"Inferences about coupling from ecological surveillance monitoring: nonlinear dynamics, information theory..."

(...and submodular functions??)

Evan Cooch

Department of Natural Resources
Cornell University

March 1, 2011

Acknowledgements

- Steve Ellner (*Cornell University*)
- James (Jim) Nichols (*Patuxent Wildlife Research Centre*)
- Jonathon Nichols (*Naval Research Labs*)
- Linda Moniz (*Johns Hopkins University*)
- Lou Pecora (*Naval Research Labs*)

science

- ‘understand ecological systems
- ‘learn stuff’

management

- apply decision-theoretic approaches
- make ‘smart’ decisions

monitoring in management

- Determine system state for state-dependent decisions
- Determine system state to assess degree to which management objectives are achieved
- Determine system state for comparison with model-based predictions to learn about system dynamics (i.e., do science)

what to monitor?

- **community** - multiple species
  - State variable: species richness
  - Vital rates: rates of extinction and colonization
- **patch** - single species
  - State variable: proportion of patches occupied
  - Vital rates: P(patch extinction/colonization)
- **population** - single species
  - State variable: abundance
  - Vital rates: P(survival, reproduction, movement)

choice depends on...

- **monitoring objectives**
  - Science: what hypotheses are to be addressed?
  - Management/conservation: what are the objectives?
- **geographic and temporal scale**
- **effort available for monitoring**
  - Required effort: species richness, patch occupancy $<$ abundance

monitoring as an 'enterprize'

- monitoring most useful when integrated into science or management
- both typically hypothesis-driven
- what about cases where
  - (near-)complete absence of information about system?
  - *surveillance monitoring* programs already established?

surveillance monitoring

- monitoring designed in the absence of guiding hypotheses about system behaviour
- scientific approach: retrospective observational
- objective: to learn inductively about a system and its dynamics by observing time series of system state variables
- new programs: should be a last resort
- existing programs: many were designed as surveillance programs

the problem(s) with surveillance monitoring

- surveillance monitoring sometimes represents a form of intellectual displacement behavior
  - easier to suggest collection of more data than to think hard about the most relevant data to collect
- at cynical worst, surveillance monitoring represents a political delaying tactic
- feeds anti-science view of science as never-ending story with few answers and little interaction with real world decision-making

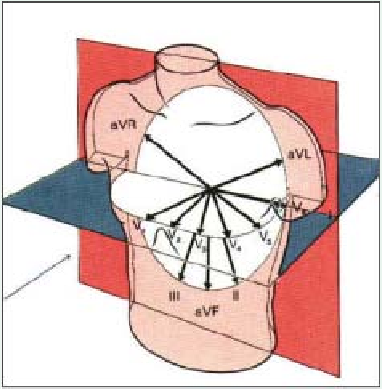
a proposed formalism for surveillance monitoring

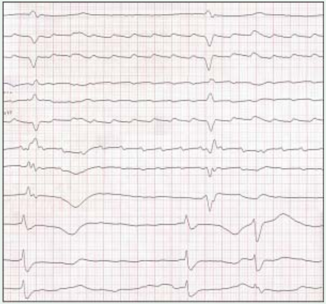
- despite inherent inefficiency: attempt to develop a reasonable approach to retrospective analyses
- view time series as sources of information and consider methods of extraction
- conceptual underpinnings reside in methods of *nonlinear dynamics* and *information theory*
- consider inductive inferential methods for:
    - system identification
    - characterization of interactions among system components
    - detection of system change and degradation

curse of non-linear, high-dimensional systems

- system dynamics complex
- dynamics often both non-linear, and 'noisy'
- where do you monitor the system?

example - cardiac function





how many variables to monitor? what variables to monitor?

example: 1 selective predator (P), 2 competing prey (H_i)

$$\frac{\partial H_1}{dt} = H_1\left(r_1 - \gamma_{11}H_1 - \gamma_{12}H_2 - \gamma_{1P}P\right)$$

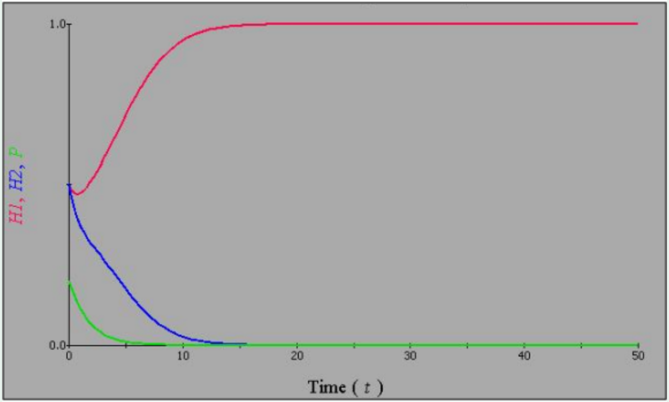
$$\frac{\partial H_2}{dt} = H_2\left(r_2 - \gamma_{22}H_2 - \gamma_{21}H_1 - \gamma_{2P}P\right)$$

