_k \rangle})_{\mathbf{x}} < m$, we can get a set of observables $\langle \bar{\bar{y}}_k \rangle$ arbitrarily close to $\langle \bar{y}_k \rangle$ such that $\operatorname{rank}(D\Phi_{\langle \bar{\bar{y}}_k \rangle})_{\mathbf{x}} = m$ for all $\mathbf{x} \in M$—i.e., $\Phi_{\langle \bar{\bar{y}}_k \rangle}$ is an immersion. Since the set of immersions is open in the set of all mappings, there is a neighborhood $\mathscr{U} \subset \mathscr{Y}^{2m+1}$ around this $\Phi_{\langle \bar{\bar{y}}_k \rangle}$ such that every $\Phi_{\langle y_k \rangle} \in \mathscr{U}$ is an immersion.

Since immersions are local embeddings, we can find a $\delta > 0$ such that on the manifold, $0 < \rho(\mathbf{x}, \mathbf{x}') \leq \delta$ implies $\Phi_{\langle \bar{y}_k \rangle}(\mathbf{x}) \neq \Phi_{\langle \bar{y}_k \rangle}(\mathbf{x}')$. Here we depart from Takens' notation and let δ denote infinitesimal separations between two points on the manifold M to avoid confusion with the later defined ε which is used to perturb the observable; ρ

is any fixed metric on M. In fact for this fixed δ, there is a subset $\mathscr{U}' \subset \mathscr{U}$ such that for any $\langle y_k \rangle$ in $\mathscr{U}'$, the associated map $\Phi_{\langle y_k \rangle}$ is an immersion, and $\rho(\mathbf{x}, \mathbf{x}') \leq \delta$ implies that $\Phi_{\langle y_k \rangle}(\mathbf{x}) \neq \Phi_{\langle y_k \rangle}(\mathbf{x}')$.

Next, we show that we can find a globally $1:1$ $\Phi_{\langle y_k \rangle} \in \mathscr{U}'$ arbitrarily close to $\Phi_{\langle \bar{\bar{y}}_k \rangle}$. To do this, we construct a finite collection of subsets $\{U_i\}_{i=1}^N$ such that the U_i are open subsets of M, the collection covers M, and diameter$(U_i) < \delta$ for every i. Then, we take a partition of unity $\{\lambda_i\}$ corresponding to these U_i, so that we can vary the value of any $\bar{\bar{y}}_k$ by an infinitesimal amount $\bar{\bar{y}}_k \to \bar{\bar{y}}_k + \varepsilon_{ki}\lambda_i$ without altering the value of $\Phi_{\bar{\bar{y}}_k}(\mathbf{x})$ for $\mathbf{x} \notin U_i$.

We now consider the mapping $\Psi : M \times M \to \mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1}$ defined as $\Psi(\mathbf{x}, \mathbf{x}') = (\Phi_{\bar{\bar{y}}_k}(\mathbf{x}), \Phi_{\bar{\bar{y}}_k}(\mathbf{x}'))$. We define the set $W \subset M \times M$ as $W = \{(\mathbf{x}, \mathbf{x}') \in M \times M | \rho(\mathbf{x}, \mathbf{x}') \geq \delta\}$, so that (by our choice of δ), the mapping $\Phi_{\langle \bar{\bar{y}}_k \rangle}$ is necessarily injective on the complement of W in $M \times M$. Furthermore, note that the intersection of $\Psi(W)$ with the diagonal of $\mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1}$ gives the set of points $\{(\mathbf{x}, \mathbf{x}') \in M \times M | \Phi_{\bar{\bar{y}}_k}(\mathbf{x}) = \Phi_{\langle \bar{\bar{y}}_k \rangle}(\mathbf{x}')\}$, and therefore $\Psi(W) \cap \triangle = \emptyset$ is equivalent to $\Phi_{\langle y_k \rangle}$ injective. Our task, then, is to perturb the manifold $\Psi(W)$ using the ε_{ki} and $\varepsilon_{ki'}$ so that it does not intersect the diagonal manifold $\triangle$.

At each $\mathbf{p} \in \Psi(W) \cap \triangle$ we know that $\rho(\mathbf{x}, \mathbf{x}') > \delta$, so $\mathbf{x}$ and $\mathbf{x}'$ cannot belong to the same U_i. Consequently, varying an ε_{ki} or $\varepsilon_{ki'}$ only alters the value of $\Phi_{\langle \bar{\bar{y}}_k \rangle}$ at either $\mathbf{x}$ or $\mathbf{x}'$ (respectively). In the tangent space $\mathscr{T}^*_{\mathbf{p}}(\mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1})$, then, the direction of the $(2m+1)+(2m+1)$ infinitesimal changes given by the ε_{ki} and $\varepsilon_{ki'}$ are all linearly independent (indeed orthogonal) and as such span $\mathscr{T}^*_{\mathbf{p}}(\mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1})$. Since the tangent spaces of $\Psi(W)$ and Δ are at most $2m$ and $2m+1$ dimensional, respectively, we can construct a vector from a linear combination of $(\frac{\partial \Psi}{\partial \varepsilon_{ki}})_{\mathbf{p}}$ and $(\frac{\partial \Psi}{\partial \varepsilon_{ki'}})_{\mathbf{p}}$ that lies outside of both $\mathscr{T}^*_{\mathbf{p}}(\Psi(W))$ and $\mathscr{T}^*_{\mathbf{p}}(\Delta)$. Therefore, an infinitesimal perturbation corresponding to this linear combination will move the sub-manifolds $\Psi(W)$ and Δ away from each

other at the point **p** without creating a new intersection at another point. By keeping the size of these perturbations sufficiently small, we ensure that we stay confined to $\mathscr{U}'$, so that $\Phi_{\langle y_k \rangle}$ is still an immersion. This is a more transparent statement of the transversality argument used in the Takens proof (1981).

Thus, we have shown that for any arbitrary set of $2m+1$ observables $\langle \bar{y} \rangle$, we can find a set of observables $\langle y_k \rangle$ arbitrarily close to $\langle \bar{y}_k \rangle$ such that $\Phi_{\langle y_k \rangle}$ is an embedding—i.e., there is a dense set of observables $\{\langle y_k \rangle\} \subset \mathscr{Y}^{2m+1}$ such that $\Phi_{\langle y_k \rangle}$ is an embedding. The set of embeddings is open in the set of all mappings, so this set is dense and open, meaning that the embedding property is generic over all mappings. □

When mappings are confined to fixed lag relationships, Takens showed it is valid to independently perturb each component of Φ at a given point of the domain by perturbing the unlagged observation function, y, in the other parts of the domain corresponding to neighborhoods of the lagged states $\phi^{-1}(\mathbf{x})$, $\phi^{-2}(\mathbf{x})$, etc. This ensures that the perturbations to Φ maintain the structure of the lag relationships and that we have not inadvertently left the subset of interest. As we now show, this allows the above result to be easily extended to families of maps having component functions that are the lags of multiple observation functions. This is the relevant case for many practical examples where lags of multiple time series (multiple variables or observation functions) are required to achieve a mechanistic reconstruction of M (e.g.[20]). It also allows information on M to be leveraged when the time series are short, as is the case in many physical and biological problems[22, 36].

Before starting the proof, however, we must clarify exactly what the "subsets of interest" are. We define these sets as follows. First, we say y_q is a lag of the observable y if we can write $y_q = \phi^b(y)$ for positive b. We consider the lags in the positive time direction only to simplify notation in the proof, noting that the results apply equally to negative lags. Let $\mathbf{r} = \{r_1, r_2, \dots\}$ be the subset of $k = 1, \dots, 2m+1$ for which y_r, $r \in \mathbf{r}$

is an unlagged observable, i.e. y_r is not a lag of another $y \in \langle y_k \rangle$. We begin with the "unlagged" observation functions, y_r, or observation functions that are not a lag of another observable in $\langle y_k \rangle$. Now define a set C_r for each $r \in \mathbf{r}$ that contains y_r and any other observation function in $\langle y_k \rangle$ which is a lag of it. That is, C_r is the set of $y_q \in \langle y_k \rangle$ that are lags of y_r given as $y_q = \phi^{b_q}(y_r)$, where the lags b_q are distinct for fixed r. This choice of $\mathbf{C} = \{C_r : r \in \mathbf{r}\}$ and $\mathbf{b} = \{b_k : k = 1, \ldots, 2m+1\}$ determine a subset $\tilde{\mathscr{Y}}^{2m+1}_{\mathbf{C},\mathbf{b}} \subset \mathscr{Y}^{2m+1}$ containing all choices of $2m+1$ observables $\langle y_k \rangle$ which obey the correct lag relationships under a dynamical system ϕ. Note that each element of $\tilde{\mathscr{Y}}^{2m+1}_{\mathbf{C},\mathbf{b}}$ can be identified by the dynamical system and the y_r. We denote such an element, then, as $(\phi, \langle y_r \rangle)$.

Theorem 2. *Consider a diffeomorphism $\phi : M \to M$ on some compact manifold M of dimension m, along with $2m+1$ observation functions $y_k : M \to \mathbb{R}$, smoothly; by "smooth" we mean at least $\mathbb{C}^2$. Restrict the y_k to have the lag relationships corresponding to a collection of sets $\mathbf{C}$ and lags $\mathbf{b}$ under the dynamical system ϕ, and impose the following generic [39, 12] properties on ϕ:*

1. *The set A of periodic points with period $p < max(b_k)$ has finitely many points,*
2. *The eigenvalues of $(D\phi^b)_{\mathbf{x}}$ at each $\mathbf{x}$ in a compact neighborhood A are distinct and not equal to 1.*

Then, for generic $\langle y_k \rangle \in \tilde{\mathscr{Y}}^{2m+1}_{\mathbf{C},\mathbf{b}}$, the mapping described by

$$\Phi_{(\phi,\langle y_r \rangle)} = (y_1(\mathbf{x}), y_2(\mathbf{x}), \ldots, y_{2m+1}(\mathbf{x}))$$

is an embedding.

Proof. The proof of this theorem closely follows the logic of the previous proof and the original argument of Takens[12]. As noted above, any perturbations to Φ via its component functions $\langle y_k \rangle$ must remain within $\tilde{\mathscr{Y}}^{2m+1}_{\mathbf{C},\mathbf{b}}$ (the set of observables having

the desired lag relationships under ϕ prescribed by the C_r and the b_q). Here we must also deal with points of M that are fixed points or periodic under the dynamical system ϕ, i.e. the points for which there exists a b such that $\phi^b(\mathbf{x}) = \mathbf{x}$ (including the fixed point case, $b = 1$). The above proof shows that the mapping $\Phi_{\langle y_k \rangle}$ is generically an immersion because the co-vectors $(D\bar{y}_k)_{\mathbf{x}} \in \mathscr{T}^*(M)$ can be independently perturbed. This is also true for non-periodic points where there are fixed lag relationships between some observables, as we can perturb y_r in the neighborhood of $\phi^{-b_q}(\mathbf{x})$ and thus perturb $y_q = y_r(\phi^{b_q}(\mathbf{x}))$ without affecting y_r in the neighborhood of $\mathbf{x}$.

Note that periodic points $\mathbf{x}$ can exist such that the period b or some integer multiple of it, $n \cdot b$, is the fixed time lag between two observables $y_{q_1}, y_{q_2} \in \langle y_k \rangle$ belonging to the same C_r. Let $V \subset M$ be a compact neighborhood of all such points. For $\mathbf{x} \in V$, the vectors $(Dy_{q_1})_{\mathbf{x}}$ and $(D(y_{q_1} \circ \phi^{n \cdot b}))_{\mathbf{x}}$ cannot necessarily be perturbed independently. Nonetheless, while $y_{q_1}(\mathbf{x}) = y_{q_1}(\phi^{n \cdot b}(\mathbf{x}))$ for such a point, it is not generally true that $(Dy_{q_1})_{\mathbf{x}} = (D(y_{q_1} \circ \phi^{n \cdot b}))_{\mathbf{x}}$. By assumption, for each $\mathbf{x} \in V$, the eigenvalues of the $(D\phi^b)_{\mathbf{x}}$ are distinct and not equal to 1. Thus, by the chain rule, it is clear that $(Dy_{q_1})_{\mathbf{x}}$ and $(D(y_{q_1} \circ \phi^{n \cdot b}))_{\mathbf{x}}$ are linearly independent. As noted above, all the other $(Dy_k)_{\mathbf{x}}$ can be perturbed independently, so we can find a set of observables $\langle \bar{y}_k \rangle$ arbitrarily near $\langle y_k \rangle$ in $\tilde{\mathscr{Y}}^{2m+1}_{\mathbf{C},\mathbf{b}}$ for which $\Phi_{(\phi, \langle \bar{y}_r \rangle)}$ is an immersion on V. Note that because the set of immersions is open, there is an open neighborhood in $\tilde{\mathscr{Y}}^{2m+1}_{\mathbf{C},\mathbf{b}}$ around this $\langle \bar{y}_k \rangle$ for which every set of observables in that neighborhood gives an immersion.

We must also satisfy $\Phi_{(\phi, \langle y_r \rangle)}$ injective. The proof above relied on the ability to independently perturb the manifold $\Psi(W) \subset \mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1}$ at any point $(\mathbf{x}, \mathbf{x}')$ by an infinitesimal amount in any coordinate direction. For a periodic point on M with perioid b and two observables related as y_q and $y_q \circ \phi^{n \cdot b}$, it is impossible to independently perturb $\Psi(W)$ locally in the coordinate $y_q(\mathbf{x})$ or $y_q(\mathbf{x}')$, as you also perturb $y_q(\mathbf{x}) \circ \phi^{n \cdot b}$ or $y_q(\mathbf{x}') \circ \phi^{n \cdot b}$. By assumption, the set V has a finite number of elements. For such a

generic ϕ and any set $\langle y_k \rangle \in \tilde{\mathscr{Y}}^{2m+1}_{\mathbf{C},\mathbf{b}}$, any neighborhood of the $\langle y_k \rangle$ will contain a set of observables $\langle \bar{y}_k \rangle$ for which the unlagged observation functions $\langle \bar{y}_r \rangle$ take distinct values at each point in V.

We first perturb the y_r to find an open neighborhood of observables which give immersions when restricted to the set V. We then further perturb the observables to find within this neighborhood a set of observables $\langle \bar{y}_k \rangle$ for which $\Phi_{(\phi,\langle \bar{y}_r \rangle)}|V$ is also injective and therefore an embedding (on $V \subset M$). Since embeddings are dense in the space of all mappings, there is a neighborhood $\mathscr{U} \subset \tilde{\mathscr{Y}}^{2m+1}_{\mathbf{C},\mathbf{b}}$ such that for all $(\phi, \langle y_r \rangle) \in \mathscr{U}$, the map $\Phi_{(\phi,\langle y_r \rangle)}|V$ is an embedding.

We now show that we can find a $(\phi, \langle \bar{\bar{y}}_r \rangle) \in \mathscr{U}$ such that $\Phi_{(\phi,\langle \bar{\bar{y}}_r \rangle)}$ is an embedding on *all* of M. We first note that at points $\mathbf{x} \in M \backslash V$, the vectors $(D\bar{y}_k)_\mathbf{x} \in \mathscr{T}^*(M)$ can be perturbed independently, so we can find $(\phi, \langle \bar{\bar{y}} \rangle) \in \mathscr{U}$ for which $\Phi_{(\phi,\langle \bar{\bar{y}}_k \rangle)}$ is an immersion. Because an immersion is a local embedding, there is a δ such that for $\mathbf{x}, \mathbf{x}' \in M$, $0 < \rho(\mathbf{x}, \mathbf{x}') < \delta$ implies that $\Phi_{(\phi,\langle \bar{\bar{y}}_r \rangle)}(\mathbf{x}) \neq \Phi_{(\phi,\langle \bar{\bar{y}}_r \rangle)}(\mathbf{x}')$. Since the set of immersions is open in the set of possible mappings, there is a neighborhood $\mathscr{U}' \subset \mathscr{U}$ such that for any $(\phi, \langle y_r \rangle) \in \mathscr{U}'$, the corresponding mapping $\Phi_{(\phi,\langle y_r \rangle)}$ is an immersion. Thus, for the same δ as above, $0 < \rho(\mathbf{x}, \mathbf{x}') \leq \delta$ implies $\Phi_{(\phi,\langle y_r \rangle)}(\mathbf{x}) \neq \Phi_{(\phi,\langle y_r \rangle)}(\mathbf{x}')$.

Now we need to show that there is a $(\phi, \langle y_r \rangle) \in \mathscr{U}'$ such that $\Phi_{(\phi,\langle y_r \rangle)}$ is also injective on M. As noted in the first proof, this is equivalent to $\Psi(V) \cap \triangle = \emptyset$ for the mapping $\Psi : M \times M \to \mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1}$ defined as $\Psi(\mathbf{x}, \mathbf{x}') = (\Phi_{(\phi,\langle y_r \rangle)}(\mathbf{x}), \Phi_{(\phi,\langle y_r \rangle)}(\mathbf{x}'))$. If $\mathbf{x}$ and $\mathbf{x}'$ are both in V or $\rho(\mathbf{x}, \mathbf{x}') \leq \delta$, we already know that $\Phi_{(\phi,\langle y_r \rangle)}(\mathbf{x}) \neq \Phi_{(\phi,\langle y_r \rangle)}(\mathbf{x}')$. Thus we restrict ourselves to the set $W = \{\mathbf{x}, \mathbf{x}' \in M \times M \mid \phi(\mathbf{x}, \mathbf{x}') > \delta \text{ and not both } \mathbf{x}, \mathbf{x}' \in \text{int}(V)\}$.

To perturb the manifold $\Psi(W)$ away from $\triangle$ at points of intersection, $\mathbf{p} \in \Psi(W) \cap \triangle$, we must be able to find variations for which the tangent vector $(\frac{\partial \Psi}{\partial \varepsilon})_\mathbf{p}$ is linearly independent from the $2m$ tangent vectors $(\frac{\partial \Psi}{\partial x_i})_\mathbf{p}$ and $(\frac{\partial \Psi}{\partial x'_i})_\mathbf{p}$ and lies outside of $\mathscr{T}^*_\mathbf{p}(\triangle)$. In

the first proof, it was obvious that each component of Ψ could be perturbed independently. Now we must be more careful. We do this by first creating a collection of N open subsets of M, $\{U_i\}$, with the following properties:

1. The $\{U_i\}$ cover the closure of $M \backslash V$.
2. For each b_q and $i = 1, \ldots, N$, the diameter of $\phi^{-b_q}(U_i)$ is less than δ.
3. For all choices of $i, j \in \{1, \ldots, N\}$, the set U_j intersects with $\phi^{-b_q}(U_i)$ for at most one b_q.
4. For $\mathbf{x}$ and $\mathbf{x}'$ such that $\phi^{-b}(\mathbf{x}) \in M \backslash \bigcup U_i$ for some $b \in \mathbf{b}$, $\mathbf{x}' \notin V$, and $\rho(\mathbf{x}, \mathbf{x}') > \delta$, no two of $\mathbf{x}, \{\phi^{b_q}(\mathbf{x})\}, \mathbf{x}', \{\phi^{b}(\mathbf{x}')\}$ belong to the same U_i.

Take a partition of unity $\{\lambda_i\}$ corresponding to this $\{U_i\}$. Because of the way we constructed the $\{U_i\}$, we can vary the value of each $\bar{\bar{y}}_k$ at any point $\mathbf{x} \in M \backslash V$ by an infinitesimal amount without altering the value of the other $\bar{\bar{y}}_k$ in the neighborhood of $\mathbf{x}$. We make this explicit as follows. To perturb the y_r, we take $\bar{\bar{y}}_r \to \bar{\bar{y}}_r + \varepsilon_{ri}\lambda_i$ for i corresponding to $\mathbf{x} \in U_i$. To perturb the other y_k ($y_k = y_r \circ \phi^{b_q}$ for some r), we perturb $\bar{\bar{y}}_r \to \bar{\bar{y}}_r + \varepsilon_{ri}\lambda_i$ for i corresponding to $\phi^{-b_q}(\mathbf{x}) \in U_i$. Consider the $2m+1$ perturbations, ε_{ri}, which are independent shifts at $\mathbf{x}$ in distinct y_k. In $\mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1}$, we note that each corresponding tangent vector $(\frac{\partial \Psi}{\partial \varepsilon_{ri}})_{\mathbf{p}}$ lies outside of $\mathscr{T}^*_{\mathbf{p}}(\triangle)$. Note the $(\frac{\partial \Psi}{\partial \varepsilon_{ri}})_{\mathbf{p}}$ together with any basis of $\mathscr{T}^*_{\mathbf{p}}(\triangle)$ form a linearly independent set of vectors. Since the dimension of span $\left((\frac{\partial \Psi}{\partial x_i})_{\mathbf{p}}, (\frac{\partial \Psi}{\partial x_{i'}})_{\mathbf{p}}\right)$ is at most $2m$, there must be a linear combination of the $(\frac{\partial \Psi}{\partial \varepsilon_{ri}})_{\mathbf{p}}$ that lies outside of both $\mathscr{T}^*_{\mathbf{p}}(\Psi(W))$ and $\mathscr{T}^*_{\mathbf{p}}(\Delta)$, which can be used to perturb $\Psi(W)$ away from Δ. By keeping variations in the ε_{ri} sufficiently small, we can find a set of $\langle y_k \rangle$ such that $(\phi, \langle y_r \rangle) \in \mathscr{U}'$ and $\Psi(\mathbf{x}, \mathbf{x}') \cap \triangle = \emptyset$ (where Ψ now corresponds to the $\Phi_{\phi, \langle y_k \rangle}$ map). This pair gives a mapping $\Phi_{(\phi, \langle y_k \rangle)}$ that is both an immersion and injective, and thus is an embedding. Because $\mathscr{U}'$ was an arbitrarily small neighborhood of any point in

$\tilde{\mathscr{Y}}^{2m+1}_{\mathbf{C},\mathbf{b}}$, this means embeddings are dense in $\tilde{\mathscr{Y}}^{2m+1}_{\mathbf{C},\mathbf{b}}$, and the set of embeddings is open in the set of mappings. Thus, the map $\Phi_{(\phi,\langle y_r\rangle)}$ given by $(\phi,\langle y_r\rangle) \in \tilde{\mathscr{Y}}^{2m+1}_{\mathbf{C},\mathbf{b}}$ is generically an embedding.

□

Just as Takens extends the original result for discrete time to dynamical systems in continuous time, we can extend our result as follows:

Corollary 3. *Consider a smooth vector field X on some compact manifold M along with $2m+1$ observables $y_k : M \to \mathbb{R}$, smoothly; by "smooth" we mean at least $\mathbb{C}^2$. Define ϕ_t as the flow on X. Suppose we restrict the y_k to have the lag relationships corresponding to a collection of sets C_r and lags b_q under the discrete dynamical system ϕ_τ, where τ is a constant. We impose the following generic properties on X:*

1. *For points $\mathbf{x}$ such that $X(\mathbf{x}) = 0$, the eigenvalues of $(D\phi_\tau)_\mathbf{x}$ are distinct and not equal to 1.*

2. *No periodic integral curve of X has integer period $\leq 2m+1$.*

Then, for generic $\langle y_k\rangle \in \tilde{\mathscr{Y}}^{2m+1}_{\mathbf{C},\mathbf{b}}$, the mapping described by

$$\Phi_{(\phi_\tau,\langle y_r\rangle)} = (y_1(\mathbf{x}), y_2(\mathbf{x}), \ldots, y_{2m+1}(\mathbf{x}))$$

is an embedding.

Proof. In this case, ϕ_τ is a discrete time dynamical system on M satisfying the conditions imposed in the theorem above, and this corollary follows directly.

□

2.2.2 A Theorem in the Style of Sauer et al.: The Prevalent Case

We now give an explicit proof of Remark 2.9 from [29] using the framework constructed in their original paper, but we extend the language to cover reconstructions using non-consecutive lags (from multiple time series). The proof uses Lemma 4.1, 4.6, and 4.11 from [29] to show that 1 : 1 mappings and immersions are prevalent in the space $\tilde{\mathscr{Y}}^{2m+1}_{\mathbf{C},\mathbf{b}}$, just as Sauer et al. use Lemma 4.6 to prove Theorem 3.3, and Lemmas 4.1 and 4.11 to prove Theorem 3.5. These lemmas are now stated (for the proofs, see their original paper).

Lemma 4. (*Originally part 2 of 4.1*) *Let n and k be positive integers,* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *distinct points in* $\mathbb{R}^k$, $u_1, \ldots, u_n$ *in* $\mathbb{R}$, *and* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *in* $\mathbb{R}^k$. *Then there exists a polynomial h in k variables of degree at most n such that for* $i = 1, \ldots, n$, $\nabla h(\mathbf{x}_i) = \mathbf{v}_i$.

Lemma 5. (*Originally 4.6*) *Let A be a compact subset of* $\mathbb{R}^k$. *Let* $\Phi_0, \Phi_1, \ldots, \Phi_t : A \to \mathbb{R}^n$ *be Lipschitz maps. For each integer* $r \geq 0$, *let* S_r *be the set of pairs* $\mathbf{x}_1 \neq \mathbf{x}_2$ *in A for which the* $n \times t$ *matrix*

$$M_{x_1 x_2} = (\Phi_1(\mathbf{x}_1) - \Phi_1(\mathbf{x}_2), \ldots, \Phi_t(\mathbf{x}_1) - \Phi_t(\mathbf{x}_2))$$

has rank r, and let $d_r = \text{lower boxdim}(\bar{S}_r)$. *Define* $\Phi_\alpha = \Phi_0 + \sum_{i=1}^t \alpha_i \Phi_i : A \to \mathbb{R}^n$. *If* $d_r < r$ *for all integers* $r \geq 0$, *then for* $\alpha = (\alpha_1, \ldots, \alpha_t)$ *outside a measure zero subset of* $\mathbb{R}^t$, *the map* Φ_α *is* $1:1$.

Lemma 6. (*Originally 4.11*) *Let A be a compact subset of a smooth manifold embedding in* $\mathbb{R}^k$. *Let* $\Phi_0, \Phi_1, \ldots, \Phi_t$ *be a set of smooth maps from an open neighborhood U of A to* $\mathbb{R}^n$. *For each positive integer r, let* S_r *be the subset of the unit tangent bundle* $S(A)$ *such that the* $n \times t$ *matrix*

$$((D\Phi_1)_{\mathbf{x}}(\mathbf{v}), \ldots, (D\Phi_t)_{\mathbf{x}}(\mathbf{v}))$$

has rank r, and let $d_r =$ *lower boxdim*$(\bar{S}_r)$. *Define* $\Phi_\alpha = \Phi_0 + \sum_{i=1}^{t} \alpha_i \Phi_i : A \to \mathbb{R}^n$. *If* $d_r < r$ *for all integers* $r \geq 0$, *then for almost every* $\alpha \in \mathbb{R}^t$, *the map* Φ_α *is an immersion on A.*

To apply these lemmas, it is necessary to restrict the dimension of the sets of periodic orbits—that is, the sets $A_p = \{\mathbf{x} \in A : \phi^p(\mathbf{x}) = \mathbf{x}\}$ for $p < \max(\{b \in \mathbf{b}\})$. For the case of consecutive lags, Sauer et al. state sufficient conditions to be $\text{boxdim}(A_p) < p/2$. A sufficient condition for non-consecutive lags is a bit more complicated. Define the constants $B_{pr} =$ number of $y_q \in C_r$ such that $b_q = m \cdot p + b_{q'}$ for at least one $b_{q'}$ and $m \in \mathbb{N}$. Also, define $B_p = \sum_r B_{pr}$. A sufficient condition on the A_p is $2 \cdot \text{boxdim}(A_p) < n - B_p$.

