Estimating Interaction Strengths in Ecological Communities
Intro to MAR modeling

Eli Holmes

FISH 507 – Applied Time Series Analysis

5 March 2019

Topics

➢Univariate and multivariate Gompertz models
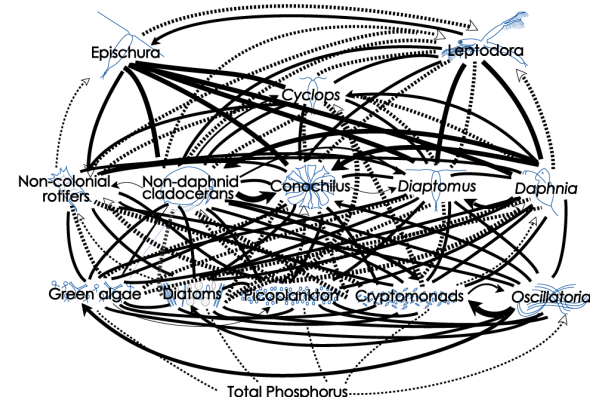
➢Example study: CA kelp forest ecosystem
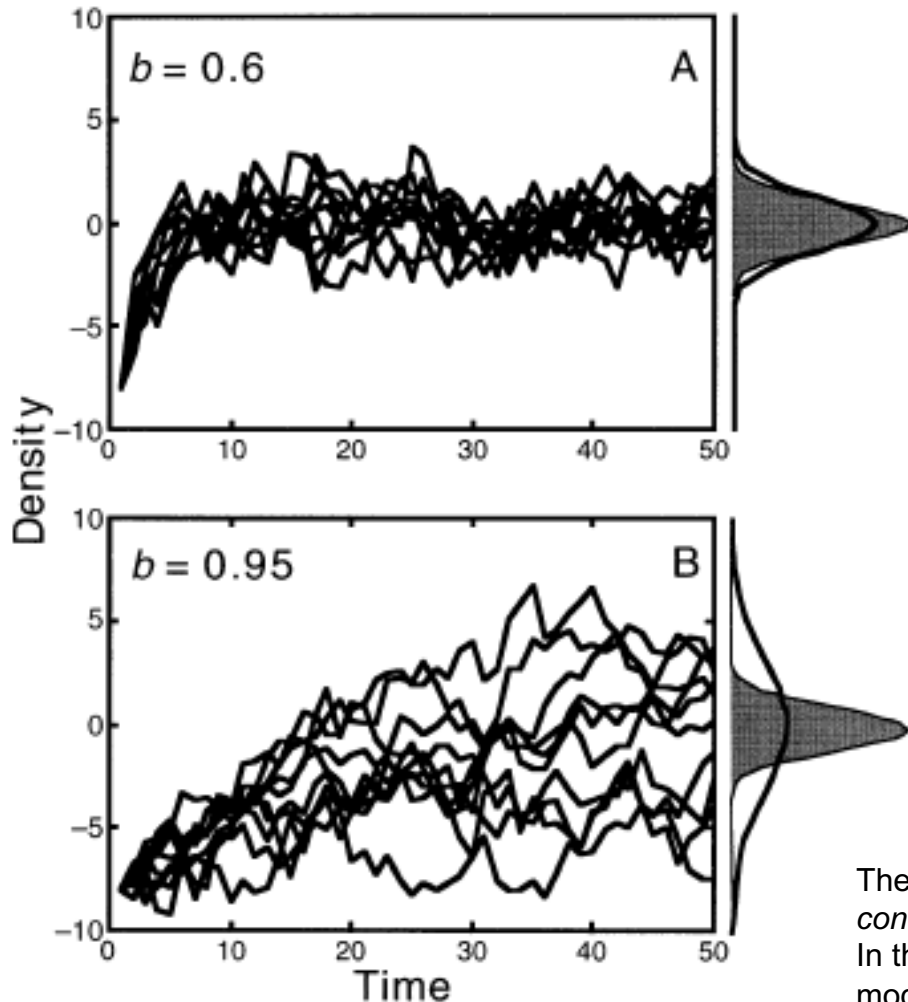
Simple 2-spp model

➢2x2 B matrix

Large multivariate models

➢Big B matrices

Mean-reverting processes

A figure showing two panels (A and B) of Density versus Time with overlaid stochastic trajectories. Panel A is labeled $b = 0.6$; Panel B is labeled $b = 0.95$. Density axis ranges from -10 to 10, Time axis from 0 to 50.

In lecture, I will talk about estimating mean-reversion in the context of density-dependence and species interactions, but mean-reverting stochastic processes are ubiquitous.

The **Ornstein-Uhlenbeck** process is the classic *continuous* time mean-reverting stochastic process. In the population dynamics literature, the **Gompertz** model is the classic discrete time mean-reverting process (although the Gompertz model also refers to a continuous time version).

Univariate Gompertz models

$$x_t = bx_{t-1} + u + w_t$$

Density dependence

univariate discrete exponential growth not in log-space

$$n_t = n_{t-1} \cdot \exp(u)$$

univariate discrete density-dependent growth not in log-space

general

Specific type

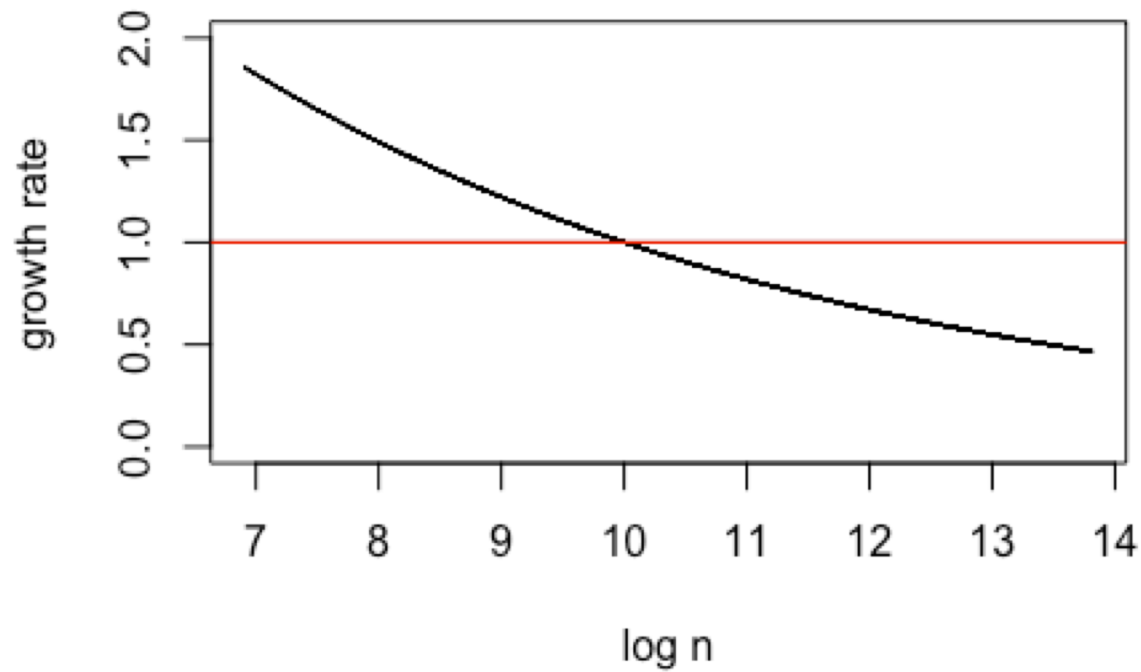$$n_t = n_{t-1} \cdot f(n_{t-1}) \qquad n_t = n_{t-1} \cdot \exp(u + f(n_{t-1}))$$

The shape of f(n_t-1) determines the dynamics of the system:
- stable or unstable equilibrium
- Speed at which equilibrium is approached
- Equilibrium level
- Sensitivity to perturbations

$$n_t = n_{t-1} \cdot \exp[u + (b-1)\ln n_{t-1}]$$

Gompertz model written in log space is AR(1)

$$n_t = n_{t-1} \cdot \exp\left[u + (b-1)\ln n_{t-1}\right]$$
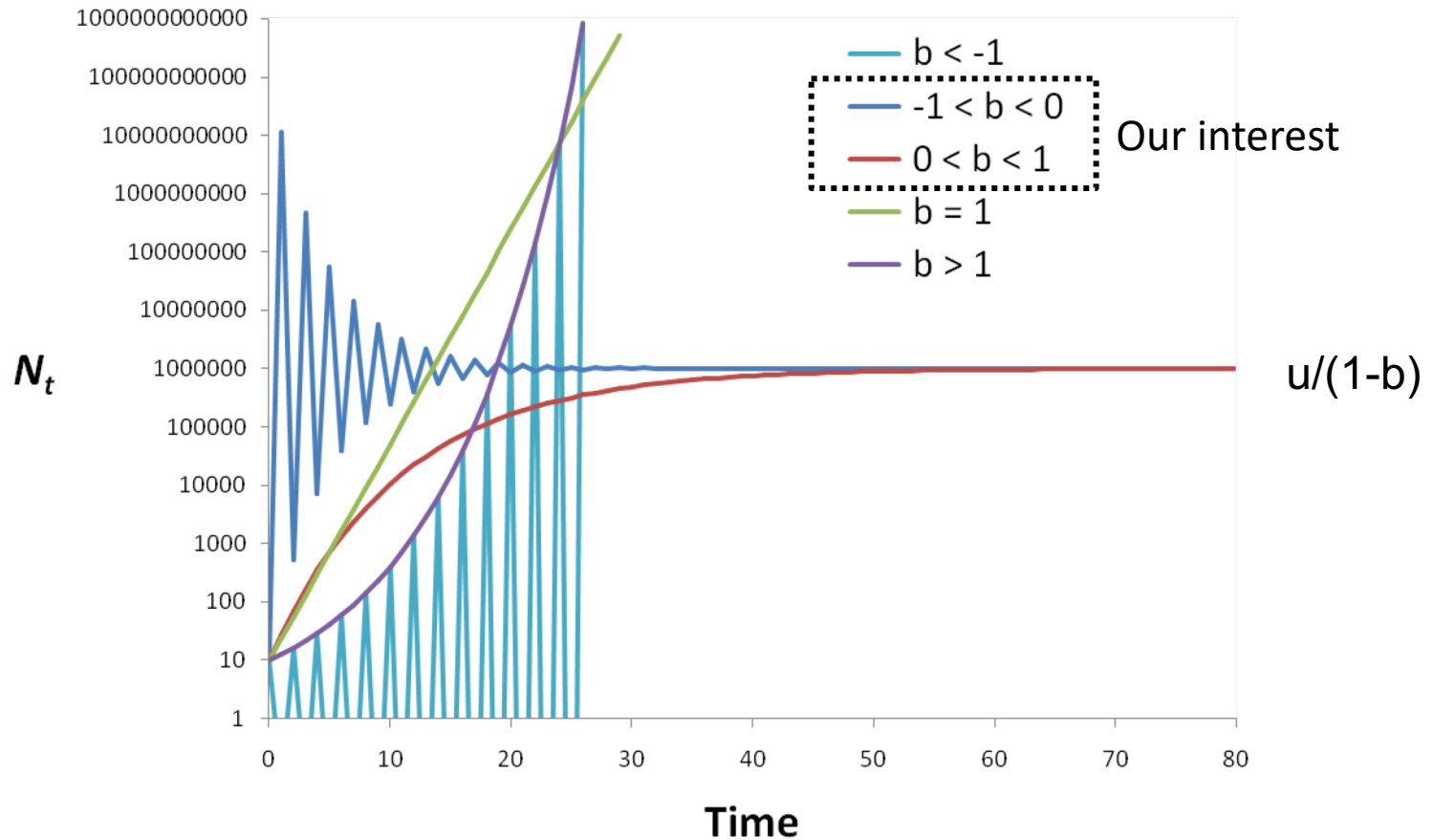
Taking the natural log of both sides

$$\ln n_t = \ln n_{t-1} + u + (b-1)\ln n_{t-1}$$

$$= \ln n_{t-1} + u + b\ln n_{t-1} - \ln n_{t-1}$$

$$= u + b\ln n_{t-1}$$

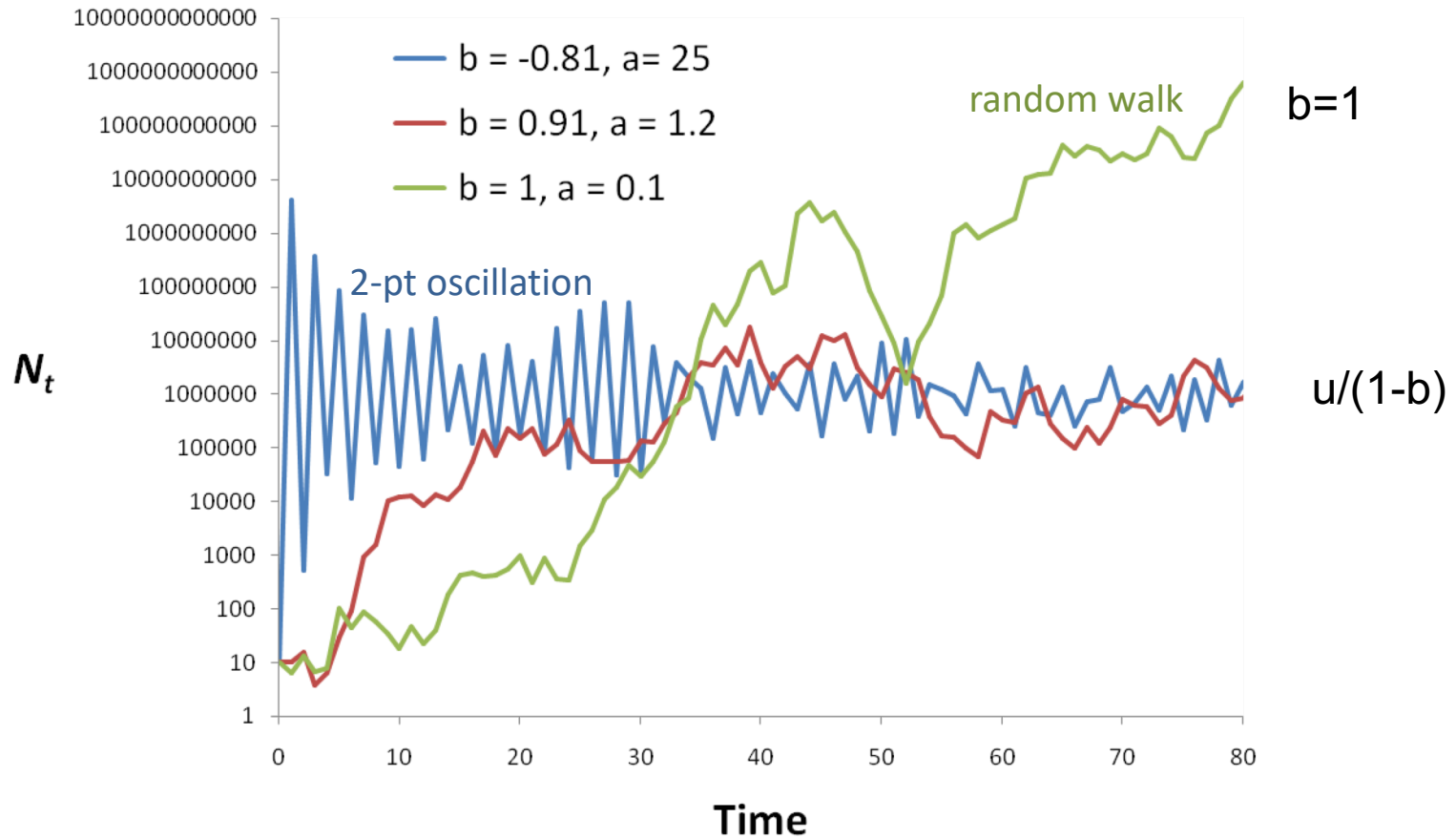Substituting x_t for $\ln n_t$
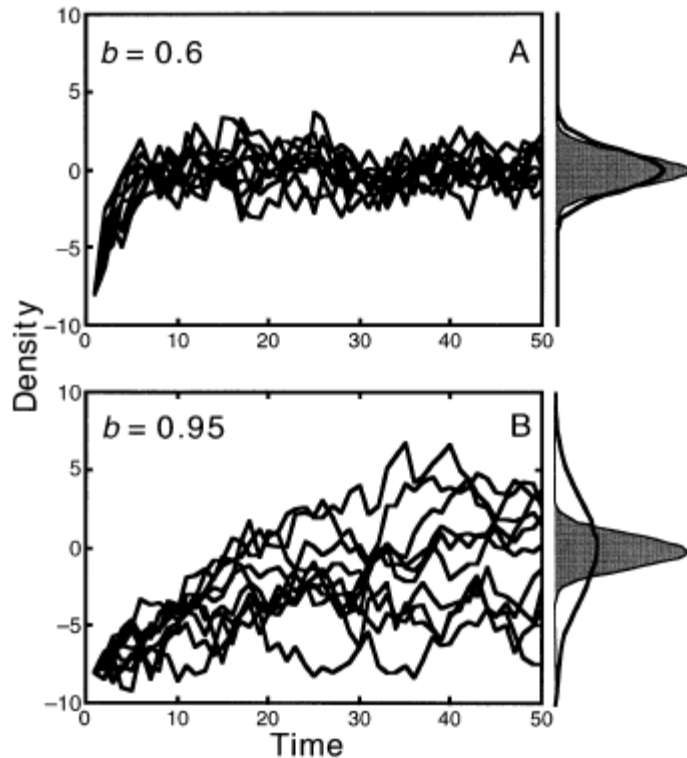
$$\boxed{x_t = u + bx_{t-1}}$$