$$\frac{\partial P}{dt} = P\left(\gamma_{P1}H_1 + \gamma_{P2}H_2 - r_P\right)$$

$$\gamma_{21} > \gamma_{12} \qquad \gamma_{P1} > \gamma_{P2}$$

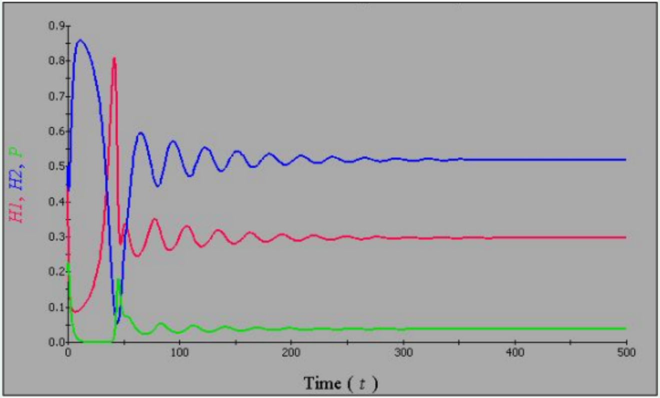
$\gamma_{1P} = \gamma_{2P}$

no selectivity – stable attractor (fixed point)



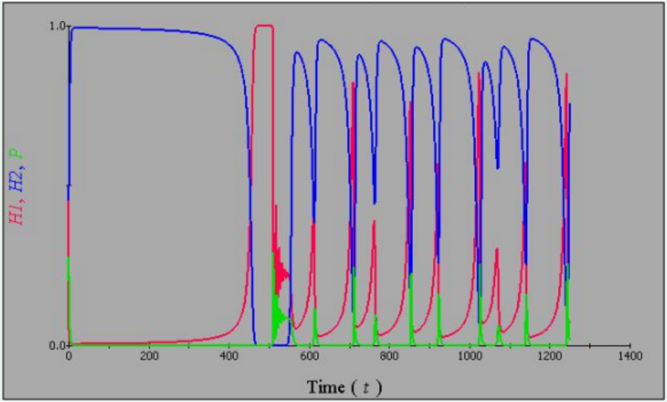
$\gamma_{1P} > \gamma_{2P}$

moderate selectivity for prey 1 – stable attractor (fixed point)

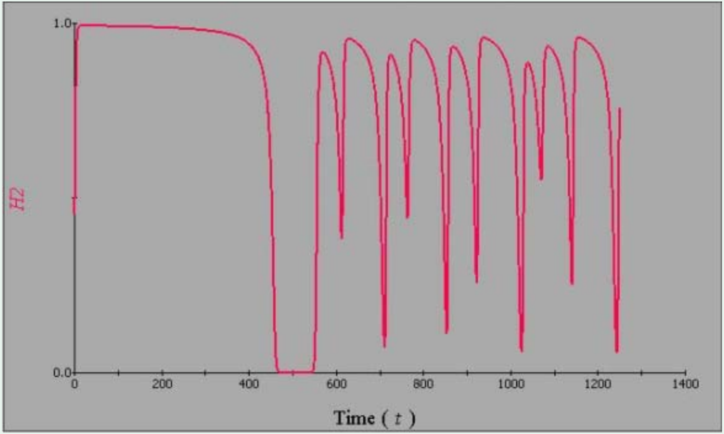


$\gamma_{1P} \gg \gamma_{2P}$

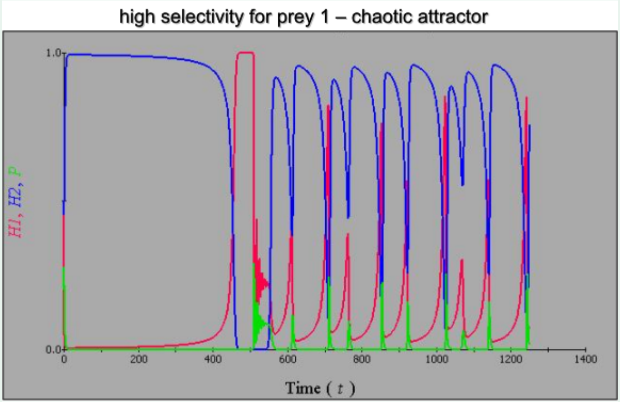
high selectivity for prey 1 – chaotic attractor



reconstruct underlying dynamics from single species?



