Introduction to Empirical Dynamics

A Quick Look Ahead at
New Approaches to Difficult Problems

Nonlinearity, Prediction, Coupling and Causation

Sugihara *SIO 276 Fall 2019*

Every one gives lipservice to nonlinearity, but few actually acknowledge it or truly understand it.

I would like to look at some pedagogical points that are often swept under the rug… but I suggest they actually redefine how we should study these systems

---

An Introduction to Empirical Dynamics:

An Inductive Data-Driven Approach

My aim here will be to speak to a particular perspective that may be of special relevance as we move away from simple 20th century reductionist toy models based on fundamental principles, toward trying to understand how messy natural systems behave. For example, while we can easily write down an accurate equation for diffusion of gases in a test tube, modeling oxygen concentrations at depth in a large lake (where biology, complex chemistry and physical currents intervene) is impractical with explicit equations. Empirical models, which infer patterns and associations from the data (instead of using hypothesized equations), represent an alternative and highly flexible approach.

All this is being made possible by the era of Big Data. 21st century holistic science is being enabled by a boon in available data, and EDM is a useful approach for data exploration. The math itself is not especially challenging, however the resonance of understanding that can be achieved with a deeper understanding of the implications of simple classical assumptions like equilibrium,

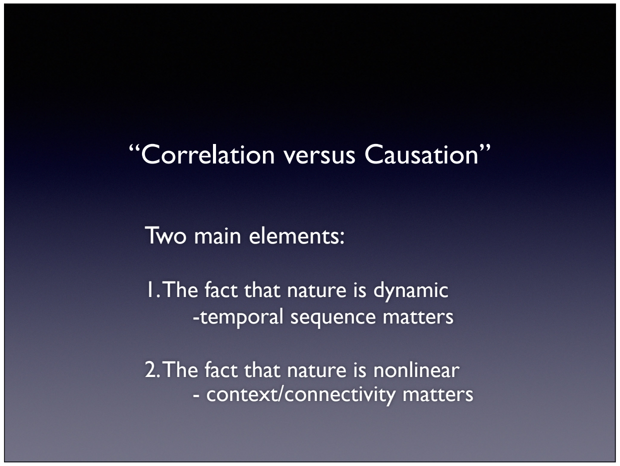
---

Two key points of emphasis for inductive data-science are as follows:

1. Detecting causation to uncover mechanism in nonlinear dynamic systems
2. Forecasting as a rigorous way to validate understanding

Two key points of emphasis are:

1. Detecting causation to uncover mechanism in natural nonlinear dynamic systems
2. Forecasting as a rigorous way to validate understanding.



1) The fact the nature is dynamic - temporal sequence matters

Nature is best understood as a movie rather than a snapshot.

2) The fact that nature is nonlinear

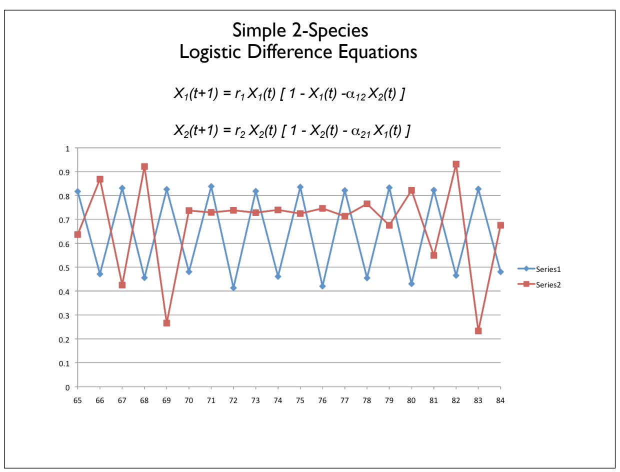
Meaning it consists of interdependent parts…that are nonseparable

– context matters

It cant't be studied as independent pieces. Rather each piece needs to be studied in the context surrounding it.

Let's start with an example….

stop



these two variables (eg.**species and a forcing variable**) are uncorrelated.

However they are in fact **deterministically coupled**. *click* *click* These time series are produced from a simple coupled logistic difference system. … an example of nonlinear dynamics.

A = [[3.5 0.1][0.02 3.8]]

So

x1(t+1) = x1(t) * (3.5 - 3.5 * x1(t) - 0.1 * x2(t))

x2(t+1) = x2(t) * (3.8 - 3.8 * x2(t) - 0.02 * x1(t))

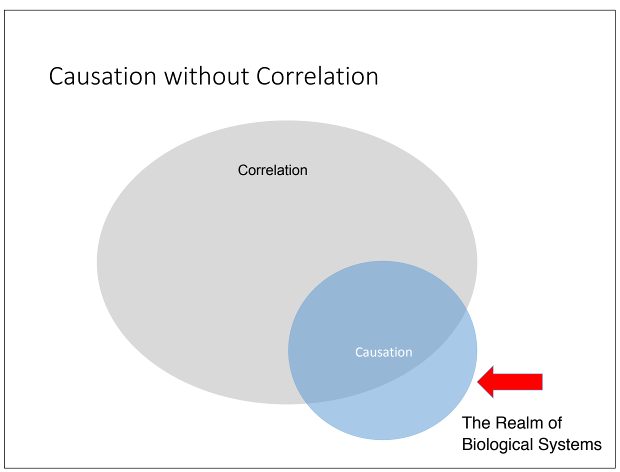
x1(0) = 0.2

x2(0) = 0.4

Bishop Berkeley: “Correlation does not imply causation.” (1710)

Converse: *“Lack of correlation does not imply lack of causation.”*

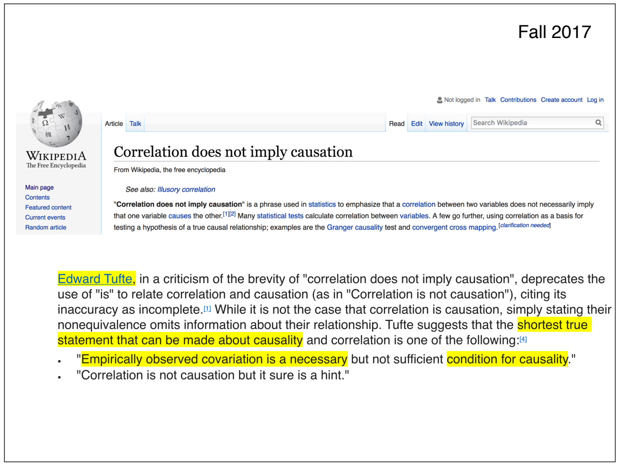
Thus not only does correlation not imply causation, but with simple **nonlinear dynamics**, the **converse is true:** Lack of correlation does not imply lack of causation.



This is interesting because this blue disc is the realm of biological systems.

stop

And **within this realm…a further consequence of nonlinearity** demonstrated in the **model example** was the phenomenon of **mirage correlation**…



Fall 2017

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

Wikipedia
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article

Correlation does not imply causation

From Wikipedia, the free encyclopedia

See also: Illusory correlation

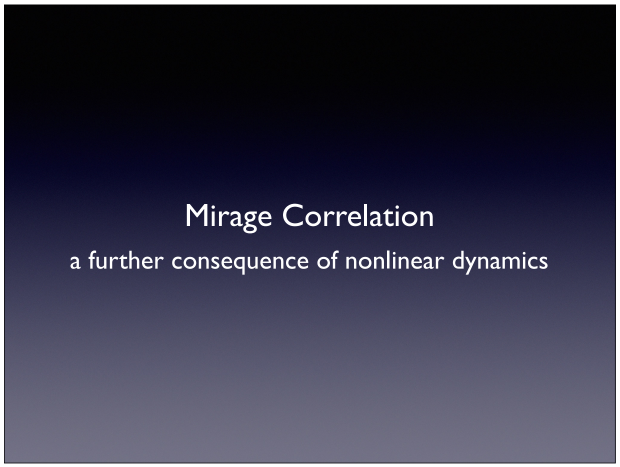
"**Correlation does not imply causation**" is a phrase used in statistics to emphasize that a correlation between two variables does not necessarily imply that one variable causes the other.[1][2] Many statistical tests calculate correlation between variables. A few go further, using correlation as a basis for testing a hypothesis of a true causal relationship; examples are the Granger causality test and convergent cross mapping.[*clarification needed*]

Edward Tufte, in a criticism of the brevity of "correlation does not imply causation", deprecates the use of "is" to relate correlation and causation (as in "Correlation is not causation"), citing its inaccuracy as incomplete.[1] While it is not the case that correlation is causation, simply stating their nonequivalence omits information about their relationship. Tufte suggests that the shortest true statement that can be made about causality and correlation is one of the following:[4]

- "Empirically observed covariation is a necessary but not sufficient condition for causality."
- "Correlation is not causation but it sure is a hint."

