Empirical dynamic modelling: who needs equations?

Andrew Edwards

Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, Canada.

Department of Biology, University of Victoria, Victoria, Canada.

Government of Canada Gouvernement du Canada

Canada

TESA ecosystem approach workshop, Nanaimo.
Tuesday 22nd November 2016

Background

“Incorporating an ecosystem approach into single-species stock assessments”

This is a different approach, more like:

“An approach for analysing ecosystem influences on single (or multiple) species.”

Outline

1. Motivation – three example papers.

2. Results from Sockeye Salmon application.

3. Some background to the methods.

4. Results from simulated data.

5. Results from time series of sardine, anchovy and temperature.

6. Suggestions for break-out group.

Motivation

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]

PNAS | Published online March 2, 2015 | E1569–E1576

Data: spawners, recruits, environmental variables.

Compare Ricker-type models with EDM.

- "EDM models produce more accurate and precise forecasts".

- EDM yields "significant improvements when environmental factors are included", unlike Ricker-type models.

Motivation

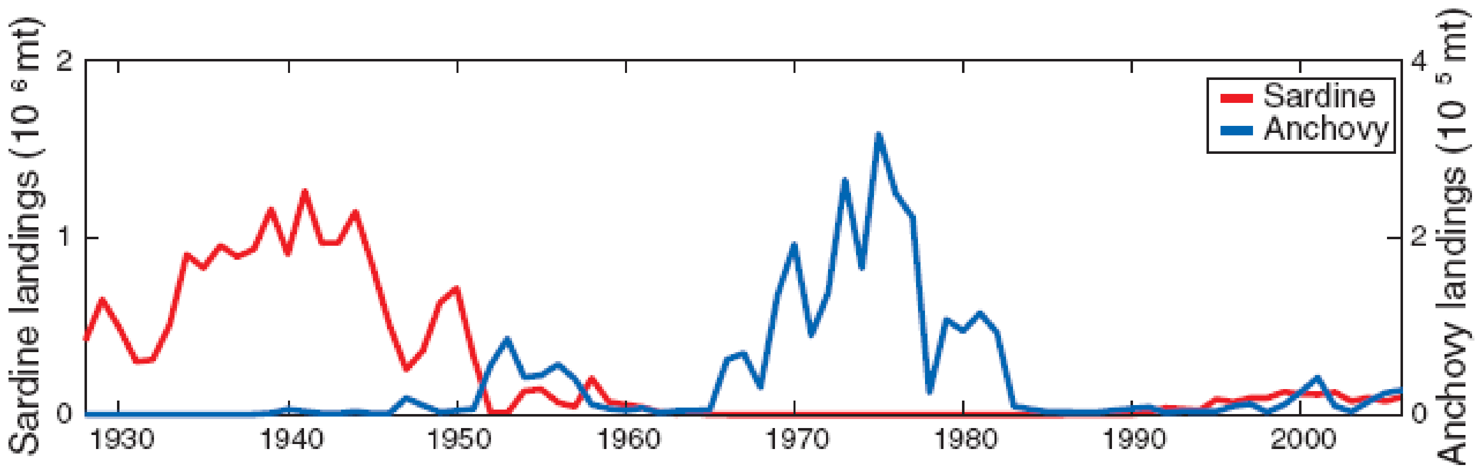
Detecting Causality in Complex Ecosystems

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

Science

AAAS

26 OCTOBER 2012 VOL 338 SCIENCE



Data from California: sardine landings, anchovy landings, SST.

EDM: not direct competition, but both influenced by SST.

Motivation

FISH and FISHERIES

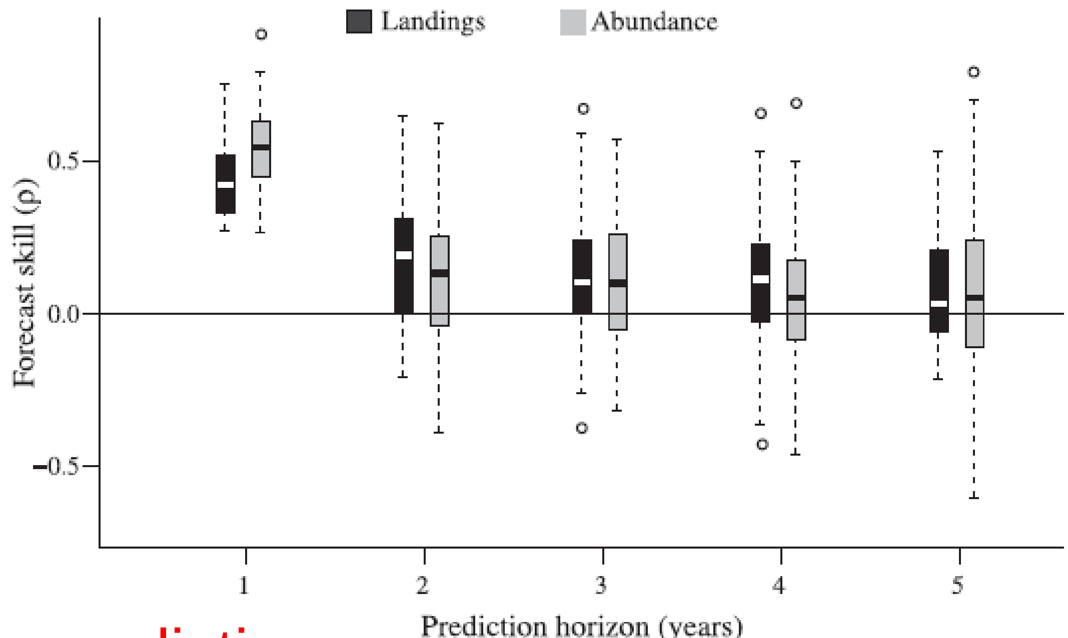
FISH and FISHERIES, 2014, **15**, 616–633

Complex dynamics may limit prediction in marine fisheries

Sarah M Glaser[1,2], Michael J Fogarty[3], Hui Liu[4], Irit Altman[5], Chih-Hao Hsieh[6], Les Kaufman[5], Alec D MacCall[7], Andrew A Rosenberg[8], Hao Ye[9] & George Sugihara[9]

Data: 206 time series of US survey indices and landings.

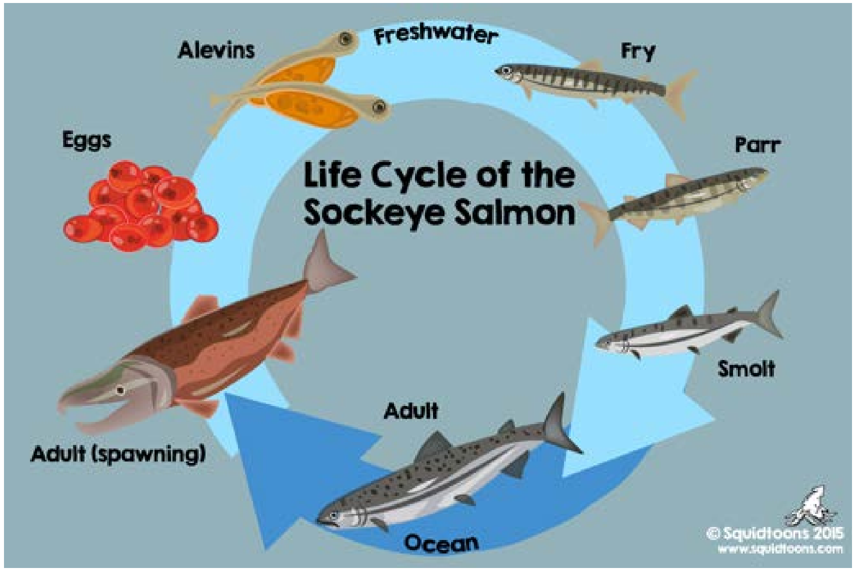
EDM: Predictability declines exponentially over 5 years.

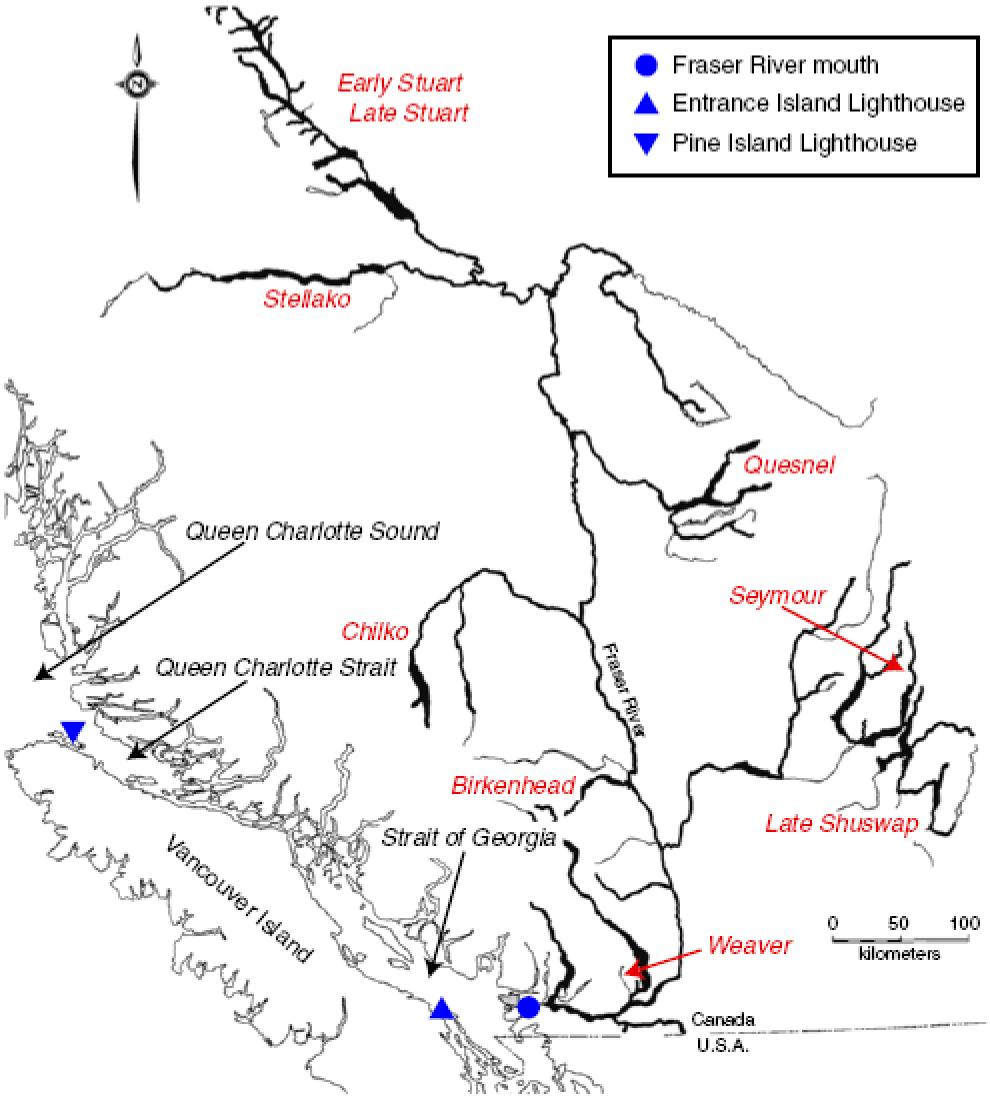


Should only make 1-2 year predictions.

Sockeye Salmon (Ye et al. 2015)

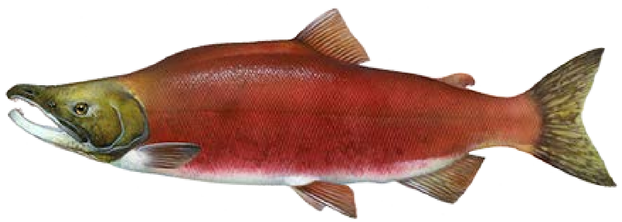
Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling

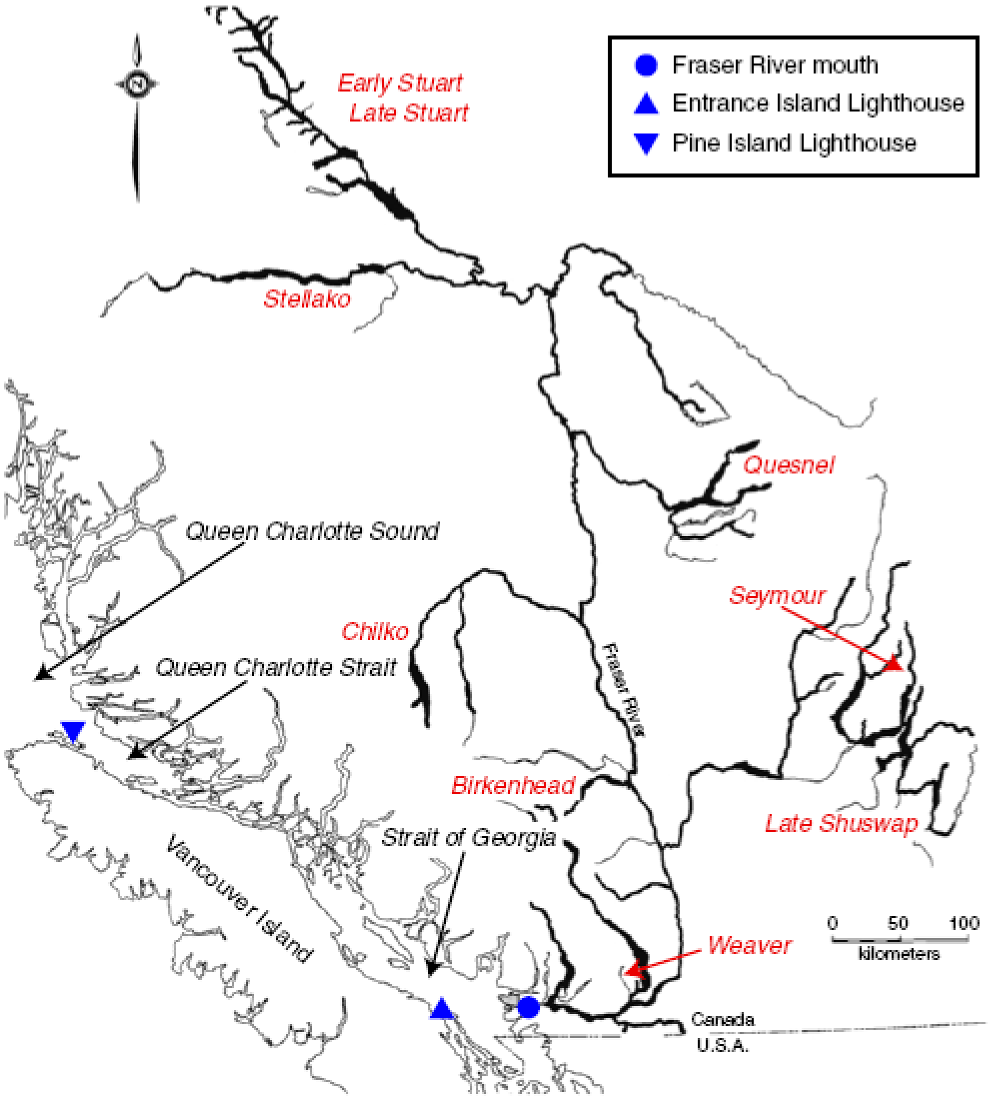




Sockeye Salmon (Ye et al. 2015)

Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling



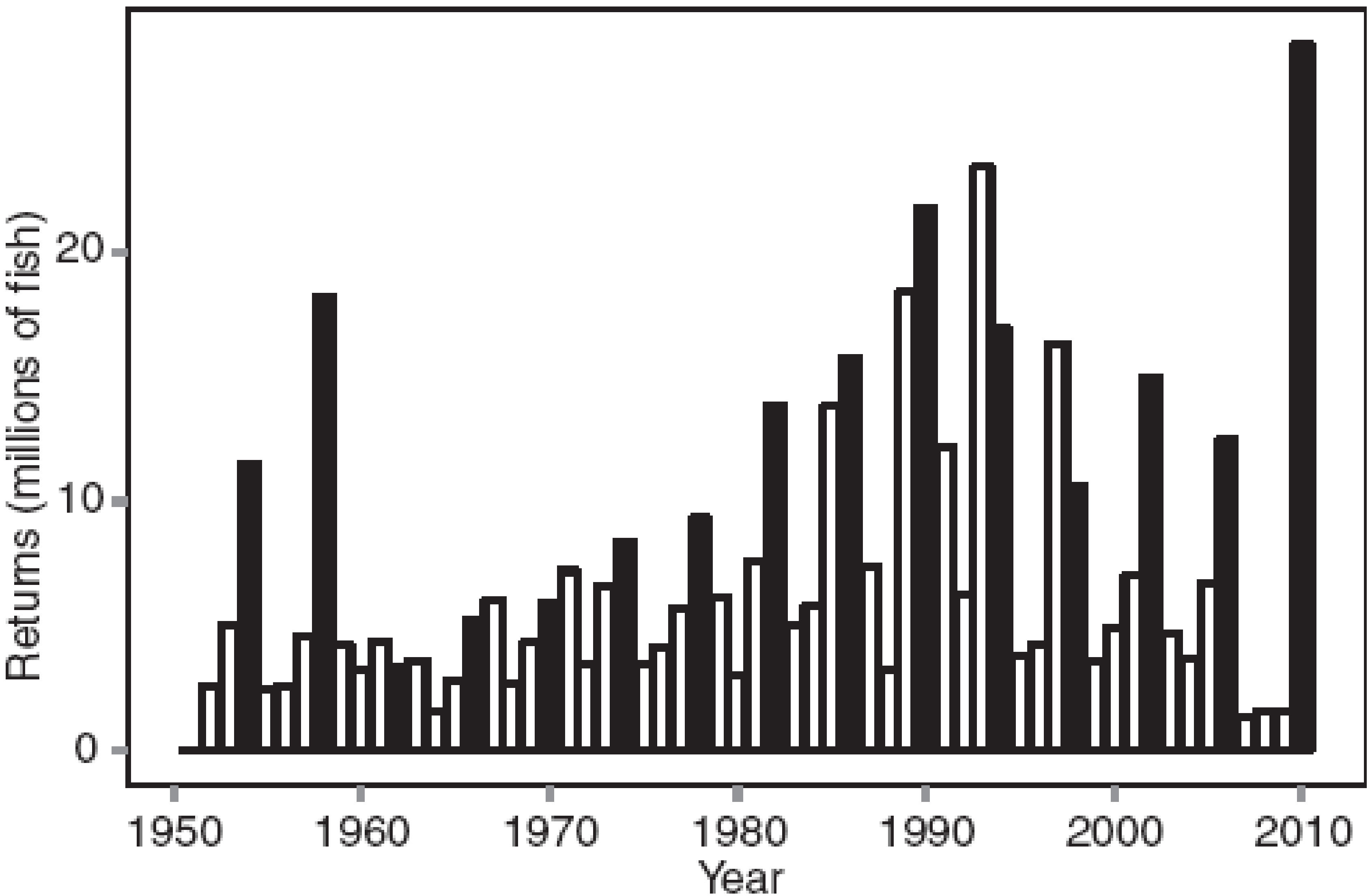


Data:

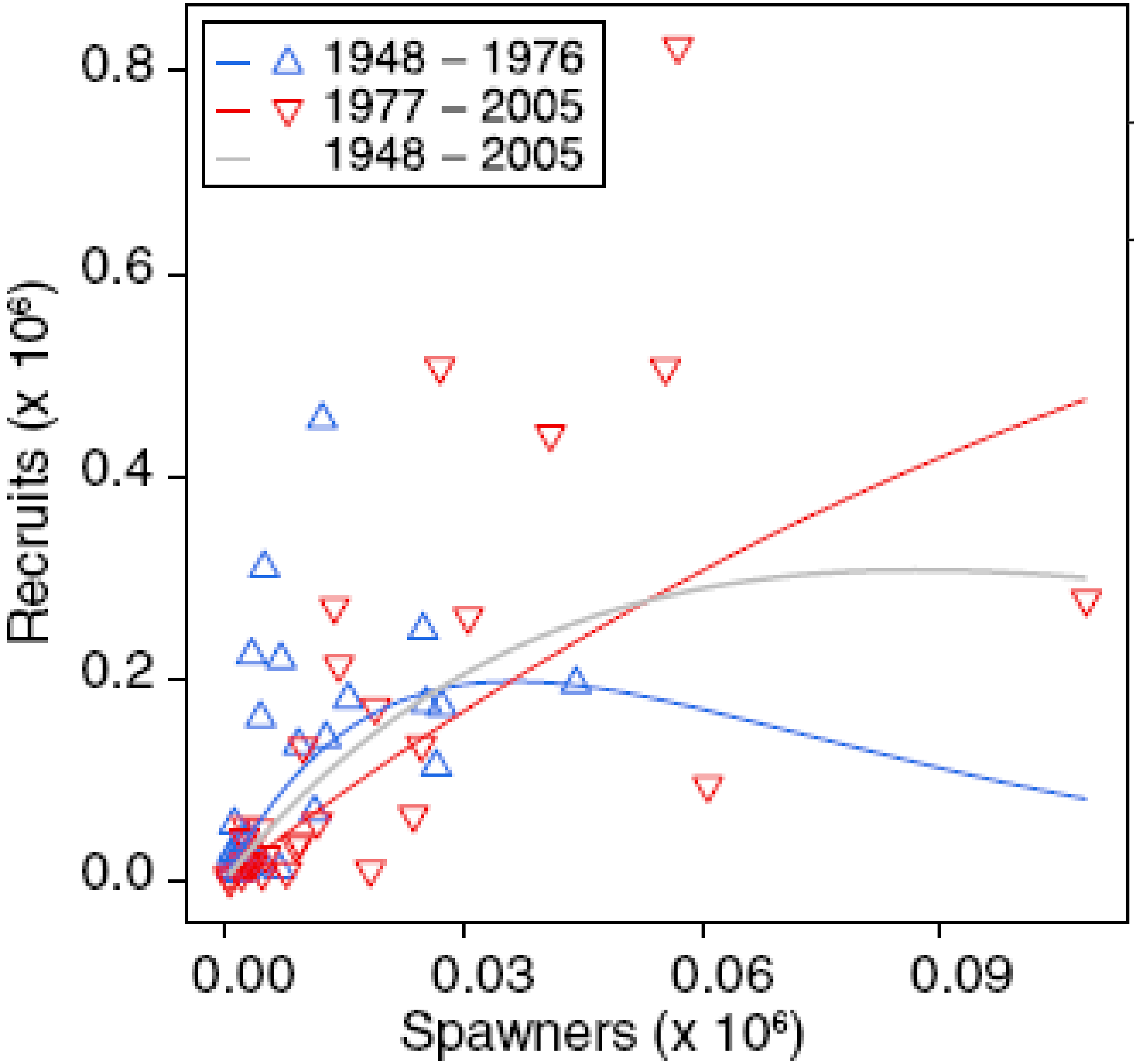
- spawners
- recruits
- SST
- Pacific Decadal Oscillation
- Fraser River discharge

Spanning 1948-2010 (or less).

Variable total returns



Seymour stock



$$\hat{R}_t = S_t \mathrm{e}^{\alpha - \beta S_t}$$

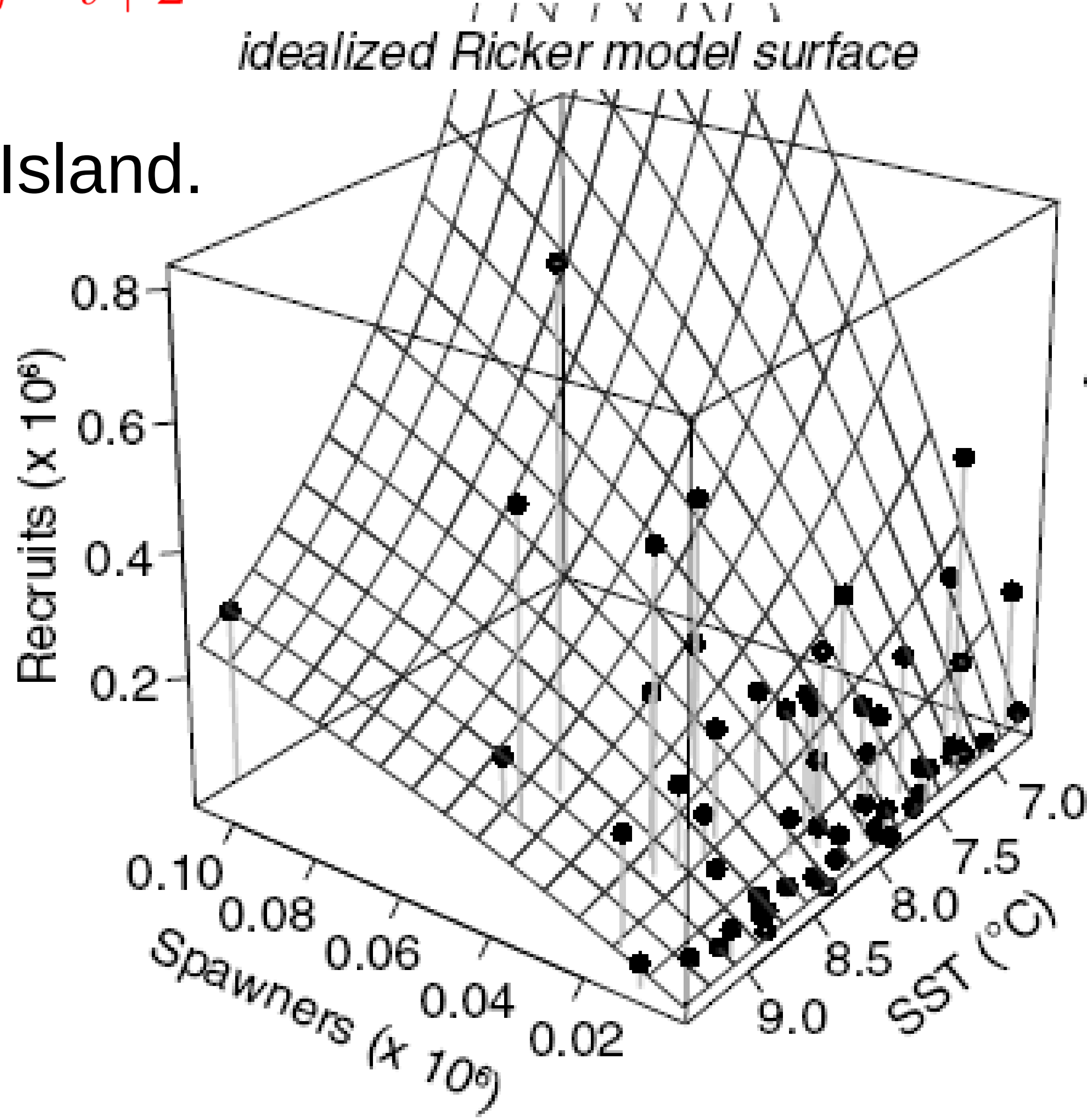
R_t is number of returning adults 4 or 5 years later.

Incorporate temperature

$$\hat{R}_t = S_t e^{\alpha - \beta S_t + \gamma U_{t+2}}$$

U_{t+2} is SST at Pine Island.

(Don't worry about *t*).