Theorem 7. *Let* ϕ *be a diffeomorphism on an open subset U of* $\mathbb{R}^m$, *and let A be a compact subset of U,* $\text{boxdim}(A) = d$. *Let* $\mathbf{C}$ *be a collection of sets and* $\mathbf{b}$ *a set of lag relationships as above, such that* $n = \sum_r n_r > 2d$. *Assume that for every positive integer* $p \leq max(\{b \in \mathbf{b}\})$, *the set* A_p *of periodic points of period p satisfies* $2 \cdot boxdim(A_p) < n - B_p$, *and that for each point of* A_p, *the Jacobian* $D\phi^p$ *has distinct eigenvalues. Then, for almost every set of n observation functions* $\{y_k\}$ *satisfying the given lag relationships, the map*

$$\Phi_{(\phi,\langle y_r \rangle)}(\mathbf{x}) = (y_1(\mathbf{x}), \ldots, y_n(\mathbf{x}))$$

is an embedding on A.

Proof. Without loss of generality, assume we have ordered the components of $\Phi_{(\phi,\langle y_r \rangle)}$ with y_{r_1} and all its lags first, then y_{r_2} and its lags, etc. That is,

$$\Phi_{(\phi,\langle y_r \rangle)}(\mathbf{x}) = (y_{r_1}(\mathbf{x}), y_{r_1}^{b_2}(\mathbf{x}), \ldots, y_{r_2}(\mathbf{x}), \ldots).$$

To show prevalence, we find a suitable probe space (see [29]). The infinite dimensional space for the univariate theorem is the observation functions $y : U \to \mathbb{R}$,

smoothly. For maps constructed from multiple lags, this becomes the sets of $s_r = \text{size}(\mathbf{r})$ unlagged observation functions. Sauer et al. take the probe space for the univariate theorem to be any set H of polynomials in m variables which include all such polynomials up to degree $2n$. It is now necessary to have a set of polynomials for each of the y_r. Thus, we take the probe space for this theorem to be the Cartesian product of s_r copies of H.

Let $\langle h_1, \ldots, h_t \rangle$ be a basis for H. We want to show that for almost all choices of $s_r \times t$ coefficients $\alpha_{r,t}$, the map $\Phi_{(\phi, \langle \bar{y}_r \rangle)}$ defined by the observation functions $\bar{y}_r = y_r + \sum_{i=1}^{t} \alpha_{r,i} h_i$ is an embedding. We first demonstrate that almost every $\Phi_{(\phi, \langle \bar{y}_r \rangle)}$ is $1:1$, proceeding as in the proof of Theorem 4.3 in [29].

To sensibly apply Lemma 5, we adopt the following convention: think of $\Phi_{(\phi, \langle \bar{y}_r \rangle)}$ as a perturbation of $\Phi_{(\phi, \langle y_r \rangle)}$, which is the summed effect of perturbations on each y_r separately. For each pair (r,i), $r \in \mathbf{r}$ and $i \in \{1, \ldots, t\}$, there is a map $\Phi_{r,i} : U \to \mathbb{R}^n$ which is $\Phi_{(\phi, \langle \tilde{y}_{r'} \rangle)}$ for $\tilde{y}_{r'} = h_i$ if $r = r'$ and 0 otherwise. The components of $\Phi_{r,i}(\mathbf{x})$ are either 0 or of the form $h_i(\phi^{b_q}(\mathbf{x}))$. Consequently, $\Phi_{\phi, \langle \bar{y}_r \rangle} = \Phi_{\phi, \langle y_r \rangle} + \sum_r \sum_{i=1}^{t} \alpha_{r,i} \Phi_{r,i}(\mathbf{x})$, which matches the structure Lemma 5.

We now check that the rank of the matrix $M_{x_1 x_2}$ satisfies the conditions of Lemma 5 for each pair of distinct $\mathbf{x}_1, \mathbf{x}_2 \in A$. Note that to avoid confusion with the previous section of this paper and Takens' original work, we continue to use row vectors to describe the transformations Φ. However, Sauer et al. [29] prefer column vectors, so it is necessary to use of transposes in several instances. Thus, we have

$$M_{x_1 x_2}^T = \begin{pmatrix} \Phi_{r_1,1}(\mathbf{x}_1) - \Phi_{r_1,1}(\mathbf{x}_2) \\ \vdots \\ \Phi_{r_1,t}(\mathbf{x}_1) - \Phi_{r_1,t}(\mathbf{x}_2) \\ \Phi_{r_2,1}(\mathbf{x}_1) - \Phi_{r_2,1}(\mathbf{x}_2) \\ \vdots \end{pmatrix}.$$

Note that $M_{x_1x_2}$ is a block diagonal matrix, and so it has rank equal to the sum of the rank of the blocks. Each of the s_r blocks can be rewritten as the product of two matrices, J_r and H_r, where the entries of H_r are values of a single polynomial h and the entries of J_r are each one of $\{1,0,-1\}$. Note, there are multiple possible choices for H_r and J_r that give the same $M_{x_1x_2}$.

Case 1: First consider $\mathbf{x}_1$ and $\mathbf{x}_2$ that do not both lie in a periodic orbit of integer period less than $\max(\mathbf{b})$. We specify H_r so that the first n_r rows, where n_r is the size of the set C_r, correspond to the $h_{r,i}(\mathbf{x_1}), h_{r,i}(\phi^{b_{r+1}}(\mathbf{x_1})), \ldots, h_{r,i}(\phi^{b_{r+n_r}}(\mathbf{x_1}))$, and the next n_r correspond to the $h_{r,i}(\phi^{b_{q_r}}(\mathbf{x_2}))$. H_r is onto, so the rank of $M_{x_1x_2}$ is just the sum of the ranks of the J_r. For this case, J_r contains a copy of I_{n_r}, and thus will have rank n_r. The entire matrix $M_{x_1x_2}$ will thus have rank $n = \sum_r n_r$, which satisfies the conditions of Lemma 5.

Case 2: Now consider $\mathbf{x}_1$ and $\mathbf{x}_2$ in separate periodic orbits with periods p_1 and p_2 such that $1 \leq B_{p_1} \leq B_{p_2}$ and $p_1, p_2 < \max(\mathbf{b})$. H_r will have $B_{p_1 r}$ fewer rows corresponding to the $b_{q_1} = m \cdot p_1 + b_{q_2}$ for some $m \in \mathbb{N}$ (there will also be a reduction in the number of rows associated with B_{p_2}). In this case, J_r will still contain the column space of $I_{(n_r - B_{p_1 r})}$ and thus $\text{rank}(J) \geq \sum_r n_r - B_{p_1 r} = n - B_{p_1}$. Again the H_r are onto, and so the rank of $M_{x_1x_2}$ is the rank of J.

The dimension of the set S of all pairs $\mathbf{x}_1$ and $\mathbf{x}_2$ is $\dim(S) = \dim(A_{p_1}) + \max(\dim(A_{p_2}))$. By the conditions placed on the size of the A_p, we can conclude that $\dim(S) < n - B_{p_1} \leq \dim(M_{x_1x_2})$, and thus that Lemma 5 applies to this case as well.

Case 3: Finally we consider $\mathbf{x}_1$ and $\mathbf{x}_2$ in the same p-periodic orbit, $p < \max(\mathbf{b})$. Now the matrix H_r becomes more complicated, since some of the $h(z)$ pertaining to x_2 may be equal to $h(z)$ pertaining to x_2. Consequently, the J_r are no longer guaranteed

to contain the column space of the identity. Each J_r does contain the column space of an $n_r - B_{pr}$ dimensional matrix with 1 along the upper diagonal and a single -1 off the diagonal in each column. Using elementary operations, it is possible to make the first m columns of J_r upper diagonal for some integer $m \geq (n_r - B_{pr})/2$. Thus, the rank of each J_r is at least $(n_r - B_{pr})/2$ and the entire matrix has $\text{rank}(J) \geq (n - B_p)/2$.

The dimension of the set S of all such $\mathbf{x}_1$ and $\mathbf{x}_2$ is just A_p. By the imposed conditions, $\dim(S) < (n - B_p)/2 \leq \dim(M_{x_1 x_2})$, and Lemma 5 applies.

Now we want show that almost every $\Phi_{(\phi, \langle \bar{y}_r \rangle)}$ is an immersion. We check that the matrix

$$\left((D\Phi_{r_1,1})_{\mathbf{x}}(\mathbf{v})^T, \ldots, (D\Phi_{r_1,t})_{\mathbf{x}}(\mathbf{v})^T, (D\Phi_{r_2,1})_{\mathbf{x}}(\mathbf{v})^T, \ldots\right)$$

has full rank and thus satisfies the conditions of Lemma 6 for each $(\mathbf{x}, \mathbf{v})$ in the tangent bundle $S(A)$. Note that this is a block diagonal matrix with s_r blocks, so it is sufficient to show that the columns of the ith block span the subspace $\mathbb{R}^{n_{r_i}}$ for $i = 1, \ldots, s_r$. We consider two cases.

Case 1: Consider first the subset S' of $\mathbf{x}$ that are not periodic with period $p < \max(\mathbf{b})$. The entries of each block are of the form $\nabla h(\phi^b(\mathbf{x}))^T (D\phi^b)_{\mathbf{x}}(\mathbf{v})$. Since ϕ is a diffeomorphism and $\mathbf{v} \neq 0$, we know that $(D\phi^b)_{\mathbf{x}}(\mathbf{v}) \neq 0$. Furthermore, the $\phi^b(\mathbf{x})$ are distinct points. Examining Lemma 4, it is clear that the columns span $\mathbb{R}^{n_r}$. The dimension of S' is at most $2d - 1$, so we may apply Lemma 6.

Case 2: Now consider the subset S' of $\mathbf{x}$ that are periodic with period $p < \max(\mathbf{b})$. By the conditions of the theorem, $(D\phi^{b_1})_{\mathbf{x}}$ has distinct eigenvalues from $(D\phi^{b_2})_{\mathbf{x}}$. Therefore, $\nabla h(\phi^{b_1}(\mathbf{x}))^T (D\phi^{b_1})_{\mathbf{x}}(\mathbf{v}) \neq \nabla h(\phi^{b_2}(\mathbf{x}))^T (D\phi^{b_2})_{\mathbf{x}}(\mathbf{v})$. Furthermore, the relationship

depends on h, and again referencing Lemma 4, it is clear that the columns span $\mathbb{R}^{n_r}$. The dimension of S' is certainly less than $2d-1$, so we can safely apply Lemma 6.

□

Theorem 7 can be extended to continuous dynamical systems (smooth vector fields on a manifold) by letting the flow ϕ_t of X be ϕ in the statement of the theorem.

2.3 Discussion

Theorem 1 and the more general result presented in Theorem 2 (and its corollary) were given proofs intended to follow those presented by Takens. The original "transversality" argument, however, has been replaced with what we reckon is a simpler and more direct argument. These clarify how perturbations to the observation functions can be constructed and highlight why $2m+1$ dimensions are necessary to have mappings that are generically embeddings. Theorem 7 is similar to Theorem 2, but takes advantage of the more powerful framework, built around the notion of prevalence, established by Sauer et al. [29]. It also provides more specific conditions on the periodic orbits than Theorem 2 and thus can be applied to certain non-generic situations that Takens' original framework would exclude. Namely, the set of periodic points need not be finite (as required in Takens' original theorem and Theorem 2), so long as the dimensionality does not exceed the bounds stated in Theorem 7. Theorem 7 is an extension of Remark 2.9 in [29], which we explicitly proved by determining a sufficient restriction for the periodic orbits when the lags composing Φ aren't necessarily consecutive.

This work also develops a language to describe a wider family of cases for reconstructing state space manifolds from multiple observational time series to encourage wider applicability of SSR in the natural sciences. For example, these results can be extended to another special case of interest for reconstructions using time derivatives

[40], when multiple observation functions are available. The argument for this case is analogous to that used by Takens [12] for the case when all the derivatives are from a single observation function. Furthermore, these theorems validate heuristic work using spatial lag reconstructions and mixed spatial and temporal lag reconstructions to study spatially coupled dynamics [41].

More importantly, in terms of future applications, Theorems 2 and 7 set the stage for practical reconstruction of state space manifolds from multiple observation functions. This is significant in answering objections to single variable state space reconstruction (SSR) concerning the excessive phenomenology of lagged-coordinate embeddings [26]. These two theorems provide proof of principle for modeling attempts of nonlinear dynamics in the natural sciences involving multiple time series (e.g. [20]), and lays bare the rather non-restrictive assumptions required in such applications for building mechanistic models from multiple time series variables. Moreover, it gives support to the notion of using multiple embeddings as a potentially efficient way of extracting information from time series data of limited length, but where there are potentially many simultaneous observations of dynamics on the same attractor manifold. By reducing correlations in noise between the reconstructed coordinates, these techniques should allow reconstructions to exceed the limitations placed on univariate methods [35], as heuristic examples have already suggested [20]. The potential information leverage provided by multiple embeddings possible from novel combinations of variables (and their lags) can pave the way for a plethora of new applied techniques to exploit the time-limited, but parallel observations of nature [36]. This paper is intended to complement the existing literature on SSR and help promote this potential growth area in the natural sciences.

Acknowledgments

We wish to thank Hao Ye, James Crutchfield, John Melack, Donald DeAngelis, Simon Levin, J. Doyne Farmer, Martin Casdagli, Tim Sauer, Sarah Glaser, Chih-hao Hsieh, Stephen Munch and Charles Peretti, Michael Fogarty, Alec MacCall, Andrew Rosenberg, Les Kaufman, and Irit Altman for helpful comments and editorial advice.

Chapter 2, in full, is a reprint of the material as it appears in Public Library of Science ONE 2011. Deyle, Ethan Robert; Sugihara. The dissertation author was the primary investigator and author of this paper

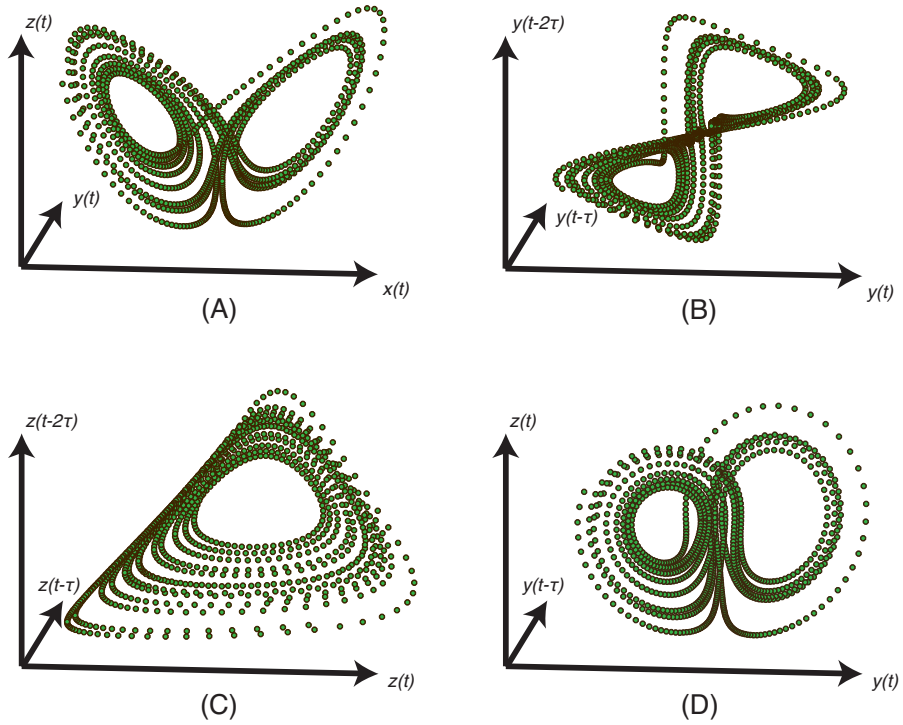

Figure 2.1: Lorenz attractor with three shadow manifolds. The Lorenz attractor [37] is shown with three shadow manifolds created from lag-coordinate transformations. The typical parameters were used: $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$, giving the three coupled equations as $\dot{x} = 10(y-x)$, $\dot{y} = 28x - xz - y$, and $\dot{z} = xy - (8/3)z$. The solution was computed using a fourth order Runge-Kutta method with a time step of $\delta t = 0.01$, and the time lag used to create the shadow manifolds was $\tau = 8\,\delta t = 0.08$. (A) The trajectory shown in the x, y, and z coordinates of the original system reveals a two-lobed manifold. (B) A univariate transformation using time lags of the y-coordinate, $\Phi = (y(t), y(t-\tau), y(t-2\tau))$, preserves this two-lobed structure (and other topological properties), verifying Takens' theorem. (C) A univariate transformation using time lags of the z-coordinate, $\Phi = (z(t), z(t-\tau), z(t-2\tau))$, does not preserve the two-lobed structure. Local neighborhoods of the original attractor are, however, preserved. Thus, though this mapping violates a genericity assumption of the original theorem and is not an embedding, it is an immersion of the original manifold. (D) A multivariate transformation using both the y- and z-coordinates, $\Phi = (y(t), y(t-\tau), z(t))$, fulfills the assumptions of Theorems 2 and 7. As predicted, it also preserves the two-lobed structure of the Lorenz and is a valid embedding.

2.4 References

[1] Casdagli M (1989) Nonlinear prediction of chaotic time series. Physica D 35: 335-356.

[2] Crutchfield JP (1979) Prediction and stability in classical mechanics. Senior thesis in physics and mathematics, University of California, Santa Cruz.

[3] Farmer JD, Sidorowich JJ (1987) Predicting chaotic time series. Physical Review Letters 59: 845-848.

[4] Jenouvrier S, Weimerskirch H, Barbraud C, Park YH, Cazelles B (2005) Evidence of a shift in the cyclicity of antarctic seabird dynamics linked to climate. Proceedings of the Royal Society of London, Series B 272: 887-895.

[5] Lo TT, Hsu HH (2010) Change in the dominant decadal patterns and the late 1980s abrupt warming in the extratropical northern hemisphere. Atmospheric Science Letters 11: 210–215.

[6] Planque B, Batten SD (2000) *Calanus finmarchicus* in the north atlantic: the year of *Calanus* in the context of interdecadal change. ICES Journal of Marine Science 57: 1528-1535.

[7] Ramanathan A, Wang C, Schreiber SL (2002) Perturbational profiling of a cell-line model of tumorigenesis by using metabolic measurements. Proceedings of the National Academy of Sciences of the United States of America 102: 5992-5997.

[8] Ruelle D (1989) Chaotic evolution and strange attractors: the statistical analysis of time series for deterministic nonlinear systems. Cambridge University Press, Cambridge.

[9] Scheffer M, Carpenter SR (2003) Catastrophic regime shifts in ecosystems: linking theory to observation. Trends in Ecology & Evolution 18: 648-656.

[10] Seth AK, Izhikevich E, Reeke GN, Edelman GM (2006) Theories and measures of consciousness: an extended framework. Proceedings of the National Academy of Sciences of the United States of America 103: 10799-10804.

[11] Soramaki K, Bech ML, Arnold J, Glass RJ, Beyeler WE (2007) The topology of interbank payment flows. Physica A 379: 317-333.

[12] Takens F (1981) Detecting strange attractors in turbulence. In: Rand DA, Young LS, editors, Symposium on Dynamical Systems and Turbulence. Berlin: Springer-Verlag, volume 898 of *Lecture Notes in Mathematics*, pp. 366-381.

[13] Wagner BK, Kitami T, Gilbert TJ, Peck D, Ramanathan A, et al. (2008) Large-scale chemical dissection of mitochondrial function. Nature Biotechnology 26: 343-351.

[14] May RM, Levin SA, Sugihara G (2008) Complex systems: Ecology for bankers. Nature 451: 893-895.

[15] Sugihara G (2010) On early warning signs. Seed Magazine Global Reset 2010.

[16] Sugihara G (2010) Nature is nonlinear. Kyoto Journal 75: 56-57.

[17] Brock WA, Sayers CL (1988) Is the business cycle characterized by deterministic chaos? Journal of Monetary Economics 22: 71-90.

[18] Brock WA, Malliaris AG (1989) Differential Equations, Stability and Chaos in Dynamic Economics, volume 27 of *Advanced Textbooks in Economics*. North-Holland.

[19] Brock WA, Hsieh DA, LeBaron BD (1991) Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence. MIT Press, Cambridge, Massachusettes.

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

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

[22] Hsieh CH, Anderson C, Sugihara G (2008) Extending nonlinear analysis to short ecological time series. American Naturalist 171: 71-80.

[23] Rodo X, Pascual M, Fuchs G, Faruque ASG (2002) ENSO and cholera: a nonstationary link related to climate change? Proceedings of the National Academy of Sciences of the United States of America 99: 12901-12906.

[24] Schaffer WM (1984) Stretching and folding in lynx fur returns: evidence for a strange attractor in nature? American Naturalist 124: 789-820.

[25] Sugihara G, Grenfell B, May RM (1990) Distinguishing error from chaos in ecological time series. Philosophical Transactions of the Royal Society of London, Series B: Biological Science 330: 235-250.

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

[27] Sugihara G (1994) Nonlinear forecasting for the classification of natural time series. Philosophical Transactions of the Royal Society of London, Series A 348: 477-495.

[28] Packard NH, Crutchfield JP, Farmer JD, Shaw RS (1980) Geomtery from a time series. Physical Review Letters 45: 712-716.

[29] Sauer T, Yorke JA, Casdagli M (1991) Embedology. Journal of Statistical Physics 65: 579-616.

[30] Hunt BR, Sauer T, Yorke JA (1992) Prevalence: a translation-invariant "almost every" on infinite-dimensional spaces. Bulletin of the American Mathematical Society 27: 217-238.

[31] Stark J, Broomhead DS, Davies ME, Huke JP (1997) Takens embedding theorems for forced and stochastic systems. Nonlinear Analysis: Theory, Methods, and Applications 30: 5303 - 5314.

[32] Stark J, Broomhead DS, Davies ME, Huke JP (2003) Delay embeddings for forced systems: II. Stochastic forcing. Journal of Nonlinear Science 13: 519-577.

[33] Stark J (1999) Delay embeddings for forced systems: I. Deterministic forcing. Journal of Nonlinear Science 9: 255-332.

[34] Gibson J, Farmer JD, Casdagli M, Eubank S (1992) An analytic approach to practical state space reconstruction. Physica D 57: 1-30.

[35] Casdagli M, Eubank S, Farmer JD, Gibson J (1991) State space reconstruction in the presence of noise. Physica D 41: 52-98.

[36] Ye H, Sugihara G, Deyle ER (2011) Leverging information from multiple time series with ensemble state space reconstructions. In preparation.

[37] Lorenz EN (1963) Deterministic nonperiodic flow. Journal of Atmospheric Science 20: 130–141.

[38] Whitney H (1936) Differentiable manifolds. Annals of Mathematics 37: 645-680.

[39] Huke JP (1993) Embedding nonlinear dynamical systems, a guide to Takens' theorem. Internal report, DRA, Malvern, UK.

[40] Crutchfield JP, McNamara BS (1987) Equations of motion from a data series. Complex Systems 1: 417.

[41] Ørstavik S, Stark J (1998) Reconstruction and cross-prediction in coupled map lattices using spatio-temporal embedding techniques. Physics Letters A 247: 145-160.

Chapter 3

Predicting climate effects on Pacific sardine

Abstract

For many marine species and habitats, climate change and overfishing present a double threat. To manage marine resources effectively, it is necessary to adapt management to changes in the physical environment. Simple relationships between environmental conditions and fish abundance have long been used in both fisheries and fishery management. In many cases, however, physical, biological, and human variables feed back on each other. For these systems, associations between variables can change as the system evolves in time. This can obscure relationships between population dynamics and environmental variability, undermining our ability to forecast changes in populations tied to physical pro- cesses. Here we present a methodology for identifying physical forcing variables based on nonlinear forecasting and show how the method provides a predictive understanding of the influence of physical forcing on Pacific sardine.

3.1 Introduction

Ecosystem-based management (EBM) is an essential challenge that places strong demands on our understanding of coupled social-ecological systems. EBM requires an understanding of how human activities such as fishing influence and are influenced by other parts of the ecosystem. This includes accounting for the effects of the physical environment on exploited populations. However, the interactions between ecosystem components can be complex, and unraveling physical-biological interactions remains a challenge. For instance, a retrospective study of 35 exploited and unexploited species in the California Current shows that fishing pressure can amplify the influence of environmental forcing on populations by truncating the age structure (1). This study and others (2-4) demonstrate that the effect of environ- mental forcing on populations can be contingent on fishing effort, current abundance, and age structure. This raises an important issue: Ecosystem variables are not separate, decomposable forces. Instead, their interactions are state-dependent, meaning that the impact of one variable on another depends on the state of the variables.

State-dependent behavior can confound many traditional statistical methods. Witness that valid correlations between physical and biological variables can be difficult to find (5) and can appear and disappear with time (6). In fact, nonlinear systems (systems with state-dependent interactions) can produce mirage correlations: variables that seem positively correlated over one period in time may seem negatively correlated or unrelated over another period (7). A meta-analysis of environment-recruitment relationships in marine populations shows that these correlations hold up poorly when retested with new data (6). Consequently, EBM requires more robust methods for identifying driving variables and understanding their influence on population and community dynamics. Here, we show that methods based on multivariate state space reconstruction (SSR) (8) offer an

alternative method for studying physical-biological interactions from observational time series. It is a robust framework for studying ecosystems empirically. For those unfamiliar with state space reconstruction, we recommend two short animations (http://simplex.ucsd.edu/Movie_s1.mov and http://simplex.ucsd.edu/Movie_s2.mov). These techniques have been successful in identifying state-dependent physical-biological interactions in larval reef fish populations (9) and in models with simple periodic forcing (10). To demonstrate the utility of multivariate SSR for ecosystem-based management, we investigate the current conundrum over the management of Pacific sardine (*Sardinops sagax*). The Pacific sardine fishery in the California Current ecosystem (CCE) is a rare example of a fishery that has been managed with explicit consideration for the environment. However, conflicting evidence concerning the effect of temperature on sardine productivity (11) led to removal of this pioneering environmental control rule in 2012.