In retrospect, after what I just showed you, this converse property should be well-known, but apparently it is not. It contradicts a currently held view that correlation is a "necessary" condition for causation.

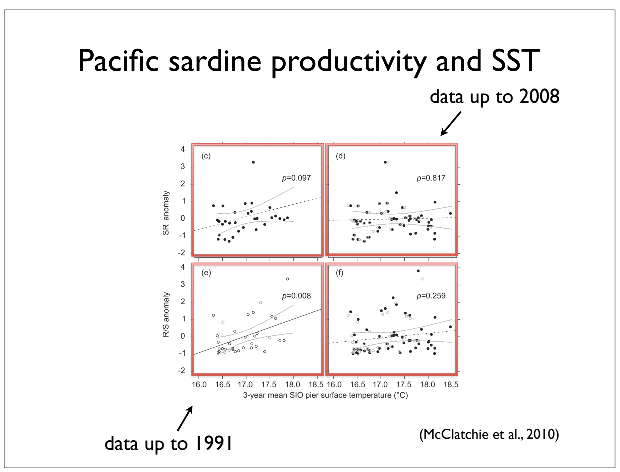
Tufte is a distinguished statistician and political scientist from Yale



These ephemeral or mirage correlations are "associations that come and go and even switch sign"

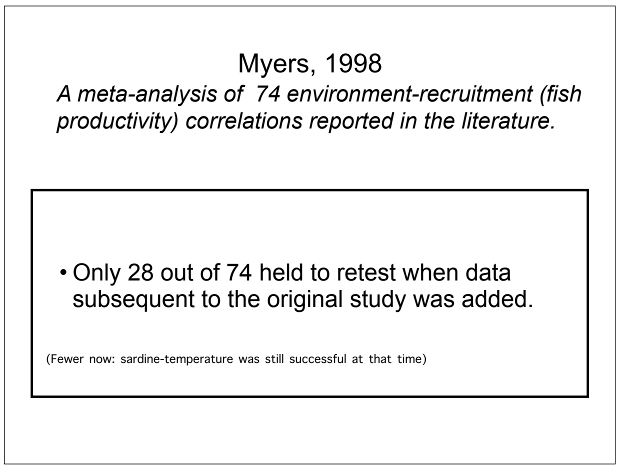
This perverse tendency of nonlinear systems is the bane of Ecology and of financial modeling …. relationships that appear then disappear as soon as you try to exploit them.

Let's see an example…



Here is another example from SoCal. *click* Using data up to 1991, a significant positive relationship was found between sea surface temperature and sardine production (true for two different measurements of productivity (recruitment)). This was reported in 1994 and was subsequently written into the state law for managing harvests.

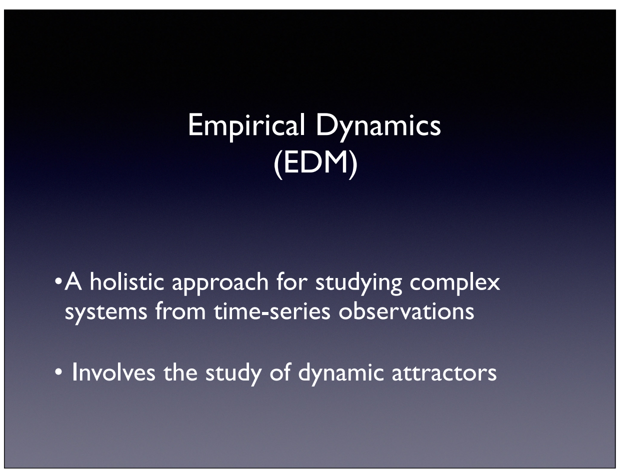
click *click* However, when data from 1992-2008 are included (17 additional data points), the correlation seemed to disappear (in both cases), causing the plan to be suspended in 2010… where it now stands.



Another famous example from fisheries-
was a meta-analysis on 74 environment-recruitment correlations that were reported in the literature. These correlations were retested using additional data obtained subsequent to the publication of each of the studies – only 28 of the 74 correlations remained.
(Certainly fewer now, since sardine-temp was among the ones that still help up at the time of Myers analysis)

Relationships we thought we understood seemed to disappear. This sort of thing is familiar in finance where relationships are uncovered but often disappear even before we try to exploit them.

(Species included Atlantic cod, Northern Anchovy, Sockeye salmon, Maine Lobster, and many others).



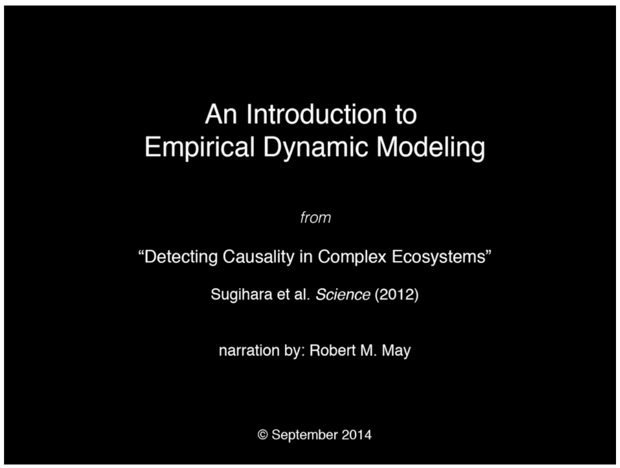
So, how to address this?
The approach I will present here is based on nonlinear state space reconstruction which I refer to here with the less technical name… empirical dynamics.
EDM is a holistic data-driven approach for studying complex systems from their attractors. It is designed to address nonlinear issues such as mirage correlation.

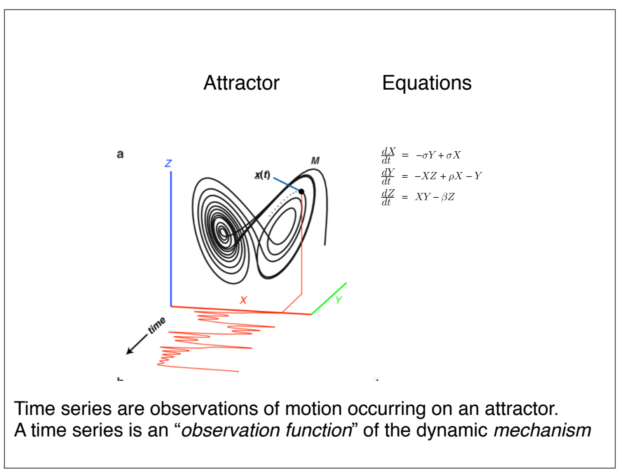
I am now going to play a brief video animation that will explain all. (my son made this for me when he was a junior at Columbia). The narration is by Robert May.
clickclick**

It is an alternative to the theoretical expedient of constancy and decomposability.
The common assumption that natural systems are in equilibrium has legitimized reductionism and the use of

linear methods. For example, to study dynamics —we can use local linear stability analysis.
-constancy in pairwise interactions- a picture of independence; dynamics are reduced to random motion around a mean. Time (sequence of events) is irrelevant

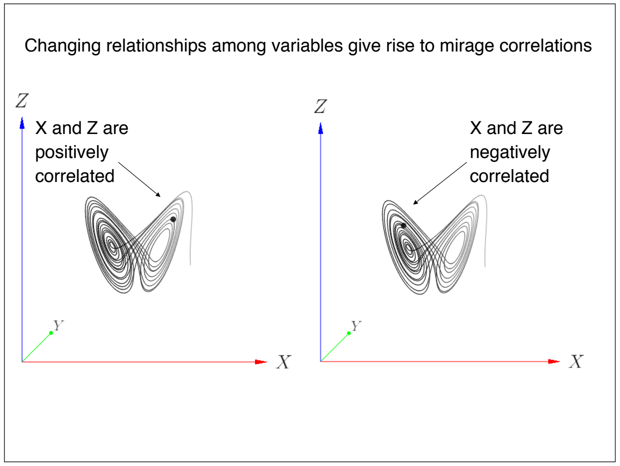
However, if we don't make this assumption then we need to account for dynamics that exhibit nonlinear state dependence
-nonlinear state dependence -> interdependence
This has important implications for how to study nature (holistically), and for identifying causal drivers and networks.





The main insight from that video is ….
to understand that a time series is a projection or observation of motion on an attractor. Indeed in the jargon term of dynamical systems a time series is an "*observation function*" for dynamics on the attractor.