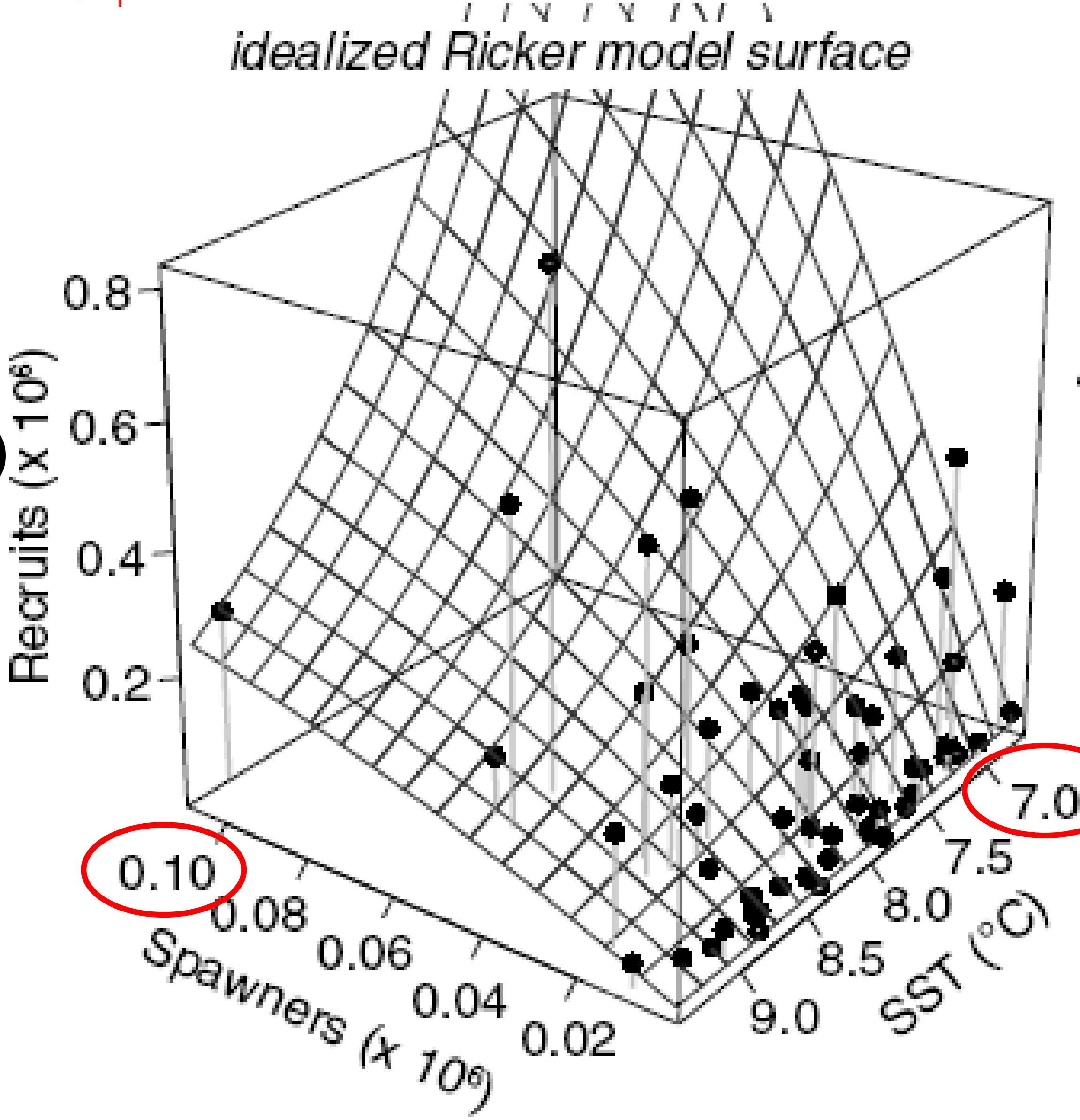


Incorporate temperature

$$\hat{R}_t = S_t e^{\alpha - \beta S_t + \gamma U_{t+2}}$$

Low (but attainable) SST, with high (but attainable) spawners, predicts ridiculously high (historically unattainable) recruits.

Prescribing above equation gives only one hypothesis for how temperature can affect recruitment.



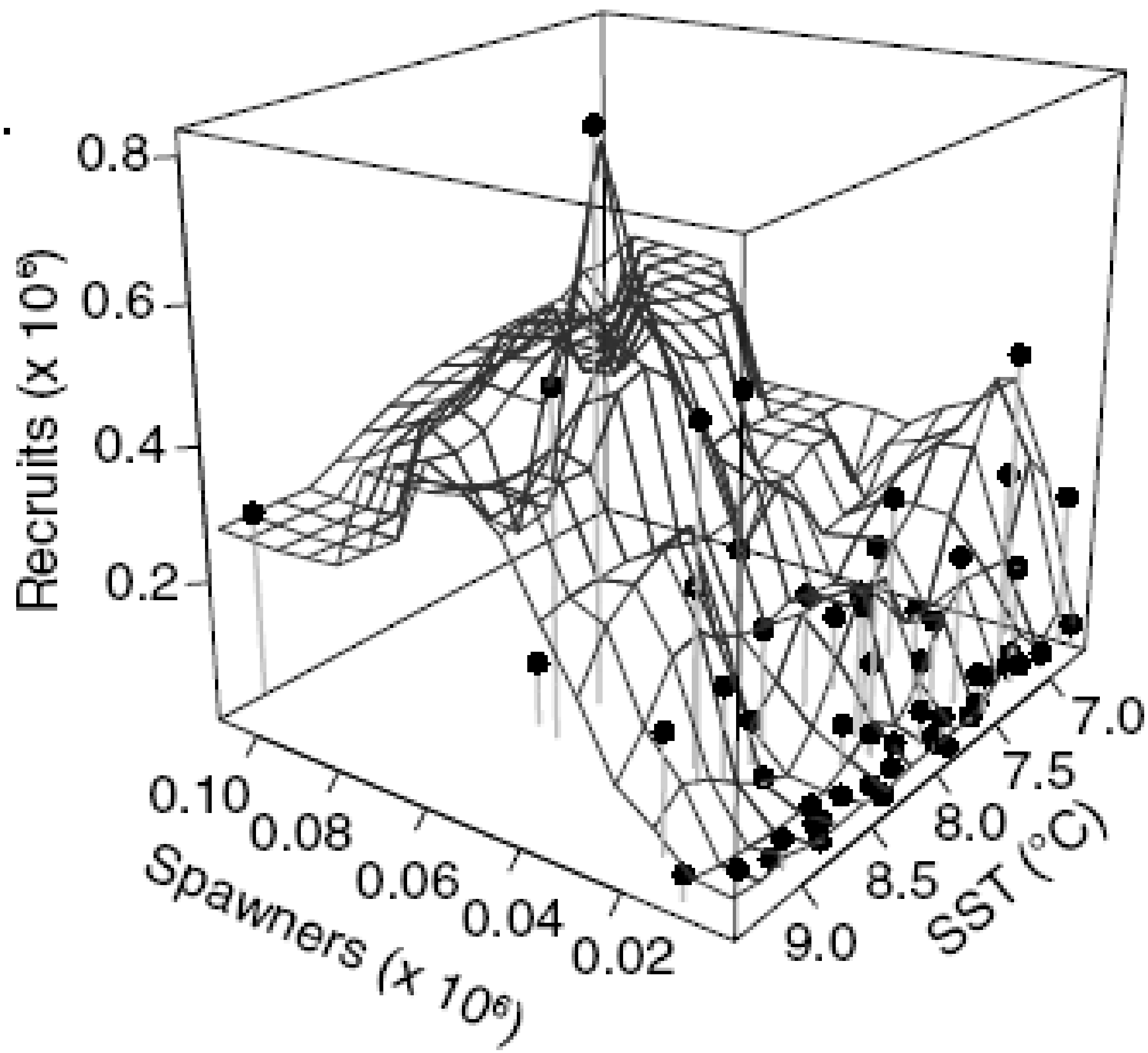
EDM approach

No equation specified.

No hypothesised relationship.

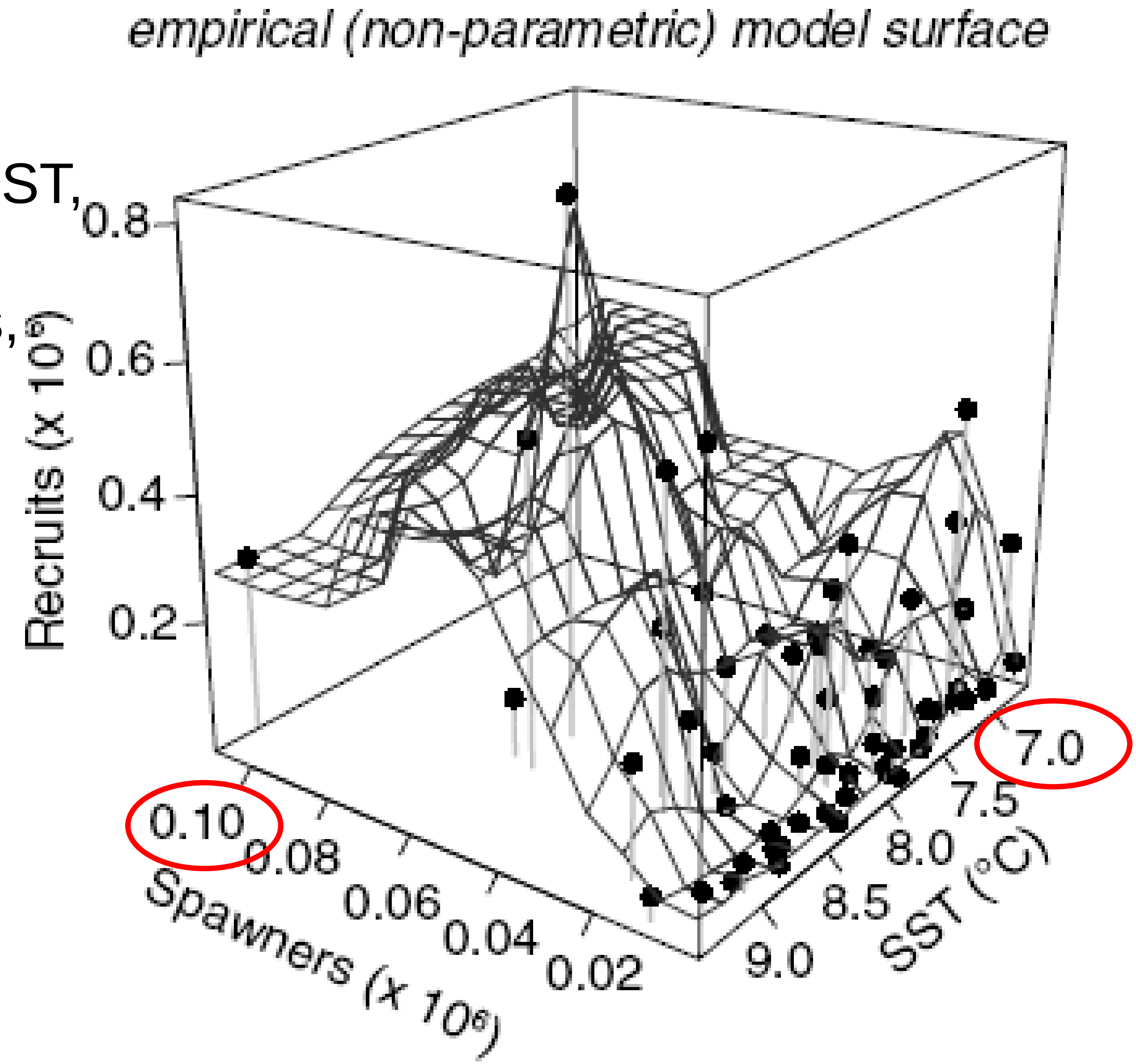
Just uses data.

empirical (non-parametric) model surface

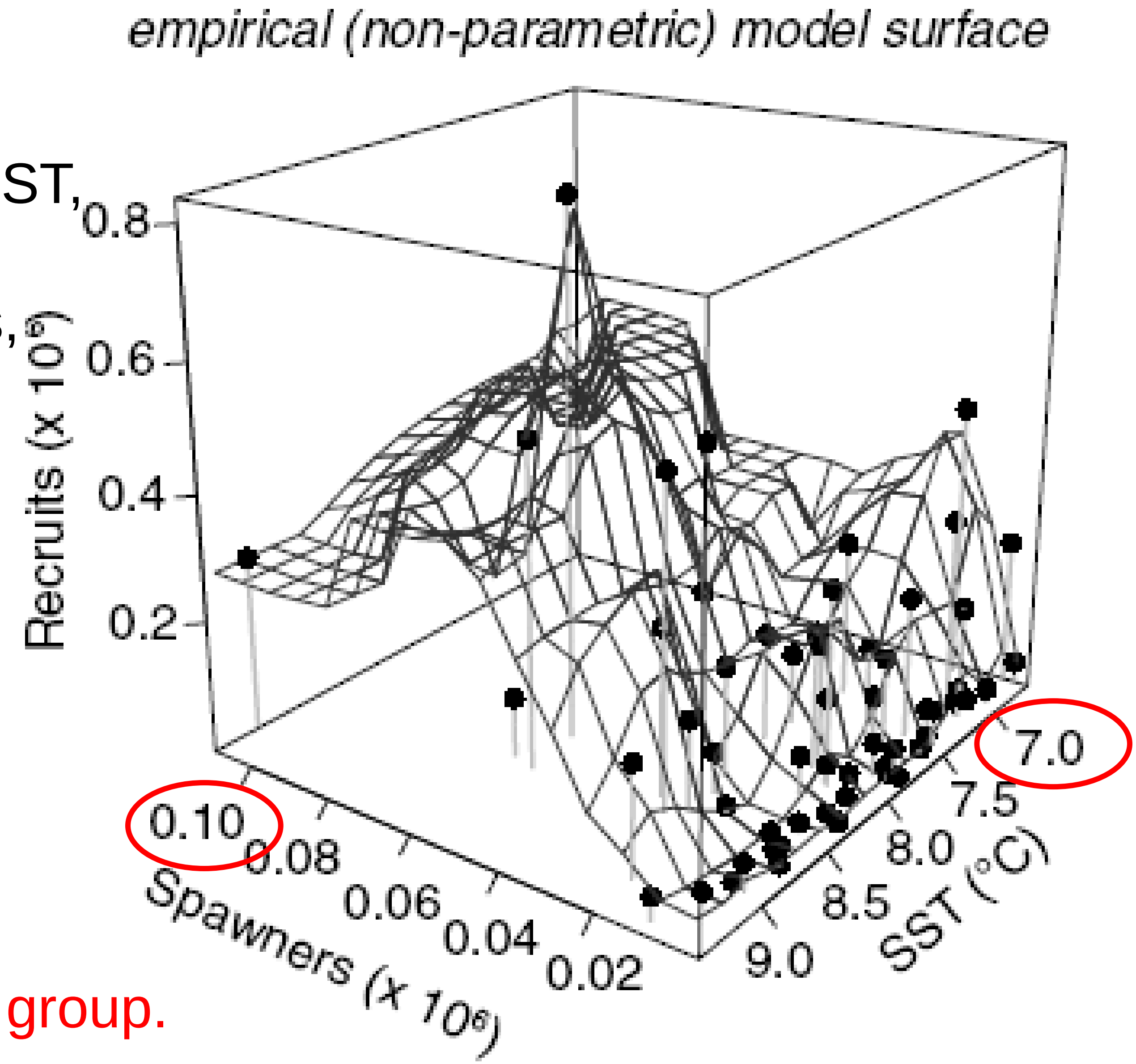


EDM approach

Low (but attainable) SST, with high (but attainable) spawners, now predicts reasonable (historically attained) recruits.