AR(1) minus the noise term

Properties of this model

Add stochasticity (process error)

Equilibrium for the stochastic Gompertz process

Fig. 1 - Ives et al. (2003)

It has a stationary distribution probability distribution of X_t as $t \to \infty$ given $|b| < 1$

Normally distributed with mean μ_∞ and variance υ_∞

Main properties

• Mean reverting, aka density-dependent

• Stationary, so it fluctuates around a mean

• Point equilibrium as opposed to a cycle equilibrium like Lotka-Volterra (Lynx & hare) models you studied in Ecology 101

• Can be seen as a **locally linear** approximation of other types of density-dependent interaction models

"locally linear" is jargon for "only holds for sure if x doesn't change too much". In our case, x = log(n) = log abundance.

Real systems do not necessarily have log-linear density-dependence

Note: If N is not spanning very large values
a linear model may suffice

Nonlinear

$\log N_{t+1}$

$\log N_t$

Linear approx.

$\log N_{t+1}$

$\log N_t$
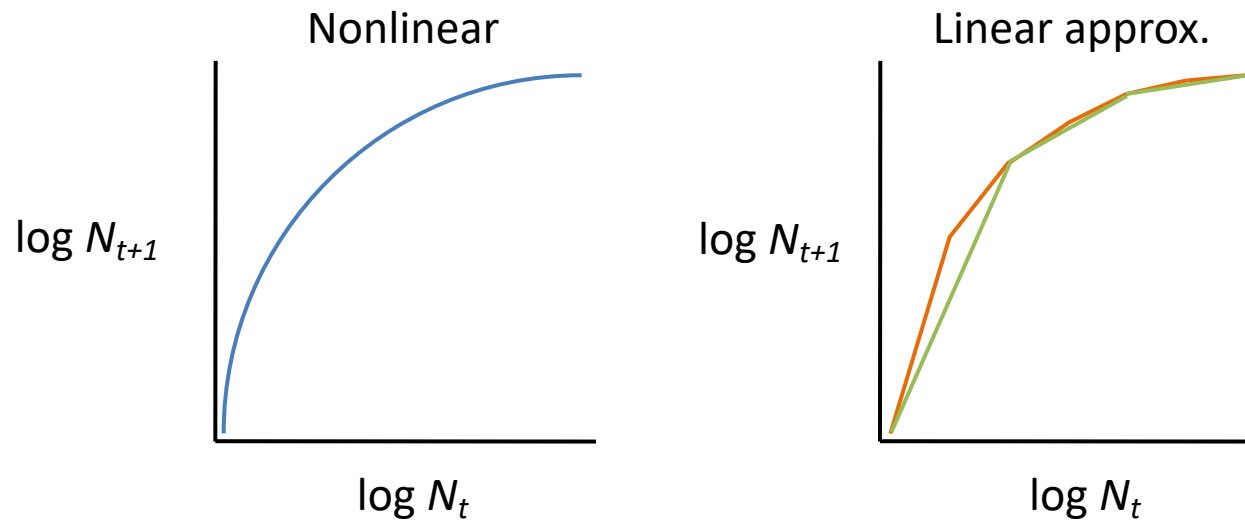
Modern literature on MAR(1) models allow B to be time-varying which allows that linear approximation to vary in time.

observation error is known a problem
obs error = spurious density-dependence

LETTER

Are patterns of density dependence in the Global Population Dynamics Database driven by uncertainty about population abundance?

Jonas Knape* and Perry de Valpine
Department of Environmental Science, Policy and Management, 137 Mulford Hall #3114, University of California, Berkeley, Berkeley, CA 94720, USA
*Correspondence: E-mail: jknape@berkeley.edu

Abstract
Density dependence in population growth rates is of immense importance to ecological theory and application, but is difficult to estimate. The Global Population Dynamics Database (GPDD), one of the largest collections of population time series available, has been extensively used to study cross-taxa patterns in density dependence. A major difficulty with assessing density dependence from time series is that uncertainty in population abundance estimates can cause strong bias in both tests and estimates of strength. We analyse 627 data sets in the GPDD using Gompertz population models and account for uncertainty via the Kalman filter. Results suggest that at least 45% of the time series display density dependence, but that it is weak and difficult to detect for a large fraction. When uncertainty is ignored, magnitude of and evidence for density dependence is strong, illustrating that uncertainty in abundance estimates qualitatively changes conclusions about density dependence drawn from the GPDD.

Keywords
Density dependence, GPDD, observation error, time series.

Ecology Letters (2011)

INTRODUCTION

Density dependence in population growth rates is a fundamental concept for ecological theory as well as for population management. Estimating density dependence in wild populations has, however, proved challenging. Ideally, density dependence in growth rates should be estimated directly from the effects of density acting on the traits contributing to population growth. Given current progress in statistical methods for jointly analysing data on both population size and demographic traits (Besbeas et al. 2005), and with long-term population studies involving demographic data becoming increasingly common, this approach holds a bright future. However, the number of such studies is currently limited and they only cover a rather narrow range of taxa. Long-term time series on population abundance are more common and can be used to estimate density dependence in population growth rates. Under this approach, density dependence is defined as a general tendency of per capita growth rates to decrease when population size is large and increase when it is small, and is identified as a statistical pattern not tied to any specific biological mechanism (Wolda & Dennis 1993).

It was noted early that estimates and tests of density dependence based on regressing log transformed current observed population size, y_t, on previous log transformed observed population size, y_{t-1}, are sensitive to uncertainty in the observations (St-Amant 1970; Kuno 1971; Itô 1972; Slade 1977). Similar concerns were aired about estimates from fisheries models of stock-recruitment data (Ludwig & Walters 1981; Walters & Ludwig 1981). Uncertainty inflates the Type I error rate of tests for density dependence (Shenk et al. 1998) and tends to bias estimates towards stronger density dependence if dynamics are under-compensatory and towards weaker density dependence if dynamics are over-compensatory (Benson 1973). Bulmer (1975) devised two tests for density dependence taking the time series nature of the data into account. One of those was designed to be robust against uncertainty about population size and has been shown to perform better than density dependence tests ignoring uncertainty in estimates of population abundance (Shenk et al. 1998). Simple procedures to correct for effects of uncertainty such as the SIMEX method have been suggested (Solow 1998; Freckleton et al. 2006) but typically require that the variance of the uncertainty about population size is known. A more direct approach to account for uncertainty is provided by state space models, first used for modelling population dynamics in the fisheries literature (e.g. Mendelssohn 1988; Sullivan 1992). State space models in these cases consist of a model of a population dynamical process combined with a model of the uncertainty in the abundance estimates, sometimes termed observation, measurement or sampling error, and may be used to estimate the variance of this uncertainty as well as to filter out its effects (de Valpine & Hastings 2002; Calder et al. 2003; Buckland et al. 2004; Dennis et al. 2006). Estimates derived from state space models tend to have smaller bias than estimates ignoring uncertainty about population abundance, but can also have large variances (Knape 2008), and the statistical properties of even simple state space model estimators are not fully understood (Dennis et al. 2006; Lebreton 2009).

The Global Population Dynamics Database (GPDD), containing over 5000 time series on population abundances obtained from various forms of population surveys, has provided an opportunity for ecologists to explore population dynamical patterns over a wide range of taxa (Inchausti & Halley 2001). Analyses using data in the GPDD have focused on, e.g., extinction risks (Fagan et al. 2001; Inchausti & Halley 2003; Brook et al. 2006), population cycles (Kendall et al. 1998; Murdoch et al. 2002) and effects of weather (Knape & de Valpine 2011) but, arguably, the studies stirring the most attention as well as debate have addressed population regulation and density dependence. These have explored patterns in the shape of density dependence (Sibly et al. 2005; Polansky et al. 2009) and in the strength of regulation and density dependence (Brook & Bradshaw 2006; Sibly et al. 2007;

ESTIMABILITY OF DENSITY DEPENDENCE IN MODELS OF TIME SERIES DATA

JONAS KNAPE[1]

Department of Theoretical Ecology, Ecology Building, Lund University, SE-223 62, Lund, Sweden

Abstract. Estimation of density dependence from time series data on population abundance is hampered in the presence of observation or measurement errors. Fitting state-space models has been proposed as a solution that reduces the bias in estimates of density dependence caused by ignoring observation errors. While this is often true, I show that, for specific parameter values, there are identifiability issues in the linear state-space model when the strength of density dependence and the observation and process error variances are all unknown. Using simulation to explore properties of the estimators, I illustrate that, unless assumptions are imposed on the process or observation error variances, the variance of the estimator of density dependence varies critically with the strength of the density dependence. Under compensatory dynamics, the stronger the density dependence the more difficult it is to estimate in the presence of observation errors. The identifiability issues disappear when density dependence is estimated from the state-space model with the observation error variance known to the correct value. Direct estimates of observation variance in abundance censuses could therefore prove helpful in estimating density dependence but care needs to be taken to assess the uncertainty in variance estimates.

Key words: density dependence; state-space models; time series analysis.

INTRODUCTION

Density dependence can be loosely defined as a quantitative influence of population size on some life history or population trait of interest. The concept is of central importance to population ecology since it determines both the limiting and the short time behavior of the dynamics of populations. Empirical estimates of density dependence are therefore important from a scientific as well as from a management perspective. Assessment of density dependence in the dynamics of natural populations has however proved to be challenging (Dennis et al. 2006).

When relevant data are available, effects of density dependence can be directly linked to life history traits. For instance, density dependence in recruitment (e.g., Crespin et al. 2006) and survival (e.g., Festa-Bianchet et al. 2003) have been estimated by mark-recapture analyses and density dependence in fecundity has been inferred from data on reproduction (e.g., Solbreck and Ives 2007). Density dependence in life history traits influences density dependence in population growth rate (Lande et al. 2002). It can be argued that density dependence in population growth is the most important form of density dependence for determining long-term behavior of populations. However, since the link from demographic traits to population change is almost never known with good precision, density dependence in

population growth rate is not easily inferred from life history data even if the effects of density dependence on several life history traits are well known. Time series analysis of population abundance data provides an alternative or complementary method that ideally could serve as a more direct way of estimating density dependence in population growth rate.

Estimates of density dependence must rely on measures of population density that are usually difficult to obtain with precision (Freckleton et al. 2006). This problem is particularly relevant to estimates of density dependence in growth rate derived from time series data on population size in that both the dependent and the independent variable are measured with uncertainty. Introducing observation error to dynamical data changes its dynamical structure (Dennis et al. 2006) and estimators relating to the dynamics of the data that do not account for observation errors are therefore often biased. Specifically, tests and estimators of density dependence based on time series data are known to be biased if observation errors are present but ignored for both direct (Kuno 1971, Walters and Ludwig 1981, Shenk et al. 1998, Freckleton et al. 2006) and delayed (Solow 2001) density dependence. An appealing method for overcoming this difficulty is provided by the state-space framework (Harvey 1990), a general term for statistical models of observations of hidden state variables that are dynamically linked through time. For time series data on population abundance, state-space models can be used for explicit modeling of both the observation and the population dynamical processes (Stenseth et al. 2003, Jamieson and Brooks 2004).

2010　　　　　　　　　　　　　2008

Estimating R matrix is not so easy, but replication helps A LOT

MARSS models (however you fit them) allow you to easily incorporate replication.