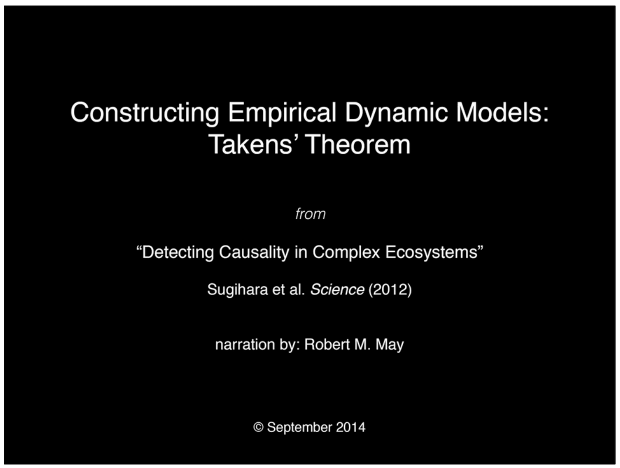
Conversely, attractors can be obtained simply by *re-plotting the relevant time series data*. Constructing attractors from time series data is the basis of the Empirical Dynamic approach.



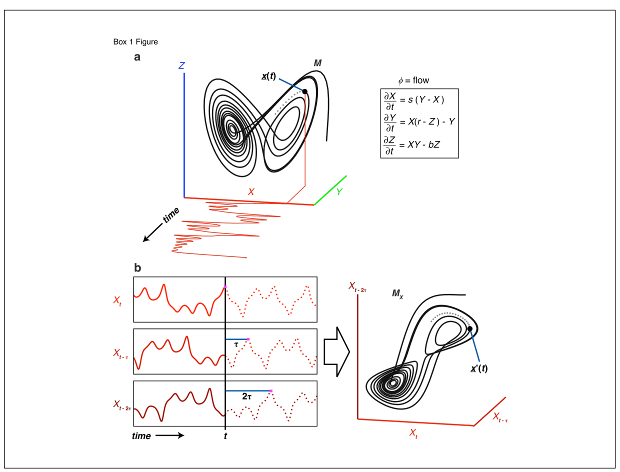
And depending on when they are viewed, relationships among variables can appear to change… giving rise to mirage correlations

Over the short term there might be correlations, but over the longer term If one were to study this system by plotting randomly sampled values of X and Z there would be no correlation. This problem only becomes coherent when temporal sequence is included.

Let's look at a real example.



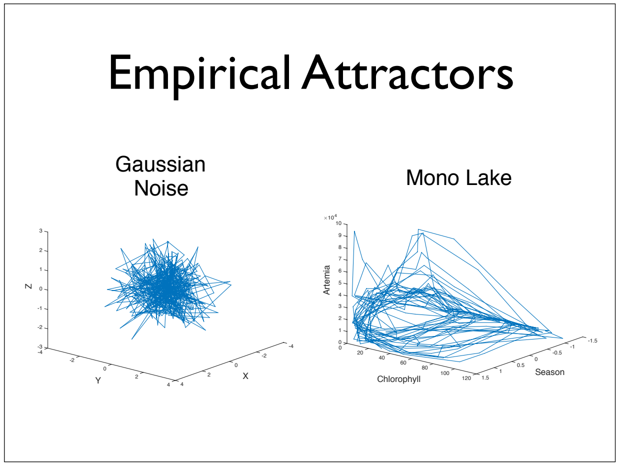
I'll now play another short video that explains a key result for EDM, related to connectivity and information sharing.



To recap, Takens theorem says any one variable contains information about the others. This allows the construction of attractors from a single variable using lags as proxy coordinates.

Constructing attractors from time series data is the basis of the Empirical Dynamic approach.
-univariate
-mulitvariate
-mixed embedding

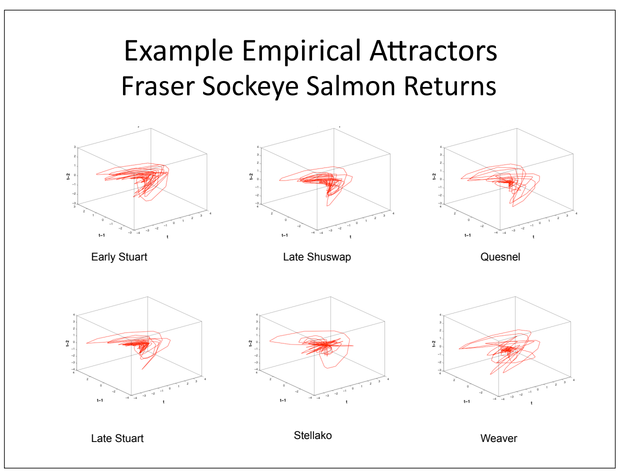
let's look at some examples…



When time series have no relationship to each other, plotting them together as a trajectory in a multivariate state space yields a tangled mess. There is no sign of structure or pattern— and there shouldn't be!

[[Click to show 2nd attractor]]

In contrast, when interrelated time series are plotted together, the trajectory forms a manifold. Here, we show three time series from Mono Lake, a saline lake in California with a simple food web. The trajectory forms a coherent pattern that we can then study to make predictions and gain insights into the interactions between the variables.



Here we have an ecological example: attractors constructed from time series for sockeye salmon returns for the Fraser River, Canada. …Again, using time-lagged coordinates.

-Each point represents a 3-year history.

-Basically, the trajectories run along consecutive 3-year histories.

The fact that **3 dimensions are sufficient to unfold** the trajectories suggests it may be **possible** to make a reasonable **3-factor multivariate model** with well-chosen mechanistically relevant time series (**eg. river discharge, SST and spawning stock abundance)**

*****Full Stop*****

State Shifts in Nanog

Stem cell transitions from the undifferentiated (pluripotent) to the differentiated state

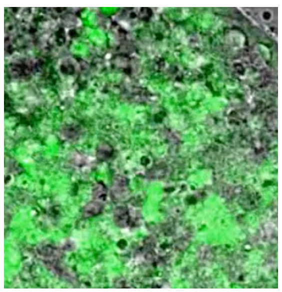
A comparison of taking a static versus a dynamic view

Verma Lab, Salk & Sugihara Lab SIO
Gerald Pao, Ethan Deye

20

This is another example.

Viewing Nanog Gene Expression from a Statistical Snapshot (Static View)



- Nanog gene manipulation
- Mouse stem cells engineered to produce GFP when Nanog gene is actively expressed
- In static snapshot, can see that some cells have high expression, some low.

Collaboration with Verma Lab, Salk
Gerald Pao, Ethan Deyle

21

Nanog is a transcription factor that keeps stem cells pluripotent.

Mouse stem cells are engineered to produce Green Florescent Protein (GFP) when the Nanog gene is actively expressed.

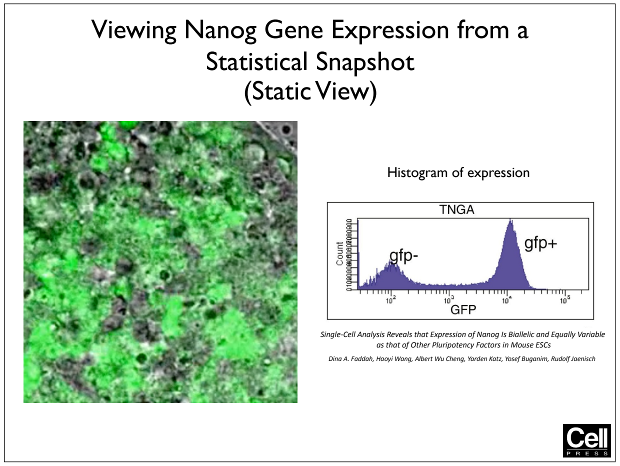
In this snapshot we see that some cells have high expression (green) and some low (dark)....low states are when the cell differentiates. We don't see much in between.

Transition to low states was believed to be stochastic

****

Nanog works to maintain pluripotence even w/o lif (leukemia inhibitory factor))

GFP was inserted into the nanog locus in one strand only

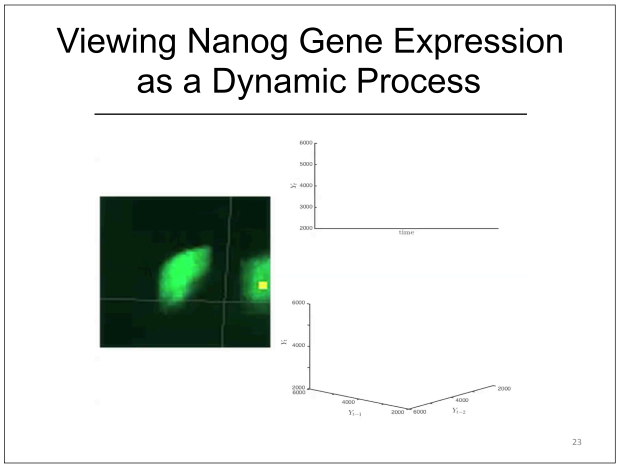


This idea of cells randomly transitioning between states comes from taking a static statistical view

That is, we assume the probabilistic average you see across this ensemble of cells represents what we expect of any individual cell.