chaotic attractor



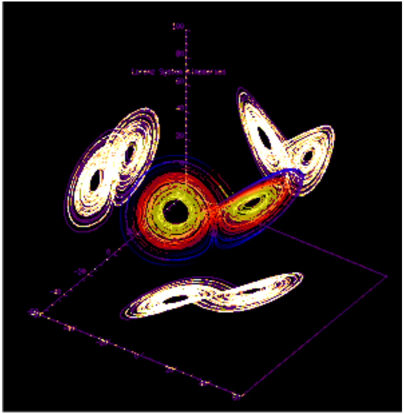
system attractor: closed set of points in state space, such that a trajectory starting on or near attractor will converge to it

Lorenz system

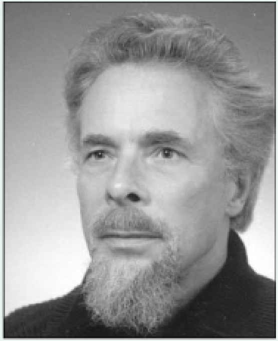
$$\frac{dx}{dt} = \sigma(y - x)$$

$$\frac{dy}{dt} = x(r - z) - y$$

$$\frac{dz}{dt} = xy - \beta z$$



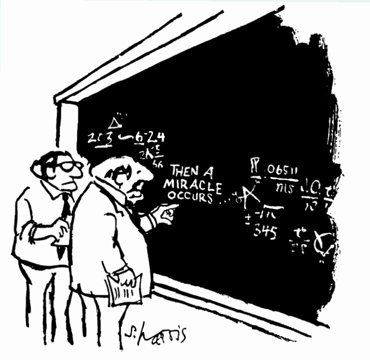
Takens' theorem



- **any** dynamical system can be reconstructed from a sequence of observations of the state of the dynamical system
- given data from single system variables, reconstruct a *diffeomorphic* copy of the attractor of the system by lagging the time-series to embed it in more dimensions

in other words...

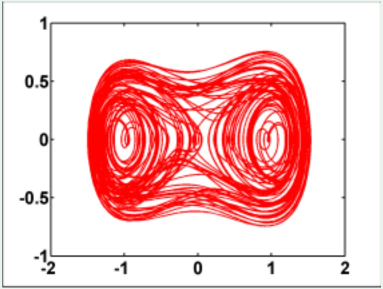
Clear as mud, eh? In other words, if we have a point $f(x, y, z, t)$ which is wandering along some strange attractor (like the Lorenz), and we can only measure $f(z, t)$, we can plot $f(z, z + N, z + 2N, t)$, and the resulting object will be topologically identical to the original attractor.



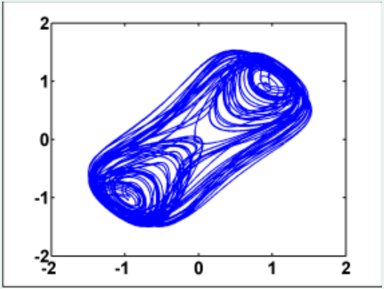
"I think you should be more explicit here in step two."

skipping some of the technical details...

actual attractor



reconstructed attractor



diffeomorphic = topological = dynamical equivalence

focus → dynamical interdependence (coupling)

- **Data**: time series of 2 different state variables
- **Questions**:
  - are they functionally related?
  - what can we learn about 1 state variable by following or knowing another?
- **Ecological applications**:
  - monitoring program design (indicator species, etc.)
  - population synchrony and its cause(s)
  - food web connectance
  - competitive interactions
  - detection of system change and degradation

coupling - old and new methods

- **linear cross-correlation**:
  - Compute ρ in usual manner based on the 2 time series, $x(t)$ and $y(t)$
- **attractor-based methods** (no restriction to linear systems):
  - if 2 state variables are dependent and belong to same system, their attractors should exhibit similar geometries
  - (1) continuity: focus on function relating 2 attractors
  - (2) mutual prediction: degree to which dynamics of 1 attractor can be used to predict dynamics of the other
- **information-based methods** (mutual information, transfer entropy)

Example 1: Pascual (1993)

- 100 patches with linear gradient in prey resource abundance, decreasing from location 0.01 to 1.00
- Prey growth (*r*) is function of resources
- both prey and predator disperse via diffusion
- simple - one-dimensional system