EDM approach



Low (but attainable) SST, with high (but attainable) spawners, now predicts reasonable (historically attained) recruits.

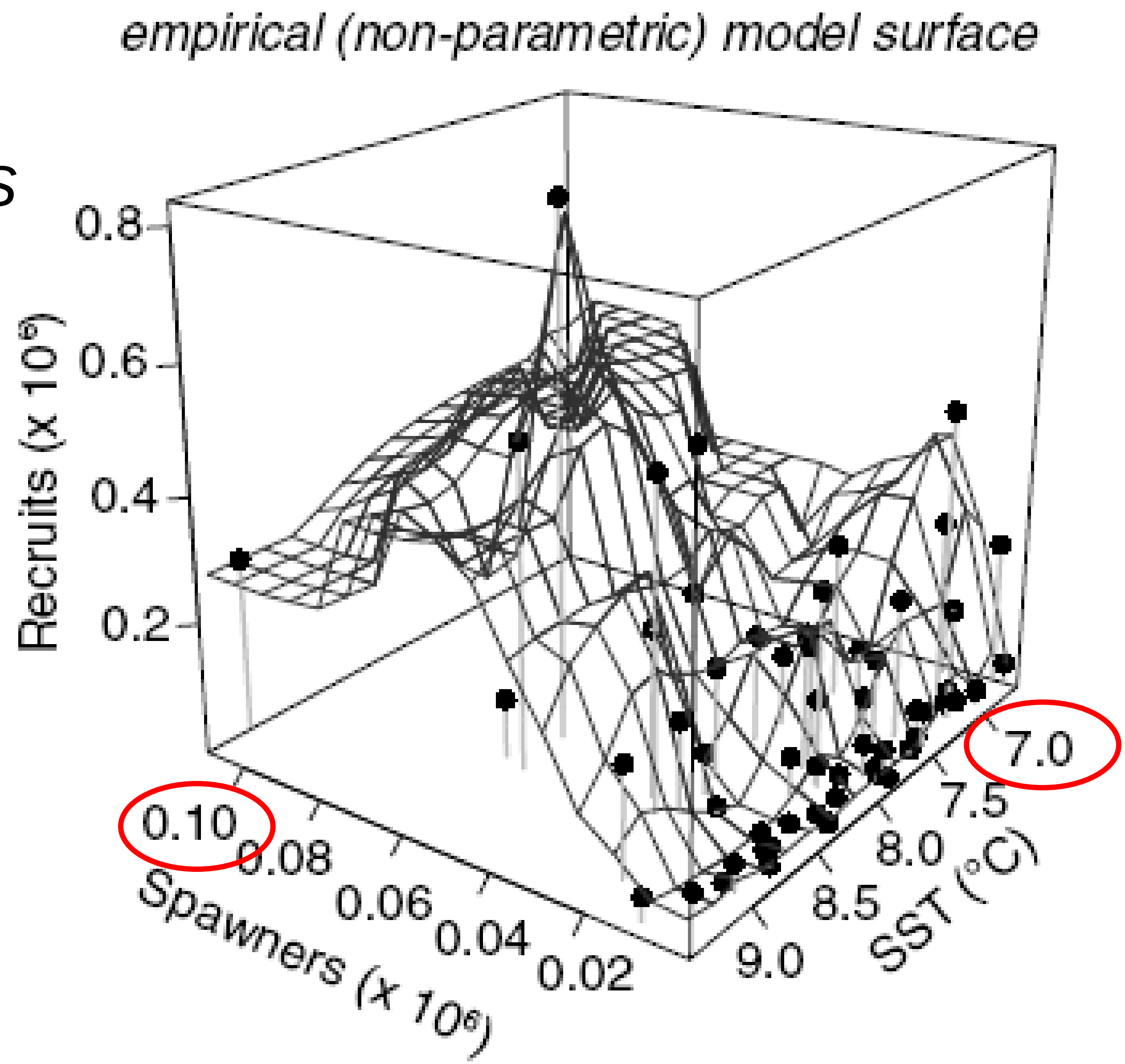
How?
Why?
Usable?

= one goal of working group.

EDM approach

Dependence of *R* on *S* changes with SST, but not in a simple way (as assumed with Ricker model).

Can reproduce (and rotate) figure from R code.



Forecast accuracy

Consider each stock separately.

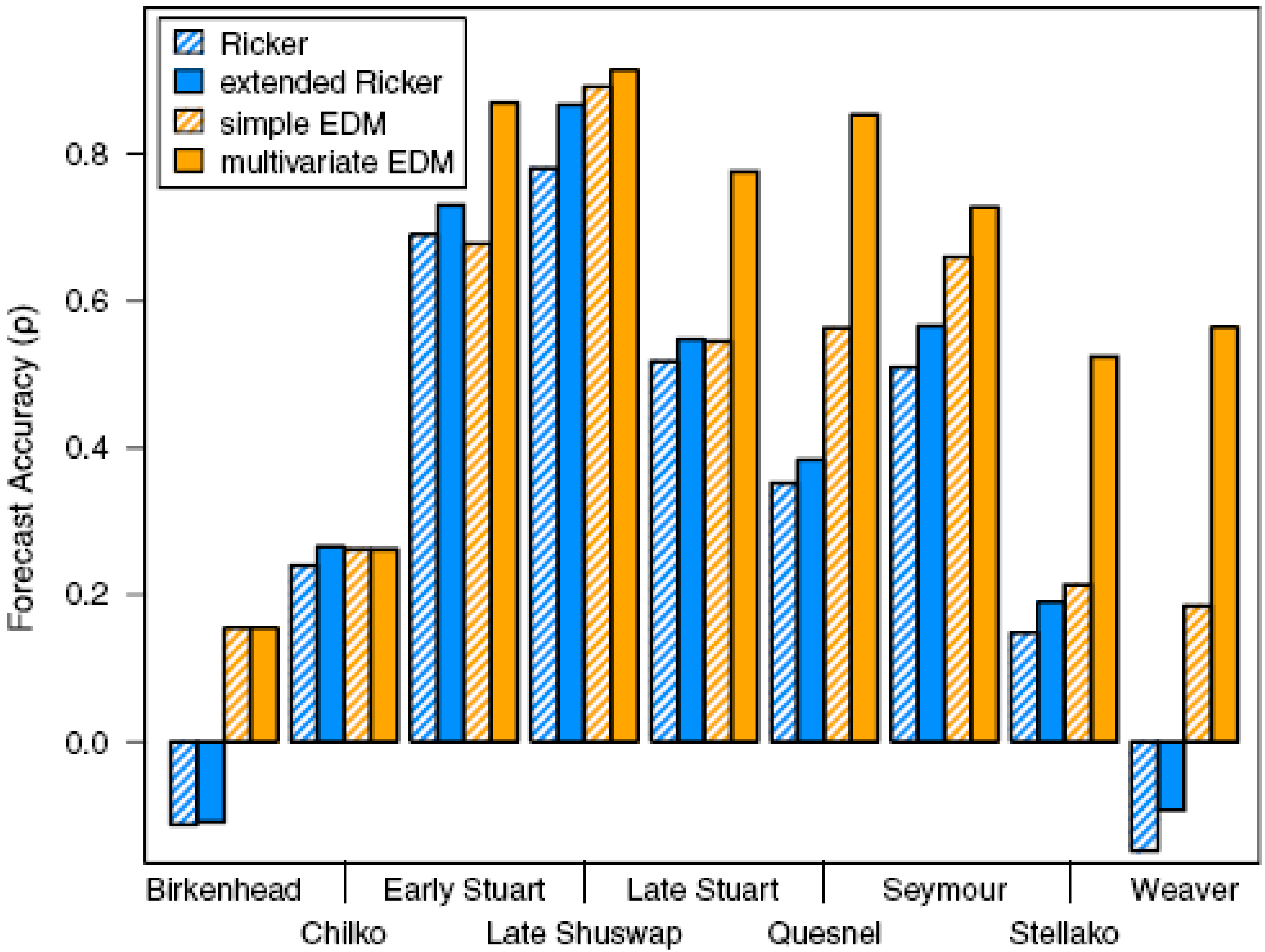
Four models: - Ricker
- extended Ricker (with environmental)
- simple EDM
- multivariate EDM (with environmental)

Accuracy: ρ = correlation between observed and predicted values.

Fourfold cross-validation:
- leave out ¼ of data
- fit to remaining ¾
- how well does the fit predict the ¼ of data.

Sequentially add environmental variable that most increases ρ.

Forecast accuracy



Multiple correct models

Some stocks, different EDM models (different variables) show similar forecast skill.

e.g. Seymour stock:

- spawners, July SST: ρ = 0.734, MAE = 0.065

- spawners, winter PDO: ρ = 0.690, MAE = 0.063.

MAE: mean absolute error.

Summary

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]

PNAS

Data: spawners, recruits, environmental variables.

Compare Ricker-type models with EDM.

- "EDM models produce more accurate and precise forecasts".

- EDM yields "significant improvements when environmental factors are included", unlike Ricker-type models.

Empirical dynamic modelling

Goal: understand system behaviour directly from the data.

No hypothesised equations.

Relationships between variables are determined empirically
from the data.

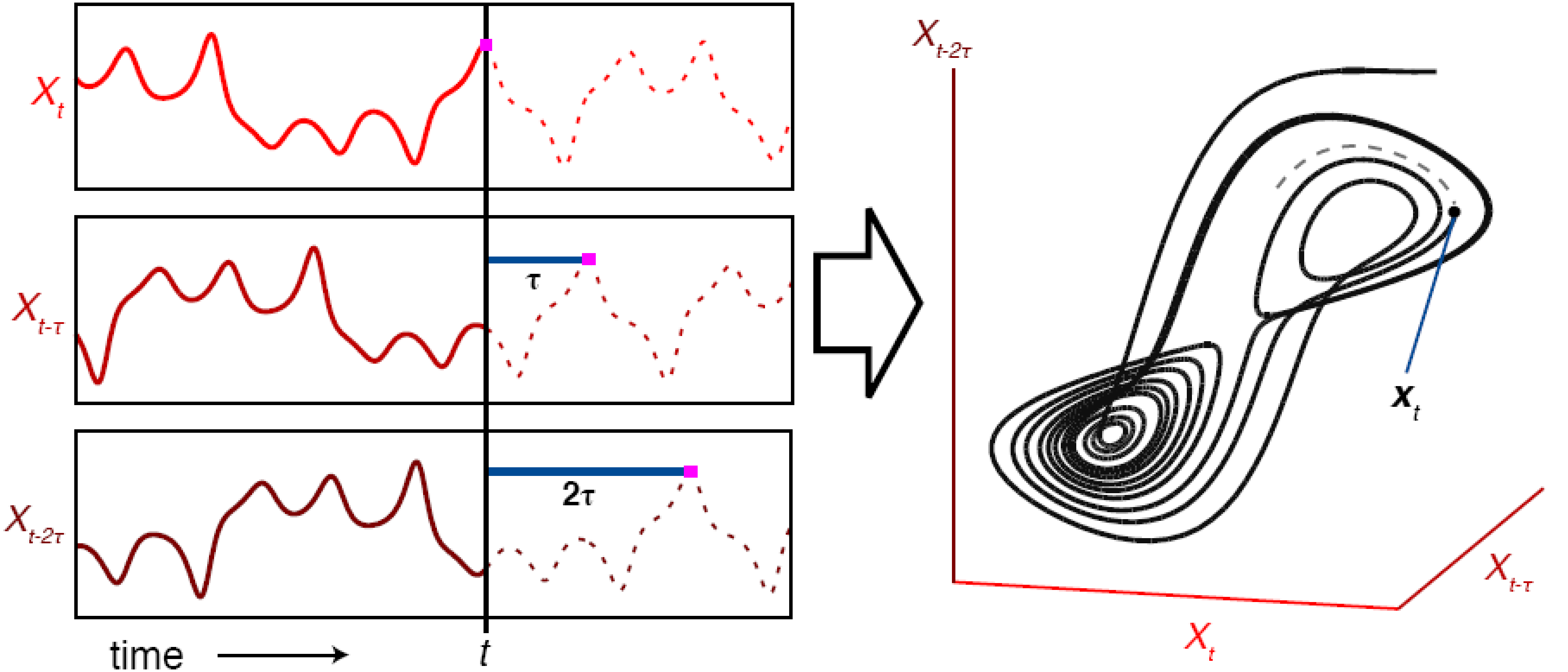
Main assumption: current state of system is not completely
random but depends on (recent) past.

Deterministic rules for how the system changes over time
- rules may be unknown
- can depend on unobserved and/or random variables
(allows stochastic drivers; not completely deterministic).