When we do this and plot the histogram we see 2 discontinuous states. (*click*)

and the reigning hypothesis is that switching states is purely random.



In fact we can draw out the attractor (in 3D here, though the 5D is the best embedding dimension for these time series data) and see that the transition between states involves visiting two different parts (lobes) of a manifold.

This suggests that the rapid transitions between states may be understandable as a nonlinear dynamic phenomenon rather than a purely stochastic one.

This is nascent work, but I think it conveys the idea that dynamic tools are useful when studying what is essentially a dynamic process.

And in particular, that understanding causal interactions in such systems really requires a nonlinear dynamic perspective.

Prediction

Two methods:
Simplex projection S-maps

Out of sample forecasting is a rigorous way to validate understanding

Model fitting is not prediction!

In my view, prediction should be the standard (measure of merit) for validation in science. (Indeed it seems odd that it is not generally so)
Fisheries models
GCM's
Hydrology etc.

Forecasting with Empirical Dynamics

- *Simplex Projection*
- *S-maps*

We will present two basic methods:
Simplex projection and S-maps.
Many other possibilities exist. These are just two very simple ones.

Simplex Projection

Sugihara and May 1990

0th order nonlinear prediction method

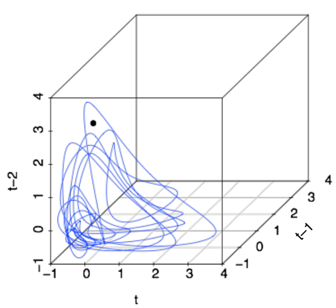
- to predict $X(t+1)$, look for point $\underline{x}(t)$ on the manifold
  - find its nearest neighbors ($\underline{x}(nn_1)$, $\underline{x}(nn_2)$, etc.)
  - see where they went ($X(nn_1+1)$, $X(nn_2+1)$, etc.)
  - take a weighted average

This is simply forecasting using nearest neighbor analogues.

Nearest neighbors on the attractor are "points with similar histories"

Let's see how this works!

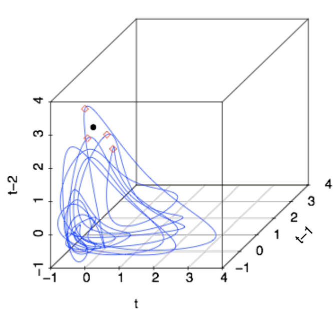
Simplex Projection



Again, each "point" on the manifold is a "history vector".... a history fragment

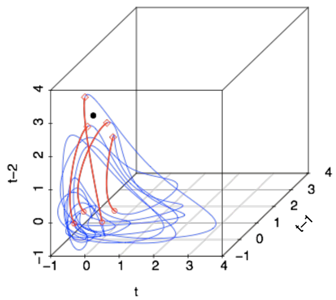
---

Simplex Projection



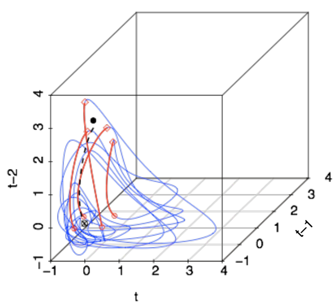
Nearest neighbors are "points with similar histories"

Simplex Projection

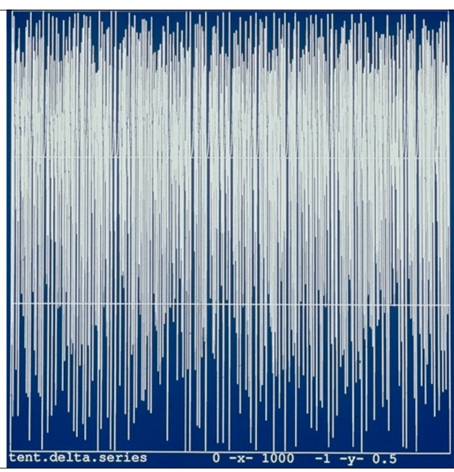


Track where they went

Simplex Projection



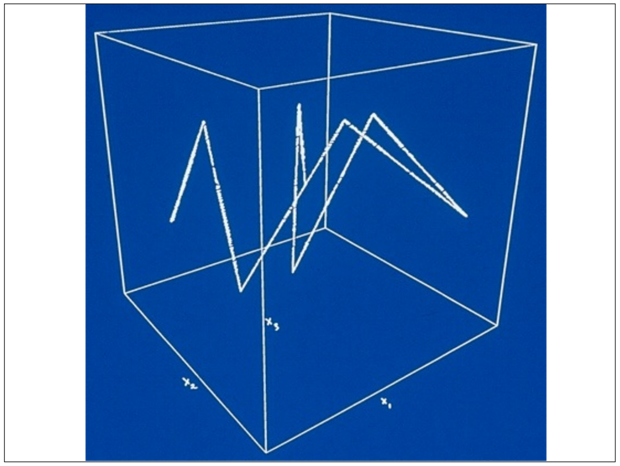
The prediction is a weighted average of the neighbors fates.



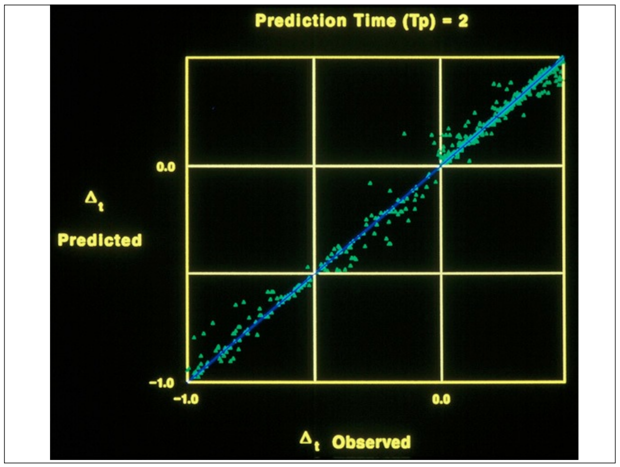
White noise (statistically not predictable)

First half second half

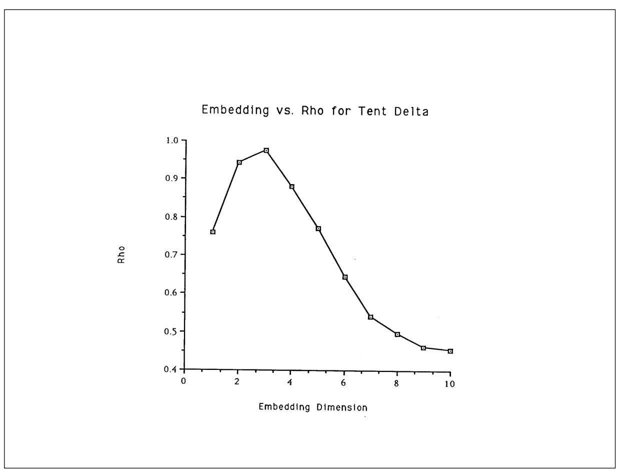
let’s see what we get with time-lagged coordinates



This is what we find in 3 dimensions (Fork with 3 tynes…. X, Y, Z).



How did I choose 3 dimensions to embed this?



How did I choose 3 dimensions to embed this?

---> Predictability

The embedding with the best predictability is the one that best resolves singularities… best unfolds the attractor.

Let me explain.

Whitney Embedding Theorem

A D-dimensional object can always be embedded in 2D+1 dimensions.

Note:
In EDM the embedding dimension gives an upper bound on the minimum number of variables required to model the system to obtain a given level of predictability. It is not absolute, but depends on the length and noisiness of the specific data

Ball of thread example

Again, we use prediction to find the embedding that best resolves singularities.