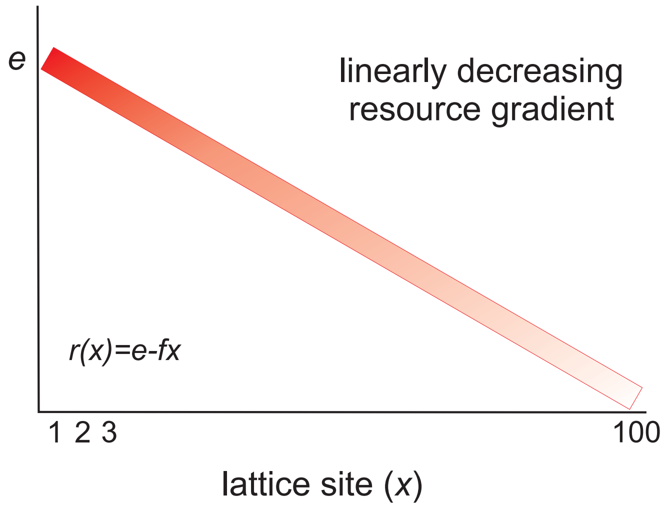
model equations

$$\frac{\partial p}{\partial t} = r(x)p(1-p) - \frac{ap}{1+bp}h + D\frac{\partial^2 p}{\partial x^2}$$

$$\frac{\partial h}{\partial t} = \frac{ap}{1+bp}h - mh + D\frac{\partial^2 h}{\partial x^2}$$

$$r(x) = e - fx$$

a = predation rate = 'species' coupling
D = diffusion rate = diffusive 'spatial' coupling


e
linearly decreasing
resource gradient
r(x)=e-fx
1 2 3
100
lattice site (x)

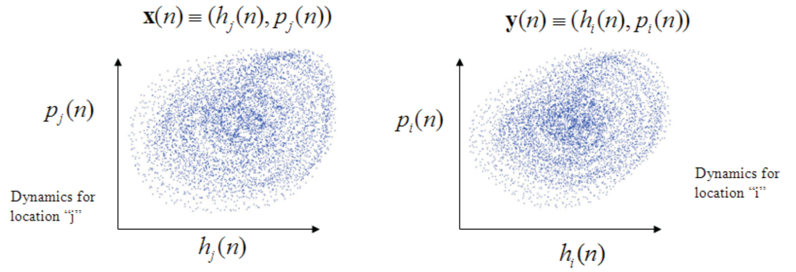
Cross-correlation: standard technique in ecology

$$c_{xy}(k) = \frac{1}{N-k} \sum_{i=1}^{N-k} (x(i) - \bar{x})\,(y(i+k) - \bar{y})$$

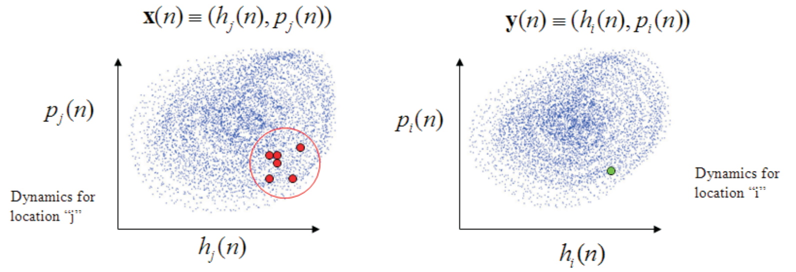
Mutual Prediction: Let one lattice site predict the dynamics of the others. Good predictions imply strong coupling

$$\gamma = \frac{1}{\sigma^2} \sum_{f=1}^{N} \|\hat{y}(f+s) - y(f+s)\|$$

mutual prediction algorithm



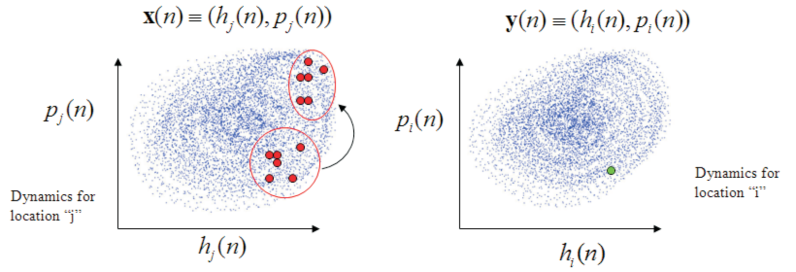
mutual prediction algorithm



Choose fiducial point on one attractor (location 2) and locate nearest neighbors within radius ϵ on other attractor (location 1)

$$x(p_j) : \|x(p_j) - y(f)\| < \epsilon$$

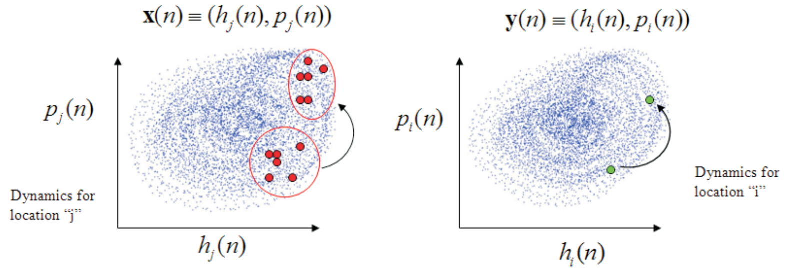
mutual prediction algorithm



Use neighborhood to make *s*-step prediction (simplest is to use average of time-evolved near neighbors)