The policy was informed by a rich history of research. Sardine populations around the world have undergone boom-bust cycles. Crashes in California have coincided with crashes in other areas of the world (12), and boom-bust dynamics appear in sedimentary records well before human exploitation began (13). These facts have led to the hypothesis that changing environmental conditions drive sardine crashes and/or shifts in distribution, al- though fishing pressure has likely exacerbated global crashes in the recent past (3, 14). In the CCE, sardine recruitment seems to peak when and where upwelling is intermediate (15). In the CCE, intermediate upwelling is associated with low ($14-15^oC$) and high ($>20^oC$) temperatures. The association between high sardine recruitment and warm episodes is puzzling, because warm episodes are associated with low productivity. This sparked a search for potential mechanisms linking water temperature and sardine abundance. One possible mechanism is that release from larval predation during warm periods outweighs the scarcity of food (16). Alternatively, sardine may respond to offshore upwelling driven by wind-stress curl, as opposed to coastal up-welling (17).

Motivated by these studies, Jacobson and MacCall (18) sought a quantitative relationship between sardine and temperature. They found a statistical correlation between log reproductive success (one way of quantifying recruitment) and the 3-year average of the Scripps Institution of Oceanography (SIO) pier sea surface temperature (SST). They then verified the relationship using a modeling approach based on generalized additive models and formulated a best-management strategy for sardine that incorporated the influence of sea temperature. In light of these findings, the Pacific Fishery Management Council modified the sardine management plan to explicitly account for SST (19). Recently, McClatchie et al. (11) repeated the correlation analysis of Jacobson and MacCall (but not the modeling approach) with the addition of 17 years of new data. They found that the statistical correlation between recruitment and SST is no longer significant when the newer data are included. They concluded that the SIO pier temperature does not influence sardine dynamics, and the temperature-based control rule was subsequently rescinded.

Another explanation for these new findings is that temperature does influence sardine, but in a state-dependent way, causing mirage correlation between temperature and sardine. Indeed, we can directly test for state dependence in sardine population dy- namics using S-maps (20). With S-maps, a family of models that range from linear ($\theta = 0$) to highly nonlinear ($\theta \geq 1$) is used to forecast the target time series. For the sardine ichthyoplankton time series, nonlinear (state-dependent) models produce better forecasts than linear models, suggesting that sardine have state-dependent dynamics (Fig. 3.6 and 3.7). Furthermore, Sugihara et al. (7) analyzed historical landings of sardine and SIO pier SST with convergent cross-mapping (an alternative to correlation that detects state-dependent interactions) and found that the SIO pier temperature does affect sardine.

Consequently, the failure of the sardine-temperature correlation to hold up to retesting does not mean that the SIO pier temperature is irrelevant to sardine. Rather, the

problem requires an analytical framework specifically suited to ecological systems with interdependent parts. To this end, the recent work (7) with convergent cross-mapping detected a cause-effect relationship between temperature and sardine. However, it is not sufficient for management to know whether a physical variable is affecting a population. It is also necessary to predict how environmental conditions will affect population dynamics. Multivariate SSR can address this need.

Thus, we seek to demonstrate multivariate SSR as a practical tool for understanding the influence of temperature on Pacific sardine population dynamics. First, we present a conceptual overview of multivariate SSR and use simple models to demonstrate the methods. We then apply the method to ichthyoplankton time series of Pacific sardine. We show that when the population dynamics of sardine are treated nonlinearly, the SIO SST and other broad-scale indicators can improve forecasts of year-to-year changes in abundance. We illustrate how multivariate SSR can predict the effect of changes in SIO SST on sardine and is thereby a useful tool for exploring possible temperature scenarios. Finally, models show how these methods can predict the outcomes of harvest targets under different climate regimes.

3.1.1 Primer on Multivariate State Space Reconstruction

SSR is based on the theory of dynamic systems. If a time series variable $X(t)$ is part of a dynamic system, a set of rules (e.g., a system of difference or differential equations) dictates how X changes in time based on the current value of X and the variables that interact with X. As an example, consider three species, X, Y, and Z, modeled with coupled logistic equations that represent a three- species food chain (21) (Appendix gives a full description). The differential equations dictate how the abundance of each species changes in time given the current state of each population.

A typical way to view the system is as separate time series of each species,

as shown for species X in Fig. 3.1D. However, we can also view the system in a multidimensional space, taking the abundances of species X, Y, and Z as the coordinate axes. This defines the state space of this simple ecosystem. Viewed in the state space, each time point corresponds to a vector and the changes in the ecosystem over time (abundances of the three populations) form a trajectory (Fig. 3.1A). Two time points nearby in the state space represent two similar states of the ecosystem and the populations will follow similar trajectories as time progresses.

Often in ecology only one or a few variables are actually measured. In these cases, Takens's theorem and its multivariate generalization (8, 22, 23) show that it is still possible to represent the system dynamics in a state space by substituting time lags of the measured variables [e.g., $X(t-1)$] for the unknown variables as coordinates. In effect, the information in the unobserved variables is encoded in the observed time series, and so a single time series can be used to reconstruct the state space. This gives a time-delayed coordinate representation (or embedding) of the system trajectories. Fig. 3.1B and C illustrate univariate (using lags of only one variable) and multivariate (using lags of two or more variables) embeddings for the three-species model. Importantly, the trajectories in the reconstructed state space are uniquely matched to trajectories in the original state space, and states (vectors) that are neighboring in one space are also neighboring in the others. Observe that at time t_1, t_2, and t_3 (shown in red) the ecosystem is in a similar state based on all three pictures.

One basic application of SSR embeddings is forecasting. Because vectors nearby in state space evolve similarly in time, the future abundance at one time point can be predicted based on the behavior of its nearest neighbors in the reconstructed state space. In this paper, we use two different types of forecasts. The first is simplex projection, where a weighted average is taken over the $E+1$ nearest neighbor vectors in the reconstructed state space (24) (E is the dimension of the state space). The second is S-maps (20, 25),

where a linear model is fit for each observed vector in the reconstructed state space using all of the remaining vectors (cross-validation). However, the vectors nearby the target in state space are given greater weight, controlled by the nonlinearity parameter θ, and so S-maps can give either linear ($\theta = 0$) or nonlinear ($\theta > 0$) forecasts. By comparing the performance of locally weighted (nonlinear) forecasts to the global linear forecasts ($\theta = 0$), S-maps can be used to test for nonlinear dynamics. A significant increase in forecast skill for the locally weighted model is taken as evidence of nonlinear dynamics (state dependence). Note that simplex projection can be the better tool for exploratory analysis (less possibility for overfitting), but S-maps can ultimately give better forecasts. A detailed mathematical description of these techniques with summary code is given in Appendix. Interested parties are encouraged to contact the authors regarding software and guidance for analysis.

3.1.2 Validating Multivariate Embeddings

Multivariate embeddings that use two variables X and Y should make good forecasts only if Y and X are interacting parts of the same system. This suggests that SSR forecasting can be used to check for interactions between variables (7). Each SSR embedding is a different representation of the same fundamental system. *A priori* it is not possible to know whether one particular choice of lagged variables will give better forecasts than another (8). Stochastically forced systems, however, are a special case. For a purely stochastic forcing variable Y, the current state cannot be inferred from past values of the forced variable. The only way to include this environmental information in SSR forecasts of X is to have $Y(t)$ as a coordinate variable, and nearest neighbor forecasts based only on lags of $X(t)$ will necessarily have greater uncertainty. By this logic, if a stochastic variable Y has an effect on X, then adding Y (with the appropriate lag) should always improve univariate forecasts of X. In this way, comparing multivariate

to univariate SSR forecasts can identify stochastic driving variables.

3.1.3 Scenario Exploration with Multivariate SSR

In ecology we are interested not only in knowing whether two variables interact, but also how they interact. Consider a Ricker model for a population S (Eq. 1). In this case, the model is exactly known, and we can simply calculate $S(t+1)$ for different hypothetical past temperatures T_1, T_2, and so on, to understand the effect temperature has on the population. When studying real populations, however, the true model is not known. Scenario exploration with multivariate SSR is a way to explore climate effects in real systems without assuming a particular model structure. Scenario exploration involves constructing a multivariate embedding to predict $S(t+1)$ using different values of T to explore the effect of temperature on the stock.

To demonstrate scenario exploration, we begin with toy models, which allow us to compare predictions based on multivariate SSR to calculations with the exact model. The temperature-driven Ricker model given by Eq. 1 in Materials and Methods is a simple model of a population that has nonlinear dynamics and is driven by temperature. We explore the effect of temperature, T, on the stock, S, using the multivariate embedding $\langle S(t), S(t-1), S(t-2), S(t-3), T(t)\rangle$. For each t, we predict the effect that an increase in past temperature $\Delta T = 0.1\sigma T$ (10% of the standard deviation of the temperature time series) would have on the population abundance the following year. That is, we use simplex projection to make a nearest-neighbor forecast of $S(t+1)$ for the state space vector $[S(t), S(t-1), S(t-2), S(t-3), T(t)+\Delta T]$. Fig. 3.2A shows the predictions of SSR scenario exploration plotted against the true values calculated from the model for 10 time series of 50 y each with different initial conditions and realizations of T and ε_{proc}. The result is robust to a wide range of growth rates (Fig. 3.8).

Importantly, SSR methods can be applied in both single- and multi-species con-

texts, even when there are no records of the other interacting species. As a demonstration we repeat the analysis above on the three-species extension of the basic Ricker model defined in Eq. 2. However, we only use the time series of the target species S_1 and the temperature T to do scenario exploration. Fig. 3.2B shows that scenario exploration can still predict the effect of temperature on a population in a multi-species complex, even if the other species are unobserved. For both the single- and multispecies models, the correlation between SSR predictions and model calculations is high: $\rho = 0.75$ and 0.86, respectively.

3.2 Results

We use multivariate SSR methods to determine whether and how sardine are affected by their environment. We first apply multivariate SSR to sardine data to verify that the SIO pier SST influences sardine dynamics and determine whether it is the best single environmental indicator variable. Table 1 displays the improvement in forecast skill ΔF of multivariate embeddings using the SIO pier SST and other environmental indicators. Including SIO pier SST in embeddings improves forecasts with simplex projection, indicating that SIO SST influences sardine, and ΔF is significantly greater than can be explained by the null model ($p < 0.05$). Several other variables show positive ΔF, including the Pacific Decadal Oscillation (PDO), North Pacific Gyre Oscillation (NPGO), and the Southern California Bight (SCB) satellite SST, suggesting these are also relevant to Pacific sardine population dynamics. Of these, only the PDO is significant ($p < 0.05$). The PDO and SIO SST are highly correlated, so this is not surprising. The method suggests three variables that are least likely important to sardine dynamics: Newport Pier SST, North Pacific Index (NPI), and Southern Oscillation Index (SOI); each of these has a strong negative effect on forecasting.

To explicitly test the notion that the influence of environment (SST) on sardine depends on population state, we compare the out-of-sample forecasts of Pacific sardine ichthyoplankton with a univariate linear forecasting model (using only lags of sardine) to both linear and nonlinear multivariate forecasting models that also account for temperature. For simplicity, all models use S-maps, which can be made linear or nonlinear by adjusting the nonlinear tuning parameter, θ. The base univariate linear ($\theta = 0$) model is equivalent to an order-E autoregressive (AR) model, whereas the multivariate linear ($\theta = 0$) model has an added linear term for the SIO pier SST. The nonlinear S-map model uses the same dimensions as the multivariate linear model, but the nonlinear parameter $\theta > 0$ is fit in-sample. Fig. 3.3 shows the percentage reduction in error of the multivariate models over the univariate linear base model. Error is quantified as mean absolute error (Fig. 3.3 left) or root-mean-squared error (right). Using either measurement, the nonlinear model that incorporates temperature does best. Furthermore, linearly including temperature increases forecasting error (percentage improvement in error is negative in both plots). Thus, accounting for temperature only improves forecasts when the model is state-dependent (nonlinear).

Because multivariate attractor reconstructions using either SIO SST or PDO give good predictions of sardine abundance, these embeddings can be used for scenario exploration to understand how past conditions affected sardine and to predict the response of sardine to future environmental scenarios. As an illustration, we do scenario exploration using the multivariate embedding that includes SIO pier SST, $\langle X(t), X(t-1), X(t-2), SST(t)\rangle$. For each historical observation, $X(t)$, we explore how $X(t+1)$ would have differed if temperatures were warmer or cooler by half the SD σ_{SST} of the historical pier temperature. That is, we predict $X(t+1)$ for the state $[X(t), X(t-1), X(t-2), SST(t) \pm \Delta T]$, where $\Delta T = 0.5\sigma_{SST}$. Fig. 3.4 shows the multivariate SSR prediction for sardine with the historic temperature (yellow circles), increase

in temperature (red triangles), and decrease in temperature (blue triangles), along with the ob- served time series (dashed black line). Note that SSR predictions were made of first differences and converted back into raw ichthyoplankton abundance.

3.3 Discussion

We demonstrated the application of multivariate SSR to studying physical-biological interactions. The predictions obtained do not assume a particular functional form for the environmental interaction; they rely only on the choice of the embedding dimension and environmental variable(s). This multivariate embed- ding can then be used in place of model equations to understand and predict the effect of changes in a driving variable on population dynamics.

Scenario exploration with multivariate embeddings motivates the development of new adaptive management schemes based on short-term forecasting. If time series of a human control variable (e.g., fishing effort or mortality) are available, management could be based on scenario exploration with multivariate embeddings that include abundance, temperature, and the control variable. Simultaneously exploring different temperature and harvesting scenarios could then reveal how temperature affects the relationship between fishing and future biomass. As an *ad hoc* illustration of this idea, take the Ricker model described by Eq. 3, which explicitly includes temperature and fishing mortality. Fig. 3.5 shows the results of using scenario exploration at a single time point to assess the relationship between fishing mortality and next year's abundance when the temperature increases by $\Delta T = 0.5\sigma_T$ (half the SD of T) from the previous year (red), remains constant (green), or decreases by $\Delta T = 0.5\sigma_T$ (blue). The results of scenario exploration (filled squares) closely match the calculations made with the exact model (dotted lines). This type of scenario exploration would tell managers the permissible level of fishing

given a target biomass and recent ocean conditions. This approach is analogous to the greedy heuristic strategies that are used in high-dimensional approximate optimization problems (26). Of course, considerable work remains to be done to develop and evaluate management plans based on these methods.

More immediately, these methods offer a data-motivated perspective on the dynamic relationship between sardine and environmental conditions in the CCE. We find that including either SIO SST or the PDO in the multivariate embedding improves forecast skill (ρ) by roughly 30%. These indicators reflect environmental forcing of sardine population dynamics that is not captured by more traditional methods. We conclude that environmental considerations should in fact play an important role in Pacific sardine management. Additional analyses demonstrating the influence of SIO SST on Pacific sardine landings using related SSR methods (7) support this assertion.

Our results further suggest that good management must reflect the complexity underlying the interaction between environmental changes and population dynamics (Fig. 3.3). Exploring the effect of SIO pier SST on sardine using multivariate SSR (shown in Fig. 3.4) suggests that, on average, increasing temperature results in higher forecasts of abundance than decreasing temperature. This corroborates the previous results used in the management plan that warm water is better for sardine. However, our empirical analysis shows that the effect of temperature depends on the specific state of the population. For example, changes in temperature seem to have little or opposite effect in the early years of the time series, when the population is at a much lower abundance, and in later years with very large abundance (e.g., 1999-2001). This suggests any temperature-sensitive control rule for sardine should be different at low, intermediate, and high sardine abundances.

Like other nonparametric methods, SSR benefits from greater time series length. Although there is no absolute rule regarding length, forecast skill improves with time series length (7); for fisheries it is usually difficult to get significant predictability with

time series shorter than 30 points. Therefore, management applications of SSR will be best suited to fisheries with longer time series. Even when time series are short, however, good predictability is possible by combining data from similar fisheries (27). For practical application to management, it is also important to acknowledge measurement error. State space forecasting is robust to measurement error (28, 29). Furthermore, we can consider extending the deterministic simplex projection (or S-map) to incorporate an estimated distribution for measurement error for each time series point. This would be extended to spatial distributions for the state vectors in the attractor reconstruction and ultimately give distributions for the forecast.

Note that we expect greater error in scenario exploration when temperature scenarios exceed the bounds of historic data. In these cases, there are no appropriate historical observations of the dynamics on which to base forecasts. The more extreme the scenario, the greater the chance of encountering dynamics never previously recorded. Parameterized methods have a potential advantage over nonparametric methods in extrapolating beyond observed behavior. However, there is uncertainty in extrapolation with any method. The grim reality of anthropogenic climate change and overfishing is that as we continue stressing ecosystems, we catalyze ecological outcomes that are increasingly surprising but less and less predictable from historical observations. Regardless of the analytical approach, the precautionary principal is critical.

Moving forward, scenario exploration with multivariate SSR can be applied to understand relationships between two (or more) dynamically interacting variables. Here, we focused on the case of abundance and temperature. It would also be possible to use catch or abundance of another species as the second variable and explore the ecological impact of harvesting one stock on another. Demonstrating the feasibility of using SSR in a multi-species context for a particular fishery is an important potential capability of these methods that remains to be shown.

Given that climate change and overfishing are dual threats to many marine species, it is critical to consider environmental conditions when formulating management strategies. Our analysis with multivariate SSR illustrates the value of explicitly considering complex dynamics in population responses to environmental forcing. For Pacific sardine, there is a clear dependence on the physical environment that can be captured using available, broad-scale indicators of conditions in the CCE. These general methods can offer immediate insight into environmental forcing of species with long time series and suggest a promising avenue of research into operational schemes for implementing ecosystem-based management.

3.4 Materials and Methods

3.4.1 Variable Identification

We compare the forecasting skill of the univariate state space

$$\langle X(t), X(t-1), \ldots, X(t-(E-1)) \rangle$$

and the multivariate state space

$$\langle X(t), X(t-1), \ldots, X(t-(E-1)), Y(t-d) \rangle ,$$

where E is the optimal univariate embedding dimension (Appendix) and d is the lag of the environmental forcing. We then evaluate the forecast skill using the correlation coefficient ρ between observed and predicted values, and the relevant quantity for identifying variables is $\Delta F = \rho_{multi} - \rho_{uni}$. A physical variable is considered relevant if ΔF is significantly greater than 0. To test significance, we use the Ebisuzaki randomization

procedure to shuffle the environmental time series values while preserving the spectrum (30); this destroys any temporal relationship with the biological time series. For our analysis of Pacific sardine, we compute ρ_{multi} for 500 randomizations of each environmental variable to produce null distributions for ΔF. Note that each environmental variable thus has its own null distribution.

3.4.2 Forecast Comparison

To explicitly investigate state dependence in the influence of temperature on sardine, we compare three S-map forecasts of the Pacific sardine ichthyoplankton. S-maps requires two parameters: the em- bedding dimensions, E (number of lags to use for forecasts), and the nonlinear tuning parameter, θ. For simplicity, we restrict all three models to use the same number of univariate lags (lags of sardine), but the multivariate models also includes SIO pier SST (averaged over the previous 3 y) as an additional dimension. As is typically done for S-map analysis (20), we use simplex projection to determine E (the optimal number of univariate lags) over the range 1-8. For the linear models, $\theta = 0$. For the nonlinear model, we must also fit $0 < \theta < 10$.

All forecasts were made out-of-sample. We restricted our forecasts to the 22 points in the time series that have eight consecutive lags. So for each target point, we find the E that minimizes the mean absolute error (MAE) of forecasts using univariate simplex projection on the remaining 21 points. We then find θ over the range $[0, 10]$ that minimizes the MAE of multivariate S-map forecasts using cross-validation over the remaining 21 points. Finally, we use these parameters to forecast the target point out-of-sample.

3.4.3 Model Examples

The most basic Ricker model only shows nonlinear dynamics (as demonstrated for Pacific sardine in Figs. 3.6 and 3.7) at values of the growth parameter, r, that are considerably higher than those usually fit in management models. However, when the species dynamics are influenced by process error, nonlinear dynamics arise at considerably lower values of r (31). Here, we include process error as additive with the growth rate, giving

$$S(t+1) = S(t)\exp\left[(r+\varepsilon(t))(1-S(t))\right]\exp\left(\psi T(t)\right) \tag{3.1}$$

where $\varepsilon(t)$ is a normally distributed random variable with $mean(\varepsilon) = 0$ and $SD(\varepsilon) = 0.2$. The temperature $T(t)$ was modeled as red noise by applying a 10-y averaging window to white noise. For Fig. 3.2A, we set $r = 2$ and $\psi = 0.3$.

Nonlinear dynamics can also arise in the Ricker model when the interactions between multiple species are considered. Thus, we also used the following three-species extension of the Ricker model:

$$S_i(t+1) = S_i(t)\exp\left[r_i\left(1-\sum_{j=1}^{3}\alpha_{ij}S_j(t)\right)\right]\exp\left(\psi_i T(t)\right), \tag{3.2}$$

$$\alpha = \begin{bmatrix} 1 & 0.2200 & 0.3490 \\ 0.0455 & 1 & 0.2670 \\ 0.3576 & 0.2900 & 1 \end{bmatrix}$$

$$r = \begin{bmatrix} 2.2327 \\ 1.8287 \\ 1.8209 \end{bmatrix}$$

$$\psi = \begin{bmatrix} 0.3 \\ 0 \\ -0.3 \end{bmatrix}$$

for $i = (1,2,3)$. Note that the growth rates r_i were drawn randomly from the interval $[1.5, 2.5]$ and the non-diagonal elements of the interaction matrix α were drawn randomly from $[0, 0.5]$.

For Fig. 3.5 we include a term for fishing mortality, F, in Eq. 1 as follows:

$$S(t+1) = S(t)\exp[(r+\varepsilon(t))(1-S(t)) - F(t)]\exp(\psi T(t)) \tag{3.3}$$

where $F(t)$ is modeled as a uniformly distributed random variable over the range $[0, 0.3]$. As in Fig. 3.2A, we set $r = 2$ and $\psi = 0.3$. To explore how temperature conditions change the effect of harvesting on the population, we did scenario exploration with the multivariate embedding $\langle S(t), S(t-1), S(t-2), S(t-3), T(t), F(t)\rangle$. The three temperature scenarios we illustrate are $T(t) = T(t-1) + 0.5\sigma_T$, $T(t) = T(t-1)$, and $T(t) = T(t-1) - 0.5\sigma_T$ (σ_T is the SD of the modeled temperature). The figure was made using a 50-y time series to predict the behavior at time $t = 50$. To get the best forecasts, we use a strongly nonlinear S-map rather than simplex projection. For simplicity, we specified $\theta = 5$ rather than explicitly fitting θ for this model example.

3.4.4 Data

We used five climate indices that have been linked to biological changes in the California Current: the SOI (32), NPI (33), PDO (34), NPGO (35), and Multivariate El Nino/Southern Oscillation Index (36). We also used daily temperature data from the SIO Pier in La Jolla, CA and the City of Newport Beach Pier in Newport Beach, CA (http:

//shorestation.ucsd.edu/data/index_data.html). For the SIO Pier, we used both surface (0 m) and bottom (5 m) temperatures. Finally, we used an index for SST in the Southern California Bight based on National Oceanic and Atmospheric Administration Extended Reconstructed Sea Surface Temperature v3 analysis (http://www.esrl.noaa.gov/psd/) averaged over four contiguous $2^o \times 2^o$ areas, following McClatchie et al. (11). All environmental indicators were averaged over a 3-y window and normalized to match the original analysis (18), but not first-differenced. For time series with daily resolution, this meant a straight daily average. For the other time series, the average was taken over monthly values.

Time series for *S. sagax* were derived from the CalCOFI ichthyoplankton surveys as in Hsieh (5) from 1950 to 2007. CalCOFI ichthyoplankton abundance provides an index of adult spawning stock biomass (37). For the normalized first differences of the sardine ichthyoplankton data, we determined the best embedding dimension to be $E = 3$ (see Appendix).

Acknowledgements

This work was funded by National Science Foundation (NSF) Grant DEB1020372, NSF-National Oceanic and Atmospheric Adminis- tration Comparative Analysis of Marine Ecosystem Organization (CAMEO) program Grant NA08OAR4320894/CAMEO, an Environmental Protection Agency Science to Achieve Results fellowship, the NSF Graduate Research Fellowship Program, a National Marine Fisheries Service/Sea Grant Population Dynamics Fellowship, the Sugihara Family Trust, the Deutsche Bank- Jameson Complexity Studies Fund, the McQuown Chair in Natural Sciences, University of California, San Diego, and National Taiwan University, National Science Council of Taiwan.

Chapter 3, in full, is a reprint of the material as it appears in Proceedings of the National Academy of Science USA, 2012. Deyle, Ethan Robert; Fogarty, Michael; Hsieh, Chih-hao; Kaufman, Les; MacCall, Alec D.; Munch, Stephan B; Perretti, Charles T.; Ye, Hao; and Sugihara, George. The dissertation author was the primary investigator and author of this paper.

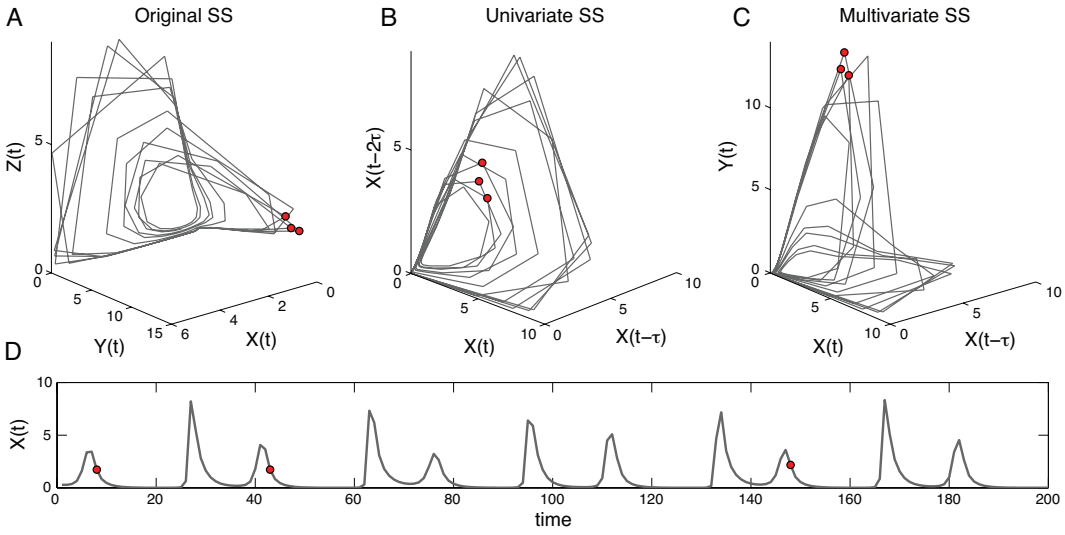

Figure 3.1: State space reconstruction demonstrated with a three-species logistic model. (A) The trajectory of the model ecosystem is plotted in three dimensions using the abundances $X(t)$, $Y(t)$, and $Z(t)$ as coordinates. (B) The trajectory is shown in the univariate SSR coordinates, $\langle X(t), X(t?1), X(t?2)\rangle$, where lags of X take the place of the other variables. (C) The trajectory is shown in multivariate SSR coordinates, $\langle X(t), X(t?1), Y(t)\rangle$. The system is in a similar state (nearby in state space) at times t_1, t_2, and t_3 (red) based on all three representations. Hence, any of these state spaces (original or reconstructed) can be used for nearest- neighbor forecasting. (D) The abundance of species X is shown as a time series

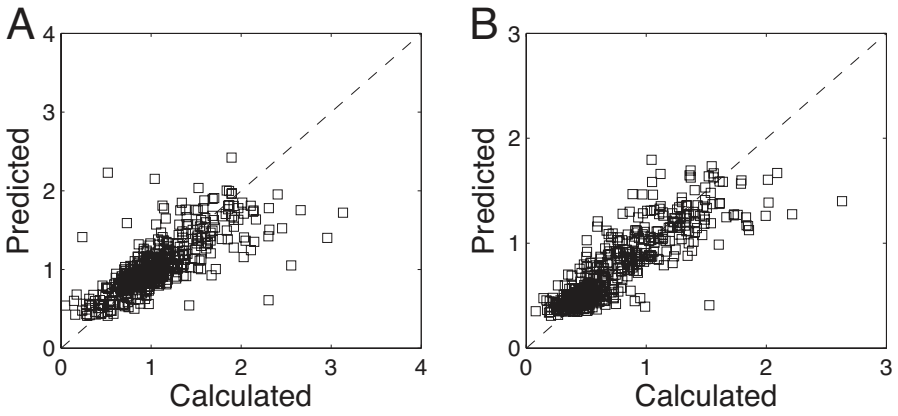

Figure 3.2: Scenario exploration illustrated for a short (50-y) time series generated with known models that are forced by temperature. For each time series point t, the effect of warming on $S(t+1)$ is predicted with multivariate SSR for warming of $\Delta T = 0.1\sigma_T$ (10% of the SD of T). SSR predictions are compared with calculations with the known model for (A) a single-species Ricker model (Eq. 1) and (B) species 1 in a multispecies Ricker model (Eq. 2) for 10 model realizations.