../edmPapers/ye15movie.mov

Empirical dynamic modelling

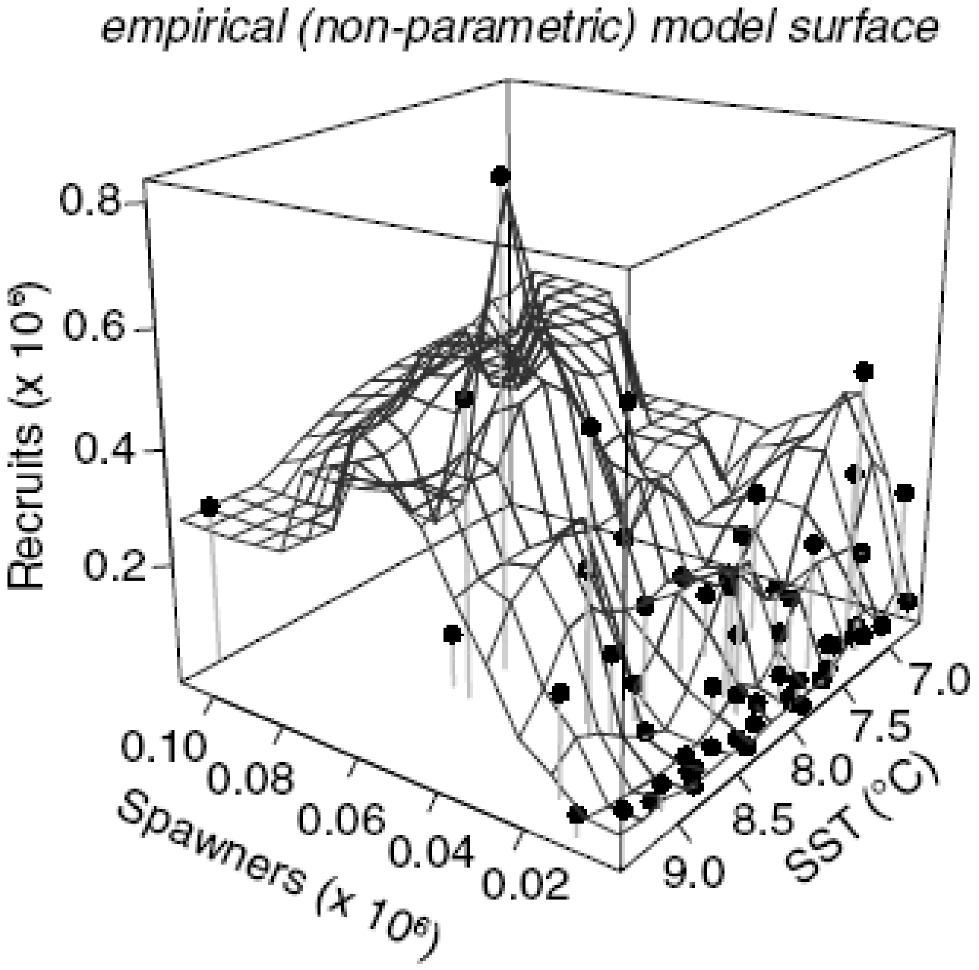
IF we know *all* relevant variables, can use their time series to reconstruct the original manifold ('surface').



Empirical dynamic modelling

But in Sockeye Salmon we do not know *all* relevant variables.

We never do in practice.



Empirical dynamic modelling

In nature, system may be highly complex (lots of components).

Time series are generally short.

Reconstruct system using successive lags of single time series.

Takens' theorem: if *enough* lags are taken, reconstruction preserves essential mathematical properties of original system.

Or: lags of a system can substitute for unknown or unobserved variables.

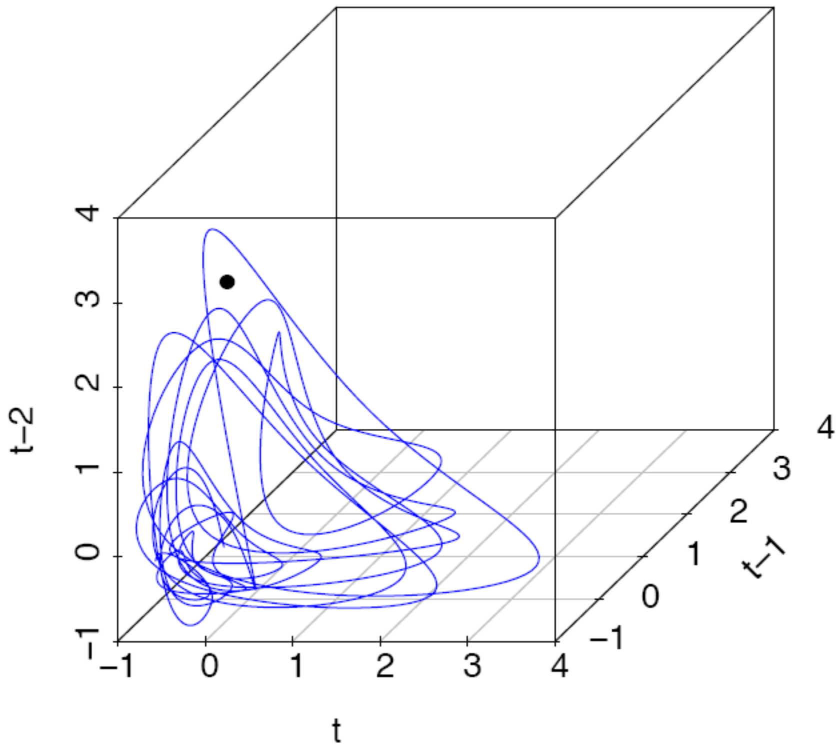
Empirical dynamic modelling

So local neighbourhoods (and their trajectories) in the reconstruction then map to local neighbourhoods (and *their* trajectories) of the original system.

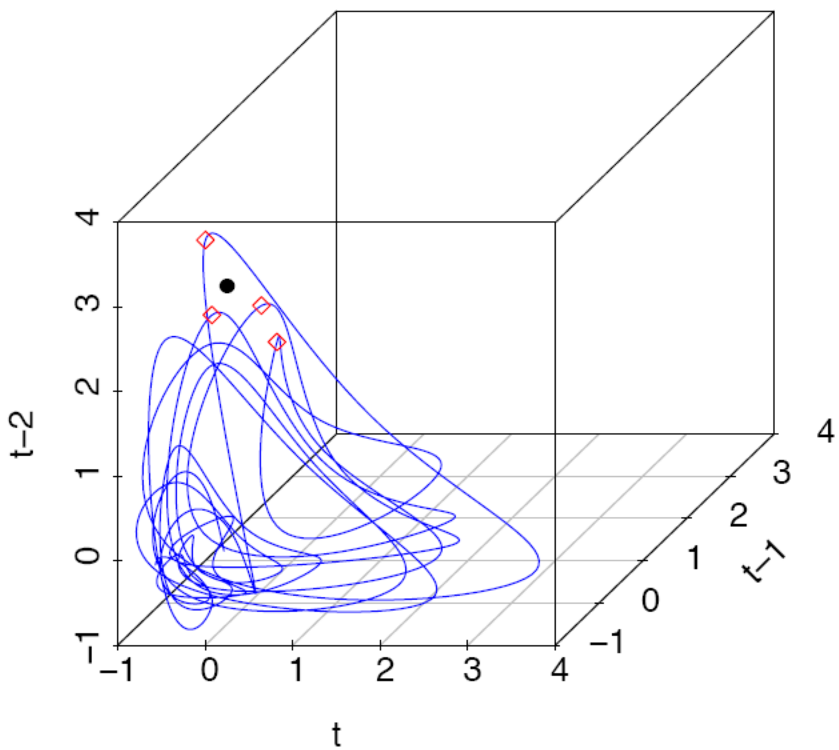
Permits forecasting: find nearest neighbours from historical record and use their behaviour to estimate evolution of system through time (e.g Simplex projection).

Identify *k* nearest neighbours. Take weighted average of mapped values.

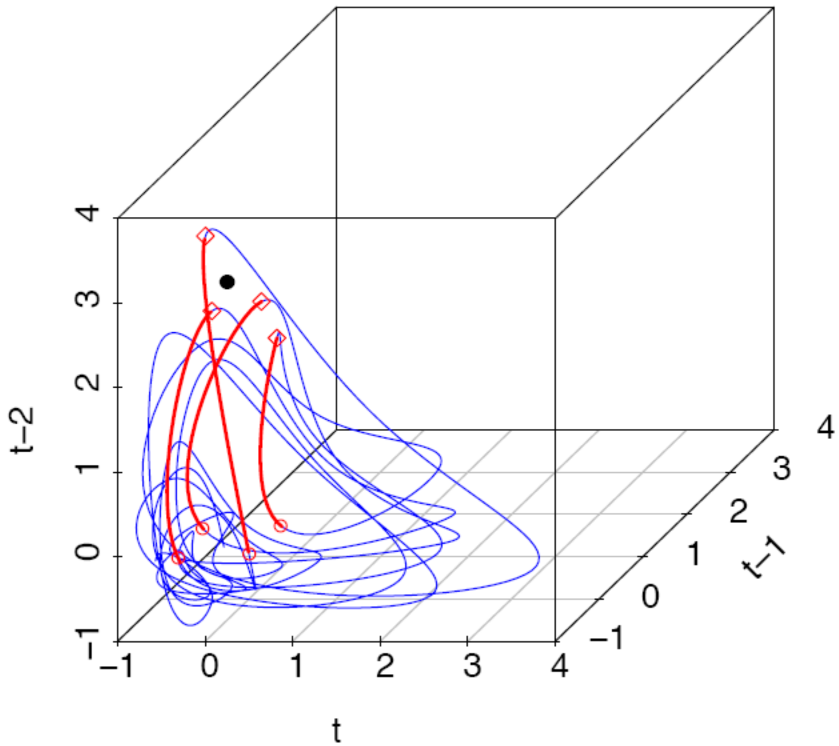
Simplex Projection



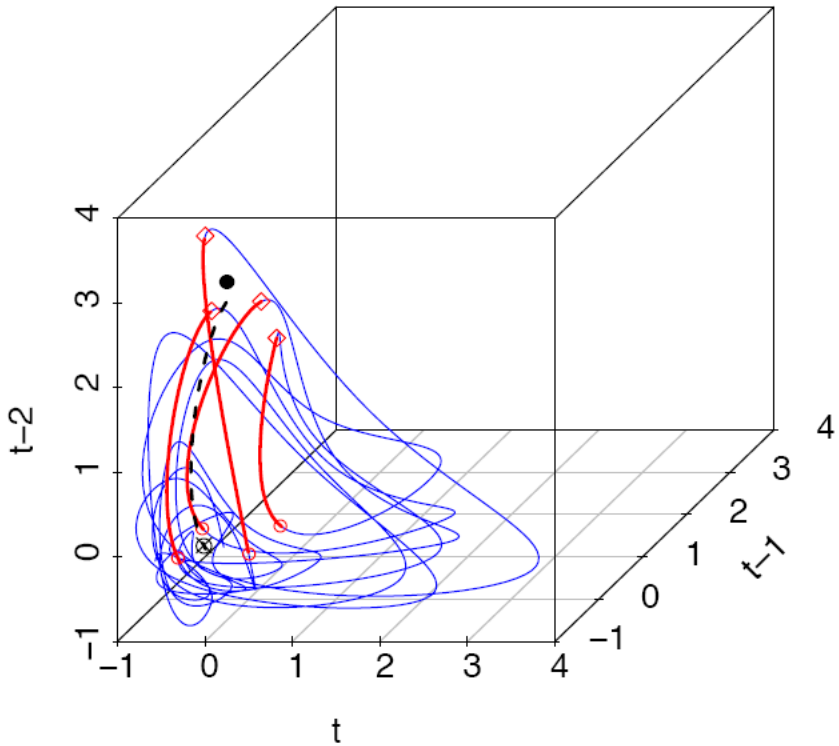
Simplex Projection



Simplex Projection



Simplex Projection



Empirical dynamic modelling

So local neighbourhoods (and their trajectories) in the reconstruction then map to local neighbourhoods (and *their* trajectories) of the original system.

Permits forecasting: find nearest neighbours from historical record and use their behaviour to estimate evolution of system through time (e.g Simplex projection).

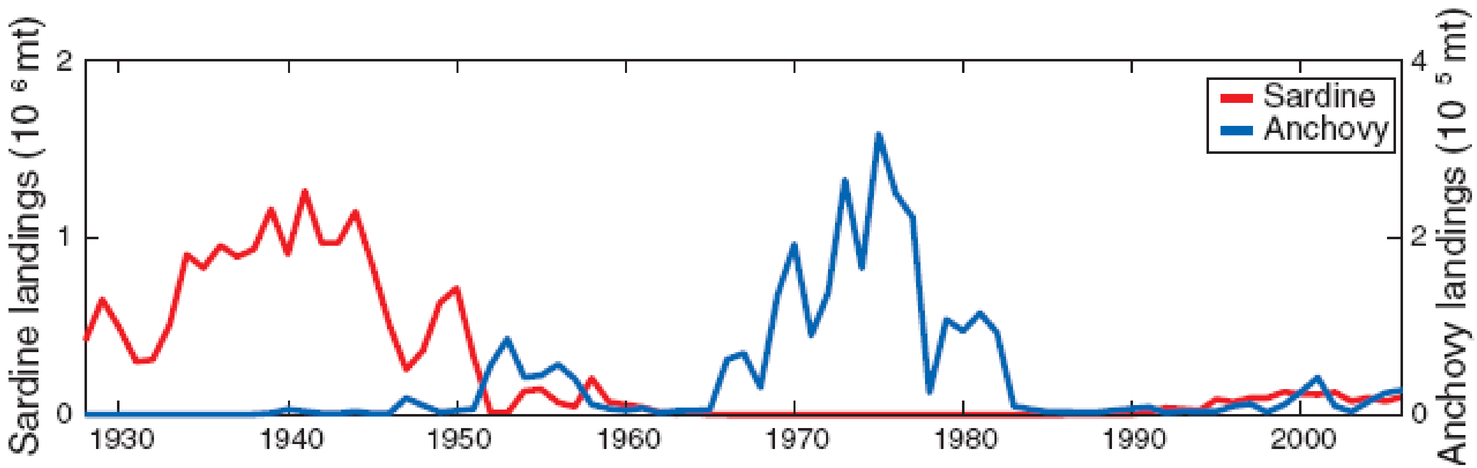
Takens' theorem: the reconstructed (time-delay) system preserves essential mathematical and geometrical properties of the original system (given enough data).