ESTIMATING DENSITY DEPENDENCE, PROCESS NOISE, AND OBSERVATION ERROR

BRIAN DENNIS,[1,5] JOSÉ MIGUEL PONCIANO,[2] SUBHASH R. LELE,[3] MARK L. TAPER,[4] AND DAVID F. STAPLES[4]

[1]Department of Fish and Wildlife Resources and Department of Statistics, University of Idaho, Moscow, Idaho 83844 USA
[2]Initiative for Bioinformatics and Evolutionary Studies (IBEST), Department of Mathematics, University of Idaho,

ESTIMABILITY OF DENSITY DEPENDENCE IN MODELS OF TIME SERIES DATA

JONAS KNAPE[1]

Department of Theoretical Ecology, Ecology Building, Lund University, SE-223 62, Lund, Sweden

*Ab
abunc
state—
densit
that,
mode

Reducing bias and improving precision in species extinction forecasts

KEVIN E. SEE[1,3] AND ELIZABETH E. HOLMES[2]

[1]Quantitative Ecology and Resource Management, University of Washington, Box 352182, Seattle, Washington 98195 USA
[2]Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration,
2725 Montlake Boulevard East, Seattle, Washington 98112 USA

Abstract. Forecasting the risk of population decline is crucial in the realm of biological conservation and figures prominently in population viability analyses (PVA). A common form of available data for a PVA is population counts through time. Previous research has suggested that improving estimates of population trends and risk from count data depends on longer observation periods, but that is often impractical or undesirable. Making multiple observations within a single time step is an alternative way to gather more data without extending the observation period. In this paper, we examine the trade-off between the length of the time period over which observations of the population have been taken and the total number of observations or samples that have been recorded through an analysis of simulated

Predator-prey interactions

2-species: Predator-Prey

Moose $\qquad x_{m,t} = u_m + b_{m\to m}x_{m,t-1} + b_{w\to m}x_{w,t-1}$
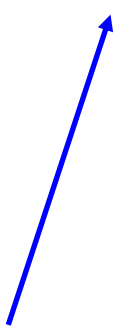
Wolf $\qquad x_{w,t} = u_w + b_{w\to w}x_{w,t-1} + b_{m\to w}x_{m,t-1}$

MAR(1): $x_t = Bx_{t-1} + u + w_t$

"spp" abundance

$$\begin{bmatrix} x_{m,t} \\ x_{w,t} \end{bmatrix} = \begin{bmatrix} b_{m \to m} & b_{w \to m} \\ b_{m \to w} & b_{w \to w} \end{bmatrix} \begin{bmatrix} x_{m,t-1} \\ x_{w,t-1} \end{bmatrix} + \begin{bmatrix} u_m \\ u_w \end{bmatrix} + \begin{bmatrix} w_m \\ w_w \end{bmatrix}$$

B = interaction matrix

Process variation
MVN(0,Q)

Adding covariates

$$\begin{bmatrix} x_{m,t} \\ x_{w,t} \end{bmatrix} = \begin{bmatrix} b_{m \to m} & b_{w \to m} \\ b_{m \to w} & b_{w \to w} \end{bmatrix} \begin{bmatrix} x_{m,t-1} \\ x_{w,t-1} \end{bmatrix} + \begin{bmatrix} u_m \\ u_w \end{bmatrix} + \begin{bmatrix} c_{1 \to m} & c_{2 \to m} & c_{3 \to m} \\ c_{1 \to 2} & c_{2 \to w} & c_{3 \to w} \end{bmatrix} \begin{bmatrix} c_{1,t} \\ c_{2,t} \\ c_{3,t} \end{bmatrix} + \begin{bmatrix} w_m \\ w_w \end{bmatrix}$$

covariates

B = interaction matrix

covariate effect

Process variation not from covariates ("unexplained")

Chapter 14: HWS18a Moose and Wolf example analysis
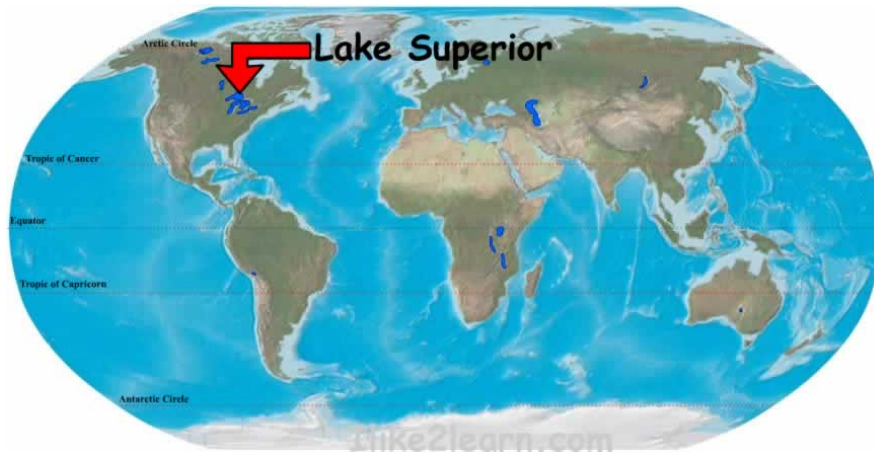
Effect of winter snow and summer heat on moose and wolves.

data and images from   www.isleroyalewolf.org
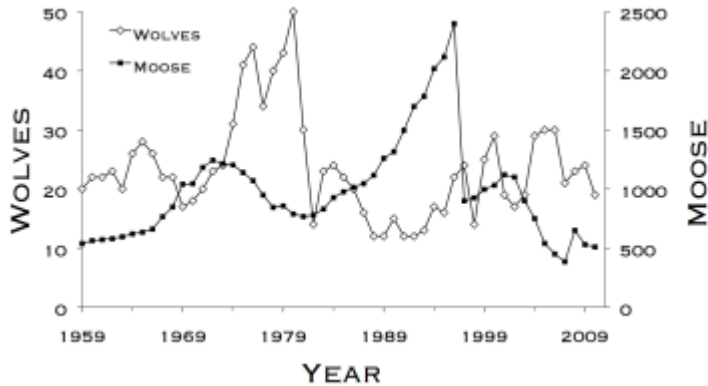
Using MAR models to understand dynamics in larger communities
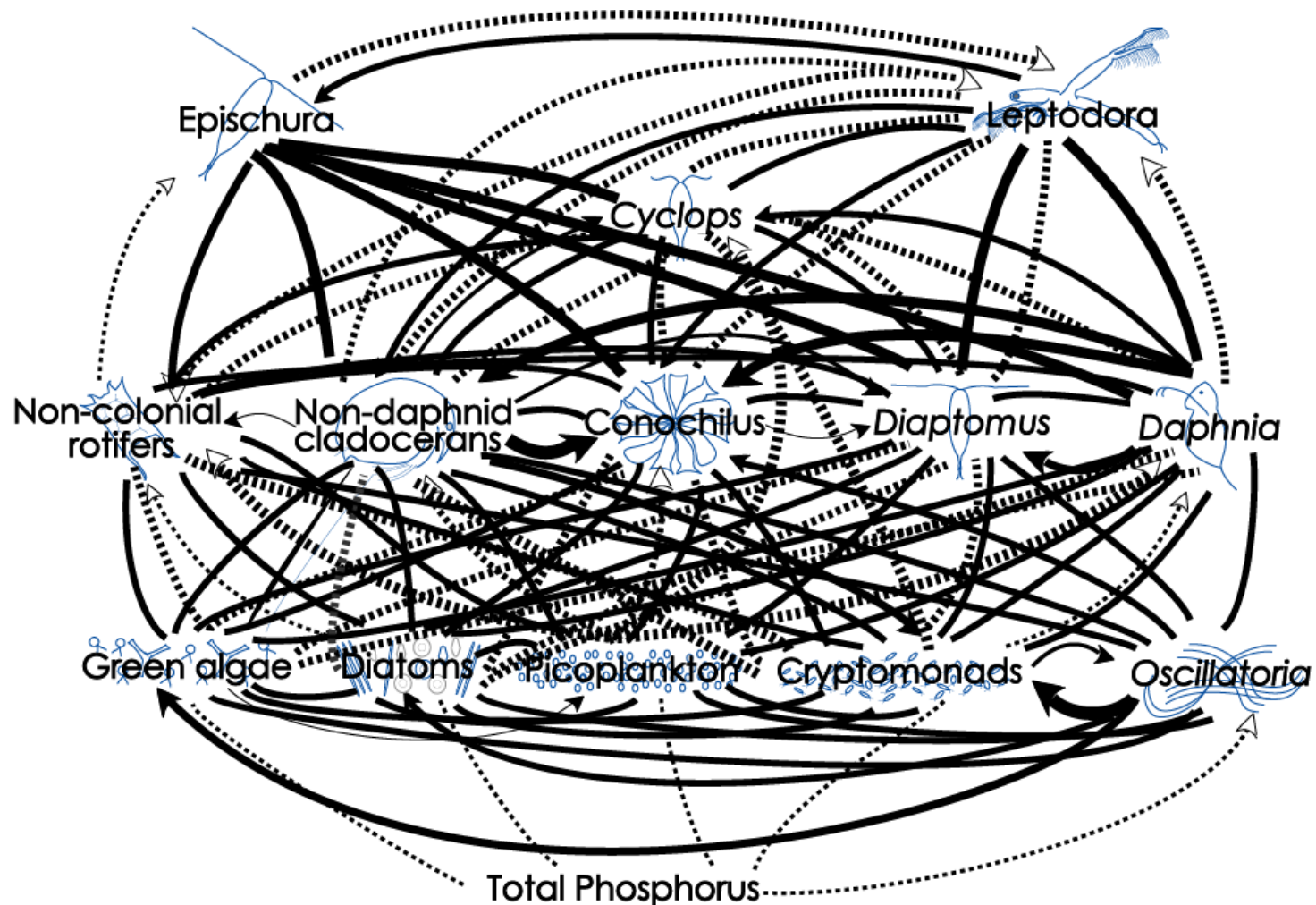
Dashed lines are (+) interactions; solid lines are (-)

Hampton et al. (2006)

Using MAR models to quantify stability
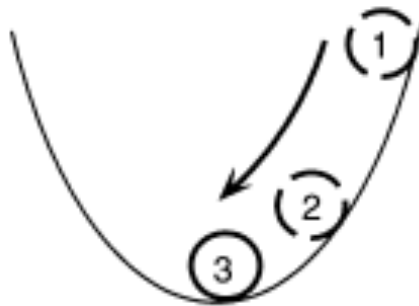
Using MAR models to identify the 'important' links
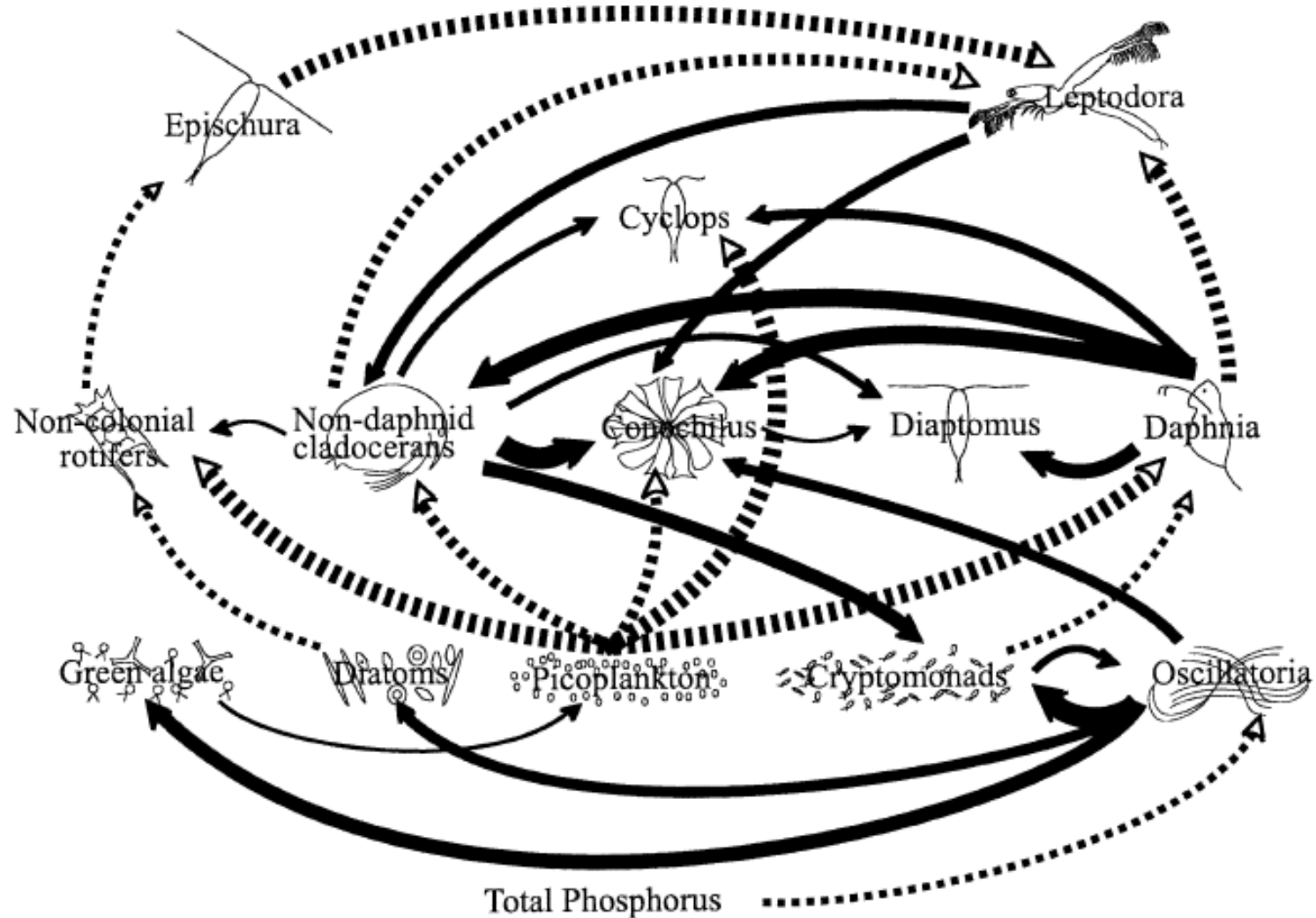
Dashed lines are (+) interactions; solid lines are (-)   Hampton et al. (2006)

MAR(1) Models

State equation

$$\mathbf{x}_t = \mathbf{a} + \mathbf{B}\mathbf{x}_{t-1} + \mathbf{C}\mathbf{c}_t + \mathbf{w}_t \qquad \mathbf{w}_t \sim \mathrm{MVN}\left(\mathbf{0}, \mathbf{Q}\right)$$

Observation equation
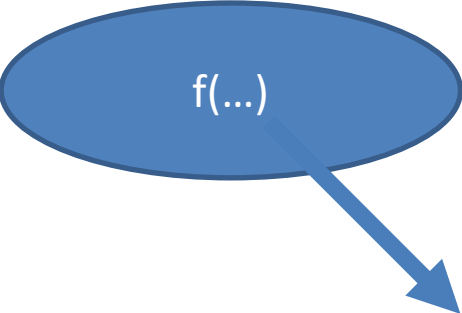
$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{v}_t \qquad \mathbf{v}_t \sim \mathrm{MVN}\left(\mathbf{0}, \mathbf{R}\right)$$

MAR(1) Models

State equation

$$\mathbf{x}_t = \mathbf{a} + \mathbf{B}\mathbf{x}_{t-1} + \mathbf{C}\mathbf{c}_t + \mathbf{w}_t \qquad \mathbf{w}_t \sim \mathrm{MVN}\big(\mathbf{0}, \mathbf{Q}\big)$$

Observation equation

$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{v}_t \qquad \mathbf{v}_t \sim$$

$f(...)$

Another observation error model

Multivariate AR(1) process

B matrix structure:

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & \cdots & \vdots \\ b_{31} & \vdots & b_{33} & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{p1} & \cdots & \cdots & \cdots & b_{pp} \end{bmatrix}$$

Inter-specific effects
(can be set to zero)

Inter-specific effects
(can be set to zero)