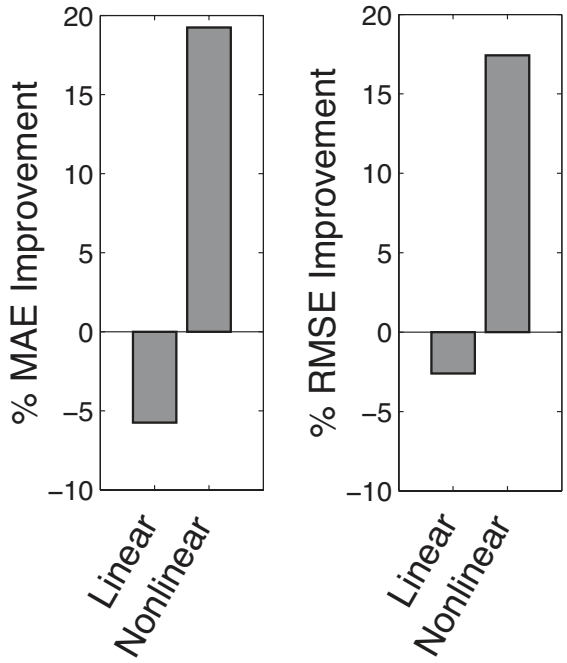

Figure 3.3: Linear and nonlinear forecasting models for Pacific sardine icthyoplankton that include SIO pier SST are compared with a base univariate linear model. The left plot shows the percent improvement in mean absolute error (MAE) of the out-of-sample forecasts for each model over the base model, $(MAE_{base} - MAE_{model})/MAE_{base}$. The right plot shows the same, but using $RMSE$. The base model errors are $MAE = 0.99$, $RMSE = 1.47$ (normalized to the SD).

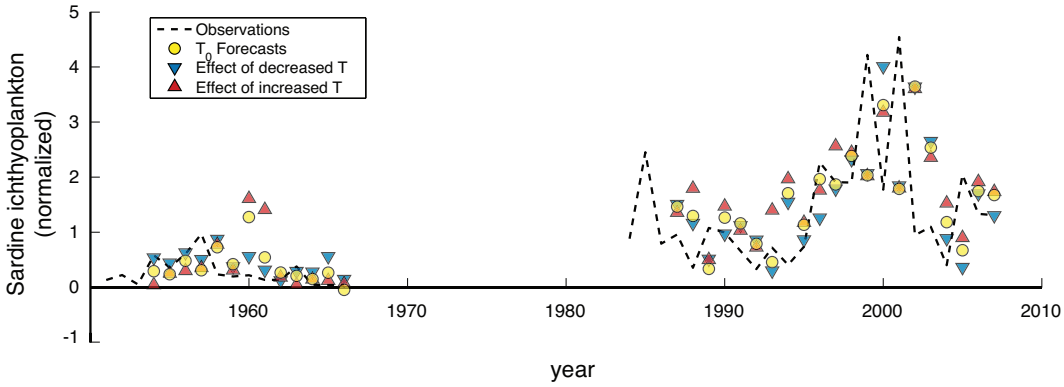

Figure 3.4: Effect of warming and cooling on sardine population calculated using scenario exploration. Using the multivariate embedding based on lags of sardine abundance and SIO pier SST (Table 1), we explore the effect of perturbing the historical temperatures (averaged over 3 y) by $\Delta T = \pm 0.5\sigma_{SST}$ on sardine abundance in the following year. The historical time series is shown as a black dashed line. Multivariate SSR forecasts based on historical temperatures (yellow circles), warming by $\Delta T = +0.5\sigma_{SST}$ (red triangles), and cooling by $\Delta T = -0.5\sigma_{SST}$ (blue triangles) are shown.

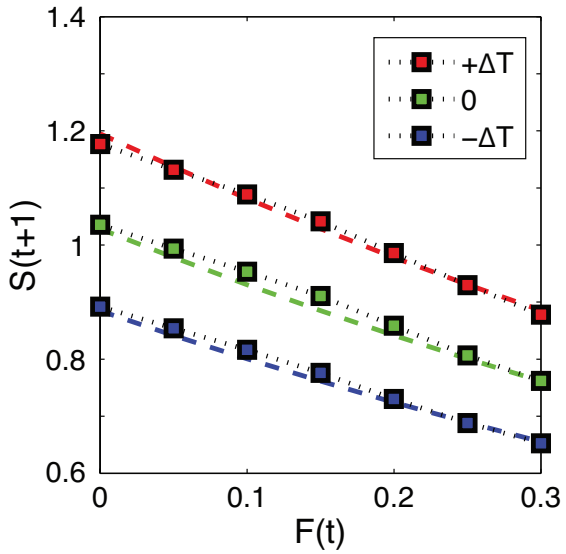

Figure 3.5: Model illustration of how simultaneous scenario exploration over temperature and fishing mortality might hypothetically be used in management. The effect of fishing mortality F on future biomass $S(t+1)$ is plotted for three temperature scenarios: T increases by $0.5\sigma_T$ (red), remains constant (green), or decreases by $0.5\sigma_T$ (blue). Even with just a 50-y time series, multivariate SSR predictions (filled squares) closely match the model calculations (dotted lines).

3.5 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.

2. Ottersen G, Hjermann DO, Stenseth NC (2006) Changes in spawning stock structure strengthen the link between climate and recruitment in a heavily fished cod (Gadus morhua) stock. Fish Oceanogr 15(3):230-243.

3. Hsieh C-H, Reiss CS, Hewitt RP, Sugihara G (2008) Spatial analysis shows that fishing enhances the climatic sensitivity of marine fishes. Can J Fish Aquat Sci 65(5):947-961.

4. Hsieh C-H, Yamauchi A, Nakazawa T, Wang W-F (2009) Fishing effects on age and spatial structures undermine population stability of fishes. Aquat Sci 72(2):165-178.

5. Hsieh C-H, Reiss CS, Watson W, Allen MJ, Hunter JR, Lea RN, Rosenblatt RH, (2005) A comparison of long-term trends and variability in populations of larvae of exploited and unexploited fishes in the Southern California region: a community approach. Progress In Oceanography 67:160-185.

6. Myers RA (1998) When do environment-recruitment correlations work? Rev Fish Biol Fish 8(3):285-305.

7. Sugihara G, May R, Ye H, Hsieh CH, Deyle E, Fogarty M, Munch S (2012) Detecting causality in complex ecosystems. Science 338:496-500.

8. Deyle ER, Sugihara G (2011) Generalized theorems for nonlinear state space reconstruction. PLoS ONE 6(3):e18295.

9. Dixon PA, Milicich MJ, Sugihara G (1999) Episodic fluctuations in larval supply. Science 283(5407):1528-1530.

10. Pascual M, Ellner SP (2000) Linking ecological patterns to environmental forcing via nonlinear time series models. Ecology 81(10):2767-2780.

11. McClatchie S, Goericke R, Auad G, Hill K (2010) Re-assessment of the stock-recruit and temperature-recruit relationships for Pacific sardine (Sardinops sagax). Can J Fish Aquat Sci 67(11):1782-1790.

12. Lluch-Belda D, Schwartzlose RA, Serra R, Parrish R, Kawasaki T, Hedgecock D, Crawford R (1989) World-wide fluctuations of sardine and anchovy stocks: The regime problem. S Afr J Mar Sci 8(1):195-205.

13. Baumgartner TR, Soutar A, Ferreira-Bartrina V (1992) Reconstruction of the history of Pacific sardine and northern anchovy populations over the past two millennia from sediments of the Santa Barbara Basin, California. CCOFI Rep 33:24-40.

14. MacCall AD (2009) Climate Change and Small Pelagic Fish (Cambridge Univ Press, Cambridge, UK), pp 1178-1254.

15. Lluch-BeldaD, Lluch-CotaDB, Hernandez-Vazquez S, Salinas-Zavala CA, Schwartzlose RA (1991) Sardine and anchovy spawning as related to temperature and upwelling in the California Current system. CCOFI Rep 32:105-111.

16. Bakun A, Broad K (2003) Environmental ÒloopholesÓ and fish population dynamics: Comparative pattern recognition with focus on El NiŰo effects in the Pacific. Fish Oceanogr 12(4-5):458-473.

17. Rykaczewski RR, Checkley DM, Jr., Checkley J (2008) Influence of ocean winds on the pelagic ecosystem in upwelling regions. Proc Natl Acad Sci USA 105(6):1965-1970.

18. Jacobson LD, MacCall AD (1995) Stock-recruitment models for Pacific sardine (Sardi- nops sagax). Can J Fish Aquat Sci 52(3):566-577.

19. Hill KT, Yaremko M, Jacobson LD, Lo NCH, Hanan DA (1998) Stock assessment and management recommendations for Pacific sardine (Sardinops sagax) in 1997. (Cal- ifornia Department of Fish and Game, La Jolla, CA), Marine Region Administrative Report No. 98-5.

20. Sugihara G (1994) Nonlinear forecasting for the classification of natural time series. Phil Trans R Soc Lond A 348(1688):477-495.

21. Gardini L, Lupini R, Messia M (1989) Hopf bifurcation and transition to chaos in Lotka- Volterra Equation. J Math Biol 27(3):259-272.

22. Takens F (1981) Dynamical Systems and Turbulence, Warwick 1980, Lecture Notes in Mathematics (Springer, Berlin), Vol 898, pp 366-381.

23. Sauer T, Yorke JA, Casdagli M (1991) Embedology. J Stat Phys 65(3-4):579-616.

24. Sugihara G, May RM (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature 344(6268):734-741.

25. Sugihara G, Allan W, Sobel D, Allan KD (1996) Nonlinear control of heart rate vari- ability in human infants. Proc Natl Acad Sci USA 93(6):2608-2613.

26. Powell WB (2010) Approximate Dynamic Programming (Wiley, Hoboken, NJ), 2nd Ed.

27. Hsieh C-H, Anderson C, Sugihara G (2008) Extending nonlinear analysis to short ecological time series. Am Nat 171(1):71-80.

28. Casdagli M, Eubank S, Farmer JD, Gibson J (1991) State space reconstruction in the presence of noise. Physica D 51(1-3):52-98.

29. Perretti CT, Sugihara G, Munch SB (2013) Nonparametric forecasting outperforms parametric methods for a simulated multi-species system. Ecology, 10.1890/12-0904.1.

30. Ebisuzaki W (1997) A method to estimate the statistical significance of a correlation when the data are serially correlated. J Clim 10(9):2147-2153.

31. Sugihara G, Beddington J, Hsieh C-H, Deyle E, Fogarty MJ, Glaser SM,Hewitt R, Hollowed AB, May RM, Munch SB, Perretti C, Rosenberg AA, Sandin S, Ye H (2011) Are exploited fish populations stable? Proc Natl Acad Sci USA 108(48):E1224-E1225, author reply E1226.

32. Trenberth K (1984) Signal versus noise in the Southern Oscillation. Mon Weather Rev 112(2):326-332.

33. Trenberth K, Hurrell JW (1994) Decadal atmosphere-ocean variations in the Pacific. Clim Dyn 9(6):303-319.

34. Mantua NJ, Hare S, Zhang Y, Wallace J, Francis RC (1997) A Pacific interdecadal climate oscillation with impacts on salmon production. Bull Am Meteorol Soc 78(6):1069-1079.

35. Di Lorenzo E, Schneider N, Cobb KM, Franks PJS, Chhak K, Miller AJ, McWilliams JC, Bograd SJ, Arango H, Curchitser E, Powell TM, Riviere P (2008) North Pacific Gyre Oscillation links ocean climate and ecosystem change. Geophys Res Lett 35(8):L08607.

36. Wolter K, Timlin M (1993) Proceedings of the Seventeenth Climate Diagnostics Workshop (US Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, Climate Analysis Center/NMC, Washington, DC), pp 52-57.

37. Moser HG, Charter RL, Watson W, Ambrose D, Hill K,Smith P,Butler J,Sandknop E, Charter S (2001) The CalCOFI ichthyoplankton time series: Potential contributions to the management of rocky-shore fishes. CCOFI Rep 42:112-128.

3.6 Appendix

3.6.1 Three-Species Logistic Model

The general form for a three-species coupled logistic model for species X_1, X_2, and X_3 in continuous time is given by

$$\frac{dX_i}{dt} = X_i \left(r_i + \sum_{j=1}^{3} \alpha_{ij} X_j \right).$$

The parameter values were from figure 5 of ref 6.

$$\alpha = \begin{pmatrix} -0.0020 & -0.4606 & -0.5051 \\ 0.2324 & -0.1920 & 2.3847 \\ 1.2949 & -0.0153 & -0.306 \end{pmatrix};$$

$$r = \begin{pmatrix} 0.9675 \\ -2.4281 \\ -0.9736 \end{pmatrix} = \begin{pmatrix} -0.0020 \\ 0.2324 \\ 1.2949 \end{pmatrix}.$$

The data were generated using a fourth-order Runge-Kutta with integration step of $h = 0.01$ and initial conditions $X_1 = 5$, $X_2 = 0.001$, and $X_3 = 11$. Fig. 4.1 shows data for $t = (101, 102, \ldots, 300)$. The three red points correspond to time indices $t_1 = 108$, $t_2 = 143$, and $t_3 = 248$.

3.6.2 Expanded Model Example of Scenario Exploration

We repeat the analysis shown in Fig. 4.2A of the main text to show that the result is robust over a wide range of the growth rate parameter, r. The model structure is the same (Eq. 1 in the main text):

$$S(t+1) = S(t)\exp\left[(r+\varepsilon(t))(1-S(t))\right]\exp(\psi T(t)),$$

where $\varepsilon(t)$ is a normally distributed random variable with $mean(\varepsilon) = 0$ and $SD(\varepsilon) = 0.2$. As in the analysis in the main text, we set $\psi = 0.3$. However, the growth parameter is varied between 1.8 and 2.8. The temperature $T(t)$ was modeled as red noise with mean 0 and SD 1 by applying a 10-y averaging window to white noise. Using scenario exploration with multivariate SSR, we predict the effect on stock size S of a 10% increase in temperature, ΔT, relative to the SD σT of the temperature time series. We then compare the SSR predictions to the exact calculations with the model. As in Fig. 4.3A of the main text, we use the multivariate embedding

$$\langle S(t), S(t-1), S(t-2), S(t-3), T(t)\rangle$$

that contains lags of population abundance and temperature. We predict the effect that an increase in temperature ΔT at time t would have on the population abundance the following year, $t+1$. That is, we make a nearest-neighbor forecast of the adult SSB for the state

$$\langle [S(t), S(t-1), S(t-2), S(t-3), T(t)+\Delta T\rangle\,.$$

Fig. 4.7 displays the results. For each value of r, we generated 10 time series of 50 y each with different initial conditions and realizations of T and ε.

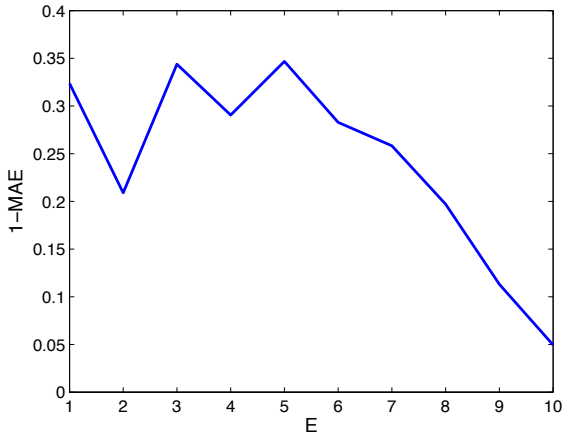

Figure 3.6: Embedding dimension for Pacific sardine. Univariate prediction skill (1 - MAE) is shown as a function of embedding dimension (E) for normalized first differences of the CalCOFI *S. sagax* ichthyoplankton time series. MAE, mean absolute error.

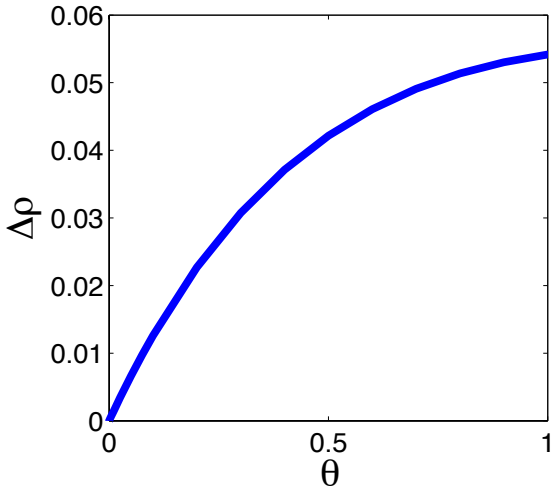

Figure 3.7: S-map analysis of CalCOFI survey abundance of Pacific sardine ichthyoplankton. The improvement in forecast skill, $\Delta\rho$, of sardine ichthyoplankton abundance for nonlinear models compared with the completely linear model ($\theta = 0$) for increasing values of the nonlinear tuning parameter, θ. Here, forecast skill is measured by the PearsonÕs correlation coefficient between model predictions (using cross-validation) and the observed values. Increasing the nonlinear parameter substantially improves forecasts, suggesting that the dynamics of Pacific sardine are state-dependent (nonlinear).

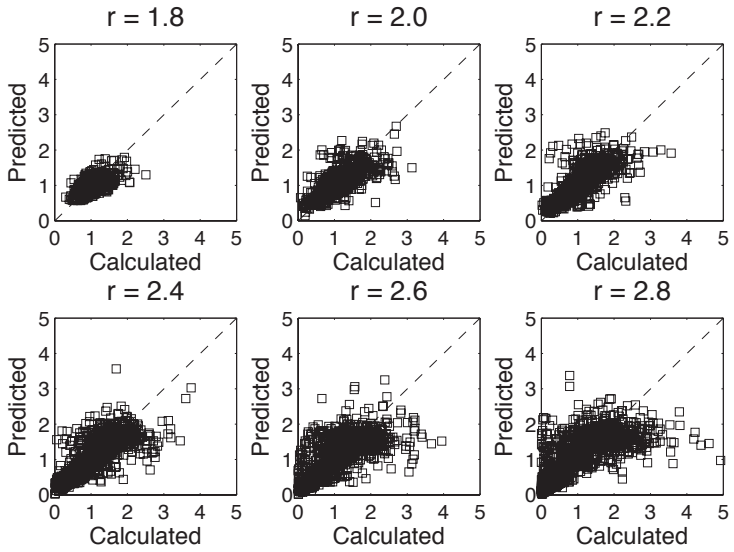

Figure 3.8: Scenario exploration illustrated for short (50-y) time series generated with a simple Ricker model forced by temperature for a range of growth rates, r. For each time series point t, the effect of warming on $S(t+1)$ is predicted with multivariate SSR for warming of $\Delta T = 0.1\sigma T$ (10% of the SD of T). Predictions are compared with the true value calculated with the model over 10 model realizations for each r.

3.7 Appendix References

1. Sugihara G, May RM (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature 344(6268):734-741.

2. Takens F (1981) Detecting strange attractors in turbulence. Dynamical Systems and Turbulence, Warwick 1980. Lecture Notes in Mathematics (Springer, Berlin), Vol 381, pp 366-381.

3. Sauer T, Yorke JA, Casdagli M (1991) Embedology. J Stat Phys 65(3-4):579-616.

4. Deyle ER, Sugihara G (2011) Generalized theorems for nonlinear state space reconstruction. PLoS ONE 6(3):e18295.

5. Sugihara G (1994) Nonlinear forecasting for the classification of natural time series. Philos Trans R Soc Lond A 348(1688):477-495.

6. Gardini L, Lupini R, Messia M (1989) Hopf bifurcation and transition to chaos in Lotka- Volterra Equation. J Math Biol 27(3):259-272.

Chapter 4

Tracking and Forecasting Ecosystem Interactions in Real Time

Abstract

Evidence shows that species interactions are not constant but change as the ecosystem shifts to new states. While 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 that uses available time series data to measure and forecast changing interactions in real systems, and identify the underlying mechanisms. The method is demonstrated with model data, data from a marine mesocosm experiment, and limnologic field data from Sparkling Lake, WI. 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.

4.1 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).While controlled experiments and model investigations demonstrate how varying interactions arise in principle, 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- a requirement 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 particular natural setting. For example, consider two species that occupy the same trophic level. If their diets overlap, we might expect mutual negative competitive effects. However, if their 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 interaction could either be positive or negative, depending on the functional forms of coupling (8) as well as the time scale of effects (9). When predators exhibit prey switching, there can be even more complicated interactions (10). The point is that in nature it is difficult at best to say which of these plausible expectations has in fact occurred.

Indeed, a dearth of quantitative tools has limited our ability to measure interactions as they are realized in natural systems. A few exceptional studies have tracked changes in

food-webs through labor-intensive field methods (11)- e.g., gut content analyses (12). But while these studies verify the complexity of interactions in principle, the high labor costs do not permit practical applications aimed at tracking (much less predicting) continually changing species interactions across many systems. A more practical alternative would be to estimate mutual interactions from time series of abundance that already exist. Although this is more readily scalable, the currently available tools assume linear constancy and are thus limited to studying fixed linear interactions (e.g. (13)) or randomly drifting ones (discussed below) without predictive capacity and without mechanism.

In this paper, we present an approach based on empirical dynamic modeling (EDM) (14-19) that is intended to measure and predict dynamic interactions as they occur in the field from readily available time series data. It is an approach that does not require assumptions of equilibrium or constancy. 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 finally data on zooplankton from Sparkling Lake, WI.

4.2 The Method: Measuring Interactions with S-Maps

The method extends the classical 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 (20,21). It represents a theoretical expedient where pair-wise interactions are treated as fixed coefficients that are thereby independent of each other.

However, ecological systems are rarely if ever in static equilibrium. Rather, they exhibit nonlinear and non-equilibrium behavior, and as a consequence the interaction

strengths (as defined below) and even the sign of the interactions can change with ecosystem state. In other words, interactions can vary with the abundances of interacting species and with changes in important environmental factors.

In principle, interactions for non-equilibrium systems can still be represented with matrices, however to do so it must be possible to recalculate the matrix continually for each successive ecosystem state. Although this may seem infeasible for a system being studied in the field, it is our aim here to show how this task is possible with time series data and empirical dynamic modeling (EDM). EDM is an equation-free mechanistic modeling approach based on the idea of reconstructing the underlying dynamical system (attractor manifold) from observational time series (14-19). The concept of building an attractor manifold from time series is clearly illustrated in a brief video animation http\://simplex.ucsd.edu/RMM_S1.mov and the rationale is as follows.

Briefly, in EDM the state of a dynamical system can be thought of as a location in a multivariate coordinate space, or state space, whose coordinate axes are ecosystem variables such as species abundance, temperature, or resource abundance. The system state changes and evolves in time according to the rules/equations that describe the system dynamics, and this traces out a trajectory. The collection of these time-series trajectories forms a geometric object known as an attractor manifold, which describes empirically (in fact) how variables relate to each other in time- hence "empirical dynamic modeling" (14-19).

To illustrate, Figure 4.1 shows a hypothetical (cartoon) ecosystem consisting of two consumers, C_1 and C_2, competing for a shared resource, R. An empirical attractor for this system can be constructed from the time series of the two consumers and the resource (Figure 4.1a) simply by taking the three time series variables as Cartesian coordinates, $\mathbf{x}(t) = [C_1(t), C_2(t), R(t)]$ and plotting out successive points in time to yield the system trajectory in the state space (Figure 4.1b).

The dynamics of the first consumer, i.e. how C_1 changes from one time point to another, is a function of the current ecosystem state $\mathbf{x}(t)$ and can be written as

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

where F represents the 3-dimensional dynamics on the attractor with respect to C_1. Note that the current system state $\mathbf{x}(t)$ is also referred to as the target point.

Consider two points on the attractor, **p** and **q**, representing specific ecosystem states. Zooming into a small neighborhood of these points we see the interactions between C_1 and the other variables are nearly linear (Fig. 4.1c & 4.1d), 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 $t+1$ (see Materials and Methods).

$$\mathbf{D}F = \left[\frac{\partial C_1(t+1)}{\partial C_1(t)}, \frac{\partial C_1(t+1)}{C_2(t)}, \frac{\partial C_1(t+1)}{\partial R(t)}\right].$$

The elements of this row (the Jacobian elements) 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 (22). Again, in a system with a globally stable equilibrium, $\mathbf{D}F$ is evaluated only at the equilibrium point fixed point. However, in our hypothetical example the interactions between variables are very different at **p** and **q**. The surface F at **p** (Fig. 4.1d) 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. In contrast, at q, $C_1(t+1)$ is not sensitive to changes in R or C_2. Consequently, F at q is flat (Fig. 4.1c) and the partial derivatives are zero. In this way, the partial derivatives or Jacobian elements corresponding to these slopes define the interaction strengths at system states **p** and **q**.