Distinguishing causality from correlation

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]

26 OCTOBER 2012 VOL 338 SCIENCE



Data from California: sardine landings, anchovy landings, SST.

EDM: not direct competition, but both influenced by SST.

Causation

If SST influences sardine population.

Which time series contains information about the other time series:

- SST?
- sardine population?

Causation

If SST influences sardine population.

Which time series contains information about the other time series:

- SST?
- sardine population?

Sardine time series – it *contains a signature* of the SST.

If you know the sardine time series this can help you infer the temperature.

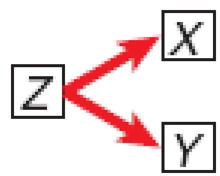
If you know the temperature time series, this will not help you infer the sardines (without a mechanistic model).

Examples of coupling

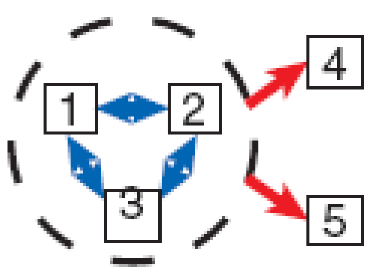
Case i:
Bidirectional coupling

Case ii:
Unidirectional coupling

Example 1:
External forcing of non-coupled variables

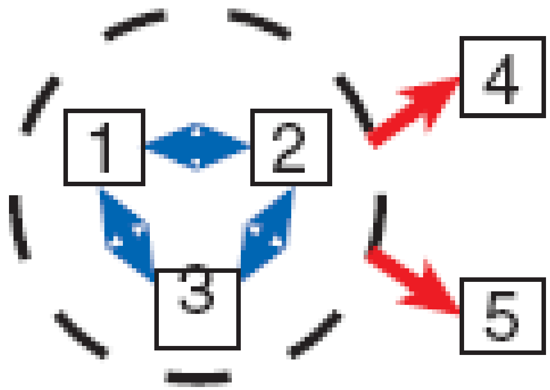


Example 2:
Complex model



Identifying influences – simulations

Example 2:
Complex model



Five-species simulation model.

Convergent cross mapping (CCM) correctly identifies influences:

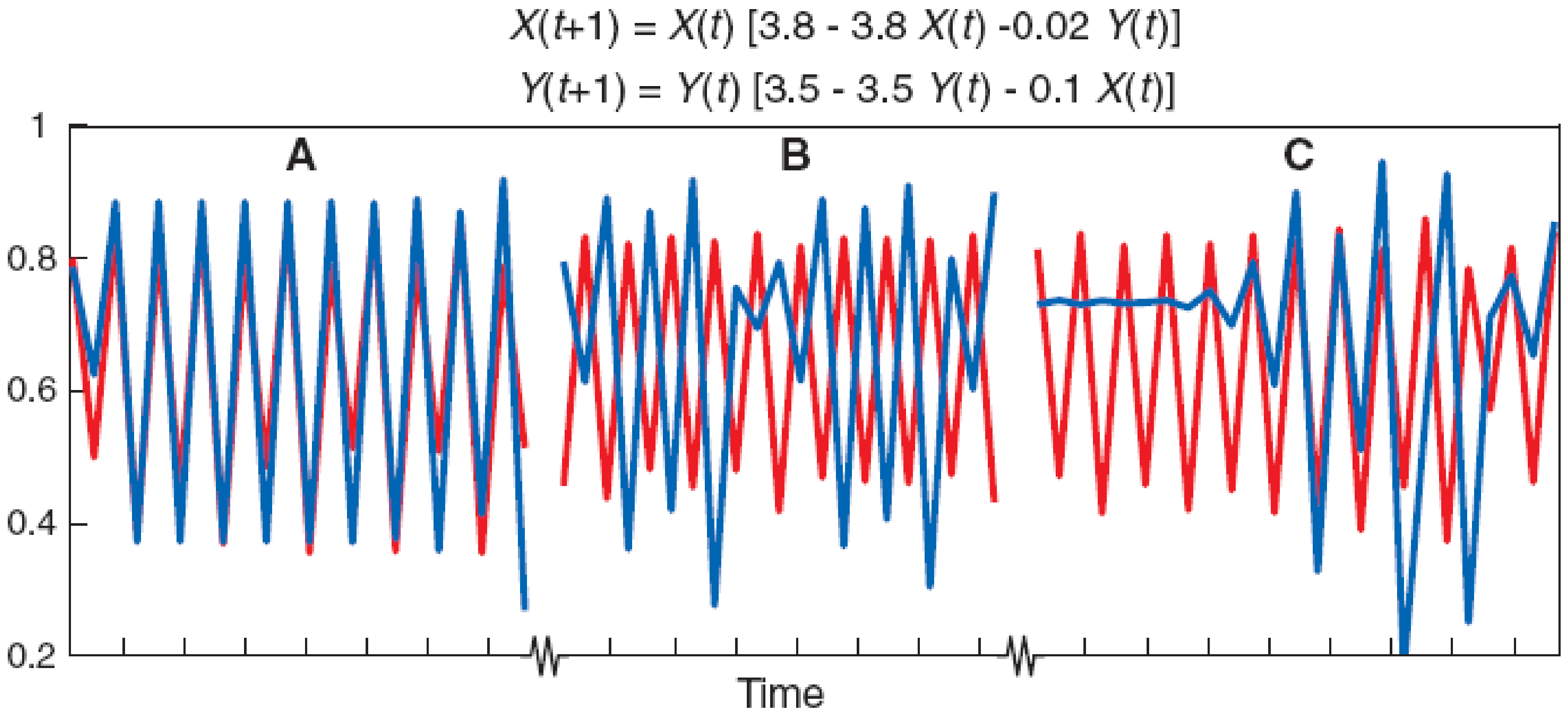
Causal links (cross map ρ):

1 → 2 (1.00)	1 → 4 (0.50)	1 → 5 (0.21)
2 → 1 (1.00)	2 → 4 (0.60)	2 → 5 (0.13)
1 → 3 (1.00)	3 → 4 (0.51)	3 → 5 (0.25)
3 → 1 (1.00)		
3 → 2 (1.00)		
2 → 3 (1.00)		

*All other links not significant

Mirage correlations – simulations

Deterministic competition model:



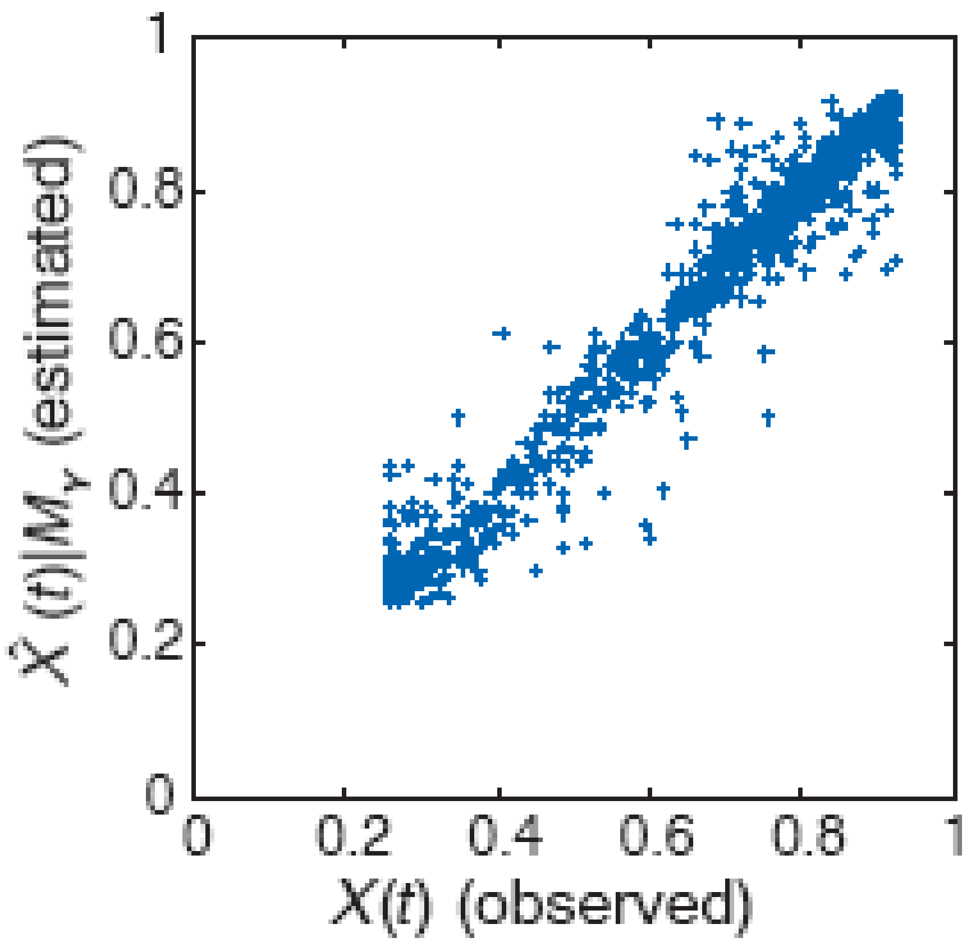
correlated anti-correlated no correlation

Convergent cross mapping (CCM) can identify influences.

Correlations

Estimate ¼ of *X* values using Simplex approach.

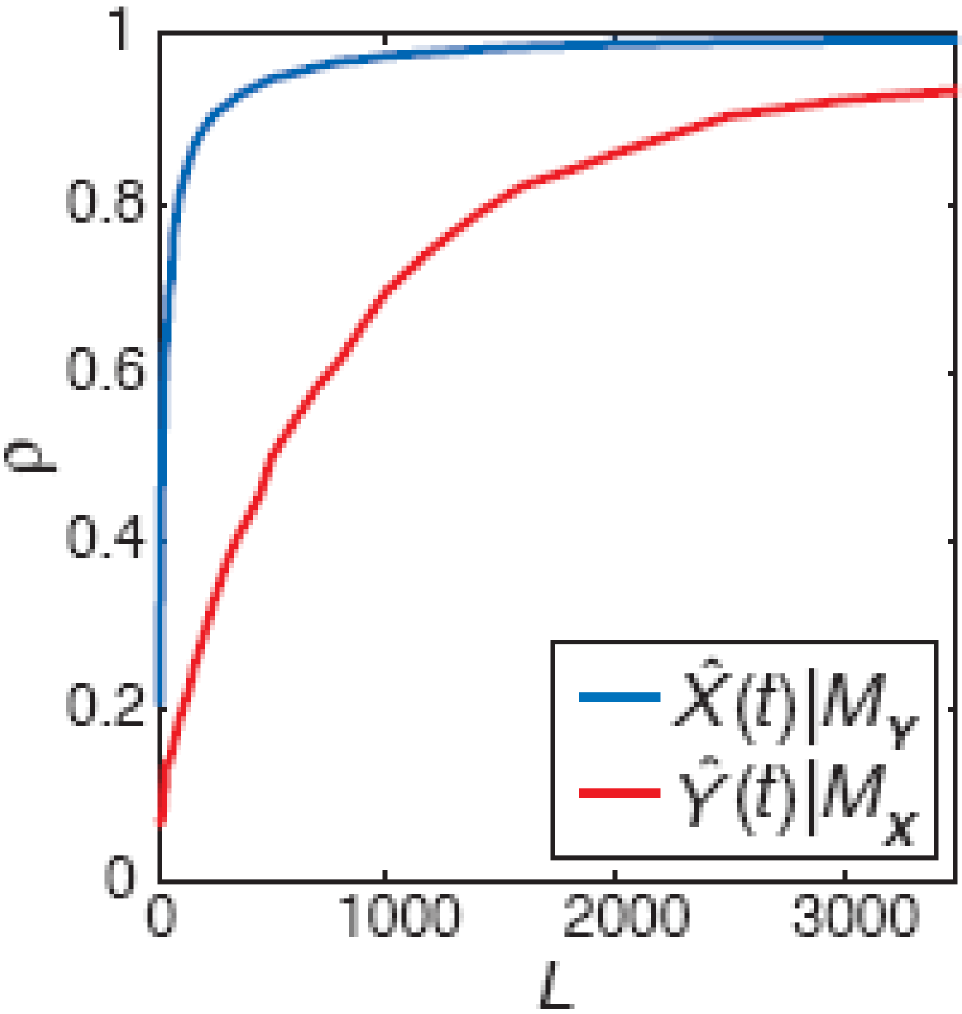
estimate of ¼ of *X*, given *Y* and ¾ of *X*



Correlation coefficient is ρ

[example is slightly different from previous slide]