S-Map

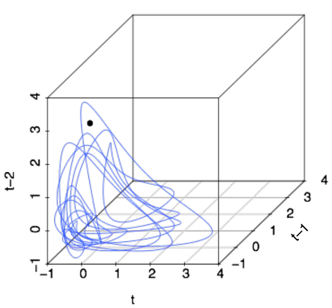
Sugihara 1994

- Weighted AR-prediction method computed over points on an attractor manifold.
- Model parameter, θ, controls weighting applied to points in local vs global state space
  - $\theta = 0$: all attractor points weighted equally (linear model, hyperplane, flat manifold, not state dependent, separable)
  - $\theta > 0$: local points weighted more (nonlinear model, nonseparable, curved manifold)

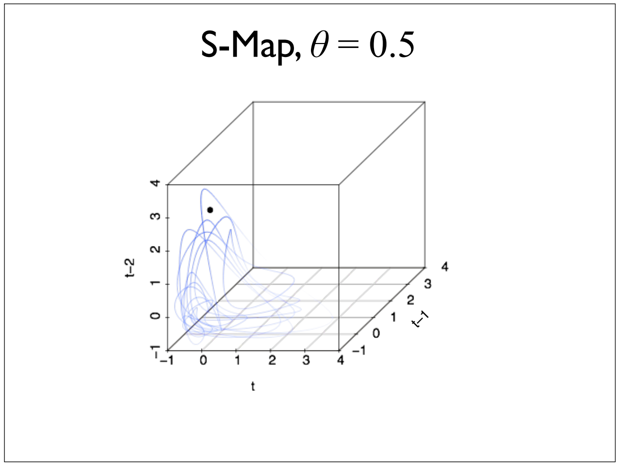
Measures State Dependence

This measures State Dependence (curvature in the manifold)

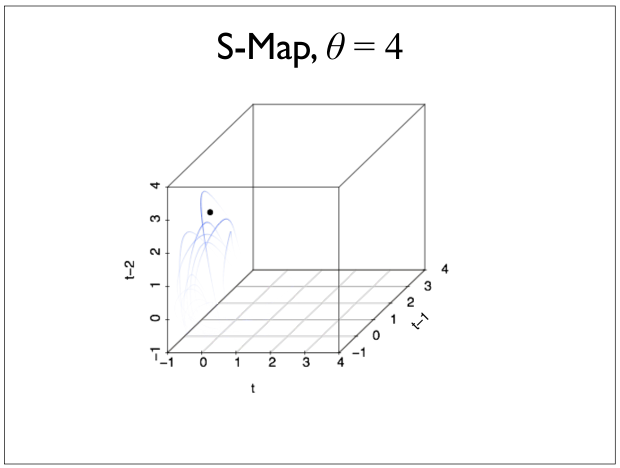
S-Map, $\theta = 0$

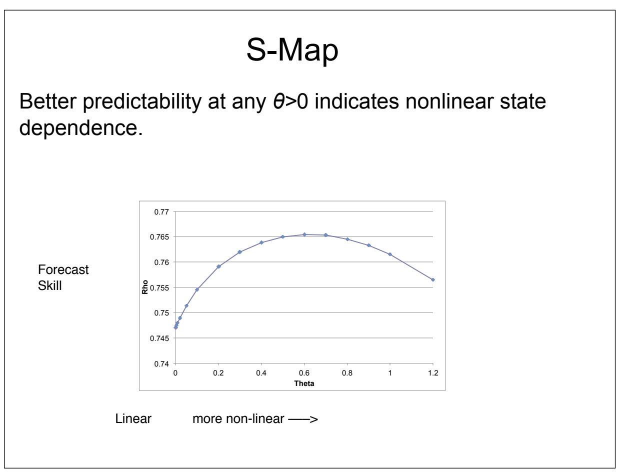


all points are weighed equally… to produce a single global linear map (fitting a single map through all the points)



as theta is tuned upwards the points close to the current state (the predictee) are weighted more heavily when computing the map. Now we no longer have a single map, but a different map at each point.





Again, S-maps are used to identify curvature in the manifold

Curvature is ubiquitous

Nonlinear Attractors Are Ubiquitous in Nature

Hsieh et. al. **Nature** Vol 435 May 2005

Table 2 | Analyses of key North Pacific biological time series

Timescale	Biological data	Best E	Best θ	Best ρ	$\Delta\rho$	Nonlinear?	N	P-value
Weekly	Scripps Pier diatom	3	0.3	0.539	0.139*	Yes	830	<0.01
Monthly	Scripps Pier diatom	4	0.05	0.542	0.083	Yes	206	0.134
Quarterly	CalCOFI coastal larval fish	7	1.6	0.715	0.031*	Yes	3,220	<0.01
Quarterly	CalCOFI coastal-oceanic larval fish	8	0.6	0.744	0.017	Yes	1,400	0.164
Quarterly	CalCOFI oceanic larval fish	8	1.4	0.678	0.020*	Yes	4,760	0.040
Biannual	CalCOFI copepod	6	1.2	0.677	0.027	Yes	1,736	0.078
Annual	CalCOFI copepod	5	0.4	0.566	0.015	Yes	868	0.322
Annual	CalCOFI coastal larval fish	5	0.6	0.603	0.060*	Yes	805	0.038
Annual	CalCOFI coastal-oceanic larval fish	4	0.2	0.502	0.092	Yes	350	0.063
Annual	CalCOFI oceanic larval fish	7	0.6	0.576	0.017	Yes	1,190	0.273
Annual	Chinook salmon	3	0.4	0.448	0.440*	Yes	63	<0.01
Annual	Coho salmon	7	0.3	0.656	0.117	Yes	63	0.213
Annual	Chum salmon	4	0.18	0.634	0.767*	Yes	63	<0.01
Annual	Steelhead trout	3	0.2	0.281	0.272	Yes	63	0.118
Annual	Sockeye salmon	4	0.7	0.484	0.168	Yes	63	0.168
Annual	Composite salmon and trout	4	0.3	0.464	0.078	Yes	315	0.148

- other examples
  - Albacore (Glaser et al. 2011)
  - Bluefin Tuna (Fromentin & Powers 2005)
  - Sheep (Grenfell et al. 1998)
  - Diatoms, Childhood diseases (Sugihara & May 1990)
  - cardiac rhythms, sunspots, gravitational flux, fruit fly behavior, neurobiology, gene expression etc.

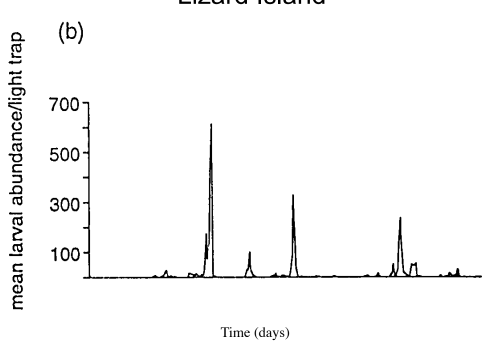
Dynamic state dependence is ubiquitous

This actually has profound implications for how we can study these systems.

An Example of Nonlinearity: Episodic Fluctuations in Larval Supply

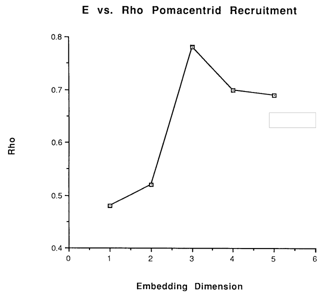
Dixon, Milicich and Sugihara ***Science (1999)***

Example of Nonlinear State Dependence Pomacentrid Larval Supply at Lizard Island



Phototropic Damsel fish larvae caught in light traps on the reef

Embedding Trial with Univariate Simplex Projection



Rho =.78 (Nonlinear Model)

Rho =.29 (Linear (AR3) Model)

What this embedding result tells us

- Pomacentrid larval supply is a low dimensional nonlinear process. (rho=0.78, n=256)
- The optimal embedding for the pomacentrid data is 3 dimensions.
- Therefore it should be possible to construct a model containing 3 variables that is similar in forecast skill as the univariate lag-coordinate model.

"Leverage with multiple timeseries" Look for a *mechanistic* model by searching parallel time series of key environmental variables.

- Construct mechanistic embedding models for prediction by a trial and error search of parallel physical time series.
- Repeat this linearly to construct the best multivariate ARMA model (AR3).

By multivariate simplex projection (nonlinear search) found that the best variables were.

- 1) %night time illumination lagged 19days.
- 2) cross shelf wind lagged 1 day (best not lagged,
  but this represents forward information).
- 3) moderate wind speeds lagged 16-19 days.
- Linear rho= 0.27,   nonlinear rho=0.82

Stars align >>>> Perfect Storm!

Using S-map to track changing interactions in real time

PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

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

Received: 5 November 2015
Accepted: 3 December 2015

Tracking and forecasting ecosystem interactions in real time

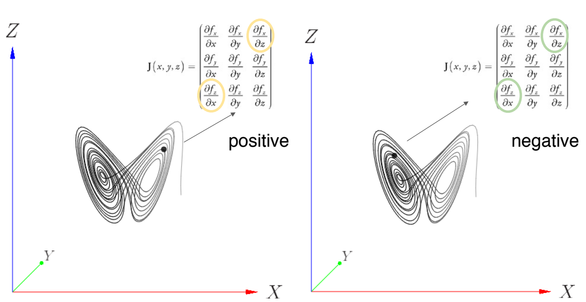
Ethan R. Deyle[1], Robert M. May[2], Stephan B. Munch[3] and George Sugihara[1]

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

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

eg. This paper that appeared last year in PRSB used S-maps to show how species interactions vary in time depending on where on the attractor you are. That is, it showed how to make real time measurements of interactions that are state-dependent.