Intra-specific effects

MAR(1) models have been widely used to study dynamics of plankton communities
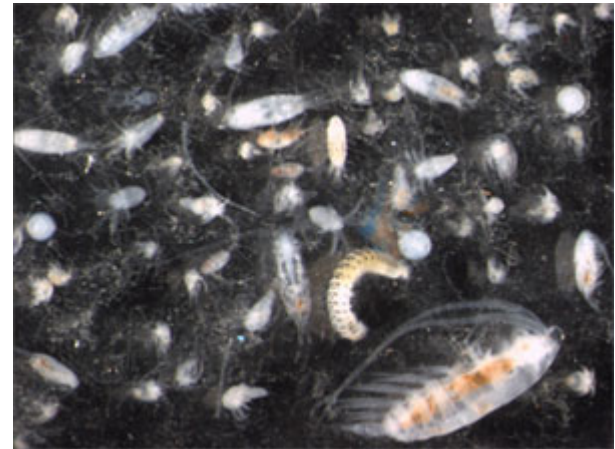
Many ecological applications are reviewed in

Hampton, S.E., E.E. Holmes, D.E. Pendleton, L.P. Scheef, M.D. Scheuerell, and E.J. Ward. 2013. Quantifying effects of abiotic and biotic drivers on community dynamics with multivariate autoregressive (MAR) models. *Ecology* 94:2663-2669

Lake Washington: a large change in sewage inputs in the late 1960s led to a dramatic change in the plankton community

Hampton et al. (2006)
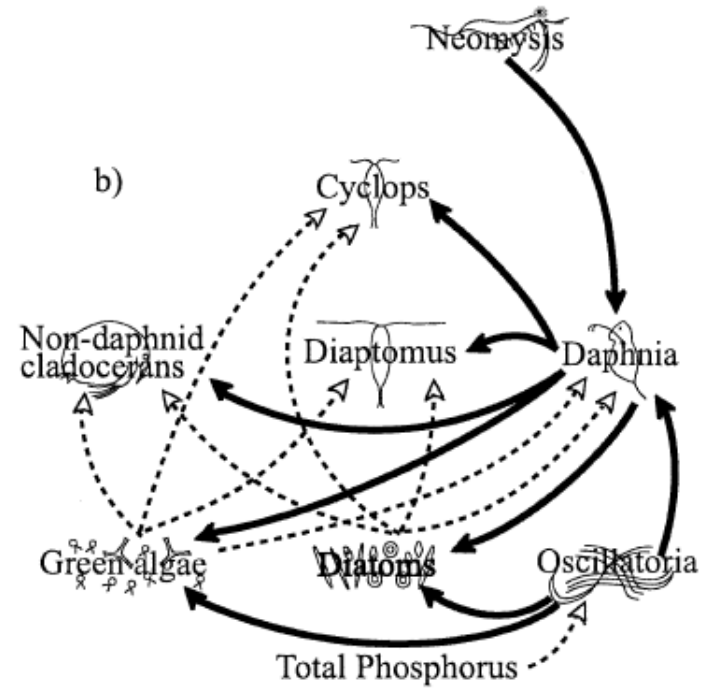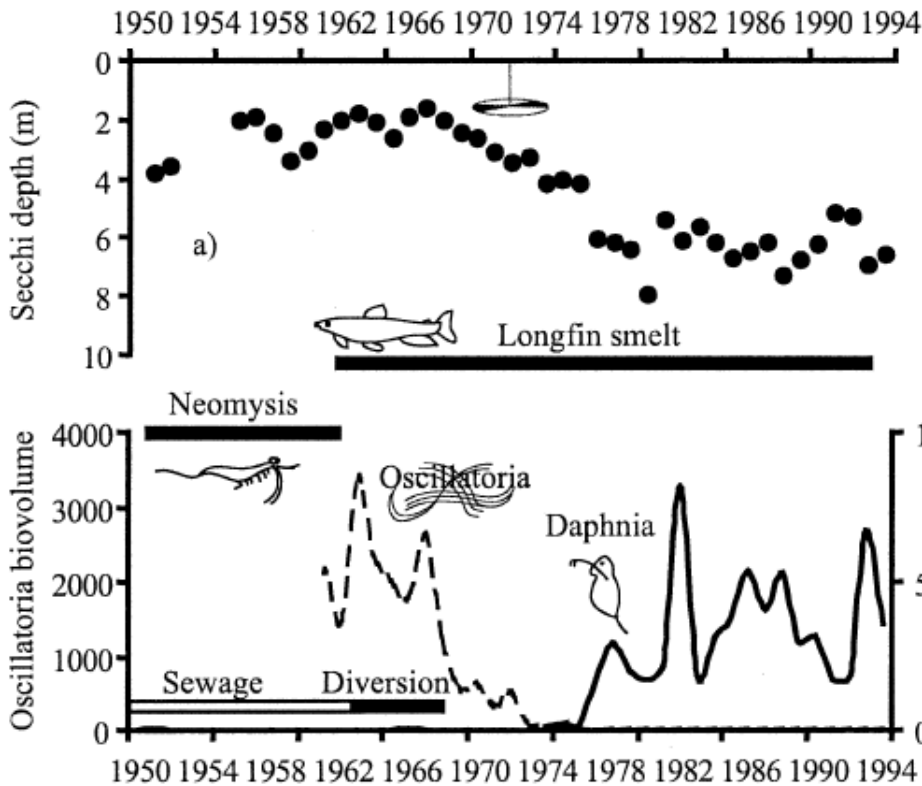
Multivariate AR(1) process, "MAR1"

Assume p different interacting species, then for each sp. i:

$$\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}_t = \begin{bmatrix} b_{11} & b_{21} & \dots & b_{n1} \\ b_{12} & b_{22} & \dots & b_{n2} \\ \dots & \dots & \dots & \dots \\ b_{1n} & b_{2n} & \dots & b_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}_{t-1} + \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{bmatrix}_t$$

B

$$\mathbf{w}_t \sim MVN(0, \mathbf{Q})$$

$X_{j,t-1}$ is abundance index of species j at time t-1

b_{ij} is effect of species j on species i

Note: when $i = j$, b_{ij} is effect of species i on itself

Multivariate AR(1) process

Can add effects of q different environmental covariates:

$$\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}_t = \mathbf{B}\mathbf{x}_{t-1} + \mathbf{u} + \begin{bmatrix} c_{11} & c_{21} \\ c_{12} & c_{22} \\ \dots & \dots \\ c_{1n} & c_{2n} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}_t + \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{bmatrix}_t$$

q=2

$c_{k,t}$ is covariate k (e.g., temperature) at time t

C_{ik} is effect of covariate k on species i

Multivariate AR(1) process

Can write all of this as:

$$\mathbf{x}_t = \mathbf{u} + \mathbf{B}\mathbf{x}_{t-1} + \mathbf{C}\mathbf{c}_t + \mathbf{w}_t \qquad \mathbf{w}_t \sim MVN\left(\mathbf{0}, \mathbf{Q}\right)$$

$\mathbf{x}_t$ is p x 1 vector of species abundance at time t

$\mathbf{u}$ is p x 1 vector of intrinsic growth rates

$\mathbf{B}$ is p x p matrix of density-dependent effects

$\mathbf{x}_{t-1}$ is p x 1 vector of species abundance at time t-1

$\mathbf{C}$ is p x q matrix of covariate effects

$\mathbf{c}_t$ is q x 1 vector of covariate values at time t

Estimating changes in species interactions & community stability in a kelp forest ecosystem

Mark Scheuerell *NOAA Northwest Fisheries Science Center*

Eric Ward *NOAA Northwest Fisheries Science Center*

Steve Katz *Washington State University*

Acknowledgments

- United States taxpayers

- Eli Holmes, Jim Thorson, Eric Buhle

- Lots of other students, post-docs & unsuspecting colleagues

Kelp forest ecosystems
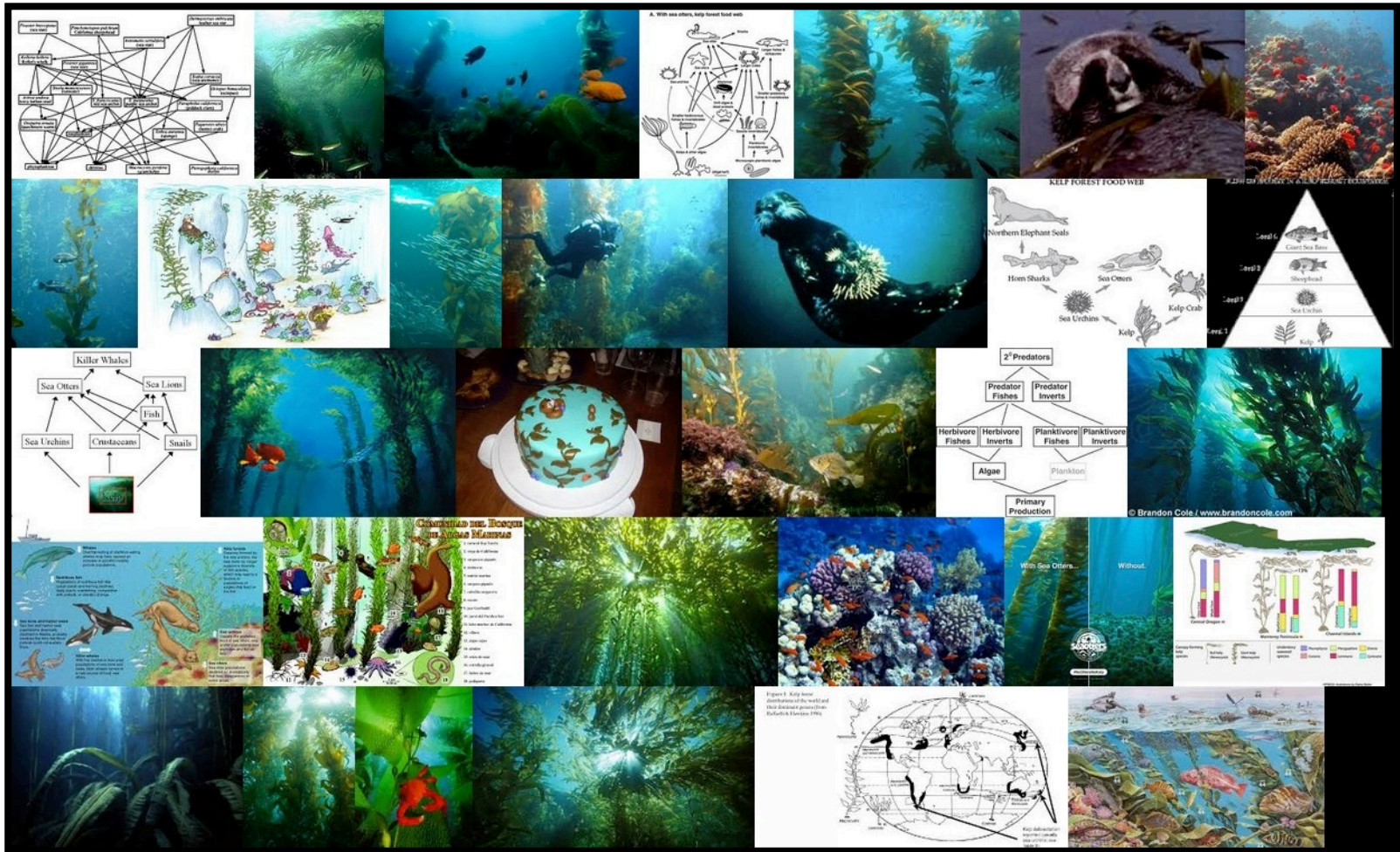
Images courtesy of Google & ImageQuilts

Conventional wisdom

Otters

Urchins

Kelps

Conventional wisdom

Otters

Urchins

Urchins

Kelps

Kelps

Images courtesy of R Schwemmer & NPS
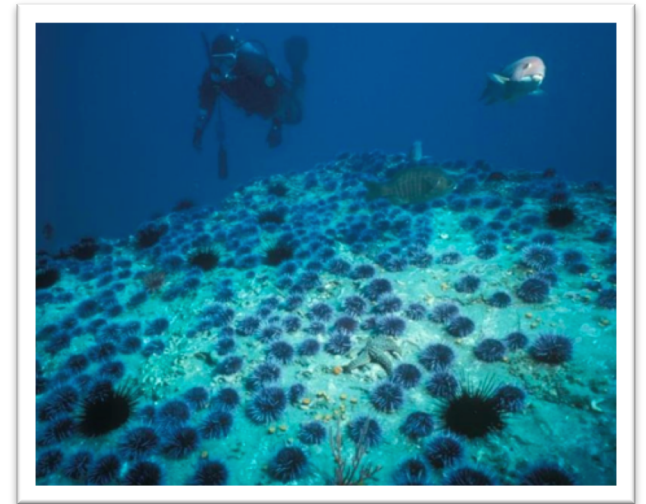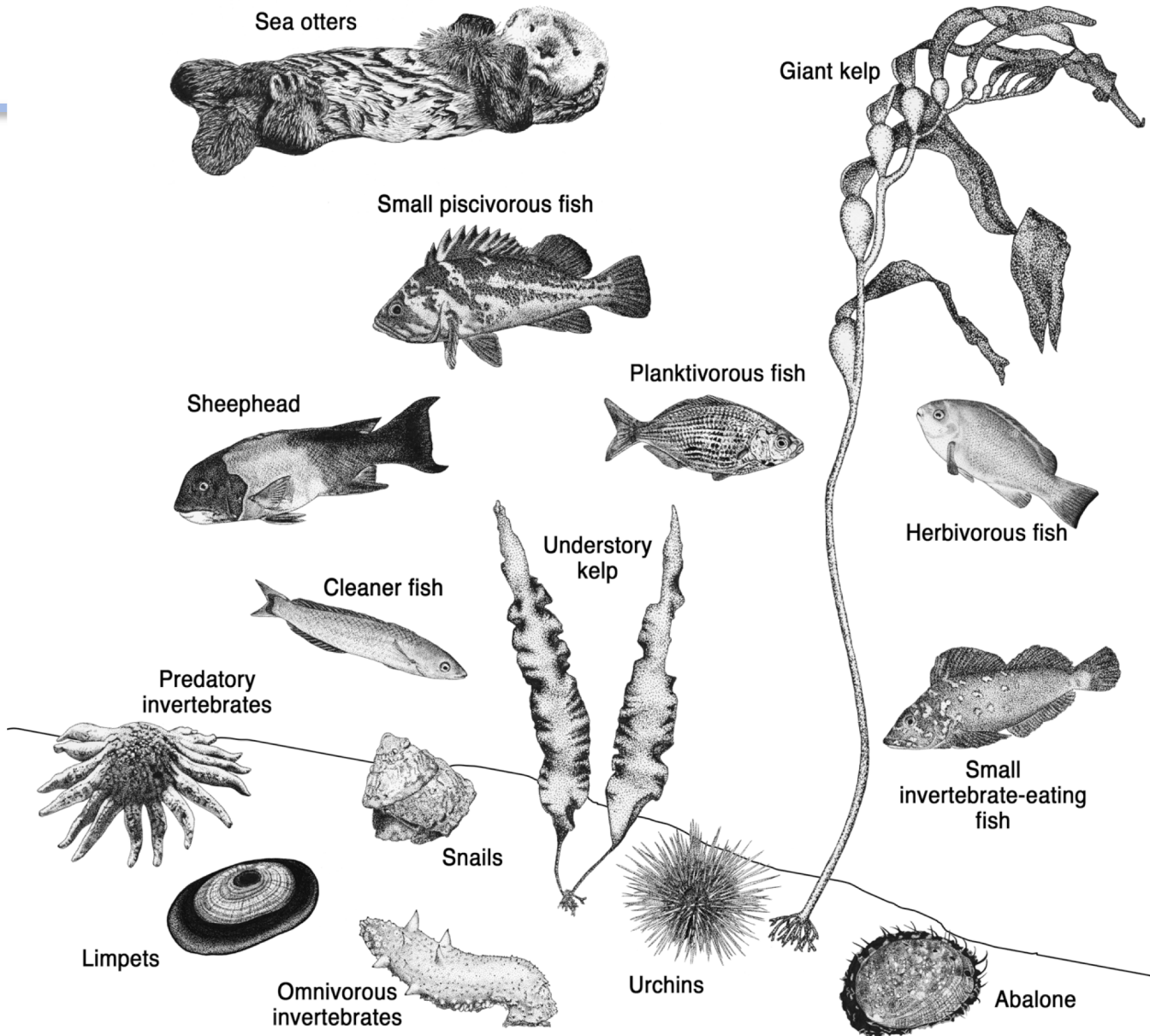
Sea otters