The key to implementing these ideas is that the Jacobian elements (and thus the

interaction strengths) can be recovered for any predicted variable, at any target point $\mathbf{x}(t)$ on the attractor using S-maps (15,16,18). S-maps is a locally weighted multivariate linear regression scheme, where points on the attractor that are near the target are given the greatest weight. As S-maps is just a specific type of weighted linear regression, it can be easily implemented in common statistical languages like MATLAB and R. Example marked-down R code is provided in the supplement, but we also now briefly outline the procedure.

Like other regression schemes, S-maps involves computing the linear model $\mathbf{C}$ that approximates F locally at $\mathbf{x}(t^*)$, so that

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

where E is the number of system variables (i.e. the embedding dimension). The linear model is fit to the other observed vectors on the empirical attractor manifold, but the vectors that are closer to the target point $\mathbf{x}(t*)$ get stronger weighting. The weighting applied to observation k is exponential by distance to the target point $\mathbf{x}(t*)$

$$w_k = \exp\left(-\theta\,||\mathbf{x}(t_k) - \mathbf{x}(t^*)||/\bar{d}\right).$$

The parameter $\theta \geq 0$ tunes how strongly the regression is localized to the region of state space around the target, and the weights are normalized to the average distance of the observed vectors on the empirical attractor to the target point $\mathbf{x}(t^*)$

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

Here $||\mathbf{x}-\mathbf{y}||$ denotes the Euclidian distance between two vectors. Note that if $\theta = 0$, the S-map model reduces to a vector autogressive (VAR) model. That is, VAR models can be thought of as a special case of S-map. For $\theta \geq 0$, the elements of $\mathbf{C}$ are

the locally weighted linear coefficients, and the larger the parameter θ, the more the coefficients are allowed to vary as the system changes state.

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

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

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 1-step ahead value of the target variable $x_i(t_k + 1)$ given by

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

Since for the target species x_i, it is therefore approximating the partial derivatives of the dynamics with respect to the . Note that with S-maps, the locally weighted linear regression is generally performed using leave-one-out cross-validation, which means that the time point at which we want to measure the interaction coefficients is not used in the fit.

S-maps have been used both as a simple test for nonlinear dynamics (16) and as a non-parametric tool for ecosystem forecasting (e.g. (18)). Here we note simply that in multivariate embeddings (i.e. native embeddings using raw variables rather than lags of a single variable) S-maps approximate the Jacobian or interaction elements at successive points along the attractor (see methods and SI). That is, S-maps generate the relevant Jacobian elements that define the interaction strengths, and as required do so sequentially (hence the name S-map, S = “Sequential”). 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.

The only general assumptions of EDM are that (1) there is a deterministic component to the ecosystem dynamics and (2) the attractor is adequately embedded by the chosen set of coordinate variables (i.e. the state space). If the attractor is not fully embedded, there will be singularities: places where trajectories cross and so the interactions are not uniquely determined. If the attractor is embedded, there are no such singularities. Do note that while we are limited to showing 3-dimensional embeddings in graphical plots, real systems often require more than 3-dimensions (which is easily done with the calculations). Importantly, these two core assumptions can be explicitly validated by nearest-neighbor prediction with S-maps. Prediction means that there are deterministic dynamics and that there are few if any singularities on the attractor (23). Testing these assumptions is a prerequisite for applying the method. Thus, in general EDM should not apply if the system cannot be properly embedded, such as would occur in a purely stochastic system with no discernible dynamics or when observational noise dominates to the extent that no predictive nonlinear manifold can be uncovered. In addition because the multivariate S-map method described here can be sensitive to the specific embedding coordinates, when possible 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 (14)) and a multivariate prediction test as shown below (e.g.(18,19)).

S-maps track the interactions as the system travels through different states (Fig. 4.1e). At point **p** (purple arrow) with steep slopes, the S-map coefficients or partial derivatives show that C_1 is affected strongly by both food $\partial C_1(t+1)/R(t) > 0$ and competition $\partial C_1(t+1)/\partial C_2(t) < 0$. At point **q** (orange arrow), the S-map coefficients are both near 0, reflecting the lack of food-limitation and competition.

Notice that because the S-map coefficients are state dependent they portray the

mechanistic conditions (system state) associated with any particular interaction strength. This is a key point that differentiates the EDM approach from non-mechanistic fitting methods developed in econometrics that allow the coefficients of a vector autoregressive model to drift over time (dynamic linear models). These methods suppose points nearby in time have similar coefficients rather than points nearby in state space (see Fig. 4.5a,b). They treat time variation as random excursions without providing a mechanistic basis for understanding why interactions are changing, and typically require forward information ($\mathbf{x}(t+1)$) to fit the drifting coefficients (e.g., Kalman filters). Consequently the non-mechanistic drift approach lacks predictive power and can fail to correctly measure interactions at all (see supplemental text, Fig. 4.5c).

4.3 Test Cases

To demonstrate the method for tracking state-dependent interactions we will apply it to three test cases: a model, an experimental mesocosm, and a natural lake ecosystem. The model (Fig. 4.2d) is a classic food web (24,25) 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 (see methods) (7,8). The model is reckoned to be a transparent example of state-dependent interactions.

Figure 4.2 shows how the EDM approach with S-maps can uncover the mechanisms that cause the interaction strength to vary, focusing on the interactions between the consumer C_1 and the other ecosystem components. For example, in Figure 4.2b competition between C_1 and C_2 only occurs at low to moderate food levels, and at high food concentrations competition tends to zero. Moreover, Figure 4.2c 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 (26) 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, and validates the approach. Finally, Figure 4.2e shows that the S-map forecasted interaction strength estimated from the time-series agrees well with the Jacobian elements computed directly from the system equations. Importantly, like applications of S-maps to simple population forecasting (27), these S-map predictions of the interaction coefficients are robust to realistic amounts of observational noise (see supplemental online material).

Next, we apply the method to the interaction between calanoid copepods and rotifers in a freely evolving marine mesocosm isolated from the Baltic sea (28,29). We focus on calanoids, rotifers, and their two main prey items, nanoflagellates and picocyanobacteria (see Fig. 4.3d). Figure 4.3a shows the interaction coefficients estimated with S-maps for the effects of rotifers, nanoflagellates, and picocyanobacteria on calanoids. As expected, interactions with the chief prey item (nanoflagellates) are always positive, and interactions with the other grazer (rotifers) are always negative, and the intensity of these interactions changes through time. Similar to the model above, we find that competition ($-\partial Cal/\partial Rot$) is strong only when the prey (nanoflagellate) concentration is near-zero (Fig. 4.3b) and that the maximum strength of competition is set by food limitation (Fig. 4.3c), as demonstrated by the 0.05 quantile regression line (dashed red). Although analogous results have been obtained from experimental evidence of saturating feeding responses (6), here they are recovered non-invasively in the freely-evolving mesocosm, and directly 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 are reasonably in line 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, focusing again on calanoid copepods and trophically similar cyclopoids. Following Beisner et al.(30), we group species by broader taxa- calanoid copepods, cyclopoid copepods, cladocerans, and rotifers. While these are relatively broad groupings, the calanoid and cyclopoid copepod groups were each dominated by single species, Leptodiaptomus minutus and Diacyclops thomasi, respectively, both of which are microplankton grazers.

Figure 4.4a shows the S-map estimates of the time-varying effects of cyclopoids, temperature, and planktivorous fish on calanoids. Note that the effect of cyclopoids on calanoids, $\partial Cal/\partial Cyc$, 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 this were the case, we would also expect competition to occur only when predation pressure is low. Indeed, plotting $\partial Cal/\partial Cyc$ against total predator biomass, Fish, shows that competition only occurs in periods with low planktivorous fish abundance (Fig. 4.4b). Similarly, plotting $\partial Cal/\partial Cyc$ against the predator prey ratio (Fig. 4.4c) 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 (here explicitly showing commensalism)- an interpretation that resonates with previous work showing that Sparkling Lake has been dominated by top-down forcing (30).

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 4.4b & 4.4c is indeed consistent with apparent mutualism mediated by common predators as a mechanism that contributes to the net interaction.

4.4 Concluding Remarks

These three demonstrations illustrate how S-maps can be used to quantify changing species interactions and 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 (Fig. 4.3b,c), as expected. Conversely, in Sparkling Lake we find that competition between copepods only occurs when predator abundance is low (Fig. 4.4b). Much of the time there is a positive net interaction that intensifies when the predator prey ratio increases, suggestive of apparent mutualism.

EDM is an equation-free mathematical framework built on minimal assumptions for studying ecosystem dynamics directly from time-series data. It accommodates non-equilibrium dynamics and does not require assumed functional forms or heuristic models based on correlations. As noted elsewhere (14), such correlations can be inappropriate in nonlinear dynamic systems, which tend to produce “mirage correlations” - ephemeral associations among variables that appear then disappear, or even change sign. As a case in point, in Sparkling Lake where the sign of $\partial Cal/\partial Cyc$ 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 auto regression did not find a significant linear-constant effect of cyclopoids on calanoids (30).

Previous work has shown how EDM can be used for population forecasting

(16,18,27) for exploring alternative environmental scenarios (19), and for detecting causal linkages (14). Here, we apply the S-map approach 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 vetted path, verified by out of sample prediction for studying biological systems as a dynamically changing interconnected whole (14). 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 programs 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 on-going and long-term data collection.

Acknowledgements

Chapter 4, in part, has been submitted for publication of the material as it may appear in Proceedings of the Royal Society B, 2015. Deyle, Ethan Robert; Munch, Stephan B; May, Robert M; Sugihara, George. The dissertation author was the primary investigator and author of this paper.

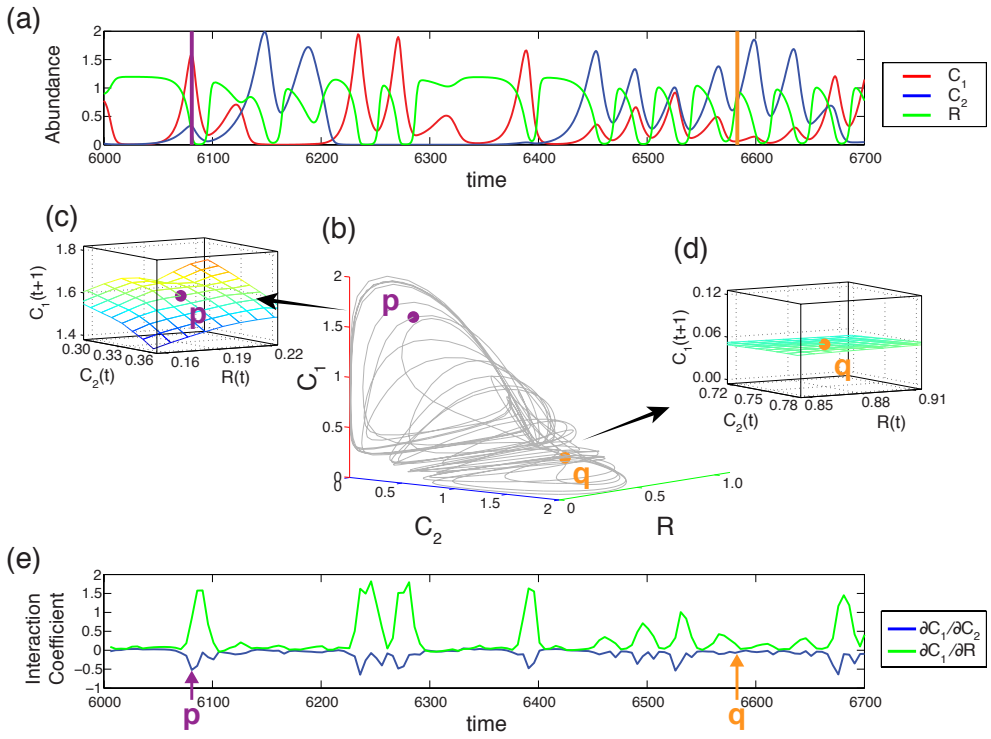

Figure 4.1: Measuring interactions in a hypothetical 3-species ecosystem. The empirical attractor is constructed by re-plotting the time series of C_1, C_2, and R (a) simultaneously in 3 dimensions (b). The attractor displays the historical relationships between variables. The magnitude of the interactions with respect to C_1 are different at the two ecosystem states (points on the attractor), **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, and these are calculated by the S-map coefficients (e). The surface at **p** is steep (c), thus the estimated interaction coefficients (e) have large magnitude (purple arrow). Conversely, the surface at **q** is flat (d), so the estimated interaction coefficients (e) are near zero (orange arrow).

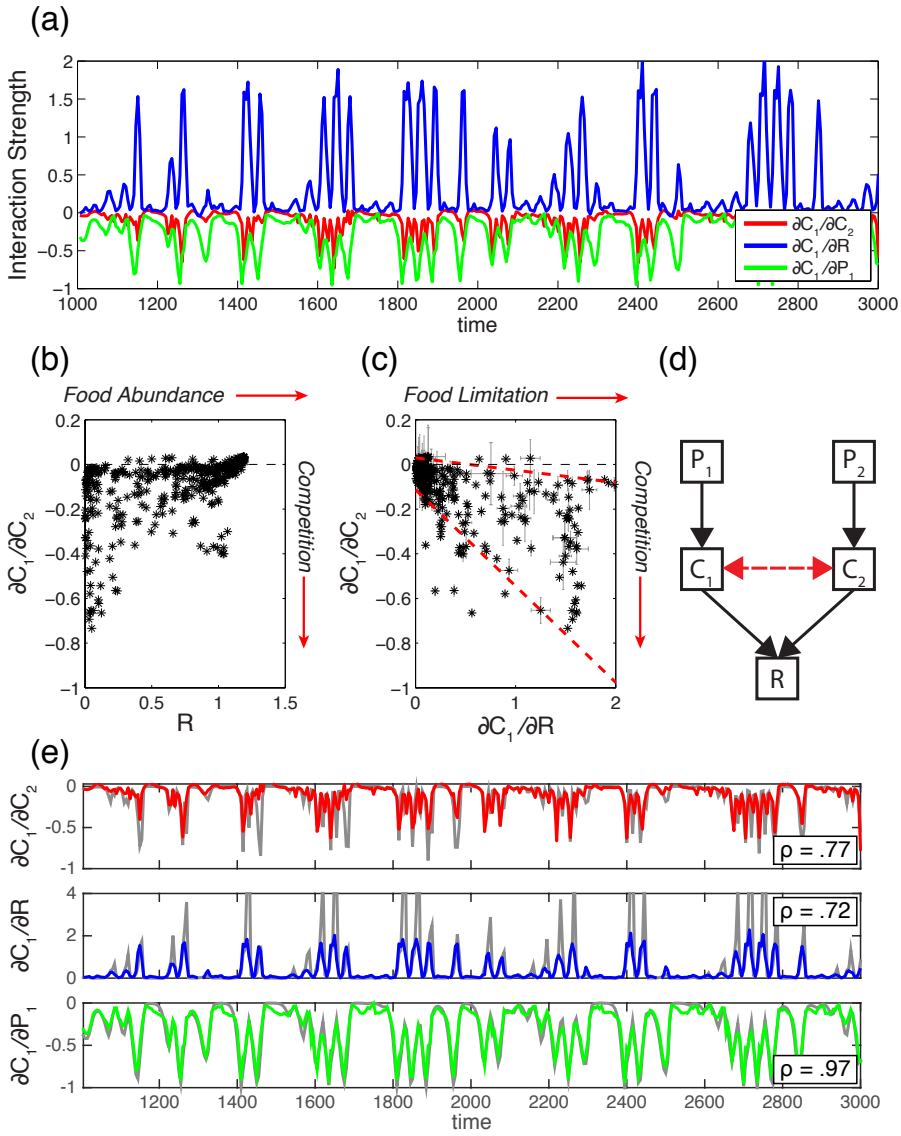

Figure 4.2: Dynamic interactions measured from the output of a 5-species model food web (d). Panel (a) shows the S-map estimated interaction coefficients over 1000 model years for the effects on consumer 1 (C_1) of the shared resource (R), predator 1 (P_1), and of consumer 2 (C_2) on C_1. Panels (b) and (c) show 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. Consumers and predators have saturating feeding responses, which leads to state-dependent dynamics and competition between consumers (red-arrow) that depends on food limitation. Panel (e) compares the estimated S-map forecasts of interaction strength to those computed directly from the model equations (correlation coefficients given in figure). The demonstrated ability to forecast interactions is important for system control and management. All axes are in normalized units.

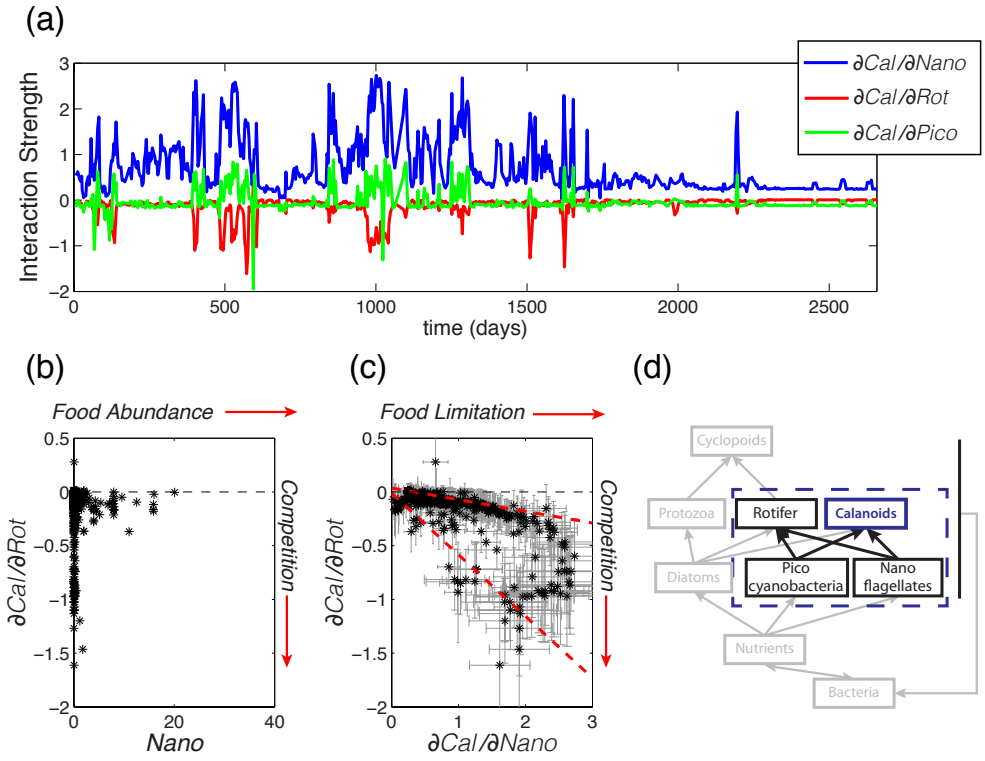

Figure 4.3: Dynamic interactions in the Baltic Sea mesocosm. (a) The S-map estimated interaction coefficients for calanoid copepods (*Cal*) with respect to the main prey item, nanoflagellates (*Nano*); secondary prey item, picocyanobacteria (*Pico*); and main competitor, rotifers (*Rot*) through the duration of the experiment. The effect of rotifers $\partial Cal/\partial Rot$ is shown as a function of (b) food abundance (*Nano*) and (c) food limitation ($\partial Cal/\partial Nano$), 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 -$\partial Cal/\partial Rot$) between rotifers and calanoids. Panel (d) shows the focal species within a summary interaction network for the mesocosm. Axes are in normalized units.

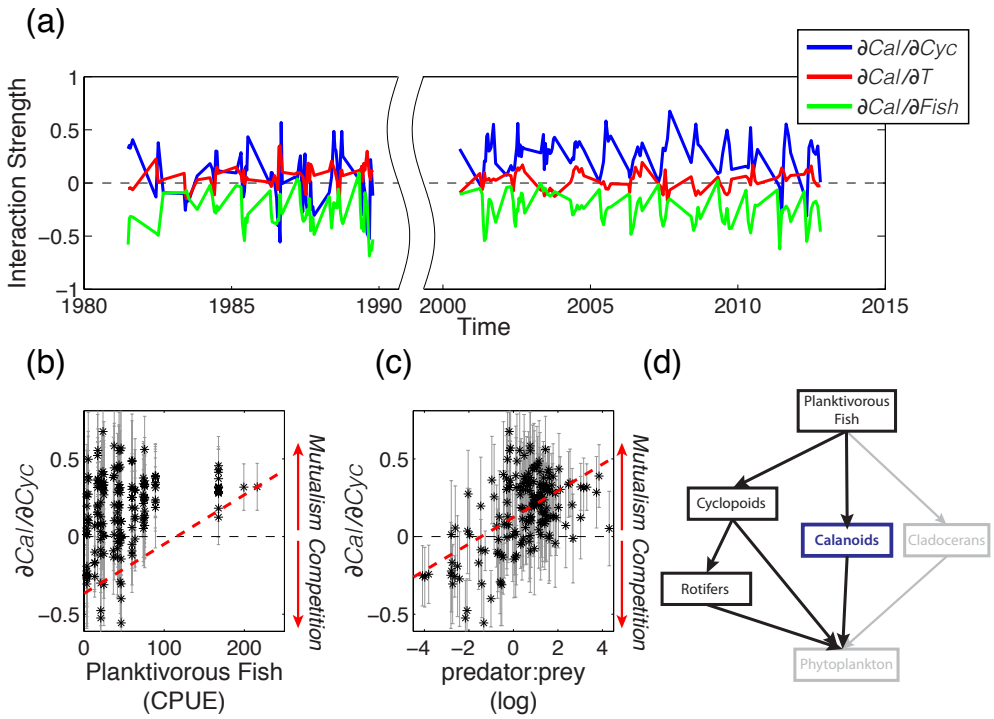

Figure 4.4: Dynamic interactions in Sparkling Lake. (a) S-map estimates of interaction coefficients quantify the changing effects of cyclopoids (Cyc), temperature (T), and planktivorous fish abundance ($Fish$) on calanoid copepods (Cal). The interaction of cyclopoids on calanoids $\partial Cal/\partial Cyc$ 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 $Fish/(1+Cal+Cyc)$.

4.5 References

1. Anderson CNK, Hsieh C-H, Sandin SA, Hewitt R, Hollowed AB, Beddington J, May RM, Sugihara G. Why fishing magnifies fluctuations in fish abundance. Nature. 2008 Apr 17;452(7189):835-9.

2. Lima M, Stenseth NC, Jaksic FM. Food web structure and climate effects on the dynamics of small mammals and owls in semi-arid Chile. Ecol Lett. 2002 Mar;5(2):273-84.

3. Lima M, Morgan Ernest SK, Brown JH, Belgrano A, Stenseth NC. Chihuahuan Desert kangaroo rats: nonlinear effects of population dynamics, competition, and rainfall. Ecology. 2008 Sep;89(9):2594-603.

4. Gratton C, Denno RF. Seasonal shift from bottom-up to top-down impact in phytophagous insect populations. Oecologia. Springer-Verlag; 134(4):487-95.

5. Werner EE. Individual behavior and higher-order species interactions. Am Nat. 1992 Nov;140:S5-S32.

6. Jeschke JM, Kopp M, Tollrian R. Consumer-food systems: why type I functional responses are exclusive to filter feeders. Biol Rev. Blackwell Publishing Ltd; 2004;79(2):337-49.

7. Abrams PA. Consumer functional response and competition in consumer-resource systems. Theor Popul Biol. 1980 Feb;17(1):80-102.

8. Abrams PA, Holt RD, Roth JD. Apparent competition or apparent mutualism? shared predation when populations cycle. Ecology. Ecological Society of America; 1998 Jan 1;79(1):201-12.

9. Chase JM, Abrams PA, Grover JP, Diehl S, Chesson P, Holt RD, Richards SA, Nisbet RM, and Case TJ. The interaction between predation and competition: a review and synthesis. Ecol Lett. 2002 Mar;5(2):302-15.

10. Holt RD, Lawton JH. The ecological consequences of shared natural enemies. Annual review of ecology and systematics. Annual Reviews; 1994 Jan 1;25 IS -:495-520.

11. Schoenly K, Cohen JE. Temporal variation in food web structure: 16 empirical cases. Ecol Monogr. JSTOR; 1991;:267-98.

12. Tavares-Cromar AF, Williams DD. The importance of temporal resolution in food web analysis: evidence from a detritus-based stream. Ecol Monogr. Ecological Society of America; 1996 Jan 1;66(1):91-113.

13. Ives A, Dennis B, Cottingham K, Carpenter S. Estimating community stability and ecological interactions from time-series data. Ecol Monogr. 2003;73(2):301-30.

14. Sugihara G, May R, Ye H, Hsieh CH, Deyle E, Fogarty M, and Munch S. Detecting causality in complex ecosystems. Science. 2012 Oct 25;338(6106):496-500.

15. Sugihara G, Allan W, Sobel D, Allan KD. Nonlinear control of heart rate variability in human infants. Proc Natl Acad Sci USA. 1996 Jan 19;93(6):2608-13.

16. Sugihara G. Nonlinear forecasting for the classification of natural time series. Philos T Roy Soc A. 1994 Sep 15;348(1688):477-95.

17. Sugihara G, May RM. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time-series. Nature. 1990;344(6268):734-41.

18. Dixon PA, Milicich MJ, Sugihara G. Episodic fluctuations in larval supply. Science. American Association for the Advancement of Science; 1999;283(5407):1528-30.

19. Deyle ER, Fogarty M, Hsieh CH, Kaufman L, MacCall AD, Munch SB, Perretti CT, Ye H, and Sugihara G. Predicting climate effects on Pacific sardine. Proc Natl Acad Sci USA. 2013 Apr 16;110(16):6430-5.

20. Levins R, Pressick ML, Heatwole H. Coexistence patterns in insular ants. Am Sci. Sigma Xi, The Scientific Research Society; 1973 Jan 1;61(4):463-72.

21. MacArthur R. Species packing and competitive equilibrium for many species. Theor Popul Biol. 1970 May;1(1):1-11.

22. Hernandez M-J. Disentangling nature, strength and stability issues in the characterization of population interactions. J Theor Biol. 2009 Nov;261(1):107-19.

23. Sugihara G, Grenfell B, May RM. Distinguishing error from chaos in ecological time series. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences. 1990 Nov;330(1257):235-51.

24. Hastings A, Powell TM. Chaos in a three-species food chain. Ecology. 1991 Jun;72(3).

25. Post DM, Conners ME, Goldberg DS. Prey preference by a top predator and the stability of linked food chains. Ecology. Ecological Society of America; 2000 Jan 1;81(1):8-14.

26. Cade BS, Noon BR. A gentle introduction to quantile regression for ecologists. Front Ecol Environ. 2003 Oct;1(8):412-20.

27. Perretti CT, Sugihara G, Munch SB. Nonparametric forecasting outperforms parametric methods for a simulated multispecies system. Ecology. Ecological Society of America; 2012 Dec 13;94(4):794-800.

28. BenincĹ E, Huisman J, Heerkloss R, Johnk KD, Branco P, van Nes EH, Schaffer M, and Ellner SP. Chaos in a long-term experiment with a plankton community. Nature. 2008 Feb 14;451(7180):822-5.