Tracking Changing Interactions with S-map



The jacobian coefficients (partial derivatives) vary depending on location on the attractor

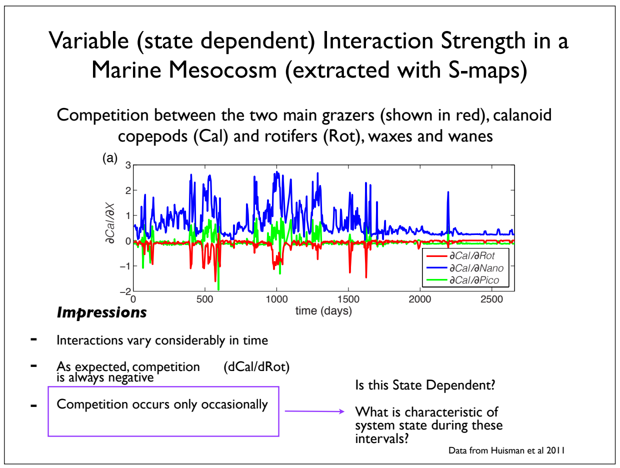
The basic idea is as follows..
The S-map involves calculating a hyperplane or surface at each point as the system travels along its attractor. This involves calculating the jacobian matrix whose elements are the partial derivatives that measure the effect of the system variables on each other.

Note the the embeddings here are multivariate – In native coordinates (not lags).

Again, the coefficients are fit "sequentially" for each location on the manifold using weighted linear regression, with strongest weight given to nearby points, as shown in

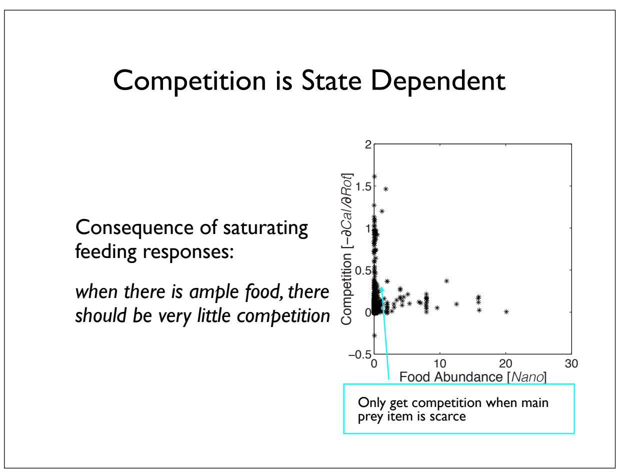
the previous slides.
In a stable equilibrium system these coefficients are fit to a single equilibrium point and are fixed and unchanging. In S-maps, however, the values are state dependent — they vary depending on location on the attractor, x(t*).
Thus, by computing sequential jacobians, the S=map tracks interactions (partial derivatives) that change with the evolving state of the system.

What is really nice about this is that it is easily accomplished with real field data.



Here is an example applied to data from a marine mesocosm. (Huisman)

Note that competition between the two main grazers (shown in red), calanoid copepods and rotifers, waxes and wanes….   competition occurs only occasionally…and very episodically   … why?



Thus, we have now have a practical tool for probing changing interactions.

Causality

Let’s now see how EDM deals with causation

Granger Causality

If the following is true

$$\sigma^2\left\{\left(Y_2\middle|\overline{U}\right)\right\} < \sigma^2\left\{\left(Y_2\middle|\overline{U - Y_1}\right)\right\}$$

Then Y_1 “Granger Causes” Y_2

U is the universe of all causal variables

Here we are trying to predict Y2 from U (left side)

If we now remove Y1, and predictability declines, Y1 was causal

The problem, however, is that for dynamic systems.... cannot remove Y1
(according to Takens information about each variable is encoded in all of the others...

Dynamic Causation

- Time series variables are causally related if they are coupled (pertubing one variable perturbs the other) and belong to the same dynamic system.
- If $X \longrightarrow Y$, then information about *X*, must be encoded in the shadow manifold of *Y*
- This can be tested with cross mapping.

In dynamic systems, time series variables are causally related if they are coupled and belong to the same dynamic system… **read slide**

“Information about the aggressor is found in the victim.” as it were

Detecting Causality in Complex Ecosystems

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

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

26 OCTOBER 2012 VOL 338 **SCIENCE** www.sciencemag.org

Featured Commentary in **Nature Physics** by Mark Buchnan November 2012

The basic idea was described in this article, and is summarized in the following video clip… last one.

Empirical Dynamics:
Convergent Cross Mapping

from

"Detecting Causality in Complex Ecosystems"

Sugihara et al. *Science* (2012)

© July 2014

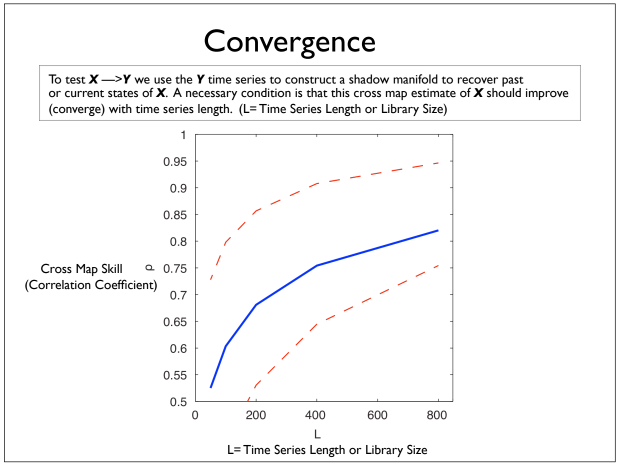
Here is another video clip

Convergent Cross Mapping (CCM)

- If *X* causes (influences) *Y* then, *Y* contains information about *X* that can be used to predict (recover) *X*.
- That is, states of *X* can be recovered from the history of *Y*.

Convergent cross mapping (CCM) involves recovering states of the causal variable from the the affected variable.

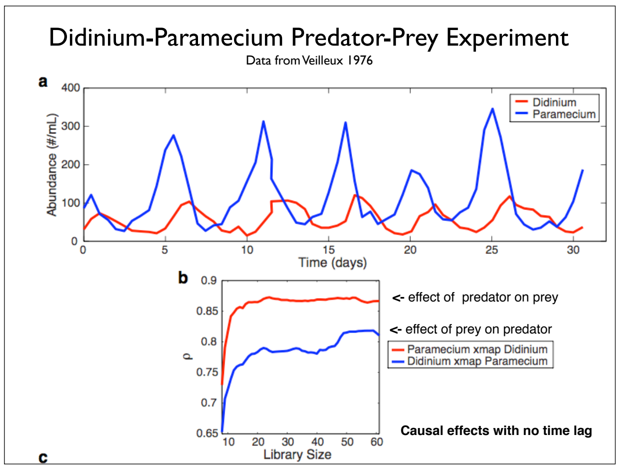
If this is possible, then causal influence is established.



A necessary condition is that the cross map estimate should improve (converge) with time series length

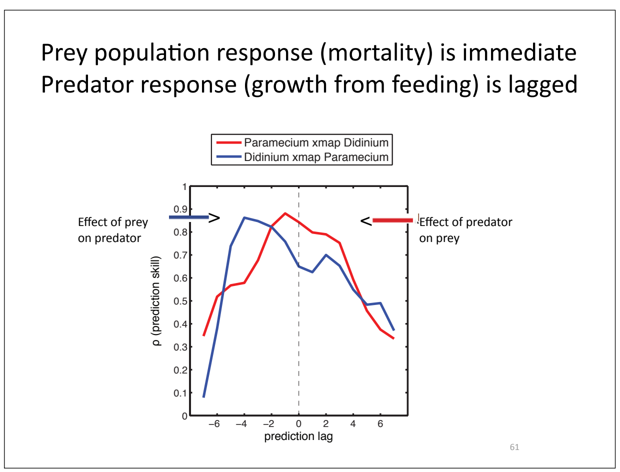
L = library size or time series length

Let's see some examples

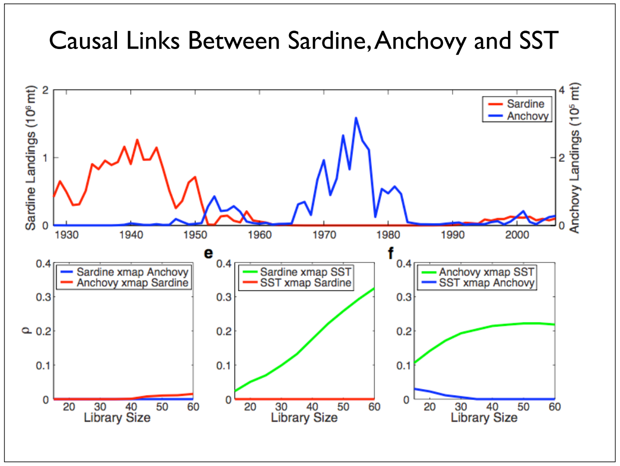


Example 1: This is the classic pred-prey experiment that Gause made famous.
didinium=rotifer predator
paramecium=prey