Small piscivorous fish

Giant kelp

Sheephead

Planktivorous fish

Herbivorous fish

Cleaner fish

Understory kelp

Predatory invertebrates

Small invertebrate-eating fish

Snails

Limpets

Omnivorous invertebrates

Urchins

Abalone

Artwork by Shannon Hennessey

Questions about the larger community

1) How do community interactions change over time?
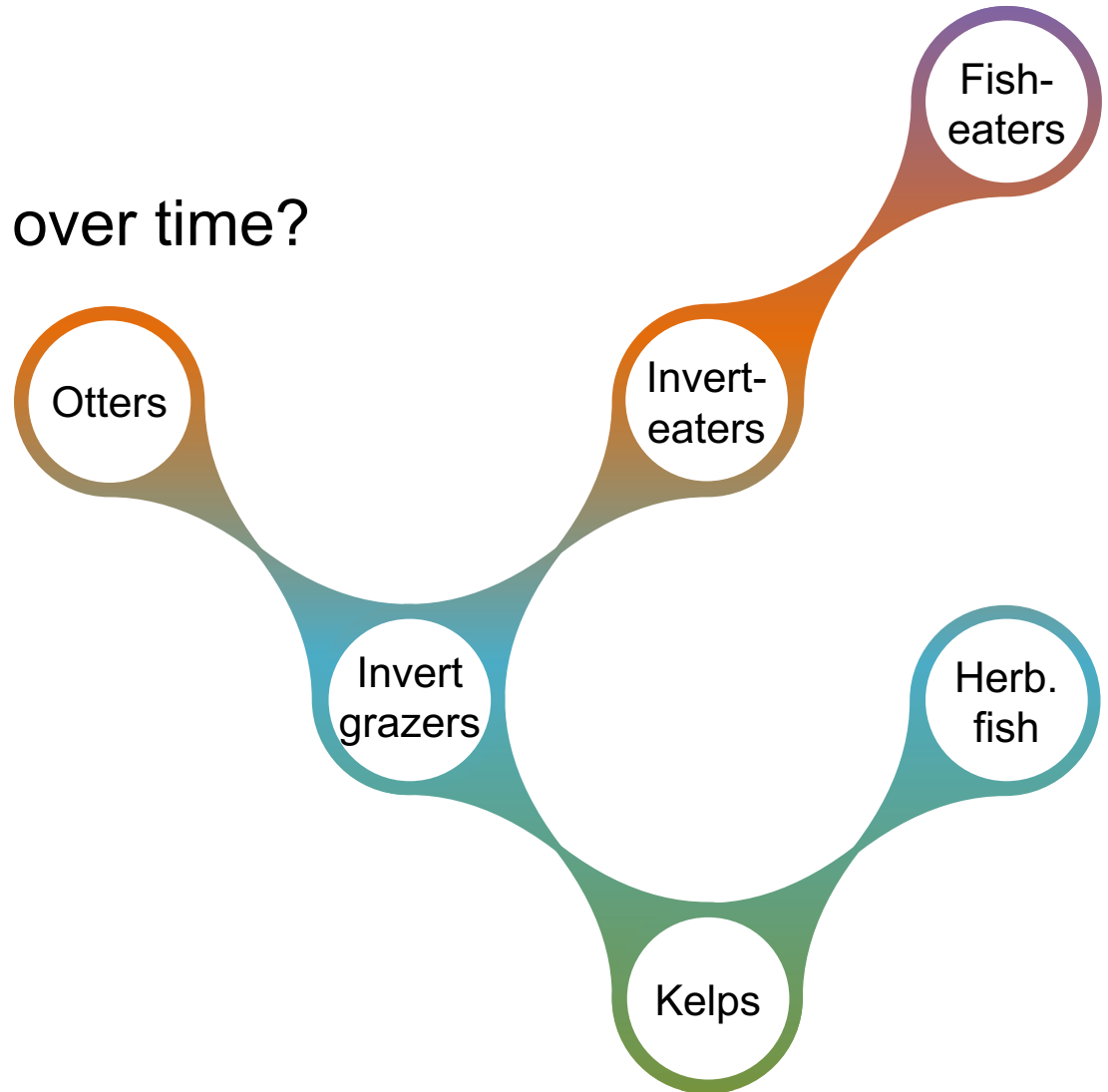
2) How do we tease out interaction changes in the face of environmental changes?

Community model in matrix form

Intrinsic
growth rates

Covariate
effects

log
densities

$$\mathbf{x}_t = \mathbf{a} + \mathbf{B}\mathbf{x}_{t-1} + \mathbf{C}\mathbf{c}_t + \mathbf{w}_t$$

Environmental
variation

Interactions
strengths

Covariates

Ives et al (2003) *Ecology*

Two issues with this model

$$\mathbf{x}_t = \mathbf{a} + \mathbf{B}\mathbf{x}_{t-1} + \mathbf{C}\mathbf{c}_t + \mathbf{w}_t$$

1) This describes a process with no sampling or observation error
2) The community interactions don't vary with time

State-space model

State equation

$$\mathbf{x}_t = \mathbf{a} + \mathbf{B}\mathbf{x}_{t-1} + \mathbf{C}\mathbf{c}_t + \mathbf{w}_t \qquad \mathbf{w}_t \sim \mathrm{MVN}\big(\mathbf{0}, \mathbf{Q}\big)$$

Observation equation

$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{v}_t \qquad \mathbf{v}_t \sim \mathrm{MVN}\big(\mathbf{0}, \mathbf{R}\big)$$

State-space model

State equation

$$\mathbf{x}_t = \mathbf{a} + \mathbf{B}\mathbf{x}_{t-1} + \mathbf{C}\mathbf{c}_t + \mathbf{w}_t \qquad \mathbf{w}_t \sim \mathrm{MVN}\big(\mathbf{0}, \mathbf{Q}\big)$$

Observation equation

$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{v}_t \qquad \mathbf{v}_t \sim \mathrm{MVN}\big(\mathbf{0}, \mathbf{R}\big)$$

Count data are the basis for many studies

log(0) = -Inf (!)

Allowing for count data

Observation of guild *g* at location *i* at time *t*

$$y_{g,i,t} \sim \text{NegBin}\left(p_{g,j,t}, r_g\right)$$

$$p_{g,j,t} = \frac{r_g}{r_g + \mu_{g,j,t}}$$

$$\log\left(\mu_{g,j,t}\right) = x_{g,j,t}$$

State

Time-varying interactions

State equation

$$\mathbf{x}_t = \mathbf{a} + \mathbf{B}_t \mathbf{x}_{t-1} + \mathbf{C}\mathbf{c}_t + \mathbf{w}_t$$

$$\text{vec}\left(\mathbf{B}_t\right) \sim \text{MVN}\left(\text{vec}\left(\mathbf{B}_{t-1}\right), \mathbf{S}\right)$$

Kelp forests in southern California

Images courtesy of Google
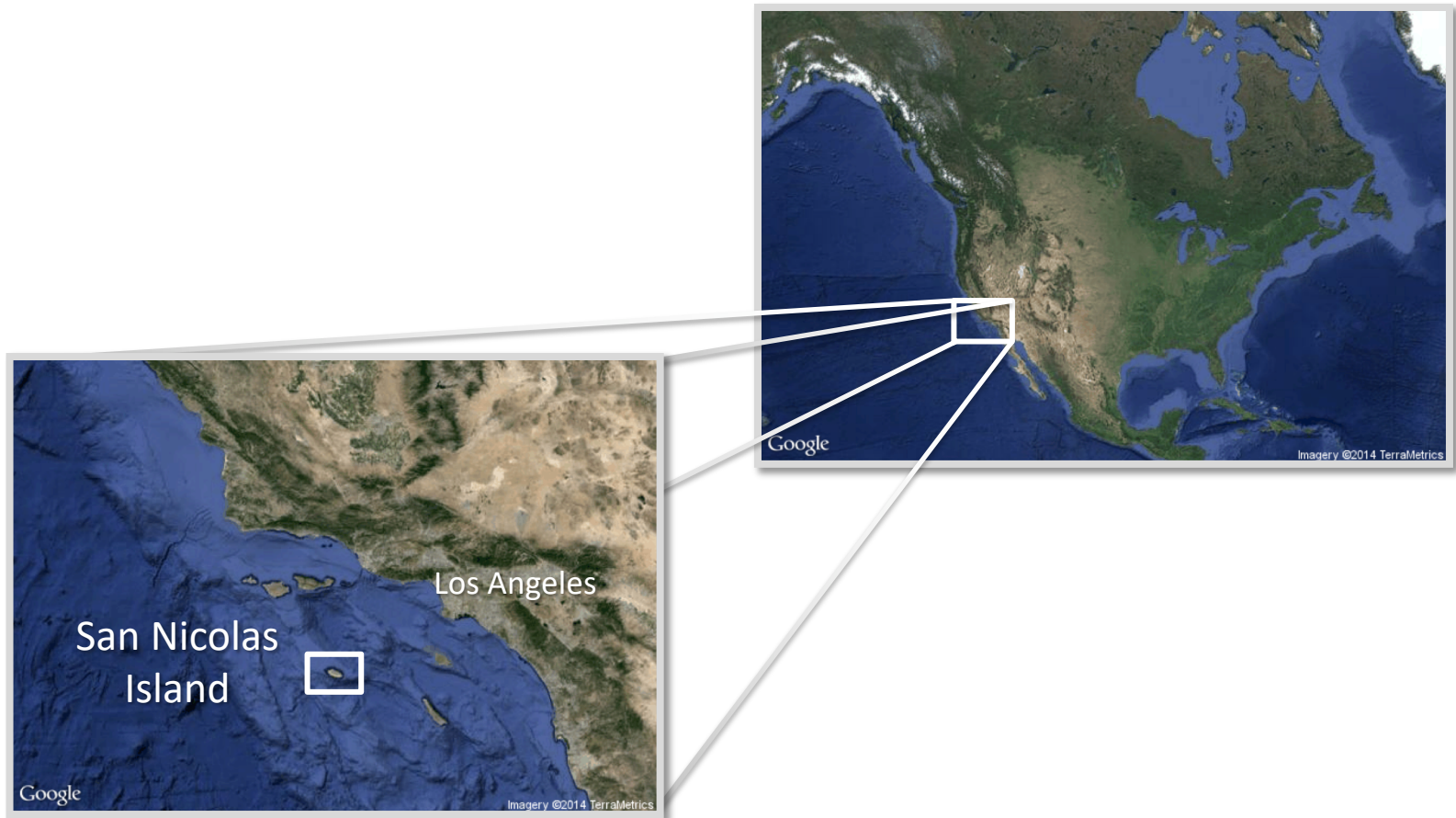
Data for San Nicolas Island

Kenner et al (2013) A multi-decade time series of kelp forest community structure at San Nicolas Island, California (USA). *Ecology* 94:2654

- 7 sites around the island

- Biannual surveys from 1980-2011 ($n = 63$)

- Divers collect data on:
  - Fish (59 spp)
  - Inverts (14 spp)
  - Kelps (6 spp)

Image courtesy of Google

Sea otters at San Nicolas

Otters absent for >100 yr prior to 1987

Image courtesy of Steve Lonhart
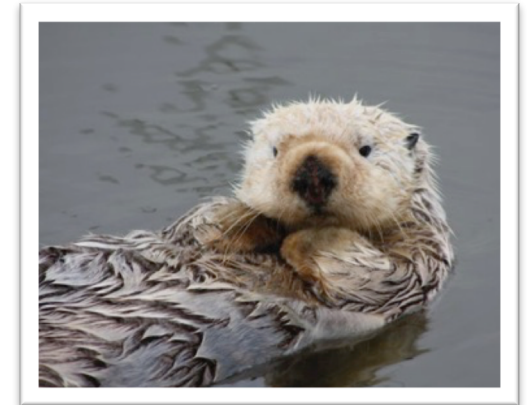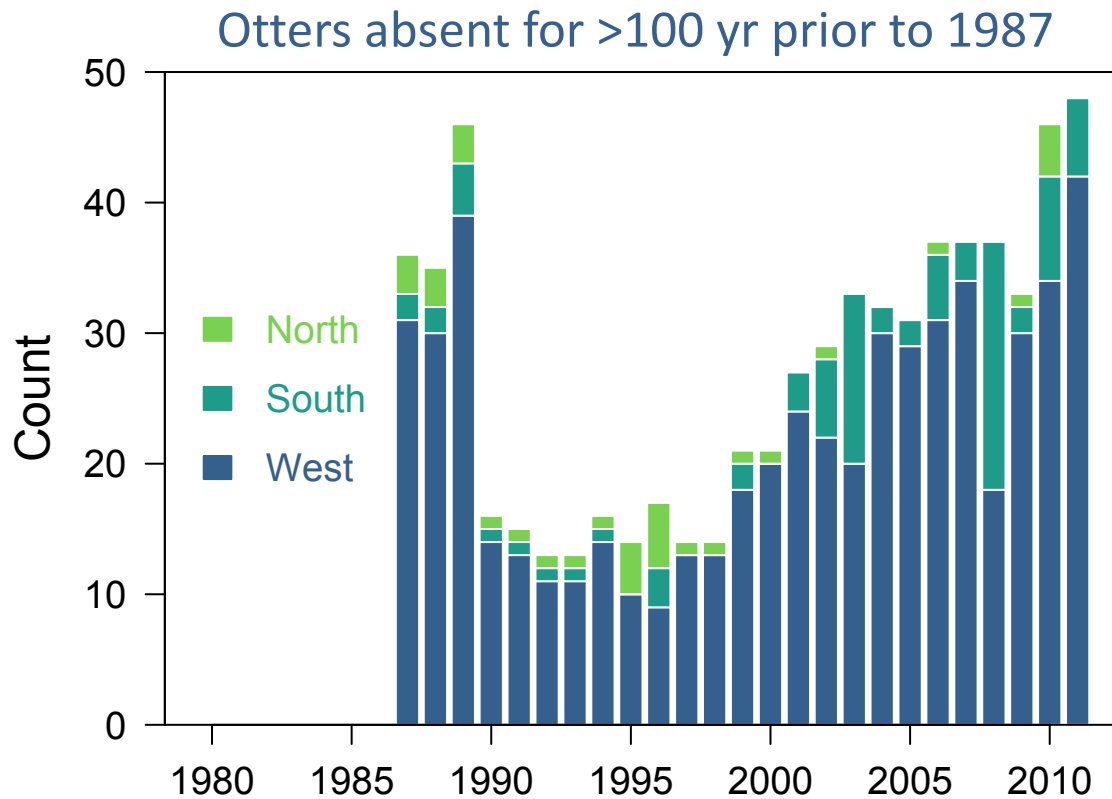
El Niño around San Nicolas