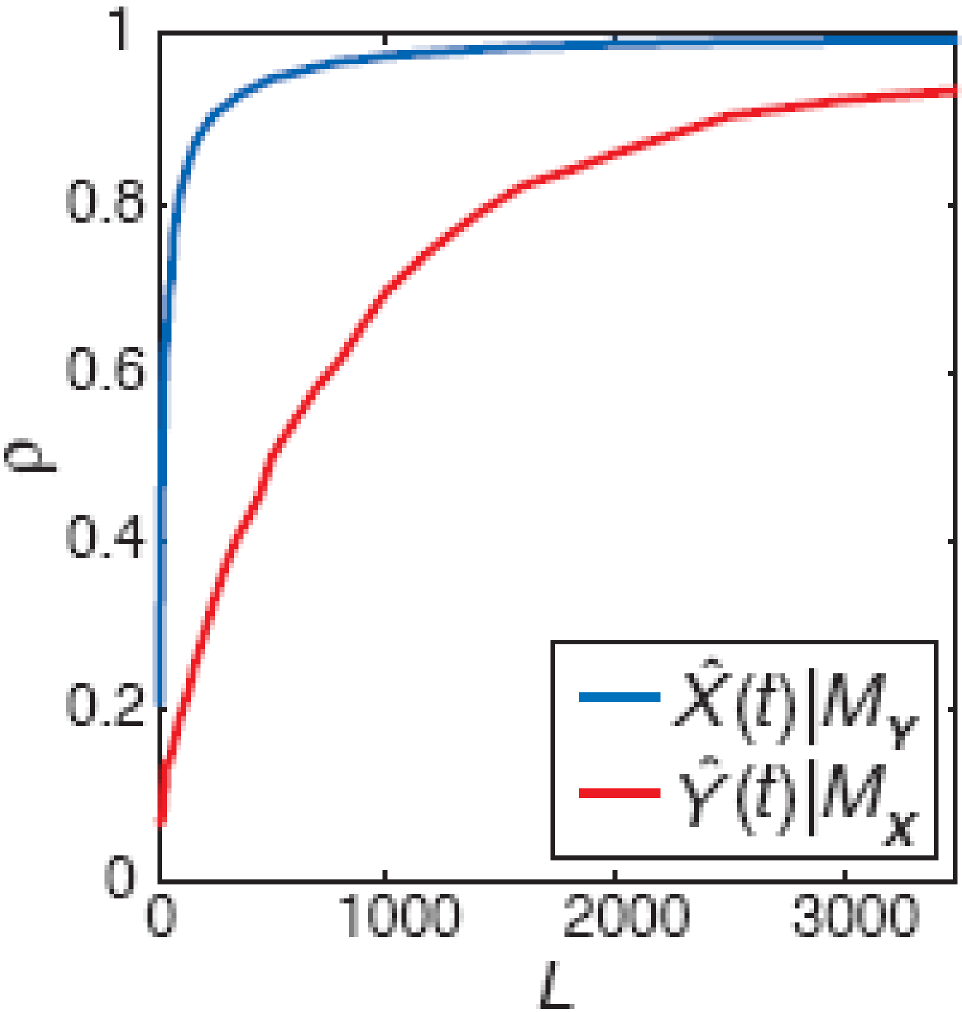
Correlations depend on length of time series



Estimates of ¼ of *X* given *Y* and ¾ of X become very good.

Remember, this has no knowledge of the equations.

Correlations depend on length of time series



Estimates of *Y* given *X* also improve with *L*, but not as fast.

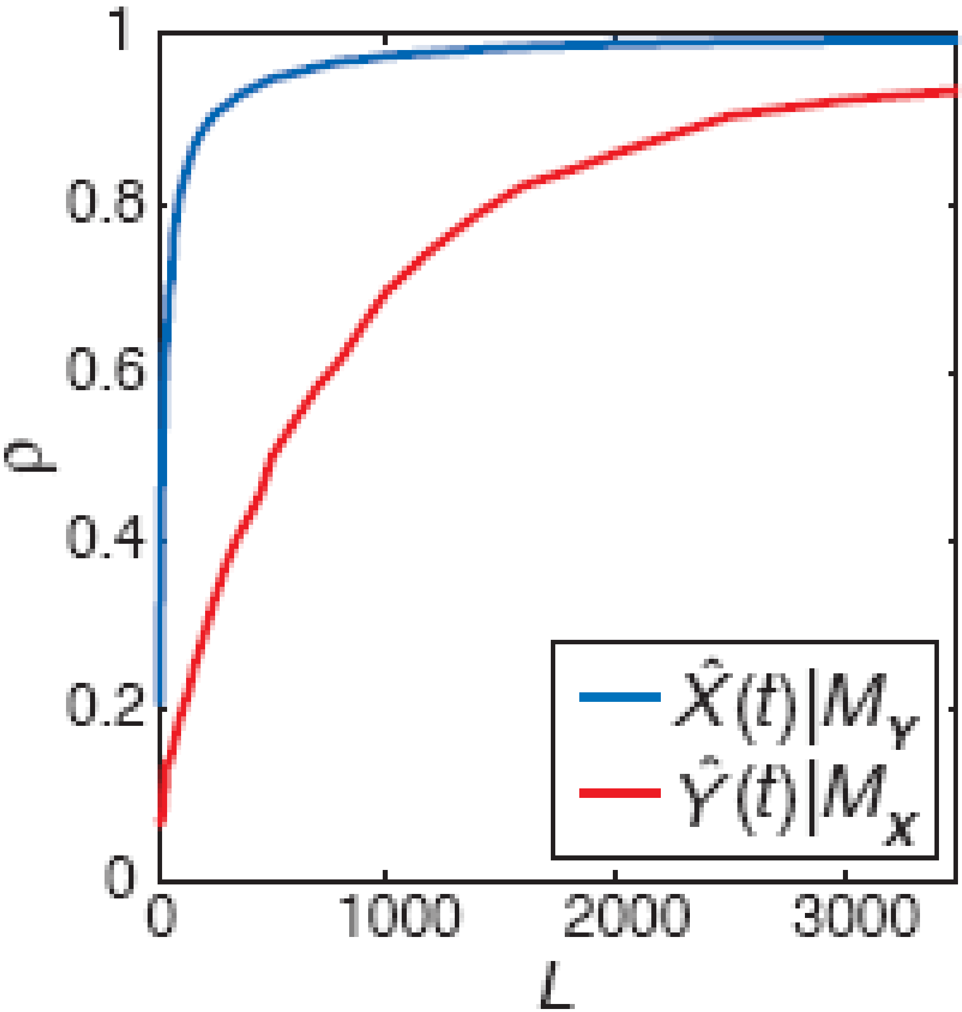
Why?

effect of *X* on *Y* (0.1)

\>

effect of *Y* on *X* (0.02)

Correlations depend on length of time series



Estimates of *Y* given *X* also improve with *L*, but not as fast.

Why?

effect of *X* on *Y* (0.1)

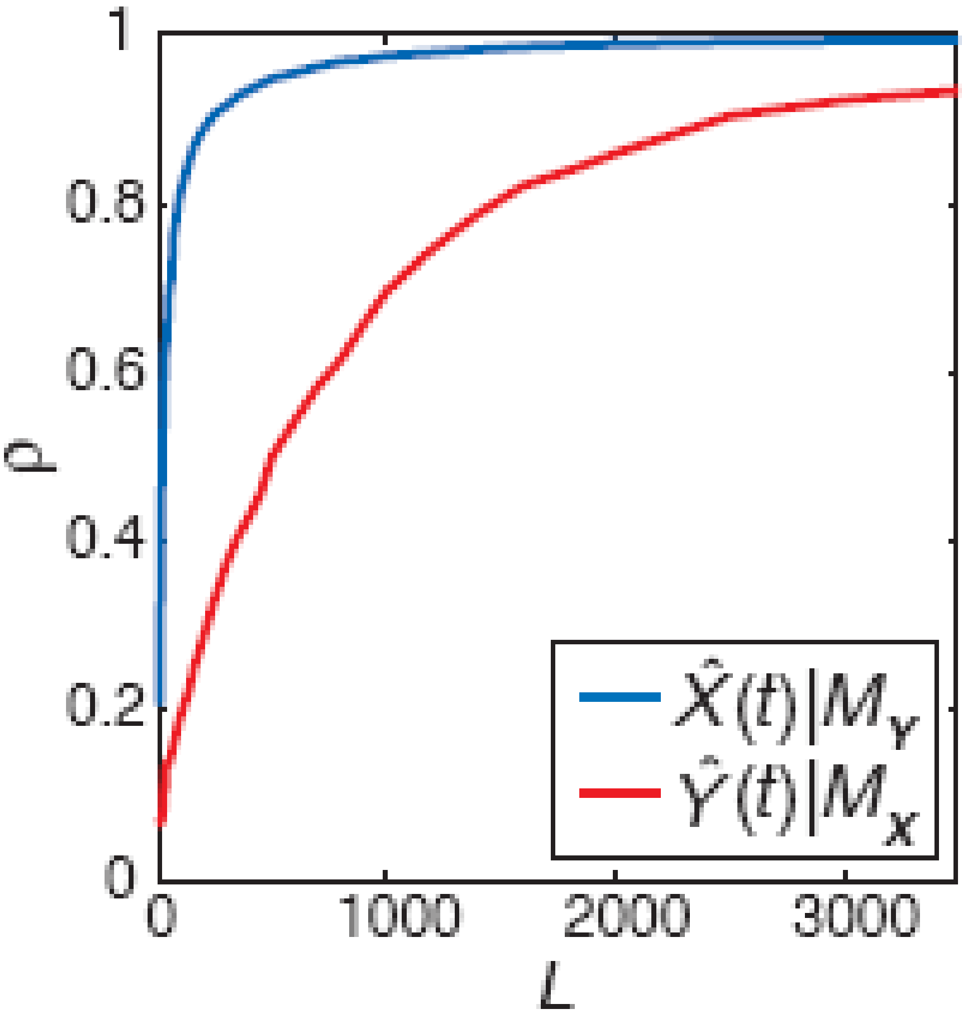
\>

effect of *Y* on *X* (0.02)

$$X(t+1) = X(t)\,[3.8 - 3.8\,X(t) - 0.02\,Y(t)]$$

$$Y(t+1) = Y(t)\,[3.5 - 3.5\,Y(t) - 0.1\,X(t)]$$

Correlations depend on length of time series



$X(t+1) = X(t)\ [3.8 - 3.8\ X(t) - 0.02\ Y(t)]$

$Y(t+1) = Y(t)\ [3.5 - 3.5\ Y(t) - 0.1\ X(t)]$

Estimates of *Y* given *X* also improve with *L*, but not as fast.

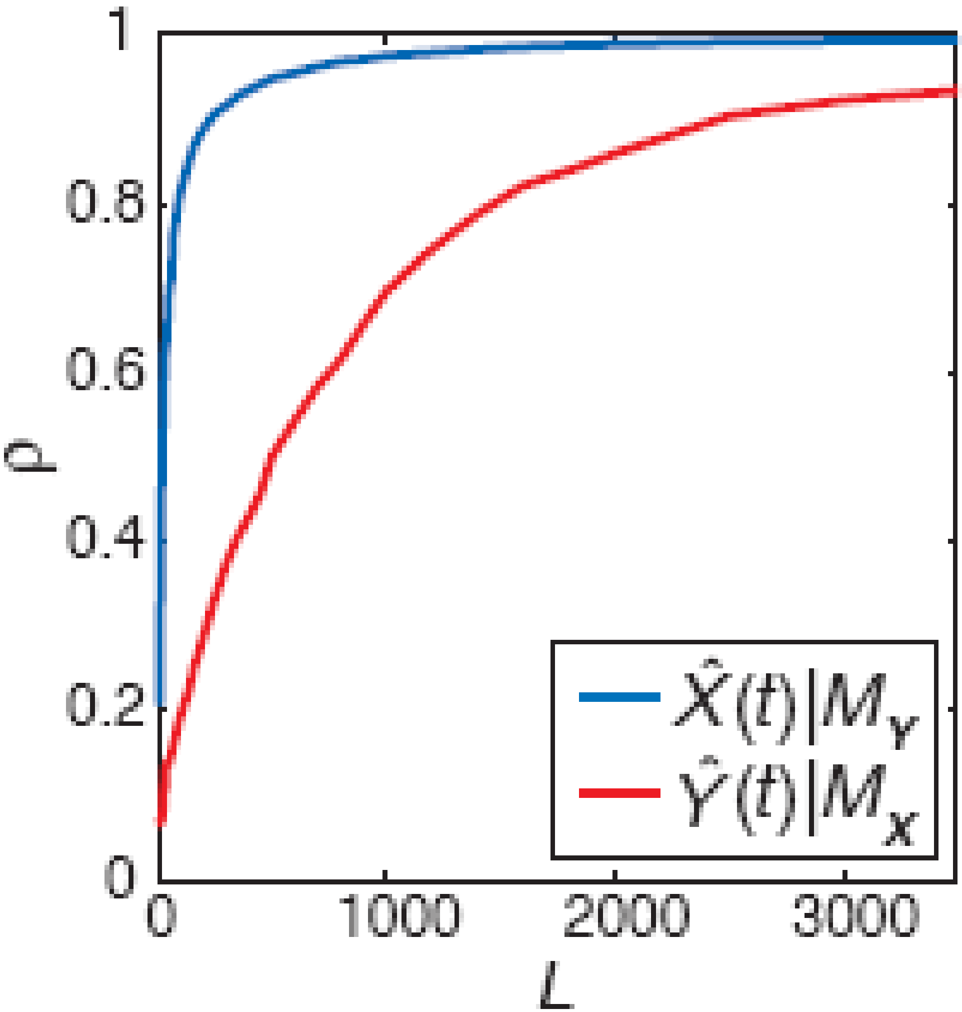
Why?

effect of *X* on *Y* (0.1)

>

effect of *Y* on *X* (0.02)

Correlations depend on length of time series



Estimates of *Y* given *X* also improve with *L*, but not as fast.