Cross mapping in both directions indicates bi-directional causation.
red = effect of pred->prey
blue = effect of prey -> pred

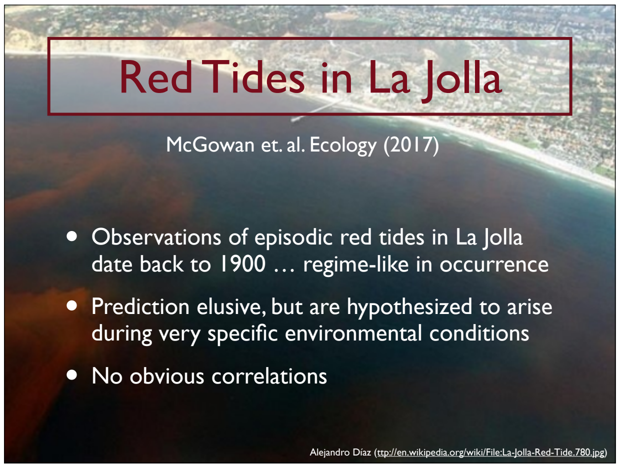


Units are in 1/2 days…. lag 2 = 1 day



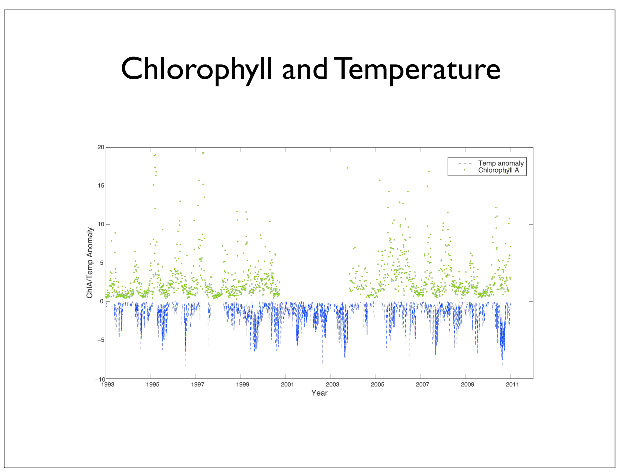
This is a field example:
Sardines and anchovies show reciprocal abundance patterns in the 20th century suggestive of competition.

With cross mapping, however *point* we see that there is no reciprocal information here. Sardines do not affect anchovies and visa versa. However, for both species we find clear evidence of convergence with SST. That is, the time series for both sardines and anchovies contain information about ocean temperatures.

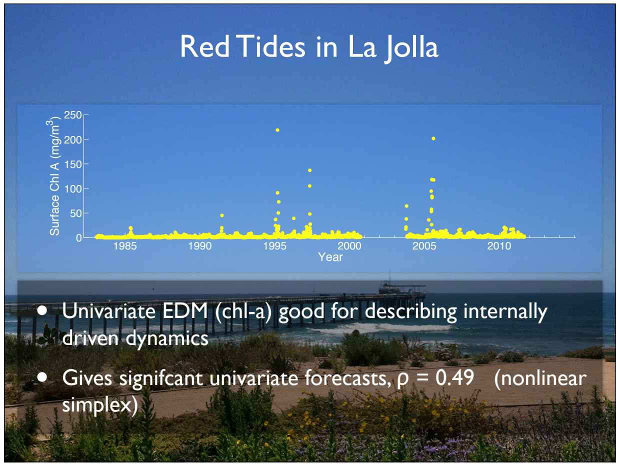


a final ecological example:
Episodic Red Tides around Scripps are a classic example of something that no one has been able to predict. They have been thought to be regime-like, and the mechanism for the rapid transition to this state remained a mystery for over a century.

Despite a half dozen or so Scripps Theses showing in principle (by experiment) that environmental drivers should be important, no obvious field correlations have been found. (between environmental variables and chlorophyll-a).



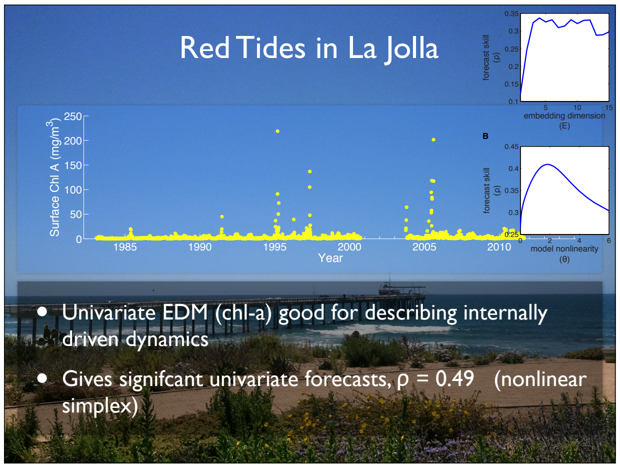
This was exactly the case with the temperature anomaly we saw earlier



The **absence of environmental correlations** suggests that the events cannot be described by linear dynamics,

and this is confirmed by an S-map test for nonlinearity and by the **significant predictability** found with nonlinear forecasting using univariate simplex projection.

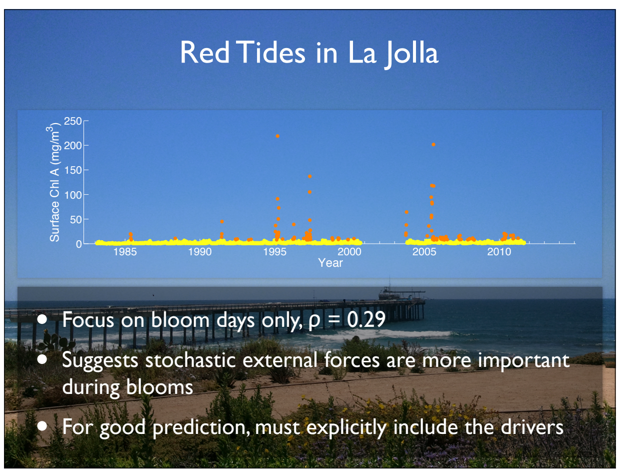
… these predictions are for *events driven by internal deterministic dynamics.*



The **absence of environmental correlations** suggests that the events cannot be described by linear dynamics,

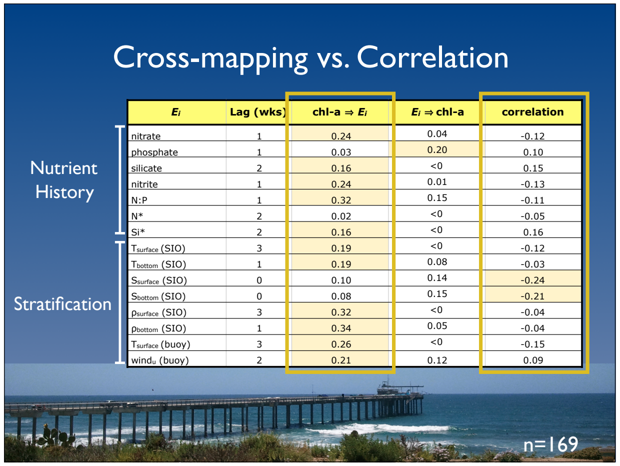
and this is confirmed by an S-map test for nonlinearity and by the **significant predictability** found with nonlinear forecasting using univariate simplex projection.

… these predictions are for *events driven by internal deterministic dynamics.*



However, when we examine just the bloom days (n=169), prediction (univariate simplex) is not nearly as skillful.

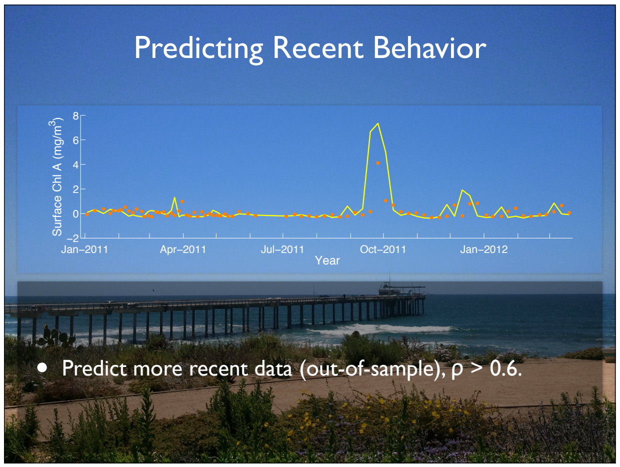
This suggests that internal dynamics alone cannot explain red tides, and that to do so we must explicitly account for stochastic external drivers.