$$\hat{y}(f+s) = \frac{1}{|n_b|} \sum_j x(p_j + s)$$

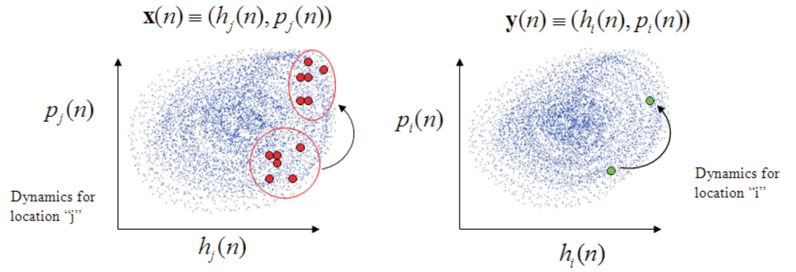
mutual prediction algorithm



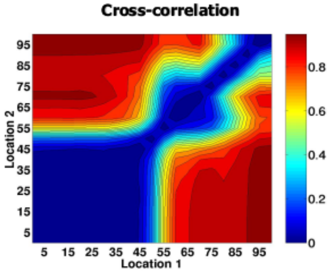
Record difference between actual and predicted values as nonlinear prediction error

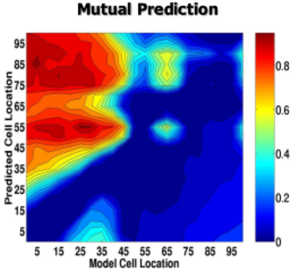
$$\gamma_f = \frac{1}{\sigma^2} \|\hat{y}(f+s) - y(f+s)\|$$

mutual prediction algorithm



good predictions $\rightarrow$ generalized synchrony $\rightarrow$ strong coupling closer coupling indicated by smaller values (blue)



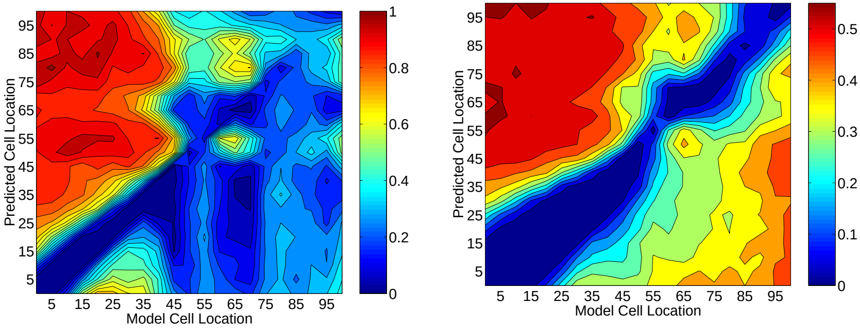


asymmetry cannot (by definition) be seen using cross-correlation function

Information about higher resource dynamics is contained in lower resource dynamics, but reverse is not true

what about Takens' theorem?

mutual prediction (2-state) | reconstructed MP (1-state)



alternatives to attractor reconstruction

- attractor-based approaches good, but other methods available
- information theoretic approaches - formal characterization of direction of information flow
- sporadic use in ecology
- most familiar use is measure of species diversity (e.g., Shannon)

Kullback entropy

- Kullback entropy, K_Y, focuses on discrepancy in information between the true probability distribution, $p(y_i)$, and a different distribution, $q(y_i)$:
- K_Y is the difference (excess) in average number of bits needed to encode draws of Y if $q(y_i)$ is used instead of $p(y_i)$

$$K_Y = \sum_y p\left(y_i\right) \log\left(\frac{p(y_i)}{q(y_i)}\right)$$

mutual information

- $I(Y, Z)$ = mutual information = average amount of information (in bits) about 1 state variable gained by knowing the value of the other state variable
- $y_i, z_i =$ discrete random variables at time i
- pdfs $[p(y_i), p(y_i, z_i)]$ estimated empirically based on "bin counting" approaches