29. BenincĹ E, JŽhnk KD, Heerkloss R, Huisman J. Coupled predator-prey oscillations in a chaotic food web. Ecol Lett. 2009 Oct 20;12(12):1367-78.

30. Beisner BE, Ives AR, Carpenter SR. The effects of an exotic fish invasion on the prey communities of two lakes. J Anim Ecol. Blackwell Science Ltd; 2003;72(2):331-42.

31. Petris G. An r package for dynamic linear models. Journal of Statistical Software. American Statistical Association; 2010;36(i12).

32. Hastie TJ, Tibshirani RJ. Generalized additive models. Boca Raton, FL: Chapman and Hall/CRC; 1990.

4.6 Appendix

4.6.1 S-maps vs. DLM

At first glance S-maps may seem similar to multivariate auto-regression (MAR) or vector auto-regression (VAR) methods that allow coefficients to change or drift in time (usually referred to under the more general umbrella of dynamic linear models or DLM). Both are based on the idea of weighted linear regression. However, there is a fundamental and critical difference between the two: DLM methods do not explicitly address state dependence. 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). States can change quite rapidly in ecosystems (e.g. outbreaks of spruce budworms or fishery collapses), and thus states nearby in time may be very dissimilar (have different interaction strengths). To illustrate the difficulty of applying DLM to measure interactions in state-dependent ecosystems, we examine the 5-species model from main text. Figure 4.5 illustrates which observations get used to determine the coefficients when weighting is determined by (a) proximity in state space (a la S-maps) versus (b) proximity in time (a la standard DLM). Note that when weighting is based on time (b), there are points on the trajectory that get weighting even though they represent very different ecosystem states. At the target point (solid red circle), resource abundance is high, but at several points nearby in time (open red circles), food abundance is much lower. Note also that there are many points on the attractor that are nearby the target point (solid red circle), but don't happen to be nearby in time. The interaction coefficients at these time points are very close to the coefficients at the target point. S-maps take advantage of this, whereas DLM do not.

To further illustrate this point, we compare the explicit calculations of interaction coefficient of DLM and S-maps to the exact calculations obtained from the known model.

The DLM is computed using the R package dlm (4) and is formulated as follows:

$$x_1(t+1) = \sum_{j=1}^{E} \left[\Theta_j(t) x_j(t)\right] + \varepsilon(t)$$

$$\Theta_j(t) = \Theta_j(t-1)\omega_j(t),$$

where $\varepsilon \propto N[0, \sigma_{obsess}$ is the observational noise, and the $\omega_j(t) \propto N[0, \alpha]$. The $\Theta_j(t)$ are the interaction coefficients between the target species x_i and the x_j. We use capital Θ for these coefficients to avoid confusion with the nonlinear S-map parameter θ above). The coefficients are assumed to follow a random walk with parameter α. We present results for $\alpha = 0.03$, which we found in our analysis maximized the correspondence with the actual competition values calculated from the model. Panel (c) of Fig. 4.5 in the main text compares DLM predictions of competition between C_1 and C_2 to S-map predictions and model calculations for a section of the time-series. The S-maps estimates match the model calculations exceedingly well, while the DLM estimates do not. Effectively, the DLM cannot account for how quickly competition changes in the model. During the rapid bursts of intense competition, there are points nearby in time in which competition is quite low. The DLM scheme gives weight to these points, and therefore misestimates the true competition. Taken over the whole time series, the S-map estimates of competition are highly accurate, $\rho = 0.75$ (Pearson's correlation), but the DLM are not, $\rho = 0.3$.

If the time series is highly over-sampled (observations are very close together in time compared to the rate of change of the system variables), weighting in time can approximate weighting in state space, and thus DLM can give similar results as S-maps. In practice, however, such over-sampling is rare in ecology.

4.6.2 CCM Analysis for Mesocosm Example

Convergent cross-mapping (CCM) is based on the theorem of Takens (1). An important consequence of the theorem is that if two species X and Y are part of the same dynamic system, then it is possible to predict the current abundance of species X using the time lag embedding of species Y (2). In this way, CCM can test if one ecosystem variable drives another. Here, we apply CCM to the species abundances in the Baltic Sea mesocosm time series to provide additional validation of the causal relationships between the species.

To calculate the cross-mapping between species, we follow the method described by Sugihara et al.(2), but with one additional consideration. The number of time lag coordinates of the predictor species Y defines the embedding dimension, E. The theorem of Takens states that an embedding dimension of $E = 2d + 1$ (where d is the dimension of the system) is sufficient, but not strictly necessary. Thus, even though the ecosystem may have d key state variables (i.e. it is d-dimensional), the optimal embedding dimension (number of time lags) E for predicting a species X from Y could be anywhere between d and $2d + 1$, depending on the particular mathematical properties of that system, the amount of data, and quality of the data (observation noise).

Fitting the embedding dimension outright to maximize cross-prediction for each predictor-target pair bears significant risk of over-fitting. Instead, it is better to use a statistically independent (or at least quasi-independent) criterion to choose E. We propose the following. CCM is typically done for simultaneous prediction, that is predicting species X at time t (i.e. $X(t)$) from the lag coordinates of Y, $\langle Y(t), Y(t-1), \ldots, Y(t-(E-1))\rangle$. However, it is also possible to predict future and past values of X (i.e. $X(t+p)$) from the same Y embedding. The optimal embedding dimension for lag coordinates of Y should be consistent for changes in the prediction time, p. Generally speaking, there will be some limited statistical dependence between these quantities, even if the two time

series X and Y are actually unrelated. Nevertheless, the rate of false positives will be much lower than simply picking E to maximize the ordinary $p = 0$ cross-prediction. 4.1 shows predicting $X_j(t)$ from lags of X_i $(p = 0)$ using the embedding dimension E that maximizes the predictions of $X_j(t-1)$ from the same lags of X_i $(p = -1)$. Statistical significance is tested using the method of surrogates. In this case, surrogate time series are created for each species by sampling values from the original time series at random (with replacement). This gives a distribution for the expected cross-mapping between species under the null hypotheses that the variables are dynamically unrelated. We find statistically significant ($p < 0.05$) evidence for bidirectional interactions among the four species focused on in the main text: calanoid copepods, rotifers, nanoflagellates, and picophytoplankton.

To have a definitive test for causality, it is not only necessary for the cross-mapping predictability to be statistically significant, but also that predictability shows convergence-that is, predictability with the amount of data used in the lag coordinate reconstruction up to a practical limit (2). This is convergent cross-mapping (CCM). Figure 4.6 shows cross-mapping convergence for the interactions between calanoid copepods and the three other focal species. CCM with filamentous diatoms, harpacticoids, and bacteria are also shown.

4.6.3 Weighting Parameter

S-maps, the sequential local linear regression scheme employed in the paper, contains a parameter, θ, which controls how strongly the points nearby the target in multidimensional state space are weighted in the regression. When $\theta = 0$, the weighting is equal across all points on the attractor (and independent of distance), and hence S-maps reduces to simple vector autoregression (VAR). This means the regression coefficients do not vary across the time series; this corresponds a system where interaction strengths are

fixed (a system at equilibrium).

For small positive θ, the regression coefficients will vary by state, but only weakly. If θ is too small, the coefficients will underestimate the true variability in interaction strength. However, the larger θ becomes, the more the regression hinges on only the most proximal points, and will therefore be more sensitive to observation error. In practice, some intermediate value of θ will optimally balance bias and uncertainty. The simplest way to chose an appropriate θ is to examine prediction error as a function of θ. Looking at the mesocosm example, the normalized mean absolute error (nMAE) between S-map predictions and observations of calanoid abundance are shown as a function of θ for the multivariate embedding $\langle Cal(t), Rot(t), Nano(t), Pico(t)\rangle$ (Fig. 4.8). Error is minimized at $\theta = 3$. For comparison, the error as a function of θ is also shown for the best univariate embedding (Fig. 4.7). To determine the best univariate embedding, we apply the same method as Glaser et al. (3), which is to minimize the error of simplex projection as a function of embedding dimension. The four dimensional embedding, i.e. $\langle Cal(t), Cal(t-1), Cal(t-2), Cal(t-3)\rangle$ is best (Fig. 4.9).

Most notably, the best multivariate embedding improves prediction (has lower error) relative to the univariate model: nMAE is 0.28 for the multivariate embedding ($\theta = 3$) and 0.31 for the univariate embedding ($\theta = 4.5$). This provides additional confirmation to CCM that the population dynamics of calanoids can be understood well in terms of calanoids, rotifers, nanoflagellates, and picophytoplankton. That is to say, the dynamic attractor is well embedded by these four variables.

A similar analysis for the multivariate embedding is applied to the Sparkling Lake case study (Fig. 4.10). We find an optimal value of $\theta = 2.8$ for the multivariate embedding analyzed in the main text, $\langle Cal(t), Cyc(t), Rot(t), T(t), Fish(t)\rangle$. Note that as discussed in the main text, univariate comparison is not possible for Sparkling Lake, due to the difficulty in taking time lags for irregularly sampled data.

4.6.4 Observation Error and Bias

Here we directly investigate effects of observation error and bias in the S-map estimations of interaction strength. We use the 5-species model from Figure 4.2 in the main text, so that we have exact calculations of interaction strength to which we can compare the S-map estimates. The model is run as described in the material and methods, but for 50,000 time points. Random time-series segments of a fixed length L are selected, then normally distributed i.i.d. observation error is added to all 5 variables, where the magnitude of the observation error is scaled to the standard deviation of the time series. We ran 200 replicates for each time-series length L and magnitude of observation error. Figure 4.13 looks at the accuracy of S-map estimates as a function of observation error for $L = 300$ and 1000. Accuracy is quantified both by Pearson's correlation between S-map estimated interaction strength and true value and by the mean error (S-map estimate - true value). The upper 0.95 quantiles and lower 0.05 quantiles (taken over the 200 replicates) of accuracy are shown in red dotted lines. Though the accuracy measured by Pearson's correlation does decline some with the magnitude of error, the effect is modest, and demonstrates the ability of S-maps to cope realistic levels of observation error.

We also note that for this model, the S-map method for estimating the interaction coefficients shows an error bias in two cases, $\partial C_1/\partial R$ and $\partial C_1/\partial P_1$. Note that in both cases, S-map is consistently under-estimating the magnitude, as the effect of R is always positive and the effect of P_1 is always negative. This is not surprising. The obvious source of bias comes from applying linear regression to a nonlinear function. If we have not "zoomed in" close enough to the manifold, there will still be some curvature (nonlinearity) in the local neighborhood that we are analyzing. We expect the most noticeable bias specifically in the areas of the manifold with greatest curvature. This is in fact suggested by Figure 4.2 (e), as the magnitude of the peaks e.g. in $\partial C_1/\partial R$ are consistently under-estimated.

As an aside for those interested, we note that S-maps were specifically designed with noisy ecological data in mind (i.e. the case studies in Sugihara:1994kn). In a noise-free system there are other ways to perform linear regression that conceivably give greater accuracy and less bias- for example, simply performing the regression over the attractor points that fall within a small epsilon ball. By giving some consideration to points outside the smallest neighborhood, S-maps have the ability to average out observational noise. As noted, the theta parameter controls this trade-off between noise averaging and potential bias in the linear model due to including distant points where the slopes are considerably different. That being said, there is clearly plenty of room to tinker with the specific scheme for performing local linear regression, for example allowing the theta parameter to change in state space based on preliminary estimates of curvature.

4.6.5 Robustness to Choice of Embedding

Though the 4-dimensional multivariate embedding for the Baltic Sea mesocosm analyzed in the main text produces good forecasts of calanoid copepod population dynamics, CCM (Table 4.1, Fig. 4.6) suggests there might be additional influences on calanoid dynamics beyond calanoids, rotifers, nanoflagellates, and picophytoplankton. Specifically, filamentous diatom and bacteria abundances cross-map well with calanoid abundance. Thus, we wish to check that the main results of the paper are robust to including either of these variables as additional state variables in the analysis shown in main text Fig. 4.3. Thus, we repeat the analysis from Fig. 4.3, but with bacteria abundance (Fig. 4.11) and filamentous diatom abundance (Fig. 4.12) each included as a 5th state variable. As in Figure 4.3 (b), competition ($-\partial Cal/\partial Rot$) is strongest only when food abundance is near zero (Fig. 4.11b, Fig. 4.12b). Furthermore, the maximum competition is controlled by the amount of food limitation, ű Cal Ú ű Nano (Fig. 4.11c,

Fig. 4.12c). We test this quantitatively by calculating the 0.95 quantile regression. In both cases, the slope of the 0.95 quantile is significantly greater than zero ($p < 0.01$).

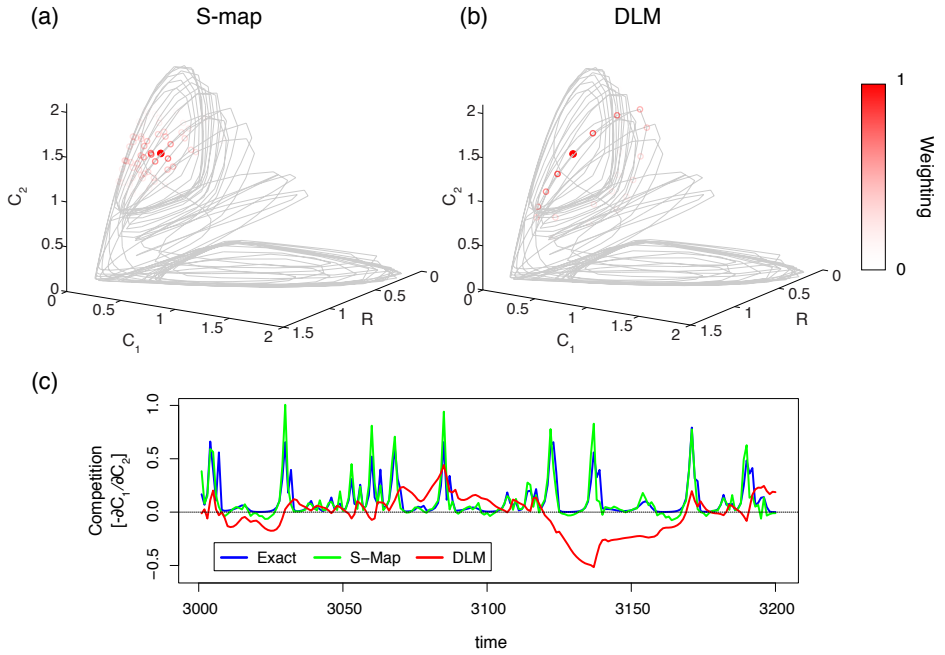

Figure 4.5: S-map versus Dynamic Linear Model (DLM). S-map and DLM are both applied to measuring competition in the 5-species food chain model described in the main text. The DLM model is a vector autoregressive (VAR) model where the linear interaction coefficients are allowed to drift in time as a random walk. Panels (a) and (b) illustrate the weighting of points (open red circles) on the empirical attractor (grey line) for measuring competition at the target point (solid red circle). In (a) weighting is determined by distance in state space (as in S-maps), while in (b) it is determined by distance in time (as in DLM). 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) that the target point (solid red). Panel (c) compares estimates of competition based on S-map, DLM, and explicit calculations from the model. DLM is unable to keep up with the rapid changes in competition observed in the real model and consistently misestimates competition.

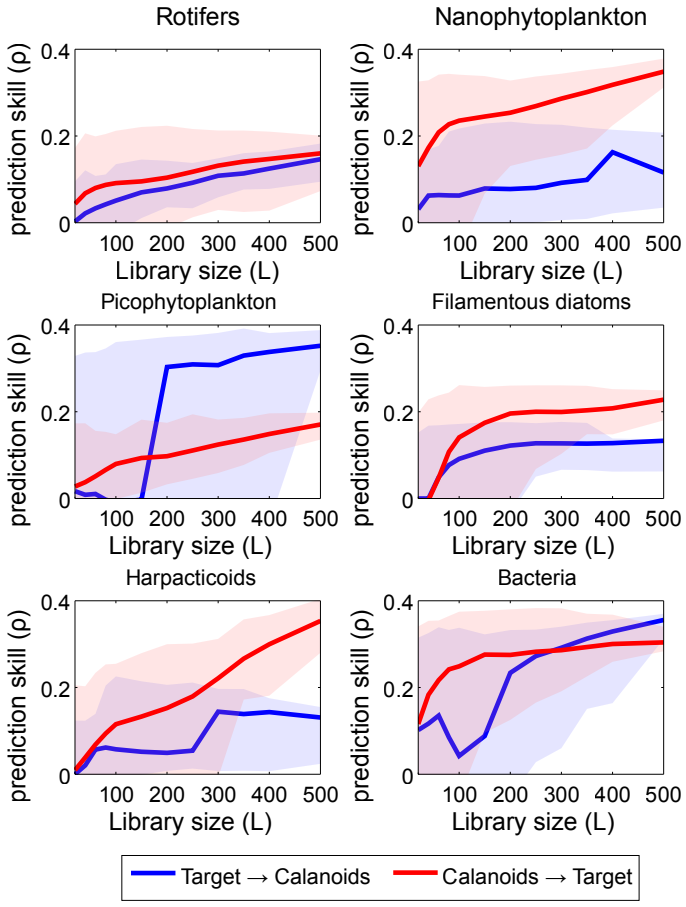

Figure 4.6: Convergent cross mapping (CCM) between calanoid copepods and the six other target population variables. If the target variable can predict calanoids (blue line), this indicates that the calanoid population has a causal influence on the target variable. Likewise, if calanoids can predict the target variable (red line), then the target variable has a causal influence on the calanoid population. Shaded region indicate 10th and 90th percentiles.

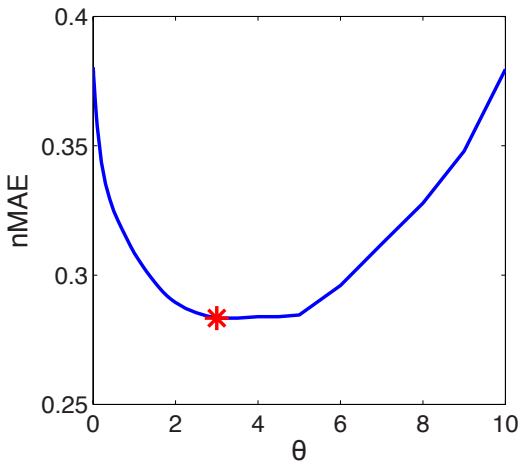

Figure 4.7: Prediction error vs. θ for multivariate forecasts of calanoid abundance. Blue line indicates normalized mean absolute error ($nMAE$) in 1-week S-maps forecasts of calanoid abundance using the multivariate embedding $\langle Cal(t), Rot(t), Nano(t), Pico(t)\rangle$. The red star indicates the optimal θ (minimizes error), which was then used for analysis in the main text.

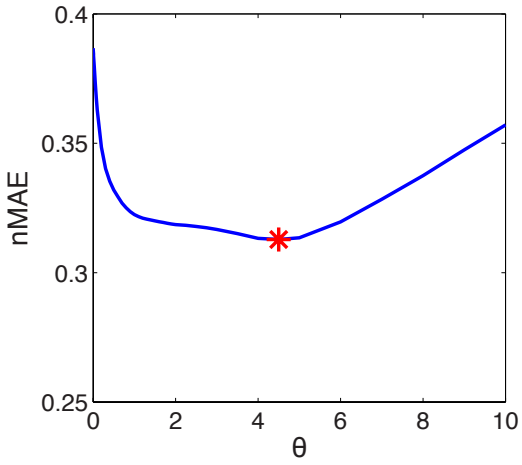

Figure 4.8: Prediction error vs. θ for multivariate forecasts of calanoid abundance. Blue line indicates normalized mean absolute error ($nMAE$) in 1-week S-maps forecasts of calanoid abundance using the multivariate embedding $\langle Cal(t), Rot(t), Nano(t), Pico(t)\rangle$. The red star indicates the optimal θ (minimizes error), which was then used for analysis in the main text.

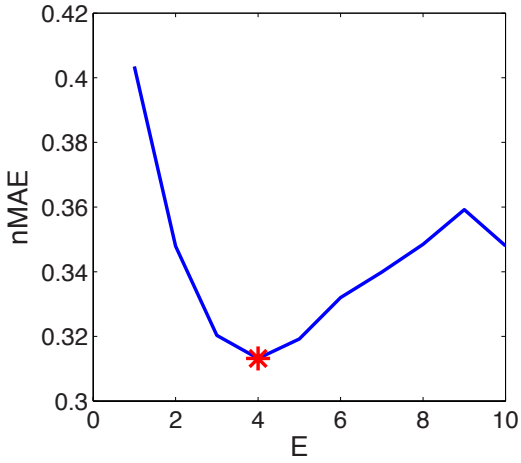

Figure 4.9: Prediction error vs. E for univariate forecasts of calanoid abundance. Simplex projection is used to forecast calanoid abundance 1 week in the future based on univariate embeddings with a range of embedding dimension, E. Normalized mean absolute error ($nMAE$) between observations and predictions is minimized with an embedding dimension $E = 4$. That is, the state space with coordinate axes $\langle Cal(t), Cal(t-1), Cal(t-2), Cal(t-3)\rangle$.

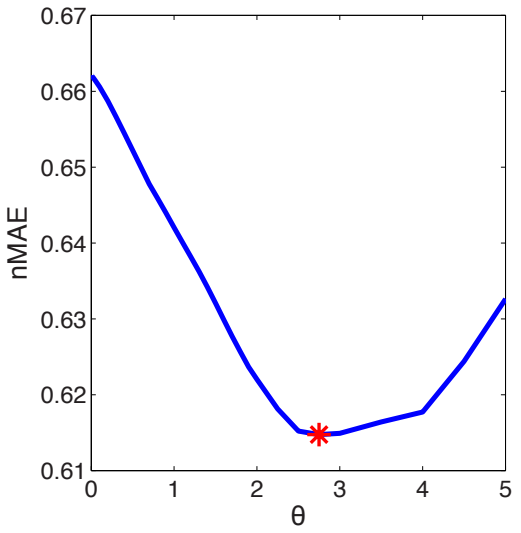

Figure 4.10: Prediction error vs. θ (nonlinearity) for multivariate EDM forecasts of calanoid abundance in Sparkling Lake. Blue line indicates normalized mean absolute error ($nMAE$) in 2-week S-maps forecasts of calanoid abundance using the five-dimensional embedding $\langle Cal(t), Cyc(t), Rot(t-2), T(t), Fish(y)\rangle$. S-maps with $\theta = 0$ represent the best linear model (VAR) with constant coefficients. As θ increases, the coefficients become increasingly more locally determined (nonlinear). Best prediction is obtained for $\theta = 2.8$, suggesting that calanoid dynamics are best understood as nonlinear.

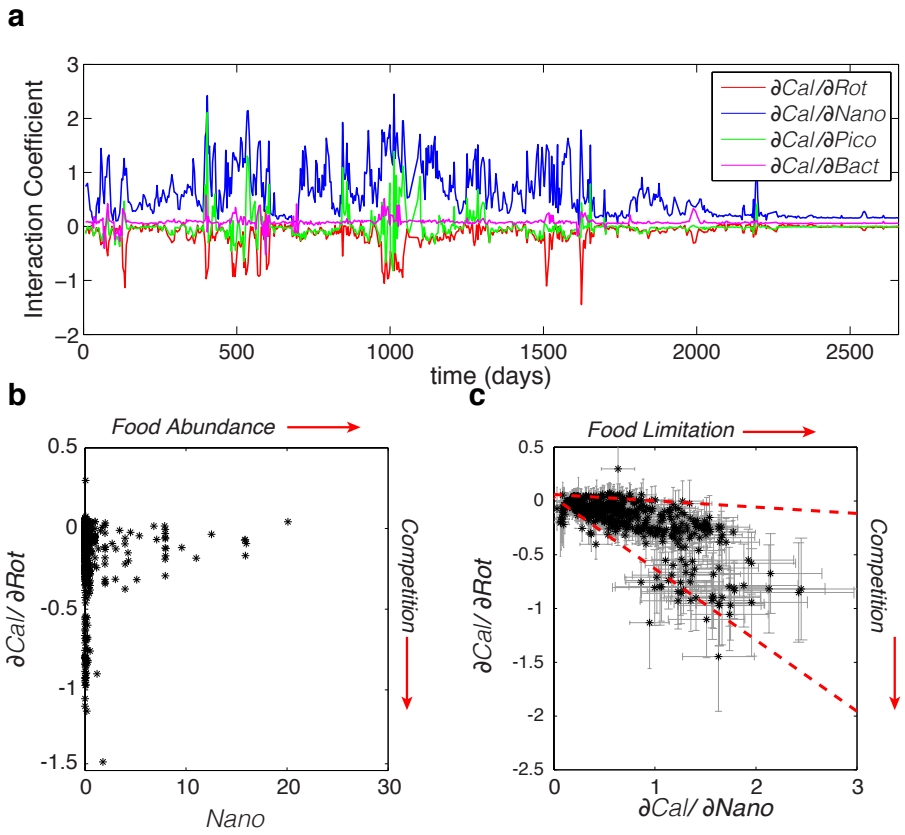

Figure 4.11: Reproduction of main text Fig. 4.3 with bacterial abundance as a 5th state variable. (a) S-map estimates of interaction coefficients capturing the effects of the nanoflagellates (*Nano*), picocyanobacteria (*Pico*), rotifers (*Rot*), and bacteria (*Bact*) on calanoid copepods (*Cal*) through the duration of the experiment. Competition ($-\partial Cal/\partial Rot$) is shown as a function of (b) food abundance (*Nano*) and (c) food limitation ($\partial Cal/\partial Nano$). Red dashed lines indicate the 0.05 and 0.95 quantile regressions. The slope of the 0.95 quantile is significantly different from 0 ($p < 0.01$). Results qualitatively match Fig. 4.3.