Why?

effect of *X* on *Y* (0.1)

>

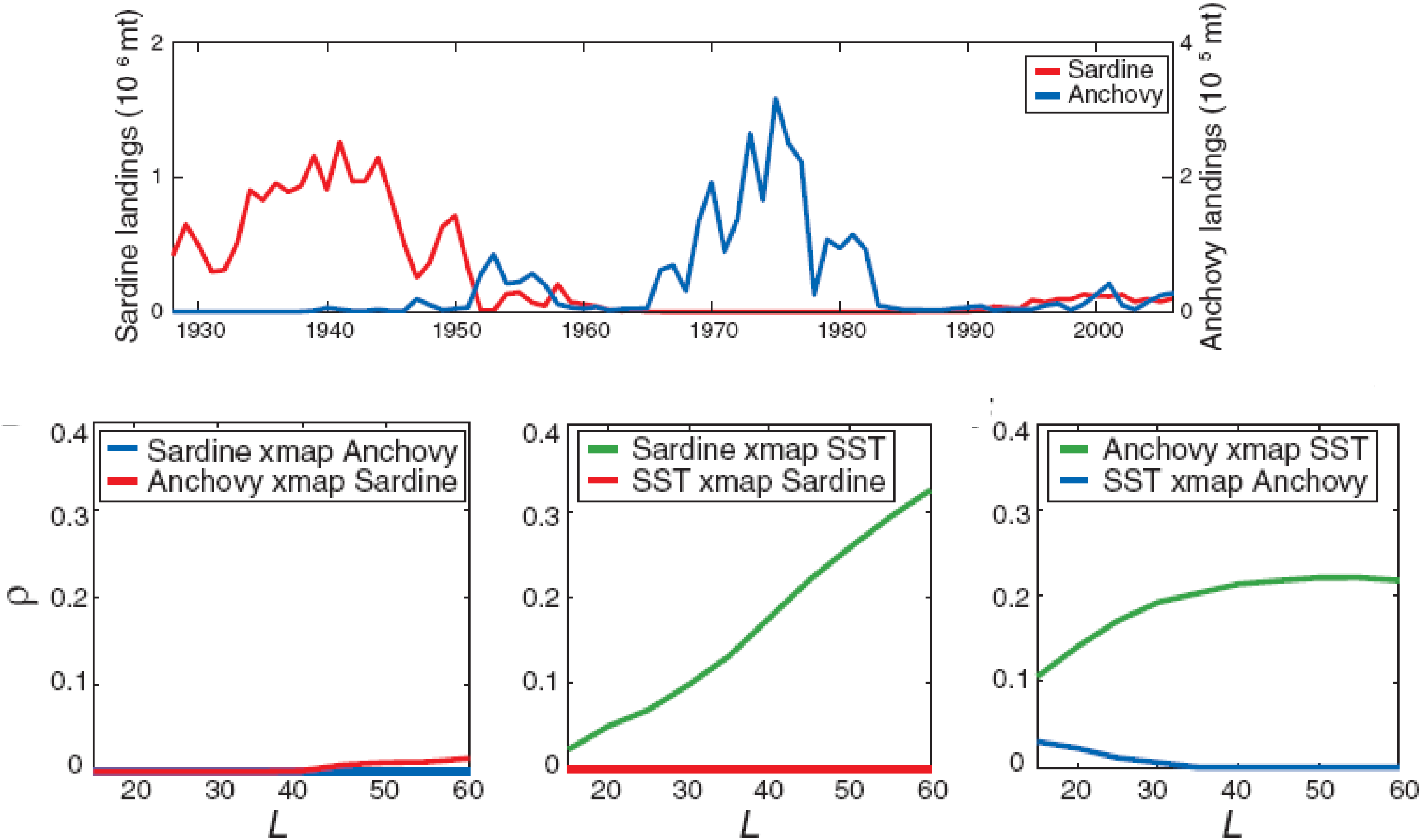
effect of *Y* on *X* (0.02)

So *Y* data contain more 'signature' of *X* data than *X* contains of *Y*.

$$X(t+1) = X(t)\,[3.8 - 3.8\,X(t) - 0.02\,Y(t)]$$

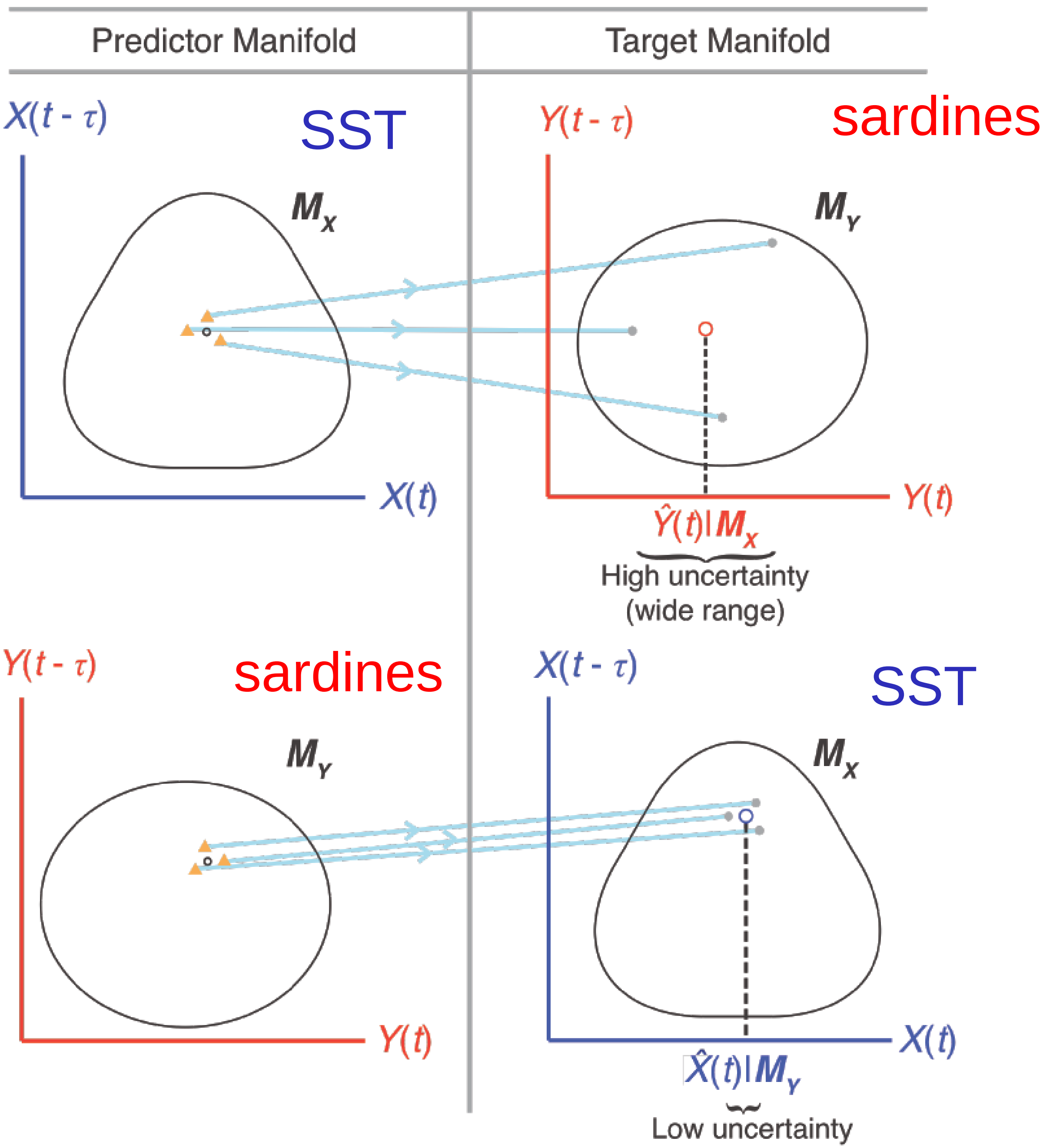
$$Y(t+1) = Y(t)\,[3.5 - 3.5\,Y(t) - 0.1\,X(t)]$$

Distinguishing causality from correlation



EDM: not direct competition, but both influenced by SST.

Asymmetric Causality, $X \Rightarrow Y$ SST influences sardines



Practical impact – management implications

“... measurable nonlinear coupling of SST to sardine stocks means that effect of SST varies with system state.”

“... fixed temperature index [as in current regulatory framework] will not suffice for sound management decisions.”

“Rather, a dynamic (state-dependent) rule involving SST is required.”

Summary

This is a different approach to using ecosystem information.

Don’t fully understand it yet, but seems worth investigating.

Gives interesting results for environmental influences on Sockeye Salmon and then sardines and anchovies.

Acknowledgments

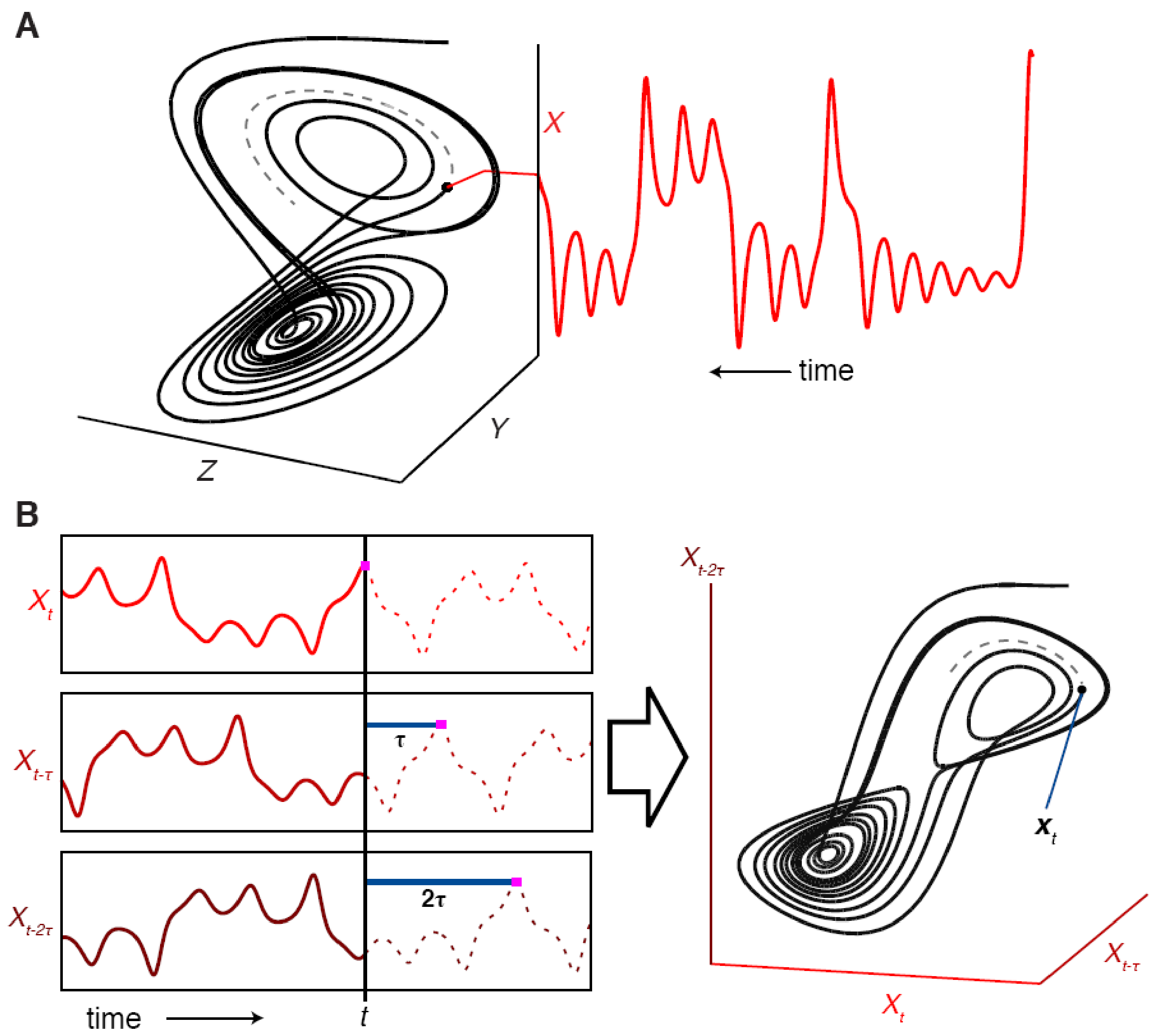
Hao Ye for sharing a recent presentation (some of which I adapted here) and providing R code with papers.

Carrie Holt.

Avenues of exploration for working group

1. Run the code (vignettes from rEDM package). Try and understand (but don't get caught up in all the details for now).
2. Sockeye Salmon results: a) understand better
   b) test other metrics of accuracy (2012 Res. Doc.)
   c) add spurious forcing, what do methods conclude?
3. Sardine example – understand better.
4. Simulated data from known models – understand better.
5. Explore other data sets, such as:
   a) Pacific Herring catches, surveys and SST
   b) North Sea bottom trawl surveys (30 years, many species, size-based).
6. I have set up a GitHub site with example code and papers.

7. Use EDM as operating model in MSE or DLMtool?!?



A
X
time
Y
Z
B
X_t
$X_{t-\tau}$
$X_{t-2\tau}$
τ
2τ
time
t
$X_{t-2\tau}$
$\mathbf{x}_t$
$X_{t-\tau}$
X_t

Data from North Sea bottom trawl surveys

Year	Species	Length class (cm)	Number (h^{-1})	α	β	Body mass (g)	Biomass (g)
1986	Smallspotted Catshark	45	0.050	0.0031	3.0290	315.46	15.77

Further pre-processing (e.g. only body masses >4 g), results in

42,298 combinations of Year / Species / Length Class.

Data from North Sea bottom trawl surveys

Year	Species	Length class (cm)	Number (h^{-1})	α	β	Body mass (g)	Biomass (g)
1986	Smallspotted Catshark	45	0.050	0.0031	3.0290	315.46	15.77
1986	Smallspotted Catshark	46	0.050	0.0031	3.0290	337.17	16.86
1986	Smallspotted Catshark	50	0.050	0.0031	3.0290	434.05	21.70
1986	Smallspotted Catshark	52	0.205	0.0031	3.0290	488.81	100.33
1986	Smallspotted Catshark	53	0.076	0.0031	3.0290	517.84	39.52
1986	Smallspotted Catshark	54	0.079	0.0031	3.0290	548.00	43.26
...	...	...	...	...	...	...	...
2015	Snakeblenny	34	0.195	0.0244	2.0439	32.93	6.42
2015	Thickback Sole	8	0.091	0.0080	3.1410	5.49	0.50
2015	Thickback Sole	14	0.273	0.0080	3.1410	31.85	8.69
2015	Thickback Sole	15	0.364	0.0080	3.1410	39.55	14.38
2015	Thickback Sole	16	0.455	0.0080	3.1410	48.44	22.02
2015	Thickback Sole	17	0.091	0.0080	3.1410	58.60	5.33

Forecast accuracy