$$I(Y, Z) = \sum_{y,z} p(y_i, z_i) \log_2 \frac{p(y_i, z_i)}{p(y_i)p(z_i)}$$

mutual information and entropy

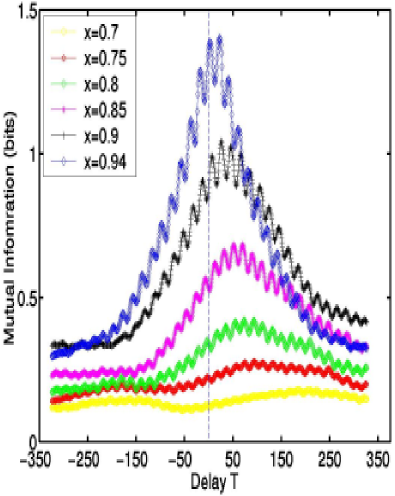
- $I(Y,Z)$ can be viewed as a Kullback entropy (excess code produced by erroneously assuming that Y and Z are independent)
- $I(Y,Z)$ focuses on the deviation of the 2-state system from independence

$$I(Y,Z) = \sum_{y,z} p(y_i, z_i) \log_2 \frac{p(y_i, z_i)}{p(y_i)p(z_i)}$$

time-lagged mutual information

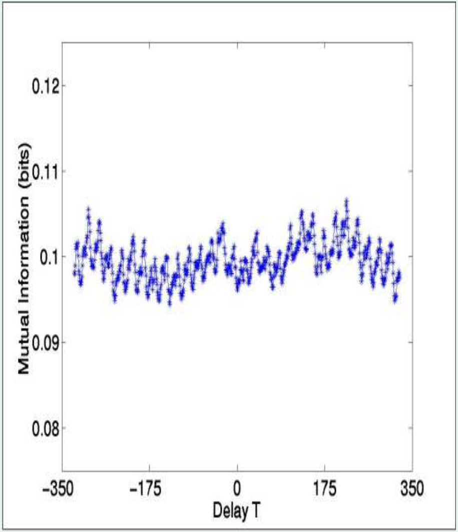
- focus on directionality of information flow
- search to find delay T at which $I(Y, Z_T)$ is maximum
- $T > 0$ suggests information transport from $Y \to Z$
- $T < 0$ suggests information transport from $Z \to Y$

$$I(Y, Z_T) = \sum_{y,z} p(y_i, z_{i+T}) \log_2 \frac{p(y_i, z_{i+T})}{p(y_i)p(z_{i+T})}$$



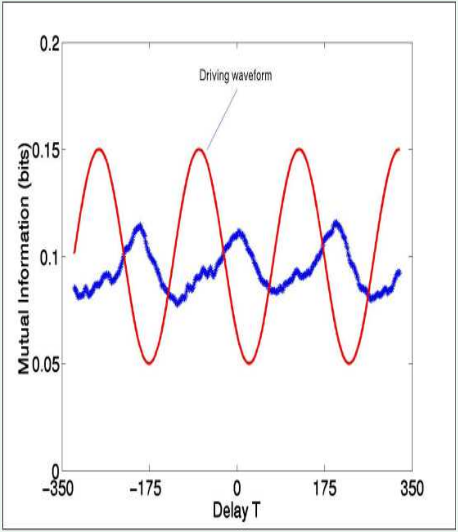
- location(x) varied between 0.7 and 0.94, target x=0.96
- as distance between data goes up, peak shifts to right (positive lag)
- information moving from high resource $\rightarrow$ low resource
- identifies critical distances for interactions ($\Delta x > 0.25$ have low mutual information exchange)

information exchange or environmental driver?



- remove dispersal ($D = 0$) - compute mutual information
- expect no strong peaks in MI in absence of information transport
- small peaks expected due to natural fluctuations as time series go in and out of phase as function of time lag

information exchange or environmental driver?



- resource abundance modeled as periodic function - no diffusion ($D = 0$)
- simulates environmental driver that can synchronize dynamics
- expect greater peaks in MI than with no periodic driver (Moran effect), yet no clear maximum because no information transport

numerical study conclusions based on mutual $I(Y, Z(T))$

- information flow for prey populations goes from high-resource to low-resource locations
- $I(Y, Z_T)$ maxima occur at small lags (T) for nearby locations and at larger lags as distance increases
- Remove dispersal and obtain no clear maximum
- Remove dispersal and add periodic driver: obtain peaks in $I(Y, Z_T)$ but again no clear maximum
- The $I(Y, Z_T)$ discriminates between information transport (dispersal) and a common environmental driver (Moran effect) for this system

time-lagged mutual information

- an *ad hoc* approach to inferences about information flow

$$I(Y,Z) = \sum_{y,z} p(y_i, z_i) \log_2 \frac{p(y_i, z_i)}{p(y_i)p(z_i)}$$

transfer entropy (Schreiber 2000)

- a formal approach that measures the degree and direction of dependence of one system variable on another