Niño 3.4 SST anomalies

Image courtesy of NOAA
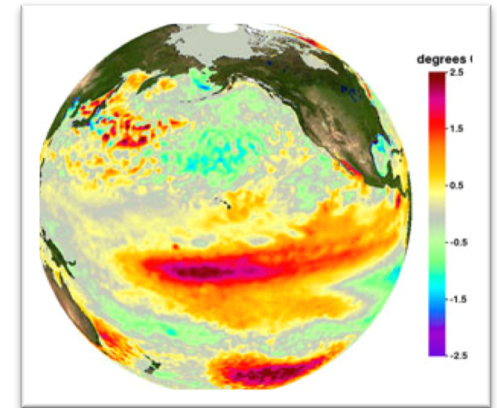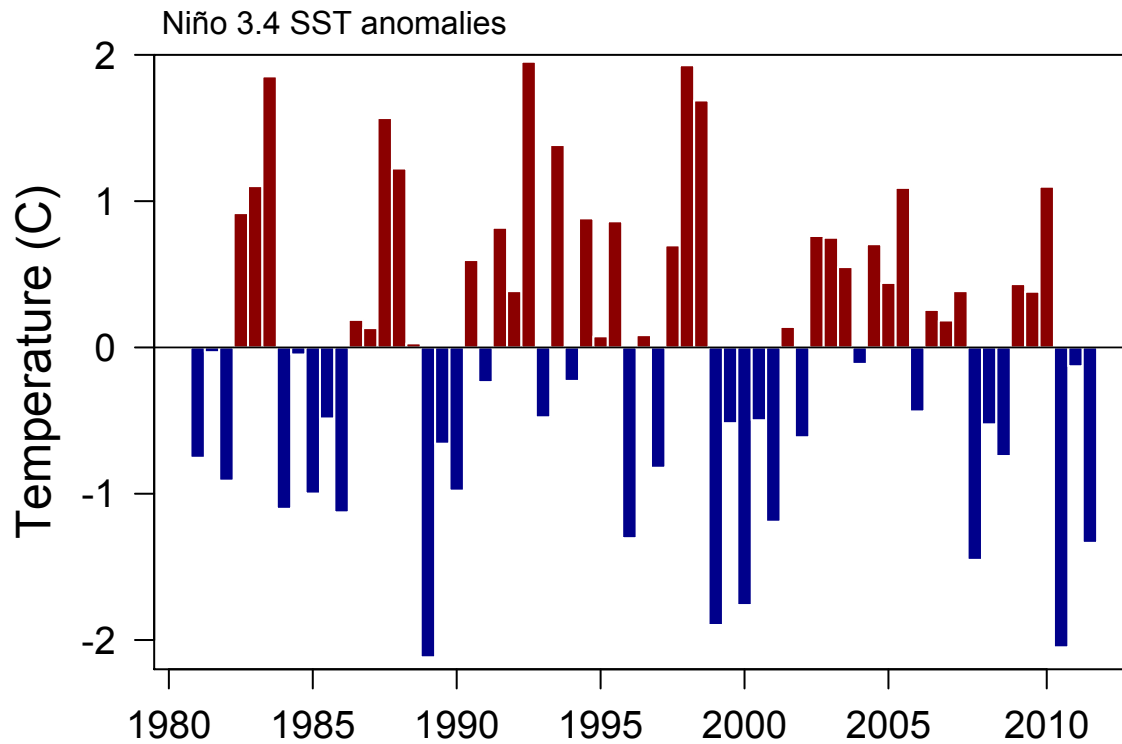
Urchin harvest at San Nicolas

Red urchin landings

Data from Calif Dept Fish Wildlife
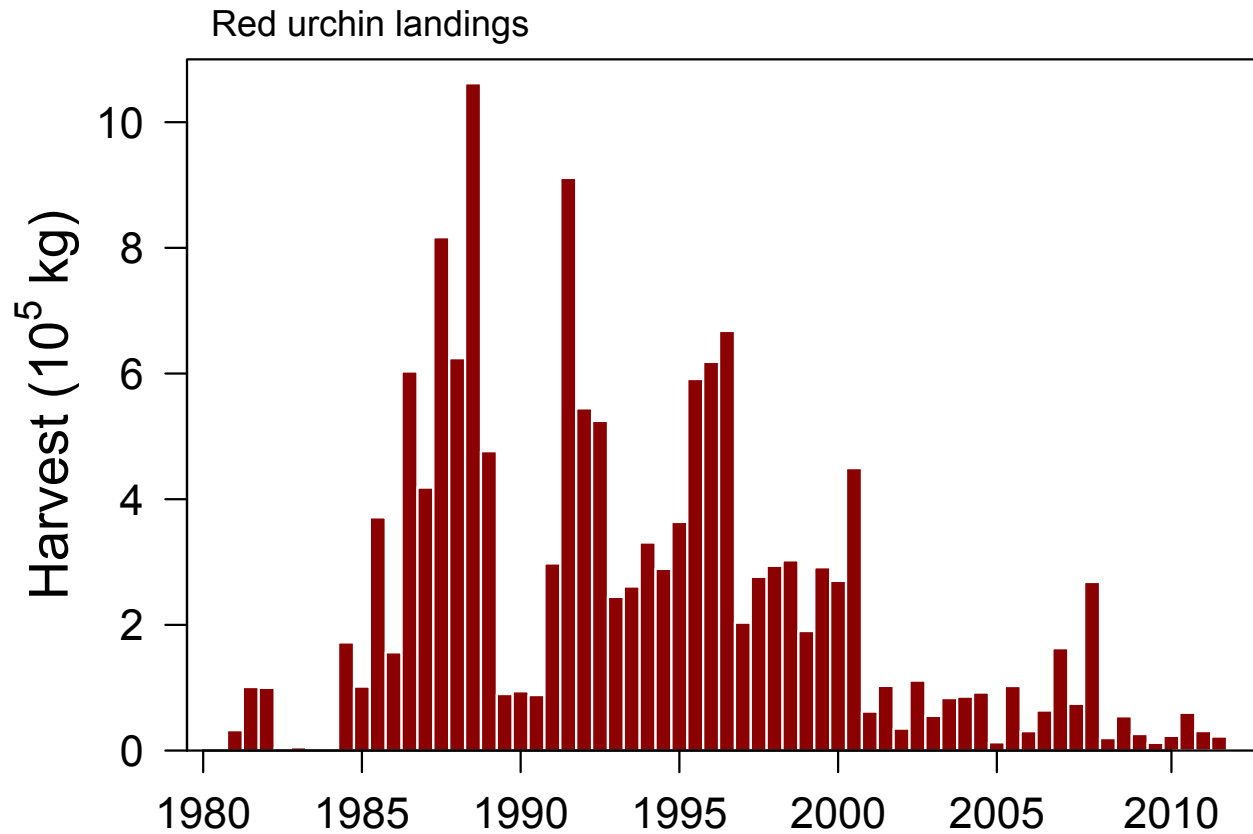
Fitting the model

- 6 observations of 4 states 1980 to 2011

- Each state has same $\mathbf{B}_t$

- MCMC via JAGS to allow observation errors with negative binomial error model.

B matrix

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & \cdots & \vdots \\ b_{31} & \vdots & b_{33} & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{p1} & \cdots & \cdots & \cdots & b_{pp} \end{bmatrix}$$

Inter-specific effects

Intra-specific effects

Changes in density & density-dependence
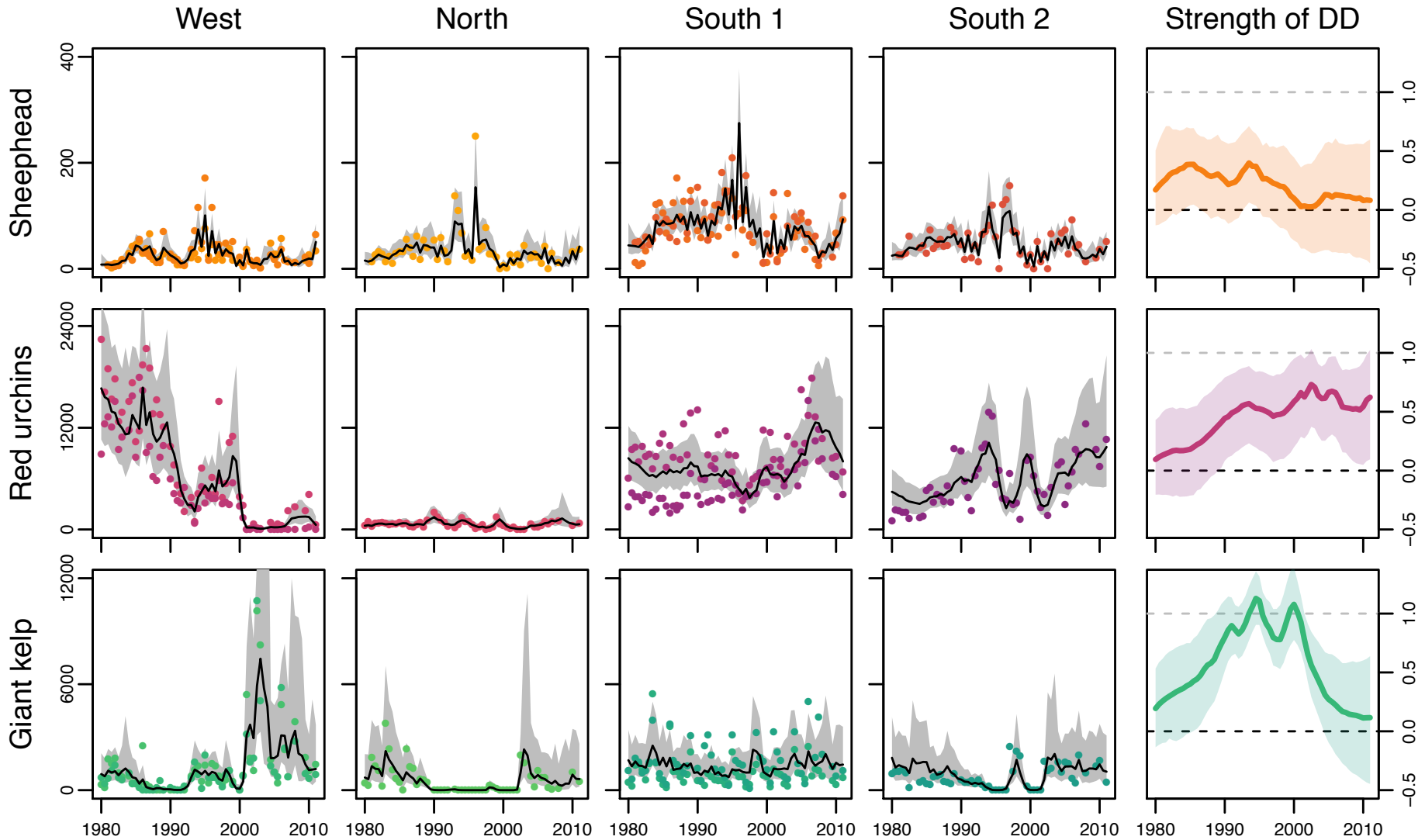
Examples of trophic effects

Top-down

Bottom-up

Red urchins on giant kelp
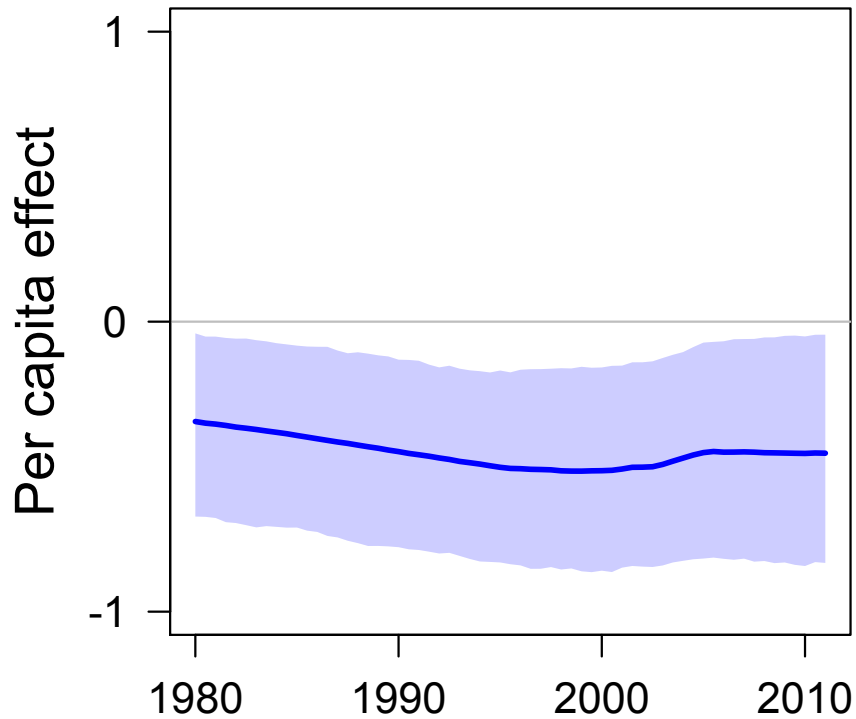
Inverts on invert-eating fish
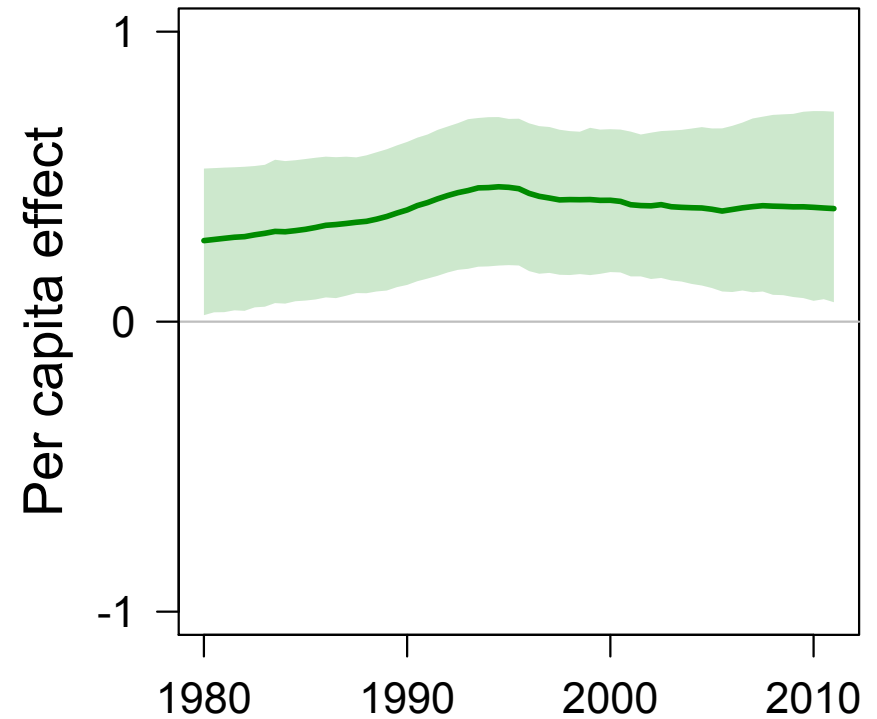
Stability