E_i	Lag (wks)	chl-a ⇒ E_i	E_i ⇒ chl-a	correlation
nitrate	1	0.24	0.04	-0.12
phosphate	1	0.03	0.20	0.10
silicate	2	0.16	<0	0.15
nitrite	1	0.24	0.01	-0.13
N:P	1	0.32	0.15	-0.11
N*	2	0.02	<0	-0.05
Si*	2	0.16	<0	0.16
$T_{surface}$ (SIO)	3	0.19	<0	-0.12
T_{bottom} (SIO)	1	0.19	0.08	-0.03
$S_{surface}$ (SIO)	0	0.10	0.14	-0.24
S_{bottom} (SIO)	0	0.08	0.15	-0.21
$\rho_{surface}$ (SIO)	3	0.32	<0	-0.04
ρ_{bottom} (SIO)	1	0.34	0.05	-0.04
$T_{surface}$ (buoy)	3	0.26	<0	-0.15
$wind_u$ (buoy)	2	0.21	0.12	0.09

The candidate variables fall into two loose categories:
1. variables that summarize nutrient history (CLICK)
2. and variables related to stratification and mixing (CLICK).

Again, (CLICK) if you look with cross correlation, there is very little suggestion of environmental forcing.
However, when you look with CCM (CLICK), you can see most of the suspected candidate variables do in fact show causal influence in the time series data from field observations.
Therefore, including these variables as coordinate axes in multivariate EDM



Story of class…
This is TRUE out of sample forecasting.
Indeed, leave-one-out cross-validation over the entire 30-year (1600 point) time series gives very few false forecasts that a bloom will occur (e.g., for some model ensembles as few as 34 false positives and 19 false negatives).

Thus, we have begun to build a valid understanding of the causal mechanisms and more importantly we can forecast red tides with some accuracy.

Some example studies

PNAS

Dynamical evidence for causality between galactic cosmic rays and interannual variation in global temperature

Anastasios A. Tsonis[a,1], Ethan R. Deyle[b], Robert M. May[c,1], George Sugihara[b,1], Kyle Swanson[a], Joshua D. Verbeten[a], and Geli Wang[d]

[a]Department of Mathematical Sciences, University of Wisconsin–Milwaukee, Milwaukee, WI 53201; [b]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093; [c]Department of Zoology, University of Oxford, Oxford OX1-3PS, United Kingdom; and [d]Key Laboratory of Middle Atmosphere and Global Environment Observations, Institute for Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China

- Experimental studies suggest that cosmic rays could affect global temperature (via cloud formation).
- CCM can distinguish between short-term dynamics (i.e., cloud formation) and long-term dynamics (i.e., climate change) by examining first-difference temperature vs. raw temperature
- Cosmic rays influence only year-to-year variations in temperature

* The increase in CR incidence in the 20th century has been used to suggest that the observed climate warming is natural and not due to man.

* This study used CCM to examine this potential effect. It found no evidence for CR causing the 20th century warming trend. But it did find an effect on interannual time scales… resonates with experiments the show how CR could affect cloud formation.

---

nature climate change | LETTERS
PUBLISHED ONLINE: 30 MARCH 2015 | DOI: 10.1038/NCLIMATE2568

Causal feedbacks in climate change

Egbert H. van Nes[1]*, Marten Scheffer[1], Victor Brovkin[2], Timothy M. Lenton[3], Hao Ye[4], Ethan Deyle[4] and George Sugihara[4]*

- Confirms by direct observation the well-established mechanism that greenhouse gases (CO_2 and CH_4) affect temperature. An immediate effect.
- Confirms the controversial link of temperature affecting greenhouse gases, producing positive feedbacks. A delayed effect.

This study involved the analysis of the Vostock ice core time series data to see if there is direct observational evidence for causal effects

PNAS

Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling

Hao Ye[a,1], Richard J. Beamish[b], Sarah M. Glaser[c], Sue C. H. Grant[d], Chih-hao Hsieh[e], Laura J. Richards[b], Jon T. Schnute[b], and George Sugihara[a,1]

[a]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093; [b]Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, BC V9T 6N7, Canada; [c]Joseph S. Korbel School of International Studies, University of Denver, Denver, CO 80210; [d]Fisheries and Oceans Canada, Delta, BC V3M 6A2, Canada; and [e]Institute of Oceanography and Institute of Ecology and Evolutionary Biology, National Taiwan University, Taipei, Taiwan 10617

Edited by Stephen R. Carpenter, University of Wisconsin–Madison, Madison, WI, and approved January 28, 2015 (received for review September 17, 2014)

- Forecasts of fisheries recruitment have been unreliable (weak stock-recruit relationship based on classic models assuming equilibrium dynamics)
- Environment is a likely factor, but does not improve forecast performance of classical models (in official DFO forecasts)
- Find that EDM models using environmental variables provide accurate forecasts with historical cross validation over 57 yrs… accurate 2014 (15 and16) forecast!

This one focused on forecasting. It was aimed at providing better production forecasts for Canada's iconic sockeye salmon industry.

Science AAAS

Home | News | Journals | Topics | Careers

Latest News | ScienceInsider | ScienceShots | Sifter | From the Magazine | About News

SHARE

Humidity and temperature both play a role in flu outbreaks.

Dirima/iStockphoto

'Messy math' from sardine studies could help fight flu outbreaks

Deyle et al. PNAS 2016

And this is an application of these methods to understand environmental drivers of flu epidemics. What is interesting here is that we were able to identify AH as causal and find a specific temperature threshold 75F below which higher AH reduces flu transmission and above which it increases flu transmission.

There are many other factors of course, but AH is certainly one of them.

2016 winner of William James Prize

Untangling Brain-Wide Dynamics in Consciousness by Cross-Embedding

Satohiro Tajima, Toru Yanagawa, Naotaka Fujii, Taro Toyoizumi

Published: November 19, 2015 • http://dx.doi.org/10.1371/journal.pcbi.1004537

- "The model-free method reveals a consciousness-related hierarchy of cortical areas, where dynamical complexity increases along with cross-area information flow.
- "This approach reveals a universality of inter-areal interactions and complexity in conscious brain dynamics, demonstrating its wide application to deciphering complex neuronal systems."

Showed how the approach can be developed to index brain states.

Closing Remark

“There is a fundamental disconnect between the biological interactions that we observe and the common (linear/ reductionist) assumptions of the framework that we use to study them.”

Now this is getting “preachy”

Static Theoretical Ideal vs. Dynamic Reality

Flip book analogy…

Summary Statement:

Static Theoretical Ideal vs. Dynamic Reality

- **Static Theoretical Ideal** (classical linear framework)
  - equilibrium
  - stable
  - separable (decomposable, study piecewise)
  - Granger
  - classic parametric models
- **Dynamic Reality** (nonlinear empirical dynamics)
  - non-equilibrium
  - non-stable
  - non-separable (interdependent, study as a whole)
  - CCM
  - empirical dynamic models

The basic dichotomy here is a contrast between what was thought to be a necessary expedient (a theoretical compromise based on a static equilibrium system of independent parts), and the reality of nonlinear interdependent ever-changing natural systems.

The explosion of data is enabling investigation at the whole systems level.

What I tried to suggest today is that it is possible and worthwhile to develop approaches where this expedient is NOT necessary.

Wordy Manifesto

Despite the known reality and ubiquity of nonlinear dynamics, and the costs associated with unanticipated threshold phenomena or tipping points, nearly all attempts to understand them in applied contexts (outside of formal studies of turbulence) have used incorrect linear statistical tools (static analytical tools based on a classical linear paradigm). This paradigm based on stable, stationary equilibrium points or cyclic equilibrium dynamics allows systems to be studied piecewise as a decomposable sum of independent parts; a tractable approach that applies robustly in designed engineering contexts. As a consequence, an extensive methodological tool chest has evolved for analyzing linear (separable) systems. Indeed the ubiquity of available tools seems to be the main reason why these methods and concepts continue to be used in non-engineering contexts, despite the obvious problem that they do not match our current views of how most real (non-engineered) systems are structured (interdependently) and actually behave (i.e, exhibiting non-stationary, non-equilibrium and non-separable state dependence).

click

PNAS

COMMENTARY

Equation-free modeling unravels the behavior of complex ecological systems

Donald L. DeAngelis[a,b,1] and Simeon Yurek[b]
[a]Southeast Ecological Science Center, US Geological Survey, Gainesville, FL 32653; and [b]Biology Department, University of Miami, Miami, FL 33124

The broader message of the Ye et al. report is that science may be moving into a period where equations do not play the central role in describing dynamic systems that they have played in the last 300 years.

79

This Commentary appeared in PNAS a year ago and was a nice confirmation of the idea.

-data science, makes this all possible…data driven discovery