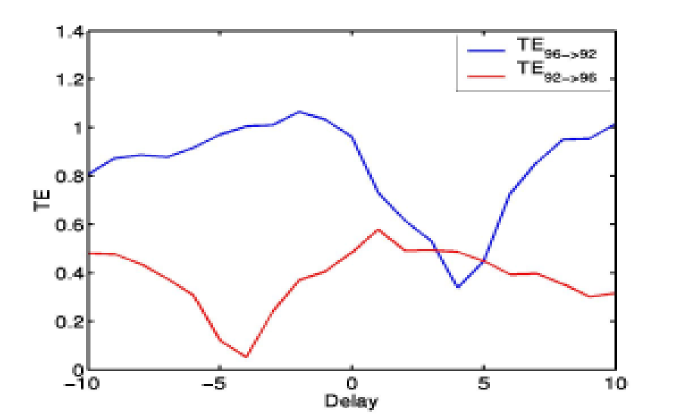
$$T_{Z\to Y} = \sum_{y,z} p\left(y_{t+1}, y_t^{(k)}, z_t^{(l)}\right) \log_2 \frac{p\left(y_{t+1}|y_t^{(k)}, z_t^{(l)}\right)}{p\left(y_{t+1}|y_t^{(k)}\right)}$$

Transfer entropy - short form...

- Consider a Markov process in which value of random variable, Y, at any time depends on past values (k time units into the past)
- Consider another possible system variable, Z, and ask whether it is related to (contributes information about) Y
- $T_{Z \to Y}$, measures the degree of dependence of Y on Z

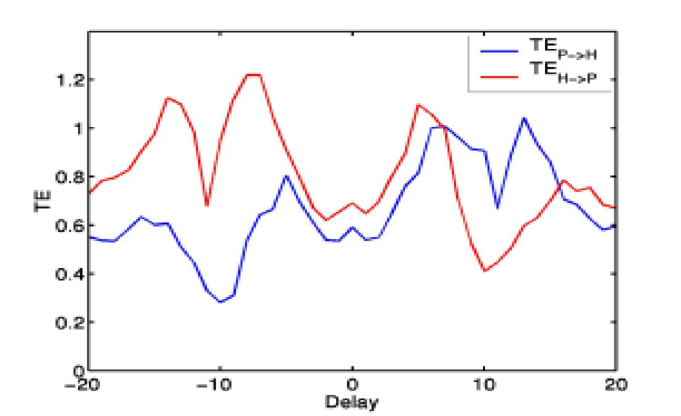
$$T_{Z \to Y} = \sum_{yz} p\left(y_{t+1}, y_t^{(k)}, z_t^{(l)}\right) \log \left(\frac{p(y_{t+1} | y_t^{(k)}, z_t^{(l)})}{p(y_{t+1} | y_t^{(k)})} \right)$$

Pascual model: prey abundance results



prey dynamics observed at $x = 0.96$ carry more additional information about site $x = 0.92$ than vice-versa

Pascual model: predator-prey information exchange



predator dynamics carry more additional information than do the prey dynamics (indicator species?)

Example 2: reconstructing a ’food web’

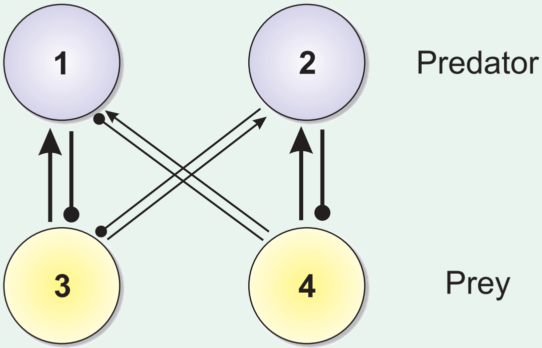
$$\frac{\partial n_1}{dt} = r_1 z_1 n_1 (1 - 0.1 n_1) - \alpha_{1,3} n_3 n_1 - \alpha_{1,4} n_4 n_1$$

$$\frac{\partial n_2}{dt} = r_2 z_2 n_2 (1 - 0.1 n_2) - \alpha_{2,3} n_3 n_2 - \alpha_{2,4} n_4 n_2$$

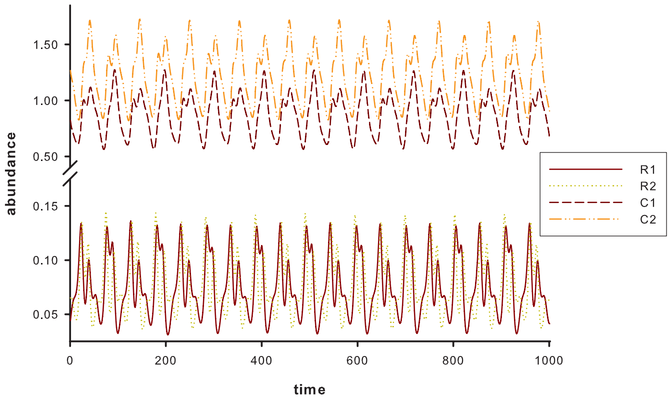
$$\frac{\partial n_3}{dt} = \alpha_{3,1} n_3 n_1 + \alpha_{3,2} n_3 n_2 - m n_3$$

$$\frac{\partial n_4}{dt} = \alpha_{4,1} n_4 n_1 + \alpha_{4,2} n_4 n_2 - m n_4$$