Stationary distribution for MAR(1)

Mean vector: $\mathbf{m}_\infty = \left(\mathbf{I} - \mathbf{B}\right)^{-1} \mathbf{u}$

Covariance matrix: $\mathrm{Vec}\left(\mathbf{V}_\infty\right) = \left(\mathbf{I} - \mathbf{B} \otimes \mathbf{B}\right)^{-1} \mathrm{Vec}\left(\mathbf{Q}\right)$

(analogous to the univariate case)

The theory in this part of the lecture draws from

Stability properties of MAR(1) models

- Our interest is in stable systems
  (i.e. all eigenvalues of $\mathbf{B}$ lie within unit circle)

- Stability can be measured in several ways, but here are three that we will use:

Stability measure	More stable when...
Variance	variance of the stationary distribution is low relative to that for the process error
Return rate	rapid approach to the stationary distribution (i.e. high return rate)
Reactivity	fewer departures from the mean of the stationary distribution (i.e. low reactivity)

Stability properties of MAR(1) models

Return rates (ie, resilience)

- In stochastic models, equilibrium is the stationary distribution
- Rate of return to the stochastic equilibrium is the rate at which the transition distribution converges to the stationary distribution from an initial, known value
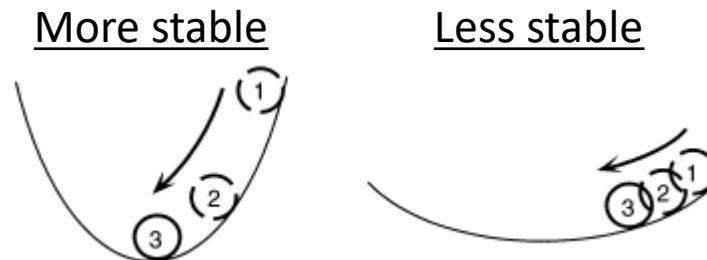- The more rapid the convergence, the more stable the system

More stable   Less stable

Fig. 2B - Ives et al. (2003)

Stability properties of MAR(1) models

Return rates

- Rate of return increases as $|b|$ gets smaller

- Rate of return to mean governed by dominant eigenvalue of $\mathbf{B} \equiv \boxed{\max(\lambda_{\mathbf{B}})}$

- Rate of return of covariance defined by dominant eigenvalue of $\mathbf{B} \otimes \mathbf{B} \equiv \max(\lambda_{\mathbf{B} \otimes \mathbf{B}})$

Take home msg: you can write the return rate as a function of B. So if we can estimate B, we can say something about the stability of the system as measure by return rate.
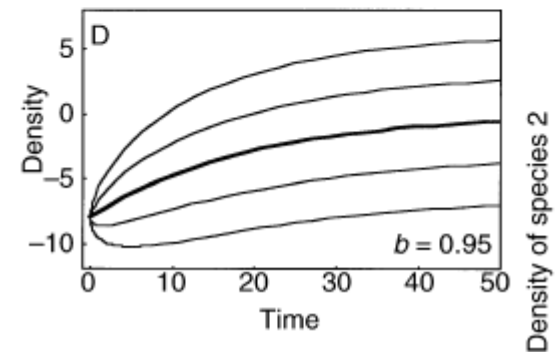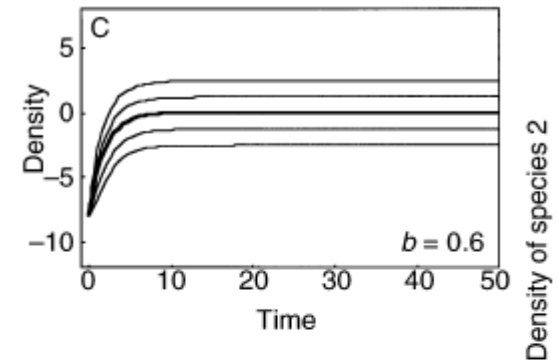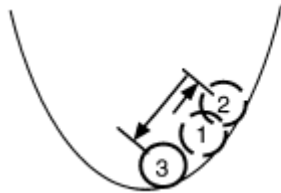

Fig. 3 - Ives et al. (2003)

Stability properties of MAR(1) models

Reactivity

A "reactive" system moves away from a stable equilibrium following a perturbation, even though the system will eventually return to equilibrium.

Low reactivity   High reactivity

Fig. 2C - Ives et al. (2003)

High reactivity occurs when species interactions greatly amplify the environmental forcing

Not how fast does it return but how far does it go?

Stability properties of MAR(1) models

Reactivity

Squared Euclidian distance for $\mathbf{X}_{t-1}$ from μ_∞

$$= \left\| \mathbf{X}_{t-1} - \mu_\infty \right\|^2$$

Squared Euclidian distance for $\mathbf{X}_t$ from μ_∞

$$= \left\| \mathbf{u} + \mathbf{B}\mathbf{X}_{t-1} - \mu_\infty \right\|^2 = \left\| \mathbf{B}\left(\mathbf{X}_{t-1} - \mu_\infty \right) \right\|^2$$

Change in squared Euclidian distance

$$= \left\| \mathbf{B}\left(\mathbf{X}_{t-1} - \mu_\infty \right) \right\|^2 - \left\| \mathbf{X}_{t-1} - \mu_\infty \right\|^2$$

Expectation of this difference

$$= -\frac{\mathrm{tr}[\mathbf{Q}]}{\mathrm{tr}[\mathbf{V}_\infty]} \leq \boxed{\max\left(\lambda_{\mathbf{B'B}} \right)} - 1$$

"worst-case" scenario

Fig. 3 - Ives et al. (2003)

Another form of reactivity is given in Neubert et al. (2009)
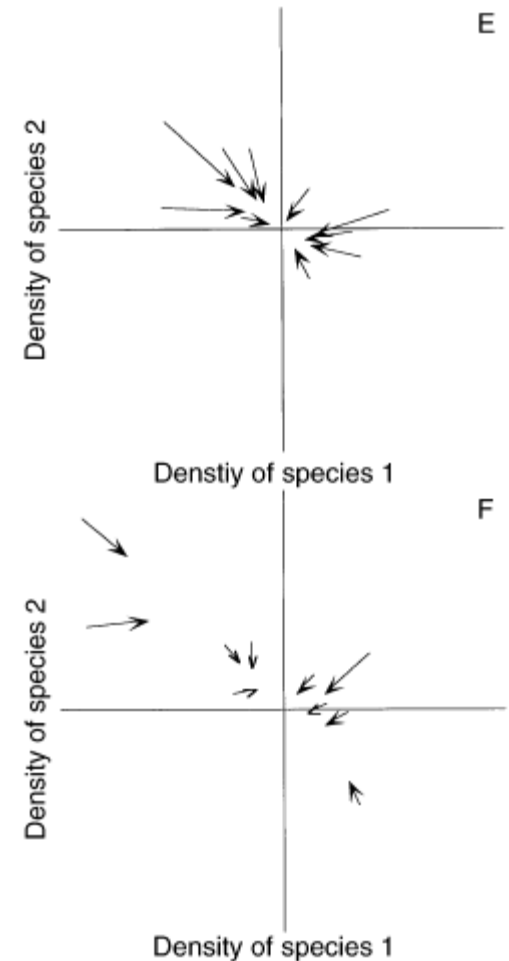
Stability properties of MAR(1) models

Variance comparison

- Generally interested in the time spent away from equilibrium
- More time → greater variance

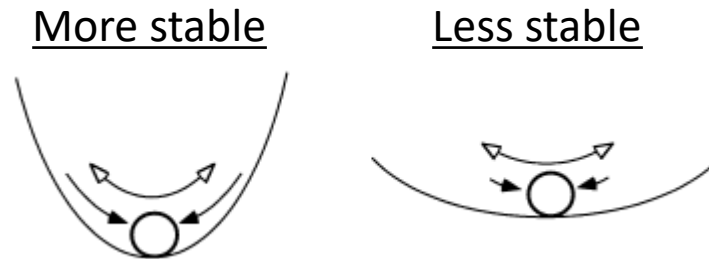More stable   Less stable

Fig. 2A - Ives et al. (2003)

Combines properties of both return rate and reactivity

Stability properties of MAR(1) models

Variance comparison

1) $\widetilde{\upsilon}_\infty = \dfrac{\psi^2}{\left(1-\lambda^2\right)}$ Smaller λ → smaller υ → more stable

2) $\det\left(\mathbf{V}_\infty - \mathbf{Q}\right) = \det\left(\mathbf{V}_\infty\right)\det\left(\mathbf{B}\right)^2$

$$\det\left(\mathbf{B}\right)^2 = \frac{\det\left(\mathbf{V}_\infty - \mathbf{Q}\right)}{\det\left(\mathbf{V}_\infty\right)} = \left(\lambda_1\lambda_2\lambda_3\cdots\lambda_p\right)^2$$

To standardize across various systems

$$\det(\mathbf{B})^{2/p} = \left(\lambda_1\lambda_2\lambda_3\cdots\lambda_p\right)^{2/p}$$

Smaller λ → $\boxed{\text{smaller } \det(\mathbf{B})^{2/p}}$ → more stable
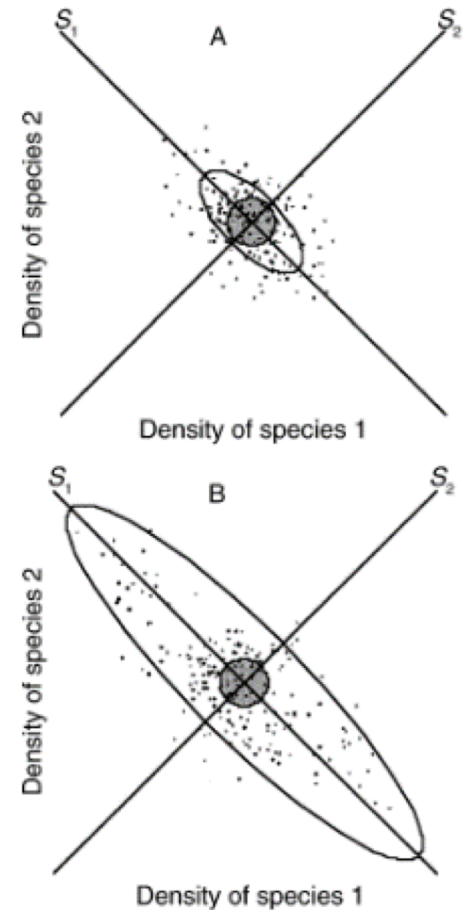
Fig. 3 - Ives et al. (2003)
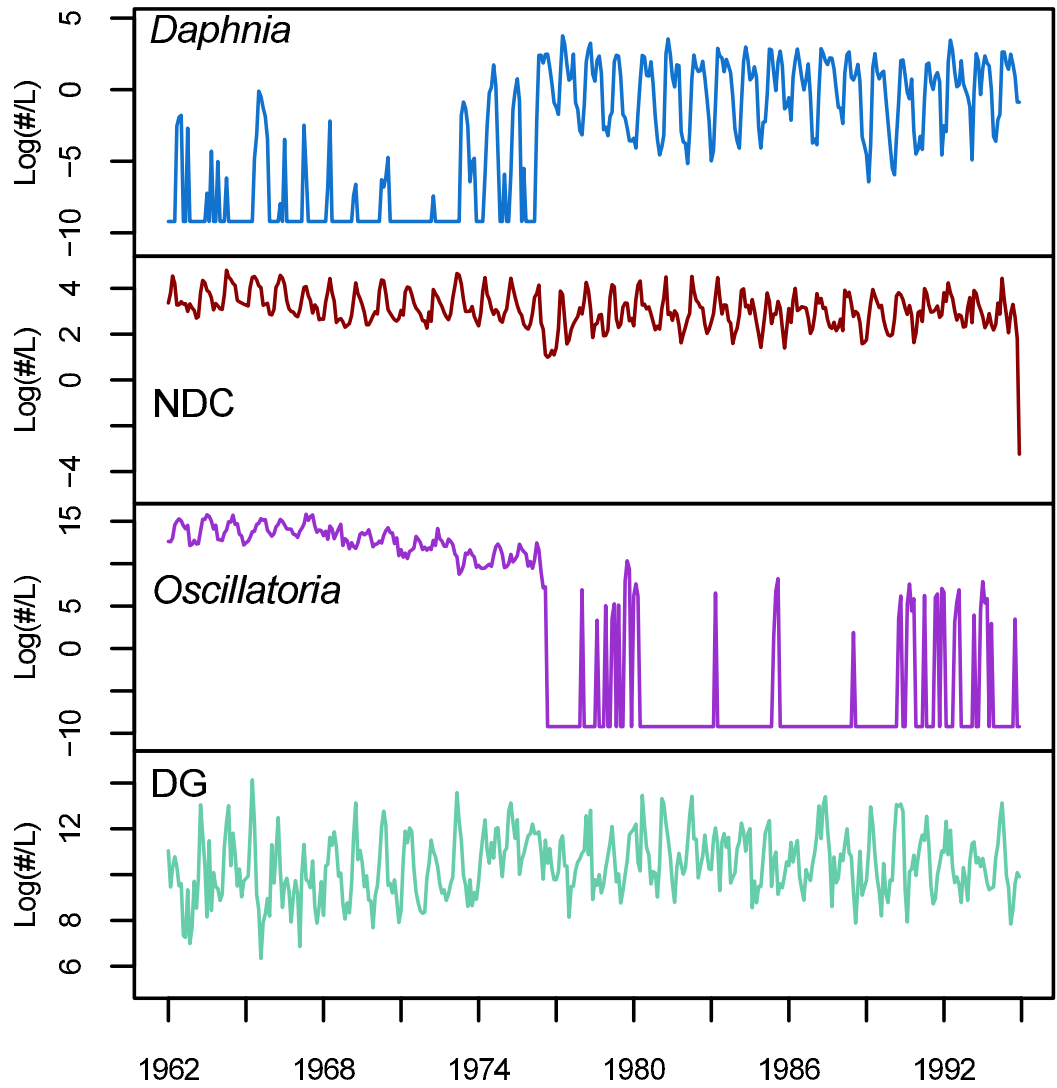
Estimating changes in stability