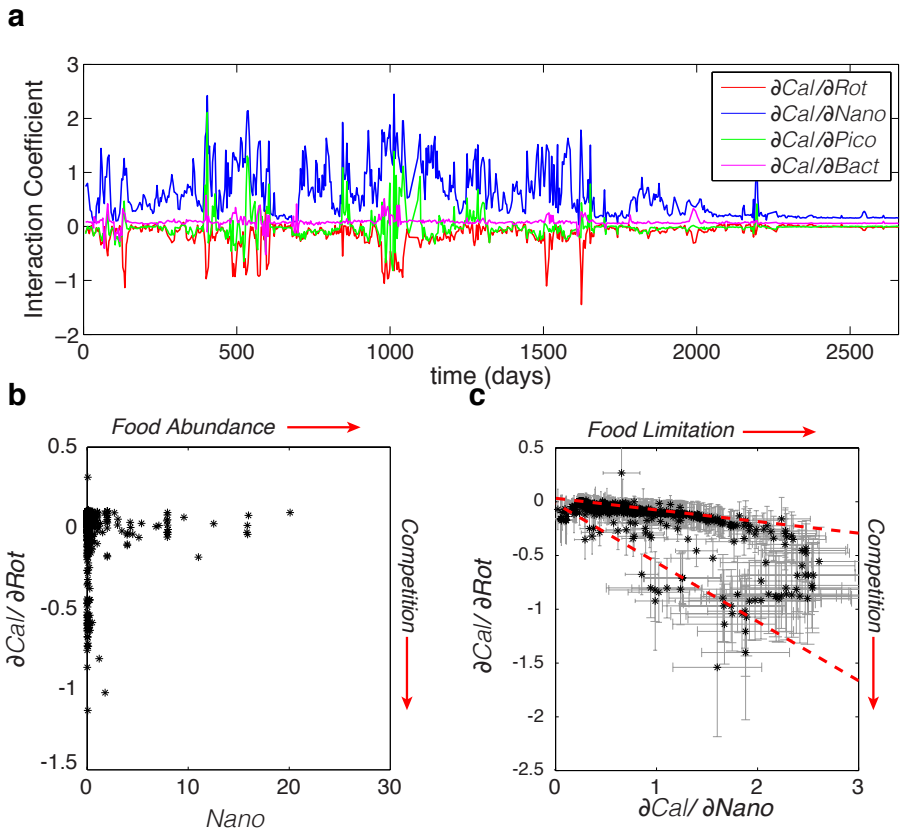

Figure 4.12: Reproduction of main text Fig. 4.3 with filamentous diatom abundance as a 5th state variable. (a) S-map estimates of interaction coefficients capturing the effects of the nanoflagellates (*Nano*), picocyanobacteria (*Pico*), rotifers (*Rot*), and diatoms (*Diat*) on calanoid copepods (*Cal*) through the duration of the experiment. Competition ($-\partial Cal/\partial Rot$) is shown as a function of (b) food abundance (Nano) and (c) food limitation ($\partial Cal/\partial Nano$). Red dashed lines indicate the 0.05 and 0.95 quantile regressions. The slope of the 0.95 quantile is significantly different from 0 ($p < 0.01$). Results qualitatively match Fig. 4.3.

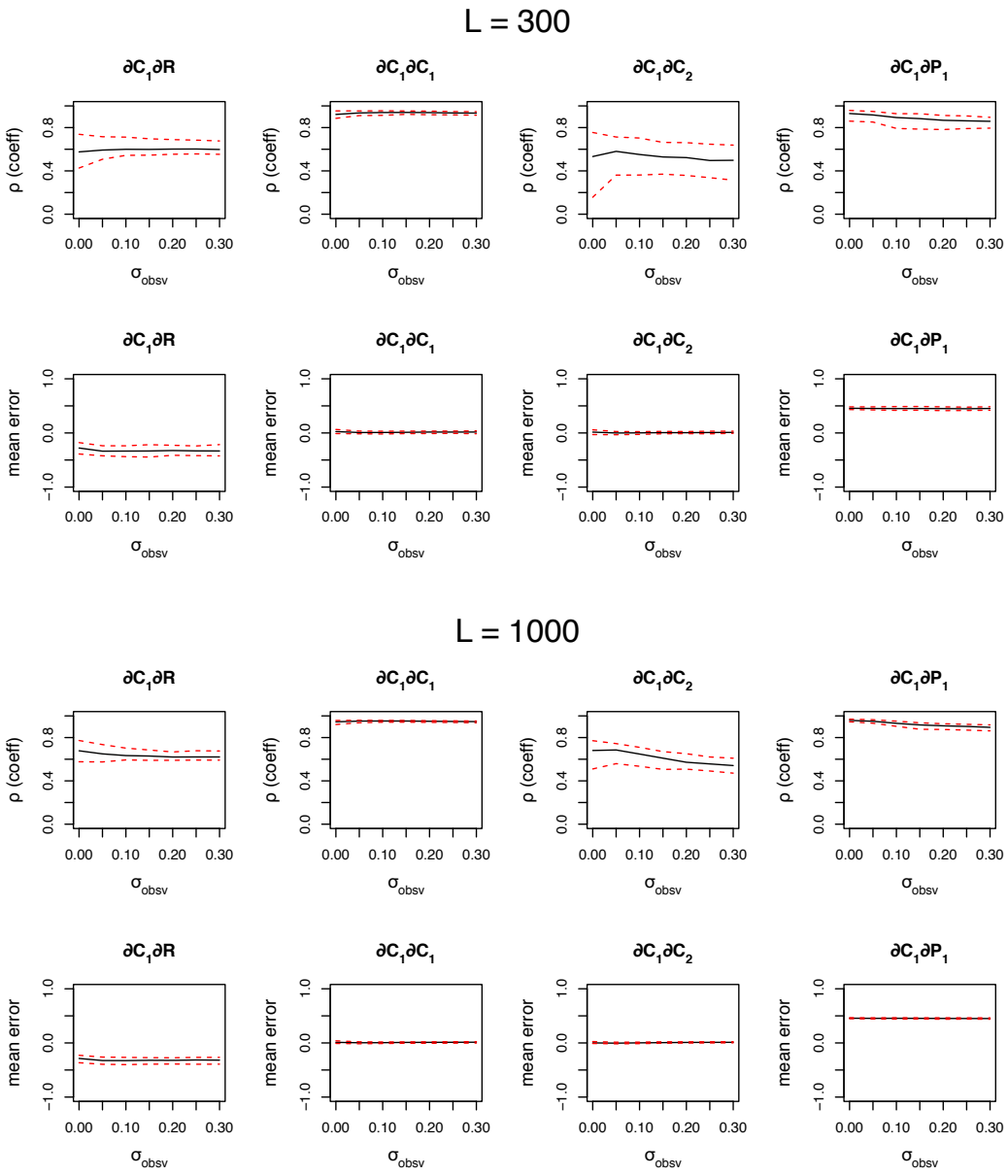

Figure 4.13: Error and bias in S-map interaction estimates are examined in the 5-species model. The correlation between S-map estimated interaction and true values, ρ_{coeff}, and the mean error (S-map estimate minus true value) are shown as a function of the level of observation error ε_{obsv} applied to all 5 variables. Here, ε_{obsv} is measured as a fraction of the standard deviation of the time-series, and thus ranges from 0 to 30%. Accuracy is calculated for 200 replicates at each ε_{obsv} and time-series length. Upper 0.95 and lower 0.05 quantiles are shown as red dotted lines.

Table 4.1: Cross-mapping between species in Baltic Sea mesocosm. If target (column) species, Y, can be predicted from time series of predictor (row) species, X, then Y causally influences X. Cross-mapping prediction skill is measured as the Pearson's correlation ? between observed and predicted values. Statistically significant cross-mapping ($p < 0.05$) is indicated by grey shading, and suggests that the column species causally affects the row species.

	Cyclopoids	Calanoids	Rotifers	Protozoa	Nanoflagellates	Picocyanobacteria	Diatoms	Ostricods	Harpacticoids	Bacteria
Cyclopoids		0.21	0.07	0.00	0.14	0.30	0.03	-0.05	0.13	0.34
Calanoids	0.24		0.16	0.06	0.11	0.37	0.10	0.04	0.13	0.37
Rotifers	0.15	0.17		-0.01	0.22	0.11	0.00	0.20	0.40	0.45
Protozoa	0.09	0.33	0.01		0.06	0.03	0.00	0.19	0.22	0.20
Nanoflagellates	0.15	0.37	0.02	-0.01		-0.02	0.00	0.33	0.03	0.20
Picocyanobacteria	0.25	0.18	0.29	0.01	0.21		0.05	0.21	0.21	0.37
Diatoms	0.07	0.25	0.11	0.01	0.23	0.03		-0.03	0.23	0.20
Ostricods	0.02	0.18	0.13	0.03	0.07	0.13	0.10		0.34	0.17
Harpacticoids	0.12	0.38	0.01	0.00	0.32	0.23	0.05	0.47		0.44
Bacteria	0.06	0.31	0.32	0.00	0.26	0.39	0.00	0.44	0.52	

4.7 Appendix References

1. Takens F. Detecting strange attractors in turbulence. In: Rand DA, Young L, editors. Dynamical systems and turbulence, Warwick 1980. New York, U.S.A.: Lector Notes in Mathematics; 1981. pp. 366-81.

2. Sugihara G, May R, Ye H, Hsieh CH, Deyle E, Fogarty M, et al. Detecting causality in complex ecosystems. Science. 2012 Oct 25;338(6106):496-500.

3. Glaser SM, Fogarty MJ, Liu H, Altman I, Hsieh C-H, Kaufman L, MacCall AD, Rosenberg AA, Ye H, and Sugihara G Complex dynamics may limit prediction in marine fisheries. Fish and Fisheries. Wiley Online Library; 2013.

4. Petris G. An r package for dynamic linear models. Journal of Statistical Software. American Statistical Association; 2010;36(i12).

Chapter 5

Global Environmental Drivers of Influenza

Abstract

In temperate countries, influenza outbreaks are well correlated to the seasonal changes in temperature and absolute humidity. However, in tropical countries where annual climate cycles are much weaker, influenza seasonality is harder to find and more difficult to explain. Here, we use a dynamical systems approach (convergent cross mapping) to directly examine global drivers of influenza outbreaks from country-level time series. By identifying causal drivers rather than correlations, we show that despite the apparent differences between temperate and tropical countries, absolute humidity and to a lesser extent temperature drive influenza outbreaks globally. We also corroborate a U-shaped relationship between absolute humidity and influenza at the population level that has been suggested in principle by experiment. The results show that there are global rules for environmental drivers of flu that apply across latitudes.

5.1 Introduction

A diverse host of plausible mechanisms have been put forward to explain the wintertime occurrence of seasonal influenza outbreaks. Low solar irradiance is thought to impair host immune functions (1, 2). Laboratory experiments show that relative humidity controls droplet size and aerosol transmission rates (3). Experiments with mammalian models showed that viral shedding by hosts increases at low temperature (4).

Recent laboratory experiments have given strong support that seasonal lows in absolute humidity might in fact be the ultimate driver of seasonal influenza outbreaks in temperate regions (5). Statistical analysis of population level data using correlation has supported absolute humidity as a driver of seasonal influenza outbreaks, but not without some ambiguity. Analysis shows a high correlation between flu peaks and absolute humidity in temperate countries (6) and individual U.S. states (7). However, absolute humidity also tends to be tightly correlated to the seasons and to other possible drivers like temperature and precipitation (8). Indeed, other regression-based analysis has suggested that it is temperature not humidity that is the most immediate drivers of flu seasonality (9).

Conversely, weather and flu have weaker seasonal patterns in the tropics, and significant correlations are much harder to find. The effect of absolute humidity and other climatic variables on influenza in the tropics continues to be questioned (8). Many tropical countries do not experience the low absolute humidity levels associated with outbreaks in temperate countries, and there has been speculation that there is a different set of "rules" for influenza in low latitudes- for example that precipitation, mediated by contact rates, drives tropical seasonality (6). However, there is laboratory and modeling evidence that flu persistence might in fact have a U-shaped response to absolute humidity (10, 11), which could potentially operate across latitudes. Host-pathogen dynamics are

widely regarded as being nonlinear. Generally speaking, correlation is poorly suited to understanding cause-effect relationships in nonlinear systems (12). However, correlation can still be useful to nonlinear systems in specific circumstances-most notably when there is synchrony between driver and response variables. That is, the effect of the driving variable is strong enough that the response becomes enslaved to the driver and the internal dynamics of the response variable cease to be important. For example, historical chickenpox infections in New York City appear to passively track the annual public school calendar of opening and closing . This fact could explain the success of correlative approaches in temperate countries and failure in the tropics. Firstly, basic host-pathogen dynamics exhibit dynamical resonance when forced by periodic drivers (13) which can cause the intrinsic nonlinear epidemiological dynamics to become synchronized (phase-locked) to the simple cyclic motion of the environmental driver. However, not all regions of the world have strong climate seasonality. While the seasonal cycle in a temperate country like Germany can explain more than 90% of the variance in absolute humidity across multiple decades, in tropical countries like Singapore it can explain less than 30%. Moreover, there is a strong correspondence between seasonality of climate and seasonality of influenza. This is illustrated in Figure 1, which shows the seasonality in absolute humidity (panel a) and influenza (panel b) across countries. The countries with the least seasonality in environment (yellow shades) also have the least seasonal influenza (Spearman's $\rho = 0.73$).

When drivers do not induce synchrony, the underlying nonlinear dynamics can cause the statistical relationship between driver and response to become very complex. Indeed, the same simple epidemiological model of (13) illustrates how the identical mechanistic effect of a driver can produce very different behavior when only the periodicity is changed. Driving the model with a strongly seasonal environmental variable produces consistent seasonal outbreaks that correlate very strongly to the driver (Figure

2a). However, replacing the driver with another signal of the exact same magnitude but much weaker seasonality results in both much weaker seasonality in flu and much less correlation between flu infections and climate (Figure 2b). The mechanisms have not changed, only the spectrum (periodicity) of the driver. This illustrates an important point: lack of seasonality in influenza and lack of correlation between flu and climate does not mean a lack of environmental forcing. Thus, to develop a global understanding of climate and influenza, it is important to consider methods that can cope with nonlinear interactions.

Here, we employ an empirical dynamical modeling framework to study climactic effects on influenza at the population level. Using convergent cross-mapping (CCM) and related methods, we show clear effects of absolute humidity and to a lesser extent temperature across latitudes. We also corroborate the U-shaped response of flu to absolute humidity at the population level that has been suggested by previous laboratory and modeling work (10, 11).

5.2 Results

First, we examine CCM between influenza and three environmental drivers: absolute humidity, temperature, and precipitation. We not only want to look for evidence of causality, but we want to distinguish actual causal relationships from shared seasonality. Thus, we compare the cross-map prediction we measure for the empirical environmental time series to the null distribution we get for randomized surrogate time series that have the same seasonality as the actual driver. Figure 3 shows box-and-whisker plots of the null distributions for cross-map skill (ρ_{CCM}) ordered according to distance from the equator (absolute latitude). The measured CCM skills are plotted on top-red stars if the value is significantly different from the null distribution ($p \leq 0.05$) and red circles otherwise.

Panels (a) and (b) show significant forcing by absolute humidity and temperature across latitude. The results have very high meta-significance (Fisher's method): $p < 5 \times 10^{-6}$ for AH and $p < 5 \times 10^{-5}$ for T, but a paired t-test does not distinguish one over the other. Panel (c) shows that there is also meta-significant evidence for forcing across latitudes by relative humidity, $p < 3 \times 10^{-4}$. However, paired t-tests with both AH and T confirm that CCM with relative humidity is weaker on average ($p < 0.01$). Finally, panel (d) shows that there is no significant evidence with CCM for driving by precipitation in any of the tropical countries examined and globally CCM with flu and precipitation is not meta-significant ($p \approx 0.39$). Note for Fisher's method, we set a small individual p-value of 0.001 for real CCM measures that are larger than all 500 surrogates. Although the absolute level of cross-map skill is lower in the tropics, the causal effect of climate (absolute humidity or temperature) is more distinct from the base seasonal signal. This dichotomy is illustrated with the model shown in Figure 2. While the magnitude of the effect is equal between the two panels, the absolute level of ρ_{CCM} is higher in (a), where the driver has strong seasonality. The seasonal cycle is trivial to predict, and so that the stronger the effect of the seasonal cycle, the easier the driver is to predict. Conversely, removing the seasonal cycle from the driver makes prediction more difficult, but it also removes the ambiguity caused by mutual seasonal correlation among variables.

This brings up another important point. In the case when "all other things are equal", the skill of CCM (ρ_{CCM}) can be used as a relative measure of interaction strength. However, this is not an applicable case to comparing CCM across latitudes, since the climate time series in tropical and temperate countries do not have the same basic levels of predictability. To compare the actual magnitude of effect across latitude, we instead can use scenario exploration with multivariate EDM.

We predict the change in resulting flu incidence, denoted Δflu, at historical points that would occur from small increases and decreases in the environmental driver. Figure

4a shows country-by-country that the magnitude of the effect of absolute humidity on flu, $\Delta flu/\Delta AH$, is roughly comparable between tropical and temperate countries. However, countries at high latitudes generally show a negative effect of absolute humidity on influenza, while low latitude countries generally show a positive effect. This would follow from a U-shaped response of flu survival to absolute humidity.

To examine this effect further, we aggregate the results of scenario exploration across all the countries. Figure 4b shows that at low AH, the effect of AH on flu is negative ($\Delta flu/\Delta AH < 0$), while at high AH the effect is positive ($\Delta flu/\Delta AH > 0$), confirming a U-shaped response of flu survival to absolute humidity. The negative effect at low AH on incidence and positive effect at high AH appear roughly equivalent, at least when weekly reported cases are normalized to average total reported cases in a year in that country. The same analysis for temperature (Fig. 4c) does not exhibit the same clear state-dependent effect (temperature variations). Temperature changes can have a positive ($\Delta flu/\Delta T > 0$) or negative ($\Delta flu/\Delta AH < 0$) effect on flu at the same temperature. To further investigate this question, we apply forecast improvement with multivariate EDM (14, 15). The results are summarized in Figure 5. Both globally and in the tropics specifically, AH and T show an effect on influenza (the mean improvement is greater than zero with $p < 0.05$). However, a paired t-test shows that the two cannot be distinguished ($p > 0.1$), as was the case with CCM. More notably, adding both AH and T together leads to an even great improvement ($p < 0.01$). This suggests that there could be some effect of temperature on influenza not mediated through absolute humidity. Unsurprisingly, in temperate countries where correlations between AH and T are extremely high (generally > 0.9), these variables contain almost identical information, and hence there is less difference in forecast skill on average between embeddings with AH, T, or both.

5.3 Discussion

Our results build on understanding from previous laboratory, statistical, and modeling studies. Prior population-level analysis has focused on seeking environmental explanations for influenza seasonality, which is most prominent in temperate countries (as is clear from Figure 5.1). By reframing the question more generally as identifying external drivers of nonlinear dynamics, we are able to provide additional insight into the global relationship between environment and influenza, showing that there are general rules that span temperate and tropical latitudes.

Cross-map analysis indicates that globally there is a causal effect of both temperature and absolute humidity on influenza that is distinct from their mutual seasonality (Figure 5.3). These results are augmented by analysis with multivariate EDM forecast improvement (Figure 5.5). Of key importance is that the results show that environmental drivers are important at all latitudes, regardless of the degree of seasonality in the environment and flu.

However, there is interest not only in the question of if environmental drivers are important across latitudes, but also which environmental drivers are most important. The tight statistical relationship between temperature and absolute humidity (stemming from reflects their fundamental physical relationship) makes this a difficult question to address at the population level. CCM shows clear evidence that temperature and absolute humidity have more direct effects on global influenza than precipitation and relative humidity (Fig. 5.3). Between absolute humidity and temperature, CCM finds a significant effect of absolute humidity in more countries, but the difference is small and not terribly compelling. Multivariate forecast improvement similarly fails to distinguish the two (Figure 5.5).

Both cross-mapping and multivariate forecast improvement of these methods

detect the bulk effect of a variable on another and do not always distinguish direct from indirect effects. A much sharper difference emerges from scenario exploration. Scenario exploration reveals a general relationship between absolute humidity and flu that is independent of country (Figure 5.5b). At low levels of absolute humidity, absolute humidity has a negative effect on influenza ($\Delta flu/\Delta AH < 0$); while at high levels of absolute humidity, it has a positive effect ($\Delta flu/\Delta AH > 0$). These results corroborate the molecular basis for a U-shaped effect of absolute humidity on influenza mediated by desiccation at high absolute humidity and disruption at low absolute humidity (11).

When the same analysis with scenario exploration is performed with temperature, there is no evidence for a simple general relationship (Fig. 5.5c). Rather, temperature appears to affect influenza both positively and negatively at the same values of temperature at different times. This suggests that the effect of temperature on influenza is strongly dependent on the state of other variables, for example if temperature has an indirect effect on influenza that is mediated by absolute humidity. Note however that there are particular temperature ranges where the variance of the effect seems to be much greater (e.g. around $82^o F$). This could indicate that there are important temperature thresholds for flu infection. It is important to keep in mind that multivariate forecast improvement showed that including both variables ultimately gave the best predictions of flu (Figure 5.5). Moreover, the molecular arguments for a U-shaped effect of absolute humidity on influenza also predict that temperature should be a control on the balance between the positive effect of absolute humidity via desiccation and the negative effect of absolute humidity via disruption (11).

With this in mind, we look at the effect of absolute humidity on influenza ($\Delta flu/\Delta AH$) as a function of temperature (Fig. 5.5d). Indeed, this is perhaps the most interesting picture to emerge, as there are a number of features that corroborate and elaborate existing ideas. (1) Temperature has a relatively loose control on ($\Delta flu/\Delta AH$)

when it is below $\approx 70^oF$, but the effect ($\Delta flu/\Delta AH$) is consistently negative. (2) The positive effect of absolute humidity appears more strongly controlled by temperature, and appears to be restricted to a narrow band of temperature between 75^oF and 85^oF. (3) At the highest temperatures, the effect of absolute humidity goes to zero in exact concordance with the laboratory finding that aerosol transmission of influenza is blocked at 30^oC (86^oF) (16).

The results reveal influenza-specific ranges for the temperature and absolute humidity effects. In particular, the balance between positive and negative effects of absolute humidity appears to shift somewhere between 70^oF and 75^oF. This is especially clear if we look back at the plot in Figure 5.4b that shows the effect of absolute humidity on influenza across the global range of absolute humidity. Figure 5.6 shows the same data, but now the points are split between two panels based on temperature. On the left are observations where temperature was below 75^oF; on the right are observations where the temperature was between 75^oF and 85^oF. The red lines represent the 0.1 and 0.9 quantile regressions. The quantile regressions indicate that the measured effect of AH on flu is almost always negative when $T < 75^oF$ and almost always positive when $75^oF \leq T \leq 85^oF$. At present, there do not appear to be modeling or laboratory results to compare these threshold results against. However, this population level result sets the stage for laboratory studies that experimentally test this threshold by varying temperature and humidity over the full range of conditions experienced globally. Notably, our analysis has sidestepped a number of important epidemiological processes, including strain dependent effects, spatial dynamics within countries (17), spatial dynamics between countries (18), and antigenic drift (19). Part of the power of EDM is that these factors-insofar as they interrelate to the deterministic dynamics in countrywide infection- are indirectly accounted for using lag-coordinate embeddings. However, the challenge for future research on flu with EDM will be to specifically incorporate these processes.

The fullest understanding surely will not emerge until all these factors can be treated integratively (20).

5.4 Methods

5.4.1 Data

Total laboratory confirmed influenza A & B cases per week were retrieved from the World Health Organization via FluNet by country (http://apps.who.int/globalatlas/dataQuery/). Ideally, we would like to analyze an index of incidence density (per capita), and thus need to account for population size and reporting rates. To account for changes in population size, we divide by linearly interpolated annual population data take from The World Bank: Health Nutrition and Population Statistics.

Accounting for changes in reporting rate over time is a more difficult issue to address. A typical approach is to divide weekly incidence by the total reported incidence for that country, that year. However, this masks all year-to-year differences in flu infections, including those that arise naturally from the nonlinear intrinsic dynamics of host-pathogen dynamics and from the state-dependent effect of climate variability. Such standardization would artificially inflate the seasonal signature in flu incidence, and hence would make the task of disentangling causality from shared seasonality harder, not easier.

However, accounting for the substantial differences in reporting rates between countries can be addressed to first order by dividing weekly incidence by the total reports per year in that country averaged over all years reported. Note that CCM is unaffected by arbitrary scaling, so this normalization only affects the comparisons between countries (Figures 5.4 & 5.5).

5.4.2 Empirical Dynamic Modeling

Empirical dynamic modeling is a general quantitative framework that centers on reconstructing and studying attractor manifolds of systems from empirical time series data. Dynamical systems are typically studied in terms of parametric equations, for example a SIR model of disease outbreak. These equations can then be solved to generate the changes of the system variables through time. When viewed in state space (multivariate space where each axis corresponds to a system variable), this becomes a trajectory that traces out the underlying dynamic attractor of the system (illustrated in the brief animation: http://simplex.ucsd.edu/RMM_S1.mov). The dynamic attractor is a complete representation of the unique system, and thus can be studied in place of parametric equations to predict and understand systems like host-pathogen dynamics.

Most immediately, historical points on the manifold represent similar system states, and similar states will follow similar trajectories forward in time. Thus, dynamics can be predicted using nearest neighbor forecasting, that is predicting the future trajectory of the system using the trajectories of historical points nearby on the manifold. Here we use simplex projection (23) and S-maps (24) for nearest neighbor forecasting.

This general framework of empirical attractor reconstruction and nearest neighbor forecasting can then be used in a number of ways. First, convergent cross-mapping (CCM) can be used to understand cause-effect relationships in nonlinear systems (12). The basic idea of CCM is that if states on the empirical manifold of X can be used to predict variable Y, this indicates that the dynamic signature of Y is present in X, and hence that Y caused X.

When driver variables can be treated as stochastic (e.g. seasonal anomalies of climactic variables), multivariate forecast improvement using S-maps can provide an additional test for causality (14, 15). A stochastic variable is considered causal if explicitly including that variable as a coordinate in the state space leads to improved

nearest neighbor forecasts.

Finally, when driver variables are explicitly included in the reconstructed state space, the idea of scenario exploration can be used to assess the dynamic the impact of environmental drivers on ecological or epidemiological dynamics (15). The key to scenario exploration is that nearest-neighbor forecasts are not constrained to predictions based on the current ecosystem state. Thus, when drivers are explicitly included, it is possible to make predictions for a given state of the biological variables (e.g. susceptible and infected individuals), but with varying values of the driver (e.g. absolute humidity). By comparing the predict effect e.g. of a small decrease and a small increase in the driver, it is possible to track the dynamic (changing in time) effect of the driver.

All EDM calculations can be done with the R package 'rEDM'. Prediction skill is always measured with Pearson's correlation (ρ) between observed and predicted values. Additional details on the exact calculations are given below.

5.4.3 Seasonality

The seasonal cycle is determined using a smoothing spline (smoothing parameter = 0.8) to the target variable across Julian day, where the spline is wrapped December 31st - January 1st. Unlike other methods for extracting seasonal cycles based on Fourier decomposition, this method works for both unbounded (e.g. temperature) and bounded (e.g. precipitation, influenza incidence) variables.

5.4.4 CCM Analysis and Seasonal Surrogates

For basic CCM analysis, we use simplex projection, which has a single parameter to select- the embedding dimension E. We select E based on the optimal prediction of cross-mapping lagged 1 week, then measure the un-lagged cross-map skill with this value

of E.

While checking for convergence in cross-map skill (i.e. that cross-map skill improves with the amount of data used) is a general way to distinguish cross-mapping from spurious correlation (12), we are concerned here with a more specific problem of distinguishing driving effects from mutual seasonality. This is more directly addressed by developing a null test with surrogate time series.