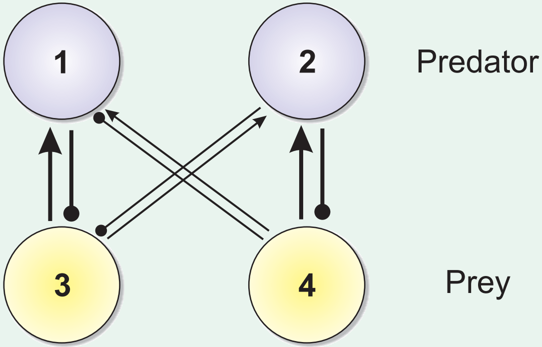
true food web



dynamics over time...



reconstructed food web (fully non-parametric)



- **surveillance monitoring programs**
    - want to infer stuff about nature of system and system change
    - **problem**: can’t measure all state variables in all places
- **indicator species**
    - lots of ‘arm-wavy’ definitions - most not based on any rigorous criterion...
    - proposed operational definition - species such that a time series of abundances (or whatever) provides more information about dynamics of overall system, or of a defined subset of the system, than that of any other species

proposed framework

- many of these methods not yet ready for ecological prime-time (clearly)
- approaches to nonlinear analysis of time series that are noisy, non-stationary and short include:
    - surrogate data sets for bootstrap-type approach to inference kernel density estimation approaches instead of “bin counting"
    - use of symbolic dynamics
    - information-based approaches for deterministic signal extraction in the presence of noise
- larger issue: retrospective versus prospective

going forward: ‘learning’

- methods (as described) based on *retrospective* analysis of exisiting time-series
- what about methods which ‘learn’ going ‘forward’ in time?
- appropriate for systems without long existing time-series of data?
- opportunities for ‘optimal learning’ about high-dimensional ‘networks’?
- do they work on the ‘real’ (ecological) world?

'similar' problem (perhaps...) – optimal sensors

- number of possible sensors $<$ number of possible sensor locations
- set V – all network associations/junctions (species interactions) – assume known (important)
- population model predicts relative degree of impact on system following perturbation
- challenge is to place sensors on this landscape (set of locations A) to minimize impact
- for each subset $A \subseteq V$ compute "sensing quality" $F(A)$
- $\max_{A \subseteq V} F(A)$, subject to $C(A) \leq B$

some basic results (Guestrin et al.)

- placement $A = \{S_1, S_2\}$, $B = \{S_1, S_2, S_3, S_4\}$
- add new sensor S' – helps more to add to A than to add to B
- i.e., for $A \subseteq B$, $F(A \cup \{S'\}) - F(A) \geq F(B \cup \{S'\}) - F(B)$
- key property – diminishing returns (*submodular*)

submodularity – 'very useful'

- want $A^* \subseteq V$ such that $A^* = \arg\max_{|A| \leq k} F(A)$ for k sensors
- typically NP-hard
- for submodular, greedy algorithm near-optimal – Nemhauser *etal*. (1978) – constant factor approximation $(F(A_{\text{greedy}}) \geq (1 - 1/e)F(A_{\text{opt}})$
- near-optimal (guarantees best unless $P = NP$)

problems in 'the real world'

- doesn't scale well
- SATURATE algorithm has very good performance but...
- ...success/performance dependent on known structure 'allowable' locations
- what about systems with a few/many hidden states (analogous to optimal salesman problem where not all possible 'bridges/barriers' are known
- can we place sensors in such a way so as to learn about the system in an optimal way (tradeoff between placement of fixed number of sensors with addition of more sensors)?

summary

- lot's of 'intriguing' tools from non-linear dynamics – many computational challenges (e.e.g, optimal banning algorithms for estimating mutual information)
- Takens' theorem allows for reconstruction – are all variables equally 'useful' in the reconstruction? Is there an optimal set of variables to be monitored?
- prospective – if 'placing sensors' is analogous to 'picking key species to monitor', how do we handle complexities of 'ecology?
- are all such problems submodular (with their nice 'properties'), or is that a 'fortunate' outcome of the 'sensor' problems that have been considered to date?
- Thanks for listening – and please 'come over and play' (translation: we need your help...).