Francis TB, Wolkovich EM, Scheuerell MD, Katz SL, Holmes EE, et al. (2014) Shifting Regimes and Changing Interactions in the Lake Washington, U.S.A., Plankton Community from 1962–1994. PLOS ONE 9(10): e110363.
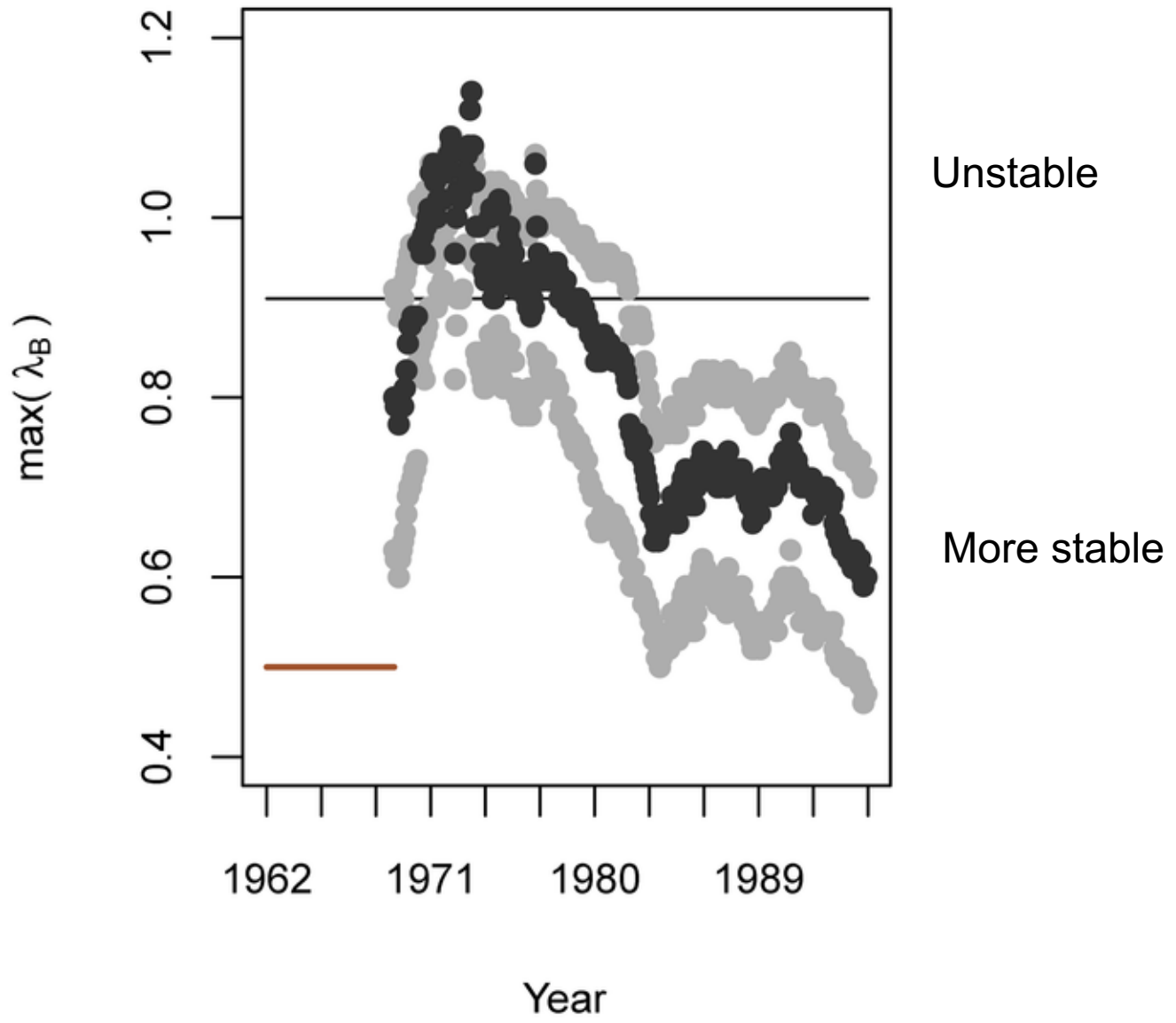https://doi.org/10.1371/journal.pone.0110363
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0110363

Monthly mean densities of Daphnia, non-Daphnia, Oscillatoria, and diatoms and green algae. Francis et al. 2014.
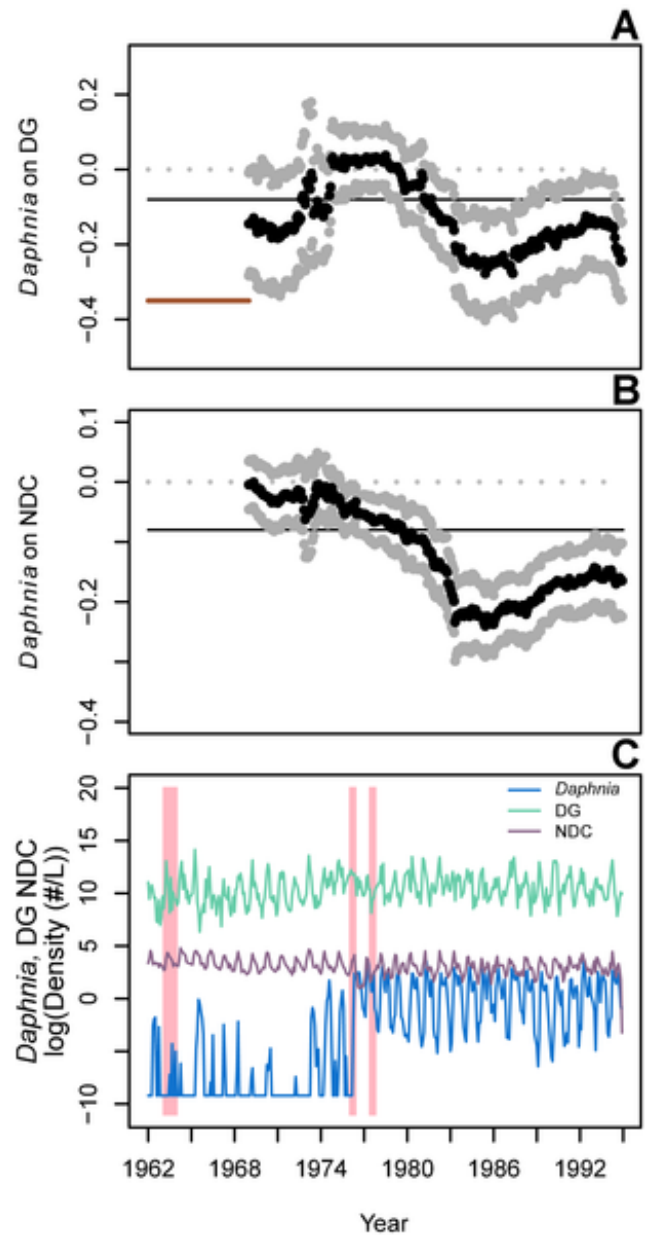
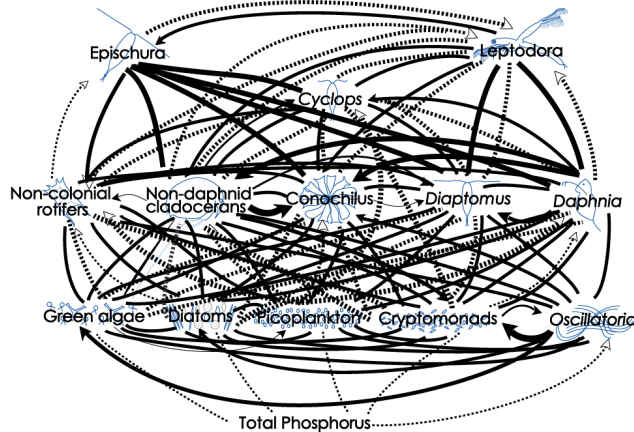Shifting effects of Daphnia on other species. Francis et al. 2014.

Unstable

More stable

Shifting stability in the community. Francis et al. 2014.
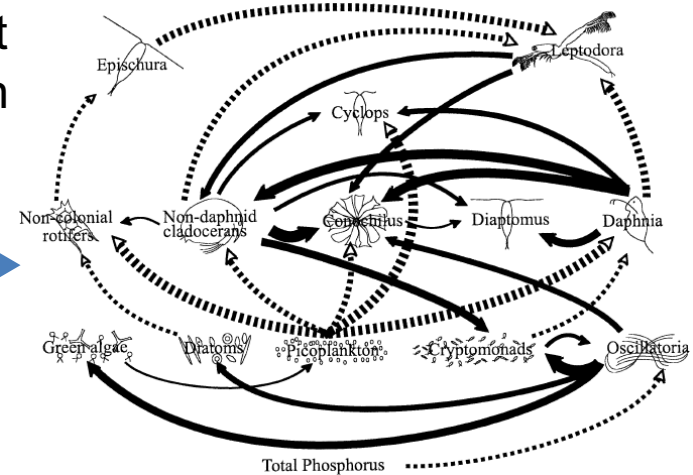
Finding the important structure

Goal is to find a simpler system to explain the observed species changes

13 x 13 B matrix

Compare model support of models with different B's set to zero

156 non-diagonal B terms. There are approximately 1e+47 (that's 1 followed by 47 zeros) unique B matrices with different 0s.

1) We cannot do a brute force fitting of all models.
2) We should try to set some of those Bs to 0 via prior knowledge about the system
3) We need a fast estimation algorithm

Search strategies

$$\mathbf{B} = \begin{bmatrix} b_{11} & 0 & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & 0 \\ b_{31} & 0 & b_{33} & b_{34} \\ 0 & b_{42} & 0 & b_{44} \end{bmatrix}$$

Search strategy for large model spaces

Best of 100 random search

1) 100 models are generated by randomly including (50/50) each b element.

2) Best model (lowest AIC or BIC) is saved

3) Repeat 100 times to get a set of 100 "best" model.

4) Remove any variables that appears in less than 15 of the 100 best models.

5) Repeat the 'Best of 100' search (steps 1-3) with the new smaller set of variables.

6) Repeat steps 1-5 until number of variables in the 100 best is stable and chose best model.

Search strategy for large model spaces

Exhaustive via "leap and bound" algorithm

1) Algorithms to find best model over all possible variable combinations without actually doing exhaustive search.

Forward step-wise

1) Start with no off-diagonal b.

2) Add the b that most improves (reduces) model AIC

3) Stop at some step-AIC threshold

Backward step-wise

1) Same as forward but start with full model (all b's)

Model diagnostics

Once selected, a MAR(1) model should be scrutinized

Useful diagnostics include:

1) Are residuals temporally autocorrelated (via ACF)?

2) Are mean or variance of the residuals correlated with any variates or covariates (via X-Y plots)?

3) Are residuals normally distributed (via normal probability plots)?

4) What proportion of the variance is explained by the model (via conditional R^2)?
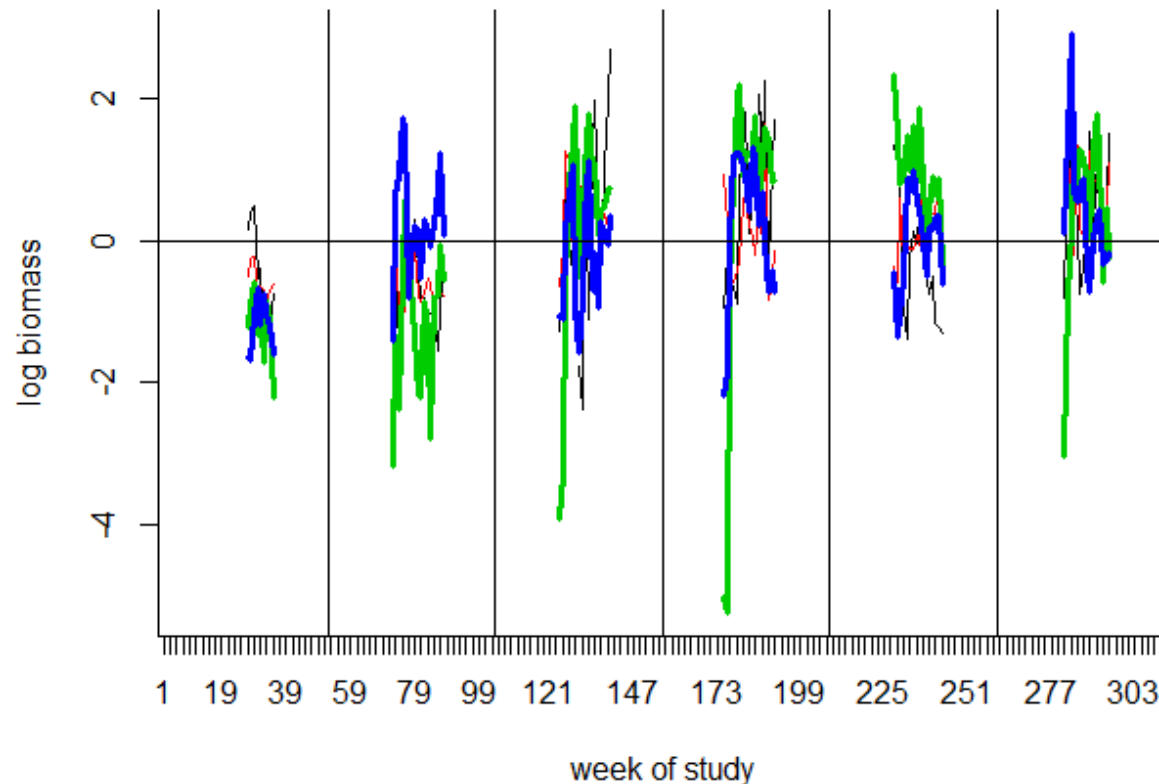
Note: conditional R^2 measures the proportion of change in log density from time t-1 to t

Model diagnostics example

$$\mathbf{x}_t = \mathbf{u} + \mathbf{B}\mathbf{x}_{t-1} + \mathbf{C}\mathbf{c}_t + \mathbf{w}_t$$

Example of demeaned plankton data, 4 spp (Ives et al 2003)
- we are fitting one B matrix (and u)
- so level changes need to be explained by the covariates

Chapter 14: HWS18a Ives et al. analysis

- Fit a MAR community model to freshwater plankton data from Ives et al. (2003)

- Add covariates to the above model

- Add an additional covariate (fish) observed with error

- Compute stability metrics

- See MAR1 package on CRAN to apply the Ives et al. 2003 and other search algorithms to data sets.