For a forcing variable $Z(t)$ (e.g. absolute humidity or temperature), we calculate the seasonal average $\bar{Z}$ as above and the anomaly from the seasonal average $\tilde{Z} = Z - \bar{Z}$. We then randomly shuffle (permute) the time indices of the seasonal anomalies. Adding the shuffled anomalies back to the season average gives a surrogate time series Z^* that has the same seasonal average as Z, but with random anomalies. If Z is in fact a driver of flu, then flu will not only be sensitive to the seasonal component of Z, but also to the anomalies. Thus, flu should better predict the real time series Z than the surrogate Z^*. In practice we repeat the shuffling procedure 500 times to produce an ensemble of surrogates.

5.4.5 Multivariate EDM: Scenario exploration

Scenario exploration with multivariate EDM (15) provides an empirical framework to assess the effect of a small increase in a physical driver (e.g. absolute humidity) on influenza incidence. We predict the effect of a small increase in absolute humidity or temperature on influenza 2-weeks later to understand the sensitivity of flu outbreaks to the environment. For each historical time point, t, we predict flu with a small increase ($+\Delta Z/2$) and a small decrease ($-\Delta Z/2$) in historically measured driver $Z(t)$. The difference in predicted flu is $\Delta flu = flu_{t+1}(Z = Z(t) + \Delta Z/2) - flu_{t+1}(Z = Z(t) - \Delta Z/2)$, and the ratio of $\Delta flu/Z$ quantifies the sensitivity of flu infection to the driver Z at time t. We use $\Delta Z = 0.2 g/m^3$ and $0.5^o F$ for absolute humidity and temperature, respectively.

These values correspond to approximately 5% of the standard deviation of these variables across all the countries analyzed. Forecasts were done using S-maps (24), with $E = 6$ and $\theta = 0.9$.

5.4.6 Multivariate EDM: Forecast Improvement

Previous implementation of forecast improvement with multivariate EDM focused on stochastic environmental drivers (15). In the case of season influenza, however, we should not regard the environmental time series as stochastic variables. Rather, in many cases the majority of the dynamics reflect the simple periodic cycle of the seasons. Therefore, information about the drivers is already contained in the univariate embedding (12, 21). Thus, we modify the method of (15) as follows.

We determine the optimal univariate embedding dimension, E^*, for each influenza time series following (22). A univariate embedding with dimension $E < E^*$ will be "under embedded", i.e. it will not contain full information about the system state and dynamics. In this case, incorporating information about a driver in a multivariate embedding will generally lead to an increase in forecast skill. Thus, we calculate the improvement in forecast skill using simplex projection of the univariate embedding with $E = E^* - 1$ and the same embedding but with the candidate environmental variable(s) included as a coordinate.

Acknowledgements

Chapter 5, in part is currently being prepared for submission for publication of the material. Mayer, M. Cyrus; Hernandez, Ryan; Basu, Sanjay; Sugihara, George. The dissertation author was the primary investigator and author of this material.

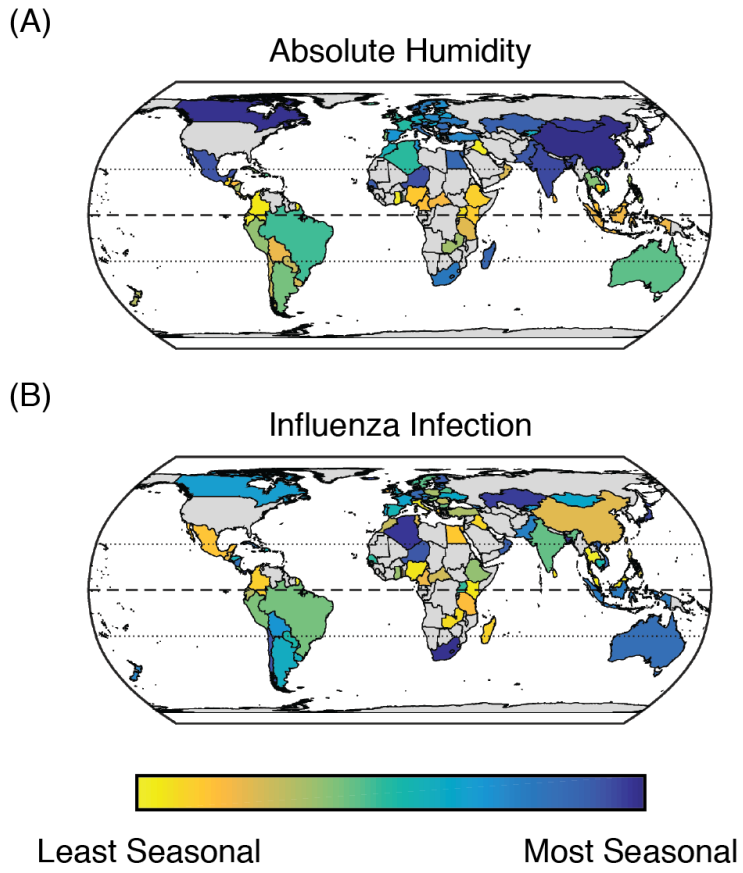

Figure 5.1: Correspondence between seasonality of environment and seasonality of influenza infection. Countries are colored from the least seasonal to most seasonal for absolute humidity (A) and influenza infection (B). The Spearman correlation between the two is high, $\rho = 0.73$.

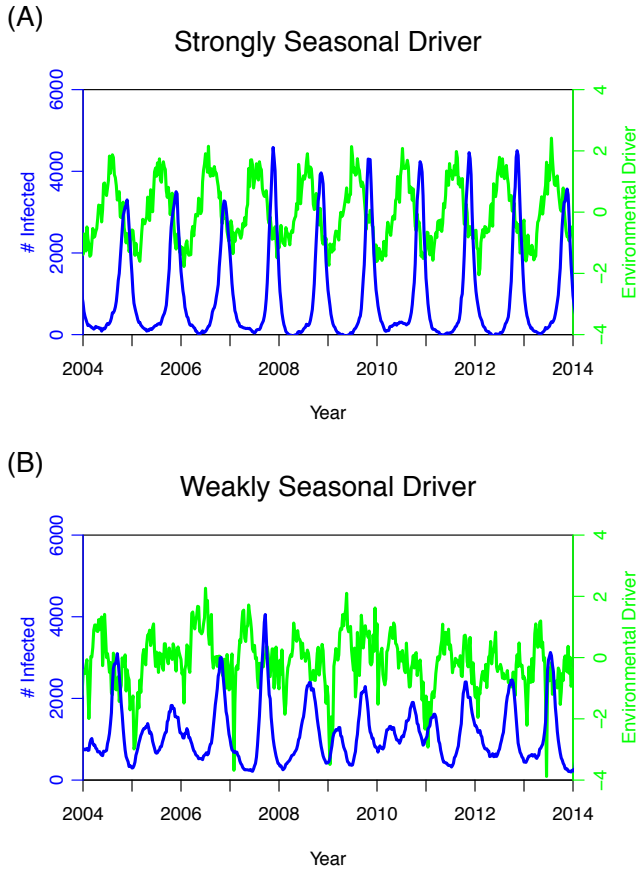

Figure 5.2: Stochastic SIRS model with strongly and weakly seasonal drivers. (A) A strongly periodic environment induces synchrony through dynamical resonance, causing peaks in infection to correlate with seasonal lows in the seasonal environment. (B) If the same SIRS model is driven by a seasonal signal with the same variance but much weaker seasonality, there is no dynamic resonance, infection peaks show much weaker seasonality, and correlation between infection and environment is much lower.

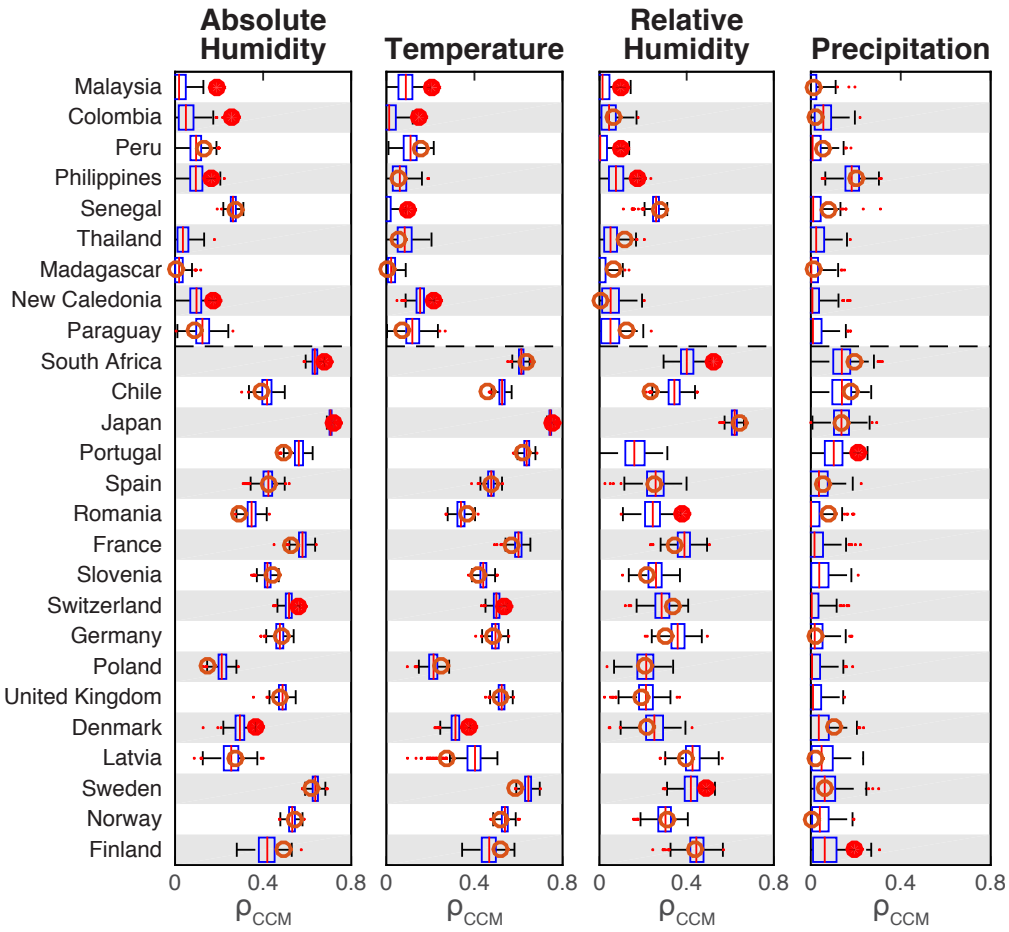

Figure 5.3: Detecting cross-map causality beyond shared seasonality of environmental drivers on influenza. Red circles show the measured cross-map skill (ρ_{CCM}) for observed influenza predicting purported seasonal drivers: absolute humidity, temperature, relative humidity, and precipitation. Together with this, box-and-whisker plots show the null distributions for ρ_{CCM} expected from random surrogate time series that shares the same seasonality as the true environmental driver. Countries are ordered according to distance from the equator (absolute latitude). Filled circles indicate that the measured ρ_{CCM} is significantly better than the null expectation ($p < 0.05$).

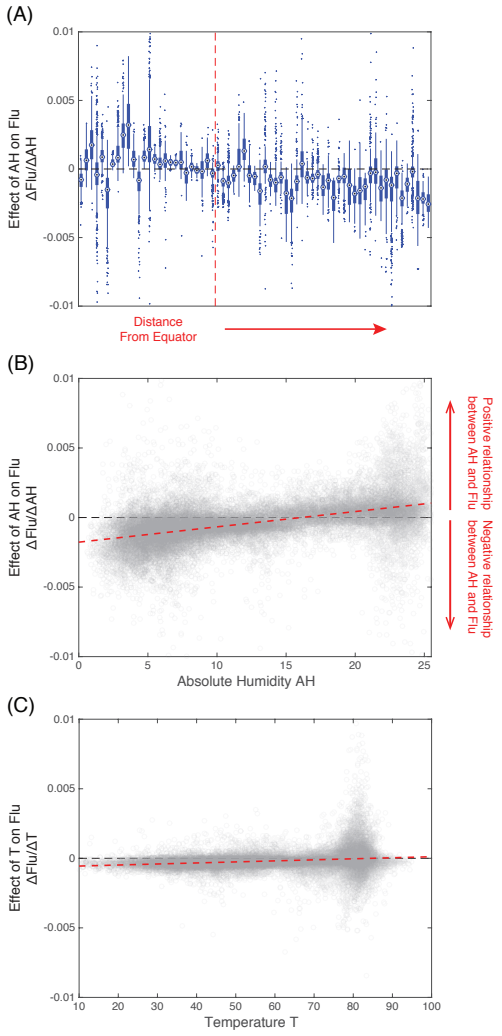

Figure 5.4: Scenario exploration with multivariate EDM. We measure the effect of environment on influenza infection by predicting the change in influenza (Δflu) that results from a small change in absolute humidity (ΔAH) or temperature (ΔT). Panel (A) shows the range of values for $\Delta flu/\Delta AH$ for each country across latitude. Countries closest to the equator tend to show a positive effect of AH on influenza infection, while countries furthest from the equator show a negative effect. Panel (B) shows the effect of absolute humidity on flu ($\Delta flu/\Delta AH$) as a function of AH grouped over all countries. At low AH (typical of high latitude countries), AH has a negative effect on flu infection, while at high AH (typical of low latitude countries), AH has a positive effect on flu. Similarly, (C) shows the effect of temperature on flu ($\Delta flu/\Delta T$) as a function of T. Evidence of a single global effect is much weaker, but it suggests there might be important temperature thresholds.

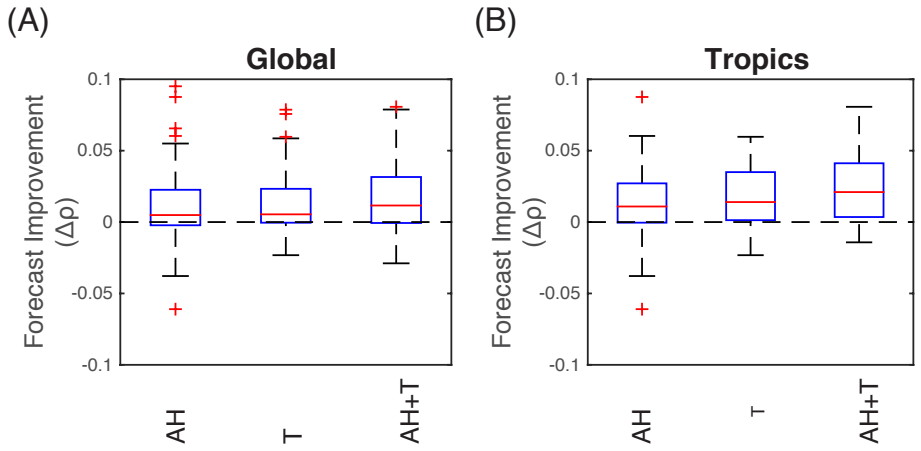

Figure 5.5: Forecast improvement with multivariate EDM. Causal effect is demonstrated if EDM forecast skill (ρ) improves when a driver variable is included in the EDM model. This is quantified by $\Delta\rho = \rho(withdriver) - \rho(withoutdriver)$, where ρ is the Pearson's correlation between observations and EDM predictions. Including either absolute humidity (AH) or temperature (T) leads to significant ($p < 0.05$) improvement in forecast skill both globally (panel A) and the tropics specifically (panel B). However, even greater improvement results from including both ($AH + T$), suggesting that there are compound effects of temperature and humidity.

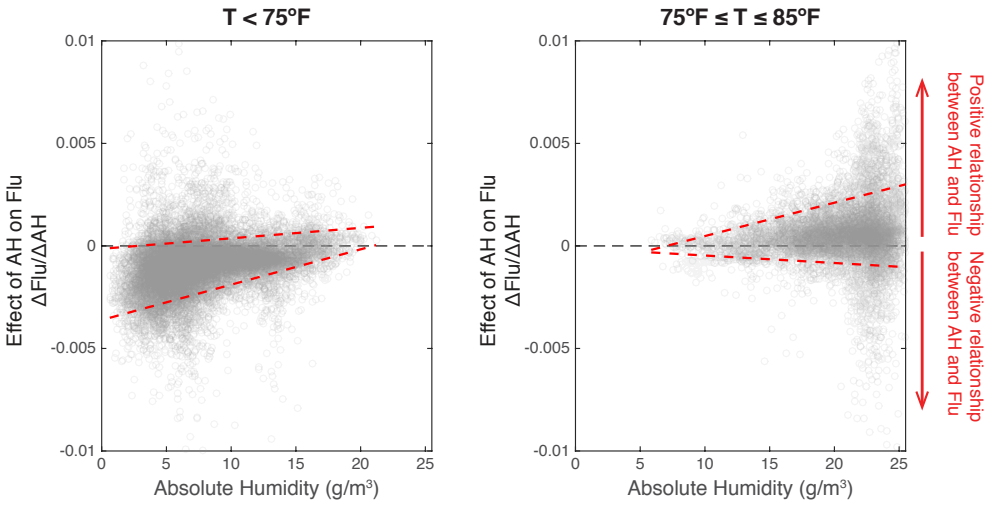

Figure 5.6: Temperature thresholds in the effect of absolute humidity on influenza. The results of scenario exploration in Figure 5.4b are re-plotted based on temperature. Values on the left correspond to observations when T was below 75^oF, while values on the left correspond to observations when T was between 75^oF and 85^oF. The red dashed lines indicate the 0.1 and 0.9 quantile regressions.

5.5 References

1. Cannell JJ, Vieth R, Umhau JC, Holick MF, Grant WB, Madronich S, Garland CF, Giovannucci E (2006) Epidemic influenza and vitamin D. Epidemiol Infect 134:1129-1140.

2. Dowell SF (2001) Seasonal variation in host susceptibility and cycles of certain infectious diseases. Emerging Infect Dis 7:369-374.

3. Xie X, Li Y, Chwang ATY, Ho PL, Seto WH (2007) How far droplets can move in indoor environments–revisiting the Wells evaporation-falling curve. Indoor Air 17:211-225.

4. Lowen AC, Mubareka S, Steel J, Palese P (2007) Influenza virus transmission is dependent on relative humidity and temperature. PLoS Pathogens 3:1470-1476.

5. Shaman J, Kohn M (2009) Absolute humidity modulates influenza survival, transmission, and seasonality. Proc Natl Acad Sci USA 106:3243-3248.

6. Tamerius JD et al. (2013) Environmental Predictors of Seasonal Influenza Epidemics across Temperate and Tropical Climates. PLoS Pathogens 9:e1003194. Tamerius, J. D.,Shaman J, Alonso WJ, Bloom-Feshbach K, Uejio CK,Comrie A, Viboud C (2013). Environmental Predictors of Seasonal Influenza Epidemics across Temperate and Tropical Climates. PLoS pathogens 9:e1003194.

7. Shaman J, Pitzer VE, Viboud C, Grenfell BT, Lipsitch M (201) Absolute Humidity and the Seasonal Onset of Influenza in the Continental United States. PLOS Biology 8:e1000316.

8. Lowen AC, Steel J (2014) Roles of Humidity and Temperature in Shaping Influenza Seasonality. Journal of Virology 88:7692-7695.

9. Azziz-Baumgartner E, Dao CN, Nasreen S, Bhuiya MU, Mah-E-Muneer S, Mamun AA, Sharker MAY, Zaman RU, Cheng PY, Klimov AI, Widdowson MA, Uyeki TM, Luby SP, Mounts A, Bresee J (2012) Seasonality, Timing, and Climate Drivers of Influenza Activity Worldwide. Journal of Infectious Diseases 206:838-846.

10. Schaffer FL, Soergel ME, Straube DC (1976) Survival of airborne influenza virus: Effects of propagating host, relative humidity, and composition of spray fluids. Archives of Virology 51:263-273.

11. Minhaz Ud-Dean SM (2010) Structural explanation for the effect of humidity on persistence of airborne virus: seasonality of influenza. J Theor Biol 264:822-829.

12. Sugihara G, May R, Ye H, Hsieh CH, Deyle E, Fogarty M, Munch S (2012) Detecting causality in complex ecosystems. Science 338:496-500.

13. Dushoff J, Plotkin JB, Levin SA, Earn DJD (2004) Dynamical resonance can account for seasonality of influenza epidemics. Proc Natl Acad Sci USA 101:16915-16916.

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

15. Deyle ER, Fogarty M, Hsieh CH, Kaufman L, MacCall AD, Munch SB, Perretti CT, Ye H, and Sugihara G. Predicting climate effects on Pacific sardine. Proc Natl Acad Sci USA. 2013 Apr 16;110(16):6430-5.

16. Lowen AC, Steel J, Mubareka S, Palese P (2008) High temperature (30 degrees C) blocks aerosol but not contact transmission of influenza virus. Journal of Virology 82:5650-5652.

17. Viboud C, Bjornstad ON, Smith DL, Simonsen L, Miller MA, Grenfell BT (2006) Synchrony, waves, and spatial hierarchies in the spread of influenza. Science 312:447-451.

18. Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, Gregory V, Gust ID, Hampson AW, Hay AJ, Hurt AC, de Jong JC, Kelso A, Klimov AI, Kageyama T, Komadina N, Lapedes AS, Lin YP, Mosterin A, Obuchi M, Odagiri, Osterhaus ADME, Rimmelzwaan GF, Shaw MW, Skepner E, Stohr K, Tashrio M, Fouchier RAM, Smith DJ (2008) The global circulation of seasonal influenza A (H3N2) viruses. Science 320:340-346.

19. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus ADME, Fouchier RAM (2004) Mapping the Antigenic and Genetic Evolution of Influenza Virus. Science 305:371-376.

20. Lofgren E, Fefferman NH, Naumov YN, Gorski J, Naumova EN (2007) Influenza seasonality: underlying causes and modeling theories. Journal of Virology 81:5429-5436.

21. Stark J, Broomhead DS, Davies ME, Huke J (2003) Delay Embeddings for Forced Systems. II. Stochastic Forcing. J Nonlinear Sci 13:519-577.

22. Glaser SM, Fogarty MJ, Liu H, Altman I, Hsieh C-H, Kaufman L, MacCall AD, Rosenberg AA, Ye H, Sugihara G(2013) Complex dynamics may limit prediction in marine fisheries. Fish and Fisheries.

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

24. Sugihara G (1994). Nonlinear forecasting for the classification of natural time series. Phil T Roy Soc A 348: 477-495.

Chapter 6

Summary

Although the chapters of this dissertation cover diverse areas of research, when taken together they are a multifaceted investigation into the use of empirical dynamic methods for ecosystem-based management. In particular, this dissertation describes several applied EDM methods for studying state-dependent interactions.

In many ways, the approach presented in Chapter 4 represents the ideal situation. When all the important variables are measured in an ecosystem, the system dynamics can be reconstructed in a so-called "native" state space. Thus, when a local linear modeling approach (using S-maps) is applied to these data, the local linear coefficients capture the partial derivatives that correspond to the dynamic interaction strengths.

However, in marine science, we rarely expect to have full-ecosystem observation. More often just a few important variables are measured, such as fish abundance and temperature. When important variables are unobserved, it is not possible to recover the native state-space dynamics. The original embedding theorem of Takens (Takens 1981) established the framework for reconstructing state-space dynamics from a single time series using time-lag coordinates in lieu of the unobserved variables.

A great deal of insight can be had from applying univariate reconstruction to

ecological data (Sugihara and May 1990, Sugihara 1994, Hsieh et al. 2005, Hsieh and Ohman 2006, Sugihara et al. 2011, Glaser et al. 2013), including the critical task of identifying causal interactions(Sugihara et al. 2012). Nevertheless, univariate embeddings lack the mechanistic insight that's possible in native embeddings. For example, it is still possible to use local linear regression with S-maps to approximate the system dynamics, but the coefficients correspond to partial derivatives that have no immediate mechanistic meaning to us, i.e. the effect of species X at time t-1 on species X at time t when all other time lags of X are held constant.

Multivariate embeddings represent a middle ground that is critical to management. They cover the common case where some but not all of the important variables are observed, and so multiple time-series are used with their lags to reconstruct the state-space dynamics. Multivariate embeddings were conjectured (Sauer et al. 1991) and heuristically found to be insightful (Dixon et al. 1999). However, Chapter 2 in this dissertation provides a formal mathematical foundation for the approach that confirms its generality.

With multivariate embeddings, scenario exploration can be used to assess the dynamic impact of environmental drivers on ecological dynamics. When drivers are explicitly included, it is possible to make predictions for a given state of the biological variables, but with varying values of the driver. In Chapter 3, this idea is explored in the case of the temperature effect on Pacific sardine, and in Chapter 5, it is applied to understanding climate drivers of seasonal influenza.

What is the upside of these results in practical management? That is an open question. Preliminary work shows that CCM and multivariate EDM can identify multi-species effects of fishing in well observed case-studies, such as the juvenile albacore tuna fishery in the United States. In theory, scenario exploration would then be possible to explore trade-offs between different multi-species harvesting strategies. This capability

gets more to the key practical questions of ecosystem management.

However, Nevertheless, multi-objective programming is extremely difficult, falling afoul of the so called “curse of dimensionality”. It is computationally intensive, and can be very sensitive to measurement or modeling error. Thus, even if a method exists, exhaustive multi-species cost-benefit analysis may be completely impractical. Looking forward, I question whether species-by-species analysis will really bear fruit for practical management. Rather, the most promise may well lie in considering management that keys to larger ecosystem indicators like resilience.

6.1 References

1. Dixon, P. A., M. J. Milicich, and G. Sugihara. 1999. Episodic fluctuations in larval supply. Science 283:1528-1530.

2. Glaser, S. M., H. Ye, and G. Sugihara. 2013. A nonlinear, low data requirement model for producing spatially explicit fishery forecasts. Fisheries Oceanography 23:45-53.

3. Hsieh, C.-H., and M. D. Ohman. 2006. Biological responses to environmental forcing: the linear tracking window hypothesis. Ecology 87:1932-1938.

4. Hsieh, C.-H., S. M. Glaser, A. J. Lucas, and G. Sugihara. 2005. Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. Nature 435:336-340.

5. Sauer, T., J. Yorke, and M. Casdagli. 1991. Embedology. Journal of Statistical Physics 65:579-616-616.

6. Sugihara, G. 1994. Nonlinear forecasting for the classification of natural time series. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 348:477-495.

7. Sugihara, G., and R. M. May. 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time-series. Nature 344:734-741.

8. Sugihara, G., J. Beddington, C.-H. Hsieh, E. Deyle, M. J. Fogarty, S. M. Glaser, R. Hewitt, A. B. Hollowed, R. M. May, S. B. Munch, C. Perretti, A. A. Rosenberg, S. Sandin, and H. Ye. 2011. Are exploited fish populations stable? Proceedings of the National Academy of Sciences of the United States of America 108:E1224-E1225.

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

10. Takens, F. 1981. Detecting strange attractors in turbulence. Pages 366-381 in D. A. Rand and L. Young, editors. Dynamical systems and turbulence, Warwick 1980. Lector Notes in Mathematics, New York, U.S.A.